Toward Precision Medicine in Intensive Care: Leveraging Electronic Health Records and Patient Similarity

by

Anis Sharafoddini

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Public Health and Health Systems

Waterloo, Ontario, Canada, 2019

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Frank Rudzicz Professor, Department of Computer Science, University of Toronto
Supervisors	Joel A. Dubin Professor, Department of Statistics and Actuarial Science; and School of Public Health and Health Systems, University of Waterloo Joon Lee Professor, Department of Community Health Sciences; and Department of Cardiac Sciences, Cumming School of Medicine, University of Calgary
Internal Members	Plinio Morita Professor, School of Public Health and Health Systems, University of Waterloo David Maslove Professor, Department of Critical Care Medicine, Queen's University
Internal-external Member	Yeying Zhu Department of Statistics and Actuarial Science, University of Waterloo

AUTHOR'S DECLARATION

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis.

This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This thesis consists in part of two manuscripts that have been published.

Exceptions to sole authorship:

Chapter 2: A. Sharafoddini, J. A. Dubin, D. M. Maslove, and J. Lee, "A New Insight Into Missing Data in Intensive Care Unit Patient Profiles: Observational Study," JMIR Med Inform, vol. 7, no. 1, p. e11605, Jan 8 2019.

Chapter 3: A. Sharafoddini, J. A. Dubin, and J. Lee, "Finding Similar Patient Subpopulations in the ICU Using Laboratory Test Ordering Patterns," Proceedings of the 7th International Conference on Bioinformatics and Biomedical Science, 2018.

As lead author of these two chapters, I was responsible for defining the research questions, conceptualizing the study design, carrying out data extraction and analysis, and drafting and submitting the manuscripts. My co-authors provided guidance during each step of the research and gave feedback on draft manuscripts.

Abstract

The growing adoption of Electronic Health Record (EHR) systems has resulted in an unprecedented amount of data. This availability of data has also opened up the opportunity to utilize EHRs for providing more customized care for each patient by considering individual variability, which is the goal of precision medicine. In this context, patient similarity (PS) analytics have been introduced to facilitate data analysis through investigating the similarities in patients' data, and, ultimately, to help improve the healthcare system.

This dissertation is presented in six chapters and focuses on employing PS analytics in data-rich intensive care units. Chapter 1 provides a review of the literature and summarizes studies describing approaches for predicting patients' future health status based on EHR and PS. Chapter 2 demonstrates the informativeness of missing data in patient profiles and introduces missing data indicators to use this information in mortality prediction. The results demonstrate that including indicators with observed measurements in a set of well-known prediction models (logistic regression, decision tree, and random forest) can improve the predictive accuracy.

Chapter 3 builds upon the previous results and utilizes these missing indicators to reveal patient subpopulations based on their similarity in laboratory test ordering being used for them. In this chapter, the Density-based Spatial Clustering of Applications with Noise method, was employed to group the patients into clusters using the indicators generated in the previous study. Results confirmed that missing indicators capture the laboratory-test-

ordering patterns that are informative and can be used to identify similar patient subpopulations.

Chapter 4 investigates the performance of a multifaceted PS metric constructed by utilizing appropriate similarity metrics for specific clinical variables (e.g. vital signs, ICD-9, etc.). The proposed PS metric was evaluated in a 30-day post-discharge mortality prediction problem. Results demonstrate that PS-based prediction models with the new PS metric outperformed population-based prediction models. Moreover, the multifaceted PS metric significantly outperformed cosine and Euclidean PS metric in k-nearest neighbors setting.

Chapter 5 takes the previous results into consideration and looks for potential subpopulations among septic patients. Sepsis is one of the most common causes of death in Canada. The focus of this chapter is on longitudinal EHR data which are a collection of observations of measurements made chronologically for each patient. This chapter employs Functional Principal Component Analysis to derive the dominant modes of variation in septic patients' EHR's. Results confirm that including temporal data in the analysis can help in identifying subgroups of septic patients.

Finally, Chapter 6 provides a discussion of results from previous chapters. The results indicate the informativeness of missing data and how PS can help in improving the performance of predictive modeling. Moreover, results show that utilizing the temporal information in PS calculation improves patient stratification. Finally, the discussion identifies limitations and directions for future research.

Acknowledgements

First and foremost, I would like to thank God Almighty for giving me the strength, ability and opportunity to undertake this research. Without his blessings, this achievement would not have been possible.

I am grateful to my parents, Maryam Salehi Sirjani and Mohammad Sharafoddini, for their love, prayers, caring and incredible sacrifices. I am extremely thankful to my husband, Ehsan Asadi, who has always patiently supported me and helped me overcome the difficulties through this journey. Thank you, Ehsan, for always being there for me. My appreciation extends to my sister Zohreh, my brother Ali, and all my family members in particular Marjan, Morteza, Ghazal, Amir Mohammad, Setayesh, Farkhondeh, Abdolazim, Nahid, Ali, and Ahmad for their support. There are not enough words in the whole world for me to describe how much I am grateful for having the loveliest family.

I also would like to thank my supervisors, Dr. Joel Dubin and Dr. Joon Lee, for guiding me to think and explore, for helping me work on my weaknesses, and for providing me with many opportunities to develop my skills and gain self-confidence. I would like to extend my thanks to Dr. Frank Rudzicz, Dr. Yeying Zhu, Dr. David Maslove and Dr. Plinio Morita for serving on my committee, for their helpful comments and valuable suggestions on improving my thesis.

I am also indebted to Mary McPherson, from Writing and Communication Center, for kindly supporting me through my PhD studies. So truly, thank you for everything. I also

would like to thank Mohaddeseh and Mahya for their friendship and help in the past few years.

My PhD journey finally came to an end, and I wish I could say thank you to all the great people I feel lucky and happy for having them in my life.

Dedication

This thesis is dedicated to my mother who took the lead to heaven before the completion of this work.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The growing amount of Electronic Health Record (EHR) data has provided the opportunity to utilize it for improving healthcare systems and patient outcomes. A recent initiative for fulfilling these objectives is precision medicine, the goal of which is to provide treatment and prevention strategies that take individual variability into consideration (Collins et al. 2015). Patient Similarity (PS) analytics, a novel approach to precision medicine, focuses on investigating the similarities in patients' data so as to provide more customized care (Parimbelli et al. 2018). This dissertation is concerned with the application of PS analytics among Intensive Care Unit (ICU) patients. It first demonstrates the informativeness of missing data among the heterogeneous populations of ICU patients particularly with respect to predictive modeling. It then utilizes the proposed data representation to identify patient subpopulations based on the similarities in laboratory tests ordered for each patient. After that, it shifts its focus to defining a multifaceted PS metric in which similarity scores from various metrics—specific to the variable under examination—are aggregated to build the final score. Finally, this dissertation takes the findings from previous sections into consideration and looks for potential subpopulations among septic patients, with a special focus on the longitudinal data collected for this patient cohort.

The purpose of this chapter is three-fold: first it presents an overview of the background concepts used throughout the dissertation, and then lays out the dissertation's organization. The background section begins with a discussion of the availability of EHR data and their potential secondary uses, followed by their challenges in practice. Then, the discussion is narrowed down to the application of EHR data in health analytics in population-based and personalized applications. With personalized health analytics in mind, PS analytics and its promise are introduced. Then, the focus is brought to ICUs, where PS analytics can play an important role. After establishing the background for this thesis, a comprehensive review of the literature on PS analytics is provided, which forms the core of this thesis. Finally, the rationale, arching objectives, and the outline for the rest of this thesis are discussed.

1.1 Background

1.1.1 Electronic Health Records and their Secondary Uses

Along with strategies that have been employed to develop and adopt health information technology and EHRs (Coiera 2009; Morrison et al. 2011), billions of dollars of investments in this field, such as the nearly 2 billion dollar funding allocated by the Health Information Technology for Economic and Clinical Health (HITECH) Act in the United States (Blumenthal 2010), have accelerated the adoption of EHR systems and resulted in an unprecedented amount of data. According to the International Organization for Standardization (2005), an EHR is a "repository of information regarding the health status of a subject of care, in computer processable form, stored and transmitted securely and

accessible by multiple authorized users, having a standardized or commonly agreed logical information model that is independent of EHR systems and whose primary purpose is the support of continuing, efficient and quality integrated health care." This availability of data has also opened up the opportunity to utilize EHR for secondary uses like assessing the efficiency of health care systems, research for expanding knowledge about diseases and treatments, improving public health and health services, etc. According to the American Medical Informatics Association (Safran et al. 2007), "Secondary use of health data can enhance health care experiences for individuals, expand knowledge about disease and appropriate treatments, strengthen understanding about effectiveness and efficiency of health care systems, support public health and security goals, and aid businesses in meeting customers' needs." Studies have also demonstrated that EHR can help in improving the quality of care (Kern et al. 2013; Cebul et al. 2011; Campanella et al. 2016) and reducing medical errors (Ammenwerth et al. 2008; Devine et al. 2010; Campanella et al. 2016).

1.1.2 EHR Data Challenges

EHR data consist of structured/coded and unstructured data. Each category also includes various data types for different purposes (Hayrinen et al. 2008). This wide range of data types highlights the challenges in EHR secondary uses (Jensen et al. 2012). Moreover, since EHR data have not been collected particularly for research purposes, the task of mining these data is not trivial due to the following challenges.

Incomplete data: EHR data comprise a variety of incomplete information and measurements. Since most standard analytical methods require complete data, missing data can be problematic.

Volume and dimensionality: EHRs include enormous amounts of administrative data, ancillary clinical data and clinical text reported as numerous variables (Jensen et al. 2012). The amount of the data gets even bigger with continuous-monitoring systems. Mining this amount of data can be complex and computationally challenging.

Data complexity: EHR data consist of information about thousands of variables and their underlying relations; therefore, EHR data are highly complex.

Data quality: Administrative data often refers to billing codes, which will then be used for financial reimbursement and sometimes suffer from biases since the amount of reimbursement depends on the assigned codes (Jensen et al. 2012). For instance, it has been observed that using a diagnosis-related group patient classification system for payment system increases secondary diagnoses (Serden et al. 2003). Ancillary clinical data also suffer from errors in measurement, collection and data entry processes (Koppel 2009). Clinical text has its own sources of imprecision, including lack of normal grammar, and is rich in spelling and typing errors (Meystre et al. 2008).

Temporal data: The temporal nature of EHR data holds the promise to provide more-detailed data (Singh 2015). However, longitudinal EHR data mining is still in its early stages,

and exploring patient trajectories can be complex and computationally demanding (Jensen et al. 2012).

Irregular sampling: EHR data are only recorded when a patient visits a hospital or during a healthcare period, which causes varying time intervals between measurements. When using non-temporal methods to analyse EHR data, this challenge is well-handled by aggregating data over the time window of interest (Singh et al. 2015). However, longitudinal analysis of the data can be challenging and requires accounting for dependency of repeated measures within the same individual.

1.1.3 Precision and Population-based Analytics

With the emergence of the ever-growing amount of EHR data, it is becoming much more demanding for clinicians to examine that data in depth and derive actionable insights from the overlapping biomedical structures and body-system interactions. Therefore, in recent decades, health analytics has been widely utilizing EHR data in the following applications to achieve the goals of secondary uses of EHR data.

Predictive modeling: In the past, medicine was largely a reactive field—a disease is treated, when it has been diagnosed (Miner et al. 2014). However, a move toward proactive medicine has been initiated (Hood et al. 2009). One particular pathway to proactive medicine employs predictive analytics, for instance, to attempt to accurately derive insights from available health data to predict disease progression and provide recommendations to optimize patient outcomes. Predictive modeling utilizes statistical and machine learning methods to

learn from clinical data and subsequently to predict patients' future health outcomes. Among the areas for prediction are graft survival in heart-lung transplants (Oztekin et al. 2009), cancer survival (Delen et al. 2005; Zolbanin et al. 2015; Gupta et al. 2014) and future hospital admissions (Peck et al. 2013; Li et al. 2009).

Patient stratification: Many clinical research studies and clinical trials largely rely on the identification of a homogeneous study population (Jensen et al. 2012). Moreover, patient stratification has been widely used to study risk factors, outcomes and prognoses within population groups to uncover underlying attributes (Hu et al. 2016a; Bose et al. 2018). Patient stratification analytics employs clustering methods to group patients based on their characteristics and similarities (Jensen et al. 2012).

Care pathway exploration: After disease diagnosis, clinicians come up with a care plan—which is a sequence of medical interventions known as a care pathway— using their knowledge and the available evidence to control or improve the patient's health status (Hu et al. 2016a). Although effective care plans hold the promise to provide best-care scenarios, developing and optimizing them is a challenging task. Care pathway analytics has been used to identify the most desirable and effective care plan by deriving and exploring care pathways and their associated outcomes, based on EHR data. This approach has been used to analyze pediatric asthma care (Basole et al. 2015), congestive heart failure care plans (Perer et al. 2013; Gotz et al. 2012a), and vasopressor intervention (Wu et al. 2017).

In general, there are two major streams in the aforementioned health analytics: population-based analytics and precision medicine analytics.

1.1.3.1 Population-Based Analytics

Many evidence-based studies have been done on large populations to answer a wide range of health-related questions, including the second version of the Acute Physiology and Chronic Health Evaluation (APACHE-II) (Knaus et al. 1985), the Framingham heart failure risk assessment (McKee et al. 1971), and a study done in southern England to devise a diabetes risk score for predicting undetected type 2 diabetes (Griffin et al. 2000). These studies provide statistically rigorous results for an average patient. However, they are also relatively expensive, time-consuming and prone to population selection bias (Miner et al. 2014). For instance, most of these studies are mainly concerned with patients who seek care. Additionally, one of the major challenges in evidence-based medicine is multimorbidity, which limits generalizing a study to many patients (Campbell-Scherer 2010; Gotz et al. 2012b).

Typically, evidence-based medicine utilizes guidelines derived from studies on a large population and provides "the average best choice" (Bellazzi et al. 2011). Therefore, physicians cannot solely rely on the evidence from the mean of the population when facing an individual with special conditions; instead, they must base their recommendations on the characteristics of that particular patient. For instance, many patients have been placed on Statins, even though only one patient among 50 may benefit (Mukherjee et al. 2002), and

certain types of patients may actually be harmed by the drug (Schork 2015; Currie et al. 2006). If medical researchers can sub-divide the general population into more homogeneous sub-groups, they should be able to find similarities among people and tailor their treatments accordingly. In other words, clinicians should be able to take individual characteristics into consideration.

1.1.3.2 Precision Medicine Analytics

Today, *personalized medicine*, which can be defined as "the tailoring of a treatment to an individual based on their unique characteristics" is gaining a lot of attention (Miner et al. 2014). The treatments may be a medication, an exercise, or any other intervention within the field of health. According to the National Research Council, *precision medicine* is a newer term for personalized medicine. The National Institutes of Health defines *precision medicine* as "an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person." In fact, there is a considerable overlap between the terms *precision medicine* and *personalized medicine*. Personalized medicine implies the design of a unique treatment for an individual. In order to avoid the misinterpretation, that treatment must be developed uniquely for an individual. In the US, the President's Council of Advisors on Science and Technology clarified this definition, by stating that precision medicine "does not literally mean the creation of drugs or medical devices that are unique to a patient, but rather the ability to classify individuals into sub-populations that differ in their susceptibility to a particular disease or their response to a

specific treatment" (PCAST 2008). Therefore, precision medicine may lead to non-personalized treatment (Khoury 2016). Another term often used in these discussions is *genomic medicine*; that is, "the use of information from genomes and their derivatives (RNA, proteins, and metabolites) to guide medical decision making." (Ginsburg et al. 2009). Unfortunately, personalized medicine, precision medicine, and genomic medicine have frequently been used interchangeably in the literature. However, the first two terms are broader than genomic medicine (Snyderman 2012). Although genomic medicine has undeniable advances and offers the promise to facilitate the move toward personalized medicine, more time is needed to translate these advances into health benefits and overcome the challenges in the practical implementation of this promise (Conti et al. 2010). One approach toward precision medicine is to consider the clinical similarity of patients, then tailor health analytics based on a cohort of such similar patients to one index patient by utilizing existing EHRs. Therefore, PS analytics can be embedded in health analytics to make personalized predictions. This approach can best manage a real-world patient with a complex health status and comorbidity.

1.1.4 Intensive Care Units and ICU Databases

ICUs provide care to acutely and severely ill patients and were primarily introduced in the 1950s, with a basis in World War II (Rodriguez 2001b) and a poliomyelitis epidemic in Denmark (Reisner-Senelar 2011). Recent studies have shown that these units target diverse critically ill patient populations and that close monitoring of these patients has generated an

enormous amount of data (Johnson et al. 2016a). Although ICUs have a higher number of staff in comparison to other departments (Johnson et al. 2016a), analysis and interpretation of this amount of data is challenging for clinicians and must be handled by data analysis methods. Therefore, ICU clinicians have adopted EHR systems for collecting and storing data (Ghassemi et al. 2015), resulting in ICU databases such as the APACHE Outcomes database (Celi et al. 2013), the Philips eICU database (Pollard et al. 2018) and the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) database—recently, renamed the Medical Information Mart for Intensive Care (Saeed et al. 2011; Johnson et al. 2016b).

In this thesis, the freely accessible MIMIC-III database (Johnson et al. 2016b)—an update of the MIMIC-II database—is used. This database, which was released in 2015, has detailed information on 53,423 distinct hospital admissions to the critical care units at the Beth Israel Deaconess Medical Center in Boston, between 2001 and 2012 (Johnson et al. 2016b). The records in this database contain data from various sources, from temporal physiological measurements to free-text hospital discharge summaries. The out-of-hospital death dates were also collected from the Social Security Administration Death Master File. The data in this database were de-identified according to the Health Insurance Portability and Accountability Act (HIPPA) standards. MIMIC-III is a relational database and has 26 tables that are linked by various identifiers. In this database, the top primary ninth version of the International Classification of Disease (ICD-9) discharge codes are: 414.01 (“Coronary

atherosclerosis of native coronary artery", 0.39.9 (Unspecified septicemia') and 419.071 ("Subendocardial infarction, initial episode of care"), accounting for 7.1%, 4.2% and 3.6% of all hospital admissions, respectively. Overall, there are three classes of tables in this database: (i) tables that track patient stays in the hospital, e.g. ADMISSION; (ii) dictionary tables, which are look-up tables that map codes to their definitions; (iii) tables that store information about patient care, such as clinical measurements and billing information.

1.2 Patient Similarity Literature Review

In recent years, the amount of literature on PS analytics has grown enormously. Three recent review studies (Parimbelli et al. 2018; Sharafoddini et al. 2017; Brown 2016) point out PS analytics' position as a core topic in precision medicine, stress its potential to fulfill its promise, and note its likelihood of remaining a trending topic in this area. Application of PS analytics has not been limited to a particular medical problem; various approaches have been taken on different EHR databases to improve medication targeting, subgroup discovery, and patient outcome prediction. However, some areas such as treatment targeting, missing patient data treatment, and longitudinal EHR data analysis have received only limited attention from the research community. Therefore, future research efforts in these areas are required to accelerate the impact of PS analytics on healthcare. In the next two subsections, a more in-depth review of the literature in terms of the preprocessing techniques (particularly predictors extracting and missing data treatments) and PS modelling is provided to identify potential gaps in this research area.

1.2.1 Predictor Extraction and Missing Data Treatment in PS Analytics Studies

Predictor extraction from raw EHR data has been one unavoidable component of PS analytics research. The variety in raw EHR formats (such as recorded signals, textual reports and laboratory measurement) has resulted in many approaches in predictor extraction. While for cross-sectional variables such as age and gender the actual value has been used, different transformation techniques have been employed to represent longitudinal/time series data. These techniques range from simple summary statistics from longitudinal data within a particular time window (such as minimum, average, maximum) (Henriques et al. 2015; Lee et al. 2015; Panahiazar et al. 2015) to more-complex transformations such as wavelets (Sun et al. 2010b; Sun et al. 2010a; Henriques et al. 2015; Lee et al. 2015a; Panahiazar et al. 2015). For instance, Sun et al. and Saeed et al. (2010b; 2006b), respectively, utilized Daubechies-4 and Harr wavelet transformations for representing patient vital signs, and concluded that wavelet coefficients are better in representing temporal data. Sun et al. (Sun et al. 2010b), derived two sets of predictors: Daubechies-4 wavelet coefficients and statistical coefficients (mean and variance) from two-hour measurements of vital signs. They defined a Mahalanobis distance by solving an optimization problem aiming at minimizing the within-class squared distances and maximizing between-class squared distances. This similarity metric was used to retrieve the three most-similar patients in their *k*-NN Classifier. This study showed that wavelet predictors are better representations of temporal measurements.

Recently, natural language processing techniques for extracting predictors from raw EHR have received attention from researchers in PS analytics (Saeed et al. 2006a; Wang et al. 2012a; Wang et al. 2015). In particular, the Term Frequency-Inverse Document Frequency (TF-IDF) technique has repeatedly been utilized to produce predictors. This method extracts the predictors in three steps: first, a list of the entire observed predictors with a certain value in the whole training data set is prepared and named the bag of predictors. For instance, a diagnosis code and a discrete glucose serum level can both be considered members in the bag of predictors. Then, the ratio of the number of times a predictor appears in a given patient profile, n_p (where p is the predictor), to the total number of predictors in the patient profile, n, will be calculated:

$$TF = \frac{n_p}{n} \tag{1.1}$$

For TF calculation, a normal clinical event such as a heart rate between 60 to 100 beats may be very common in a patient's profile during a stay in the hospital, and consequently has a fairly high TF value. Therefore, IDF is defined as follows to overcome this challenge:

$$IDF = \log_2 \frac{N}{N_p} \tag{1.2}$$

where N is the number of patient profiles in the training set, and N_p is the number of patient profiles that contain the predictor p. Again, a normal clinical observation arguably appears in every patient's profile during his or her stay, and consequently the ratio (N/N_p) would be

close to one, and the IDF value would be close to zero for that particular predictor. Therefore, the IDF value is low for common clinical observations and high for rare observations. Finally, the TF-IDF value is defined as the simple product of TF and IDF as shown below:

$$TF\_IDF = \frac{n_p}{n} \times \log_2 \frac{N}{N_p} \quad (1.3)$$

Applying this technique to derive predictors could help boost the accuracy of similarity assessment for patients with rare conditions; that is, the IDF coding favours patients with rare conditions. Wang et al. (2012a) represented patient profiles using the TF-IDF method and utilized three bags of predictors: diagnosis codes, medication codes and lab tests. Additionally, Saeed et al. (2006a) employed IDF to identify patients with rare conditions.

While feature extraction has been of interest in PS analytics, missing data in EHR, which is one of the most-common challenges in this area to date, has received scant attention in the research literature. Most studies choose to omit patients' data entirely when missingness (some missing records) is a feature, and a few studies simply imputed the missing data with the average value for that particular variable (Sharafoddini et al. 2017). Sun et al. (2010a) took this challenge more seriously and used linear regression to predict missing values based on the available values. Since missing data in clinical contexts are considerable and inevitable (Weiskopf et al. 2013a; Wells et al. 2013; Little et al. 2012), this lack of evidence on missing data treatment in PS analytics highlights a research gap in this area and indicates

a need to address the missing data challenge so as to establish an appropriate ground for PS analytics.

1.2.2 The PS Modelling Algorithms

1.2.2.1 Neighbourhood-based methods

While various modelling techniques have been used in PS analytics, neighbourhood-based methods have received more attention in recent years. The very first study on these methods goes back to 1998 (Jurisica et al.). Neighbourhood techniques refer to studies in which a group of patients similar to a new patient is retrieved and a prognosis, diagnosis or recommendation is provided by a model trained on the data from those similar patients. This category of techniques is comparable to memory-based techniques in collaborative filtering, in which a new product or a movie is suggested to the customer based on the history of similar consumers (Su et al. 2009). One of the most common methods in this category is k-nearest neighbors (KNN). The overall structure of these techniques is demonstrated in Figure 1-1.

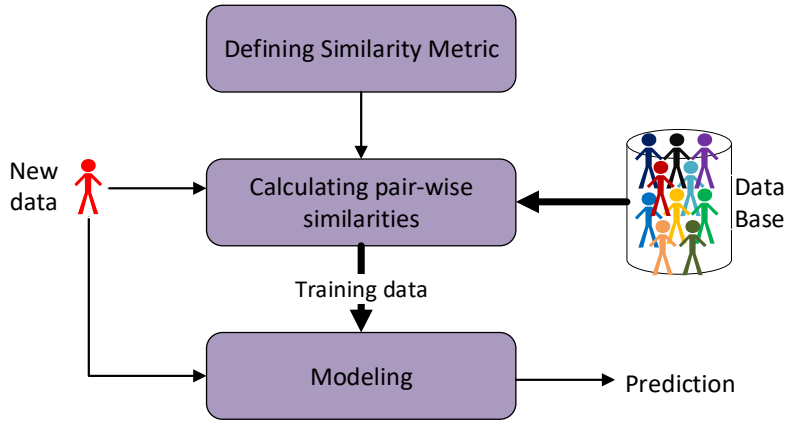

Figure 1-1 An overview of neighbourhood-based PS analytics.

As can be seen in Figure 1-1, the similarity metric is one of the main components in these methods. Various types of similarity metrics including absolute distance (Chattopadhyay et al. 2008), Euclidean distance and its family (Bobrowski 2006; Henriques et al. 2015; David et al. 2011; Park et al. 2006; Hielscher et al. 2014), Mahalanobis distance and its family (Sun et al. 2010b; David et al. 2011; Wang et al. 2012a; Wang et al. 2015; Lowsky et al. 2013; Han et al. 2015; Sun et al. 2010a), correlation-based similarity metric (Saeed et al. 2006a) and cosine similarity metric (Lee et al. 2015a) have been employed in the literature which will be discussed in more detail in this section.

An early example of research into neighbourhood-based PS modeling is the study by Bobrowski (2006) in which a linear data transformation mapped patients to a space in which similar patients are closer to each other, and the distance between different patients is greater. Then, a KNN model with the Euclidean distance was employed on the mapped data. This

two-level PS analytics procedure outperformed the classic KNN. However, the optimal number of similar patients (k) remained unanswered in this study. Around the same time, Park et al. (2006) investigated the optimal number of similar patients for an index patient. The proposed statistical case-based reasoning technique in this study consists of two major steps. First, the distribution of pair-wise distances in the training set was derived, and then an optimum cut-off which defined a distance threshold was found using a grid search. These techniques not only outperformed the one-size fits-all logistic regression (LR), decision tree (DT) and KNN, but also the conventional case-based reasoning technique. David et al. (2011) employed the Euclidean distance metric in their PS analytics. They first assigned random weights to their predictors, then mapped the data to a weighted space. In the next step, the Euclidian distance was used to identify patients similar to an index patient within a particular threshold. Last, they mapped the data to a lower-dimensional space using singular value decomposition (SVD) and examined the discriminative power of the weights. These steps were repeated several times to achieve a set of discriminative weights. Unfortunately, no comparison method was used in this study, and they only reported the level of agreement between their algorithm results and data labeled by a reviewer.

While some studies utilized only a simple distance metric such as absolute difference (Chattopadhyay et al. 2008), other researchers focused on more-complex distance/similarity calculation techniques. Sun et al. (2010b) defined a Mahalanobis distance optimized to minimize within-class distances and maximize between-class distances. The proposed metric

outperformed the conventional Euclidean metric. When they tested the proposed Mahalanobis distance on a lower dimensional space mapped by Principal Component Analysis (PCA) (Sun et al. 2010b) and linear discriminant analysis (Sun et al. 2010a), it outperformed the other models.

Recently, a number of researchers have started working on integrating multiple PS metrics learned separately without the need to share the data that have been used to calculate them. In other words, in these methods each care facility requires to share the output labels from the identified neighborhood for an index patient not the raw data used for similarity calculations. These techniques are very helpful when a number of health facilities want to share insights without sharing their patient data. Wang et al. (2011) used this approach and defined a quadratic optimization problem aiming to minimize in-class scatterness and maximize between-class distances for the weighted sum of the similarity matrices. In another study, they proposed a Mahalanobis PS metric learned from a human expert's ideas (Wang et al. 2012a). In this study, they first calculated the pair-wise Euclidean distances and then applied the expert's idea on these PS scores by using a similarity matrix and dissimilarity matrix. To incorporate expert's knowledge, two matrices are defined based on the expert's opinion: similarity matrix and dissimilarity matrix. The (i, j)-th entry of these matrices will be -1, if patient i and j are similar/dissimilar. Otherwise, the entry is equal to the number of patients that are similar/dissimilar to the patient i based on the expert opinion. If expert knowledge is available, these matrices are considered in the global optimization problem. In the same vein,

Wang et al. (2015) introduced a two-fold objective function for PS learning: a part focused on humans' expert knowledge and the other based on the available EHR data. In this study, the proposed two-fold PS metric outperformed the one based only on patient data. Similarly, Huai et al. (2018) considered the idea of learning PS without directly accessing the EHR data in their uncorrelated PS learning method. In this study, PS learning was formulated as a maximum likelihood problem in which two regularization terms were considered for assuring the selection of the most relevant and uncorrelated features. The proposed PS metric outperformed all competing metrics, including cosine and Euclidean in a KNN setting.

Unlike previous studies, that by Lowsky et al. (2013) demonstrated that in survival probability prediction using a Kaplan-Meier survival curve, neighbourhood-based PS analytics based on the Mahalanobis distance do not show a consistent advantage over the Cox model as the baseline. They also found that more-complex models such as Random Survival Forest (Ishwaran et al. 2008) outperformed the proposed PS analytics in all scenarios.

A number of authors have considered using PS analytics in feature selection and predictive modelling. Hielscher et al. (2014) utilized a two-step PS analytics. They first split the training data set into two data sets based on gender. Then, after performing correlation-based feature selection to identify the most important features for each group, they used KNN and weighted KNN to predict liver fat concentration level for new patients. In this study, weighted KNN outperformed the conventional KNN, and it was found that feature selection,

by reducing dimensionality, helps predictive modelling performance. Moreover, the fact that only a few of the most-predictive predictors within each subgroup were common highlights the efficiency of PS analytics and customized predictors. Han et al. (2015) followed the same approach for diabetes. They first retrieved a cohort of similar patients using a Mahalanobis distance and then selected a subset of predictors common between the new patient and the cohort of similar patients. Then, an LR model was trained on the cohort of similar patients to make predictions for the new patients. They acknowledged that personalized predictive modeling can perform better than conventional models. In their further exploration using clustering analysis on the risk factors, they found that patients with similar risk factors were grouped together and also that a large number of risk factors were not identified by their global model, whereas the PS-based model highlighted them. These results acknowledged the importance of using PS analytics in feature selection and modeling in clinical applications.

A Mahalanobis distance metric has also been used in unsupervised settings for finding similar patient subpopulations. Panahiazar et al. (2015) utilized this method for recommending medication to patients with a history of heart failure. Hierarchical clustering was used to find a subpopulation of similar patients, and the cluster was labeled by the most commonly used medication in that group. Then, for a new patient, the medication of the most similar cluster was recommended.

Besides PS measures such as Mahalanobis that have been widely used in the literature, some researchers have focused on similarity metrics such as correlation-based (Saeed et al. 2006a) and cosine similarity metrics. Lee et al. (2015a) examined the performance of PS-based predictive modelling in comparison to one-size fits-all methods in a systematic set of experiments using cosine similarity metrics. They compared the performance of personalized and global KNN, LR, and DT models. In all experiments, the PS settings not only outperformed the global setting but also the well-known medical scoring systems. It was also demonstrated that the size of the cohort of similar patients matters; a very small cohort can suffer from small sample size effect and decrease the performance. This study did not compare the performance of cosine metric to other metrics.

1.2.2.2 Other methods

Other non-conventional PS metrics have also been introduced in the literature. For instance, one study (Houeland 2011b) devised a PS metric by combining Euclidean distance and random forest. A random forest with decision trees with the height of five (16 terminal nodes) was trained first. Then, in the first stage of PS calculation, a cohort of similar patients (half of the size of the training set) to an index patient was retrieved using Euclidean distance. In the second stage, PS was further investigated using the random forest. Patients were sorted based on the number of trees in which they were assigned to the same node as the index patient. Then the data from the most similar patients in the second stage were used for prediction. This method outperformed the simple Euclidean distance.

While some researchers argue that the computational burden of neighbourhood-based PS analytics increases as the number of patients increases and requires more memory, a group of researchers have introduced the idea of implementing all steps of PS analytics ranging from predictor extraction to PS calculation in a database using Structured Query Language (SQL), which will be independent of Random Access Memory (RAM) size (Wiese et al. 2018; Tashkandi et al. 2018). In a comprehensive evaluation framework, this study demonstrated that PS calculation is more time-efficient in the database systems than the data mining tools (ELKI and Apache Mahout). Investigating the performance of cosine similarity and Euclidean distance, it was shown that Euclidean distance calculation is more time-consuming in the databased systems.

Although the literature (Perlman et al. 2011) suggests using various similarity metrics for different variables (for instance an ECG similarity metric, age similarity metric, and gender similarity metric)—because using just one similarity metric for all predictors may miss information relevant to prediction—only one study (Gottlieb et al. 2013) has taken this into consideration and defined eight similarity metrics between hospitalizations and two for ICD codes. All these similarity scores then formed the hospitalization-discharge code associations. Then, the researchers combined these measures into 16 hospitalization-discharge code associations.

PS has also been used with other types of similarity. Zhang et al. (2014b) combined PS analytics with drug similarity analytics to provide personalized drug recommendations on

hypercholesterolemia treatment. Three sets of Jaccard similarities were considered in this study: a) patient-patient similarity based on ICD-9 codes b) drug-drug similarity based on chemical structure and c) patient-drug similarity based on the ICD-9 diagnosis codes of patients and ICD-9 format drug indications from the MEDI database. Then, a label-propagation algorithm was employed to measure the efficiency of each of the available four drugs for a given patient in three settings: employing PS, drug similarity, and a combination of both similarities. The latter setting outperformed the others, suggesting that combining PS with drug similarity can help achieve personalized medicine.

Some studies have used predictive modeling for PS calculation. Wang (2015) introduced an Adaptive Semi-Supervised Recursive Tree Partitioning (ART) method to calculate pairwise PS with less computational burden. In this method, the tree is constructed based on two objective functions: a term based on expert knowledge and a term based on information from EHR data. This study also provided a kernelized tree-construction framework. The proposed method performed better than all methods they compared it with.

Zhang et al. (2018) utilized a Gaussian process in which a kernel function measured the similarity between the available patient data and that of new patients and reported a weighted average of all retrieved diagnoses. In this study, the inverse of similarity scores was used as the weights. The proposed method outperformed not only linear regression and DT, but also ranked the highest in the Alzheimer's Disease Big Data DREAM Challenge.

With the availability of big EHR databases, PS analytics has received attention from researchers in the deep learning field. To date, several researchers have investigated the idea of measuring PS via a deep neural network. Zhu et al. (2017) employed medical concept embedding (word2vec) to demonstrate the medication events, and then a convolutional neural network (CNN) was modified to derive pair-wise PS scores. In their network, patients' features are filtered through the convolutional layer, and then feature maps demonstrating patients' clinical characteristics are used for measuring PS. Suo and colleagues (Suo et al. 2017) also employed CNNs for PS calculation from ICD-9 codes. Their time-fusion CNN also learned the local temporal relationships between the codes and accounts for the time intervals between the codes. The proposed method outperformed the conventional PS metrics in KNN, weighted sampling, and personalized LR settings.

The evidence reviewed in this section suggests that PS analytics can often outperform one-size fits-all models. The contradictory evidence to this hypothesis suggests the need for further investigation of PS analytics' performance compared to population-based analytics. Together these studies provide insights into how PS can be employed in various stages of modelling. There remain several domains in PS that would benefit from further research, such as PS visualization, missing data treatment, and longitudinal data processing.

1.3 Thesis Rationale

As described in sub-section 1.1 Background, PS analytics is an emerging field and has repeatedly been used in various applications. While many studies have been done in this area,

most of them still suffer from poor handling of missing data and longitudinal data, and lack of interpretability in their PS calculation. There has been some research on missing data treatment in EHR databases; however, in-depth research is lacking on the potential informativeness of this missingness and how it can be converted to useful information. Therefore, a move away from traditional missing data treatment toward more-informative methods is needed. Moreover, studies on PS calculation have shown remarkable results in predicting outcomes; however, such predictive performance has been achieved at the expense of trading off the explainability of the method. In particular, studies in which deep learning has been used to calculate PS suffer from lack of interpretability and transparency. Although researchers in the field of explainable artificial intelligence are working on making these networks more interpretable, we should bear in mind that one of the PS analytics promises is to help clinicians in their decision making. Therefore, any PS calculation must be able to answer at least one question for clinicians: "why are these two patients similar?". Besides interpretability of PS calculation, there is also a need for more research on methods for including longitudinal data in PS analytics. Therefore, the overall goal of this thesis is to learn more from missing data in EHRs and use this information to identify patient subpopulations, introduce an explainable multifaceted PS metric, and take PS analytics into application, with a special focus on longitudinal data processing. Moreover, this thesis utilizes data from ICU patients, since there is a wealth of high-resolution data available in

many ICUs and PS analytics can be helpful in deriving insights from these complex and heterogenous ICU patient data.

1.4 Thesis Organization

1.4.1 Thesis Outline

The six chapters of this thesis comprise this introduction; four studies, all written in the form of journal or conference papers; and a general conclusion. To date, the first two studies have already been published, and the other two are very close to being submitted. The first study (Chapter 2) focuses on missing data, one of the main challenges in working with EHR data. This study, by utilizing a new data representation method (missingness indicators), investigates the informativeness of missing data in ICU EHR and how we can learn from missing data and use them in predictive modeling (Sharafoddini et al. 2019). The second study (Chapter 3) builds upon the first and utilizes the missingness indicators to capture laboratory-test-ordering patterns in ICU and uses this information to find subpopulations of similar patients in the ICU (Sharafoddini et al. 2018). The third study (Chapter 4) takes PS analytics further and introduces a multifaceted PS metric by which the PS is explainable. The proposed metric considers ICU PS from different aspects and investigates whether PS analytics is always a better alternative to population-based analytics. This study focuses on 30-day ICU mortality prediction. The last study (Chapter 5) applies all findings from the previous studies to a specific application, with a special focus on longitudinal data, and investigates the possibility of having subpopulations in septic patients.

Finally, Chapter 6 summarizes the findings from the four studies and discusses their limitations. This chapter also provides some possible future directions for interested researchers.

1.4.2 Overarching Objectives

Although each of the aforementioned studies had its own distinct objectives, the overarching goals of this thesis are to:

1. Explore the informativeness of missing data in an ICU database and introduce a new data representation that can measure PS in terms of missingness (Chapter 2).
2. Utilize missingness similarity in the context of PS analytics to identify subpopulations of ICU patients with similar laboratory test ordering patterns (Chapter 3).
3. Aggregate various aspects of PS in one PS metric to capture the similarity of ICU patients from different perspectives (vital signs, laboratory tests, etc.) in order to improve 30-days post-discharge mortality prediction (Chapter 4).
4. Utilize the multifaceted PS to investigate the heterogeneity of a cohort of ICU septic patients and identify groups of similar patients while including the vital signs trajectories in PS calculation (Chapter 5).

Chapter 2

A New Insight Into Missing Data in Intensive Care Unit Patient Profiles: Observational Study

This chapter investigates the informativeness of missing data in patient profiles and introduces a new data representation technique that can be used in PS similarity calculations for evaluating similarity in terms of missingness. This chapter was originally published in January 2019 and revisions have been made to the current copy based on the thesis committee's reviews. The full citation is as follows: Sharafoddini, A., Dubin, J. A., Maslove, D. M., & Lee, J. (2019). A New Insight Into Missing Data in Intensive Care Unit Patient Profiles: Observational Study. *JMIR Med Inform, 7*(1), e11605. doi:10.2196/11605

2.1 Abstract

Background: The data missing from patient profiles in intensive care units (ICUs) are substantial and unavoidable. However, this incompleteness is not always random or because of imperfections in the data collection process.

Objective: This study aimed to investigate the potential hidden information in data missing from electronic health records (EHRs) in an ICU and examine whether the presence or missingness of a variable itself can convey information about the patient health status.

Methods: Daily retrieval of laboratory test (LT) measurements from the Medical Information Mart for Intensive Care III database was set as our reference for defining complete patient profiles. Missingness indicators were introduced as a way of representing presence or absence of the LTs in a patient profile. Thereafter, various feature selection methods (filter and embedded feature selection methods) were used to examine the predictive power of missingness indicators. Finally, a set of well-known prediction models (logistic regression [LR], decision tree, and random forest) were used to evaluate whether the absence status itself of a variable recording can provide predictive power. We also examined the utility of missingness indicators in improving predictive performance when used with observed laboratory measurements as model input. The outcome of interest was in-hospital mortality and mortality at 30 days after ICU discharge.

Results: Regardless of mortality type or ICU day, more than 40% of the predictors selected by feature selection methods were missingness indicators. Notably, employing missingness indicators as the only predictors achieved reasonable mortality prediction on all days and for all mortality types (for instance, in 30-day mortality prediction with LR, we achieved area under the curve of the receiver operating characteristic [AUROC] of 0.6836±0.012). Including indicators with observed measurements in the prediction models also improved the AUROC; the maximum improvement was 0.0426. Indicators also improved the AUROC for Simplified Acute Physiology Score II model—a well-known ICU

severity of illness score—confirming the additive information of the indicators (AUROC of 0.8045±0.0109 for 30-day mortality prediction for LR).

Conclusions: Our study demonstrated that the presence or absence of LT measurements is informative and can be considered a potential predictor of in-hospital and 30-day mortality. The comparative analysis of prediction models also showed statistically significant prediction improvement when indicators were included. Moreover, missing data might reflect the opinions of examining clinicians. Therefore, the absence of measurements can be informative in ICUs and has predictive power beyond the measured data themselves. This initial case study shows promise for more in-depth analysis of missing data and its informativeness in ICUs. Future studies are needed to generalize these results.

Keywords: Electronic Health Records; Clinical Laboratory Tests; Imputation Methods; Feature Selection Methods; Machine Learning; Mortality Prediction.

2.2 Introduction

2.2.1 Background

The increased adoption of electronic health record (EHR) systems has boosted interest in the secondary use of EHR data (Weiskopf et al. 2013c). Although the literature has introduced various dimensions for EHR data quality, completeness and correctness have been reported as the fundamental dimensions (Weiskopf et al. 2013c; Chan et al. 2010). Although these issues can also be observed in paper-based records, EHR brought us the opportunity to

identify them faster and helped us with addressing them. The data missing from clinical contexts are substantial (Weiskopf et al. 2013a; Wells et al. 2013) and unavoidable (Little et al. 2012); many studies have focused on resolving this issue (Sterne et al. 2009; Haukoos et al. 2007; Newgard et al. 2007). Although many researchers treat missing data as a challenge (Pringle et al. 1995; Thiru et al. 2016; Forster et al. 2008; Jones et al. 1986; Porcheret et al. 2004; Soto et al. 2002; Tang et al. 1999; Jensen et al. 2009; Botsis et al. 2010; Sharafoddini et al. 2017), others continue to debate whether lack of completeness also provides useful information (Wells et al. 2013; Rusanov et al. 2014; Weiskopf et al. 2013b; Agniel et al. 2018). Researchers do agree that a part of this incompleteness is not random or because of imperfections in the data collection process (Kuhn et al. 2013; Agniel et al. 2018). Recently, Angiel et al. (2018) demonstrated that the laboratory ordering time (ie, the interval between 2 orders of a laboratory test; LT) for some LT is more informative than the actual values in predicting 3-year survival. Our study focuses on systematically investigating the implications or possible value of lack of data, particularly in intensive care units (ICUs) and proposes a representation method for missing data to capture hidden information. In general, two reasons are given for missing data in EHRs:

- No intention to collect: the clinical variable was never measured because there was no clinical indication to do so—the patient was not suffering from a relevant symptom or comorbidity (Wells et al. 2013), or it could not be measured (Rusanov et al. 2014).

- Intention to collect: records are missing although the variables were measured (Wells et al. 2013).

Therefore, the health care process (e.g., clinicians' decision to order a test and nurse data entry) affects the recorded EHR and can cause incompleteness in data.

Incomplete EHR data can complicate or prohibit the data analysis process, as many machine learning (ML) algorithms assume that there are no missing data in the dataset or require users to clean the data in the preprocessing stage and so provide a complete dataset. Therefore, from a research perspective, the ideal situation is to increase the amount and accuracy of EHR documentation by employing approaches that focus on intention to collect, such as reducing the error in data entry or increasing data documentation in terms of resolution. Although the current amount of testing and bloodwork has been reported as actually redundant in ICUs (Lee et al. 2015b; Oliveira et al. 2014; Cismondi et al. 2013) and requires extra time and work from clinicians (Wells et al. 2013), these approaches suffer from their own shortcomings. Besides analytical methods that can handle missing data (that are missing at random) such as decision trees (DTs) or mixed-effects models for longitudinal data, other approaches usually assume missing data are missing completely at random. In general, the literature proposes 3 analytical approaches: complete case analysis (CCA) or deletion, available case analysis (ACA), and imputation.

CCA starts with the list of variables included in the analysis and discards records with missing data on any of the variables. However, this subsample might not be a random sample

of the population. Although researchers argue that sample selection based on the predefined eligibility criteria in randomized clinical trials can limit the external generalizability of these studies (Rothwell 2005), CCA in studies using EHR data can also potentially threaten the external validity of a study (Rusanov et al. 2014) and cause bias as the literature shows a statistically significant relationship between severity of illness and data completeness (Weiskopf et al. 2013b). For example, a study (Rusanov et al. 2014) on 10,000 EHRs from patients receiving anesthetic service showed that patients with an anesthesiologists physical status (ASA) (Delegates 2014) class 4 fitness rating had 5.05 more days with laboratory results and 6.85 more days with medication orders than patients with ASA class 1, suggesting more data are recorded for sicker patients than healthier patients. Thus, imposing complete case requirements when using EHR data for secondary use can cause bias toward selecting patients with more severe conditions (or several comorbidities). Despite this drawback, CCA has been identified as the leading approach in studies on ICU data (Vesin et al. 2013). That said, CCA provides valid inference only when data are missing completely at random (MCAR), which is unlikely in practice (Fitzmaurice et al. 2011).

The ACA (or pairwise deletion) utilizes all available data for a given analysis. In other words, it maximizes the availability of data by an analysis-by-analysis basis (Baraldi et al. 2010). The advantage of this method is that more data are included in each analysis than with CCA. It also allows for valid inference by likelihood-based models when missing data are ignorable—often the case when the data are missing at random (MAR) (Fitzmaurice et al.

2011). In each analysis, the ACA method utilizes the data points that are needed for that particular analysis. For instance, in correlation analysis to measure the correlation for each pair of variables, ACA uses the data points that have the information for those variables in each analysis resulting in varying samples. Although ACA is an improvement to CCA (Baraldi et al. 2010), it also has limitations. As different samples are being used in each analysis, not only is comparison of various analyses impossible (Stockdale et al. 2016) but also, using different samples for estimating the parameters of interest has occasionally led to biased or mathematically inconsistent results (Myers 2011; Pigott 2010; Roth 1994).

Imputation methods, which try to draw inferences from incomplete data, rely on knowing the mechanism of missingness, which cannot be validated from the available data. Single imputation methods suffer from 2 problems. First, an inference based on imputed data can be biased if the underlying assumptions are not valid. Second, because imputed data are assumed to be true, the model's statistical precision is overstated. Multiple imputation methods, in spite of their promising performance, rely on parametric assumptions that, if not valid, can lead to incorrect imputation. Due to these limitations, imputation methods should be used with caution, and checking underlying assumptions with clinicians is highly recommended (Little et al. 2012). However, Gorelick (2006) in a simulation study, demonstrated that either CCA or imputation could cause bias in predictive modeling, and that assuming missing values to be normal when missingness rates are high and substituting them with normal values would also cause substantial bias. In brief, if primary assumptions are not

fully satisfied, neither considering complete or available cases nor imputing missing data is likely to yield reliable results. Furthermore, these statistical methods on their own are not sufficient to capture the hidden information about the patient health status and care process in the complex EHR data. Alternatively, we can try to learn from what is missing rather than only dealing with missingness as a deficiency.

2.2.2 Objectives

This case study provides evidence that 'missing data' in ICU might be missing because of the patient's health status or health care process and introduces a new method for representing patient profiles. In this representation, auxiliary variables, called indicators, are used to represent the presence or absence of a measurement and might convey the possible hidden information in the missing data. Then, by employing various analytical methods, this study attempts to demonstrate the informativeness of missing data. In the rest of the study, the term missing data is used to describe not-at-random missing information in patient profiles. In other words, the potential informativeness of data that has not been recorded by choice is of interest.

2.3 Methods

2.3.1 Measurement Protocol and Data Collection

As patient monitoring strongly relies on clinical needs, no universal standards for ICU data completeness have been established (Schulman et al. 2010; Asiimwe et al. 2014; Cardona-

Morrell et al. 2015). However, a study by Frassica in 2005 published a list of the top 80% of LTs common to all ICU patients within a university teaching hospital. We revised this list based on the presence of these tests in our database and updated it with input from an ICU clinician to reflect current practices (Table 2-1).

Table 2-1 Thirty six laboratory tests used in investigating informativeness of missing data.

Variable Category	**Variables**
Top 80% laboratory tests and profiles common to all ICUs (Frassica 2003) reviewed and revised by domain expert	Alanine Aminotransferase (ALT); Alkaline Phosphatase (ALK); Aspartate Aminotransferase (AST); Arterial blood gases: pH, PCO2, PO2, Base Excess (BE); Basic metabolic panel: Sodium (Na), Potassium (K), Chloride (Cl), Bicarbonate (HCO3), Anion Gap (AG), Blood Glucose (BG), Blood Urea Nitrogen (BUN), Creatinine (Cr); Complete blood count: White blood cells (WBCs), red blood cells (RBCs), Hemoglobin (HGB), Hematocrit (HCT), Mean corpuscular volume (MCV), Mean corpuscular hemoglobin(MCH), Mean corpuscular hemoglobin concentration (MCHC), Red cell distribution width (RDW), Platelet count (PLT), Absolute Monocytes (MO), Absolute Eosinophils (EO), Absolute Basophils (BA), Absolute Lymphocytes (LY), Absolute Neutrophils (NE); Lactate (Lac); Calcium (Ca); Magnesium (Mg); Phosphorus/ Phosphate (Phos); Partial Thromboplastin Time (PTT); Prothrombin Time (PT); Total Bilirubin (TBil).

The data for this study were collected from the Medical Information Mart for Intensive Care III (MIMIC-III) (Johnson et al. 2016b) database which contains data from 38,597

distinct adult patients admitted to the Beth Israel Deaconess Medical Center in Boston, Massachusetts, between 2001 and 2012. For patient cohort selection, a tailored version of the generalized cohort selection heuristics for retrospective EHR studies introduced by Harrell et al. (2016) was used. The data for first admission to 1 of the 5 ICUs—medical ICU, surgical ICU, cardiac care unit, cardiac surgery recovery unit, and trauma surgical ICU—were extracted for adult patients (aged 15 years or older). Included patients must have had at least one data point in any of the variable categories during the first, second, and third days of their ICU stay.

2.3.2 Data Preprocessing and Missing Data Representation

Each day's extracted data were mapped into a matrix with columns for measurements and rows for patients. Therefore, we had a column for each daily measurement of LTs, resulting in 36 columns for LTs. An auxiliary matrix was generated to store binary values reflecting the presence (0) or absence (1) of measurements. Since many well-performing ML algorithms are designed to work with a complete data matrix, two methods—Predictive Mean Matching (PMM) (Little 1988) and Hot Deck (HD)—were used to impute missing laboratory test measurements. PMM is a commonly used and well-accepted imputation method in public health research (Zhou et al. 2001) and is also robust against model misspecification (Buuren 2012). HD imputation is used commonly in applied data analysis when missing data exist (Andridge et al. 2010).

Given that imputed values are indistinguishable to the ML algorithm from true values, we combined the original matrix and auxiliary matrix to form an augmented matrix that directly indicates where values were imputed. This was done to mitigate the risk of treating imputed values the same as actual values, in a setting where the underlying reason for missing data is not fully known (Figure 2-1). Missing data indicators in this augmented matrix might also provide extra information about the reliability of the values (actual and imputed values) and potentially preserve any meaningful missing data patterns. Missingness indicators have been used as a method of handling missing data in epidemiological and clinical studies. However, in the current use of indicators, missing values are set to a fixed value (0 or the normal value for the variable) and the indicators are used as dummy variables in analytical models to indicate that a value was missing (Abraham et al. 2004; Groenwold et al. 2012). Studies have shown that this method causes bias as the missing values are imputed with a single value (Knol et al. 2010). In our study, we are not using indicators as dummy variables; instead, we are introducing them as a source of information to be used besides imputation methods.

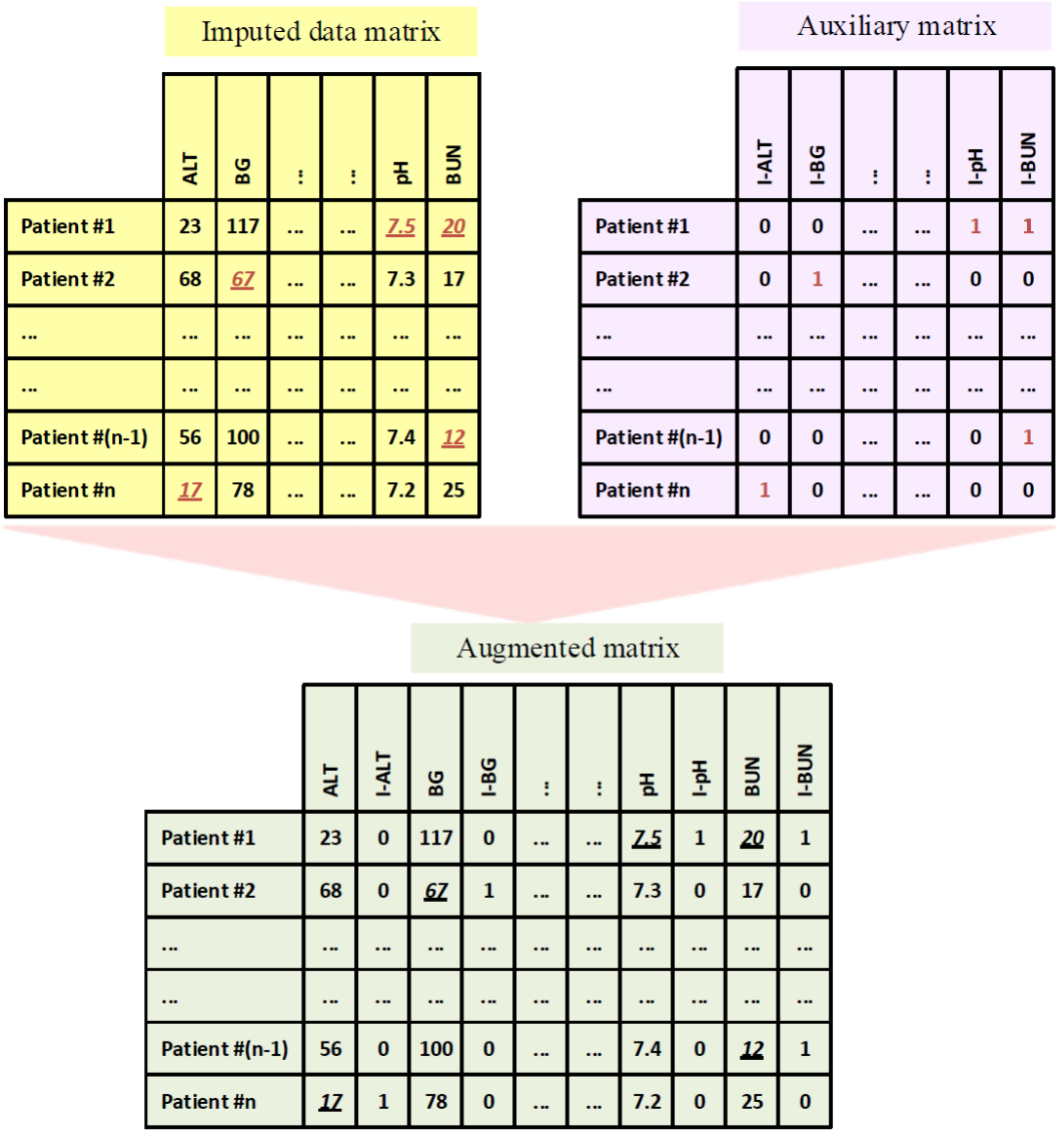

Figure 2-1 An example of the augmented data matrix, the imputed data matrix (imputed values are underlined and italicized) and the auxiliary matrix (containing the missingness indicators: 0-present, 1-absent).

2.3.3 Validation

Several validation techniques are available in medical research. In this study, for all experiments where applicable, we used cross-validation technique (10-fold cross-validation). We also repeated the cross-validation procedure several times (20 times) to acquire more stable results as suggested in the literature (Steyerberg 2009).

2.3.4 Assessments

2.3.4.1 Exploratory Analysis

First, the trends of missingness among LTs were visualized for comparison. Afterward, pairwise correlation among indicators, using Phi coefficient, was done to explore the general behavior of missingness. The Elixhauser (1998) and the Charlson (1987) comorbidity indices are the most common comorbidity scores in clinical applications. The literature has shown that the Elixhauser Comorbidity Index (ECI) in general has the best performance in predicting mortality (Menendez et al. 2014; Southern et al. 2004; Farley et al. 2006; Sharabiani et al. 2012). This better performance can be the result of (1) including new comorbidities in ECI, (2) the differences in the coding of variables common between both indices, or (3) a combination of the first and second factors (Southern et al. 2004). The Simplified Acute Physiology Score II (SAPS-II) (Le Gall et al. 1993) scoring system that has been widely used by most ICUs for predicting illness severity was also chosen. Therefore, the association of missingness rates with ECI and SAPS-II was investigated using Spearman

correlation. Besides the clinical information, SAPS-II also has the information about type of admission (scheduled surgical, medical, or unscheduled surgical) and presence of 3 chronic diseases (metastatic cancer, hematologic malignancy, and AIDS).

2.3.4.2 Feature Selection

After exploratory analyses, we assessed the importance of the indicators as potential predictors. First, we used feature selection methods, which are widely used to determine which predictors should be used in a model, particularly for high-dimensional data (Kuhn et al. 2013). Two copies of the augmented matrix (derived from HD and PMM imputation) were fed to various feature selection methods. Our study considered in-hospital and 30-day post-discharge mortality as outcomes. Overall, we used 2 categories of supervised feature selection methods described below.

First, filter techniques evaluated the importance of a predictor by looking at data properties. Filter methods, in general, use a metric to identify irrelevant features and filter out the redundant predictors form the data matrix (Saeys et al. 2007). We selected 3 different metrics: LR beta value, Relief algorithm (Robnik-Šikonja et al. 2003), and Information Gain (InfGain) (Peng et al. 2005). The Relief algorithm examines the relevance of predictors based on their power to distinguish between similar patients with the same and different outcome. InfGain measures the reduction in entropy of the class variable achieved by partitioning the data based on the index predictor; relevant predictors receive a high InfGain value (Mitchell 1997). This ensemble of the scoring methods was then used to determine the normalized

informativeness of all predictors. Aggregating these methods in one score provides a tool for comparing predictors from different aspects.

Second, we used embedded techniques to search for the optimal set of predictors. In these techniques, feature selection is embedded in the model's construction and interacts with the classifier. Least absolute shrinkage and selection operator (LASSO), used in this study, is a penalizing method in this category. LASSO regression in its objective functions considers a penalty that equals to the sum of the absolute values of the coefficients. As absolute function (L1 norm) is not differentiable, the estimated coefficients are close to 0, and some will be exactly 0 resulting in an automatic variable selection. For this and the next experiments, 10-fold cross-validation with 20 repeats was used (leading to 200 repetitions in total). This number of repetitions is recommended to achieve desired accuracy for prediction performance estimation (Steyerberg 2009).

2.3.4.3 Predictive Modeling

In the last assessment, we first trained group of classification models, including DT, logistic regression (LR), and random forest (RF), on the indicator and imputed data matrices and evaluate their performance for predicting desired outcomes using the area under the curve of the receiver operating characteristic (AUROC) validation metric. Thereafter, new models were trained using the augmented data matrix and their performance was compared with that of the original to determine whether the indicators have predictive power and can boost the models' predictive accuracy. We also investigated the predictive performance of SAPS-II

score, and then we added indicators to these scores to examine the impact of indicators beyond SAPS-II score. It is worth mentioning that in this assessment, the absolute accuracy of the models is not of our interest, instead, the relative improvement in the performance when including indicators as input. That is, achieving the best possible mortality prediction AUROC is not the objective of this study.

2.4 Results

2.4.1 Population

The analyses of the first 24 hours after admission to ICU included 32,618 patients but decreased to 20,381 for the second 24-hour interval, as many patients were discharged after 24 hours. The third 24-hour period included 13,670 patients. Of these groups, 10.99% (3586/32,618), 13.59% (2769/20,381), and 16.19% (2213/13,670) experienced death in-hospital and 15.12% (4933/32,618), 18.26% (3722/20,381), and 21.32% (2915/ 13,670) experienced death within 30 days of discharge, respectively. Figure 2-2 demonstrates the retrospective study design.

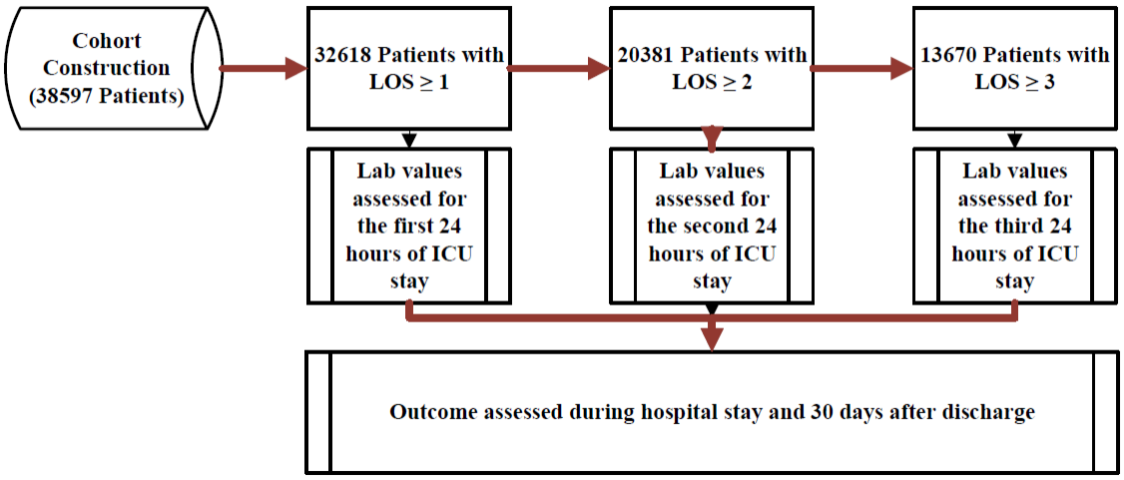

Figure 2-2 The retrospective cohort study design. LOS: length of stay.

2.4.2 Exploratory Analysis

Missingness rates for LTs range from 1.36% (445/32,618) to 88.27% (12066/13,670) in the first 72 hours after admission. Figure 2-3 shows the missingness rate for LTs over 3 days. Absolute basophils (BA), absolute eosinophils (EO), absolute monocytes (MO), absolute lymphocytes (LY), absolute neutrophils (NE), alanine aminotransferase (ALT), alkaline phosphatase (ALK), aspartate aminotransferase (AST), total bilirubin (TBil), and lactate (Lac) were among the less-common LTs and were missing in the profiles of more than 60% of patients.

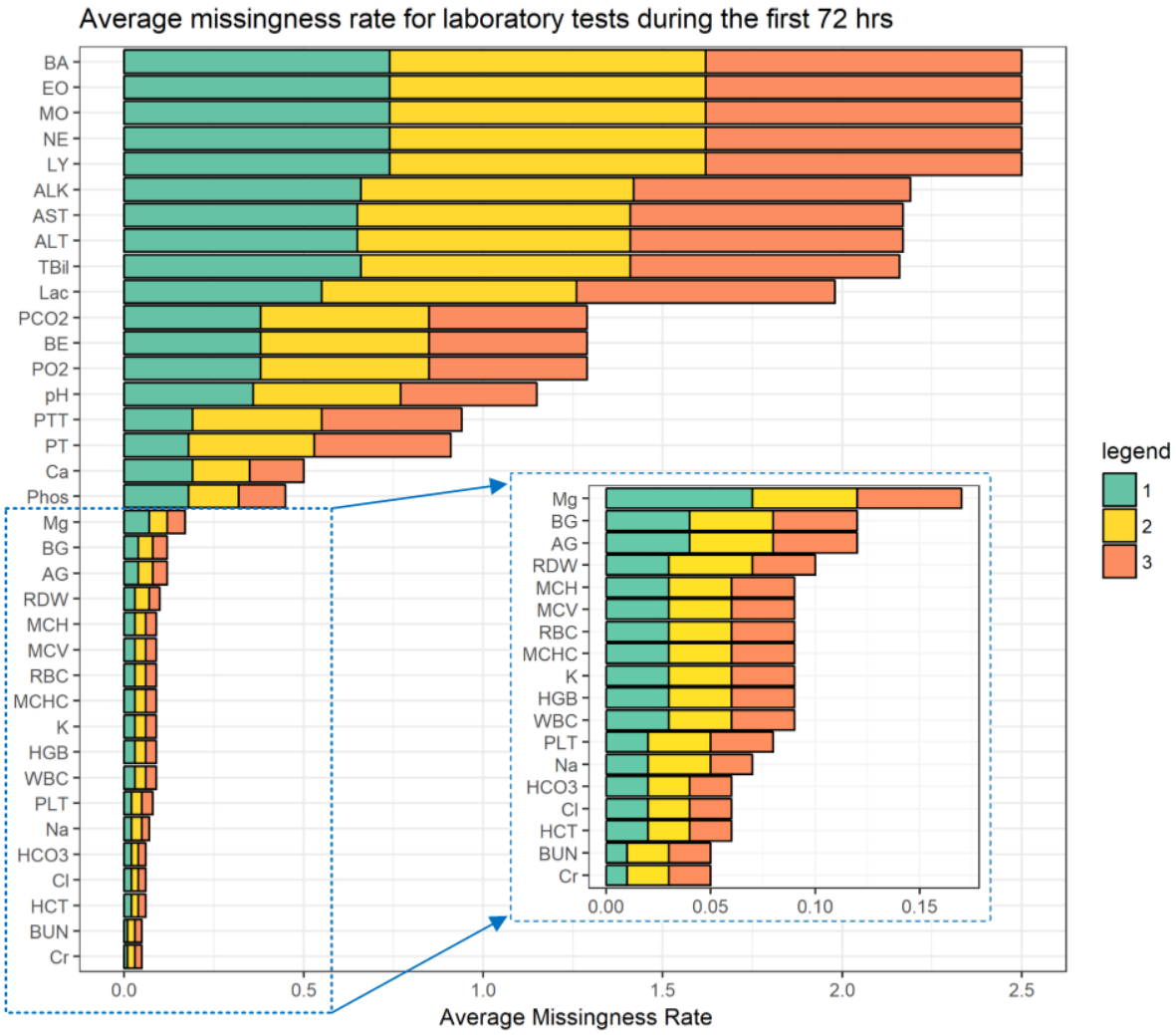

Figure 2-3 The average missingness rate among patients for laboratory tests in the first 72 hours of admission.

We calculated the association between each indicator and the mortality flag. Although association values were small, on day 1, ALT, ALK, AST, and TBil stand out as the top LTs associated with both types of mortality. On days 2 and 3, partial pressure of carbon dioxide

(PCO2), partial pressure of oxygen (PO2) and base excess (BE) were the top LTs associated with both mortality types. Lac also joined the top tests on day two for 30-day mortality. Detailed association values are provided in Appendix C.

Figure 2-4 visualizes the pairwise correlations among indicators. In total, 7 major groups of highly correlated ($\rho \geq .95$) indicators were observed in the results using Phi coefficient: (1) BA, MO, NE, EO, and LY; (2) mean corpuscular hemoglobin concentration (MCHC), red cell distribution width (RDW) mean corpuscular volume (MCV), red blood cell (RBC), and mean corpuscular hemoglobin (MCH); (3) BE, PCO2, and PO2; (4) TBil, ALT, AST, and ALK; (5) Blood urea nitrogen (BUN) and creatinine (Cr); (6) chloride (Cl) and bicarbonate (HCO3); (7) partial thromboplastin time (PTT) and prothrombin time (PT).

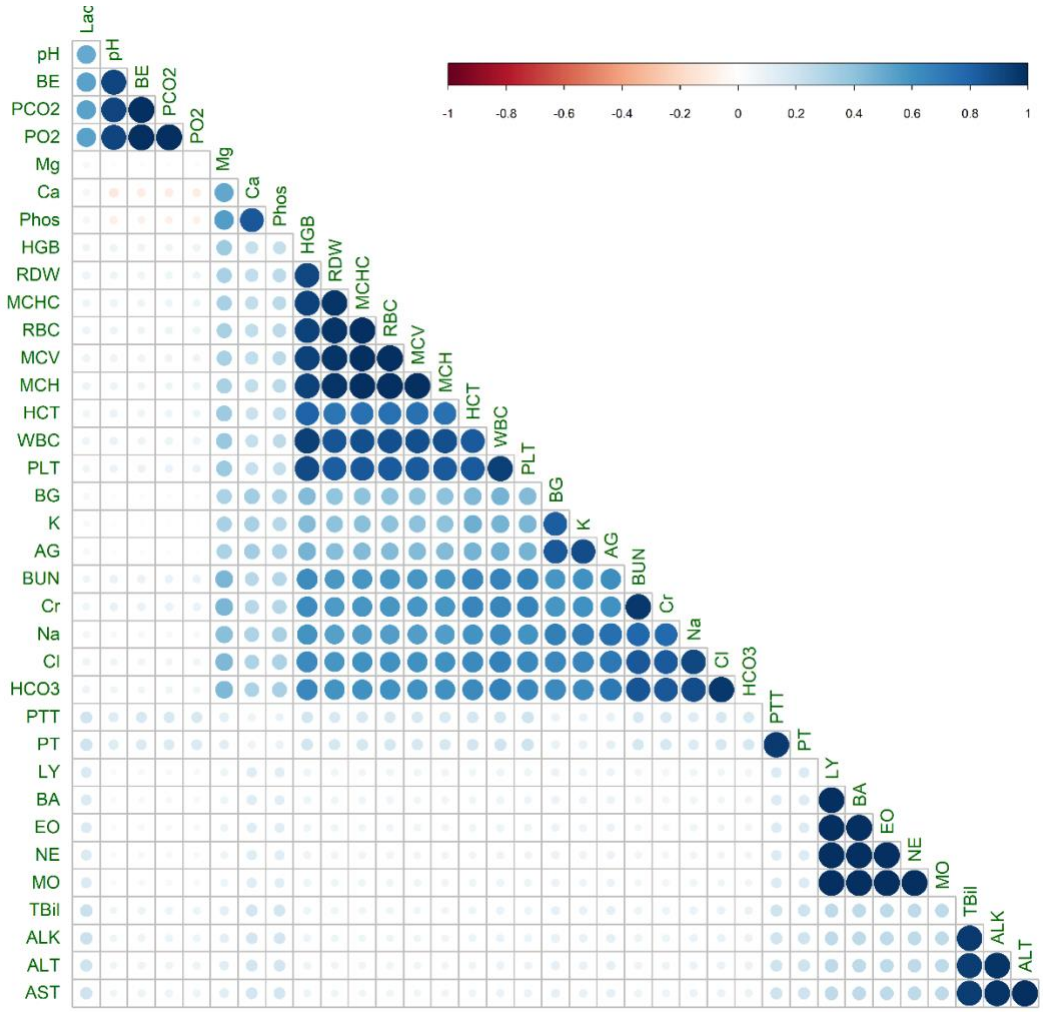

Figure 2-4 Visualization of the correlation matrix for variable indicators in first 72 hours.

The Spearman correlation between missingness rates and ECI was also calculated daily. Results show a statistically significant correlation between these variables (day 1: ρ=–.233; day 2: ρ=–.196; day 3: ρ=–.184; P<.001). The same assessment was done using SAPS-II. The results were in line with the previous one and demonstrate higher correlation (day 1: ρ=–

.315; day 2: ρ=–.277; day 3=–.234; P<.001). These findings are interesting as they confirm that the missingness of data is associated with patient severity of illness.

2.4.3 Feature Selection: Missing Data Indicators as Important Predictors

Each of the imputation methods was applied to the original dataset, and the potential informativeness of missingness indicators in comparison with actual variables was investigated using an ensemble of the most representative filter selection methods (Aggarwal 2014): LR beta value, relief, and InfGain. Table 2-2 shows the top 18 variables selected on each day based on the PMM-generated imputed matrix predicting 30-day mortality. BUN, RDW, and anion gap (AG) were among the top variables in all 3 days. Indicators for TBil, phosphate (Phos), calcium (Ca), and Lac were selected on the first day, whereas indicators for Lac, BE, PO2, and PCO2 were among the top features on the second and third days. PTT and pH indicators were also among the important indicators on the third day.

Table 2-2 The top 18 variables selected on each day after employing predictive mean matching imputation with regard to 30-day mortality. The 'I' at the beginning of the variables' names means indicator. Numbers represent the ranking after aggregating the ranking results from the 3 different feature selection methods.

Day 1		**Day 2**		**Day 3**	
Variable	Score	Variable	Score	Variable	Score
BUN	0.762397	AG	0.795419	RDW	0.748997
RDW	0.680087	HCO3	0.783337	BUN	0.666667
MCHC	0.668965	BUN	0.77677	HCO3	0.544964
AG	0.540484	BE	0.609532	BE	0.540542
I-Ca	0.436429	RDW	0.608711	pH	0.488433
Cr	0.436071	I-PO2	0.587151	AG	0.450426
HCO3	0.416741	I-PCO2	0.585947	I-Lac	0.418716
PO2	0.404289	I-BE	0.585592	I-pH	0.40463
MCV	0.386964	Cl	0.53158	Cr	0.400008
I-Phos	0.374431	PT	0.462085	Phos	0.387661
PTT	0.353913	Lac	0.461869	I-PCO2	0.387019
HGB	0.342786	Cr	0.451999	I-PO2	0.386739
pH	0.32767	PTT	0.424956	I-BE	0.385935
Lac	0.320339	Na	0.422474	PCO2	0.367257
BE	0.320299	Phos	0.419171	NE	0.360791
I-Lac	0.318216	I-Lac	0.415475	MCV	0.351266
PCO2	0.316668	MCV	0.368343	I-PTT	0.338352
I-TBil	0.31277	MCHC	0.363146	Lac	0.331205

Similar results were observed when using the HD imputation method, except that ALT and Phos were also selected on the first and second day, respectively. Moreover, PTT and pH indicators were not among the important indicators on the third day. Detailed results of this assessment can be found in Appendix C.

Results for in-hospital mortality were slightly different (Table 2-3). Although the selected indicators were almost the same as for 30-day mortality, more indicators were selected on the first day for in-hospital mortality, implying that indicators are more associated with in-hospital mortality than 30-day mortality. Detailed results are available in Appendix C.

Table 2-3 The top 18 variables selected on each day after employing predictive mean matching imputation with regard to in-hospital mortality. The 'I' at the beginning of the variables names means indicator. Numbers represent the ranking after aggregating the ranking results from the 3 different feature selection methods.

Day 1		**Day 2**		**Day 3**	
Variable	**Score**	**Variable**	**Score**	**Variable**	**Score**
BUN	0.825715	BUN	1	RDW	0.75246
AG	0.668918	RDW	0.711852	BUN	0.635729
RDW	0.573188	HCO3	0.684191	BE	0.633926
HCO3	0.531746	AG	0.664339	HCO3	0.62367
MCHC	0.507343	BE	0.528778	I-BE	0.595553
PCO2	0.489483	MCHC	0.503805	I-PCO2	0.595238
Cr	0.480181	PT	0.453111	I-PO2	0.594924
BE	0.452599	Cl	0.429405	pH	0.556242
I-LAC	0.436382	I-LAC	0.425279	Phos	0.494694
LAC	0.415773	Cr	0.395266	AG	0.492864
HGB	0.414263	I-PO2	0.382404	I-pH	0.470007
pH	0.402466	I-PCO2	0.381737	I-LAC	0.469215
I-TBil	0.399363	I-BE	0.381448	Cr	0.415249
I-Ca	0.395278	PTT	0.357339	LAC	0.396136
I-ALT	0.376004	Phos	0.352738	NE	0.338372
I-AST	0.375944	Na	0.345109	PT	0.326491
LY	0.375163	I-PT	0.333936	LY	0.319146
I-ALK	0.366346	BG	0.320947	MCV	0.314868

To validate our previous results, we assessed the predictive power of the indicators using embedded feature selection methods. Each day, a LASSO model was trained on the augmented data from HD and PMM imputation using 10-fold cross-validation with 20 repeats. In general, the AUROC of mortality prediction (in-hospital and 30-day postdischarge) and number of selected variables decreased from days 1 to 3 (Table 2-4).

Moreover, prediction of in-hospital mortality resulted in higher AUROCs than 30-day mortality. Regardless of mortality type, on all days, more than 40% of the predictors selected by the best-performing model were indicators. Moreover, more than 61% of selected predictors were indicators on the third day. Sliding lambda to compromise the predictor number and model performance led to almost the same results. Generally, more than 40% of the selected predictors were indicators, and on the third day, this number increased to 61%.

Table 2-4 Results from feature selection by least absolute shrinkage and selection operator (LASSO) for 3 days (area under the curve of the receiver operating characteristics are reported with the SE). The best performing model refers to the model with a lambda value associated with minimum cross-validation error. The adjusted model refers to a LASSO model with the largest value of lambda such that the error remains within 1 SE of the minimum.

Outcome	Imputation method	Day 1	Day 2	Day 3
AUROC for best performing model				
30-day mortality	**HD**	0.7858 (0.0033)	0.7685 (0.0041)	0.7302 (0.0043)
	PMM	0.7876 (0.0039)	0.7708 (0.0046)	0.7391 (0.0053)
In-hospital mortality	**HD**	0.7983 (0.0040)	0.7804 (0.0046)	0.7476 (0.0042)
	PMM	0.8007 (0.0047)	0.7838 (0.0049)	0.7582 (0.0054)
Indicators among selected predictors by the best performing model, n (%)				
30-day mortality	**HD**	43% (23/53)	48% (24/50)	70% (19/27)
	PMM	45% (26/58)	47% (26/55)	68% (17/25)
In-hospital mortality	**HD**	46% (28/61)	48% (29/61)	60% (21/35)
	PMM	47% (29/62)	49% (27/55)	62% (24/39)
AUROC for adjusted model				
30-day mortality	**HD**	0.7826 ± 0.0034	0.7646 ± 0.0043	0.7262 ± 0.0041
	PMM	0.7840 ± 0.0038	0.7667 ± 0.0045	0.7339 ± 0.0044
In-hospital mortality	**HD**	0.7944 ± 0.0043	0.7762 ± 0.0047	0.7439 ± 0.0041
	PMM	0.7961 ± 0.0049	0.7793 ± 0.0050	0.7536 ± 0.0045

Indicators among selected predictors by the adjusted model, n (%)				
30-day mortality	**HD**	45% (20/44)	48% (16/33)	67% (22/33)
	PMM	45% (19/42)	52% (16/31)	62% (31/50)
In-hospital mortality	**HD**	47% (20/43)	42% (13/31)	64% (16/25)
	PMM	50% (18/36)	41% (11/27)	62% (16/26)

Results in this section once more confirm the informativeness of missing data as missingness indicators have been selected by various feature selection methods. The high percentage of selected indicators also implies that the actual value of an LT is not always required in outcome prediction; instead, knowledge about whether the test was performed would suffice.

2.4.4 Predictive Modeling: Missing Data Indicators in Predictive Modeling

In the second assessment, we compared the performance of a set of 3 classification models (DT, LR, and RF) using the indicators, imputed and augmented data matrices, and SAPS-II score with or without indicators with 10-fold cross-validation over 20 repeats. We investigated whether including indicators can improve prediction and whether indicators alone have predictive power. For our LR, the iteratively reweighted least square method was used to fit the model. The complexity parameter (CP) for DT was tuned based on the model performance. On the basis of some preliminary model fitting, we set the CP value to vary from 0 (including all variables and having a large tree) to .02 for each model and then we picked the best performance model. In all models, the best-tuned model had a CP greater

than 0. Figure 2-5 shows the AUROC with 95% CI for all 3 days with regard to 30-day mortality (Appendix C provides the AUROC values for 30-day mortality and in-hospital mortality).

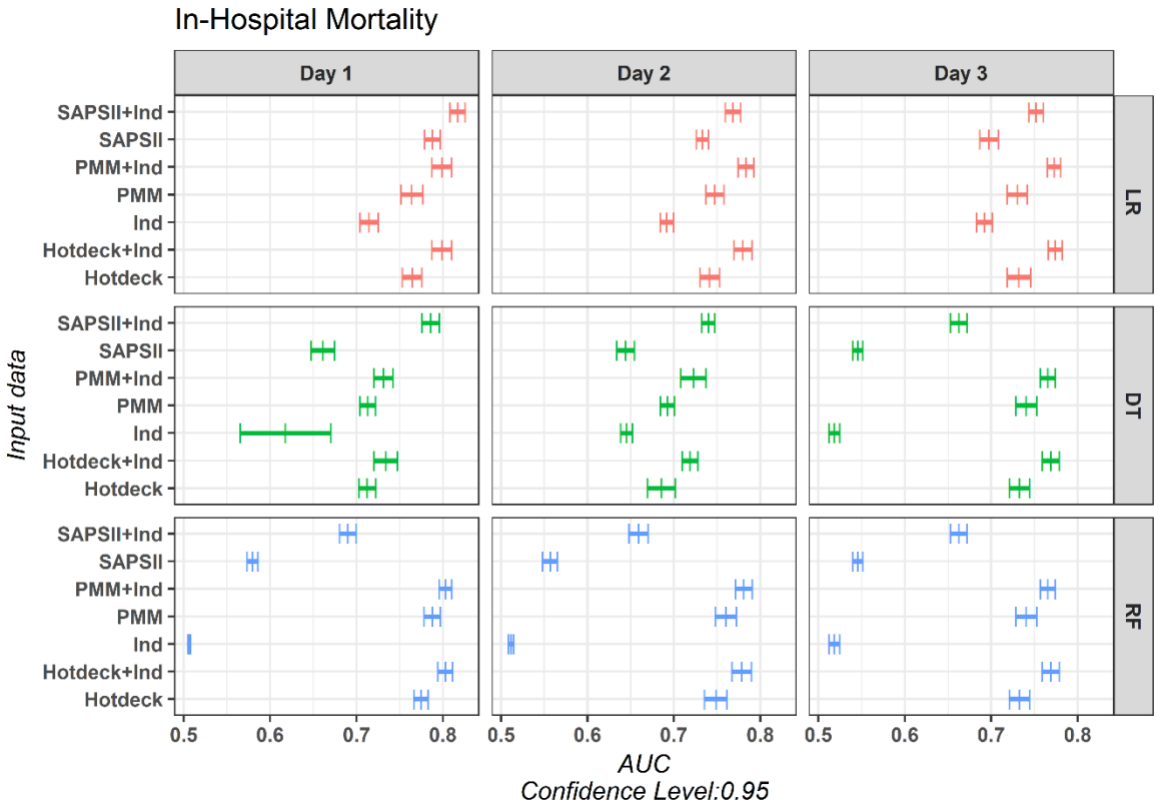

Figure 2-5 The 95% confidence intervals of the AUROC for LR, DT and RF models on missingness indicators, SAPS-II score and actual variables with and without the missingness indicators.

Including indicators improved the AUROC in all modeling techniques, on average by 0.0511; the maximum improvement was 0.1209 (Figure 2-5). AUROC has been demonstrated as an insensitive metric, for which an increase of 0.01 suggests meaningful improvement and is clinically of interest (Martens et al. 2016; Cook 2007; Pencina et al. 2012). Although using only indicators demonstrated reasonable performance in all scenarios (AUROC=0.6019 [0.0862]>0.5), conventional scores such as SAPS II perform better (AUROC=0.6390 [0.0853]) on their own. Therefore, models trained only on indicators are not sufficient. However, including indicators with conventional scores can improve the performance (AUROC=0.7263 [0.0578]). The SAPS-II score has information for age, heart rate, systolic blood pressure, Glasgow coma scale, temperature, mechanical ventilation administration, partial pressure of oxygen in the arterial blood (PaO2), fraction of inspired oxygen (FiO2), urine output, BUN, sodium (Na), potassium (K), HCO3, TBil, white blood cells (WBCs), presence of chronic diseases, and type of admission. These results demonstrate that indicators have information beyond that included in SAPS-II.

Figure 2-6 demonstrates the AUROC curves for LR 30-day mortality prediction on day 1.

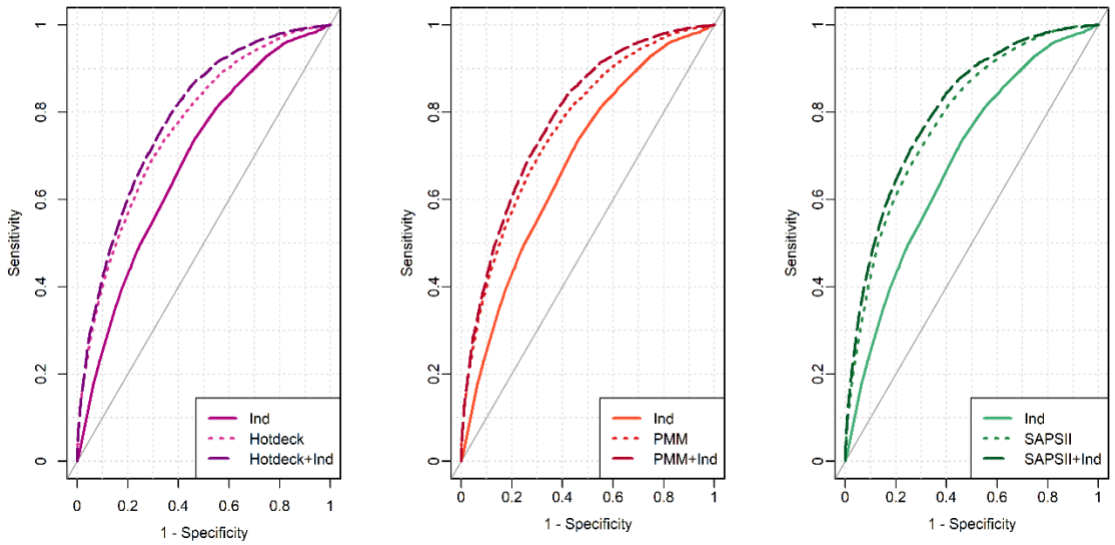

Figure 2-6 The ROC curves for LR 30-day mortality prediction on day one.

This combination of findings provides more support for the informativeness of missing data. Employing the missing indicators in mortality prediction modeling can improve the results in comparison to not including them.

2.5 Discussion

2.5.1 Principal Findings

We used missingness indicators to represent missing information in patient profiles in ICU. The informativeness of these indicators was demonstrated in 3 sets of assessments. First, our exploratory analysis confirms that the missingness of data is associated with patient severity of illness or comorbidities. Afterward, by means of feature selection methods, the predictive power of the presence of an LT in the patient profile was found to be more than the actual

measured value. Finally, missingness indicators noticeably improved the performance of mortality prediction models. The high correlation observed among some of the variable indicators suggests that all the variables in a set are typically measured or ordered together. Therefore, if a patient is missing 1 variable of a set, he or she will likely be missing the others as well. This fact is well represented in all 7 groups. The first group comprises the differential WBC counts (BA, MO, NE, eosinophil; EO, and LY), which itemizes the number of basophils, monocytes, neutrophils, eosinophils, and lymphocytes among present WBCs. The second group (RDW, MCHC, MCV, RBC, and MCH) comprises tests that are used to measure the actual number of RBCs and their physical characteristics. The third group (BE, PCO2, and PO2) consists of blood gas components and focuses on oxygen and carbon dioxide pressure as well as excess or deficit of base levels in the blood. Tbil, ALT, AST, and ALK in the fourth group are liver enzymes (Gowda et al. 2009) that are ordered when a patient is suffering from or showing symptoms of a liver-related comorbidity. BUN and Cr mainly focus on kidney function. Bicarbonate; HCO3 and chloride; Cl are the primary measured anions in the blood. PT along with PTT are used for investigating hemostasis and are the starting points for looking into potential bleeding or clotting complications. Therefore, the presence of a clinical variable in a patient profile can represent a comorbidity in the patient. Although LTs are mainly ordered for diagnostic and prognostic reasons, studies have shown widely diverse test-ordering behavior among clinicians for similar symptoms (Wennberg 1984; Daniels et al. 1977; Solomon et al. 1998). Therefore, indicators

could also reflect the opinions, preconceptions, and biases of the treating clinicians. In other words, by using the missingness indicators, we are learning from practice patterns rather than physiologic patterns. Therefore, indicators as introduced in this study can then be used for modeling health care process in various applications such as clinical care, clinical research, health care economics, and health care policy (Agniel et al. 2018; Sharafoddini et al. 2018).

Filter methods verified the importance of some indicators with regard to our outcomes. Results also demonstrated that indicators become more and more important on ICU days 2 and 3 (Tables 2-2 and 2-3). This observation aligns with clinical practice in which ICU clinicians might try to get a complete dataset on day 1 to fully investigate the patient and understand the situation but are likely to be more selective with LT ordering on subsequent days. The Lac indicator was associated with 30-day and in-hospital mortality on the second and third day. Lactate is usually used as a biomarker for shock states. The literature has constantly reported an association between lactate levels and mortality rates among critically ill patients (Zhang et al. 2014c). Our study demonstrated that just the presence of this information could represent the severity of a patient's illness, as patients with profound shock have a very high mortality rate in hospitals and ICUs (Levinson et al. 2011). Moreover, BUN (Beier et al. 2011; Cauthen et al. 2008; Kajimoto et al. 2015), RDW (Bazick et al. 2011; Hunziker et al. 2012; Patel et al. 2010; Purtle et al. 2014; Senol et al. 2013) and AG (Ahn et al. 2014; Kim et al. 2017; Lee et al. 2016; Sahu et al. 2006) have been repeatedly determined as a risk factor of all-cause mortality and their indicators received a high score in our

analysis. These results are consistent with those of Agniel et al's (2018) who demonstrated that the presence of these tests have significant association with odds of 3-years survival.

The LASSO model selected indicators among the clinical predictors of in-hospital mortality and 30-day mortality, implying the predictive power of indicators. More indicators than clinical variables were selected on the third day (60%-70% of selected predictors were indicators); the assessment demonstrates that indicators from the third day are more informative than those from the first, again supporting the idea that the practice patterns diverge later during ICU stays, so there is more variability in what gets measured. In other words, care on the first day is likely to be highly protocolized—all patients get the same tests regardless of their condition because their trajectory is still unclear. As time goes on, the patterns become more evident and ordering and prescribing practices change according to clinical need. This high percentage of selected indicators suggests that clinical variables are not always required in outcome prediction; instead, information about their presence would suffice.

The last assessment demonstrated that models trained on indicators alone in some scenarios have reasonable performance (for instance, in 30-day mortality prediction with LR, we achieved AUROC of 0.6836 [0.012]). Reasonable performance in this study was defined as the AUROC above 0.6. These results imply that by considering missing data as noise or a random artifact, we can lose valuable information about patient outcomes. Moreover, indicators improved the AUROCs in most scenarios. Researchers in this field are looking for

predictors that can be included in the models to improve the prediction results. Having a low-dimensional set of typical predictors plus these missing data indicators can actually lead to performance comparable with that achieved using typical predictors plus other potentially useful predictors identified a priori by medical researchers: First, in comparison with including extra numeric predictors, the computational load for performing mathematical calculations on binary values such as indicators is usually less. Second, binary data require less computational memory than numbers when performing data mining techniques. Finally, for some important clinical variables, storing the missing data indicators instead of the actual value better protects patient privacy while preserving predictive power. In other words, less privacy concern is expected in a situation when the type of test is disclosed rather than the actual test result. The comparative analyses on the predictive models showed that missing data indicators could improve the prediction models' performance (please refer to Tables C4 and C5 in Appendix C). Although literature considers a small increase (0.01) in AUROC meaningful and of clinical interest (because of insensitivity of AUROC) (Martens et al. 2016; Pencina et al. 2012), including the indicators in our study could improve the average AUROC by 0.0511. Thus, missing data indicators can be introduced as informative predictors and be used to learn from. In other words, these indicators can be representative of physicians' and patients' opinions during the health care process. Furthermore, the overall model performance decreased over time perhaps implying that patients' data on the first 24-hour has the highest level of information. The same pattern was also observed in the previous

assessment. According to these observations, we can infer that presence or absence of a variable can be used in predicting patients' severity of illness.

2.5.2 Strengths and Limitations of the Project

A significant strength of this study is its new insight on missing data in a real-world ICU database. The results confirm the predictive power of some indicators and their advantage over actual values in predictive modeling. The findings further clarify the factors associated with lack of data collection such as the healthier status of a patient or practice patterns of clinicians. These insights, in turn, can be used to design models that consider missing data and benefit from the hidden information. On the basis of our results, missingness indicators can be introduced as potential predictors of ICU patients' outcomes.

Despite the strength, significance, and novel nature of this study, there also exist limitations that cannot be overlooked. First, because of the nature of ICUs, the amount of missing data in MIMIC is less than that from a general ward. Therefore, our study may not fully demonstrate the informativeness of these indicators. Moreover, adding the indicators of interest to the actual data matrix increases the dimension of the matrix and may become computationally burdensome. Using other imputation methods, the power of missing data indicators may vary but this was beyond the scope of our study, which focused on providing evidence on missing data informativeness.

2.5.3 Perspectives for Future Work

Although our study demonstrates that missingness indicators are informative and have predictive power in mortality prediction in ICU, further studies are required to investigate their power in predicting other clinical outcomes. Future researchers can investigate the association between missingness patterns and patient diagnosis. They can also consider more sensitive criteria such as net reclassification or integrated discrimination improvements while preserving improvement in the AUROC as the first criterion. Moreover, as this study looked at the 3 days in the ICU independently, one can investigate if the missing data on a particular day are still informative given all the clinical and indicator variables from previous days. These future studies should also investigate the effect of missing rate on the predictive power of indicators. Another area of future work is examining the test-ordering behavior among clinicians, by using missingness indicators.

While prediction uses data to guess a value, estimation utilizes data to guess parameters. This study focused on capabilities of missing indicators in prediction, however, they can result in bias in parameter estimation. Therefore, an open area of research is to investigate approaches for incorporating missing indicators in a way that better prediction performance can be achieved while minimizing bias in parameter estimation. Last but not least, future researchers can incorporate administrative data into their analyses. These data cover information about physician services, hospital services and can be used with missing indicators to understand the care practices in the hospital.

2.5.4 Conclusions

Our study has demonstrated that the missingness of data itself might be informative in ICU and might have added predictive value beyond observed data alone. Moreover, indicators for variables with higher missingness rates had more predictive power. In practice, the lack of a set of symptoms might lead health professionals to conclude that a particular set of tests is not required at the current stage. Therefore, these missing data are not a random occurrence. This study showed that the number of comorbidities is associated with a decreased rate of missing data. Therefore, rudimentary treatments of missing data (eg, CCA) can cause bias toward sicker patients. The study is also notable because it provided new insight about the informativeness of missing data and described how this information could be used in predicting mortality.

2.5.5 Acknowledgments

This study was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (RGPIN-2014-04743, RGPIN-2014-05911) and Early Researcher Award (Ministry of Research and Innovation, Government of Ontario).

2.5.6 Authors' Contributions

Study conception and design were conducted by AS, JAD, DMM, and JL. AS extracted data and performed the data analysis. Interpretation of the results was provided by all authors. All authors contributed in writing the paper and approved the final version of the review.

2.5.7 Conflicts of Interest

None declared.

Chapter 3

Finding Similar Patient Subpopulations in the ICU Using Laboratory Test Ordering Patterns

This chapter utilizes the missing indicators introduced in Chapter 2 to capture the laboratory test ordering patterns in ICU. Then these patterns are utilized to identify patient subpopulations in ICUs. This chapter was originally published in June 2018 and revisions have been made to the current copy based on the thesis committee's reviews. The full citation is as follows: Sharafoddini, A., Dubin, J. A., & Lee, J. (2018). Finding Similar Patient Subpopulations in the ICU Using Laboratory Test Ordering Patterns. Paper presented at the Proceedings of the 2018 7th International Conference on Bioinformatics and Biomedical Science, Shenzhen, China.

3.1 Abstract

In this paper, we focus on phenotyping critically ill patients in intensive care units (ICUs). Various data types have been used to cluster patients. We introduce laboratory-test-ordering patterns as a source of information for finding clinically similar patients. We employed Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering method to find patient subpopulations based on the first 24 hours of laboratory test ordered. The

DBSCAN identified 25 clusters, and we utilized t-Distributed Stochastic Neighbor Embedding (t-SNE) to visualize the subpopulations. Then, we evaluated the clinical interpretability of the clusters by using cluster characteristics and two outcomes: in-hospital mortality and 30-days post-discharge mortality. Our results demonstrate that laboratory-test-ordering patterns are informative and can be used to identify patient subpopulations.

Keywords: Phenotyping; Intensive Care Unit (ICU); DBSCAN clustering; t-SNE

3.2 Introduction

Intensive care units (ICUs) provide care to acutely ill patients and were primarily introduced in the 1950s (Rodriguez 2001a; Reisner-Senelar 2011). These units target diverse critically ill patient populations and that close monitoring of these patients has generated an enormous amount of data (Johnson et al. 2016a). Although ICUs have a higher number of staff than other departments (Johnson et al. 2016a), analysis and interpretation of this amount of data is challenging for clinicians and must be handled by data analysis methods. ICUs have a heterogeneous population with various health status dynamics and similar needs for constant care (Prin et al. 2012; 'Critical Care Statistics'). This heterogeneity adds to the importance of finding similar patients and detecting the underlying phenotypic groups. Recently, efforts have been made to phenotype patients, a term widely used in the literature with various meanings (Robinson 2012). We focus on phenotyping as in the study of identifying subpopulations of patients, suggested in (Shivade et al. 2014). Finding precise phenotypes from population-scale electronic health records is a core task in developing precision

medicine (Lasko et al. 2013). Traditionally, the task of phenotype discovery was based on one specific question using supervised learning; for instance, stratifying patients into five predefined heart failure risk levels. Although these methods were successful for decades, they are limited to only a set of predefined phenotypes and cannot help when the goal is new phenotype discovery (Lasko et al. 2013). Recently unsupervised learning methods (clustering) have been used to discover phenotypes from data. Many research groups have utilized these methods on varied information in EHRs data, such as demographics, vital signs (Pimentel et al. 2013), laboratory test results, and discharge summaries (Dai et al. 2017b), to identify patient subpopulations (Shivade et al. 2014). Owing to the sensitive nature of patient data, most of these groups have developed their own methods, resulting in a lack of standard tools (Shivade et al. 2014).

Laboratory testing is a fundamental part of day-to-day practice in ICUs and it supports 70% of the decision making in medicine (either for diagnosis or treatment) (Cadogan et al. 2015). Laboratory tests may be ordered for various purposes, including diagnosis, treatment monitoring or severity scoring. However, the decision to order a test is influenced by many hospital-, caregiver-, patient- or disease-related factors. For instance, it has been observed that older and younger patients received less testing in comparison to middle-age patients in ICUs (Zimmerman et al. 1997) or more complete laboratory testing was performed for severely ill patients (Weiskopf et al. 2013b; Rusanov et al. 2014). In this work, we propose a data-driven framework for discovering patient subpopulations, using the laboratory test

ordering patterns from the EHR data. The presence or absence of a laboratory test may be a valuable indicator of patients' characteristics (age group or ethnicity) or health status (symptoms or severity of illness). Therefore, such phenotype discovery can shed light on some characteristics, possibly even latent characteristics, of each subpopulation. Moreover, these phenotypes can help in customizing predictive modeling by training a model for each subpopulation instead of using a one-size-fits-all model.

3.3 Methodology

3.3.1 Data and Data Representation

This study utilizes the freely accessible MIMIC-III (Medical Information Mart for Intensive Care) database (Johnson et al. 2016b)—an update of the MIMIC-II database—which was released in 2015. This database has detailed information on 38,597 distinct patients in the critical care units at the Beth Israel Deaconess Medical Center in Boston, between 2001 and 2012 (Johnson et al. 2016b). Adult patients (age>=15 years) admitted to one of the five following ICUs were included in this study: medical ICU (MICU), surgical ICU (SICU), cardiac care unit (CCU), cardiac surgery recovery unit (CSRU), and trauma surgical ICU (TSICU). Therefore, we focused on the top 80% laboratory tests and profiles common to all ICUs (Frassica 2003) which were available in our database. Data for patients having at least one measurement for any of the following 36 laboratory tests in the first 24 hours of admission was extracted: Alanine Aminotransferase (ALT); Alkaline Phosphatase (ALP); Aspartate Aminotransferase (AST); Arterial blood gases: pH, PCO2, PO2, Base Excess (BE);

Basic metabolic panel: Sodium (Na), Potassium (K), Chloride (Cl), Bicarbonate (HCO3), Anion Gap (AG), Blood Glucose (BG), Blood Urea Nitrogen (BUN), Creatinine (Cr); Complete blood count: WBC, RBC, HGB, HCT, MCV, MCH, MCHC, RDW, Platelet count (PLT), Absolute MO no., Absolute EO no., Absolute BA no., Absolute LY no., Absolute NE no.; Lactate (Lac); Calcium (Ca); Magnesium (Mg); Phosphorus/ Phosphate (Phos); Partial Thromboplastin Time (PTT); Prothrombin Time (PT); Total Bilirubin (TBil). Therefore, whether a laboratory test was performed at least once in the first 24 hours constructs the laboratory ordering test patterns.

3.3.2 Clustering

Various clustering methods have been widely used to find subpopulations. However, many of these methods focus only on spherical-shaped clusters and are sensitive to the presence of noise or outliers (Ester et al. 1996). However, these limitations are more highlighted when working with EHR data in which information inaccuracy is frequently observed (Botsis et al. 2010). DBSCAN (Ester et al. 1996), one of the most common clustering methods, utilizes the idea that the clusters are located where data have a high density and are separated by regions with a lower density of data (the density is evaluated based on two user-defined parameters). Unlike in other methods, clusters in DBSCAN can have any arbitrary shape and the algorithm is not sensitive to noise. DBSCAN received the ACM SIGKDD Test of Time Award due to its outstanding influence in data mining.

In this study, we will use DBSCAN to cluster the laboratory test ordering patterns, as it does not require the number of clusters to be set. However, it is important to acknowledge that the DBSCAN algorithm fails to identify clusters if density varies. Since DBSCAN uses a single global parameter epsilon (ε) to identify clusters, it is impossible to detect clusters with varied densities simultaneously. The only two parameters that must be tuned for DBSCAN are ε and minimum points (MinPts). MinPts (the minimum number of neighbors required to consider a point as a core point) is usually set to the dimensionality of data pulse one or higher. After setting MinPts, epsilon (the radius of the neighbourhood) can be set using k-distance tuning method. In this method, first, the distance of each point from its k'th nearest neighbor is calculated. This distance is called kdist. After calculating and sorting kdists for all data points, the distance values will be plotted in a k-distance graph for a particular value of k. Then, the value for which the graph shows a strong bend—the knee point—will be chosen for epsilon.

The DBSCAN algorithm also requires a distance metric to cluster patients. Since our data are binary, we use the Jaccard distance (Deza et al. 2009) to calculate the similarity between points.

3.3.3 Graphical Representation

To visualize the data set, a dimensionality reduction algorithm, which creates two-dimensional visualization of all data points is required. We have chosen t-Distributed Stochastic Neighbor Embedding (t-SNE) (Maaten et al. 2008), a dimensionality reduction

technique that is used for mapping high dimensional datasets into two-dimensional space in order to see the data structure. Numerous nonlinear dimensionality reduction methods have been introduced for data visualization while preserving the structure of data. However, t-SNE has been shown to perform better in preserving the structure of data than other widely used techniques (Maaten et al. 2008) such as Sammon mapping, curvilinear components analysis, Stochastic Neighbor Embedding, Isomap, Maximum Variance Unfolding, Locally Linear Embedding, and Laplacian Eigenmaps.

Here, t-SNE starts with converting the pair-wise Jaccard distances (d) of a data point into probabilities (p_{ji}) that represent the probability that a data point (x_i) will choose (x_j) as its neighbor. For close data points, this probability is relatively high. Therefore, first, it centers a Gaussian distribution over x_i and measures the density of other points under this Gaussian distribution. The joint probability of p_{ji} is calculated using $p_{ji} = (p_{j|i} + p_{i|j})/2N$, where

$$P_{j|i} = \frac{\exp\left(-\frac{d(x_i - x_j)^2}{2\sigma_i^2}\right)}{\sum_{k \neq i} \exp\left(-\frac{d(x_i - x_k)^2}{2\sigma_i^2}\right)} \tag{3.1}$$

Similar conditional probability can be defined for the low-dimensional mappings of x_i and x_j, denoted as y_i and y_j. This probability measures the similarities of data points in low-dimensional space. t-SNE utilizes a heavy-tailed Student-t distribution with one degree of freedom for low-dimensional spaces since it wants the dissimilar points to be too far apart in the map.

$$q_{j|i} = \frac{\left(1 + d(y_i - y_j)^2\right)^{-1}}{\sum_{k \neq i}(1 + d(y_i - y_k)^2)^{-1}} \quad (3.2)$$

The goal is for $q_{j|i}$ to reflect $p_{j|i}$ as well as possible. t-SNE uses Kullback-Leiber divergence to measure the difference between these probabilities. Therefore, it moves the data points to minimize $KL(P||Q) = \sum_i \sum_{j \neq i} p_{ij} \log \frac{p_{ij}}{q_{ij}}$. Using gradient descent, the result is a map that reflects data points in a low-dimensional space while preserving their local structure.

3.3.4 Evaluation Design

One of the most challenging parts in cluster analysis as an exploratory analysis is the evaluation of results. In this work we employed two evaluation methods: (i) exploring the clinical characteristics of the clusters; and (ii) the *silhouette index* (Rousseeuw 1987), which evaluates the suitability of assigning a patient to a group rather than to another by considering cluster cohesion and cluster separation. Therefore, we employed silhouette index, which is a number between -1 and 1, to evaluate our clustering. Besides the total silhouette value for our proposed method, we calculated the silhouette index for each cluster by averaging the silhouette widths for the patients in that cluster. Silhouette values close to one indicate precise clustering, while small values close to zero represent observations that lie between two clusters. Negative values represent wrong patient assignment.

3.4 Results

Based on the inclusion criteria, 32618 patients were retrieved from the MIMIC-III ICU database described in Section 3.3.1. To determine the parameters of DBSCAN, after testing various values for MinPts (including 37, 50, 100, 150, 200, 300, 400, 500) and observing little variation in clustering, we set MinPts to 37, which equals the dimensionality of data plus one. Then using the k-distance graph the radius associated with the sharp change was chosen as ε. Figure 3-1 shows the k-distance graph for k=37. The value of ε determined from this graph was 0.125.

Utilizing DBSCAN with tuned parameters on the laboratory test ordering data resulted in 25 clusters, with some data points as outliers (cluster 0).

We represented each of the 32618 critically ill patients in a two-dimensional space using t-SNE. Although the author of t-SNE has claimed that this algorithm is not very sensitive to the perplexity parameter (Maaten et al. 2008), there is evidence showing otherwise (Wattenberg et al. 2016). Therefore, first, we tested representing data with various values for perplexity: changing perplexity from 5 to 50 (as suggested by (Maaten et al. 2008)) and then testing it for higher values of 100, 150, and 200. Robust behaviors were observed from the topological aspect. However, the relative size of clusters and distance between them were unstable and meaningless, as was expected from literature (Wattenberg et al. 2016).

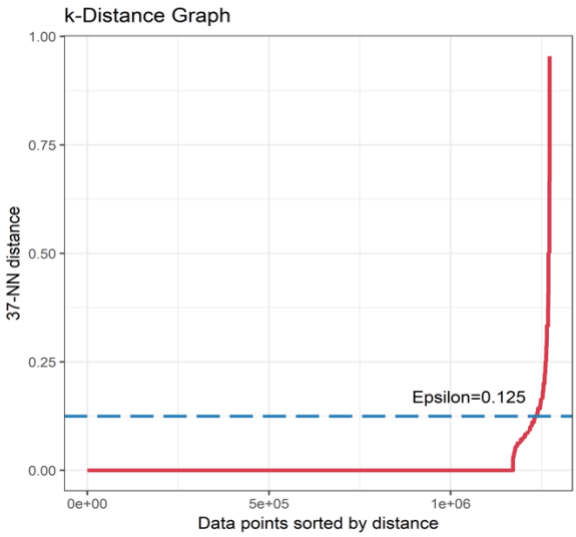

Figure 3-1 k-Distance graph (k=37) for tuning epsilon in DBSCAN.

Figure 3-2 demonstrates the two-dimensional representation of patient data points with colored cluster assignment. The cluster populations range from 39 to 10486 patients. The in-hospital mortality rate was from 0.83 to 27.95 while the 30-days hospital mortality rate was higher ranging from 1.65 to 33.29. The highest mortality rate was in cluster number 2 (light green), in which 92% of patients had an emergency admission.

This cluster had the highest 30-days post-discharge mortality rate as well. We employed the Simplified Acute Physiology Score (SAPS-II) (Le Gall et al. 1993)—a well-known and widely-used ICU severity scoring system—to report the illness severity in each cluster. In Cluster #2, 49.2% of the patients received a SAPS-II score of 49, which equals the maximum SAPS-II score observed in the whole dataset. Interestingly, all the 36 laboratory tests listed above were ordered for all patients in this cluster, implying the severity of illness in this

cohort which resulted in higher mortality rate. Cluster #19 had the lowest mortality rates. The average number of tests per patient for this cluster was 3.4 (median=2, sd= 3.78). The second and third clusters with highest mortality rates were clusters #7 and 11, for which the average numbers of tests per patient were 35 and 31, respectively. The average age in clusters ranges from 49.21 to 67.44. The mortality rate in the younger cohort was approximately half of the mortality rate in the older cohort. Length of stay (LOS) ranges from 6.62 to 16.42, with a higher mortality rate associated with long LOS.

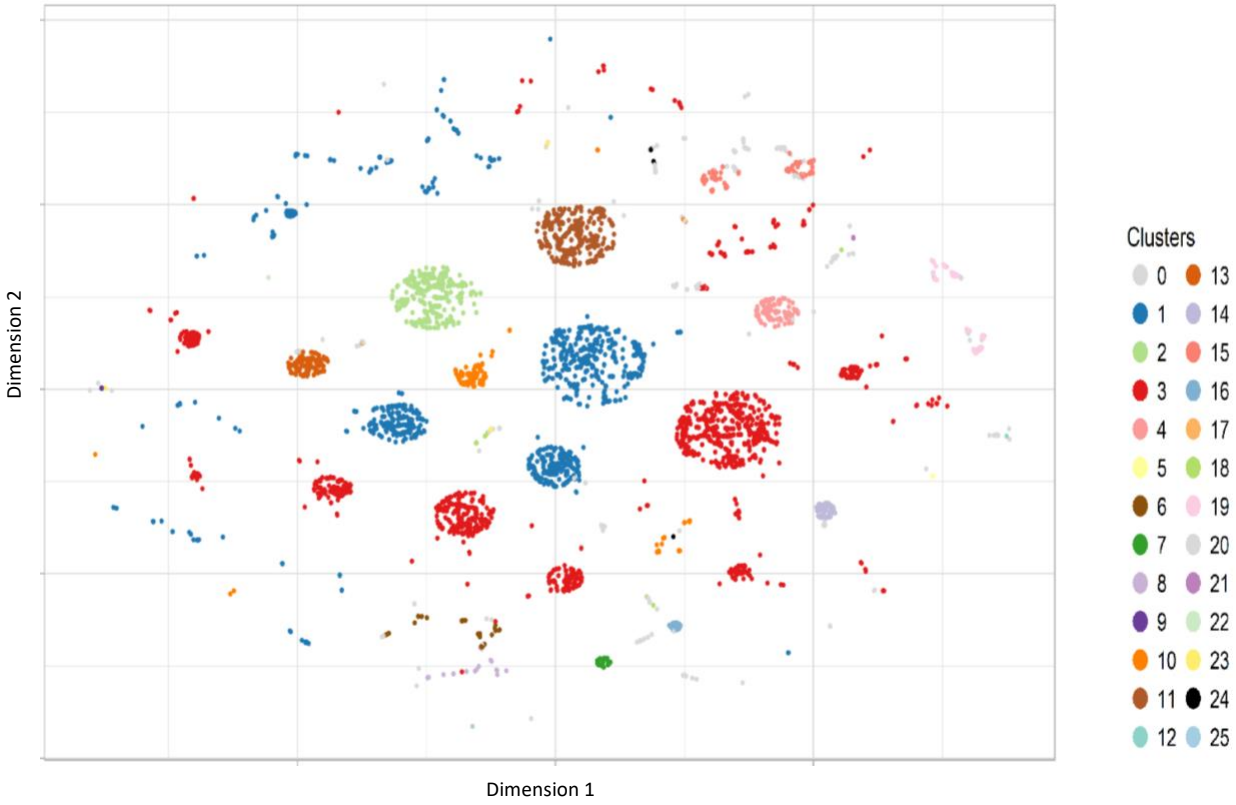

Figure 3-2 Clustering critically ill patients using laboratory test ordering patterns. DBSCAN clustering method with MinPts=37 and ε =0.125 was used. Data were mapped to a two-dimensional space using t-SNE with perplexity=30 and 1000 iteration.

For each cluster, the population, gender distribution, average age, mortality rates, average number of tests performed per patient and the dominant primary ICD-9 codes for cluster are shown in Table 3-1.

Table 3-1 Population characteristics for each cluster.

#	Size	Gender (%M)	Age	Mortality (%)		LOS	Average number of tests per patient	Dominant ICD-9 diagnosis code
				In-Hospital	30-days			
0	1096	54.74	63.42	13.78	18.89	12.13	28.16	41401 - Coronary atherosclerosis of native coronary artery
1	8754	54.31	63.54	5.37	9.71	8.03	22.47	41071 - Subendocardial infarction, initial episode of care
2	2472	55.66	61.53	27.95	33.29	14.52	36	0389 - Unspecified septicemia
3	10486	60.91	64.73	8.27	11.18	10.98	25.27	41401 - Coronary atherosclerosis of native coronary artery
4	1065	52.49	63.85	15.4	20.19	13.11	32	0389 - Unspecified septicemia
5	237	48.95	64.75	10.13	21.52	9.9	32	0389 - Unspecified septicemia
6	260	63.85	65.38	7.31	8.85	12.14	28.32	41401 - Coronary atherosclerosis of native coronary artery
7	436	54.82	62.69	22.02	26.83	13.48	35	51881 - Acute respiratory failure
8	346	52.89	61.13	11.27	14.45	10.5	28.5	0389 - Unspecified septicemia

9	146	43.15	64.73	13.7	20.55	11.12	30	0389 - Unspecified septicemia
10	1444	52.29	65.9	5.89	12.12	8.36	26.84	0389 - Unspecified septicemia
11	2301	59.76	60.95	21.82	25.9	15.47	31	0389 - Unspecified septicemia
12	90	50	64.5	13.33	22.22	12.75	34	0389 - Unspecified septicemia
13	1015	56.65	62.45	9.36	15.86	9.52	31	41071 - Subendocardial infarction, initial episode of care
14	590	57.8	61.96	19.66	24.58	12.98	30	41071 - Subendocardial infarction, initial episode of care
15	377	53.85	63.88	9.81	13.26	9.37	14.66	41071 - Subendocardial infarction, initial episode of care
16	370	46.22	65	15.95	20.54	11.41	31	51881 - Acute respiratory failure
17	39	61.54	49.21	7.69	7.69	6.62	10.28	41519 - Other pulmonary embolism and infarction
18	121	63.64	66.44	0.83	1.65	9.36	28.87	41401 - Coronary atherosclerosis of native coronary artery
19	467	51.18	62.75	19.49	23.34	7.26	3.4	41401 - Coronary atherosclerosis of native coronary artery
20	70	42.86	61.63	4.29	8.57	11.08	33	03842 - Septicemia due to escherichia coli

21	145	75.86	66.76	2.07	4.14	9.69	30	41401 - Coronary atherosclerosis of native coronary artery
22	53	58.49	59.95	16.98	22.64	12.14	32	20500 - Acute myeloid leukemia, without mention of having achieved remission
23	103	41.75	64.49	18.45	20.39	8.78	28.85	51881 - Acute respiratory failure
24	94	55.32	61.53	5.32	8.51	9.33	28.83	486 - Pneumonia, organism unspecified
25	41	60.98	67.44	12.2	12.2	16.42	30	41401 - Coronary atherosclerosis of native coronary artery

The total silhouette value was 0.81. Evaluating individual clusters, we observed that except for the group of outliers which received a silhouette value of -0.56, the average silhouette value for all clusters were positive. Figure 3-3 demonstrates the average silhouette values for all clusters. The interpretations of silhouette values are color coded based on the ranges proposed in (Kaufman et al. 2009). Green indicates that a strong structure has been found for that particular cluster. Yellow and red indicate reasonable and weak structures, respectively. Based on this interpretation, only three clusters had weak structure and alternative methods are suggested for them (cluster #6, 15, 18). Other clusters had acceptable structures. These results support the applicability of test-ordering patterns for patient phenotyping.

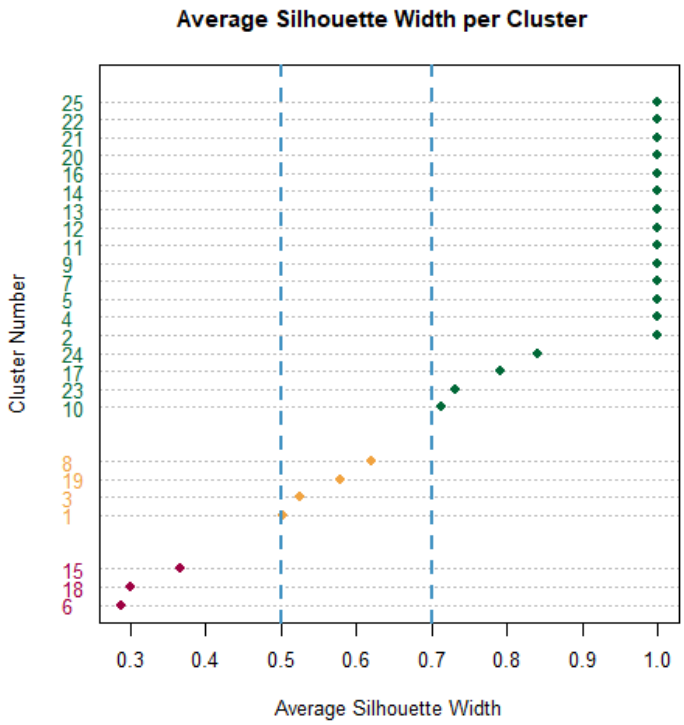

Figure 3-3 The average silhouette values for each cluster. Green, yellow and red respectively represent strong, reasonable and weak structures.

3.5 Discussion

Our results strongly support the introduction of lab ordering test as a source of information for phenotyping patients. Our approach can cluster patients into subpopulation following two rationales.

First, lab ordering patterns are informative about physician's opinion and biases. Various factors affect physicians' decisions to order a test. One of these factors is whether a patient is

showing a symptom or not (Litchfleld et al. 2014; Wells et al. 2013). Therefore, these binary data can easily be used as an indicator of patients' symptoms.

Second, lab ordering patterns are informative about patient illness severity. Intensive care physicians tend to order more tests for patients who are severely ill, as they need to monitor them closely and comprehensively. Therefore, the number of tests being ordered can be associated with the severity of illness.

These facts are well represented in our results. Clusters #2, 7 and 11 had higher mortality rates as an implication of severe health conditions. In these clusters, more than 30 lab tests (out of 36) were performed on average per patient.

Our results are of interest from two aspects. While many phenotyping methods use huge amounts of data to find subpopulations, our method works in a 36-dimensional binary space. This will result in less computation burden and ease of implementing the method on personal computers. One trending research area is predicting modeling in health care. While many researchers are investing on one-fits-all models, a hierarchical approach where patients are first grouped into subpopulations and then models are trained on each group has been studied and showed promising results. Our approach can also contribute to this field as a fast and memory efficient way of subgrouping patients for predictive modelling.

There are some limitations to this study. First, we limited our laboratory tests to the most common ones in ICUs and their availability in our database; therefore, abundant room remains for considering more-comprehensive lists. Second, this study focused on the

laboratory-test patterns of the first 24 hours of ICU admission; however, including data from all days of an ICU stay can provide information about the evolution of patient health status.

3.6 Conclusion

In this paper, we used laboratory test ordering patterns to discover subpopulations among critically ill patients. We used DBSCAN to reveal these subpopulations. t-SNE was utilized to represent results in a low-dimensional space. Our results demonstrate that lab ordering patterns are able to reveal information about patient health status and can be used to identify clinically meaningful subpopulations. Future researchers can focus on employing weighted distance for eliminating the effect of the lab tests that are routinely ordered as part of hospital protocols and highlighting laboratory tests that are more patient-specific and representative of their special health condition. These patterns can also be used with other methods to increase the performance of phenotyping patients.

3.7 Acknowledgments

This work was supported by NSERC Discovery Grants (RGPIN-2014-04743, RGPIN-2014-05911) and an Early Researcher Award from the Ontario Ministry of Research, Innovation and Science.

Chapter 4

Multifaceted Patient Similarity Metric

This chapter builds upon the previous chapters and introduces a multifaceted PS metric that accounts for various EHR data and integrate different similarity metrics. This metric also considers the similarity in terms of missing data using the indicators introduced in Chapter 2. Various experiments are performed to evaluate the performance of the proposed metric in comparison to the conventional methods in mortality prediction application.

4.1 Introduction

In the past decade, with the emergence of precision medicine, patient similarity (PS) analytics has received special attention and become the core component of health analytics in, among other areas, predictive modeling, patient stratification, and clinical pathway analytics (Hu et al. 2016a). EHR data are present in a plethora of formats, including demographic data, vital signs, lab results, diagnosis/symptom/procedural codes, and clinicians' notes, and reflect information about different aspects of patient health status. The volume of existing EHR data makes it difficult for physicians to assess the similarity of two patients. Therefore, health analytics has focused on devising similarity metrics, using various approaches (Jurisica et al. 1998; Bobrowski 2006; Park et al. 2006; Saeed et al. 2006a; Chattopadhyay et al. 2008; Han et al. 2015; Sun et al. 2010b; Sun et al. 2010a; David et al. 2011; Wang et al. 2011; Campillo-Gimenez et al. 2013; Lowsky et al. 2013; Hielscher et al.

2014; Henriques et al. 2015; Lee et al. 2015a; Wang et al. 2015). Although studies have shown the utility of using PS in health analytics, assessing PS using one similarity metric for all the various clinical variables/predictors may not sufficiently capture the similarities.

The ICD coding system is a widely used coding system for classifying diagnoses and procedures. MIMIC-III contains the ninth version of ICD codes (ICD-9 codes) for diagnosis and procedures. The ICD-9 coding system consists of a set of trees that represent the hierarchical relation between the codes. Whereas, higher-level nodes represent more general concepts, lower-level nodes are more detailed. Figure 4-1 shows a snippet of a hierarchy tree.

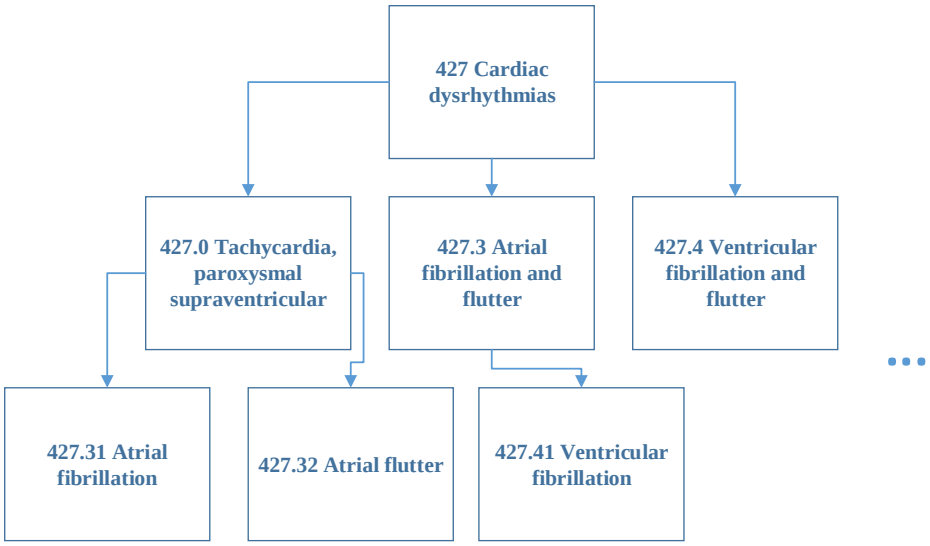

Figure 4-1 A snippet of an ICD-9 hierarchy tree.

ICD-9 codes have been used in healthcare analytics for various applications, including 30 days post-discharge mortality prediction (Lee et al. 2015a). However, the ICD-9 codes are usually treated as a categorical variable, and the hierarchal nature of the codes is ignored.

The hierarchical nature of the ICD-9 code can be a potentially valuable source of information. For instance, considering the hierarchical nature of ICD-9 codes has led to better automated ICD-9 code assignment from discharge summaries (Perotte et al. 2014). Moreover, leveraging the hierarchy in the feature construction or model building or both significantly improved predictive modeling for chronic kidney disease and heart failure applications (Singh et al. 2014). Although Gottlieb et al. (Gottlieb et al. 2013) considered the hierarchy in their similarity assessment, the superiority of it over conventional methods remained undetermined. A recent study has also investigated the effect of leveraging the hierarchy in the similarity assessment. Girardi et al. (Girardi et al. 2016) proposed a new semantic similarity metric for ICD-10 codes in which the hierarchical characteristic was taken into consideration. They proposed the following measure for assessing the distance between two ICD codes in a hierarchy tree:

$$dist_{ICD}(ICD_1, ICD_2) = \frac{MinEdge(ICD_1, ICD_2)}{p(ICD_1) + p(ICD_2)} \quad (4.1)$$

where MinEdge is the minimum number of edges between two ICD codes, and $p(ICD_1) + p(ICD_2)$ denotes the longest path between the codes in the tree. They compared the proposed method to the Jaccard distance—a measure of the distance between sets—and Haase-Li (Haase et al. 2004; Yuhua et al. 2003) distance—a measure of the distance between individual hierarchical codes—, and it demonstrated better performance at detecting similarities.

Therefore, constructing a similarity score by utilizing appropriate similarity metrics for specific clinical variables can contribute to more accurate PS assessment and data retrieval in clinical care.

4.1.1 Literature review

Overall, various types of similarity metrics can be used in computing the similarity between patients.

- Distance-based similarity metrics: Absolute distance (Chattopadhyay et al. 2008), Euclidean distance and its family (Bobrowski 2006; Henriques et al. 2015; David et al. 2011; Park et al. 2006; Hielscher et al. 2014), Mahalanobis distance and its family (Sun et al. 2010b; David et al. 2011; Wang et al. 2012a; Wang et al. 2015; Lowsky et al. 2013; Han et al. 2015; Sun et al. 2010a).
- Correlation-based similarity metrics (Saeed et al. 2006a)
- Cosine-based similarity metrics (Lee et al. 2015a)
- Model-based similarity metrics (Houeland 2011a; Zhu et al. 2017; Wang 2015; Zhang et al. 2018; Suo et al. 2017)

Most of the current studies in PS analytics utilize one universal metric to assess the similarities among patients. However, researchers have recently focused on capturing multiple aspects of drug similarity (Zhang et al. 2014a; Gottlieb et al. 2011; Li et al. 2012; Moghadam et al. 2016; Yan et al. 2014)—including genetic (target proteins), phenotypic

(side effects) and chemical aspects—and disease similarities (Zhang et al. 2014a; Gottlieb et al. 2011; Li et al. 2016; Moghadam et al. 2016; Yan et al. 2014)—inclusive of disease symptom, ontology and gene aspects—by employing various metrics (Zhang et al. 2014a; Gottlieb et al. 2011) and combining the similarity scores to achieve one unified score.

Although this kind of multifaceted similarity assessment is gaining attention in drug and disease similarities, only a very few studies exist on multifaceted PS assessment using EHR data. Gottlieb et al. (2013) utilized various similarity metrics for different variables:

- ICD codes:
    - ICD similarity:

$$Sim(ICD_1, ICD_2) = \frac{Nearest\ Common\ Ancestor\ (ICD_1, ICD_2)}{\#of\ levels\ in\ ICD\ hierarchy} \qquad (4.2)$$

      This metric uses the levels of ICD codes in the ICD coding hierarchy to calculate the similarity. For instance, for ICD codes 427.31 and 427.41 the nearest common ancestor is 427 (level 3) and the number of levels in the ICD-9 hierarchy is five. Therefore, the similarity between these two codes is 3/5.

    - *Empirical co-occurrences between ICD codes*

      This metric computes the number of co-occurrences of an ICD pair in patient profiles across the dataset as a measure of similarity between two codes.

    - *Bipartite graph over ICD codes*

To calculate the similarity score for two sets of ICD codes, this method constructs the bipartite graph over the codes in the two profiles in which an edge represents the similarity of two codes calculated based on one of the above methods. Then, the final similarity between two sets is calculated using maximal matching in graph theory.

- Blood test and ECG records: Euclidean distance and absolute difference
- Age:

$$Sim(Age_1, Age_2) = \frac{|Age_1 - Age_2|}{\max(Age_1, Age_2)} \quad (4.3)$$

- Sex: XOR distance (returns zero for patients of the same gender and one otherwise)

The authors then used these scores to construct a feature vector for a regression model. Although their study considered various similarity metrics, they did not justify their use of various metrics. In addition, the advantage of this approach was not investigated in comparison to conventional methods. Moreover, regardless of the nature of the variables and frequency of measurement for blood test measurements and ECG timeseries, only the Euclidean distance between the first measurement of each variable was used for similarity calculation. Thus, the unique characteristics of each variable were not the main focus of the study. Using a variable-specific similarity metric can overcome the limitations of using only one particular similarity metric (utilizing just one similarity metric for all predictors may miss information relevant to clinical similarity assessment). The work most relevant to this

study is a system called Advanced Analytics for Information Management (AALIM), devised by the IBM Almaden Research Center for cardiac diagnosis (Syeda-Mahmood et al. 2007; Amir et al. 2010). AALIM computes cardiac PS in each of the following modalities: electrocardiogram (ECG or EKG), heart auscultation sounds and cardiac echo videos. In this study, the authors employed a weighted linear combination method to fuse all the similarity scores into one. However, published evidence on this work covering more details about the methodology is very limited.

4.1.2 Study Objectives

The study presented here is motivated by the need to use the various types of information in EHR data more effectively. The purpose of this study is to devise a multifaceted PS metric in which the characteristics of individual clinical variables are considered. Moreover, in the context of predictive modeling, this study compares the performance of the multifaceted PS metric with that of conventional methods (specifically, cosine and Euclidean distance).

4.2 Materials and Methods

4.2.1 Data and Analytical Dataset Preparation

Data for this study were extracted from the MIMIC-III database since it has a rich variety of EHR data. To manage computational time for the pair-wise similarity calculations and at the same time preserving the heterogeneity of the ICU patients, only adult patients admitted to the medical ICU and surgical ICU were included in this study. To be included, patients must

have had at least one measurement for any of the variables in Table 4-1. PMM imputation were used to address missing data. While imputing a missing value, an indicator was generated to distinguish between an imputed value and an actual measurement (missing measurement indicators). In case of multiple ICU admissions, only the first ICU admission was considered. Moreover, patients must have stayed in the ICU for at least 24 hours, and data for laboratory tests, vital signs, urine output and therapeutic interventions were extracted only from the first day of stay in the ICU. Data extraction was performed using Structured Query Language (SQL) in PostgreSQL. In the MIMIC III database, the birthdate for patients older than 89 years old are shifted to obscure their true age. These patients appear in the dataset with ages of over 300 years, therefore, the age value for these patients was treated as missing data. Moreover, min-max normalizer was used to linearly rescale each predictor to the [0,1] interval.

Table 4-1 Data types in a patient profile.

Category	Variables
Demographics and Admission Information	Age, Gender, Admission type (Emergency, Elective, Urgent), The first ICU type in which the patient was cared for (MICU, SICU)
Laboratory Tests and Urine Output	Albumin, Anion Gap, Bicarbonate, Bilirubin, Creatinine, Chloride, Glucose, Hematocrit, Hemoglobin, Lactate, Platelet, Potassium, PTT, INR, PT, Sodium, Bun, WBC, Urine Output
Vital Signs	Heart Rate Systolic Blood Pressure, Diastolic Blood Pressure, Respiratory Rate, Temperature, SpO_2, Glasgow Coma Score
Discharge code	ICD-9
Therapeutic interventions	Dialysis (Yes/No), Mechanical Ventilation (Yes/No), Fluid (Colloids and Crystalloid) Administration (Yes/No)
Missing Measurement Indicators	Albumin, Anion Gap, Bicarbonate, Bilirubin, Creatinine, Chloride, Glucose, Hematocrit, Hemoglobin, Lactate, Platelet, Potassium, PTT, INR, PT, Sodium, Bun, WBC, Urine Output

4.2.2 Methodology

4.2.2.1 Multifaceted PS Metric

Figure 4-2 shows the overall structure of the methodology of this study. For each new patient profile, its similarity to the profiles of other patients is assessed from different perspectives. In other words, a similarity score is generated for each variable in the profile. Besides the variables' values, the similarity of the patient to others, in terms of missing measurements, is

calculated using the representation introduced in Chapter 2 (Sharafoddini et al. 2019). Then, all these similarity scores are aggregated to build a multifaced PS score.

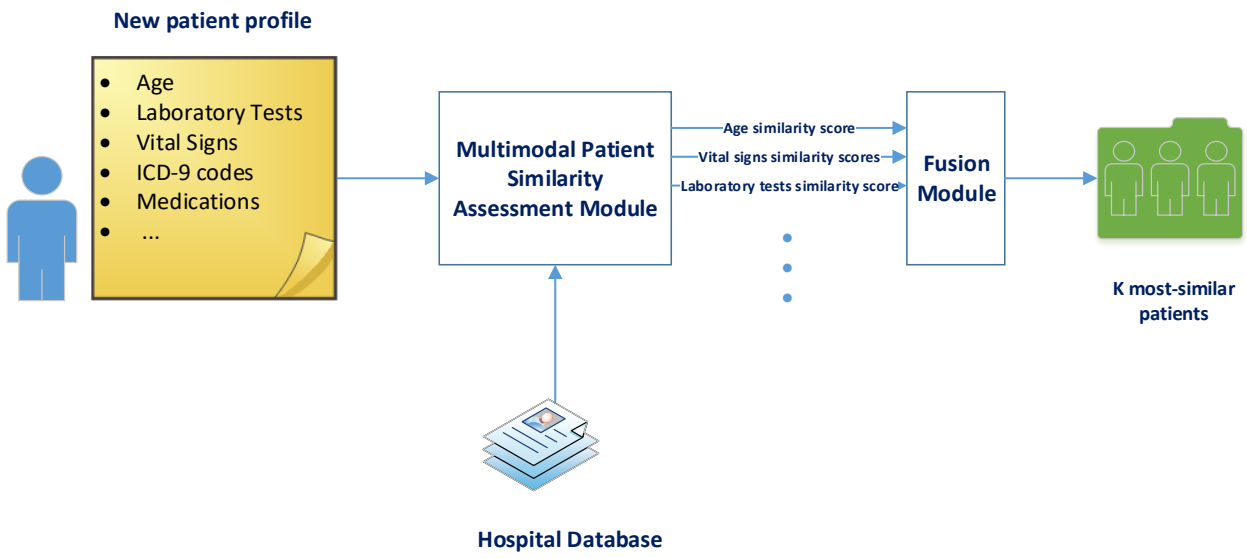

Figure 4-2 The overall structure of the proposed method.

4.2.2.1.1 Individual Similarity Metrics

4.2.2.1.1.1 Demographics and Admission Information

For categorical demographic data the well-known Simple Matching Coefficient (SMC) is used to measure PS (Sokal 1958). SMC is a statistic for calculating the similarity of two sets.

$$SMC = \frac{number\ of\ matching\ predictors}{number\ of\ predictors} \quad (4.4)$$

For age, simple absolute difference was used.

4.2.2.1.1.2 Therapeutic Interventions

Since therapeutic interventions in ICUs are very broad, this study focused only on the presence of the following significant interventions in a patient's profile: mechanical ventilation, fluid administration, and dialysis. These interventions are among the ones that require special nursing service—based on the Simplified Therapeutic Intervention Scoring System (Reis Miranda et al. 1996)—and also have been used in scoring systems for critically ill patients as a marker of their health status (Vincent et al. 2010; Rao et al. 2008). For this category of variables, SMC was used to calculate pair-wise PS.

4.2.2.1.1.3 Laboratory Tests and Urine Output

While in general wards, phlebotomy (the taking of blood samples) is performed on average once a day for each patient, in ICUs this number increases to nearly three to four blood samples drawn per day (Smoller et al. 1986; Low et al. 1995). Therefore, the maximum and minimum values of all laboratory tests in Table 4-1 were extracted per day—since they can represent the worst conditions (Lee et al. 2015a). For urine output, the total amount during the first 24 hours was extracted. Then, the Mahalanobis distance was employed to calculate the similarity of patients in terms of laboratory results and urine output.

4.2.2.1.1.4 ICD-9 Code

This study employed the hierarchical distance introduced by Gottlieb et al. (2013) for calculating the distance between ICD-9 codes.

$$Sim(ICD_1, ICD_2) = \frac{Nearest\ Common\ Ancestor\ (ICD_1, ICD_2)}{5} \qquad (4.5)$$

4.2.2.1.1.5 Vital Signs

Vital signs are collected more frequently for patients than other data. Thus, the maximum and minimum of each vital sign were extracted during non-overlapping 6-hour periods—which exceed the longest acceptable gap between charting vital signs (Cahill 2010) and the commonly-accepted frequency of charting vital signs (Schulman et al. 2010; Miltner et al. 2014; Johnson et al. 2014). Then, the Mahalanobis distance was used to calculate the similarity of patients in terms of vital signs.

4.2.2.1.1.6 Absence Indicators

Absence indicators are binary data, and the literature has introduced many distances for comparing the similarity of two sets of binary values of the same or different lengths (Deza et al. 2009; Choi et al. 2010). In this study, SMC was used for comparing patients in terms of their missing measurements.

4.2.2.2 Combining Similarity Scores

After calculating the individual, variable-specific similarity scores, the next step is to fuse all the scores into one. Similarity scores are based on multiple sources of information and can be seen as distinct pieces of evidence. In this study, after normalizing each metric, simple averaging was used for aggregating all the scores.

Figure 4-3 demonstrates an example of PS calculation by multifaceted PS metrics in the dataset. The similarity of the new patient (ICUStay_Id: 210282) from test set is compared to the profile of a patient in the training set (ICUStay_Id: 243452). The similarity is measured from different perspectives and then all the scores are aggregated into one similarity score. Based on the outcome of the patient on the right and its similarity to the new patient, same outcome is expected for the new patient. Data confirm a negative value for the 30-days post-discharge mortality flag for the new patient.

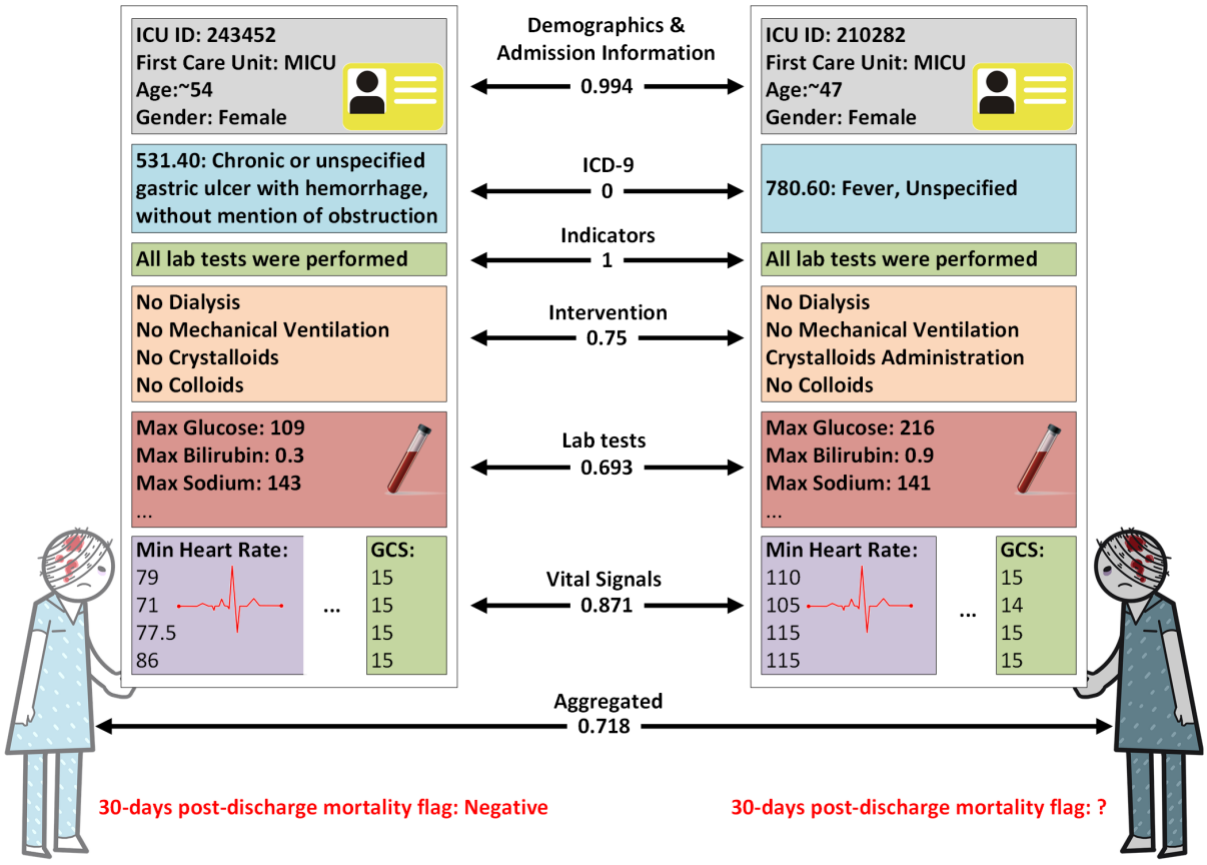

Figure 4-3 A visual representation of multifaceted PS metric calculation.

Table 4-2 shows a summary of calculations performed in Figure 4-6. All numerical variables were normalized and categorical variables were one-hot encoded before calculation.

Table 4-2 An overview of individual PS calculation methodologies and example input vectors with similarity calculations.

Category	PS calculation methodology	Input vectors and similarity calculations
Demographics and Admission Information	1- (SMC+ Absolute Difference)/2	1-(SMC([MICU1, ..., Female1],[MICU2, ..., Female2]) + (|Age1-Age2|))/2
ICD-9	$Sim(ICD_1, ICD_2) = \frac{Nearest\ Common\ Ancestor\ (ICD_1, ICD_2)}{5}$	0/5
Laboratory Tests	1-Mahalanobis distance	1-Mahalanobis ([min_Glucose1, max_Glucose1, ..., max Albumin],[min_Glucose2, max_Glucose2, ..., max Albumin2])
Vital Signs	1-Mahalanobis distance	1-Mahalanobis ([min_GCS1_6hr, max_GCS1_6hr1, ..., max_HR1_6hr4],[min_GCS2_6hr, max_GCS2_6hr1, ..., max_HR2_6hr4])
Indicators	1-SMC	1-SMC([0,0,...,0,0],[0,0,...,0,0])
Intervention	1-SMC	1-SMC([no_Dialysis,no_MV, ...,no_crystalloids], [no_Dialysis,no_MV, ...,crystalloids])

4.2.2.3 Experiments

To benchmark the following experiments, the predictive performance of SAPS-II and SOFA scores were investigated in a 10-fold cross-validation set-up.

4.2.2.3.1 Evaluating the Validity of the Multifaceted PS Calculation

Quantifying the accuracy and correctness of a PS metric in an unsupervised problem is challenging, since subjective evaluation for such a huge number of pair-wise similarity scores seems impossible. This study utilizes the objective evaluation method introduced by Keogh et al. (2003). The idea is to use a nearest neighbor classifier on labelled data to evaluate the accuracy of a PS metric. It has been repeatedly demonstrated that the performance of k-nearest neighbor (KNN) critically depends on the distance metric (Hu et al. 2016b; Wang et al. 2012b; Keogh et al. 2003); therefore, the accuracy of classification is a proxy of metric accuracy. Here, this methodology is performed over various values of K to eliminate the effect of number of similar patients. For each new patient, data from a cohort of similar patients in the training set were retrieved using the proposed multifaceted PS metric; then a prediction was made for a new patient based on a majority vote of the *k* most similar patients. An equivalent approach was used to investigate the performance of the following conventional metrics

- Euclidean distance

$$dist(X,Y) = \sqrt[2]{\frac{\sum_{i=1}^{m}(x_i - y_i)^2}{m}} \tag{4.6}$$

- Cosine similarity

$$sim(X,Y) = \frac{\vec{X}.\vec{Y}}{|\vec{X}||\vec{Y}|} \quad (4.7)$$

In calculating the similarities, all numerical predictors were rescaled to the [0,1] interval and one-hot encoder was used for categorical variables.

4.2.2.3.2 Multifaceted PS metric in predictive modelling

In this experiment, after retrieving a cohort of similar patients, a logistic regression (LR) and decision tree (DT) were trained on the retrieved cohort, and a prediction was made for the new patient. The same procedure was followed using a cosine PS metric. AUROC was used for analyzing the performance of the prediction models in each scenario. Once overall performances are revealed, PS-based scenarios are also compared to a population-based approach, in which prediction models are trained on the whole training dataset without using PS. In all scenarios, 10-fold cross validation was performed.

4.3 Results

4.3.1 Data

From 38,597 distinct adult patients in MIMIC III database, 33,276 patients have LoS≥1 for their first ICU admission. 15,017 patients were excluded since the first ICU type to which they were admitted was not medical ICU or surgical ICU. Finally, after excluding patients who did not have at least one measurement for any of the variables in Table 4-1, 17,547 patients were included in this study. Of these patients, 11,963 were admitted to medical ICU

and 5,583 were admitted to surgical ICU. Moreover, 2,927 patients (16.68%) experienced death within 30 days of discharge.

4.3.2 Evaluating the Validity of the Multifaceted PS Calculation

SAPS-II and SOFA achieved a mean AUROC of 0.764 (sd=0.014) and 0.681 (sd=0.024), respectively.

Figure 4-4 illustrates the AUROC of KNN as a function of *K* (the number of similar patients from the training set). The results confirm our hypothesis that a multifaceted PS metric is more accurate in retrieving similar patients than the conventional PS metrics. The maximum AUROC of 0.69 (SD=0.016) and 0.699 (SD=0.018) for the Euclidean and cosine PS metrics was achieved with 200 and 130 similar patients, respectively. For the multifaceted PS metric, the maximum AUROC of 0.789 (SD=0.015) was achieved with 870 patients. Although the maximum AUROC for the multifaceted PS metric was achieved with a higher value of K, this method outperforms the conventional metrics in all values of K. Appendix D reports more detailed AUROC results. The best AUROC was significantly better than the AUROC associated with k=5000 (the maximum number of similar patients) for the Euclidean ($p<10^{-4}$), cosine ($p<10^{-4}$) and multifaceted ($p= 0.031$) PS metrics.

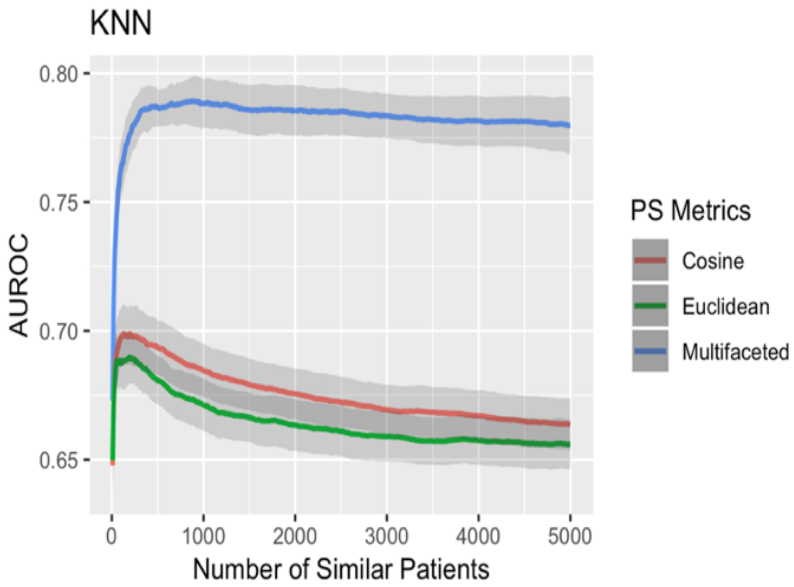

Figure 4-4 Accuracy evaluation of PS Metrics using KNN. The bands demonstrate the 95% confidence intervals.

The predictive performance worsened rapidly as the number of similar patients increased for the cosine and Euclidean metrics and more gradually for the multifaceted metric.

4.3.3 Multifaceted PS Metric in Predictive Modelling

Figure 4-5 demonstrates the predictive performance of PS-based LR as a function of a similar patient-cohort size. Table 4-2 summarizes the detailed results.

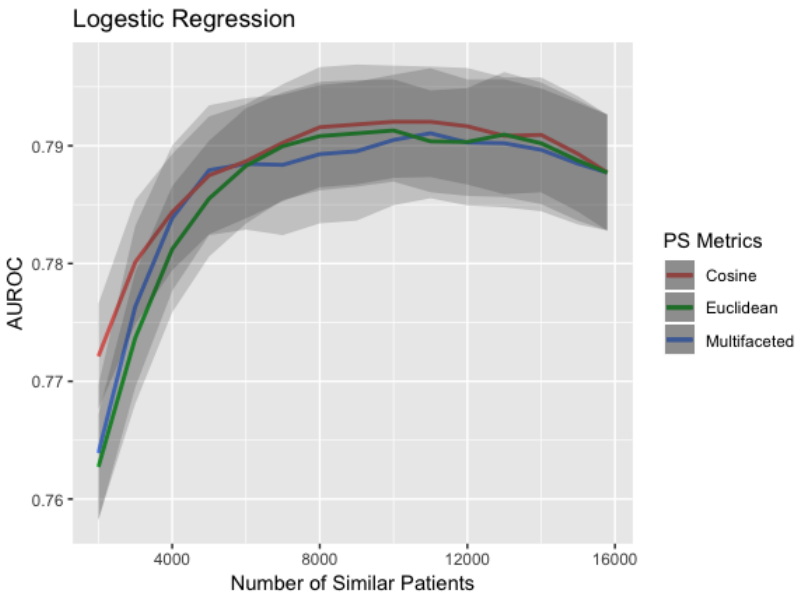

Figure 4-5 Mortality prediction performance of PS-based LR. The bands demonstrate the 95% confidence intervals.

Table 4-3 Detailed mortality prediction performance of PS-based LR.

Number of Similar Patients	Multifaceted PS AUROC (Mean [95% CI])	Cosine PS AUROC (Mean [95% CI])
2000	0.764 [0.758, 0.758]	0.772 [0.768, 0.777]
3000	0.776 [0.77, 0.77]	0.78 [0.775, 0.785]
4000	0.784 [0.778, 0.778]	0.784 [0.779, 0.789]
5000	0.788 [0.782, 0.782]	0.787 [0.782, 0.792]
6000	0.788 [0.783, 0.783]	0.789 [0.784, 0.794]
7000	0.788 [0.782, 0.782]	0.79 [0.785, 0.795]
8000	0.789 [0.783, 0.783]	0.792 [0.786, 0.797]
9000	0.79 [0.784, 0.784]	0.792 [0.787, 0.797]
10000	0.79 [0.785, 0.785]	0.792 [0.787, 0.797]
11000	0.791 [0.786, 0.786]	0.792 [0.787, 0.797]
12000	0.79 [0.785, 0.785]	0.792 [0.787, 0.797]
13000	0.79 [0.785, 0.785]	0.791 [0.786, 0.796]
14000	0.79 [0.784, 0.784]	0.791 [0.786, 0.796]
15000	0.788 [0.783, 0.783]	0.789 [0.784, 0.794]
15792	0.788 [0.783, 0.783]	0.788 [0.783, 0.793]

The maximum AUROC of 0.792 (SD=0.008) and 0.791 (SD=0.009) were achieved when using data from 10000 and 11000 similar patients in training based on cosine and multifaceted PS metrics. None of the PS metrics outperform the others in this setting, and their results were not significantly better than that resulting from a population-based setup (AUROC of 0.788 [SD=0.008]).

Figure 4-6 shows the predictive performance of PS-based DT as the number of similar patients increases for cosine and multifaceted PS metrics. Table 4-4 summarizes the detailed results.

Figure 4-6 Mortality prediction performance of PS-based LR. Detailed mortality prediction performance of PS-based LR.

Table 4-4 Detailed mortality prediction performance of PS-based DT.

Number of Similar Patients	Multifaceted PS AUROC (Mean [95% CI])	Cosine PS AUROC (Mean [95% CI])
2000	0.7204 [0.7078, 0.7078]	0.7099 [0.694, 0.7257]
3000	0.7039 [0.6949, 0.6949]	0.6931 [0.6849, 0.7013]
4000	0.6827 [0.6762, 0.6762]	0.6804 [0.6721, 0.6887]
5000	0.6728 [0.6664, 0.6664]	0.6759 [0.6682, 0.6835]
6000	0.6637 [0.6577, 0.6577]	0.6687 [0.6628, 0.6747]
7000	0.6604 [0.6555, 0.6555]	0.6663 [0.6602, 0.6724]
8000	0.6584 [0.653, 0.653]	0.6608 [0.6538, 0.6678]
9000	0.6551 [0.6501, 0.6501]	0.6553 [0.6492, 0.6615]
10000	0.6566 [0.6509, 0.6509]	0.6525 [0.6473, 0.6577]
11000	0.6568 [0.6523, 0.6523]	0.6516 [0.6468, 0.6563]
12000	0.6528 [0.648, 0.648]	0.6512 [0.6463, 0.6561]
13000	0.6525 [0.6484, 0.6484]	0.6506 [0.6463, 0.6549]
14000	0.6506 [0.6477, 0.6477]	0.6499 [0.6469, 0.653]
15000	0.6533 [0.6499, 0.6499]	0.6501 [0.6467, 0.6535]

Results demonstrate that a smaller number of more-similar patients in the training set can result in better performance than when all patients are included. The maximum AUROC of 0.72 (SD= 0.02) and 0.71 (SD= 0.026) were achieved with 2000 similar patients for the multifaceted and cosine PS metrics, respectively. The maximum performances were significantly better than those for the model that used 15792 patients ($p<10^{-5}$).

4.4 Discussion

While precision medicine aims to provide more accurate and personalized treatment for patients, a PS metric is a fundamental component for identifying patients who are clinically similar. Physicians generally utilize their knowledge, available research, and experience from previous patients to make a decision for new patients, and PS analytics can help them in this information-retrieval process. Recently, many researchers have introduced various PS measurement techniques. However, a challenge in this area is how to define a PS metric that can best capture the search intent (Wongsuphasawat et al. 2009). This study introduced a multifaceted PS metric in which the similarities of patients are investigated from various aspects and then combined to provide one similarity score. The multi-layer nature of the multifaceted PS metric provides physicians with the opportunity to customize their similarity measure definition. If the similarity in terms of demographics is not of interest or if the search intent is to find all similar patients regardless of their gender, the associated part can easily be removed from a multifaceted PS metric without changing the system pipeline. In other words, the multifaceted PS metric has a modular design, thus, each part can perform independently.

In the current study, comparing a multifaceted PS metric with conventional PS metrics (i.e., cosine and Euclidean) showed that the former is more accurate in measuring the similarity between patients in KNN setting. Although the proposed method outperformed the conventional methods in every value of K, the best performance was achieved by including a

higher number of patients in the training set in comparison to conventional metrics. Moreover, the advantage of multifaceted PS metric was not observed in the DT and LR settings. This can be due to the fact that DT and LR are more invariant to the training set in comparison to KNN. Therefore, it may be the case that using different PS metrics resulting in different training sets does not significantly change the prediction performance of LR and DT.

The results of this study did not show that the performance of predictive models always significantly improves when PS analytics was utilized in defining the training set. Although the results from including only similar patients in training a DT demonstrate statistically significant improvement in AUROC in comparison to a situation when a DT is trained on the whole training dataset, results from the LR model did not strongly support this fact. The best AUROC for LR was achieved when 11000 similar patients were included in the training set rather than when all the training set was used; however, the AUROC was not significantly higher than the population-based setting. This outcome is contrary to that of Lee et al. (2015a), who found that PS-based LR significantly outperformed the population-based LR on their cohort of patients from MIMIC-II. Although our results are in line with those of Lee et al. (2015c), a possible explanation for this lack of strong evidence for the LR model might be the difference between the groups of patients included in the studies. The aforementioned study used data from patients in the MIMIC-II database who did not have data missing from their profile, whereas the present work focused on MICU and SICU units in MIMIC-III and

did not limit itself to data from patients with complete profiles. A recent case study on the reproducibility of studies on mortality prediction using the MIMIC database demonstrated how subtle differences in the inclusion/exclusion criteria can impact results (Johnson et al. 2017).

The current study found that various predictive models can show different behaviours when only similar patients are being used for their training. For instance, when fewer but more similar patients were included in training a DT, the performance improved. However, LR was more invariant to changes in the training set. These results match those observed in the earlier study by Lee et al. (2015a) in which improvement in AUROC for DT was three times greater than that for LR (the improvements in AUROC for LR and DT were 0.02 and 0.062, respectively).

It is important to mention that although the PS analytics often lead to better prediction in comparison to population-based techniques, the reduced sample size in this approach leads to greater variation in model estimation and prediction accuracy measures. The most clinically important finding was the interpretability of the multifaceted PS metric and its suitability for visualization analytics. Via a multifaceted PS metric, physicians will have the ability to identify the drivers of similarity.

Our study has limitations. First, the scope was limited to two ICU units, and only from an ICU database coming from a single (albeit large) research hospital. Second, since the focus was on introducing a new multifaceted PS metric, only averaging was used to fuse the

similarity scores. Therefore, further research is needed to investigate the effect of fusion methods on the accuracy of the multifaceted PS. Moreover, the pair-wise similarity calculation comes at the expense of high computational time, however, future researchers can benefit from big data technologies such as Spark and Hadoop or cloud computational resources to minimize the computation time. Last but not least, this study investigated the performance of PS analytics being used with DT and LR. To investigate the performance of PS-based predictive modeling in comparison to population-based methods, a more-comprehensive list of models, such as random forest and gradient boosting, must be included.

Several questions still remain to be answered. A natural progression of this work would be to analyze the effect of other fusion methods on the accuracy of the multifaceted PS metric. A further study could investigate PS variation over time and monitor patient progression over time. Last but not least, future researchers can focus on tuning the multifaceted PS metric in terms of the metrics used for each variable.

Chapter 5

Identifying Subpopulations of Septic Patients: A Temporal Data-Driven Approach

This chapter utilizes the multifaceted PS metric introduced in previous chapter to identify septic patient subpopulations. It also investigates the extra information in the temporal aspect of EHR data for phenotyping patient similarity. The identified subpopulations can provide insights into customizing care for septic patients.

5.1 Introduction

In ICUs, it is very important to monitor patient health status over time. This necessity results in multiple measurements of a particular clinical variable across a given patient's stay. Due to the time varying nature of these measurements, they are usually referred to as temporal EHR. Since temporal EHR data have valuable information about the evolution of patient health status, researchers have been able to detect improvements in predictive modeling (Sha et al. 2016; Sun et al. 2010b; Sun et al. 2010a) and patient stratification (Singh et al. 2015) through using this information. However, analyzing longitudinal data is accompanied with multiple challenges such as irregular sampling rates and varying lengths of available measurements, as well as the inherent correlation of repeated measurements of a variable over time within the same patient.

This chapter focuses on using the information in temporal vital signs data in our multifaceted PS metric through functional data analysis (Ramsay 2005). Then, cluster analysis will be employed for identifying a subpopulation of septic patients with similar clinical needs and trajectories. This information may be used to design customized care platforms for patients who share similar needs.

5.1.1 Literature Review

Longitudinal data are an important part of EHR data and have initiated interest among researchers, many of whom agree that the time-varying aspect of EHR data may contain additional information about patient health status (Lehman et al. 2014; Lehman et al. 2015b; Lehman et al. 2015a; Singh et al. 2015; Agarwal et al. 2016; Pimentel et al. 2013). ICUs have a heterogeneous population with various health status dynamics and similar needs for constant care ('Critical Care Statistics' ; Prin et al. 2012). This heterogeneity adds to the importance of finding similar patients and detecting the underlying phenotypic groups. Recently, efforts have been made to employ the temporal information of heterogeneous EHR data to reveal subpopulations.

A large and growing body of literature has investigated vital-sign trajectories to discover patient subpopulations in order to identify the underlying pathophysiology of diseases and suggest customized care pathways. Pimentel et al. (2013) focused on the evolution of vital signs in post-operative patients using an unsupervised Gaussian process model. Their study revealed four dominant underlying physiological behaviors in their population.

Lehman et al. (2014) investigated the discriminative bivariate dynamics of heart rate and blood pressure in 450 ICU patients, using a switching vector autoregressive process approach. Their study demonstrated that the temporal evolution of the vital signs has additional predictive value for sepsis detection beyond the non-temporal approaches. Moreover, their study revealed ten prevalent underlying physiological modes for patient health status, each of which was correlated with different sepsis risk levels. In addition, their exploratory analysis revealed that high-risk modes have less variability in their trajectories. Building on this study, they suggested that the observed patterns are due to a patient's health status, undergoing clinical intervention, and possible artifacts. Therefore, similarity can be investigated based on the underlying dynamics of vital signs and help healthcare providers manage short- and long-term outcomes with appropriate interventions (2015a).

Lehman et al. (2015b) extended their work and compared the changes in vital signs dynamics before and after applying vasopressor treatment. The study demonstrated distinguishable differences in the dynamics among survivors and non-survivors. However, they did not elaborate on the potentially discriminative nature of vasopressor treatment patterns. Agarwal et al. (2016) used a functional clustering model to find sub-populations in a cohort of patients with chronic kidney disease, based on creatinine measurement trajectories. They found two sub-populations, each with a dominant creatinine trajectory. Exploratory analysis of the clusters revealed various discriminative factors between the clusters, including the presence of comorbidities and adherence to medication.

Along with previous studies, the Latent Class Mixed Models (LCMMs) have been widely used for analysis of change over time and for uncovering subpopulations in a heterogeneous patient population (Gill et al. 2010; Maggs et al. 2004; Proust-Lima et al. 2009b; Proust-Lima et al. 2009a; McCulloch et al. 2002; Quantin et al. 1999; Rubin et al. 1997; Muthén et al. 1999) for various clinical applications related to hospital reimbursement (Quantin et al. 1999), schizophrenic behaviour (Rubin et al. 1997), alcohol dependence in youth (Muthén et al. 1999) and prostate cancer (Gill et al. 2010; Maggs et al. 2004; Proust-Lima et al. 2009b; Proust-Lima et al. 2009a; McCulloch et al. 2002). In these models, two sub-models based on the predictors are defined: the probability function of latent class membership and the class-specific trajectory function. LCMM refers to models that have an unobserved subpopulation structure.

Lin et al. (2000), applying LCMM on data from the Nutritional Prevention of Cancer study, focused on the binary outcome of prostate cancer and the prostate-specific antigen (PSA) as the predictor. In this study, the longitudinal sub-model was the Linear Mixed Model (LMM), and the Expectation Maximization (EM) algorithm was used for coefficient estimation. The fitted model uncovered sub-populations, and the PSA trajectories were explicitly different between these classes. Building on the previous study, they extended the similar model to predict a survival outcome (prostate cancer onset) (Lin et al. 2002). McCulloch et al. (2002) reviewed the utility of these models in a health context. Moreover, they extended the previous studies and proposed a model for binary (incidence of prostate

cancer after 7 years), continuous longitudinal (PSA readings over time), and survival (the time until prostate cancer diagnosis) outcome. Considering all this evidence, it seems that longitudinal EHR is a valuable resource for finding PS in order to customize care delivery.

The Third International Consensus Definitions for Sepsis and Sepsis Shock (Sepsis-3) defined sepsis as "life threatening organ dysfunction caused by a dysregulated host response to infection," which respectively costs Canada and the United States more than $300 million and $20 billion annually, respectively (Singer et al. 2016). Along with this definition, Sepsis-3 introduced criteria to detect septic cases with the goal of facilitating clinical care. However, there are various sepsis guidelines available that may be more helpful for other purposes (Rhee et al. 2019). Overall, there are six widely-used criteria for identifying septic patients.

1. Sepsis-3 criteria: suspected infection with sequential organ failure assessment (SOFA) score of greater or equal to two (Singer et al. 2016)
2. Explicit sepsis: having at least one of the following ICD-9 codes:
    a. 785.52: septic shock
    b. 995.92: severe sepsis
3. Angus criteria: having at least one of the ICD-9 codes proposed by Angus et al. (2001).
4. Martin criteria: having at least one of the ICD-9 codes proposed by Martin et al. (2003).

5. The Centers for Medicare & Medicaid Services (CMS) criteria: uses a combination of ICD-9 codes, Systemic Inflammatory Response Syndrome (SIRS) criteria and specific thresholds for organ dysfunction (Medicare et al. 2012).
6. The Centers for Disease Control and Prevention (CDC) complete surveillance criteria: suspected infection with organ dysfunction criteria similar to SOFA (Seymour et al. 2016).

A number of interventions are used for treating sepsis. The antibiotics treatment usually starts immediately after diagnosis. Moreover, patients often receive intravenous fluid. If the low blood pressure persists, patients may also receive vasopressor medications, which make blood vessels constrict and helps to increase blood pressure. Understanding the various interventions for sepsis, and the administration and outcomes, has been a trending topic (Wu et al. 2017; Fialho et al. 2013). However, there is still considerable heterogeneity in the outcomes of sepsis treatments, a phenomenon known as "treatment effect heterogeneity" (Kravitz et al. 2004). Even after many attempts to explain this heterogeneity, there is no consensus for much of the variability in outcome of a particular treatment. Many researchers have focused on investigating sepsis-related research questions by using EHR data. Johnson et al., in their comprehensive study, demonstrated that even in one hospital, there are various groups of patients with a diagnosis of "sepsis" who have highly variable outcomes.

Fialho et al. (2013) suggested the idea of disease-specific models instead of one-size-fits-all models in ICUs. They focused on fluid resuscitation therapy and attempted to predict the

need for vasopressor therapy after failed fluid resuscitation. This study demonstrated that the response of fluid resuscitation of each predefined population—pneumonia and pancreatitis patients in an ICU—led to a different model. Salgado et al. (2016) used fuzzy ensemble models to predict vasopressor dependence. They first found sub-populations in the dataset using an unsupervised clustering method and then trained a fuzzy model on each sub-population. Researchers have also focused on leveraging temporal data in answering sepsis-related questions. One study utilized the longitudinal measurement of heartrate, mean blood pressure and respiratory rate for predicting the onset of septic shock with coupled hidden Markov models (Ghosh et al. 2017). They also compared their method to conventional approaches such as SVMs. According to their results, methods that account for the temporal aspect of data tend to perform better than conventional methods. On the same application, Khoshnevisan et al.(2018) demonstrated that using recent temporal patterns with various classification methods consistently outperform atemporal approaches. These results support the idea of leveraging temporal data when considering septic patients.

5.1.2 Objectives

Building on the previous studies, in this research, multifaceted PS is employed to investigate the presence of similar subpopulations among septic patients, with a special focus on taking the trajectory of their vital signs into consideration. Identifying subpopulations of septic patients with similar clinical needs, trajectories, and health status is the main objective of this

study. The results may provide a framework for more customized care for each subgroup of septic patients.

5.1.3 Materials and Method

5.1.3.1 Study Sample

This study utilized the data from patients admitted to the critical care units of the Beth Israel Deaconess Medical Center between 2008 and 2012 (MIMIC III database) provided in (Johnson et al. 2018a) and data extraction was done using the code provided by the authors (Johnson et al. 2018b). From 23,620 ICU admissions that were initially included, three nonadult patients were excluded. 7,536 admission were excluded to only focus on the first admission of patients with multiple admissions. Patients who admitted to the cardiothoracic surgical service were also excluded since their postoperative physiologic disorders do not have the same mortality risk as the other ICU patients (2,298 patients). Moreover, 18 admissions were removed because they had no charted data. Patients suspected of infection more than 24 hours after or before ICU admission were excluded to only focus on patients who admitted to ICU with sepsis (824 patients). Moreover, 2270 patients who stayed in the ICU for less than 24 hours and 209 patients who had less than two measurements for their vital signs in 24 hours were excluded. Finally, 9321 adult septic patients based on any of the following sepsis criteria were included in the study: Sepsis-3 criteria, explicit sepsis, Angus criteria, Martin criteria, CMS and CDC criteria.

5.1.3.2 Feature Extraction

For each patient, the following sets of predictors from the first 24 hour of the ICU admission were extracted:

1. Admission and demographic data: ICU service type (MICU, SICU, TSICU, CCU), admission type (emergency, elective, urgent), gender (female, male) and age
2. Minimum and maximum of the following variables: blood urea nitrogen (BUN), hematocrit, creatinine, bicarbonate, lactate, potassium, sodium, glucose, platelets, white blood cells, and Glasgow Coma Scale
3. Hourly measurement of vital signs: mean blood pressure (MBP), systolic blood pressure (SysBP), heartrate (HR), respiratory rate (RR), SpO_2 and body temperature (Temp)
4. Daily total urine output
5. Interventions: Duration and dosage of each of the following vasopressor administration: norepinephrine, epinephrine, phenylephrine, vasopressin, dobutamine and dopamine, presence of mechanical ventilation and dialysis.

To include the temporal aspect of vital signs, in addition to the simple statistical characteristics (including maximum, minimum, median, mode, mean, standard deviation and number of measurements) that can capture magnitude and variability of variables, functional principal components were used to identify the dominant modes of variation in vital signs. In the last decade, functional principal component analysis (FPCA) has been widely used in the

statistics and machine learning (ML) community for various application, including public health and biomedical applications, to reduce the dimensionality of data while preserving the information on variability over time (Ullah et al. 2013). The FPCA was applied for sparsely or densely observed vital signs via the Principal Analysis by Conditional Estimation (PACE) algorithm provided in the R package fdapace (Dai et al. 2017a). In the PACE, functional principal components are defined as conditional expectations. The expectation for the trajectory $X_i(t)$ for the ith patient when using only the first p eigenfunctions ($\hat{\phi}_k(t)$) is

$$\hat{X}_i^p(t) = \hat{\mu} + \sum_{k=1}^{p} \hat{\xi}_{ik}\, \hat{\phi}_k(t) \tag{4.8}$$

where $\hat{\mu}$ is the estimate of the mean function $E\big(X(t)\big) = \mu(t)$, and $\hat{\xi}_{ik}$ represents the functional principal component scores. The first p functional principal components were extracted for vital signs in such a way that the components cumulatively explain 98% of the total variation in the trajectory.

For laboratory tests, the maximum and minimum measurement during 24 hours was used. For vasopressor administration, the duration and total amount of each variable was extracted for the first 24 hours. For mechanical ventilation and dialysis only the presence of the administration (a binary variable) was utilized.

Vital signs in this study were considered as sparse data, therefore, no missing data treatment was performed for them. However, for laboratory tests, predictive mean matching (PMM) imputation method was employed to address missing data.

5.1.3.3 Cluster Analysis

A common practice when clustering datasets with large numbers of variables is to reduce dimensionality. This study utilized t-SNE to map patients to a two-dimensional space. The conventional t-SNE method utilizes Euclidean distance to map data into a two-dimensional space. In this study, the Mahalanobis distance and multifaceted PS were used instead of Euclidean distance in mapping data points in the following experiments. Then, DBSCAN clustering was used to find subpopulations in the cohort of septic patients. The parameters of DBSCAN (epsilon [ε] and minimum points [MinPts]) were tuned using the k-distance tuning method discussed in Section 3.4. MinPts were set to dimensionality of data pulse one. Then, the distance of each point from its k'th (k=MinPts) nearest neighbor was calculated (kdist). After calculating and sorting kdists for all data points, the k-distance graph was plotted and the value for which the graph showed a strong bend—the knee point—was selected for epsilon.

The silhouette index (Rousseeuw 1987), which evaluates the suitability of assigning a patient to a group rather than to another was utilized to evaluate the clustering method. The silhouette index is a number between -1 and 1, where a high value means the patient is strongly matched to its own cluster and weakly matches other clusters. The silhouette index

was calculated for each cluster by averaging the silhouette indices of the patients in that cluster.

For each of the following two scenarios, the same approach was utilized.

5.1.3.3.1 Investigating the informativeness of temporal data over cross-sectional data

In the first scenario, statistical characteristics (including maximum, minimum, median, mode, mean, standard deviation and number of measurements) and the functional principal component scores of vital sign measurements were utilized to calculate the pair-wise Mahalanobis distances. The resulting distance matrix was used as the input to our cluster analysis pipeline. Then, only the average over 24 hours for the vital sign measurements was used, in which the variability over time was lost. The pair-wise Mahalanobis distances were then used as the input. The results of both analyses were compared to identify the informativeness of temporal data over cross-sectional data.

5.1.3.3.2 Finding subpopulations in the septic patient cohort

In the second scenario, the focus was on finding septic patient subpopulations using the multifaceted PS metric introduced in Chapter 4. For the vital sign distance calculation, the Mahalanobis distance was applied to the Euclidean vector of statistical characteristics and functional principal component scores. After evaluating the clusters using the Silhouette method, the clusters were compared in terms of the patient and hospitalization characteristics. Then, the average severity of illness and prevalence of comorbidities were examined in each

cluster using Logistic Organ Dysfunction Score (LODS) (Le Gall et al. 1996), SOFA, SIRS and Elixhauser scores (Elixhauser et al. 1998). Moreover, correlation analysis was employed to investigate potential association between overall patient outcome (in-hospital and 30-days mortality rates, and ICU and hospital length of stays [LOSs]) and average age in clusters. The same analysis was performed for investigating the association between average severity of illness and patient outcomes in clusters. Moreover, the three most-common diagnoses in each cluster were reported.

5.2 Results

5.2.1 Data

This study was not limited to septic patients identified only by Sepsis-3 guidelines, as it was of interest to observe how a machine-learning driven approach can identify different groups of septic patients. However, the ML-driven clusters were compared with those derived by various sepsis definitions. Table 5-1 displays the descriptive characteristic of the included patients.

Table 5-1 Descriptive characteristics of the study cohort.

Patient Characteristics	
Age (yrs)	63.31 ± 18.42
Gender	
Female	4211 (45.22%)

Male	5101 (54.78%)
Admission Type	
Elective	657 (7.06%)
Emergency	8552 (91.84%)
Urgent	103 (1.11%)
First Care Unit	
MICU	4247 (45.61%)
SICU	1954 (20.98%)
TSICU	1491 (16.01%)
CCU	1293 (13.89%)
30-day Mortality Count	1211 (13%)
In-hospital Mortality Count	920 (9.88%)
Hospital Length of Stay (day)	8.67 ± 8.66
ICU Length of Stay (day)	3.96 ± 5.09

5.2.2 Functional Principal Component Score Extraction

The estimate of mean function using local linear smoothing is shown in Figure 5-1, revealing the overall decreasing trend for HR, MBP, SysBP and SpO_2, and the increasing trend for RR and body temperature.

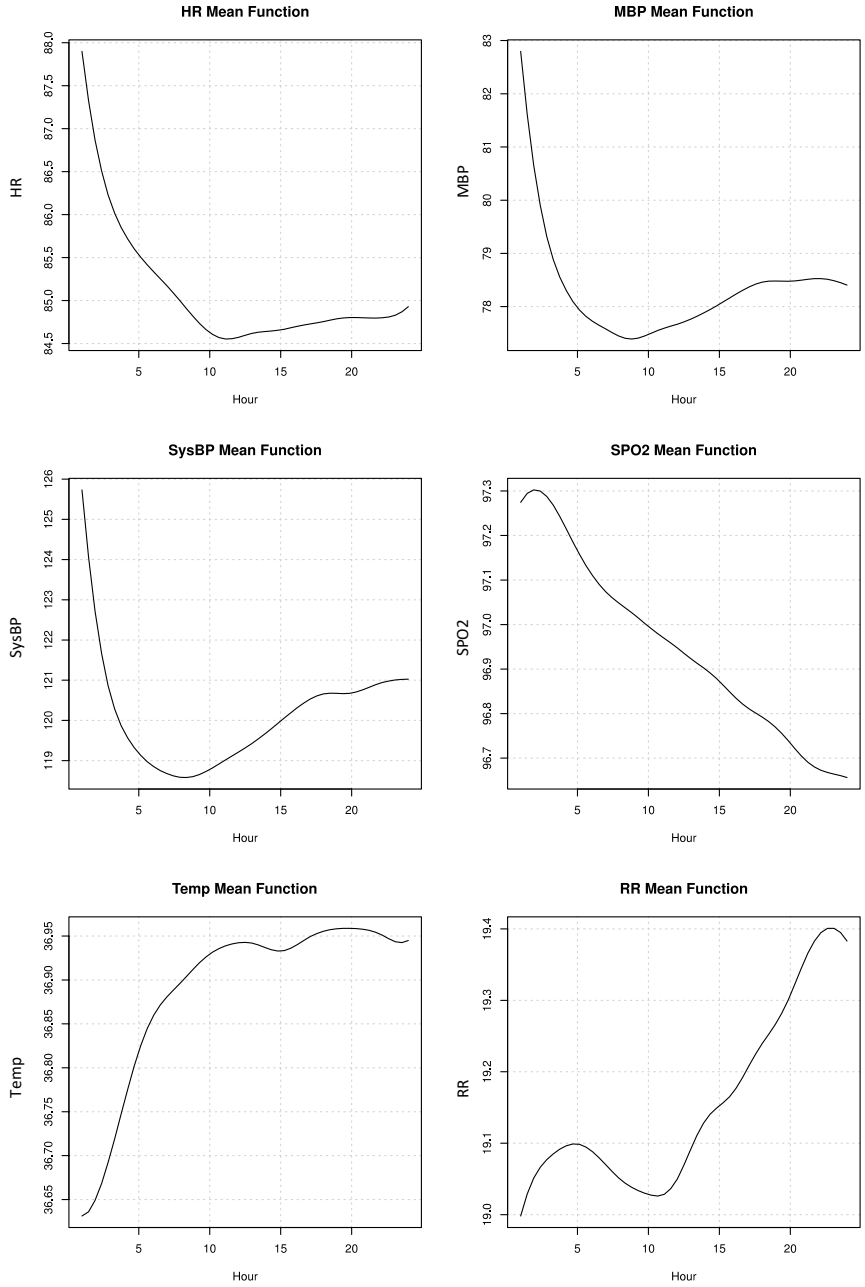

Figure 5-1 Smooth estimate of the mean functions for vital signs.

In Figure 5-2, the first estimated eigenfunction for each vital sign is provided.

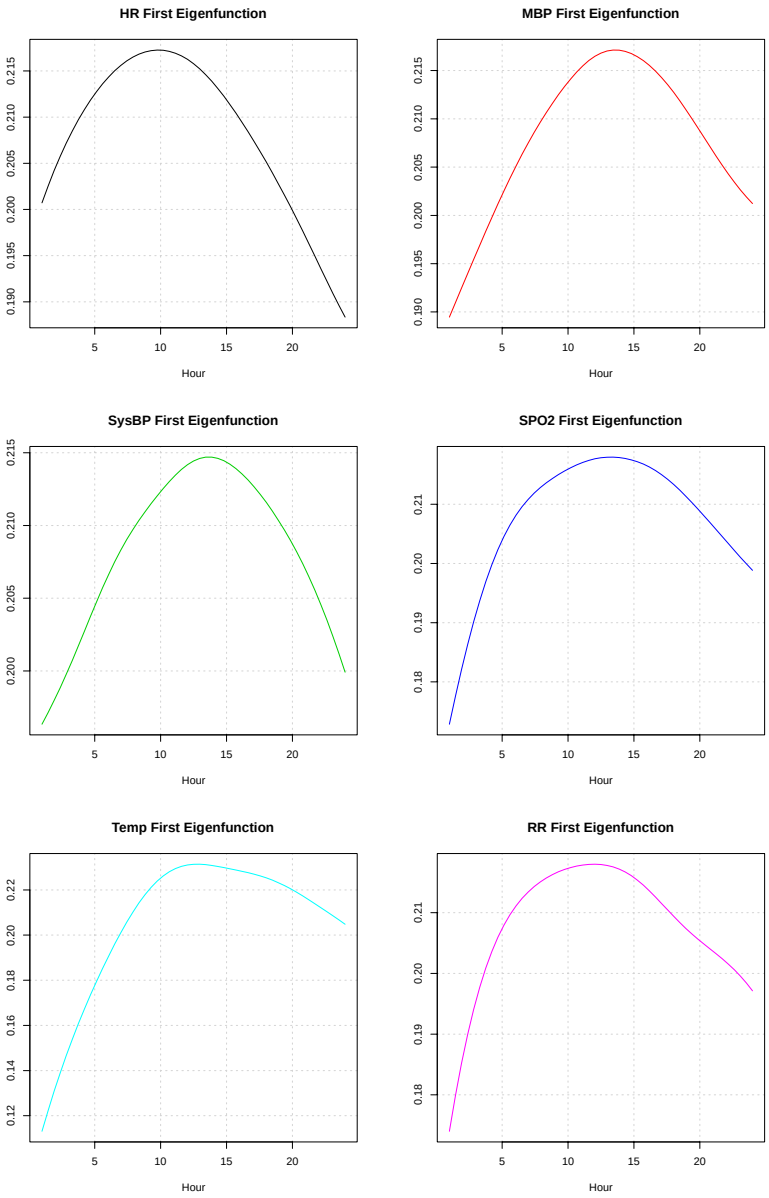

Figure 5-2 Estimate of the first eigenfunction for various vital signs in the entire cohort.

Figure 5-3 provides the scree-plots for each vital sign, demonstrating the fraction of total variance in data as explained by each functional principal component. The first eigenfunction for HR, SysBP and MBP explains more than 80% of the total variation of the data, and for SpO_2 and Temp accounts for 78.41% and 67.98% of the total variation, respectively. Overall, more than 98% of the variation in HR and RR can be explained by the first three eigenfunctions. For MBP, SysBP and SpO_2 this number increases to four. The first five eigenfunctions for Temp explains more than 98% of the variation. The scores for these eigenfunctions were extracted for each patient to be used as features for the clustering phase.

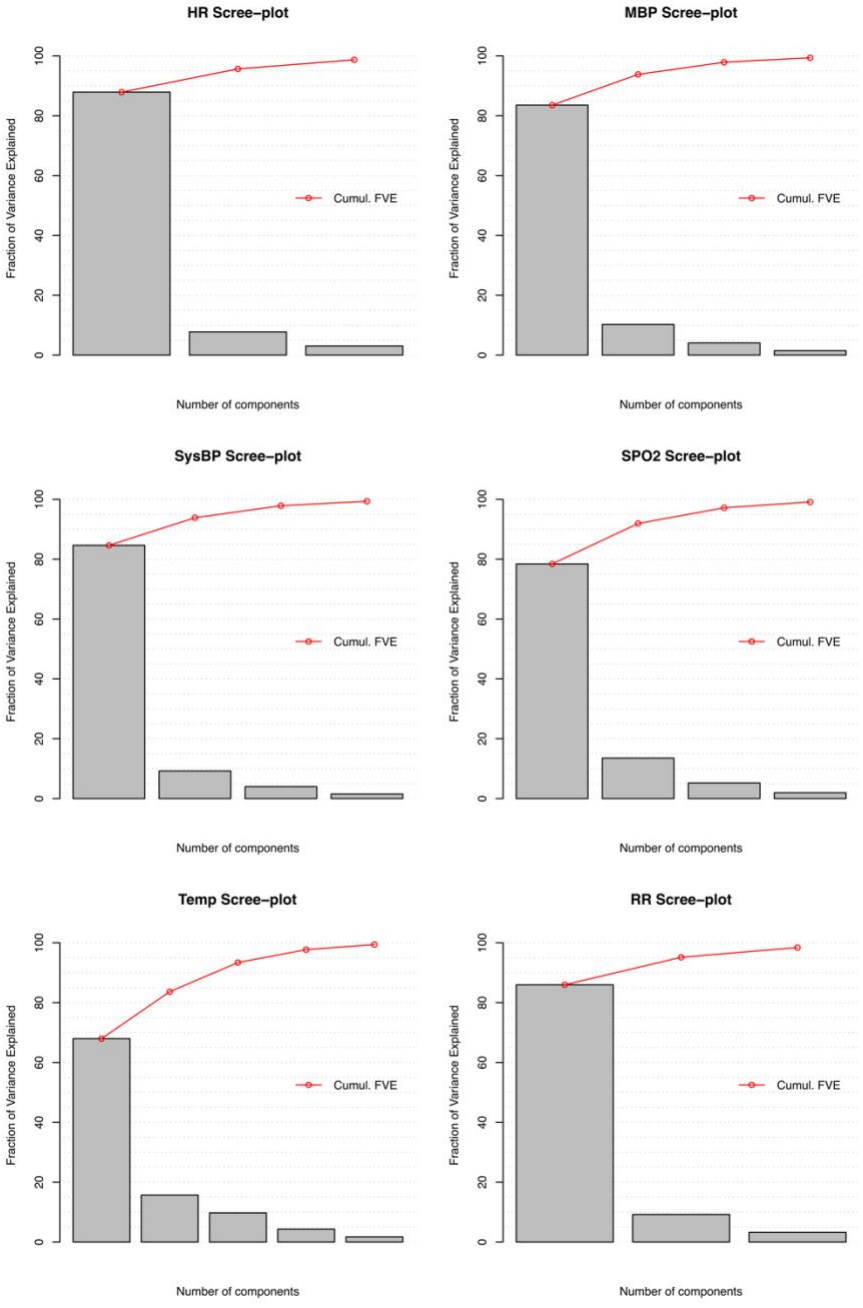

Figure 5-3 The scree plots show the portion of variance explained by each functional principal component.

5.2.3 Cluster Analysis

5.2.3.1 Investigating the informativeness of temporal data over cross-sectional data

Figure 5-4 demonstrates clustering results when only the average of vital signs is used. As can be seen, all patients are considered in one cluster and similar to each other, with a few patients labeled as outliers.

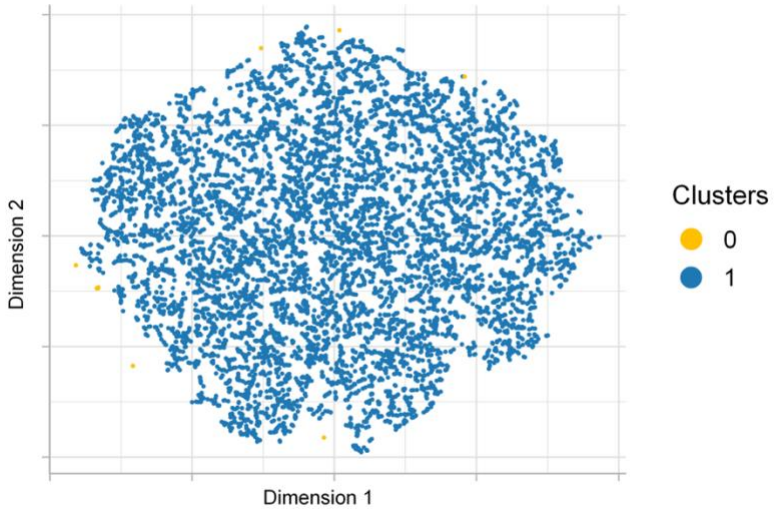

Figure 5-4 Clustering septic patients using the average value of vital signs during the first 24 hours of admission to ICU. DBSCAN clustering method with MinPts=25 and ε =2.25 was used. Data were mapped to a two-dimensional space using t-SNE with perplexity=25 and 1000 iteration.

By including the functional principal component scores, small clusters start emerging, while excluding them resulted in one cohort. Figure 5-5 demonstrates five clusters identified

by DBSCAN after including the information about the trajectories of vital signs. Moreover, there is another small cluster emerging at the left end in Figure 5-5, which was not identified by DBSCAN clustering method.

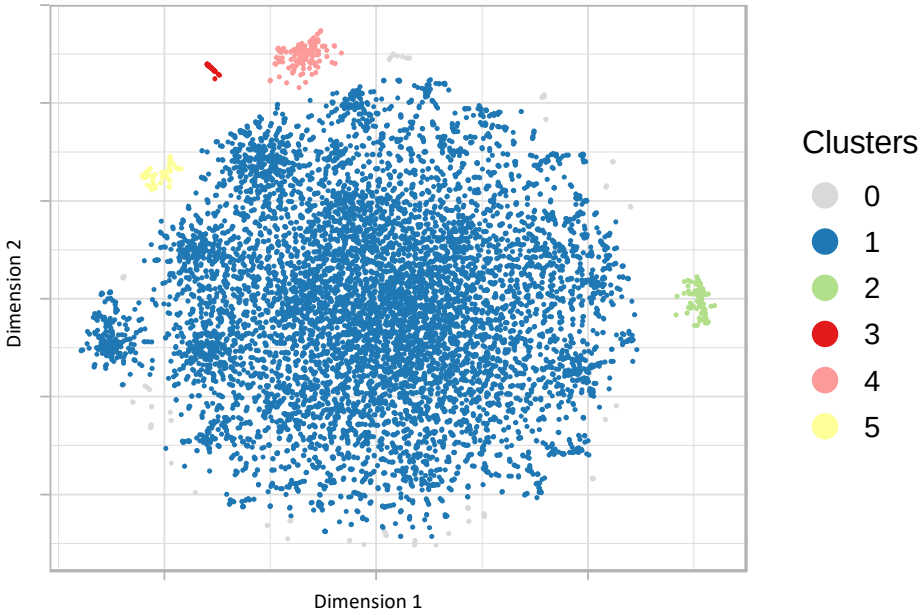

Figure 5-5 Clustering septic patients using hourly vital signs during the first 24 hours of admission to ICU along with the functional principal component scores. DBSCAN clustering method with MinPts=25 and ε =2.25 was used. Data were mapped to a two-dimensional space using t-SNE with perplexity=25 and 1000 iteration.

For a fair comparison, all the parameters were kept the same and only changed how vital signs are represented. These results posit that the temporal aspect of vital signs has additive information and can be helpful in clustering septic patients.

5.2.3.2 Finding subpopulations in the septic patient cohort

Cluster analysis was implemented using the multifaceted PS metric on a comprehensive list of information from patients provided in Section 5.1.3.1. After tuning the parameters of DBSCAN, 9 main clusters and cluster 0 representing outlier samples were identified.

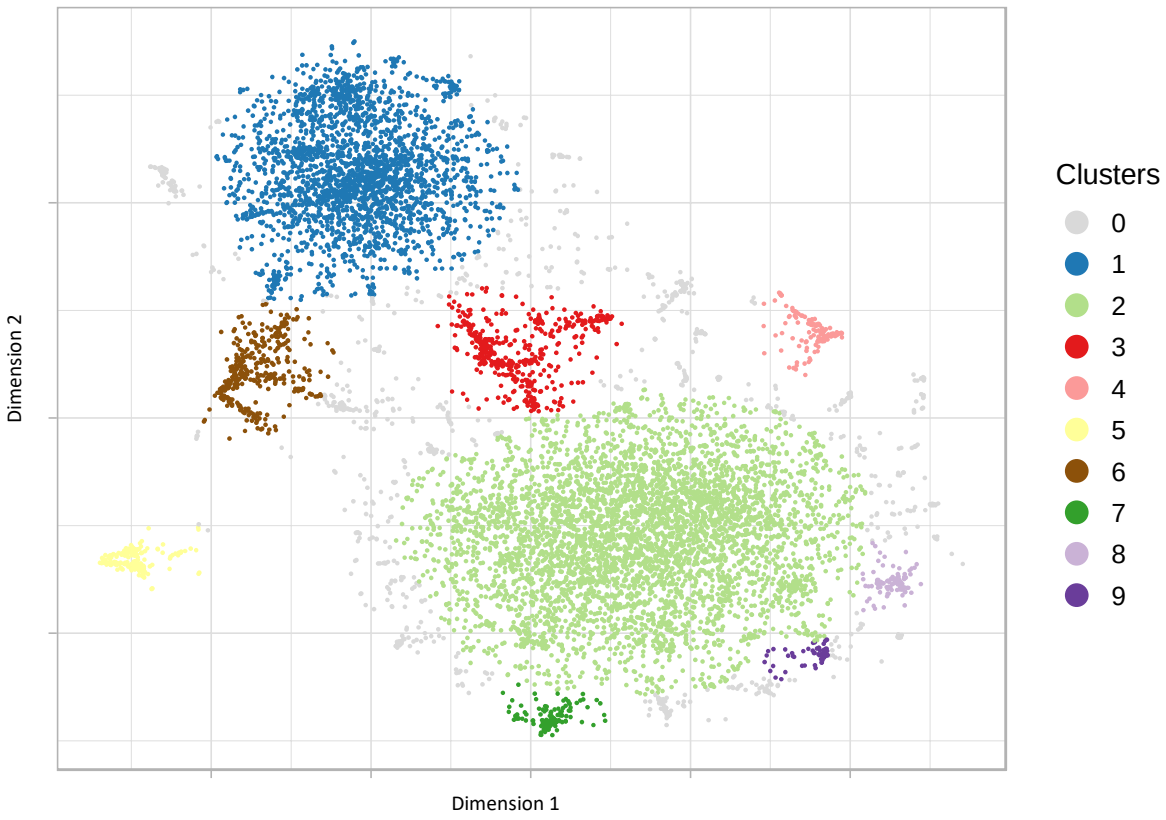

Figure 5-6 Clustering septic patients using multifaceted PS metric on patient profiles. DBSCAN clustering method with MinPts=103 and ε =1.4 was used. Data were mapped to a two-dimensional space using t-SNE with perplexity=100 and 1000 iteration.

The clusters were evaluated using the silhouette statistics. It was observed that except for the group of outliers which received a silhouette value of -0.63, the average silhouette value for all clusters were positive. Figure 5-7 demonstrates the average silhouette values for all clusters in which only one cluster had a weak structure, all other clusters had reasonable and strong structures. The silhouette values are color coded based on the ranges proposed in (Kaufman et al. 2009). Green ([0.71,1]) indicates that a strong structure has been found for that particular cluster. Yellow ([0.51,0.70]) and red (≤0.50) indicate reasonable and weak structures, respectively.

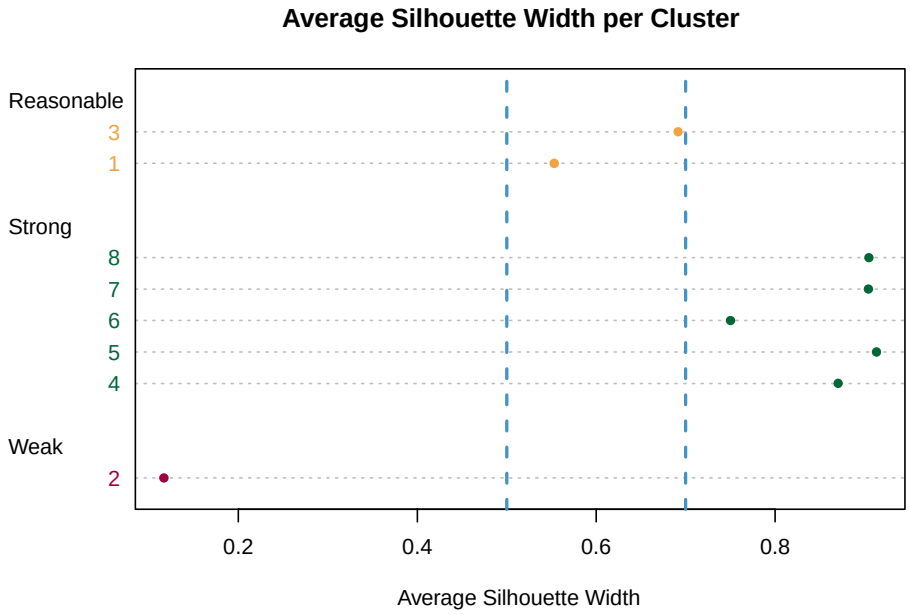

Figure 5-7 The average silhouette values for each cluster. Green, yellow and red respectively represent strong, reasonable and weak structures.

The distribution of cluster memberships, patient and hospitalization characteristics are shown in Table 5-2. The smallest cluster size is 97 and the biggest is 4522. In hospital mortality ([3.45%, 24.43%]) and 30-days mortality ([6.9%, 28.24%]) rates varied significantly between clusters, implying the various outcomes of septic patients. Clusters # 2, 8, and 9 had the lowest mortality rates.

Table 5-2 Patient and hospitalization characteristics in each cluster.

#	Size	Age (y)	Male (%)	In-hospital mortality (%)	30-day mortality (%)	Hospital LoS (d)	ICU LoS (d)
0	916	62.94	55.13	19.76	22.05	11.13	5.1
1	2417	60.41	56.76	13.53	15.72	10.24	5.15
2	4522	64.01	54.14	3.78	7.41	6.79	2.61
3	461	68.5	50.54	19.31	24.08	11.56	6.79
4	131	73.85	52.67	24.43	28.24	7.95	5.73
5	220	61.48	57.73	12.73	13.18	12.21	5.31
6	354	65.34	56.78	17.51	19.77	11.56	5.99
7	136	71.94	50.74	17.65	25.74	8.51	3.12
8	97	60.87	53.61	4.12	8.25	4.64	2.28
9	58	47.68	43.10	3.45	6.9	5.57	2.41

It was also observed that the average age was positively correlated with 30-day mortality. This association was not as significant for in-hospital mortality. This result may imply that death after discharge is more likely among older age groups (Figure 5-8).

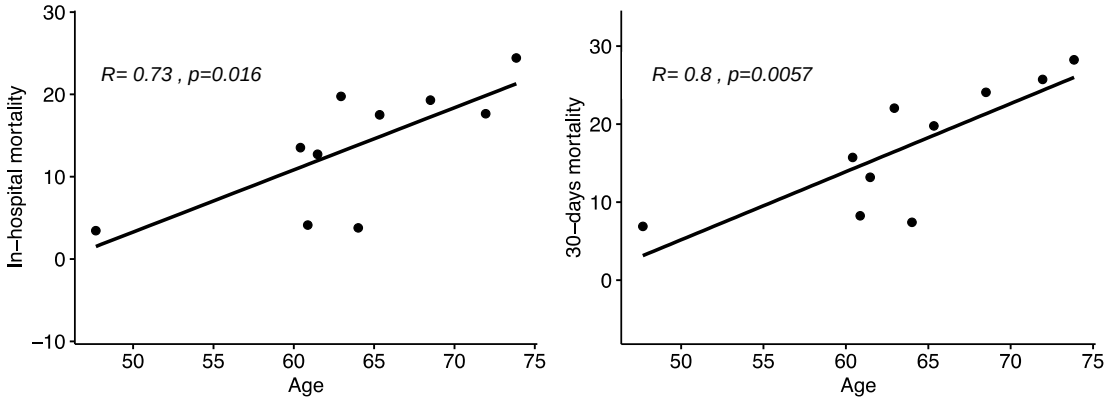

Figure 5-8 The association between mortality rates and age groups.

Table 5-3 demonstrates the average Elixhauser comorbidity score in each cluster. The average scores of SOFA, LODS and SIRS are also presented for each cluster. The 30-days mortality rate was significantly correlated with average SOFA (ρ=0.76, p-value<0.02), LODS (ρ=0.87, p-value<0.005) and SIRS (ρ=0.72, p-value<0.02). The same results were observed for in-hospital mortality, hospital LoS, and ICU LoS. These results are in line with the fact that sicker patients are more vulnerable and, hence, more likely to experience death.

Table 5-3 A summary of comorbidities and severity of illness scores in each cluster.

#	Average Elixhauser	SOFA	LODS	SIRS
0	3.61	6.03	5.53	2.95
1	2.33	4.11	4.3	2.91
2	2.14	2.68	2.86	2.53
3	5.49	8.25	6.23	3.16
4	3.25	7.07	6.17	2.79
5	4.18	7.29	5.95	2.73
6	3	4.61	5.13	3.08
7	5.85	5.26	5.56	2.76
8	1.34	2.26	2.37	2.3
9	1.64	2.36	2.59	2.47

Table 5-4 presents the top three most-common diagnoses assigned by the admitting clinicians in each cluster. Clusters number 0 and 3, which have patients with a sepsis diagnosis also have very high rates of mortality

Table 5-4 Top three most-common diagnoses assigned by the admitting clinicians in each cluster (#: number of patients with a particular diagnosis in each cluster).

Cluster 0	**#**	**Cluster 1**	**#**	**Cluster 2**	**#**
Sepsis	45	Intracranial hemorrhage	145	Pneumonia	203
Pneumonia	33	Pneumonia	101	Upper GI bleed	129
Altered mental status	25	Altered mental status	99	Chest pain	93
Cluster 3	**#**	**Cluster 4**	**#**	**Cluster 5**	**#**
Sepsis	77	Congestive heart failure	9	Acute renal failure	9
Pneumonia	35	Bradycardia	8	Congestive heart failure	8
Hypotension	23	Chest pain	5	Sepsis	7
Cluster 6	**#**	**Cluster 7**	**#**	**Cluster 8**	**#**
Intracranial hemorrhage	13	Stroke, telemetry, transient ischemic attack	15	Stemi	3
Chest pain	12	Altered mental status	10	Alcohol withdrawal	3
Chest pain\cardiac cath	9	Pneumonia	10	Bradycardia	3
Cluster 9	**#**				
Diabetic ketoacidosis	30				
Hyperglycemia	2				
Upper gi bleed	2				

Table 5-5 presents the distribution of sepsis diagnosis based on six sepsis definitions (explicit, Angus, Martin, CMS, CDC and Sepsis-3 methodologies) in each cluster. Except for cluster numbers 3 and 6, Sepsis-3 criteria identified a higher number of septic patients in each cluster. For the two aforementioned clusters, CDC method covered most of the patients.

Table 5-5 The distribution of sepsis diagnosis based on six sepsis definitions in each cluster.

#	Angus (%)	Martin (%)	Explicit (%)	CDC (%)	CMS (%)	Sepsis-3 (%)
0	44.1	29.37	22.82	55.13	25	64.08
1	37.98	11.17	5.79	37.9	6.41	61.11
2	20.63	11.26	5.04	20.17	7.7	37.13
3	78.96	63.77	56.62	94.58	59.65	94.36
4	44.27	24.43	17.56	64.12	20.61	62.6
5	45.91	30	18.18	46.82	23.64	70.45
6	34.18	17.23	12.99	80.23	13.84	71.75
7	40.44	26.47	13.97	31.62	22.06	56.62
8	17.53	6.19	3.09	10.31	5.15	20.62
9	20.69	8.62	6.9	12.07	6.9	25.86

Table 5-6 shows the percentage of patients in each cluster that underwent dialysis or mechanical ventilation. It also includes the average administration duration of various vasopressor administration. No statistically significant correlation was observed between the level of medication that patients received and their outcomes.

Table 5-6 Statistical description of medications (average duration in minutes) and interventions (percentage of patients who received the intervention) utilized in each cluster.

#	Mechanical Ventilation (%)	Dialysis (%)	Norepinephrine Duration (m)	Epinephrine Duration (m)	Phenylephrine duration (m)	Vasopressin Duration (m)	Dobutamine Duration (m)	Dopamine Duration (m)
0	45.52	3.93	4.46	0.64	3.63	4.71	0.5	0.44
1	100	0	0.21	0	0.3	0.03	0	0.02
2	0.04	0	0.12	0	0.06	0.01	0	0.02
3	60.52	0.87	22.4	0.03	1.22	0.35	0.01	0.38
4	45.8	6.11	3.15	0.15	1.85	1.29	0	21.06
5	39.55	100	3.06	0.02	0.53	0.08	0	0.15
6	80.23	1.13	0.42	0	19.61	0.11	0	0.03
7	2.94	0.74	0.88	0	0.52	0.8	0	0.02
8	1.03	0	0.43	0.01	0.11	0.15	0.01	0
9	0	0	0.1	0	0	0	0	0

Table 5-7 represents the average dosage of vasopressors for each cluster.

Table 5-7 The average dosage of vasopressors utilized in each cluster.

#	Norepinephrine Dosage (mg)	Epinephrine Dosage (mg)	Phenylephrine Dosage (mg)	Vasopressin Dosage (mg)	Dobutamine Dosage (mg)	Dopamine Dosage (mg)
0	4.77	0.16	41.87	10.86	10.76	18.35
1	0.1	0	1.2	0.05	0.02	0.72
2	0.06	0	0.37	0.01	0.06	0.48
3	13.76	0.01	8.86	0.81	0.26	11.87
4	3.15	0.06	15.78	2.94	0	683.35
5	1.82	0.01	2.78	0.17	0	3.63
6	0.25	0	76.04	0.19	0	0.66
7	0.69	0	3.21	1.84	0	0.31
8	0.5	0	1.14	0.36	0.13	0
9	0.01	0	0	0	0	0

Finally, Figure 5-9 presents the first eigenfunction for each vital sign in each cluster. It can be clearly seen that some clusters have completely different trends for their first eigenfunctions. This observation highlights the importance of including the variation explained by temporal data to capture different evolutions of patient health status.

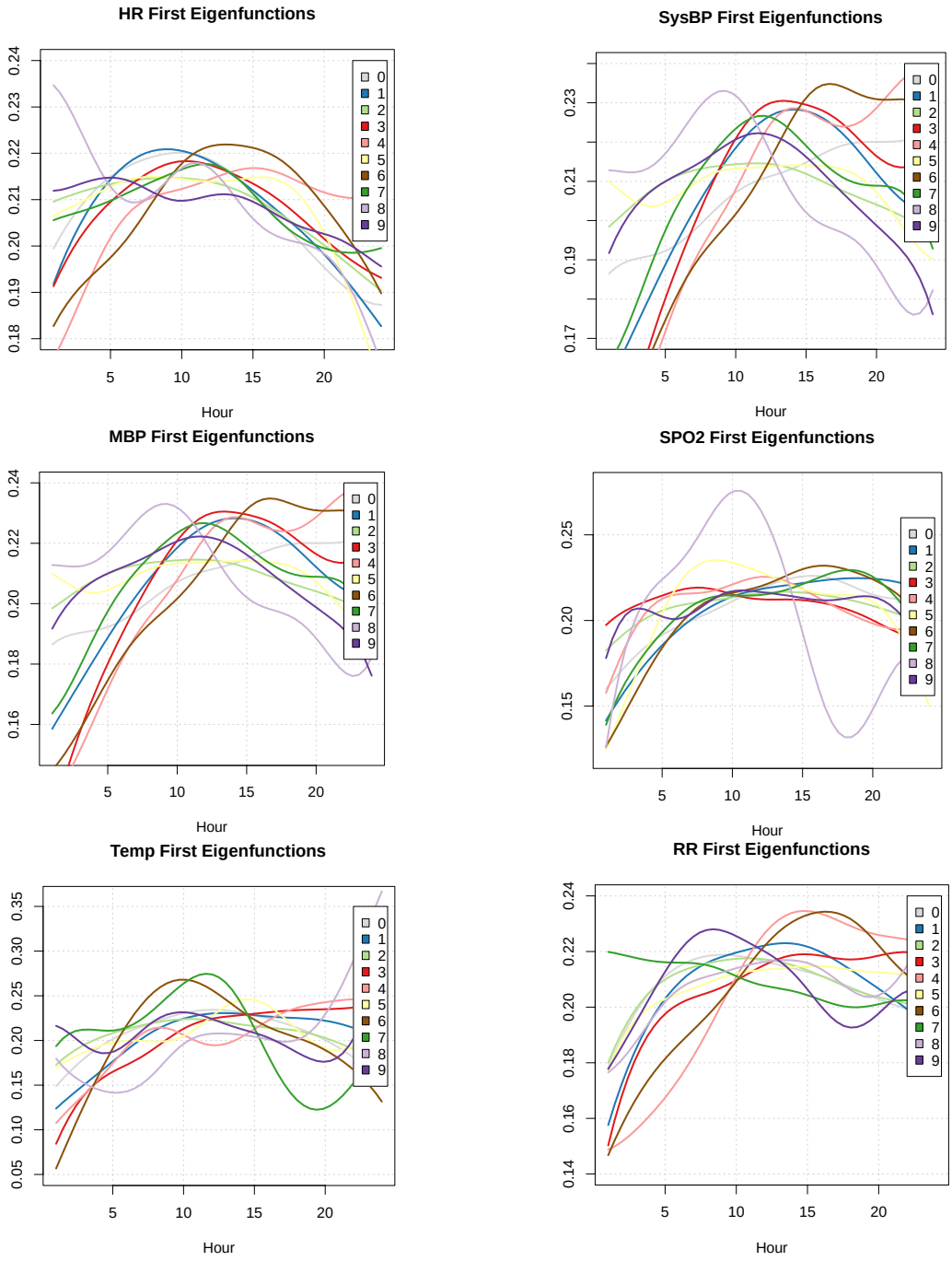

Figure 5-9 First eigen function estimates for HR, SysBP, MBP, SpO_2, Temp RR data in each cluster.

5.3 Discussion

In this study the focus was on leveraging temporal data in clustering septic patients. Summarizing vital sign measurements by their average not only eliminates the information on patient health status changes over time, it cannot be helpful in phenotyping septic patients. Moreover, incorporating the information on the vital sign trajectories using functional data analysis can help in finding distinction among septic patients.

The results suggest that although septic patients share similar underlying physiological condition, grouping them as one cohort results in a loss of information about their unique characteristics. It was demonstrated that clustering analysis enabled identification of nine clinically distinguishable subpopulations from a cohort of septic patients. The results also showed that the dominant mode of variation for each vital sign differs from one subpopulation to another. For instance, the first eigenfunction for SpO_2 in cluster #8 shows high variation over time while cluster #3 has a near-flat first eigenfunction. It is worth mentioning that cluster #8 is among subpopulations with low mortality rates (30-day mortality: 8.25%, in-hospital mortality: 4.12%) while cluster #3 has high mortality rates (30-day mortality: 24.08%, in-hospital mortality: 24.43%). Thus, the variations in vital signs trajectories can be informative of patient future outcome.

The characteristics of identified clusters were in line with clinical outcomes such as increased mortality rates in groups with a higher average age. These results were in line with the cluster analysis done by Vranas et al. (2017) that focused on clustering ICU patients and

identified clinically distinguishable clusters with sepsis as their most-common admitting diagnoses, and thus highlights the potential heterogeneity among septic patients.

The results also demonstrated that septic patients have different care needs; however, a higher level of care was not correlated with poorer outcome. This observation is supported by the fact that there are precisely predefined care plans for septic patients at various levels of severity of illness, and they are usually effective in controlling the disease. However, understanding why therapeutic efficiency differs between subpopulations is critically important in caring for septic patients. For instance, cluster #8 has low mortality rates (30-day mortality: 8.25%, in-hospital mortality: 4.12%) while cluster #7 with similar therapeutic patterns has much higher mortality rates (30-day mortality: 25.74%, in-hospital mortality: 17.65%). A further study with more focus on treatment effectiveness in each subpopulation is therefore suggested.

In addition, the results showed that the Sepsis-3 criteria are more inclusive than the others, except for two clusters for which CDC was more inclusive. Both Sepsis-3 and CDC criteria use treatments as proxy of organ failure. This result is supported by that of Johnson et al. (2018a) in which Sepsis-3 criteria identified a higher number of patients, followed by CDC. However, this finding also suggests more studies are needed on the characteristics of septic patients based on Sepsis-3 and its inclusiveness.

This study has some limitations. First, the data used in this work was from only a single medical center (albeit a large one) in the analysis. Second, the vital signs trajectories were

examined in a 24-hour window; however, data from a longer period would better capture patient health changes over time.

In this study we identified groups of septic patients with similar characteristics and needs which can facilitate the move toward precision medicine by considering the differences between subpopulations to support customized therapy approaches. Therefore, future research should focus on investigating the key differences between these subpopulations using medical expert opinion and identifying the best practice for each group. Another future direction is to use these latent sepsis phenotypes for PS-based predictive modeling by training a model for each subpopulation instead of using a one-size-fits-all model.

Chapter 6

Conclusions and Discussion

The goal of the work presented in this thesis was to introduce new approaches to employing PS for more-efficient prediction and clustering of ICU data in order to fulfill the promise of precision medicine. To that end, it was demonstrated that missing data in EHR are informative and can be used to improve prediction performance. Moreover, this information can help in finding similar patients in the ICU. The PS metric itself was one of the focuses of this work. A new multifaceted PS metric was introduced that is not only more accurate than conventional methods, but is also more interpretable and adaptive to the different use cases. Last but not least, it was showed that the temporal information in longitudinal EHR data can be used to improve the performance of clustering methods and identify patient subpopulations in a focused application of sepsis patients.

In Chapter 2, it was showed that the hidden information in missing data has implications for patient health status and outcomes and that this information can be employed to improve the performance of predictive modeling. The performance of different predictive models (DT, LR and RF) was investigated when missingness indicators are added to the feature set. It was showed that models in which the hidden information of missing data was included

outperformed not only SAPS-II but also similar methods without missing data information in predicting in-hospital mortality and 30-day mortality.

In Chapter 3, an unsupervised study was performed to better understand the informativeness of missingness indicators and how they can be used to identify similar patients. Missingness indicators were used as representative of lab-test ordering and used them for phenotyping patients. Results demonstrated that these indicators capture physicians' opinions about patient needs and symptoms. In the results of clustering, it was observed that clusters with higher rates of mortality had higher numbers of performed laboratory tests. These results were in line with the fact that ICU physicians tend to order more tests for patients who are severely ill, as they need to be monitored closely and comprehensively.

Chapter 4 focused on introducing a multifaceted PS metric in which the similarity of every aspect of the EHR data is considered individually and then various similarity scores are combined to make a unified score for two patients. The proposed method significantly outperformed the conventional and most-commonly-used metrics, including the cosine PS metric, in terms of similarity accuracy, with focus on 30-day mortality. The performance of this metric was also investigated when used with other prediction models, including DT and LR. Although the multifaceted PS metric outperformed the cosine metric when used with DT, similar results were not observed for LR. These results highlight the need for in-depth investigation of the performance of various predictive models in a PS-based framework on different databases.

Finally, in Chapter 5, the results from previous chapters were utilized to identify subpopulations among septic patients with special focus on temporal EHR. The results demonstrate that temporal data have additive information and can be helpful in clustering patients.

Notwithstanding the contributions that this thesis work has provided, it is important to acknowledge the known shortcomings of the research. First, this work has resulted in findings that are limited to only one large research healthcare facility, with a special focus on its ICU. This focus is largely a consequence of the unavailability of other publicly available EHR databases. Moreover, the findings from the study in Chapter 4 were limited to a simple averaging combining method. Investigations into the impacts of other fusion methods on the accuracy and predictive power of the multifaceted PS metric would be of interest, though were beyond the aim of the current study.

Overall, each of the five studies in this thesis contributes to current knowledge about missing data in EHR, PS analytics in patient stratification, PS metric in predictive modeling and informativeness of temporal EHR. Chapter 2 provided a new insight into missing data in EHR and was novel in that it is the first study comprehensively examining the informativeness of missing data in ICU EHRs. Chapter 3 built on the previous study by exploring ICU patient subpopulations using the information in missing data. The findings supported the informativeness of missing data and promoted the new insight into missing data. Chapter 4 introduced the idea of multifaceted PS and investigated its validity. Finally,

Chapter 5 provided evidence on additive information of temporal EHR and how this information can help in identifying subpopulations of septic people.

The methods provided in this dissertation can easily be implemented in the back-end of the various interactive data visualization products at care facilities to help clinicians better understand the data. Moreover, these techniques can easily be implemented on the cloud as an application programming interface (API) and be employed by a simple call to the API at any time for various applications."

Despite these promising contributions, questions remain. Although much knowledge exists on imputation methods, as evidenced in Chapter 2, more research is needed to explore the additive information of missing data in various care units and particularly general wards where missing data is more substantial than ICUs. The findings from the exploratory work in Chapter 3 provided only insights into phenotyping ICU patients using missing data. Future research should be undertaken to examine the test-ordering behavior among clinicians using missing data information. Moreover, to develop a full picture of the multifaceted PS introduced in Chapter 4, additional studies will be needed to investigate its capabilities on various datasets in predicting other clinical outcomes than mortality. Last but not least, while Chapter 5 demonstrated the informativeness of temporal EHR data in phenotyping septic patients, there is abundant room for further analysis of these subpopulations and identifying more efficient care plan for patients who demonstrate similar trajectories. Moreover, in this work, we used AUROC discrimination measure to evaluate the performance of the predictive

models. Besides other discrimination measure such as area under the precision recall curve, future researchers can also perform evaluation from calibration perspective (Steyerberg 2009).

Bibliography

Abraham, W. T., and D. W. Russell. 2004. 'Missing data: a review of current methods and applications in epidemiological research', *Current Opinion in Psychiatry*, 17: 315-21.

Agarwal, Vibhu, and Nigam H Shah. 2016. 'Learning Attributes of Disease Progression from Trajectories of Sparse Lab Values.' in, *Biocomputing 2017* (World Scientific).

Aggarwal, Charu C. 2014. *Data Classification: Algorithms and Applications* (Chapman and Hall/CRC).

Agniel, Denis, Isaac S. Kohane, and Griffin M. Weber. 2018. 'Biases in electronic health record data due to processes within the healthcare system: retrospective observational study', *BMJ*, 361.

Ahn, S. Y., J. Ryu, S. H. Baek, J. W. Han, J. H. Lee, S. Ahn, K. I. Kim, H. J. Chin, K. Y. Na, D. W. Chae, K. W. Kim, and S. Kim. 2014. 'Serum anion gap is predictive of mortality in an elderly population', *Exp Gerontol*, 50: 122-7.

Amir, A., D. Beymer, J. Grace, H. Greenspan, D. Gruhl, A. Hobbs, K. Pohl, T. Syeda-Mahmood, J. Terdiman, and F. Wang. 2010. 'AALIM: a cardiac clinical decision support system powered by advanced multi-modal analytics', *Studies in health technology and informatics*, 160: 846-50.

Ammenwerth, E., P. Schnell-Inderst, C. Machan, and U. Siebert. 2008. 'The effect of electronic prescribing on medication errors and adverse drug events: a systematic review', *J Am Med Inform Assoc*, 15: 585-600.

Andridge, R. R., and R. J. Little. 2010. 'A Review of Hot Deck Imputation for Survey Non-response', *Int Stat Rev*, 78: 40-64.

Angus, D. C., W. T. Linde-Zwirble, J. Lidicker, G. Clermont, J. Carcillo, and M. R. Pinsky. 2001. 'Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care', *Crit Care Med*, 29: 1303-10.

Asiimwe, S. B., S. Okello, and C. C. Moore. 2014. 'Frequency of vital signs monitoring and its association with mortality among adults with severe sepsis admitted to a general medical ward in Uganda', *PLoS One*, 9: e89879.

Baraldi, A. N., and C. K. Enders. 2010. 'An introduction to modern missing data analyses', *J Sch Psychol*, 48: 5-37.

Basole, R. C., M. L. Braunstein, V. Kumar, H. Park, M. Kahng, D. H. Chau, A. Tamersoy, D. A. Hirsh, N. Serban, J. Bost, B. Lesnick, B. L. Schissel, and M. Thompson. 2015. 'Understanding variations in pediatric asthma care processes in the emergency department using visual analytics', *J Am Med Inform Assoc*, 22: 318-23.

Bazick, H. S., D. Chang, K. Mahadevappa, F. K. Gibbons, and K. B. Christopher. 2011. 'Red cell distribution width and all-cause mortality in critically ill patients', *Crit Care Med*, 39: 1913-21.

Beier, K., S. Eppanapally, H. S. Bazick, D. Chang, K. Mahadevappa, F. K. Gibbons, and K. B. Christopher. 2011. 'Elevation of blood urea nitrogen is predictive of long-term mortality in critically ill patients independent of "normal" creatinine', *Crit Care Med*, 39: 305-13.

Bellazzi, Riccardo, Fulvia Ferrazzi, and Lucia Sacchi. 2011. 'Predictive data mining in clinical medicine: a focus on selected methods and applications', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1: 416-30.

Blumenthal, D. 2010. 'Launching HITECH', *N Engl J Med*, 362: 382-5.

Bobrowski, L. 2006. 'Induction of similarity measures and medical diagnosis support rules through separable, linear data transformations', *Methods of Information in Medicine*, 45: 200-03.

Bose, E. L., G. Clermont, L. Chen, A. W. Dubrawski, D. Ren, L. A. Hoffman, M. R. Pinsky, and M. Hravnak. 2018. 'Cardiorespiratory instability in monitored step-down unit patients: using cluster analysis to identify patterns of change', *J Clin Monit Comput*, 32: 117-26.

Botsis, T., G. Hartvigsen, F. Chen, and C. Weng. 2010. 'Secondary Use of EHR: Data Quality Issues and Informatics Opportunities', *Summit Transl Bioinform*, 2010: 1-5.

Brown, S. A. 2016. 'Patient Similarity: Emerging Concepts in Systems and Precision Medicine', *Front Physiol*, 7: 561.

Buuren, Stef van. 2012. *Flexible Imputation of Missing Data* (CRC press).

Cadogan, S. L., J. P. Browne, C. P. Bradley, and M. R. Cahill. 2015. 'The effectiveness of interventions to improve laboratory requesting patterns among primary care physicians: a systematic review', *Implement Sci*, 10: 167.

Cahill, Ken. 2010. 'Royal Prince Alfred Hospital Patient Observation (Vital Signs) Policy - Adult'.

Campanella, P., E. Lovato, C. Marone, L. Fallacara, A. Mancuso, W. Ricciardi, and M. L. Specchia. 2016. 'The impact of electronic health records on healthcare quality: a systematic review and meta-analysis', *Eur J Public Health*, 26: 60-4.

Campbell-Scherer, Denise. 2010. 'Multimorbidity: a challenge for evidence-based medicine', *Evidence Based Medicine*, 15: 165-66.

Campillo-Gimenez, B., W. Jouini, S. Bayat, and M. Cuggia. 2013. 'Improving Case-Based Reasoning Systems by Combining K-Nearest Neighbour Algorithm with Logistic Regression in the Prediction of Patients' Registration on the Renal Transplant Waiting List', *PLoS One*, 8.

Cardona-Morrell, M., M. Nicholson, and K. Hillman. 2015. 'Vital Signs: From Monitoring to Prevention of Deterioration in General Wards.' in Jean-Louis Vincent (ed.), *Annual Update in Intensive Care and Emergency Medicine 2015* (Springer International Publishing: Cham).

Cauthen, C. A., M. J. Lipinski, A. Abbate, D. Appleton, A. Nusca, A. Varma, E. Goudreau, M. J. Cowley, and G. W. Vetrovec. 2008. 'Relation of blood urea nitrogen to long-term mortality in patients with heart failure', *Am J Cardiol*, 101: 1643-7.

Cebul, R. D., T. E. Love, A. K. Jain, and C. J. Hebert. 2011. 'Electronic health records and quality of diabetes care', *N Engl J Med*, 365: 825-33.

Celi, L. A., R. G. Mark, D. J. Stone, and R. A. Montgomery. 2013. '"Big data" in the intensive care unit. Closing the data loop', *Am J Respir Crit Care Med*, 187: 1157-60.

Chan, K. S., J. B. Fowles, and J. P. Weiner. 2010. 'Review: electronic health records and the reliability and validity of quality measures: a review of the literature', *Med Care Res Rev*, 67: 503-27.

Charlson, M. E., P. Pompei, K. L. Ales, and C. R. MacKenzie. 1987. 'A new method of classifying prognostic comorbidity in longitudinal studies: development and validation', *J Chronic Dis*, 40: 373-83.

Chattopadhyay, S., P. Ray, H. S. Chen, M. B. Lee, and H. C. Chiang. 2008. 'Suicidal risk evaluation using a similarity-based classifier', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5139 LNAI: 51-61.

Choi, Seung-Seok, Sung-Hyuk Cha, and Charles C Tappert. 2010. 'A survey of binary similarity and distance measures', *Journal of Systemics, Cybernetics and Informatics*, 8: 43-48.

Cismondi, F., L. A. Celi, A. S. Fialho, S. M. Vieira, S. R. Reti, J. M. Sousa, and S. N. Finkelstein. 2013. 'Reducing unnecessary lab testing in the ICU with artificial intelligence', *Int J Med Inform*, 82: 345-58.

Coiera, E. 2009. 'Building a National Health IT System from the middle out', *J Am Med Inform Assoc*, 16: 271-3.
Collins, F. S., and H. Varmus. 2015. 'A new initiative on precision medicine', *N Engl J Med*, 372: 793-5.
Collins, Francis S. 2015. "NIH Precision Medicine Initiative." In *NIH Medline Plus*, 19-21. North Carolina: StayWell.
Conti, Rena, David L. Veenstra, Katrina Armstrong, Lawrence J. Lesko, and Scott D. Grosse. 2010. 'Personalized Medicine and Genomics: Challenges and Opportunities in Assessing Effectiveness, Cost-Effectiveness, and Future Research Priorities', *Medical Decision Making*, 30: 328-40.
Cook, Nancy R. 2007. 'Response to Letters Regarding Article, "Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction"', *Circulation*, 116: E134-E34.
'Critical Care Statistics'. Society of Critical Care Medicine, Accessed 2019-05-03. https://www.sccm.org/Communications/Critical-Care-Statistics.
Currie, G. P., D. K. Lee, and B. J. Lipworth. 2006. 'Long-acting beta2-agonists in asthma: not so SMART?', *Drug Saf*, 29: 647-56.
Dai, X, PZ Hadjipantelis, H Ji, HG Mueller, and JL Wang. 2017a. "fdapace: Functional data analysis and empirical dynamics. R package version 0.4. 0." In.
Dai, Y., S. Lokhandwala, W. Long, R. Mark, and L. W. H. Lehman. 2017b. "Phenotyping hypotensive patients in critical care using hospital discharge summaries." In *4th IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2017*, 401-04. Institute of Electrical and Electronics Engineers Inc.
Daniels, Marcia, and Steven A. Schroeder. 1977. 'Variation among Physicians in Use of Laboratory Tests II. Relation to Clinical Productivity and Outcomes of Care', *Med Care*, 15: 482-87.
David, G., L. Bernstein, and R. R. Coifman. 2011. 'Generating evidence based interpretation of hematology screens via anomaly characterization', *Open Clinical Chemistry Journal*, 4: 10-16.
Delegates, ASA House of. 2014. 'ASA Physical Status Classification System', American Society of Anesthesiologists (ASA). https://www.asahq.org/resources/clinical-information/asa-physical-status-classification-system.
Delen, D., G. Walker, and A. Kadam. 2005. 'Predicting breast cancer survivability: a comparison of three data mining methods', *Artif Intell Med*, 34: 113-27.
Devine, E. B., R. N. Hansen, J. L. Wilson-Norton, N. M. Lawless, A. W. Fisk, D. K. Blough, D. P. Martin, and S. D. Sullivan. 2010. 'The impact of computerized provider order entry on medication errors in a multispecialty group practice', *J Am Med Inform Assoc*, 17: 78-84.
Deza, Michel Marie, and Elena Deza. 2009. 'Encyclopedia of distances.' in, *Encyclopedia of Distances* (Springer).
Elixhauser, A., C. Steiner, D. R. Harris, and R. M. Coffey. 1998. 'Comorbidity measures for use with administrative data', *Med Care*, 36: 8-27.
Ester, Martin, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. "A density-based algorithm for discovering clusters in large spatial databases with noise." In *Kdd*, 226-31.
Farley, J. F., C. R. Harley, and J. W. Devine. 2006. 'A comparison of comorbidity measurements to predict healthcare expenditures', *Am J Manag Care*, 12: 110-9.
Fialho, A. S., L. A. Celi, F. Cismondi, S. M. Vieira, S. R. Reti, J. M. Sousa, and S. N. Finkelstein. 2013. 'Disease-based modeling to predict fluid response in intensive care units', *Methods Inf Med*, 52: 494-502.

Fitzmaurice, Garrett M., Nan M. Laird, and James H. Ware. 2011. *Applied longitudinal analysis* (Wiley: Hoboken, N.J.).
Forster, M., C. Bailey, M. W. Brinkhof, C. Graber, A. Boulle, M. Spohr, E. Balestre, M. May, O. Keiser, A. Jahn, M. Egger, and Art-Linc collaboration of International Epidemiological Databases to Evaluate AIDS. 2008. 'Electronic medical record systems, data quality and loss to follow-up: survey of antiretroviral therapy programmes in resource-limited settings', *Bull World Health Organ*, 86: 939-47.
Frassica, J. J. 2003. 'Frequency of laboratory test utilization in the intensive care unit and its implications for large scale data collection efforts', *AMIA Annu Symp Proc*, 12: 844.
Ghassemi, M., L. A. Celi, and D. J. Stone. 2015. 'State of the art review: the data revolution in critical care', *Crit Care*, 19: 118.
Ghosh, S., J. Li, L. Cao, and K. Ramamohanarao. 2017. 'Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns', *J Biomed Inform*, 66: 19-31.
Gill, T. M., E. A. Gahbauer, L. Han, and H. G. Allore. 2010. 'Trajectories of disability in the last year of life', *N Engl J Med*, 362: 1173-80.
Ginsburg, G. S., and H. F. Willard. 2009. 'Genomic and personalized medicine: foundations and applications', *Transl Res*, 154: 277-87.
Girardi, D., S. Wartner, G. Halmerbauer, M. Ehrenmuller, H. Kosorus, and S. Dreiseitl. 2016. 'Using concept hierarchies to improve calculation of patient similarity', *J Biomed Inform*, 63: 66-73.
Gorelick, M. H. 2006. 'Bias arising from missing data in predictive models', *J Clin Epidemiol*, 59: 1115-23.
Gottlieb, A., G. Y. Stein, E. Ruppin, R. B. Altman, and R. Sharan. 2013. 'A method for inferring medical diagnoses from patient similarities', *BMC Med*, 11: 194.
Gottlieb, A., G. Y. Stein, E. Ruppin, and R. Sharan. 2011. 'PREDICT: a method for inferring novel drug indications with application to personalized medicine', *Mol Syst Biol*, 7: 496.
Gotz, D., and K. Wongsuphasawat. 2012a. 'Interactive intervention analysis', *AMIA Annu Symp Proc*, 2012: 274-80.
Gotz, David, Harry Stavropoulos, Jimeng Sun, and Fei Wang. 2012b. "ICDA: a platform for Intelligent Care Delivery Analytics." In *AMIA*.
Gowda, S., P. B. Desai, V. V. Hull, A. A. Math, S. N. Vernekar, and S. S. Kulkarni. 2009. 'A review on laboratory liver function tests', *Pan Afr Med J*, 3: 17.
Griffin, S. J., P. S. Little, C. N. Hales, A. L. Kinmonth, and N. J. Wareham. 2000. 'Diabetes risk score: towards earlier detection of Type 2 diabetes in general practice', *Diabetes/Metabolism Research and Reviews*, 16: 164-71.
Groenwold, R. H., I. R. White, A. R. Donders, J. R. Carpenter, D. G. Altman, and K. G. Moons. 2012. 'Missing covariate data in clinical research: when and when not to use the missing-indicator method for analysis', *CMAJ*, 184: 1265-9.
Gupta, S., T. Tran, W. Luo, D. Phung, R. L. Kennedy, A. Broad, D. Campbell, D. Kipp, M. Singh, M. Khasraw, L. Matheson, D. M. Ashley, and S. Venkatesh. 2014. 'Machine-learning prediction of cancer survival: a retrospective study using electronic administrative records and a cancer registry', *BMJ Open*, 4: e004007.
Haase, Peter, Ronny Siebes, and Frank van Harmelen. 2004. 'Peer Selection in Peer-to-Peer Networks with Semantic Topologies.' in Mokrane Bouzeghoub, Carole Goble, Vipul Kashyap and Stefano Spaccapietra (eds.), *Semantics of a Networked World. Semantics for Grid Databases* (Springer Berlin Heidelberg: Berlin, Heidelberg).

Han, Y., Y. Yang, Y. Yan, Z. Ma, N. Sebe, and X. Zhou. 2015. 'Semisupervised feature selection via spline regression for video semantic recognition', *IEEE Trans Neural Netw Learn Syst*, 26: 252-64.

Harrell, M., D. Fabbri, and M. Levy. 2016. "Evaluating EHR Data Availability for Cohort Selection in Retrospective Studies." In *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, 380-87.

Haukoos, J. S., and C. D. Newgard. 2007. 'Advanced statistics: missing data in clinical research--part 1: an introduction and conceptual framework', *Acad Emerg Med*, 14: 662-8.

Hayrinen, K., K. Saranto, and P. Nykanen. 2008. 'Definition, structure, content, use and impacts of electronic health records: a review of the research literature', *Int J Med Inform*, 77: 291-304.

Henriques, J., P. Carvalho, S. Paredes, T. Rocha, J. Habetha, M. Antunes, and J. Morais. 2015a. 'Prediction of Heart Failure Decompensation Events by Trend Analysis of Telemonitoring Data', *IEEE Journal of Biomedical and Health Informatics*, 19: 1757-69.

Hielscher, T., M. Spiliopoulou, H. Volzke, and J. P. Kuhn. 2014. "Using Participant Similarity for the Classification of Epidemiological Data on Hepatic Steatosis." In *Computer-Based Medical Systems (CBMS), 2014 IEEE 27th International Symposium on*, 1-7.

Hood, Leroy, and David J Galas. 2009. 'Systems biology and emerging technologies will catalyze the transition from reactive medicine to predictive, personalized, preventive and participatory (P4) medicine', *Interdisciplinary Bio Central*, 1: 6.

Houeland, T. G. 2011a. "An efficient random decision tree algorithm for case-based reasoning systems." In *Proceedings of the 24th International Florida Artificial Intelligence Research Society, FLAIRS - 24*, 401-06.

Houeland, Tor Gunnar. 2011b. "An Efficient Random Decision Tree Algorithm for Case-Based Reasoning Systems." In *FLAIRS Conference*.

Hu, Jianying, Adam Perer, and Fei Wang. 2016a. 'Data Driven Analytics for Personalized Healthcare.' in, *Healthcare Information Management Systems*.

Hu, L. Y., M. W. Huang, S. W. Ke, and C. F. Tsai. 2016b. 'The distance function effect on k-nearest neighbor classification for medical datasets', *SpringerPlus*, 5: 1304.

Huai, M., C. Miao, Q. Suo, Y. Li, J. Gao, and A. Zhang. 2018. "Uncorrelated patient similarity learning." In *SIAM International Conference on Data Mining, SDM 2018*, 270-78.

Hunziker, S., L. A. Celi, J. Lee, and M. D. Howell. 2012. 'Red cell distribution width improves the simplified acute physiology score for risk prediction in unselected critically ill patients', *Crit Care*, 16: R89.

Ishwaran, Hemant, Udaya B. Kogalur, Eugene H. Blackstone, and Michael S. Lauer. 2008. 'Random survival forests', *The Annals of Applied Statistics*, 2: 841-60.

Jensen, P. B., L. J. Jensen, and S. Brunak. 2012. 'Mining electronic health records: towards better research applications and clinical care', *Nat Rev Genet*, 13: 395-405.

Jensen, R. E., K. S. Chan, J. P. Weiner, J. B. Fowles, and S. M. Neale. 2009. 'Implementing electronic health record-based quality measures for developmental screening', *Pediatrics*, 124: e648-54.

Johnson, A. E., M. M. Ghassemi, S. Nemati, K. E. Niehaus, D. A. Clifton, and G. D. Clifford. 2016a. 'Machine Learning and Decision Support in Critical Care', *Proc IEEE Inst Electr Electron Eng*, 104: 444-66.

Johnson, A. E., T. J. Pollard, L. Shen, L. W. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark. 2016b. 'MIMIC-III, a freely accessible critical care database', *Sci Data*, 3: 160035.

Johnson, A. E. W., J. Aboab, J. D. Raffa, T. J. Pollard, R. O. Deliberato, L. A. Celi, and D. J. Stone. 2018a. 'A Comparative Analysis of Sepsis Identification Methods in an Electronic Database', *Crit Care Med*, 46: 494-99.

Johnson, Alistair E. W., Tom J. Pollard, and Roger G. Mark. 2017. "Reproducibility in critical care: a mortality prediction case study." In *Proceedings of the 2nd Machine Learning for Healthcare Conference*, edited by Doshi-Velez Finale, Fackler Jim, Kale David, Ranganath Rajesh, Wallace Byron and Wiens Jenna, 361--76. Proceedings of Machine Learning Research: PMLR.

Johnson, Alistair, and Tom Pollard. 2018b. "sepsis3-mimic." In.: Zenodo.

Johnson, K. D., C. Winkelman, C. J. Burant, M. Dolansky, and V. Totten. 2014. 'The factors that affect the frequency of vital sign monitoring in the emergency department', *J Emerg Nurs*, 40: 27-35.

Jones, R. B., and A. J. Hedley. 1986. 'A computer in the diabetic clinic.Completeness of data in a clinical information system for diabetes', *Practical Diabetes International*, 3: 295-96.

Jurisica, Igor, John Mylopoulos, Janice Glasgow, Heather Shapiro, and Robert F. Casper. 1998. 'Case-based reasoning in IVF: prediction and knowledge mining', *Artif Intell Med*, 12: 1-24.

Kajimoto, K., N. Sato, T. Takano, and registry investigators of the Acute Decompensated Heart Failure Syndromes. 2015. 'Relation between elevated blood urea nitrogen, clinical features or comorbidities, and clinical outcome in patients hospitalized for acute heart failure syndromes', *Int J Cardiol*, 201: 311-4.

Kaufman, Leonard, and Peter J Rousseeuw. 2009. *Finding groups in data: an introduction to cluster analysis* (John Wiley & Sons).

Keogh, Eamonn, and Shruti Kasetty. 2003. 'On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration', *Data Mining and Knowledge Discovery*, 7: 349-71.

Kern, L. M., Y. Barron, R. V. Dhopeshwarkar, A. Edwards, R. Kaushal, and Hitec Investigators. 2013. 'Electronic health records and ambulatory quality of care', *J Gen Intern Med*, 28: 496-503.

Khoshnevisan, F., J. Ivy, M. Capan, R. Arnold, J. Huddleston, and M. Chi. 2018. "Recent Temporal Pattern Mining for Septic Shock Early Prediction." In *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, 229-40.

Khoury, Muin J. 2016. "The Shift From Personalized Medicine to Precision Medicine and Precision Public Health: Words Matter!" In. Genomics and Health Impact Blog.

Kim, M. J., Y. H. Kim, I. S. Sol, S. Y. Kim, J. D. Kim, H. Y. Kim, K. W. Kim, M. H. Sohn, and K. E. Kim. 2017. 'Serum anion gap at admission as a predictor of mortality in the pediatric intensive care unit', *Sci Rep*, 7: 1456.

Knaus, W. A., E. A. Draper, D. P. Wagner, and J. E. Zimmerman. 1985. 'APACHE II: a severity of disease classification system', *Crit Care Med*, 13: 818-29.

Knol, M. J., K. J. Janssen, A. R. Donders, A. C. Egberts, E. R. Heerdink, D. E. Grobbee, K. G. Moons, and M. I. Geerlings. 2010. 'Unpredictable bias when using the missing indicator method or complete case analysis for missing confounder values: an empirical example', *J Clin Epidemiol*, 63: 728-36.

Koppel, Ross. 2009. 'EMR entry error: not so benign', U.S. Department of Health and Human Services, Accessed 2019-05-03. https://psnet.ahrq.gov/webmm/case/199/emr-entry-error-not-so-benign.

Kravitz, R. L., N. Duan, and J. Braslow. 2004. 'Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages', *Milbank Q*, 82: 661-87.

Kuhn, Max, and Kjell Johnson. 2013. *Applied predictive modeling* (Springer).
Lasko, Thomas A., Joshua C. Denny, and Mia A. Levy. 2013. 'Computational Phenotype Discovery Using Unsupervised Feature Learning over Noisy, Sparse, and Irregular Clinical Data', *PLoS One*, 8: e66341.
Le Gall, J. R., J. Klar, S. Lemeshow, F. Saulnier, C. Alberti, A. Artigas, and D. Teres. 1996. 'The Logistic Organ Dysfunction system. A new way to assess organ dysfunction in the intensive care unit. ICU Scoring Group', *JAMA*, 276: 802-10.
Le Gall, J. R., S. Lemeshow, and F. Saulnier. 1993. 'A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study', *JAMA*, 270: 2957-63.
Lee, J, D M Maslove, and J A Dubin. 2015a. 'Personalized mortality prediction driven by electronic medical data and a patient similarity metric', *PLoS One*, 10: e0127428.
Lee, J., and D. M. Maslove. 2015b. 'Using information theory to identify redundancy in common laboratory tests in the intensive care unit', *BMC Med Inform Decis Mak*, 15: 59.
Lee, S. W., S. Kim, K. Y. Na, R. H. Cha, S. W. Kang, C. W. Park, D. R. Cha, S. G. Kim, S. A. Yoon, S. Y. Han, J. H. Park, J. H. Chang, C. S. Lim, and Y. S. Kim. 2016. 'Serum Anion Gap Predicts All-Cause Mortality in Patients with Advanced Chronic Kidney Disease: A Retrospective Analysis of a Randomized Controlled Study', *PLoS One*, 11: e0156381.
Lehman, L. W., R. P. Adams, L. Mayaud, G. B. Moody, A. Malhotra, R. G. Mark, and S. Nemati. 2015a. 'A physiological time series dynamics-based approach to patient monitoring and outcome prediction', *IEEE J Biomed Health Inform*, 19: 1068-76.
Lehman, L. w. H., S. Nemati, and R. G. Mark. 2015b. "Hemodynamic monitoring using switching autoregressive dynamics of multivariate vital sign time series." In *2015 Computing in Cardiology Conference (CinC)*, 1065-68.
Lehman, L. w. H., S. Nemati, G. B. Moody, T. Heldt, and R. G. Mark. 2014. "Uncovering clinical significance of vital sign dynamics in critical care." In *Computing in Cardiology 2014*, 1141-44.
Levinson, Andrew T, Brian P Casserly, and Mitchell M Levy. 2011. "Reducing mortality in severe sepsis and septic shock." In *Seminars in respiratory and critical care medicine*, 195-205. Thieme Medical Publishers.
Li, J., L. Guo, and N. Handly. 2009. "Hospital Admission Prediction Using Pre-hospital Variables." In *2009 IEEE International Conference on Bioinformatics and Biomedicine*, 283-86.
Li, J., and Z. Lu. 2012. "A new method for computational drug repositioning using drug pairwise similarity." In *2012 IEEE International Conference on Bioinformatics and Biomedicine*, 1-4.
Li, P., Y. Nie, and J. Yu. 2016. 'Fusing literature and full network data improves disease similarity computation', *BMC Bioinformatics*, 17: 326.
Lin, H., C. E. McCulloch, B. W. Turnbull, E. H. Slate, and L. C. Clark. 2000. 'A latent class mixed model for analysing biomarker trajectories with irregularly scheduled observations', *Stat Med*, 19: 1303-18.
Lin, Haiqun, Bruce W. Turnbull, Charles E. McCulloch, and Elizabeth H. Slate. 2002. 'Latent Class Models for Joint Analysis of Longitudinal Biomarker and Event Process Data', *Journal of the American Statistical Association*, 97: 53-65.
Litchfield, I. J., R. J. Lilford, L. M. Bentham, and S. M. Greenfield. 2014. 'A qualitative exploration of the motives behind the decision to order a liver function test in primary care', *Quality in Primary Care*, 22: 201-10.
Little, R. J., R. D'Agostino, M. L. Cohen, K. Dickersin, S. S. Emerson, J. T. Farrar, C. Frangakis, J. W. Hogan, G. Molenberghs, S. A. Murphy, J. D. Neaton, A. Rotnitzky, D. Scharfstein, W. J.

Shih, J. P. Siegel, and H. Stern. 2012. 'The prevention and treatment of missing data in clinical trials', *N Engl J Med*, 367: 1355-60.
Little, Roderick J. A. 1988. 'Missing-Data Adjustments in Large Surveys', *Journal of Business & Economic Statistics*, 6: 287-96.
Low, L. L., G. R. Harrington, and D. P. Stoltzfus. 1995. 'The effect of arterial lines on blood-drawing practices and costs in intensive care units', *Chest*, 108: 216-9.
Lowsky, D. J., Y. Ding, D. K. K. Lee, C. E. McCulloch, L. F. Ross, J. R. Thistlethwaite, and S. A. Zenios. 2013. 'A K-nearest neighbors survival probability prediction method', *Statistics in Medicine*, 32: 2062-69.
Maaten, Laurens van der, and Geoffrey Hinton. 2008. 'Visualizing data using t-SNE', *Journal of Machine Learning Research*, 9: 2579-605.
Maggs, J. L., and J. E. Schulenberg. 2004. 'Trajectories of alcohol use during the transition to adulthood', *Alcohol Research & Health*, 28: 195-201.
Martens, F. K., E. C. Tonk, J. G. Kers, and A. C. Janssens. 2016. 'Small improvement in the area under the receiver operating characteristic curve indicated small changes in predicted risks', *J Clin Epidemiol*, 79: 159-64.
Martin, G. S., D. M. Mannino, S. Eaton, and M. Moss. 2003. 'The epidemiology of sepsis in the United States from 1979 through 2000', *N Engl J Med*, 348: 1546-54.
McCulloch, C. E., H. Lin, E. H. Slate, and B. W. Turnbull. 2002. 'Discovering subpopulation structure with latent class mixed models', *Stat Med*, 21: 417-29.
McKee, P. A., W. P. Castelli, P. M. McNamara, and W. B. Kannel. 1971. 'The natural history of congestive heart failure: the Framingham study', *N Engl J Med*, 285: 1441-6.
Medicare, Centers for, and Medicaid Services. 2012. "Implementation of severe sepsis and septic shock: management bundle measure (NQF# 0500)." In *National Quality Forum*.
Menendez, M. E., V. Neuhaus, C. N. van Dijk, and D. Ring. 2014. 'The Elixhauser comorbidity method outperforms the Charlson index in predicting inpatient death after orthopaedic surgery', *Clin Orthop Relat Res*, 472: 2878-86.
Meystre, S. M., G. K. Savova, K. C. Kipper-Schuler, and J. F. Hurdle. 2008. 'Extracting information from textual documents in the electronic health record: a review of recent research', *Yearb Med Inform*: 128-44.
Miltner, R. S., K. D. Johnson, and R. Deierhoi. 2014. 'Exploring the frequency of blood pressure documentation in emergency departments', *J Nurs Scholarsh*, 46: 98-105.
Miner, Linda, Pat Bolding, Joseph Hilbe, Mitchell Goldstein, Thomas Hill, Robert Nisbet, Nephi Walton, and Gary Miner. 2014. *Practical Predictive Analytics and Decisioning Systems for Medicine: Informatics Accuracy and Cost-Effectiveness for Healthcare Administration and Delivery Including Medical Research* (Academic Press).
Mitchell, Tom M. 1997. *Machine Learning* (McGraw-Hill: New York).
Moghadam, H., M. Rahgozar, and S. Gharaghani. 2016. 'Scoring multiple features to predict drug disease associations using information fusion and aggregation', *SAR QSAR Environ Res*, 27: 609-28.
Morrison, Zoe, Ann Robertson, Kathrin Cresswell, Sarah Crowe, and Aziz Sheikh. 2011. 'Understanding Contrasting Approaches to Nationwide Implementations of Electronic Health Record Systems: England, the USA and Australia', *Journal of Healthcare Engineering*, 2: 25-42.
Mukherjee, D., and E. J. Topol. 2002. 'Pharmacogenomics in cardiovascular diseases', *Prog Cardiovasc Dis*, 44: 479-98.

Muthén, Bengt, and Kerby Shedden. 1999. 'Finite Mixture Modeling with Mixture Outcomes Using the EM Algorithm', *Biometrics*, 55: 463-69.
Myers, Teresa A. 2011. 'Goodbye, Listwise Deletion: Presenting Hot Deck Imputation as an Easy and Effective Tool for Handling Missing Data', *Communication Methods and Measures*, 5: 297-310.
Newgard, C. D., and J. S. Haukoos. 2007. 'Advanced statistics: missing data in clinical research--part 2: multiple imputation', *Acad Emerg Med*, 14: 669-78.
Oliveira, Anderson Magalhães, Marcio Vasconcelos Oliveira, and Claudio Lima Souza. 2014. 'Prevalence of unnecessary laboratory tests and related avoidable costs in intensive care unit', *Jornal Brasileiro de Patologia e Medicina Laboratorial*, 50: 410-16.
Oztekin, A., D. Delen, and Z. J. Kong. 2009. 'Predicting the graft survival for heart-lung transplantation patients: an integrated data mining methodology', *Int J Med Inform*, 78: e84-96.
Panahiazar, M., V. Taslimitehrani, N. L. Pereira, and J. Pathak. 2015. "Using EHRs for Heart Failure Therapy Recommendation Using Multidimensional Patient Similarity Analytics." In *26th Medical Informatics in Europe Conference, MIE 2015*, edited by R. Cornet, L. Stoicu-Tivadar, R. Cornet, C. L. Parra Calderon, S. K. Andersen, A. Horbst and M. Hercigonja-Szekeres, 369-73. IOS Press.
Parimbelli, E., S. Marini, L. Sacchi, and R. Bellazzi. 2018. 'Patient similarity for precision medicine: A systematic review', *J Biomed Inform*, 83: 87-96.
Park, Yoon-Joo, Byung-Chun Kim, and Se-Hak Chun. 2006. 'New knowledge extraction technique using probability for case-based reasoning: application to medical diagnosis', *Expert Systems*, 23: 2-20.
Patel, K. V., R. D. Semba, L. Ferrucci, A. B. Newman, L. P. Fried, R. B. Wallace, S. Bandinelli, C. S. Phillips, B. Yu, S. Connelly, M. G. Shlipak, P. H. Chaves, L. J. Launer, W. B. Ershler, T. B. Harris, D. L. Longo, and J. M. Guralnik. 2010. 'Red cell distribution width and mortality in older adults: a meta-analysis', *J Gerontol A Biol Sci Med Sci*, 65: 258-65.
PCAST. 2008 "Priorities for Personalized Medicine." In.: President's Council of Advisors on Science and Technology.
Peck, J. S., S. A. Gaehde, D. J. Nightingale, D. Y. Gelman, D. S. Huckins, M. F. Lemons, E. W. Dickson, and J. C. Benneyan. 2013. 'Generalizability of a simple approach for predicting hospital admission from an emergency department', *Acad Emerg Med*, 20: 1156-63.
Pencina, M. J., R. B. D'Agostino, K. M. Pencina, A. C. Janssens, and P. Greenland. 2012. 'Interpreting incremental value of markers added to risk prediction models', *American Journal of Epidemiology*, 176: 473-81.
Peng, H., F. Long, and C. Ding. 2005. 'Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy', *IEEE Trans Pattern Anal Mach Intell*, 27: 1226-38.
Perer, Adam, and David Gotz. 2013. "Data-driven exploration of care plans for patients." In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*, 439-44. Paris, France: ACM.
Perlman, L., A. Gottlieb, N. Atias, E. Ruppin, and R. Sharan. 2011. 'Combining drug and gene similarity measures for drug-target elucidation', *Journal of Computational Biology*, 18: 133-45.
Perotte, A., R. Pivovarov, K. Natarajan, N. Weiskopf, F. Wood, and N. Elhadad. 2014. 'Diagnosis code assignment: models and evaluation metrics', *J Am Med Inform Assoc*, 21: 231-7.

Pigott, Therese D. 2010. 'A Review of Methods for Missing Data', *Educational Research and Evaluation*, 7: 353-83.
Pimentel, M, D Clifton, Lei Clifton, and Lionel Tarassenko. 2013. "Modelling patient time-series data from electronic health records using Gaussian processes." In *Advances in Neural Information Processing Systems: Workshop on Machine Learning for Clinical Data Analysis*, 1-4. Harrahs and Harveys, Lake Tahoe.
Pollard, T. J., A. E. W. Johnson, J. D. Raffa, L. A. Celi, R. G. Mark, and O. Badawi. 2018. 'The eICU Collaborative Research Database, a freely available multi-center database for critical care research', *Sci Data*, 5: 180178.
Porcheret, M., R. Hughes, D. Evans, K. Jordan, T. Whitehurst, H. Ogden, P. Croft, and Network North Staffordshire General Practice Research. 2004. 'Data quality of general practice electronic health records: the impact of a program of assessments, feedback, and training', *J Am Med Inform Assoc*, 11: 78-86.
Prin, M., and H. Wunsch. 2012. 'International comparisons of intensive care: informing outcomes and improving standards', *Curr Opin Crit Care*, 18: 700-6.
Pringle, M., P. Ward, and C. Chilvers. 1995. 'Assessment of the completeness and accuracy of computer medical records in four practices committed to recording data on computer', *Br J Gen Pract*, 45: 537-41.
Proust-Lima, C., and J. M. Taylor. 2009a. 'Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach', *Biostatistics*, 10: 535-49.
Proust-Lima, Cécile, Pierre Joly, Jean-François Dartigues, and Hélène Jacqmin-Gadda. 2009b. 'Joint modelling of multivariate longitudinal outcomes and a time-to-event: A nonlinear latent class approach', *Computational Statistics & Data Analysis*, 53: 1142-54.
Purtle, S. W., T. Moromizato, C. K. McKane, F. K. Gibbons, and K. B. Christopher. 2014. 'The association of red cell distribution width at hospital discharge and out-of-hospital mortality following critical illness*', *Crit Care Med*, 42: 918-29.
Quantin, C., E. Sauleau, P. Bolard, C. Mousson, M. Kerkri, P. Brunet Lecomte, T. Moreau, and L. Dusserre. 1999. 'Modeling of High-cost Patient Distribution within Renal Failure Diagnosis Related Group', *Journal of Clinical Epidemiology*, 52: 251-58.
Ramsay, James. 2005. 'Functional data analysis', *Encyclopedia of Statistics in Behavioral Science*.
Rao, M H, Prashanti Marella, and Bipphy Kath. 2008. 'Assessment of Severity and Outcome of Critical Illness', *Indian Journal Of Anaesthesia*, 2008: 652-62.
Reis Miranda, Dinis, Angelique de Rijk, and Wilmar Schaufeli. 1996. 'Simplified Therapeutic Intervention Scoring System', *Crit Care Med*, 24: 64-73.
Reisner-Senelar, L. 2011. 'The birth of intensive care medicine: Bjorn Ibsen's records', *Intensive Care Med*, 37: 1084-6.
Rhee, C., Z. Zhang, S. S. Kadri, D. J. Murphy, G. S. Martin, E. Overton, C. W. Seymour, D. C. Angus, R. Dantes, L. Epstein, D. Fram, R. Schaaf, R. Wang, M. Klompas, and C. D. C. Prevention Epicenters Program. 2019. 'Sepsis Surveillance Using Adult Sepsis Events Simplified eSOFA Criteria Versus Sepsis-3 Sequential Organ Failure Assessment Criteria', *Crit Care Med*, 47: 307-14.
Robinson, P. N. 2012. 'Deep phenotyping for precision medicine', *Hum Mutat*, 33: 777-80.
Robnik-Šikonja, Marko, and Igor Kononenko. 2003. 'Theoretical and empirical analysis of ReliefF and RReliefF', *Machine Learning*, 53: 23-69.

Rodriguez, R. 2001a. 'Demystifying critical care: a new series provides a succinct, modern approach aimed at primary care physicians', *West J Med*, 175: 366-7.
Rodriguez, Robert. 2001b. 'Demystifying critical care: A new series provides a succinct, modern approach aimed at primary care physicians', *Western Journal of Medicine*, 175: 366-67.
Roth, Philip L. 1994. 'Missing Data: A Conceptual Review for Applied Psychologists', *Personnel Psychology*, 47: 537-60.
Rothwell, P. M. 2005. 'External validity of randomised controlled trials: "to whom do the results of this trial apply?"', *Lancet*, 365: 82-93.
Rousseeuw, Peter J. 1987. 'Silhouettes: A graphical aid to the interpretation and validation of cluster analysis', *Journal of Computational and Applied Mathematics*, 20: 53-65.
Rubin, Donald B., and Ying Nian Wu. 1997. 'Modeling Schizophrenic Behavior Using General Mixture Components', *Biometrics*, 53: 243-61.
Rusanov, A., N. G. Weiskopf, S. Wang, and C. Weng. 2014. 'Hidden in plain sight: bias towards sick patients when sampling patients with sufficient electronic health record data for research', *BMC Med Inform Decis Mak*, 14: 51.
Saeed, M., M. Villarroel, A. T. Reisner, G. Clifford, L. W. Lehman, G. Moody, T. Heldt, T. H. Kyaw, B. Moody, and R. G. Mark. 2011. 'Multiparameter Intelligent Monitoring in Intensive Care II: a public-access intensive care unit database', *Crit Care Med*, 39: 952-60.
Saeed, Mohammed, and Roger Mark. 2006a. 'A Novel Method for the Efficient Retrieval of Similar Multiparameter Physiologic Time Series Using Wavelet-Based Symbolic Representations', *AMIA Annual Symposium Proceedings*, 2006: 679-83.
Saeed, Mohammed, and Roger G Mark. 2006b. "A novel method for the efficient retrieval of similar multiparameter physiologic time series using wavelet-based symbolic representations." In *AMIA*.
Saeys, Y., I. Inza, and P. Larranaga. 2007. 'A review of feature selection techniques in bioinformatics', *Bioinformatics*, 23: 2507-17.
Safran, C., M. Bloomrosen, W. E. Hammond, S. Labkoff, S. Markel-Fox, P. C. Tang, D. E. Detmer, and Panel Expert. 2007. 'Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper', *J Am Med Inform Assoc*, 14: 1-9.
Sahu, A., H. A. Cooper, and J. A. Panza. 2006. 'The initial anion gap is a predictor of mortality in acute myocardial infarction', *Coron Artery Dis*, 17: 409-12.
Salgado, Cátia M., Susana M. Vieira, Luís F. Mendonça, Stan Finkelstein, and João M. C. Sousa. 2016. 'Ensemble fuzzy models in personalized medicine: Application to vasopressors administration', *Engineering Applications of Artificial Intelligence*, 49: 141-48.
Schork, N. J. 2015. 'Personalized medicine: Time for one-person trials', *Nature*, 520: 609-11.
Schulman, C. S., and L. Staul. 2010. 'Standards for frequency of measurement and documentation of vital signs and physical assessments', *Crit Care Nurse*, 30: 74-6.
Senol, K., B. Saylam, F. Kocaay, and M. Tez. 2013. 'Red cell distribution width as a predictor of mortality in acute pancreatitis', *Am J Emerg Med*, 31: 687-9.
Serden, L., R. Lindqvist, and M. Rosen. 2003. 'Have DRG-based prospective payment systems influenced the number of secondary diagnoses in health care administrative data?', *Health Policy*, 65: 101-7.
Seymour, C. W., C. M. Coopersmith, C. S. Deutschman, F. Gesten, M. Klompas, M. Levy, G. S. Martin, T. M. Osborn, C. Rhee, D. K. Warren, R. S. Watson, and D. C. Angus. 2016. 'Application of a Framework to Assess the Usefulness of Alternative Sepsis Criteria', *Crit Care Med*, 44: e122-30.

Sha, Ying, Janani Venugopalan, and May D. Wang. 2016. "A Novel Temporal Similarity Measure for Patients Based on Irregularly Measured Data in Electronic Health Records." In *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics - BCB '16*, 337-44. Seattle, WA, USA: ACM.

Sharabiani, M. T., P. Aylin, and A. Bottle. 2012. 'Systematic review of comorbidity indices for administrative data', *Med Care*, 50: 1109-18.

Sharafoddini, A., J. A. Dubin, and J. Lee. 2017. 'Patient Similarity in Prediction Models Based on Health Data: A Scoping Review', *JMIR Med Inform*, 5: e7.

Sharafoddini, A., J. A. Dubin, D. M. Maslove, and J. Lee. 2019. 'A New Insight Into Missing Data in Intensive Care Unit Patient Profiles: Observational Study', *JMIR Med Inform*, 7: e11605.

Sharafoddini, Anis, Joel A. Dubin, and Joon Lee. 2018. 'Finding Similar Patient Subpopulations in the ICU Using Laboratory Test Ordering Patterns', *Proceedings of the 7th International Conference on Bioinformatics and Biomedical Science*.

Shivade, C., P. Raghavan, E. Fosler-Lussier, P. J. Embi, N. Elhadad, S. B. Johnson, and A. M. Lai. 2014. 'A review of approaches to identifying patient phenotype cohorts using electronic health records', *J Am Med Inform Assoc*, 21: 221-30.

Singer, M., C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J. D. Chiche, C. M. Coopersmith, R. S. Hotchkiss, M. M. Levy, J. C. Marshall, G. S. Martin, S. M. Opal, G. D. Rubenfeld, T. van der Poll, J. L. Vincent, and D. C. Angus. 2016. 'The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)', *JAMA*, 315: 801-10.

Singh, A., G. Nadkarni, O. Gottesman, S. B. Ellis, E. P. Bottinger, and J. V. Guttag. 2015. 'Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration', *J Biomed Inform*, 53: 220-8.

Singh, Anima. 2015. 'Exploiting hierarchical and temporal information in building predictive models from EHR data', Massachusetts Institute of Technology.

Singh, Anima, Girish Nadkarni, John Guttag, and Erwin Bottinger. 2014. 'Leveraging hierarchy in medical codes for predictive modeling': 96-103.

Smoller, B. R., and M. S. Kruskall. 1986. 'Phlebotomy for diagnostic laboratory tests in adults. Pattern of use and effect on transfusion requirements', *N Engl J Med*, 314: 1233-5.

Snyderman, R. 2012. 'Personalized health care: from theory to practice', *Biotechnol J*, 7: 973-9.

Sokal, Robert R. 1958. 'A statistical method for evaluating systematic relationship', *University of Kansas science bulletin*, 28: 1409-38.

Solomon, Daniel H., Hideki Hashimoto, Lawren Daltroy, and Matthew H. Liang. 1998. 'Techniques to Improve Physicians' Use of Diagnostic Tests', *JAMA*, 280: 2020-27.

Soto, C. M., K. P. Kleinman, and S. R. Simon. 2002. 'Quality and correlates of medical record documentation in the ambulatory care setting', *BMC Health Serv Res*, 2: 22.

Southern, D. A., H. Quan, and W. A. Ghali. 2004. 'Comparison of the Elixhauser and Charlson/Deyo methods of comorbidity measurement in administrative data', *Med Care*, 42: 355-60.

Standards, International Organization for. 2005. "ISO/TR 20514:2005 Health informatics -- Electronic health record -- Definition, scope and context." In, 27. International Organization for Standards.

Sterne, J. A., I. R. White, J. B. Carlin, M. Spratt, P. Royston, M. G. Kenward, A. M. Wood, and J. R. Carpenter. 2009. 'Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls', *BMJ*, 338: b2393.

sSteyerberg, Ewout W. 2009. *Clinical prediction models : a practical approach to development, validation, and updating* (Springer: New York, NY).
Stockdale, Myrah, and Kenneth Royal. 2016. 'Missing data as a validity threat for medical and healthcare education research: problems and solutions', *International Journal of Healthcare*, 2: 67.
Su, Xiaoyuan, and Taghi M. Khoshgoftaar. 2009. 'A Survey of Collaborative Filtering Techniques', *Advances in Artificial Intelligence*, 2009: 1-19.
Sun, J., D. Sow, J. Hu, and S. Ebadollahi. 2010a. "A system for mining temporal physiological data streams for advanced prognostic decision support." In *Proceedings - IEEE International Conference on Data Mining, ICDM*, 1061-66.
Sun, Jimeng, Daby Sow, Jianying Hu, and S. Ebadollahi. 2010b. "Localized Supervised Metric Learning on Temporal Physiological Data." In *Pattern Recognition (ICPR), 2010 20th International Conference on*, 4149-52.
Suo, Q., F. Ma, Y. Yuan, M. Huai, W. Zhong, A. Zhang, and J. Gao. 2017. "Personalized disease prediction using a CNN-based similarity learning method." In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 811-16.
Syeda-Mahmood, T., F. Wang, D. Beymer, A. Amir, M. Richmond, and S. N. Hashmi. 2007. "AALIM: Multimodal mining for cardiac decision support." In *2007 Computers in Cardiology*, 209-12.
Tang, P. C., M. P. LaRosa, and S. M. Gorden. 1999. 'Use of Computer-based Records, Completeness of Documentation, and Appropriateness of Documented Clinical Decisions', *Journal of the American Medical Informatics Association*, 6: 245-51.
Tashkandi, Araek, Ingmar Wiese, and Lena Wiese. 2018. 'Efficient In-Database Patient Similarity Analysis for Personalized Medical Decision Support Systems', *Big Data Research*, 13: 52-64.
Thiru, K., S. de Lusignan, and N. Hague. 2016. 'Have the completeness and accuracy of computer medical records in general practice improved in the last five years? The report of a two-practice pilot study', *Health informatics journal*, 5: 224-32.
Ullah, S., and C. F. Finch. 2013. 'Applications of functional data analysis: A systematic review', *Bmc Medical Research Methodology*, 13: 43.
Vesin, A., E. Azoulay, S. Ruckly, L. Vignoud, K. Rusinova, D. Benoit, M. Soares, P. Azeivedo-Maia, F. Abroug, J. Benbenishty, and J. F. Timsit. 2013. 'Reporting and handling missing values in clinical studies in intensive care units', *Intensive Care Med*, 39: 1396-404.
Vincent, J. L., and R. Moreno. 2010. 'Clinical review: scoring systems in the critically ill', *Crit Care*, 14: 207.
Vranas, K. C., J. K. Jopling, T. E. Sweeney, M. C. Ramsey, A. S. Milstein, C. G. Slatore, G. J. Escobar, and V. X. Liu. 2017. 'Identifying Distinct Subgroups of ICU Patients: A Machine Learning Approach', *Crit Care Med*, 45: 1607-15.
Wang, F., J. Hu, and J. Sun. 2012a. "Medical prognosis based on patient similarity and expert feedback." In *Proceedings - International Conference on Pattern Recognition*, 1799-802.
Wang, F., J. Sun, and S. Ebadollahi. 2011a. "Integrating distance metrics learned from multiple experts and its application in patient similarity assessment." In *11th SIAM International Conference on Data Mining, SDM 2011*, 59-70. Mesa, AZ.
Wang, Fei. 2015. 'Adaptive semi-supervised recursive tree partitioning: The ART towards large scale patient indexing in personalized healthcare', *Journal of Biomedical Informatics*, 55: 41-54.
Wang, Fei, and Jimeng Sun. 2015. 'PSF: A Unified Patient Similarity Evaluation Framework through Metric Learning with Weak Supervision', *IEEE J Biomed Health Inform*, 19: 1-1.

Wang, Xiaoyue, Abdullah Mueen, Hui Ding, Goce Trajcevski, Peter Scheuermann, and Eamonn Keogh. 2012b. 'Experimental comparison of representation methods and distance measures for time series data', *Data Mining and Knowledge Discovery*, 26: 275-309.
Wattenberg, Martin, Fernanda Viégas, and Ian Johnson. 2016. 'How to use t-sne effectively', *Distill*, 1: e2.
Weiskopf, N. G., G. Hripcsak, S. Swaminathan, and C. Weng. 2013a. 'Defining and measuring completeness of electronic health records for secondary use', *J Biomed Inform*, 46: 830-6.
Weiskopf, N. G., A. Rusanov, and C. Weng. 2013b. 'Sick patients have more data: the non-random completeness of electronic health records', *AMIA Annu Symp Proc*, 2013: 1472-7.
Weiskopf, N. G., and C. Weng. 2013c. 'Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research', *J Am Med Inform Assoc*, 20: 144-51.
Wells, B. J., K. M. Chagin, A. S. Nowacki, and M. W. Kattan. 2013. 'Strategies for handling missing data in electronic health record derived data', *EGEMS (Wash DC)*, 1: 1035.
Wennberg, J. E. 1984. 'Dealing with medical practice variations: a proposal for action', *Health Aff (Millwood)*, 3: 6-32.
Wiese, Ingmar, Nicole Sarna, Lena Wiese, Araek Tashkandi, and Ulrich Sax. 2018. 'Concept acquisition and improved in-database similarity analysis for medical data', *Distributed and Parallel Databases*.
Wongsuphasawat, K., and B. Shneiderman. 2009. "Finding comparable temporal categorical records: A similarity measure with an interactive visualization." In *2009 IEEE Symposium on Visual Analytics Science and Technology*, 27-34.
Wu, M., M. Ghassemi, M. Feng, L. A. Celi, P. Szolovits, and F. Doshi-Velez. 2017. 'Understanding vasopressor intervention and weaning: risk prediction in a public heterogeneous clinical time series database', *J Am Med Inform Assoc*, 24: 488-95.
Yan, Y., X. Shao, and Z. Jiang. 2014. 'Inferring novel indications of approved drugs via a learning method with local and global consistency', *PLoS One*, 9: e107100.
Yuhua, Li, Z. A. Bandar, and D. McLean. 2003. 'An approach for measuring semantic similarity between words using multiple information sources', *Ieee Transactions on Knowledge and Data Engineering*, 15: 871-82.
Zhang, H., F. Zhu, H. H. Dodge, G. A. Higgins, G. S. Omenn, Y. Guan, and Initiative Alzheimer's Disease Neuroimaging. 2018. 'A similarity-based approach to leverage multi-cohort medical data on the diagnosis and prognosis of Alzheimer's disease', *GigaScience*, 7: giy085-giy85.
Zhang, P., F. Wang, and J. Hu. 2014a. 'Towards drug repositioning: a unified computational framework for integrating multiple aspects of drug similarity and disease similarity', *AMIA Annu Symp Proc*, 2014: 1258-67.
Zhang, Ping, Fei Wang, Jianying Hu, and Robert Sorrentino. 2014b. 'Towards Personalized Medicine: Leveraging Patient Similarity and Drug Similarity Analytics', *AMIA Summits on Translational Science Proceedings*, 2014: 132-36.
Zhang, Z., and X. Xu. 2014c. 'Lactate clearance is a useful biomarker for the prediction of all-cause mortality in critically ill patients: a systematic review and meta-analysis*', *Crit Care Med*, 42: 2118-25.
Zhou, X. H., G. J. Eckert, and W. M. Tierney. 2001. 'Multiple imputation in public health research', *Stat Med*, 20: 1541-9.
Zhu, Z., C. Yin, B. Qian, Y. Cheng, J. Wei, and F. Wang. 2017. "Measuring patient similarities via a deep architecture with medical concept embedding." In *Proceedings - IEEE International Conference on Data Mining, ICDM*, 749-58.

Zimmerman, J. E., M. G. Seneff, X. Sun, D. P. Wagner, and W. A. Knaus. 1997. 'Evaluating laboratory usage in the intensive care unit: patient and institutional characteristics that influence frequency of blood sampling', *Crit Care Med*, 25: 737-48.
Zolbanin, Hamed Majidi, Dursun Delen, and Amir Hassan Zadeh. 2015. 'Predicting overall survivability in comorbidity of cancers: A data mining approach', *Decision Support Systems*, 74: 150-61.

Appendix A

An Overview of Prediction Models

Generally, prediction models can be broadly grouped into two categories: classification and regression. Although both groups aim to discover the underlying relationship between predictors and outcomes, they differ in types of outcomes. Discussed below is the set of commonly used machine learning algorithms that were named in this dissertation.

- Logistic Regression (LR): LR is a type of regression analysis that seeks the relation between predictors and a binary outcome. In other words, this model predicts the probability of an outcome occurring.
- Decision Tree (DT): DT model is a non-parametric method used for classification and regression. A DT algorithm recursively splits data set into smaller homogenous subdivisions based on a set of criteria. In DT, each node represents a test on a predictor, and the derived branches relate to the possible values for the predictor. Then, instances are classified by starting from the root node and following the branches based on their predictor values.
- Linear Discriminant Analysis (LDA): This algorithm searches for a linear combination of predictors that separates instances in different classes. LDA assumes Gaussian distribution density within each class.
- Support Vector Machine (SVM): An SVM algorithm finds a separating hyperplane or set of hyperplanes in predictors' space that has the maximum distance from instances

on each side. In other words, SVM tries to maximize the margins between the instances in each class.

- Naive Bayes (NB): This classification method is based on Bayes' rule and assumes that predictors are independent of one another. The basic model has been modified, and various versions have been introduced to improve performance.
- k-Nearest Neighbors (KNN): This method is based on the assumption that similar instances have similar characteristics. For each instance, KNN finds a cohort of k nearest instances based on a distance metric and classifies the new instance based on the majority vote of the cohort.
- Random Forest (RF): This algorithm constructs a set of DT models on different sub-samples of the data, and the output class will be the aggregation of the classes of all trees. In other words, RF is an ensemble of DTs.

Appendix B

An Overview of Clustering Algorithms

Below, a brief description of the clustering algorithms discussed in this dissertation is provided.

- Partition-based clustering algorithms: the core idea of this type of method is to consider the center of data points in each cluster as the representative (center) of the cluster. K-means is one of the most famous clustering methods. K-means iteratively updates the representatives of the clusters until some criteria of convergence are satisfied. This method is timely and computationally efficient. However, it is sensitive to the number of clusters and outliers.
- Hierarchical clustering algorithms: hierarchical algorithms successively construct clusters by merging or splitting previously built clusters. These algorithms are further categorized into agglomerative (bottom-up) or divisive (top-down) algorithms, and are suitable for datasets that have clusters with arbitrary shapes.
- Fuzzy clustering algorithms: in fuzzy clustering, an instance can be a member of more than one clusters. A set of membership degrees is assigned to each data point and used to determine to which degree it belongs to each cluster. This group of methods is suitable for datasets with overlapping clusters.

- Density-based clustering algorithms: the idea of these algorithms is that the clusters are located where data has a high density and are separated by regions with a lower density of data.

Appendix C

Detailed Results of Chapter 2

Table C1 The association between indicators (missing=1 and not missing =0) and mortality flag (deceased=1, alive=0) using Phi coefficient.

	In-hospital Mortality			30-days Mortality		
Indicator	Day1	Day2	Day3	Day1	Day2	Day3
ALT	-0.14	-0.12	-0.12	-0.14	-0.12	-0.11
ALK	-0.14	-0.12	-0.11	-0.14	-0.12	-0.1
pH	-0.11	-0.13	-0.15	-0.08	-0.1	-0.11
PCO	-0.11	-0.16	-0.18	-0.08	-0.13	-0.15
PO	-0.11	-0.16	-0.18	-0.08	-0.13	-0.15
BE	-0.11	-0.16	-0.18	-0.08	-0.13	-0.15
AST	-0.14	-0.12	-0.12	-0.14	-0.12	-0.11
Na	0.01	0.03	0.04	0.01	0.02	0.05
K	0	0.01	0.01	-0.01	0	0.03
Cl	0.03	0.03	0.04	0.02	0.03	0.06
HCO	0.03	0.03	0.04	0.02	0.03	0.05
AG	-0.01	0	0.01	-0.02	0	0.01
BG	-0.02	0	0.01	-0.02	-0.01	0.02
BUN	0.04	0.04	0.06	0.04	0.04	0.07
Cr	0.04	0.05	0.06	0.04	0.04	0.07
Ca	-0.1	-0.04	-0.04	-0.12	-0.05	-0.04
WBC	0.03	0.04	0.06	0.02	0.04	0.06
RBC	0.02	0.03	0.04	0.02	0.03	0.05
HGB	0.03	0.03	0.05	0.02	0.03	0.05
HCT	0.04	0.05	0.05	0.03	0.05	0.06

MCV	0.02	0.03	0.04	0.02	0.03	0.05
MCH	0.02	0.03	0.04	0.02	0.03	0.05
MCHC	0.02	0.03	0.05	0.02	0.03	0.05
RDW	0.02	0.03	0.05	0.02	0.03	0.05
PLT	0.03	0.04	0.05	0.03	0.04	0.05
NE	-0.1	-0.07	-0.08	-0.11	-0.08	-0.07
LY	-0.1	-0.07	-0.08	-0.11	-0.08	-0.07
MO	-0.1	-0.07	-0.08	-0.11	-0.08	-0.07
EO	-0.1	-0.07	-0.08	-0.11	-0.08	-0.07
BA	-0.1	-0.07	-0.08	-0.11	-0.08	-0.07
LAC	-0.12	-0.15	-0.15	-0.1	-0.13	-0.12
Mg	-0.03	0.01	0.02	-0.05	0.01	0.03
PTT	-0.04	-0.09	-0.09	-0.04	-0.09	-0.09
Phos	-0.1	-0.05	-0.04	-0.12	-0.06	-0.04
PT	-0.04	-0.09	-0.08	-0.04	-0.09	-0.09
TBil	-0.14	-0.12	-0.12	-0.14	-0.12	-0.11

Table C2 Detailed results for predictor importance evaluation with regard to 30-day mortality. Numbers represent the ranking after aggregating the ranking results from the three different feature-selection methods.

Day One				**Day Two**				**Day Three**			
Hot Deck		**PMM**		**Hot Deck**		**PMM**		**Hot Deck**		**PMM**	
BUN	0.742936	BUN	0.762397	BUN	0.9007	AG	0.795419	RDW	0.693305	RDW	0.748997
RDW	0.682517	RDW	0.680087	RDW	0.851829	HCO3	0.783337	BUN	0.666667	BUN	0.666667
MCHC	0.677031	MCHC	0.668965	HCO3	0.655154	BUN	0.77677	HCO3	0.543871	HCO3	0.544964

AG	0.524767	AG	0.540484	AG	0.580852	BE	0.609532	BE	0.49406	BE	0.540542
I-Ca	0.475564	I-Ca	0.436429	MCHC	0.448053	RDW	0.608711	AG	0.449286	pH	0.488433
I-Phos	0.46459	Cr	0.436071	Cr	0.405085	I-PO2	0.587151	Cr	0.398543	AG	0.450426
PO2	0.447435	HCO3	0.416741	Cl	0.382846	I-PCO2	0.585947	I-PCO2	0.375211	I-LAC	0.418716
HCO3	0.444744	PO2	0.404289	MCV	0.375821	I-BE	0.585592	I-PO2	0.374931	I-pH	0.40463
Cr	0.428755	MCV	0.386964	I-LAC	0.359897	Cl	0.53158	I-BE	0.374127	Cr	0.400008
I-LAC	0.38719	I-Phos	0.374431	Na	0.358051	PT	0.462085	PCO2	0.359771	Phos	0.387661
HGB	0.373706	PTT	0.353913	PTT	0.356926	LAC	0.461869	NE	0.357608	I-PCO2	0.387019
MCV	0.369112	HGB	0.342786	Phos	0.337663	Cr	0.451999	MCHC	0.331802	I-PO2	0.386739
LY	0.367866	pH	0.32767	PT	0.333779	PTT	0.424956	PT	0.328478	I-BE	0.385935
PTT	0.33424	LAC	0.320339	I-PO2	0.3335	Na	0.422474	LAC	0.289964	PCO2	0.367257
RBC	0.333482	BE	0.320299	I-PCO2	0.332814	Phos	0.419171	LY	0.287681	NE	0.360791
I-TBil	0.320728	I-LAC	0.318216	I-BE	0.332517	I-LAC	0.415475	pH	0.283643	MCV	0.351266
BG	0.319001	PCO2	0.316668	BE	0.290359	MCV	0.368343	MCV	0.281627	I-PTT	0.338352
I-ALT	0.315839	I-TBil	0.31277	I-Phos	0.284644	MCHC	0.363146	I-LAC	0.270327	LAC	0.331205

I-AST	0.314876	I-ALK	0.305529	I-TBil	0.283694	I-pH	0.352443	Phos	0.262758	MCHC	0.329817
PT	0.303447	I-ALT	0.303033	I-pH	0.273527	I-PT	0.338144	I-pH	0.239426	PT	0.329586
I-ALK	0.302031	I-AST	0.302074	I-PTT	0.261553	I-TBil	0.325466	BG	0.221239	I-PT	0.29994
LAC	0.277614	PT	0.296326	I-PT	0.258242	I-PTT	0.325362	I-PTT	0.221211	RBC	0.238642
Cl	0.276099	RBC	0.293754	PLT	0.252963	pH	0.312481	WBC	0.217481	I-Phos	0.237935
pH	0.273849	Phos	0.288846	I-Ca	0.248052	PCO2	0.307902	I-PT	0.214963	I-TBil	0.237811
PLT	0.273301	LY	0.285421	NE	0.247463	BG	0.301989	I-TBil	0.213362	BG	0.222807
I-PT	0.256103	ALK	0.284528	I-ALK	0.236376	I-ALK	0.30149	I-ALT	0.205211	WBC	0.219308
PCO2	0.255203	BG	0.282668	LAC	0.232278	PO2	0.297596	PTT	0.204495	I-ALT	0.21435
I-PTT	0.249578	PLT	0.279892	I-ALT	0.229095	PLT	0.29116	I-AST	0.202047	LY	0.212783
MCH	0.232707	NE	0.266646	I-AST	0.223172	I-ALT	0.285826	I-ALK	0.198413	I-AST	0.209672
Phos	0.218656	Cl	0.223353	BG	0.218129	TBil	0.283123	PLT	0.198322	PO2	0.208279
BE	0.208477	TBil	0.216197	HGB	0.208915	I-AST	0.281695	Cl	0.188406	PTT	0.206123
Na	0.207431	MCH	0.212788	PCO2	0.201961	LY	0.271601	PO2	0.180645	PLT	0.200699
I-NE	0.200687	I-NE	0.196231	WBC	0.201917	MCH	0.261553	I-Phos	0.167356	Cl	0.195587

I-MO	0.200687	I-MO	0.196231	Ca	0.193641	I-Phos	0.2614	I-Ca	0.165253	I-ALK	0.19224
I-EO	0.200687	I-EO	0.196231	BA	0.186936	Ca	0.242739	Mg	0.154738	Ca	0.185077
I-BA	0.200687	I-BA	0.196231	RBC	0.181905	WBC	0.242647	Na	0.146887	Mg	0.156356
I-LY	0.200608	I-LY	0.196152	pH	0.168043	I-Ca	0.200865	Ca	0.144049	Na	0.149534
I-PCO2	0.187662	WBC	0.171699	HCT	0.16619	HGB	0.193331	TBil	0.140346	MCH	0.146659
I-PO2	0.187662	Ca	0.167312	LY	0.159886	RBC	0.188082	RBC	0.107089	I-Ca	0.144385
I-BE	0.187534	AST	0.162532	Mg	0.131304	ALK	0.184185	I-NE	0.097467	TBil	0.142264
NE	0.173115	I-PCO2	0.156981	TBil	0.126985	BA	0.169845	I-LY	0.097467	EO	0.142107
WBC	0.166962	I-PO2	0.156981	MCH	0.120035	AST	0.16688	I-MO	0.097467	BA	0.142031
Ca	0.163823	I-BE	0.156853	I-NE	0.103639	Mg	0.159173	I-EO	0.097467	I-NE	0.099961
I-pH	0.159875	BA	0.149717	I-LY	0.103639	NE	0.150139	I-BA	0.097467	I-LY	0.099961
I-Mg	0.14009	Na	0.146076	I-MO	0.103639	HCT	0.117954	K	0.089782	I-MO	0.099961
ALK	0.131004	I-pH	0.144973	I-EO	0.103639	EO	0.114959	AST	0.084185	I-EO	0.099961
TBil	0.111702	I-Mg	0.111008	I-BA	0.103639	I-NE	0.113355	ALK	0.082093	I-BA	0.099961
EO	0.109363	I-BG	0.106693	ALK	0.094855	I-LY	0.113355	I-Mg	0.070701	I-RDW	0.094193

BA	0.108785	EO	0.106064	MO	0.09224	I-MO	0.113355	BA	0.069968	K	0.092206
I-BG	0.103206	I-AG	0.10267	K	0.075067	I-EO	0.113355	I-Cl	0.069821	I-RBC	0.08586
I-AG	0.098421	I-PT	0.092205	EO	0.063887	I-BA	0.113355	I-Cr	0.069027	I-MCV	0.08586
AST	0.085218	I-PTT	0.08084	I-Mg	0.063662	MO	0.108044	I-BUN	0.06785	I-MCH	0.08586
I-K	0.085122	MO	0.062942	PO2	0.055418	K	0.103865	I-BG	0.065149	ALK	0.084326
ALT	0.061017	ALT	0.055905	I-HCT	0.055202	ALT	0.078095	MO	0.056612	AST	0.081458
Mg	0.052026	Mg	0.044178	I-MCHC	0.045723	I-RDW	0.077468	I-HCT	0.054567	I-MCHC	0.080782
I-BUN	0.048802	HCT	0.044135	I-RBC	0.045597	I-MCHC	0.076733	I-PLT	0.052654	I-Mg	0.065196
HCT	0.047372	I-PLT	0.035278	I-MCV	0.045597	I-RBC	0.076601	I-HCO3	0.050589	I-Cr	0.059682
I-Cr	0.045343	K	0.034445	I-MCH	0.045597	I-MCV	0.076601	I-K	0.049123	I-BUN	0.058504
MO	0.037529	I-K	0.034074	I-RDW	0.045242	I-MCH	0.076601	EO	0.048483	I-WBC	0.055585
K	0.0359	I-HGB	0.033298	AST	0.04516	I-HGB	0.069039	MCH	0.048214	I-HGB	0.055267
I-Na	0.028494	I-MCV	0.033014	I-BUN	0.031937	I-HCT	0.063231	I-WBC	0.04705	MO	0.055048
I-Cl	0.026307	I-MCH	0.033014	I-PLT	0.03054	I-PLT	0.057216	I-Na	0.046679	I-HCT	0.054935

I-PLT	0.023585	I-MCHC	0.03284	ALT	0.030451	I-WBC	0.055654	ALT	0.045415	I-PLT	0.048878
I-HCO3	0.023289	I-RDW	0.032836	I-WBC	0.028815	I-Cr	0.043887	I-HGB	0.039608	I-Cl	0.044026
I-WBC	0.015726	I-RBC	0.032737	I-Cr	0.028358	I-BUN	0.041846	I-RDW	0.035126	I-K	0.043233
I-HCT	0.012493	I-WBC	0.028194	I-HGB	0.027524	I-Cl	0.034206	I-MCHC	0.032941	I-Na	0.04292
I-HGB	0.011658	I-HCT	0.023648	I-K	0.027506	I-HCO3	0.033703	I-RBC	0.032401	I-HCO3	0.041208
I-MCV	0.011373	I-Cr	0.01678	I-AG	0.021307	I-Na	0.031393	I-MCV	0.032401	I-BG	0.038748
I-MCH	0.011373	I-BUN	0.016679	I-BG	0.020366	I-Mg	0.022961	I-MCH	0.032401	I-AG	0.030697
I-MCHC	0.011198	I-Na	0.014031	I-Cl	0.019271	I-BG	0.017067	HGB	0.0278	HGB	0.030518
I-RDW	0.011194	I-Cl	0.007722	I-HCO3	0.01926	I-AG	0.015687	I-AG	0.025759	HCT	0.02312
I-RBC	0.011095	I-HCO3	0.005457	I-Na	0.013649	I-K	0.013539	HCT	0.020315	ALT	0

Table C3 Detailed results for predictor importance evaluation with regard to in-hospital mortality. Numbers represent the ranking after aggregating the ranking results from the three different feature-selection methods.

Day One				**Day Two**				**Day Three**			
Hot Deck		**PMM**		**Hot Deck**		**PMM**		**Hot Deck**		**PMM**	
BUN	0.792686	BUN	0.825715	BUN	0.871227	BUN	1	BE	0.66376	RDW	0.75246
AG	0.66198	AG	0.668918	AG	0.856826	RDW	0.711852	BUN	0.640534	BUN	0.635729
RDW	0.599006	RDW	0.573188	RDW	0.810929	HCO3	0.684191	HCO3	0.626034	BE	0.633926
HCO3	0.590773	HCO3	0.531746	HCO3	0.802246	AG	0.664339	RDW	0.61847	HCO3	0.62367
MCHC	0.584486	MCHC	0.507343	I-PO2	0.594496	BE	0.528778	I-BE	0.587481	I-BE	0.595553
BG	0.53102	PCO2	0.489483	I-PCO2	0.593258	MCHC	0.503805	I-PCO2	0.587166	I-PCO2	0.595238
PO2	0.521196	Cr	0.480181	I-BE	0.592893	PT	0.453111	I-PO2	0.586851	I-PO2	0.594924
Cr	0.487362	BE	0.452599	I-LAC	0.548438	Cl	0.429405	pH	0.515275	pH	0.556242
MCV	0.417346	I-LAC	0.436382	Cl	0.529452	I-LAC	0.425279	AG	0.494856	Phos	0.494694
I-LAC	0.411539	LAC	0.415773	PTT	0.511771	Cr	0.395266	I-LAC	0.489909	AG	0.492864
I-ALT	0.387559	HGB	0.414263	Phos	0.497639	I-PO2	0.382404	PCO2	0.438274	I-pH	0.470007
I-AST	0.387498	pH	0.402466	MCHC	0.485717	I-PCO2	0.381737	I-pH	0.437812	I-LAC	0.469215
I-ALK	0.385754	I-TBil	0.399363	Na	0.473093	I-BE	0.381448	Cr	0.416895	Cr	0.415249
I-TBil	0.384603	I-Ca	0.395278	Cr	0.461062	PTT	0.357339	Phos	0.355405	LAC	0.396136
I-Phos	0.381937	I-ALT	0.376004	I-pH	0.460122	Phos	0.352738	PT	0.328079	NE	0.338372
I-Ca	0.380717	I-AST	0.375944	Ca	0.422747	Na	0.345109	I-PTT	0.311758	PT	0.326491
LY	0.373098	LY	0.375163	PT	0.386761	I-PT	0.333936	NE	0.304336	LY	0.319146

PTT	0.36056	I-ALK	0.366346	PLT	0.371176	BG	0.320947	LAC	0.290642	MCV	0.314868
PLT	0.359362	RBC	0.36628	BE	0.365499	I-pH	0.317841	MCV	0.286645	PCO2	0.304013
Phos	0.358443	Phos	0.360009	I-PT	0.362492	LAC	0.307212	I-PT	0.280568	MCHC	0.297485
HGB	0.353079	BG	0.359947	BG	0.356136	PO2	0.295944	I-TBil	0.276302	RBC	0.280764
LAC	0.345373	PO2	0.333154	LAC	0.343797	MCV	0.293408	I-ALT	0.273609	I-AST	0.274608
I-PCO2	0.340782	PLT	0.332843	I-PTT	0.342374	HGB	0.289974	I-AST	0.269524	I-TBil	0.274146
I-PO2	0.340782	MCV	0.330676	MCV	0.341198	PCO2	0.287056	MCHC	0.257425	I-ALT	0.270957
I-BE	0.340621	PT	0.324238	RBC	0.30348	I-PTT	0.286421	BG	0.256086	I-ALK	0.262906
PCO2	0.336776	I-Phos	0.324092	I-TBil	0.300585	NE	0.283794	PLT	0.254084	I-PTT	0.260668
PT	0.328778	I-PCO2	0.31347	PCO2	0.300358	I-TBil	0.282065	I-ALK	0.25055	I-PT	0.25894
RBC	0.313951	I-PO2	0.31347	I-ALK	0.295056	LY	0.264846	WBC	0.240404	BG	0.254835
pH	0.312077	I-BE	0.313312	I-ALT	0.293048	TBil	0.25852	RBC	0.229375	PLT	0.253083
BE	0.301225	PTT	0.313247	I-AST	0.288134	RBC	0.257641	Cl	0.224759	WBC	0.239337
I-pH	0.293267	I-pH	0.312705	HGB	0.287403	MCH	0.256053	I-Ca	0.220419	Cl	0.230215
NE	0.291994	ALK	0.258992	I-Phos	0.278502	pH	0.254797	I-Phos	0.211873	PTT	0.201343
Cl	0.282029	Na	0.228743	PO2	0.247892	I-ALK	0.245638	PTT	0.202375	Mg	0.188997
MCH	0.259909	TBil	0.227722	WBC	0.245929	I-ALT	0.23875	Mg	0.190069	PO2	0.173802
WBC	0.225897	AST	0.226631	LY	0.230429	I-AST	0.237073	Ca	0.182171	BA	0.164709
I-NE	0.218566	WBC	0.221189	pH	0.21219	BA	0.229466	LY	0.163745	Ca	0.163818
I-MO	0.218566	I-NE	0.214159	MCH	0.204618	PLT	0.227216	PO2	0.160079	MO	0.161944
I-EO	0.218566	I-MO	0.214159	I-Ca	0.202655	WBC	0.193095	Na	0.150253	MCH	0.155433

I-BA	0.218566	I-EO	0.214159	HCT	0.195806	ALK	0.187838	TBil	0.139763	Na	0.149719
I-LY	0.218487	I-BA	0.214159	NE	0.194221	HCT	0.180302	MCH	0.131544	TBil	0.139005
Ca	0.209441	I-LY	0.214082	BA	0.189756	Ca	0.171374	K	0.117619	EO	0.136577
Na	0.195981	Ca	0.211277	Mg	0.159471	I-Phos	0.14342	I-NE	0.114392	K	0.117024
I-PT	0.185589	NE	0.204053	TBil	0.136629	MO	0.142681	I-LY	0.114392	I-Phos	0.115488
I-PTT	0.183623	BA	0.18897	I-NE	0.125957	Mg	0.134089	I-MO	0.114392	I-NE	0.113881
TBil	0.138903	EO	0.17213	I-LY	0.125957	AST	0.119857	I-EO	0.114392	I-LY	0.113881
AST	0.134067	MCH	0.170869	I-MO	0.125957	I-NE	0.112565	I-BA	0.114392	I-MO	0.113881
EO	0.127119	I-PT	0.170498	I-EO	0.125957	I-LY	0.112565	BA	0.110613	I-EO	0.113881
ALK	0.110338	I-PTT	0.160825	I-BA	0.125957	I-MO	0.112565	AST	0.108952	I-BA	0.113881
BA	0.108381	Cl	0.154443	K	0.096208	I-EO	0.112565	MO	0.083459	AST	0.112826
MO	0.099866	ALT	0.098121	AST	0.079471	I-BA	0.112565	EO	0.074052	I-Ca	0.088959
I-Mg	0.09522	I-BG	0.095042	MO	0.073539	I-Ca	0.097019	I-Mg	0.072413	ALK	0.066881
ALT	0.086944	I-AG	0.081084	EO	0.059083	EO	0.090394	ALK	0.067309	I-RDW	0.059171
Mg	0.05808	I-Mg	0.06207	ALK	0.050719	I-BG	0.077374	ALT	0.06125	I-MCHC	0.057317
K	0.055639	I-K	0.054604	I-MCHC	0.049679	I-AG	0.070263	I-WBC	0.047277	I-RBC	0.056806
I-HCT	0.038074	K	0.054598	I-RBC	0.049543	K	0.068477	I-Cr	0.046987	I-MCV	0.056806
I-AG	0.03646	Mg	0.041903	I-MCV	0.049543	I-K	0.064423	I-BUN	0.043822	I-MCH	0.056806
I-BG	0.034724	I-RBC	0.039129	I-MCH	0.049543	ALT	0.043276	I-PLT	0.040581	I-Cr	0.050183
I-BUN	0.028065	I-MCV	0.039069	I-HCT	0.047944	I-HCT	0.036985	I-HCT	0.03936	I-PLT	0.047133
I-K	0.027928	I-MCH	0.039069	I-HGB	0.047702	I-RDW	0.03124	I-RDW	0.039224	I-WBC	0.047106

I-Cr	0.027837	I-MCHC	0.038744	I-RDW	0.046801	I-PLT	0.029356	I-MCHC	0.037362	I-BUN	0.047027
HCT	0.024159	I-HCT	0.036842	ALT	0.0414	I-WBC	0.027678	I-RBC	0.036849	I-HCT	0.045927
I-PLT	0.022738	I-RDW	0.036652	I-BUN	0.040751	I-Cr	0.027393	I-MCV	0.036849	I-HGB	0.038626
I-WBC	0.019027	I-Cr	0.0327	I-PLT	0.03946	I-BUN	0.026789	I-MCH	0.036849	I-Cl	0.034929
I-RDW	0.018714	I-PLT	0.03036	I-Cr	0.039221	I-MCHC	0.023899	I-HGB	0.035416	I-HCO3	0.034152
I-HGB	0.016144	I-BUN	0.030341	I-WBC	0.034229	I-RBC	0.023775	I-Cl	0.031698	I-AG	0.030166
I-RBC	0.015815	I-HCO3	0.021465	I-Cl	0.02632	I-MCV	0.023775	I-HCO3	0.03092	I-K	0.027199
I-MCV	0.015754	I-Cl	0.021105	I-HCO3	0.025802	I-MCH	0.023775	HGB	0.027294	HGB	0.027109
I-MCH	0.015754	I-WBC	0.018852	I-Na	0.020485	I-HGB	0.022524	I-Na	0.023554	I-Na	0.026809
I-MCHC	0.015422	HCT	0.018361	I-Mg	0.019784	I-HCO3	0.022448	HCT	0.020428	I-Mg	0.024299
I-HCO3	0.01364	I-Na	0.016962	I-K	0.005176	I-Mg	0.020097	I-BG	0.007874	HCT	0.020301
I-Cl	0.013272	I-HGB	0.016188	I-AG	0.00514	I-Cl	0.018554	I-AG	0.003873	ALT	0.017294
I-Na	0.009183	MO	0.008405	I-BG	0.003625	I-Na	0.016975	I-K	0.003757	I-BG	0.010681

Table C4 Detailed AUROC values for all three days models for 30-days mortality.

		Day1		Day2		Day3	
		AUROC	AUROC-SD	AUROC	AUROC-SD	AUROC	AUROC-SD
Logistic Regression	Indicator Only	0.683639	0.012024	0.662923	0.010076	0.658562	0.016731
	HD	0.764109	0.007059	0.742167	0.015646	0.734406	0.009016
	HD + Indicator	0.785046	0.008317	0.76738	0.013904	0.761211	0.01254
	PMM	0.765781	0.00824	0.749135	0.013908	0.733725	0.012114
	PMM + Indicator	0.786277	0.010084	0.772205	0.012432	0.760894	0.013726
	SAPS II	0.781272	0.010389	0.734669	0.016996	0.704888	0.014114
	SAPS II +Indicator	0.804469	0.010902	0.758655	0.014826	0.738983	0.013721
Decision Tree	Indicator Only	0.650103	0.030717	0.626416	0.013378	0.629829	0.029985
	HD	0.710146	0.008507	0.686409	0.018551	0.664421	0.023127
	HD + Indicator	0.721343	0.009624	0.699798	0.017404	0.673743	0.010208
	PMM	0.707519	0.010437	0.683301	0.024034	0.660748	0.024769
	PMM + Indicator	0.714665	0.016703	0.695593	0.022077	0.676769	0.023449
	SAPS II	0.781272	0.010389	0.64003	0.020941	0.631917	0.014237
	SAPS II +Indicator	0.804469	0.010902	0.733759	0.019449	0.716765	0.015101
Random Forest	Indicator Only	0.505464	0.006078	0.514214	0.011948	0.530299	0.009883
	HD	0.773757	0.008081	0.7451	0.011768	0.72928	0.007846
	HD + Indicator	0.792436	0.008205	0.766697	0.011666	0.754279	0.010198
	PMM	0.778668	0.009603	0.757371	0.00969	0.741238	0.014159

	PMM + Indicator	0.790357	0.01027	0.77025	0.010431	0.751982	0.011704
	SAPS II	0.598363	0.007601	0.582176	0.013028	0.581908	0.010418
	SAPS II +Indicator	0.702174	0.009598	0.669676	0.016399	0.662755	0.014661

Table C5 Detailed AUROC values for all three days models for in-hospital mortality.

		Day1		Day2		Day3	
		AUROC	AUROC-SD	AUROC	AUROC-SD	AUROC	AUROC-SD
Logistic Regression	Indicator Only	0.714672	0.017386	0.691756	0.012501	0.692288	0.015078
	HD	0.764219	0.018237	0.741566	0.01794	0.732094	0.0221
	HD + Indicator	0.798866	0.01876	0.779967	0.01663	0.774248	0.013158
	PMM	0.764211	0.019857	0.747549	0.016608	0.73011	0.018976
	PMM + Indicator	0.798734	0.018334	0.783444	0.015014	0.772708	0.012686
	SAPS II	0.787666	0.014567	0.733107	0.011018	0.697324	0.017352
	SAPS II +Indicator	0.817011	0.014251	0.768525	0.013978	0.751957	0.013731
Decision Tree	Indicator Only	0.617608	0.084701	0.645272	0.010946	0.518569	0.010091
	HD	0.712456	0.016019	0.685595	0.025948	0.732607	0.018908
	HD + Indicator	0.733814	0.021929	0.718591	0.014642	0.768959	0.015584
	PMM	0.712825	0.014675	0.692577	0.013314	0.740328	0.019537
	PMM + Indicator	0.731189	0.017454	0.72262	0.023738	0.765453	0.014443
	SAPS II	0.661025	0.021642	0.64404	0.016882	0.545543	0.009066
	SAPS II +Indicator	0.785751	0.016167	0.740022	0.012199	0.662513	0.015397
Random Forest	Indicator Only	0.506492	0.002377	0.511513	0.004844	0.518569	0.010091
	HD	0.774549	0.01358	0.748636	0.021041	0.732607	0.018908
	HD + Indicator	0.802737	0.013752	0.778569	0.018496	0.768959	0.015584
	PMM	0.787579	0.015326	0.759913	0.019799	0.740328	0.019537
	PMM + Indicator	0.802973	0.01194	0.780929	0.016078	0.765453	0.014443
	SAPS II	0.579601	0.010443	0.556754	0.013962	0.545543	0.009066

	SAPS II +Indicator	0.689969	0.015505	0.658816	0.017864	0.662513	0.015397

Appendix D

Detailed Results of Chapter 4

Table D1 Detailed mortality prediction performance of PS-based KNN.

Number of Similar Patients	Multifaceted PS AUROC (Mean [95% CI])	Cosine PS AUROC (Mean [95% CI])	Euclidean PS AUROC (Mean [95% CI])
1	0.5522 [0.5442, 0.5442]	0.5499 [0.544, 0.5559]	0.5452 [0.5354, 0.5549]
10	0.6729 [0.6631, 0.6631]	0.6477 [0.6369, 0.6586]	0.6497 [0.6368, 0.6626]
20	0.7096 [0.6979, 0.6979]	0.6717 [0.6583, 0.685]	0.6706 [0.6569, 0.6844]
30	0.7269 [0.7138, 0.7138]	0.6815 [0.6691, 0.6938]	0.6796 [0.6667, 0.6924]
40	0.7355 [0.7221, 0.7221]	0.6884 [0.6758, 0.701]	0.6854 [0.674, 0.6969]
50	0.743 [0.7318, 0.7318]	0.6904 [0.6777, 0.7031]	0.6885 [0.6779, 0.699]
60	0.7488 [0.7374, 0.7374]	0.6925 [0.6796, 0.7055]	0.6881 [0.6784, 0.6978]
70	0.7545 [0.7446, 0.7446]	0.6934 [0.6814, 0.7054]	0.6879 [0.678, 0.6979]
80	0.7571 [0.7469, 0.7469]	0.6957 [0.6834, 0.7081]	0.6874 [0.6769, 0.6979]
90	0.7608 [0.7494, 0.7494]	0.6966 [0.6855, 0.7076]	0.6885 [0.6782, 0.6989]
100	0.7628 [0.7519, 0.7519]	0.6972 [0.6856, 0.7088]	0.6883 [0.6783, 0.6983]
110	0.7648 [0.7541, 0.7541]	0.6983 [0.687, 0.7096]	0.6874 [0.6769, 0.6979]
120	0.7651 [0.7542, 0.7542]	0.6983 [0.6864, 0.7101]	0.6881 [0.6775, 0.6987]
130	0.7675 [0.7567, 0.7567]	0.6989 [0.6874, 0.7103]	0.6885 [0.6773, 0.6997]
140	0.7691 [0.7587, 0.7587]	0.6982 [0.6868, 0.7097]	0.6884 [0.6774, 0.6993]
150	0.7714 [0.7608, 0.7608]	0.6986 [0.687, 0.7101]	0.6884 [0.6773, 0.6995]
160	0.7726 [0.7619, 0.7619]	0.6979 [0.6868, 0.7091]	0.6888 [0.6781, 0.6996]
170	0.7729 [0.7622, 0.7622]	0.6984 [0.6872, 0.7096]	0.689 [0.6789, 0.6991]
180	0.7745 [0.7643, 0.7643]	0.6977 [0.6867, 0.7088]	0.6893 [0.6788, 0.6997]
190	0.7758 [0.7656, 0.7656]	0.698 [0.6867, 0.7093]	0.6897 [0.6794, 0.6999]

200	0.7768 [0.7667, 0.7667]	0.6988 [0.6875, 0.7102]	0.6897 [0.6797, 0.6996]
210	0.7773 [0.7675, 0.7675]	0.698 [0.6866, 0.7093]	0.6891 [0.6794, 0.6988]
220	0.7788 [0.7689, 0.7689]	0.6983 [0.6867, 0.7098]	0.6893 [0.6794, 0.6991]
230	0.7799 [0.7704, 0.7704]	0.6976 [0.6859, 0.7094]	0.6893 [0.6796, 0.699]
240	0.78 [0.7705, 0.7705]	0.698 [0.6863, 0.7097]	0.6892 [0.6798, 0.6986]
250	0.7805 [0.7709, 0.7709]	0.6978 [0.686, 0.7096]	0.6886 [0.6792, 0.6981]
260	0.7812 [0.7717, 0.7717]	0.698 [0.6861, 0.7099]	0.688 [0.6785, 0.6974]
270	0.7821 [0.7727, 0.7727]	0.6978 [0.6859, 0.7098]	0.6883 [0.6786, 0.698]
280	0.7823 [0.7733, 0.7733]	0.6978 [0.6859, 0.7097]	0.6879 [0.6782, 0.6976]
290	0.7834 [0.7743, 0.7743]	0.6972 [0.6859, 0.7086]	0.6874 [0.6777, 0.6972]
300	0.7839 [0.7751, 0.7751]	0.6973 [0.6861, 0.7085]	0.687 [0.6771, 0.6969]
310	0.785 [0.7762, 0.7762]	0.697 [0.686, 0.7081]	0.6868 [0.6771, 0.6966]
320	0.7855 [0.7764, 0.7764]	0.6968 [0.6858, 0.7078]	0.6865 [0.6769, 0.6961]
330	0.7858 [0.7772, 0.7772]	0.6966 [0.6854, 0.7078]	0.686 [0.6763, 0.6957]
340	0.786 [0.7773, 0.7773]	0.6967 [0.6855, 0.7078]	0.6854 [0.6756, 0.6951]
350	0.7862 [0.7777, 0.7777]	0.6964 [0.6852, 0.7075]	0.6852 [0.6753, 0.6952]
360	0.7864 [0.7778, 0.7778]	0.6958 [0.6848, 0.7069]	0.6853 [0.6756, 0.6951]
370	0.7861 [0.7772, 0.7772]	0.6957 [0.6847, 0.7066]	0.6854 [0.676, 0.6948]
380	0.7861 [0.7772, 0.7772]	0.695 [0.6839, 0.7062]	0.6846 [0.6751, 0.6941]
390	0.7862 [0.7773, 0.7773]	0.695 [0.6836, 0.7064]	0.6845 [0.675, 0.694]
400	0.7865 [0.7774, 0.7774]	0.6945 [0.6833, 0.7058]	0.6841 [0.6747, 0.6935]
410	0.7869 [0.7779, 0.7779]	0.6947 [0.6837, 0.7057]	0.6835 [0.6739, 0.6931]
420	0.7865 [0.7778, 0.7778]	0.6947 [0.6838, 0.7057]	0.6832 [0.6737, 0.6927]
430	0.7868 [0.7785, 0.7785]	0.6946 [0.6835, 0.7057]	0.6828 [0.6732, 0.6925]
440	0.7873 [0.7795, 0.7795]	0.6946 [0.6835, 0.7057]	0.6824 [0.6728, 0.6921]
450	0.7871 [0.7792, 0.7792]	0.6943 [0.6832, 0.7055]	0.6822 [0.6726, 0.6918]
460	0.787 [0.7794, 0.7794]	0.6943 [0.6831, 0.7054]	0.6817 [0.6722, 0.6913]
470	0.7869 [0.7793, 0.7793]	0.6941 [0.683, 0.7051]	0.6816 [0.6722, 0.691]
480	0.7874 [0.7796, 0.7796]	0.6943 [0.6836, 0.7051]	0.6812 [0.6717, 0.6908]
490	0.787 [0.7796, 0.7796]	0.6938 [0.6831, 0.7046]	0.6811 [0.6717, 0.6904]

500	0.7868 [0.7795, 0.7795]	0.6935 [0.6829, 0.7041]	0.6808 [0.6715, 0.6901]
510	0.7866 [0.7795, 0.7795]	0.6934 [0.6829, 0.7039]	0.6806 [0.6713, 0.6899]
520	0.7865 [0.7794, 0.7794]	0.6932 [0.6827, 0.7038]	0.6805 [0.6711, 0.6899]
530	0.7866 [0.7796, 0.7796]	0.6931 [0.6827, 0.7035]	0.6803 [0.671, 0.6897]
540	0.7867 [0.7795, 0.7795]	0.693 [0.6823, 0.7036]	0.6801 [0.6707, 0.6895]
550	0.7869 [0.7796, 0.7796]	0.6931 [0.6823, 0.7039]	0.6799 [0.6705, 0.6894]
560	0.787 [0.7796, 0.7796]	0.6929 [0.682, 0.7038]	0.6797 [0.6702, 0.6892]
570	0.7868 [0.7795, 0.7795]	0.6926 [0.6819, 0.7034]	0.6792 [0.6697, 0.6886]
580	0.787 [0.7797, 0.7797]	0.6921 [0.6812, 0.703]	0.6789 [0.6695, 0.6884]
590	0.7869 [0.7796, 0.7796]	0.6919 [0.6808, 0.7029]	0.6786 [0.669, 0.6882]
600	0.7872 [0.7797, 0.7797]	0.6915 [0.6805, 0.7026]	0.6783 [0.6686, 0.6879]
610	0.7873 [0.7796, 0.7796]	0.6913 [0.6802, 0.7023]	0.6778 [0.6682, 0.6874]
620	0.7876 [0.7798, 0.7798]	0.6912 [0.68, 0.7023]	0.6776 [0.6681, 0.6871]
630	0.7877 [0.7798, 0.7798]	0.6909 [0.6798, 0.7019]	0.6771 [0.6679, 0.6864]
640	0.7879 [0.7802, 0.7802]	0.6906 [0.6796, 0.7017]	0.6766 [0.6674, 0.6858]
650	0.7884 [0.7806, 0.7806]	0.6906 [0.6796, 0.7016]	0.6764 [0.6672, 0.6856]
660	0.7882 [0.7804, 0.7804]	0.6904 [0.6793, 0.7016]	0.6762 [0.6669, 0.6855]
670	0.7877 [0.7799, 0.7799]	0.6901 [0.6789, 0.7012]	0.6761 [0.6668, 0.6853]
680	0.7878 [0.7801, 0.7801]	0.6899 [0.6788, 0.701]	0.6759 [0.6665, 0.6852]
690	0.7876 [0.7798, 0.7798]	0.6897 [0.6786, 0.7007]	0.6754 [0.666, 0.6848]
700	0.7878 [0.7798, 0.7798]	0.6894 [0.6783, 0.7004]	0.6752 [0.6656, 0.6847]
710	0.7881 [0.7797, 0.7797]	0.6893 [0.6783, 0.7003]	0.6753 [0.6657, 0.6848]
720	0.7882 [0.7797, 0.7797]	0.689 [0.678, 0.7]	0.6753 [0.6656, 0.685]
730	0.7879 [0.7795, 0.7795]	0.6888 [0.6777, 0.6999]	0.6752 [0.6655, 0.6849]
740	0.7881 [0.78, 0.78]	0.6886 [0.6776, 0.6995]	0.6751 [0.6653, 0.6849]
750	0.7883 [0.7802, 0.7802]	0.6884 [0.6774, 0.6993]	0.6749 [0.6651, 0.6846]
760	0.7883 [0.78, 0.78]	0.6881 [0.6774, 0.6988]	0.6749 [0.6652, 0.6846]
770	0.7885 [0.7803, 0.7803]	0.688 [0.6772, 0.6988]	0.6747 [0.665, 0.6845]
780	0.7887 [0.7805, 0.7805]	0.6878 [0.6772, 0.6984]	0.6745 [0.6647, 0.6843]
790	0.7887 [0.7804, 0.7804]	0.6875 [0.6769, 0.698]	0.6743 [0.6646, 0.6839]

800	0.7889 [0.7803, 0.7803]	0.6872 [0.6767, 0.6978]	0.6741 [0.6643, 0.6839]
810	0.789 [0.7803, 0.7803]	0.687 [0.6765, 0.6975]	0.6738 [0.664, 0.6836]
820	0.7891 [0.7804, 0.7804]	0.6869 [0.6763, 0.6974]	0.6739 [0.6641, 0.6837]
830	0.7889 [0.7801, 0.7801]	0.687 [0.6765, 0.6976]	0.6739 [0.664, 0.6838]
840	0.7891 [0.7801, 0.7801]	0.6867 [0.6762, 0.6973]	0.6735 [0.6636, 0.6835]
850	0.7891 [0.78, 0.78]	0.6865 [0.6759, 0.697]	0.6735 [0.6636, 0.6835]
860	0.7891 [0.7797, 0.7797]	0.6866 [0.6761, 0.6971]	0.6734 [0.6635, 0.6833]
870	0.7893 [0.7798, 0.7798]	0.6864 [0.676, 0.6969]	0.6733 [0.6632, 0.6834]
880	0.7893 [0.7797, 0.7797]	0.6865 [0.676, 0.6971]	0.6733 [0.6632, 0.6835]
890	0.7892 [0.7794, 0.7794]	0.6864 [0.6759, 0.697]	0.6734 [0.6634, 0.6833]
900	0.7893 [0.7794, 0.7794]	0.6861 [0.6756, 0.6967]	0.6731 [0.663, 0.6832]
910	0.7891 [0.7792, 0.7792]	0.6861 [0.6756, 0.6967]	0.6728 [0.6627, 0.6829]
920	0.7892 [0.7795, 0.7795]	0.6859 [0.6754, 0.6964]	0.6725 [0.6624, 0.6825]
930	0.7891 [0.7792, 0.7792]	0.6857 [0.6753, 0.6961]	0.6723 [0.6621, 0.6825]
940	0.789 [0.7791, 0.7791]	0.6855 [0.6752, 0.6959]	0.6724 [0.6622, 0.6825]
950	0.7891 [0.7794, 0.7794]	0.6854 [0.675, 0.6958]	0.6722 [0.6621, 0.6823]
960	0.7885 [0.7786, 0.7786]	0.6853 [0.6748, 0.6958]	0.672 [0.662, 0.682]
970	0.7886 [0.7787, 0.7787]	0.6852 [0.6747, 0.6957]	0.6718 [0.6618, 0.6818]
980	0.7884 [0.7784, 0.7784]	0.685 [0.6746, 0.6954]	0.6717 [0.6616, 0.6818]
990	0.7885 [0.7787, 0.7787]	0.6849 [0.6745, 0.6953]	0.6716 [0.6613, 0.6818]
1000	0.7885 [0.7789, 0.7789]	0.6847 [0.6744, 0.6951]	0.6714 [0.6611, 0.6818]
1010	0.7884 [0.7788, 0.7788]	0.6844 [0.674, 0.6947]	0.671 [0.6609, 0.6812]
1020	0.7885 [0.7789, 0.7789]	0.6843 [0.674, 0.6947]	0.6709 [0.6608, 0.681]
1030	0.7886 [0.779, 0.779]	0.6841 [0.6739, 0.6943]	0.6709 [0.6608, 0.681]
1040	0.7888 [0.7791, 0.7791]	0.6841 [0.674, 0.6943]	0.6709 [0.6608, 0.681]
1050	0.7886 [0.7789, 0.7789]	0.684 [0.6739, 0.6941]	0.6706 [0.6605, 0.6806]
1060	0.7887 [0.779, 0.779]	0.6839 [0.6738, 0.6939]	0.6702 [0.6602, 0.6803]
1070	0.7883 [0.7787, 0.7787]	0.6837 [0.6736, 0.6938]	0.67 [0.6599, 0.6801]
1080	0.7883 [0.7786, 0.7786]	0.6834 [0.6733, 0.6934]	0.6698 [0.6597, 0.6799]
1090	0.7884 [0.7787, 0.7787]	0.6831 [0.673, 0.6931]	0.6696 [0.6595, 0.6797]

1100	0.788 [0.7784, 0.7784]	0.6829 [0.6728, 0.693]	0.6698 [0.6597, 0.6799]
1110	0.7882 [0.7787, 0.7787]	0.6828 [0.6726, 0.693]	0.6695 [0.6595, 0.6795]
1120	0.7881 [0.7786, 0.7786]	0.6827 [0.6725, 0.693]	0.6692 [0.6592, 0.6793]
1130	0.7881 [0.7788, 0.7788]	0.6826 [0.6723, 0.6929]	0.6691 [0.659, 0.6791]
1140	0.7882 [0.7788, 0.7788]	0.6826 [0.6723, 0.693]	0.6687 [0.6585, 0.6789]
1150	0.788 [0.7786, 0.7786]	0.6825 [0.6721, 0.6929]	0.6685 [0.6583, 0.6788]
1160	0.7882 [0.7788, 0.7788]	0.6823 [0.6719, 0.6928]	0.6683 [0.6582, 0.6785]
1170	0.788 [0.7785, 0.7785]	0.6823 [0.6718, 0.6928]	0.6682 [0.6581, 0.6782]
1180	0.7878 [0.7783, 0.7783]	0.6822 [0.6717, 0.6927]	0.6682 [0.6582, 0.6783]
1190	0.7878 [0.7782, 0.7782]	0.682 [0.6715, 0.6926]	0.6682 [0.6582, 0.6783]
1200	0.7878 [0.7783, 0.7783]	0.6818 [0.6714, 0.6923]	0.6683 [0.6584, 0.6782]
1210	0.788 [0.7786, 0.7786]	0.6817 [0.6714, 0.6921]	0.6683 [0.6585, 0.678]
1220	0.7881 [0.7786, 0.7786]	0.6818 [0.6714, 0.6921]	0.6682 [0.6585, 0.6779]
1230	0.7881 [0.7786, 0.7786]	0.6818 [0.6715, 0.6921]	0.6679 [0.6582, 0.6776]
1240	0.7882 [0.7787, 0.7787]	0.6817 [0.6714, 0.6919]	0.6681 [0.6583, 0.6778]
1250	0.788 [0.7786, 0.7786]	0.6818 [0.6715, 0.692]	0.6678 [0.658, 0.6776]
1260	0.788 [0.7786, 0.7786]	0.6817 [0.6714, 0.6919]	0.6677 [0.658, 0.6774]
1270	0.788 [0.7785, 0.7785]	0.6814 [0.6712, 0.6916]	0.6677 [0.658, 0.6774]
1280	0.7879 [0.7785, 0.7785]	0.6812 [0.671, 0.6914]	0.6678 [0.658, 0.6776]
1290	0.7875 [0.7781, 0.7781]	0.6812 [0.671, 0.6914]	0.6677 [0.6579, 0.6774]
1300	0.7875 [0.7781, 0.7781]	0.681 [0.6708, 0.6912]	0.6675 [0.6577, 0.6772]
1310	0.7876 [0.7782, 0.7782]	0.681 [0.6708, 0.6912]	0.6674 [0.6575, 0.6773]
1320	0.7874 [0.7781, 0.7781]	0.6809 [0.6708, 0.6911]	0.6674 [0.6575, 0.6774]
1330	0.7871 [0.7777, 0.7777]	0.6808 [0.6707, 0.6909]	0.6671 [0.6572, 0.6771]
1340	0.7871 [0.7777, 0.7777]	0.6807 [0.6705, 0.6909]	0.6671 [0.6572, 0.677]
1350	0.7873 [0.778, 0.778]	0.6806 [0.6703, 0.6909]	0.6671 [0.6573, 0.6768]
1360	0.7871 [0.7777, 0.7777]	0.6805 [0.6702, 0.6909]	0.6672 [0.6576, 0.6768]
1370	0.787 [0.7775, 0.7775]	0.6804 [0.6702, 0.6906]	0.667 [0.6575, 0.6766]
1380	0.7867 [0.7771, 0.7771]	0.6803 [0.67, 0.6905]	0.667 [0.6574, 0.6767]
1390	0.7864 [0.7769, 0.7769]	0.6802 [0.6699, 0.6904]	0.667 [0.6574, 0.6767]

1400	0.7864 [0.7769, 0.7769]	0.6802 [0.67, 0.6904]	0.6671 [0.6575, 0.6768]
1410	0.7867 [0.7773, 0.7773]	0.6802 [0.67, 0.6904]	0.6668 [0.6571, 0.6765]
1420	0.7865 [0.7768, 0.7768]	0.6801 [0.67, 0.6903]	0.6669 [0.6572, 0.6765]
1430	0.7863 [0.7768, 0.7768]	0.6799 [0.6697, 0.6902]	0.6669 [0.6573, 0.6766]
1440	0.7862 [0.7766, 0.7766]	0.6799 [0.6696, 0.6901]	0.6669 [0.6573, 0.6766]
1450	0.7863 [0.7766, 0.7766]	0.6797 [0.6695, 0.6899]	0.6669 [0.6572, 0.6765]
1460	0.7862 [0.7767, 0.7767]	0.6797 [0.6695, 0.6898]	0.6666 [0.6569, 0.6763]
1470	0.7863 [0.7766, 0.7766]	0.6794 [0.6694, 0.6895]	0.6666 [0.6569, 0.6763]
1480	0.7862 [0.7765, 0.7765]	0.6794 [0.6694, 0.6895]	0.6666 [0.657, 0.6762]
1490	0.7861 [0.7764, 0.7764]	0.6792 [0.6692, 0.6892]	0.6664 [0.6568, 0.6761]
1500	0.7862 [0.7765, 0.7765]	0.6791 [0.6691, 0.6892]	0.6664 [0.6568, 0.6761]
1510	0.7861 [0.7766, 0.7766]	0.6791 [0.6691, 0.6891]	0.6662 [0.6566, 0.6758]
1520	0.786 [0.7763, 0.7763]	0.6789 [0.6689, 0.6889]	0.6663 [0.6566, 0.6759]
1530	0.7859 [0.7763, 0.7763]	0.6788 [0.6688, 0.6888]	0.6661 [0.6566, 0.6757]
1540	0.786 [0.7763, 0.7763]	0.6788 [0.6688, 0.6888]	0.6659 [0.6563, 0.6755]
1550	0.786 [0.7761, 0.7761]	0.6787 [0.6687, 0.6887]	0.6658 [0.6563, 0.6753]
1560	0.786 [0.7762, 0.7762]	0.6786 [0.6687, 0.6885]	0.6655 [0.6559, 0.6752]
1570	0.7861 [0.7762, 0.7762]	0.6786 [0.6686, 0.6885]	0.6655 [0.6557, 0.6753]
1580	0.7858 [0.7758, 0.7758]	0.6784 [0.6684, 0.6884]	0.6653 [0.6555, 0.6752]
1590	0.786 [0.776, 0.776]	0.6783 [0.6683, 0.6883]	0.6654 [0.6554, 0.6753]
1600	0.7859 [0.7758, 0.7758]	0.6782 [0.6683, 0.6882]	0.6655 [0.6556, 0.6754]
1610	0.786 [0.7759, 0.7759]	0.6782 [0.6682, 0.6881]	0.6654 [0.6555, 0.6752]
1620	0.7859 [0.7758, 0.7758]	0.6781 [0.6682, 0.6881]	0.6652 [0.6555, 0.675]
1630	0.7861 [0.776, 0.776]	0.6781 [0.6681, 0.688]	0.6651 [0.6553, 0.6749]
1640	0.786 [0.7761, 0.7761]	0.678 [0.6681, 0.6879]	0.6652 [0.6555, 0.6749]
1650	0.7863 [0.7763, 0.7763]	0.6779 [0.668, 0.6878]	0.665 [0.6551, 0.6748]
1660	0.7862 [0.7763, 0.7763]	0.6779 [0.6679, 0.6878]	0.6649 [0.655, 0.6748]
1670	0.786 [0.776, 0.776]	0.6777 [0.6678, 0.6876]	0.6649 [0.6549, 0.6749]
1680	0.7862 [0.7762, 0.7762]	0.6777 [0.6678, 0.6875]	0.665 [0.655, 0.6749]
1690	0.7861 [0.7762, 0.7762]	0.6776 [0.6678, 0.6875]	0.665 [0.6551, 0.675]

1700	0.7863 [0.7764, 0.7764]	0.6776 [0.6679, 0.6874]	0.6651 [0.6551, 0.675]
1710	0.7863 [0.7766, 0.7766]	0.6776 [0.6678, 0.6874]	0.6649 [0.655, 0.6749]
1720	0.7863 [0.7768, 0.7768]	0.6777 [0.6679, 0.6874]	0.665 [0.655, 0.675]
1730	0.7864 [0.7768, 0.7768]	0.6776 [0.6679, 0.6874]	0.6651 [0.6551, 0.675]
1740	0.7863 [0.7768, 0.7768]	0.6776 [0.6678, 0.6874]	0.6652 [0.6552, 0.6752]
1750	0.786 [0.7765, 0.7765]	0.6776 [0.6678, 0.6873]	0.6652 [0.6552, 0.6752]
1760	0.786 [0.7765, 0.7765]	0.6773 [0.6675, 0.6871]	0.6651 [0.6551, 0.6752]
1770	0.7859 [0.7765, 0.7765]	0.6772 [0.6673, 0.6871]	0.6652 [0.6552, 0.6753]
1780	0.7861 [0.7768, 0.7768]	0.6772 [0.6674, 0.6871]	0.6653 [0.6553, 0.6753]
1790	0.7862 [0.7768, 0.7768]	0.6772 [0.6674, 0.687]	0.6651 [0.6551, 0.6752]
1800	0.7861 [0.7768, 0.7768]	0.677 [0.6672, 0.6869]	0.665 [0.655, 0.675]
1810	0.7862 [0.777, 0.777]	0.6769 [0.6671, 0.6867]	0.6649 [0.6549, 0.6749]
1820	0.7861 [0.7767, 0.7767]	0.6768 [0.6669, 0.6866]	0.6648 [0.6548, 0.6747]
1830	0.7862 [0.7768, 0.7768]	0.6766 [0.6668, 0.6864]	0.6645 [0.6546, 0.6745]
1840	0.786 [0.7766, 0.7766]	0.6766 [0.6667, 0.6864]	0.6645 [0.6546, 0.6745]
1850	0.786 [0.7765, 0.7765]	0.6766 [0.6668, 0.6865]	0.6647 [0.6547, 0.6746]
1860	0.7859 [0.7764, 0.7764]	0.6766 [0.6668, 0.6864]	0.6645 [0.6546, 0.6745]
1870	0.7859 [0.7763, 0.7763]	0.6765 [0.6667, 0.6864]	0.6643 [0.6543, 0.6743]
1880	0.7858 [0.776, 0.776]	0.6763 [0.6664, 0.6862]	0.6642 [0.6542, 0.6742]
1890	0.7858 [0.7759, 0.7759]	0.6763 [0.6665, 0.6862]	0.6643 [0.6542, 0.6743]
1900	0.7858 [0.776, 0.776]	0.6761 [0.6662, 0.6861]	0.6642 [0.6542, 0.6741]
1910	0.7859 [0.776, 0.776]	0.676 [0.666, 0.686]	0.664 [0.654, 0.674]
1920	0.7859 [0.776, 0.776]	0.6759 [0.6658, 0.6859]	0.6639 [0.6538, 0.674]
1930	0.786 [0.7761, 0.7761]	0.6758 [0.6658, 0.6859]	0.6637 [0.6534, 0.6739]
1940	0.7858 [0.7762, 0.7762]	0.6758 [0.6658, 0.6859]	0.6636 [0.6534, 0.6738]
1950	0.7858 [0.7761, 0.7761]	0.6758 [0.6658, 0.6859]	0.6636 [0.6533, 0.6739]
1960	0.7859 [0.7761, 0.7761]	0.6758 [0.6659, 0.6857]	0.6637 [0.6535, 0.6738]
1970	0.7855 [0.7756, 0.7756]	0.6757 [0.6658, 0.6856]	0.6635 [0.6535, 0.6736]
1980	0.7858 [0.7759, 0.7759]	0.6756 [0.6657, 0.6855]	0.6635 [0.6534, 0.6736]
1990	0.7857 [0.7758, 0.7758]	0.6756 [0.6657, 0.6855]	0.6634 [0.6532, 0.6736]

2000	0.7858 [0.776, 0.776]	0.6755 [0.6656, 0.6855]	0.6633 [0.6532, 0.6734]
2010	0.7857 [0.7758, 0.7758]	0.6755 [0.6656, 0.6854]	0.6634 [0.6534, 0.6733]
2020	0.7856 [0.7759, 0.7759]	0.6754 [0.6654, 0.6854]	0.6633 [0.6534, 0.6732]
2030	0.7856 [0.7758, 0.7758]	0.6753 [0.6653, 0.6854]	0.6632 [0.6532, 0.6731]
2040	0.7858 [0.7762, 0.7762]	0.6752 [0.6652, 0.6853]	0.6629 [0.6529, 0.6729]
2050	0.7859 [0.7762, 0.7762]	0.6752 [0.6651, 0.6852]	0.663 [0.6531, 0.6729]
2060	0.7859 [0.7762, 0.7762]	0.6751 [0.665, 0.6851]	0.6629 [0.653, 0.6729]
2070	0.7859 [0.7762, 0.7762]	0.675 [0.665, 0.685]	0.6629 [0.653, 0.6729]
2080	0.7858 [0.776, 0.776]	0.675 [0.6649, 0.685]	0.6629 [0.653, 0.6728]
2090	0.7855 [0.7757, 0.7757]	0.6749 [0.6649, 0.685]	0.6629 [0.653, 0.6729]
2100	0.7856 [0.7758, 0.7758]	0.6748 [0.6648, 0.6848]	0.6629 [0.653, 0.6728]
2110	0.7857 [0.776, 0.776]	0.6747 [0.6647, 0.6848]	0.6628 [0.6528, 0.6728]
2120	0.7856 [0.7759, 0.7759]	0.6747 [0.6646, 0.6848]	0.6628 [0.6528, 0.6727]
2130	0.7854 [0.7758, 0.7758]	0.6746 [0.6645, 0.6847]	0.6627 [0.6527, 0.6727]
2140	0.7853 [0.7758, 0.7758]	0.6745 [0.6644, 0.6846]	0.6627 [0.6528, 0.6726]
2150	0.7852 [0.7757, 0.7757]	0.6744 [0.6644, 0.6845]	0.6626 [0.6526, 0.6725]
2160	0.7852 [0.7756, 0.7756]	0.6743 [0.6642, 0.6844]	0.6625 [0.6526, 0.6723]
2170	0.7852 [0.7757, 0.7757]	0.6743 [0.6642, 0.6843]	0.6623 [0.6524, 0.6722]
2180	0.7857 [0.7763, 0.7763]	0.6741 [0.6641, 0.6842]	0.6624 [0.6524, 0.6724]
2190	0.7856 [0.7761, 0.7761]	0.674 [0.664, 0.6839]	0.6621 [0.6522, 0.6721]
2200	0.7856 [0.7763, 0.7763]	0.674 [0.664, 0.684]	0.662 [0.652, 0.672]
2210	0.7855 [0.7762, 0.7762]	0.6738 [0.6638, 0.6838]	0.6621 [0.6521, 0.6722]
2220	0.7855 [0.7762, 0.7762]	0.6737 [0.6636, 0.6837]	0.6621 [0.6521, 0.6721]
2230	0.7855 [0.7762, 0.7762]	0.6737 [0.6636, 0.6837]	0.6622 [0.6522, 0.6722]
2240	0.7854 [0.7761, 0.7761]	0.6737 [0.6637, 0.6837]	0.6621 [0.6521, 0.672]
2250	0.7853 [0.7761, 0.7761]	0.6735 [0.6634, 0.6836]	0.6621 [0.6522, 0.6721]
2260	0.7851 [0.7759, 0.7759]	0.6735 [0.6635, 0.6836]	0.6621 [0.6522, 0.672]
2270	0.7852 [0.7762, 0.7762]	0.6736 [0.6635, 0.6836]	0.6621 [0.6522, 0.6721]
2280	0.7853 [0.7761, 0.7761]	0.6735 [0.6635, 0.6835]	0.6621 [0.6522, 0.672]
2290	0.7853 [0.7763, 0.7763]	0.6734 [0.6633, 0.6835]	0.662 [0.652, 0.6719]

2300	0.7853 [0.7762, 0.7762]	0.6733 [0.6631, 0.6835]	0.662 [0.652, 0.6719]
2310	0.7854 [0.7763, 0.7763]	0.6734 [0.6632, 0.6836]	0.6619 [0.652, 0.6718]
2320	0.7853 [0.7763, 0.7763]	0.6733 [0.6631, 0.6835]	0.6619 [0.652, 0.6718]
2330	0.7854 [0.7764, 0.7764]	0.6733 [0.6631, 0.6834]	0.662 [0.6522, 0.6718]
2340	0.7854 [0.7764, 0.7764]	0.6732 [0.663, 0.6833]	0.6619 [0.652, 0.6718]
2350	0.7855 [0.7765, 0.7765]	0.6731 [0.6629, 0.6833]	0.6618 [0.652, 0.6716]
2360	0.7854 [0.7764, 0.7764]	0.673 [0.6628, 0.6832]	0.6618 [0.652, 0.6716]
2370	0.785 [0.7759, 0.7759]	0.673 [0.6628, 0.6832]	0.6617 [0.6518, 0.6716]
2380	0.7853 [0.7762, 0.7762]	0.6728 [0.6626, 0.6829]	0.6616 [0.6517, 0.6715]
2390	0.7852 [0.7761, 0.7761]	0.6727 [0.6626, 0.6828]	0.6615 [0.6517, 0.6714]
2400	0.7851 [0.7759, 0.7759]	0.6727 [0.6626, 0.6829]	0.6614 [0.6515, 0.6713]
2410	0.7849 [0.7758, 0.7758]	0.6727 [0.6626, 0.6828]	0.6614 [0.6516, 0.6713]
2420	0.7851 [0.776, 0.776]	0.6726 [0.6625, 0.6827]	0.6614 [0.6516, 0.6712]
2430	0.7853 [0.7761, 0.7761]	0.6726 [0.6624, 0.6828]	0.6614 [0.6517, 0.6711]
2440	0.7851 [0.7759, 0.7759]	0.6726 [0.6625, 0.6827]	0.6613 [0.6517, 0.671]
2450	0.7851 [0.7759, 0.7759]	0.6726 [0.6626, 0.6826]	0.6615 [0.6517, 0.6712]
2460	0.7852 [0.7759, 0.7759]	0.6726 [0.6627, 0.6825]	0.6614 [0.6517, 0.6712]
2470	0.7851 [0.7758, 0.7758]	0.6726 [0.6627, 0.6825]	0.6612 [0.6514, 0.671]
2480	0.7851 [0.7758, 0.7758]	0.6725 [0.6626, 0.6824]	0.6611 [0.6512, 0.6709]
2490	0.7855 [0.7761, 0.7761]	0.6724 [0.6625, 0.6823]	0.661 [0.6512, 0.6708]
2500	0.7853 [0.7761, 0.7761]	0.6723 [0.6624, 0.6822]	0.6611 [0.6513, 0.6709]
2510	0.7853 [0.776, 0.776]	0.6721 [0.6623, 0.682]	0.6609 [0.6511, 0.6708]
2520	0.7854 [0.7761, 0.7761]	0.6721 [0.6622, 0.6819]	0.661 [0.6512, 0.6708]
2530	0.7854 [0.7762, 0.7762]	0.6722 [0.6624, 0.682]	0.6608 [0.6511, 0.6706]
2540	0.7852 [0.776, 0.776]	0.6722 [0.6624, 0.6821]	0.6607 [0.651, 0.6705]
2550	0.7852 [0.7761, 0.7761]	0.6721 [0.6622, 0.682]	0.6607 [0.6509, 0.6705]
2560	0.7852 [0.776, 0.776]	0.672 [0.6622, 0.6819]	0.6605 [0.6507, 0.6702]
2570	0.785 [0.7759, 0.7759]	0.672 [0.6623, 0.6817]	0.6604 [0.6506, 0.6702]
2580	0.785 [0.776, 0.776]	0.672 [0.6623, 0.6817]	0.6604 [0.6507, 0.6701]
2590	0.7848 [0.7759, 0.7759]	0.6719 [0.6621, 0.6816]	0.6603 [0.6505, 0.6701]

2600	0.7846 [0.7757, 0.7757]	0.6718 [0.6621, 0.6815]	0.6601 [0.6503, 0.6698]
2610	0.7846 [0.7758, 0.7758]	0.6718 [0.6621, 0.6815]	0.66 [0.6502, 0.6698]
2620	0.7847 [0.776, 0.776]	0.6719 [0.6622, 0.6815]	0.6599 [0.65, 0.6698]
2630	0.7847 [0.7758, 0.7758]	0.6718 [0.6622, 0.6813]	0.66 [0.6501, 0.6699]
2640	0.7846 [0.7758, 0.7758]	0.6717 [0.6621, 0.6812]	0.6599 [0.6501, 0.6697]
2650	0.7845 [0.7757, 0.7757]	0.6716 [0.6621, 0.6811]	0.6599 [0.6501, 0.6697]
2660	0.7845 [0.7756, 0.7756]	0.6715 [0.662, 0.681]	0.6598 [0.65, 0.6696]
2670	0.7844 [0.7755, 0.7755]	0.6713 [0.6618, 0.6807]	0.6598 [0.6501, 0.6696]
2680	0.7845 [0.7757, 0.7757]	0.6712 [0.6617, 0.6806]	0.6598 [0.6499, 0.6696]
2690	0.7843 [0.7754, 0.7754]	0.6711 [0.6617, 0.6805]	0.6598 [0.65, 0.6696]
2700	0.7844 [0.7755, 0.7755]	0.6712 [0.6619, 0.6806]	0.6598 [0.6499, 0.6696]
2710	0.7844 [0.7755, 0.7755]	0.6711 [0.6617, 0.6805]	0.6597 [0.6499, 0.6695]
2720	0.7844 [0.7756, 0.7756]	0.671 [0.6616, 0.6804]	0.6596 [0.6498, 0.6694]
2730	0.7842 [0.7752, 0.7752]	0.6709 [0.6615, 0.6803]	0.6594 [0.6497, 0.6692]
2740	0.7842 [0.7753, 0.7753]	0.6708 [0.6614, 0.6802]	0.6594 [0.6496, 0.6692]
2750	0.7841 [0.7752, 0.7752]	0.6708 [0.6615, 0.6802]	0.6594 [0.6496, 0.6693]
2760	0.7841 [0.7752, 0.7752]	0.6708 [0.6615, 0.6802]	0.6594 [0.6497, 0.6692]
2770	0.784 [0.7751, 0.7751]	0.6708 [0.6614, 0.6802]	0.6593 [0.6496, 0.6691]
2780	0.7839 [0.775, 0.775]	0.6707 [0.6613, 0.6802]	0.6592 [0.6495, 0.669]
2790	0.7839 [0.7751, 0.7751]	0.6707 [0.6612, 0.6802]	0.6592 [0.6495, 0.669]
2800	0.7837 [0.775, 0.775]	0.6707 [0.6611, 0.6803]	0.6593 [0.6495, 0.6691]
2810	0.7839 [0.7752, 0.7752]	0.6708 [0.6612, 0.6803]	0.6594 [0.6496, 0.6692]
2820	0.7838 [0.7751, 0.7751]	0.6706 [0.661, 0.6803]	0.6594 [0.6497, 0.6692]
2830	0.7841 [0.7753, 0.7753]	0.6705 [0.6609, 0.6802]	0.6594 [0.6497, 0.6691]
2840	0.784 [0.7752, 0.7752]	0.6705 [0.6609, 0.6801]	0.6594 [0.6496, 0.6691]
2850	0.7841 [0.7752, 0.7752]	0.6704 [0.6608, 0.68]	0.6594 [0.6496, 0.6692]
2860	0.7841 [0.7752, 0.7752]	0.6702 [0.6606, 0.6798]	0.6593 [0.6496, 0.6691]
2870	0.7842 [0.7752, 0.7752]	0.6702 [0.6607, 0.6797]	0.6593 [0.6494, 0.6691]
2880	0.7841 [0.7752, 0.7752]	0.6701 [0.6605, 0.6797]	0.6592 [0.6495, 0.669]
2890	0.7841 [0.7752, 0.7752]	0.67 [0.6604, 0.6796]	0.6592 [0.6494, 0.669]

2900	0.7841 [0.7751, 0.7751]	0.67 [0.6605, 0.6796]	0.6592 [0.6493, 0.669]
2910	0.784 [0.7751, 0.7751]	0.67 [0.6603, 0.6796]	0.6592 [0.6494, 0.6691]
2920	0.7838 [0.7749, 0.7749]	0.6699 [0.6602, 0.6796]	0.6591 [0.6492, 0.669]
2930	0.7837 [0.775, 0.775]	0.6697 [0.66, 0.6794]	0.6591 [0.6492, 0.669]
2940	0.7837 [0.7749, 0.7749]	0.6696 [0.6599, 0.6792]	0.659 [0.6492, 0.6689]
2950	0.7837 [0.7751, 0.7751]	0.6695 [0.6598, 0.6791]	0.659 [0.6493, 0.6688]
2960	0.784 [0.7755, 0.7755]	0.6695 [0.6598, 0.6792]	0.659 [0.6493, 0.6688]
2970	0.7838 [0.7752, 0.7752]	0.6695 [0.6598, 0.6792]	0.6591 [0.6493, 0.6688]
2980	0.7839 [0.7752, 0.7752]	0.6695 [0.6598, 0.6792]	0.6591 [0.6494, 0.6688]
2990	0.7837 [0.775, 0.775]	0.6694 [0.6598, 0.6791]	0.6589 [0.6492, 0.6685]
3000	0.7837 [0.7749, 0.7749]	0.6694 [0.6597, 0.6791]	0.659 [0.6493, 0.6687]
3010	0.7837 [0.7748, 0.7748]	0.6693 [0.6596, 0.679]	0.659 [0.6493, 0.6688]
3020	0.7836 [0.7749, 0.7749]	0.6692 [0.6594, 0.679]	0.6591 [0.6493, 0.6688]
3030	0.7836 [0.7749, 0.7749]	0.6692 [0.6594, 0.679]	0.659 [0.6493, 0.6687]
3040	0.7835 [0.7749, 0.7749]	0.6692 [0.6594, 0.679]	0.659 [0.6493, 0.6687]
3050	0.7833 [0.7748, 0.7748]	0.6691 [0.6594, 0.6789]	0.6589 [0.6491, 0.6687]
3060	0.7833 [0.7747, 0.7747]	0.669 [0.6592, 0.6789]	0.6589 [0.6491, 0.6686]
3070	0.7833 [0.7747, 0.7747]	0.669 [0.6593, 0.6788]	0.6589 [0.6492, 0.6686]
3080	0.7833 [0.7747, 0.7747]	0.669 [0.6592, 0.6788]	0.6588 [0.6491, 0.6686]
3090	0.7832 [0.7747, 0.7747]	0.6689 [0.6591, 0.6787]	0.6588 [0.6492, 0.6685]
3100	0.7832 [0.7745, 0.7745]	0.6688 [0.659, 0.6786]	0.6588 [0.6491, 0.6685]
3110	0.7832 [0.7745, 0.7745]	0.6687 [0.6588, 0.6786]	0.6588 [0.649, 0.6686]
3120	0.7831 [0.7745, 0.7745]	0.6688 [0.6588, 0.6787]	0.6589 [0.6493, 0.6686]
3130	0.7831 [0.7745, 0.7745]	0.6687 [0.6589, 0.6786]	0.6588 [0.6492, 0.6684]
3140	0.7829 [0.7744, 0.7744]	0.6686 [0.6587, 0.6785]	0.6587 [0.6491, 0.6683]
3150	0.7831 [0.7746, 0.7746]	0.6687 [0.6588, 0.6785]	0.6587 [0.6492, 0.6683]
3160	0.783 [0.7745, 0.7745]	0.6687 [0.6589, 0.6784]	0.6586 [0.6491, 0.6682]
3170	0.783 [0.7745, 0.7745]	0.6687 [0.659, 0.6784]	0.6586 [0.6491, 0.6681]
3180	0.7831 [0.7746, 0.7746]	0.6687 [0.659, 0.6785]	0.6585 [0.6491, 0.668]
3190	0.7831 [0.7746, 0.7746]	0.6688 [0.6591, 0.6785]	0.6585 [0.649, 0.6679]

3200	0.7832 [0.7747, 0.7747]	0.6688 [0.6591, 0.6785]	0.6585 [0.6491, 0.668]
3210	0.7833 [0.7748, 0.7748]	0.6688 [0.6591, 0.6785]	0.6583 [0.6489, 0.6677]
3220	0.7831 [0.7747, 0.7747]	0.6687 [0.659, 0.6784]	0.6582 [0.6487, 0.6676]
3230	0.7831 [0.7745, 0.7745]	0.6687 [0.659, 0.6785]	0.6582 [0.6487, 0.6677]
3240	0.783 [0.7744, 0.7744]	0.6688 [0.659, 0.6786]	0.6581 [0.6487, 0.6675]
3250	0.783 [0.7745, 0.7745]	0.6689 [0.659, 0.6787]	0.658 [0.6487, 0.6673]
3260	0.7828 [0.7744, 0.7744]	0.6688 [0.659, 0.6786]	0.6579 [0.6485, 0.6672]
3270	0.7827 [0.7742, 0.7742]	0.6688 [0.659, 0.6786]	0.6578 [0.6485, 0.6672]
3280	0.7826 [0.774, 0.774]	0.6688 [0.659, 0.6786]	0.6578 [0.6485, 0.6672]
3290	0.7828 [0.7742, 0.7742]	0.6687 [0.6589, 0.6785]	0.6578 [0.6485, 0.6672]
3300	0.7827 [0.774, 0.774]	0.6688 [0.6589, 0.6786]	0.6577 [0.6485, 0.667]
3310	0.7825 [0.7738, 0.7738]	0.6689 [0.659, 0.6787]	0.6576 [0.6483, 0.6669]
3320	0.7824 [0.7736, 0.7736]	0.6689 [0.6591, 0.6788]	0.6575 [0.6482, 0.6668]
3330	0.7823 [0.7734, 0.7734]	0.669 [0.6592, 0.6788]	0.6575 [0.6482, 0.6669]
3340	0.7822 [0.7733, 0.7733]	0.669 [0.6592, 0.6788]	0.6574 [0.6481, 0.6668]
3350	0.7823 [0.7735, 0.7735]	0.6689 [0.659, 0.6788]	0.6574 [0.6481, 0.6667]
3360	0.7823 [0.7734, 0.7734]	0.6688 [0.6587, 0.6788]	0.6574 [0.6481, 0.6666]
3370	0.7823 [0.7732, 0.7732]	0.6688 [0.6587, 0.6789]	0.6573 [0.648, 0.6666]
3380	0.7822 [0.7732, 0.7732]	0.6688 [0.6587, 0.6788]	0.6573 [0.6481, 0.6665]
3390	0.782 [0.7729, 0.7729]	0.6688 [0.6588, 0.6788]	0.6574 [0.6481, 0.6666]
3400	0.782 [0.7729, 0.7729]	0.6687 [0.6587, 0.6787]	0.6573 [0.648, 0.6665]
3410	0.7821 [0.773, 0.773]	0.6687 [0.6586, 0.6788]	0.6572 [0.648, 0.6664]
3420	0.7822 [0.7731, 0.7731]	0.6686 [0.6586, 0.6786]	0.6573 [0.648, 0.6665]
3430	0.7822 [0.773, 0.773]	0.6685 [0.6585, 0.6785]	0.6572 [0.6479, 0.6665]
3440	0.7823 [0.7731, 0.7731]	0.6685 [0.6586, 0.6785]	0.6573 [0.648, 0.6667]
3450	0.7824 [0.7732, 0.7732]	0.6685 [0.6586, 0.6785]	0.6572 [0.6479, 0.6665]
3460	0.7823 [0.7731, 0.7731]	0.6685 [0.6586, 0.6785]	0.6572 [0.6479, 0.6665]
3470	0.7823 [0.7731, 0.7731]	0.6684 [0.6585, 0.6784]	0.6572 [0.6479, 0.6665]
3480	0.7822 [0.773, 0.773]	0.6683 [0.6584, 0.6782]	0.6573 [0.648, 0.6666]
3490	0.7822 [0.7731, 0.7731]	0.6683 [0.6584, 0.6783]	0.6574 [0.6482, 0.6666]

3500	0.7822 [0.7729, 0.7729]	0.6684 [0.6586, 0.6783]	0.6572 [0.648, 0.6664]
3510	0.782 [0.7728, 0.7728]	0.6684 [0.6586, 0.6783]	0.6571 [0.6478, 0.6664]
3520	0.7819 [0.7728, 0.7728]	0.6683 [0.6584, 0.6782]	0.6572 [0.648, 0.6665]
3530	0.782 [0.7729, 0.7729]	0.6682 [0.6583, 0.6781]	0.6571 [0.6479, 0.6664]
3540	0.782 [0.7728, 0.7728]	0.6683 [0.6584, 0.6782]	0.6573 [0.648, 0.6665]
3550	0.7821 [0.7727, 0.7727]	0.6683 [0.6583, 0.6782]	0.6572 [0.648, 0.6665]
3560	0.7821 [0.7727, 0.7727]	0.6681 [0.6582, 0.678]	0.6573 [0.648, 0.6666]
3570	0.7819 [0.7726, 0.7726]	0.6682 [0.6583, 0.6781]	0.6571 [0.6479, 0.6664]
3580	0.7818 [0.7725, 0.7725]	0.6681 [0.6583, 0.678]	0.6571 [0.6479, 0.6664]
3590	0.7817 [0.7724, 0.7724]	0.6681 [0.6582, 0.6781]	0.6571 [0.6479, 0.6664]
3600	0.7816 [0.7722, 0.7722]	0.6682 [0.6582, 0.6781]	0.6571 [0.6479, 0.6663]
3610	0.7817 [0.7723, 0.7723]	0.6681 [0.6581, 0.6781]	0.6571 [0.6479, 0.6664]
3620	0.7816 [0.7723, 0.7723]	0.6681 [0.6581, 0.6781]	0.6572 [0.6479, 0.6665]
3630	0.7816 [0.7723, 0.7723]	0.6681 [0.6581, 0.6781]	0.6571 [0.6478, 0.6664]
3640	0.7818 [0.7724, 0.7724]	0.6681 [0.6581, 0.6781]	0.6571 [0.6478, 0.6664]
3650	0.7817 [0.7724, 0.7724]	0.668 [0.658, 0.678]	0.6572 [0.6478, 0.6666]
3660	0.7817 [0.7723, 0.7723]	0.668 [0.658, 0.6779]	0.6574 [0.648, 0.6667]
3670	0.7818 [0.7724, 0.7724]	0.6679 [0.658, 0.6779]	0.6574 [0.6481, 0.6667]
3680	0.7817 [0.7723, 0.7723]	0.6679 [0.658, 0.6778]	0.6574 [0.648, 0.6668]
3690	0.7817 [0.7721, 0.7721]	0.668 [0.6581, 0.6778]	0.6574 [0.6481, 0.6668]
3700	0.7817 [0.7721, 0.7721]	0.6679 [0.6581, 0.6778]	0.6575 [0.6481, 0.6668]
3710	0.7815 [0.772, 0.772]	0.668 [0.6581, 0.6778]	0.6575 [0.6482, 0.6669]
3720	0.7816 [0.772, 0.772]	0.668 [0.6581, 0.6779]	0.6576 [0.6482, 0.6669]
3730	0.7815 [0.7719, 0.7719]	0.668 [0.6581, 0.6779]	0.6576 [0.6483, 0.667]
3740	0.7814 [0.7718, 0.7718]	0.668 [0.6581, 0.6778]	0.6577 [0.6483, 0.667]
3750	0.7814 [0.7718, 0.7718]	0.6679 [0.658, 0.6778]	0.6577 [0.6484, 0.667]
3760	0.7813 [0.7717, 0.7717]	0.6679 [0.658, 0.6778]	0.6577 [0.6483, 0.6672]
3770	0.7813 [0.7716, 0.7716]	0.6679 [0.658, 0.6778]	0.6577 [0.6483, 0.6671]
3780	0.7814 [0.7718, 0.7718]	0.6679 [0.6581, 0.6777]	0.6578 [0.6484, 0.6672]
3790	0.7814 [0.7717, 0.7717]	0.6678 [0.6579, 0.6776]	0.6578 [0.6484, 0.6672]

3800	0.7813 [0.7716, 0.7716]	0.6678 [0.6581, 0.6776]	0.6578 [0.6484, 0.6673]
3810	0.7814 [0.7716, 0.7716]	0.6678 [0.658, 0.6776]	0.6579 [0.6484, 0.6674]
3820	0.7814 [0.7716, 0.7716]	0.6677 [0.6581, 0.6774]	0.658 [0.6485, 0.6674]
3830	0.7815 [0.7718, 0.7718]	0.6677 [0.658, 0.6773]	0.658 [0.6486, 0.6674]
3840	0.7816 [0.7719, 0.7719]	0.6676 [0.658, 0.6773]	0.6579 [0.6485, 0.6672]
3850	0.7816 [0.7719, 0.7719]	0.6675 [0.6578, 0.6773]	0.6579 [0.6485, 0.6673]
3860	0.7816 [0.7718, 0.7718]	0.6674 [0.6577, 0.6772]	0.658 [0.6486, 0.6673]
3870	0.7817 [0.772, 0.772]	0.6675 [0.6577, 0.6772]	0.6579 [0.6486, 0.6673]
3880	0.7817 [0.772, 0.772]	0.6674 [0.6575, 0.6772]	0.6579 [0.6484, 0.6673]
3890	0.7818 [0.772, 0.772]	0.6674 [0.6576, 0.6772]	0.6577 [0.6483, 0.6672]
3900	0.7817 [0.772, 0.772]	0.6674 [0.6576, 0.6771]	0.6577 [0.6481, 0.6672]
3910	0.7816 [0.7718, 0.7718]	0.6673 [0.6574, 0.6771]	0.6577 [0.6482, 0.6672]
3920	0.7817 [0.7719, 0.7719]	0.6672 [0.6574, 0.677]	0.6576 [0.6481, 0.6671]
3930	0.7816 [0.7719, 0.7719]	0.6671 [0.6573, 0.6769]	0.6577 [0.6482, 0.6671]
3940	0.7817 [0.772, 0.772]	0.6672 [0.6574, 0.677]	0.6576 [0.6481, 0.6671]
3950	0.7816 [0.772, 0.772]	0.6671 [0.6573, 0.6769]	0.6575 [0.648, 0.6671]
3960	0.7816 [0.7719, 0.7719]	0.6671 [0.6573, 0.677]	0.6576 [0.648, 0.6672]
3970	0.7816 [0.7719, 0.7719]	0.6671 [0.6572, 0.677]	0.6576 [0.648, 0.6672]
3980	0.7816 [0.7719, 0.7719]	0.6671 [0.6571, 0.677]	0.6576 [0.648, 0.6672]
3990	0.7816 [0.7719, 0.7719]	0.667 [0.657, 0.677]	0.6575 [0.6479, 0.667]
4000	0.7816 [0.7719, 0.7719]	0.667 [0.657, 0.677]	0.6575 [0.6479, 0.667]
4010	0.7815 [0.7717, 0.7717]	0.667 [0.6571, 0.6769]	0.6574 [0.6478, 0.6669]
4020	0.7815 [0.7718, 0.7718]	0.667 [0.657, 0.677]	0.6574 [0.6478, 0.6669]
4030	0.7814 [0.7717, 0.7717]	0.667 [0.657, 0.677]	0.6574 [0.648, 0.6668]
4040	0.7814 [0.7717, 0.7717]	0.667 [0.657, 0.677]	0.6574 [0.6479, 0.6668]
4050	0.7812 [0.7715, 0.7715]	0.667 [0.657, 0.677]	0.6572 [0.6478, 0.6667]
4060	0.7813 [0.7715, 0.7715]	0.6669 [0.6568, 0.6769]	0.6572 [0.6477, 0.6667]
4070	0.7811 [0.7714, 0.7714]	0.6668 [0.6568, 0.6768]	0.6572 [0.6477, 0.6667]
4080	0.7811 [0.7714, 0.7714]	0.6667 [0.6567, 0.6767]	0.6571 [0.6475, 0.6666]
4090	0.7811 [0.7714, 0.7714]	0.6665 [0.6566, 0.6765]	0.657 [0.6476, 0.6665]

4100	0.7812 [0.7715, 0.7715]	0.6664 [0.6565, 0.6764]	0.657 [0.6476, 0.6665]
4110	0.7811 [0.7715, 0.7715]	0.6664 [0.6564, 0.6764]	0.6571 [0.6477, 0.6666]
4120	0.7812 [0.7715, 0.7715]	0.6664 [0.6565, 0.6764]	0.657 [0.6475, 0.6665]
4130	0.7813 [0.7716, 0.7716]	0.6664 [0.6564, 0.6764]	0.657 [0.6475, 0.6665]
4140	0.7814 [0.7717, 0.7717]	0.6664 [0.6564, 0.6763]	0.657 [0.6475, 0.6665]
4150	0.7813 [0.7716, 0.7716]	0.6663 [0.6562, 0.6763]	0.6571 [0.6476, 0.6666]
4160	0.7813 [0.7716, 0.7716]	0.6662 [0.6562, 0.6763]	0.6572 [0.6477, 0.6666]
4170	0.7812 [0.7715, 0.7715]	0.6662 [0.6561, 0.6762]	0.6571 [0.6476, 0.6666]
4180	0.7812 [0.7715, 0.7715]	0.6662 [0.6562, 0.6762]	0.6572 [0.6476, 0.6667]
4190	0.7811 [0.7714, 0.7714]	0.666 [0.6561, 0.676]	0.6571 [0.6476, 0.6667]
4200	0.7812 [0.7715, 0.7715]	0.666 [0.656, 0.6759]	0.6571 [0.6475, 0.6667]
4210	0.7812 [0.7715, 0.7715]	0.6659 [0.6559, 0.6758]	0.6571 [0.6474, 0.6667]
4220	0.7813 [0.7716, 0.7716]	0.6659 [0.6559, 0.6758]	0.6571 [0.6474, 0.6667]
4230	0.7813 [0.7717, 0.7717]	0.6658 [0.6559, 0.6758]	0.657 [0.6473, 0.6667]
4240	0.7812 [0.7716, 0.7716]	0.6658 [0.6559, 0.6758]	0.657 [0.6473, 0.6667]
4250	0.7813 [0.7717, 0.7717]	0.6658 [0.6559, 0.6757]	0.6569 [0.6472, 0.6665]
4260	0.7812 [0.7716, 0.7716]	0.6656 [0.6557, 0.6756]	0.6568 [0.6472, 0.6665]
4270	0.7814 [0.7717, 0.7717]	0.6656 [0.6557, 0.6756]	0.6568 [0.6472, 0.6664]
4280	0.7814 [0.7717, 0.7717]	0.6656 [0.6557, 0.6756]	0.6568 [0.6472, 0.6665]
4290	0.7813 [0.7716, 0.7716]	0.6656 [0.6557, 0.6755]	0.6569 [0.6473, 0.6665]
4300	0.7812 [0.7715, 0.7715]	0.6656 [0.6557, 0.6755]	0.657 [0.6474, 0.6666]
4310	0.7813 [0.7717, 0.7717]	0.6656 [0.6557, 0.6755]	0.657 [0.6474, 0.6666]
4320	0.7814 [0.7717, 0.7717]	0.6655 [0.6556, 0.6755]	0.6571 [0.6475, 0.6668]
4330	0.7814 [0.7718, 0.7718]	0.6654 [0.6555, 0.6753]	0.6569 [0.6472, 0.6665]
4340	0.7814 [0.7717, 0.7717]	0.6654 [0.6555, 0.6753]	0.6569 [0.6473, 0.6666]
4350	0.7812 [0.7715, 0.7715]	0.6655 [0.6556, 0.6754]	0.6569 [0.6472, 0.6666]
4360	0.7813 [0.7717, 0.7717]	0.6655 [0.6556, 0.6754]	0.6569 [0.6472, 0.6665]
4370	0.7814 [0.7717, 0.7717]	0.6655 [0.6556, 0.6754]	0.6568 [0.6471, 0.6665]
4380	0.7815 [0.7718, 0.7718]	0.6655 [0.6556, 0.6754]	0.6568 [0.6471, 0.6665]
4390	0.7814 [0.7716, 0.7716]	0.6655 [0.6556, 0.6754]	0.6568 [0.6471, 0.6664]

4400	0.7814 [0.7716, 0.7716]	0.6654 [0.6554, 0.6753]	0.6567 [0.6471, 0.6664]
4410	0.7814 [0.7716, 0.7716]	0.6654 [0.6555, 0.6753]	0.6569 [0.6472, 0.6665]
4420	0.7813 [0.7714, 0.7714]	0.6652 [0.6553, 0.6752]	0.6569 [0.6473, 0.6666]
4430	0.7812 [0.7714, 0.7714]	0.6653 [0.6553, 0.6752]	0.6569 [0.6472, 0.6665]
4440	0.7813 [0.7716, 0.7716]	0.6651 [0.6551, 0.6751]	0.6568 [0.6471, 0.6665]
4450	0.7814 [0.7716, 0.7716]	0.6651 [0.655, 0.6751]	0.6567 [0.647, 0.6664]
4460	0.7814 [0.7716, 0.7716]	0.665 [0.655, 0.675]	0.6567 [0.647, 0.6664]
4470	0.7813 [0.7716, 0.7716]	0.665 [0.6549, 0.6751]	0.6566 [0.6471, 0.6662]
4480	0.7813 [0.7715, 0.7715]	0.665 [0.6549, 0.6751]	0.6565 [0.647, 0.6661]
4490	0.7814 [0.7716, 0.7716]	0.6649 [0.6549, 0.675]	0.6564 [0.6467, 0.6661]
4500	0.7813 [0.7715, 0.7715]	0.6649 [0.6548, 0.6749]	0.6565 [0.6467, 0.6662]
4510	0.7814 [0.7715, 0.7715]	0.6648 [0.6548, 0.6748]	0.6564 [0.6468, 0.6661]
4520	0.7814 [0.7714, 0.7714]	0.6647 [0.6546, 0.6747]	0.6564 [0.6466, 0.6662]
4530	0.7814 [0.7714, 0.7714]	0.6646 [0.6546, 0.6747]	0.6564 [0.6465, 0.6663]
4540	0.7812 [0.7712, 0.7712]	0.6645 [0.6544, 0.6746]	0.6564 [0.6465, 0.6663]
4550	0.7813 [0.7713, 0.7713]	0.6645 [0.6545, 0.6746]	0.6563 [0.6464, 0.6662]
4560	0.7814 [0.7713, 0.7713]	0.6645 [0.6545, 0.6746]	0.6562 [0.6463, 0.6661]
4570	0.7813 [0.7712, 0.7712]	0.6646 [0.6546, 0.6746]	0.6562 [0.6463, 0.6662]
4580	0.7812 [0.7711, 0.7711]	0.6646 [0.6546, 0.6747]	0.6564 [0.6465, 0.6663]
4590	0.7812 [0.7711, 0.7711]	0.6646 [0.6546, 0.6746]	0.6564 [0.6465, 0.6663]
4600	0.7812 [0.7711, 0.7711]	0.6647 [0.6547, 0.6746]	0.6563 [0.6464, 0.6662]
4610	0.7811 [0.7709, 0.7709]	0.6646 [0.6546, 0.6746]	0.6562 [0.6464, 0.666]
4620	0.781 [0.7708, 0.7708]	0.6646 [0.6546, 0.6747]	0.6562 [0.6464, 0.666]
4630	0.7809 [0.7707, 0.7707]	0.6645 [0.6545, 0.6746]	0.6562 [0.6464, 0.666]
4640	0.7809 [0.7706, 0.7706]	0.6645 [0.6545, 0.6746]	0.6562 [0.6465, 0.666]
4650	0.7808 [0.7706, 0.7706]	0.6644 [0.6544, 0.6744]	0.6561 [0.6464, 0.6659]
4660	0.7809 [0.7706, 0.7706]	0.6644 [0.6544, 0.6744]	0.6562 [0.6464, 0.666]
4670	0.7809 [0.7705, 0.7705]	0.6643 [0.6544, 0.6743]	0.6561 [0.6464, 0.6658]
4680	0.7808 [0.7704, 0.7704]	0.6644 [0.6545, 0.6743]	0.6561 [0.6464, 0.6658]
4690	0.7809 [0.7705, 0.7705]	0.6644 [0.6545, 0.6743]	0.656 [0.6464, 0.6657]

4700	0.7809 [0.7707, 0.7707]	0.6644 [0.6545, 0.6743]	0.6562 [0.6465, 0.6658]
4710	0.781 [0.7708, 0.7708]	0.6645 [0.6546, 0.6744]	0.6562 [0.6466, 0.6658]
4720	0.7809 [0.7707, 0.7707]	0.6645 [0.6545, 0.6744]	0.6562 [0.6465, 0.6658]
4730	0.7808 [0.7706, 0.7706]	0.6644 [0.6545, 0.6743]	0.6561 [0.6466, 0.6656]
4740	0.7808 [0.7705, 0.7705]	0.6644 [0.6545, 0.6743]	0.6561 [0.6466, 0.6657]
4750	0.7807 [0.7703, 0.7703]	0.6643 [0.6545, 0.6742]	0.656 [0.6464, 0.6656]
4760	0.7804 [0.77, 0.77]	0.6643 [0.6544, 0.6742]	0.6559 [0.6463, 0.6655]
4770	0.7804 [0.7699, 0.7699]	0.6642 [0.6543, 0.6741]	0.6559 [0.6462, 0.6656]
4780	0.7804 [0.7699, 0.7699]	0.6642 [0.6543, 0.6741]	0.656 [0.6462, 0.6657]
4790	0.7802 [0.7697, 0.7697]	0.6641 [0.6542, 0.674]	0.656 [0.6462, 0.6657]
4800	0.7802 [0.7698, 0.7698]	0.6642 [0.6542, 0.6741]	0.656 [0.6462, 0.6657]
4810	0.7804 [0.77, 0.77]	0.6641 [0.6541, 0.674]	0.656 [0.6462, 0.6657]
4820	0.7803 [0.7699, 0.7699]	0.6641 [0.6542, 0.674]	0.6561 [0.6463, 0.6658]
4830	0.7803 [0.7699, 0.7699]	0.664 [0.6541, 0.674]	0.6562 [0.6464, 0.6659]
4840	0.7802 [0.7698, 0.7698]	0.664 [0.6541, 0.6739]	0.6562 [0.6464, 0.6661]
4850	0.7803 [0.7698, 0.7698]	0.664 [0.6541, 0.6738]	0.6563 [0.6464, 0.6661]
4860	0.7804 [0.7699, 0.7699]	0.6639 [0.6541, 0.6738]	0.6562 [0.6463, 0.6661]
4870	0.7803 [0.7698, 0.7698]	0.6639 [0.654, 0.6737]	0.6563 [0.6465, 0.6662]
4880	0.7803 [0.7697, 0.7697]	0.6638 [0.6539, 0.6738]	0.6563 [0.6464, 0.6661]
4890	0.7804 [0.7698, 0.7698]	0.6638 [0.6539, 0.6737]	0.6564 [0.6466, 0.6662]
4900	0.7804 [0.7697, 0.7697]	0.6638 [0.6539, 0.6738]	0.6563 [0.6466, 0.6661]
4910	0.7804 [0.7696, 0.7696]	0.6638 [0.6539, 0.6737]	0.6562 [0.6465, 0.6658]
4920	0.7803 [0.7695, 0.7695]	0.6638 [0.6539, 0.6738]	0.6561 [0.6465, 0.6658]
4930	0.7803 [0.7695, 0.7695]	0.6638 [0.6539, 0.6737]	0.6562 [0.6466, 0.6658]
4940	0.7802 [0.7694, 0.7694]	0.6639 [0.654, 0.6738]	0.6561 [0.6464, 0.6657]
4950	0.7802 [0.7693, 0.7693]	0.6639 [0.654, 0.6738]	0.656 [0.6464, 0.6657]
4960	0.7801 [0.7691, 0.7691]	0.6639 [0.654, 0.6738]	0.656 [0.6463, 0.6657]
4970	0.7798 [0.7688, 0.7688]	0.6639 [0.6539, 0.6738]	0.6559 [0.6462, 0.6656]
4980	0.7797 [0.7687, 0.7687]	0.6638 [0.6538, 0.6737]	0.6559 [0.6463, 0.6655]
4990	0.7798 [0.7687, 0.7687]	0.6637 [0.6538, 0.6737]	0.656 [0.6464, 0.6655]

5000	0.7798 [0.7688, 0.7688]	0.6637 [0.6537, 0.6736]	0.656 [0.6464, 0.6655]