Forecasting Medical Patient Length of Stay at Presentation in an Emergency Department Using Machine Learning

Ming-Tse Tsai

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2017

Committee:

Thomas Payne, Chair

Neil Abernethy

Steven Mitchell

Program Authorized to Offer Degree:

Biomedical and Health Informatics

University of Washington

Abstract

Forecasting Medical Patient Length of Stay at Presentation in an Emergency Department Using Machine Learning

Ming-Tse Tsai

Chair of the Supervisory Committee:
Thomas Payne, MD, FACP
Professor, Department of Medicine
Adjunct Professor, Department of Biomedical Informatics & Medical Education
Adjunct Professor, Department of Health Services

Emergency Department (ED) overcrowding has become common across the globe. Among many proposed measures, ED length of stay (LOS) remains the most commonly reported outcome resulting from overcrowding. Predicting patients' ED LOS, especially as early as at presentation, could provide valuable information for both patients and providers: it could not only improve resource allocation, but also could facilitate decision-making. In addition, understanding the influence of each associated predictor enables better operational management on this complex and harmful situation. In this paper, data available at patient presentation were identified based on operational

data and patients' demographic data from an ED, and subsequently predictive modeling was attempted. Overall, the resulting model suffered from high bias, but it performed well in the subgroup of ED LOS between 1 hour and 8 hours. In addition, it was able to capture the trend of ED loading. Furthermore, influential predictors were identified, which serve to inform future more sophisticated modeling.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I'd like to express my gratitude to Dr. Kuang-Chau Tsai, the director of Emergency Medicine in Far Eastern Memorial Hospital, for his support in this study. Also, Dr. Jen-Tang Sun, M.D, Ms. Wen-Chueh Hsu and Ms. Shu-Yueh Chen helped in collecting and verifying data.

I am grateful to my committee members. Prof. Thomas Payne, along with his optimistic spirit, always encouraged me and helped me navigate in the uncertain future. Prof. Neil Abernethy proposed questions to expand my thoughts, as well as gave me timely suggestions. Dr. Steve Mitchell served as not only the clinical expert but also my spiritual mentor. Without their help, I couldn't make it. I'd like to express my gratitude to Prof. Noah Simon and Prof. Ali Shojaie, the professors in Department of Biostatistics at UW. They both provided me useful advice to curate my project.

I'd like to give credits to my family, especially my beloved wife, Hsing-Ju Hsieh, for their warm and strong supports. I thank them for allowing me the time and energy to this project. I also admired their courage for going abroad with me and staying with me.

DEDICATION

This article is dedicated to those who suffered from ED overcrowding, patients or providers, their family, and to those who ever despaired of the healthcare system.

Chapter 1. INTRODUCTION

1.1 OVERVIEW OF EMERGENCY DEPARTMENT (ED) OVERCROWDING

ED overcrowding has increased for decades. In 2003, a systematic review stated “ED overcrowding is widespread in US cities and has reportedly reached crisis proportions” [1]. However, despite efforts to reduce it, recent studies show it has become common across the globe [2-9]. ED overcrowding negatively impacts efficient patient care [10], patient outcomes [11-13] and efficiency of ED processes [9]. It also results in job dissatisfaction of providers [14] and further provider burnout [15].

Several factors were described as the reasons of ED overcrowding, including an aging population [16], patient complexity [7], ED patient flow protocol [17], delayed tests and consultants [6, 8], the lack of inpatient beds [6, 8, 18-20], and others [4]. The reasons can be further divided into 3 categories: input, throughput, and output factors, which are based on the model (Figure 1.1) created by Asplin et al [21] in 2003. Several strategies have been proposed, targeting on these 3 factors. System-level interventions were suggested in some literature, for example, the Ontario approach [2].

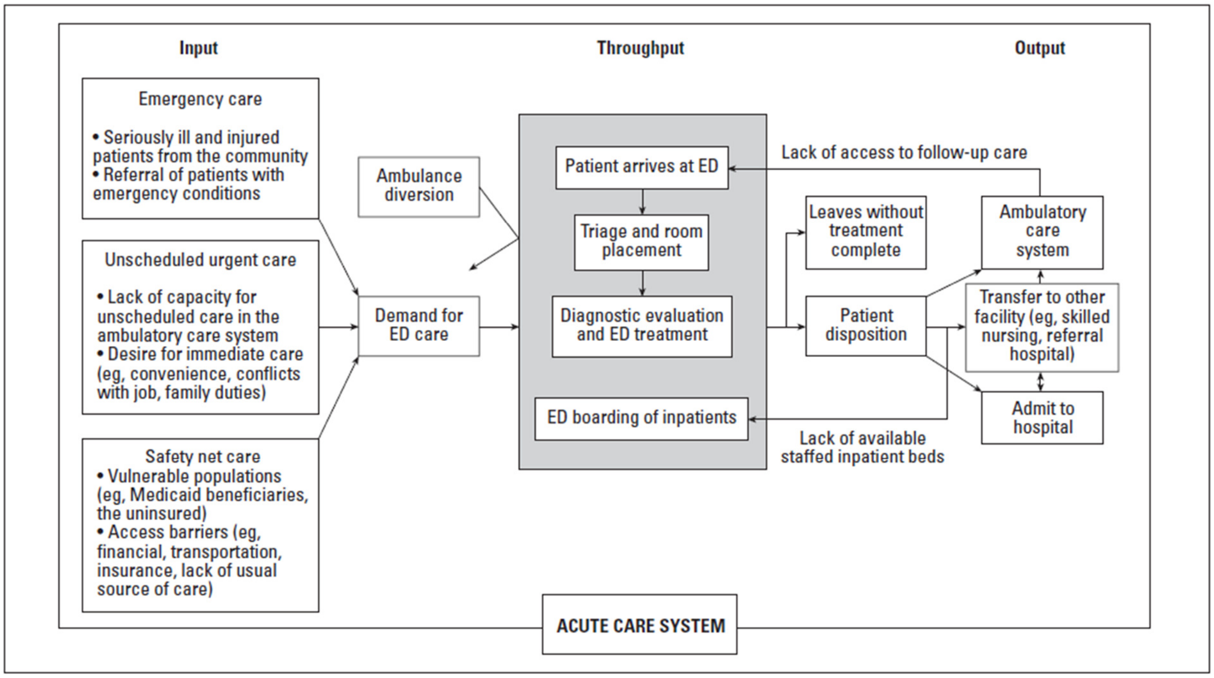

Figure 1.1. The input-throughput-output conceptual model of ED overcrowding, proposed by Asplin et al [21].

1.2 DEFINITION/MEASURES OF ED OVERCROWDING

Although ED overcrowding is a universal problem, there is no consistent definition of this phenomenon [4, 22, 23]. In 2009, Moskop et al [18] observed that "debate continues about the most appropriate definition and measures for (over)crowding," and this situation still exists. Despite the lack of the best definition, the policy statement from American College of Emergency Physicians (ACEP), revised in 2013, illustrates the problem. It stated that "(over)crowding occurs when the identified need for emergency services exceeds available resources for patient care in the emergency department, hospital, or both" [24].

ED overcrowding is characterized by increasing numbers of ED visits, longer lengths of stay in the ED, and the common practice of ED boarding [25], but there is no agreement on a single measure or a set of measures of overcrowding [4]. Persistent efforts by researchers to identify explanatory variables to determine the existence and the extent of ED overcrowding have resulted in several scoring systems available, including the National Emergency Department Overcrowding Score (NEDOCS, USA) [26], the Severely overcrowded-Overcrowded-Not overcrowded Estimation Tool (SONET) [27], and the International Crowding Measure in Emergency Departments (ICMED, UK) [28]. The required parameters are listed in Table 1.1, categorized by the input, throughput, output measures correspondent with Asplin's model (Figure 1.1). There are three critical points worthy of attention. The first one is that derivation, validation and even comparison of these systems are heavily dependent on provider perceptions [26-29], which indicate provider perceptions remain the reference standards of ED overcrowding. The second point is that there is no such set of universal measures to evaluate the extent of ED overcrowding. Despite its widespread nature, the measures and the related strategies should be tailored to different health systems.

Furthermore, Pines et al [30] found even within the same country, such as the USA, there is no single set of ED efficiency benchmarks available to account for factors that cannot be controlled by individual sites. Last but not least, among these 3 categories, the throughput measures are more relevant to EDs because they directly relate to how the ED is organized and managed. This viewpoint was also supported by Morris et al [4].

Category	Parameter	NEDOCS, USA†	SONET, USA†	ICMED, UK‡
Input measure	Ability of ambulances to offload			✓
	Patients who leave without being seen or treated			✓
	Time until Triage			✓
Throughput measure	Number of ED beds	✓	✓*	✓**
	Total Patients in the ED	✓	✓*	✓**
	Number of Respirators in the ED	✓	✓	
	Number of Acuity Level-3 Patients in the Waiting Room		✓	
	Number of Acuity Level-2 Patients Occupying an ED bed		✓	
	Longest Admit Time	✓		
	Longest Waiting Room Time of Last Patient put in bed	✓		
	Patients' Total Length of Stay in ED			✓
	Time to See a Physician			✓
Output measure	ED Boarding Time			✓
	Number of Hospital beds	✓		
	Total Admits in the ED	✓		✓

†The National Emergency Department Overcrowding Score [26].
†The Severely overcrowded-Overcrowded-Not overcrowded Estimation Tool [27].
‡The International Crowding Measure in Emergency Departments [28].
*In the SONET system, these 2 parameters are incorporated into a new parameter, which is called total patient index. It is defined as the total number of ED patients divided by the total number of ED beds [27].
**In the ICMED system, these 2 parameters are slightly differently explained and incorporated into a new parameter, which is called ED occupancy rate. This rate is the total volume of patients in the ED compared with the total number of officially designated ED treatment spaces [28].

Table 1.1. The required parameters in different scoring systems of ED overcrowding.

1.3 ED LENGTH OF STAY

Among these proposed measures, ED Length of Stay (LOS) is a common focus in the literature of ED overcrowding [20, 31-35]. In these articles, ED LOS is used either as a surrogate of ED overcrowding or as a reference candidate when interpreting the efficacy of other scoring systems. McCarthy et al [36] further emphasized that ED overcrowding resulted in lengthening ED LOS, even in those high-acuity patients. In fact, Elder et al [37] concluded ED LOS was the most commonly reported outcome variable in his review of key strategies designed to improve patient flow through the emergency department.

Furthermore, some countries have implemented a national standard of ED LOS as a target to control ED overcrowding. In 2000, the National Health Service (NHS) in United Kingdom initiated the "four-hour target" to address this issue [38]. Several years later, New Zealand and Australia introduced similar strategies, namely the Shorter Stay in ED target [3] and the National Emergency Access Target [39], to their own ED systems respectively. Ontario, Canada started a provincial program in 2008, aiming at addressing chronic ED overcrowding and improving system performance. Among mandatorily reported data, ED LOS was the main target which must be disclosed publicly [2]. In the US, the Centers for Medicare and Medicaid Services (CMS) Hospital Compare website displays data on publicly reported and audited quality measures from a national cohort of hospitals. This organization began 'pay-for-reporting' of ED timeliness measures in 2012, including median ED LOS for patients who are admitted to the hospital for inpatient care [40].

Longer ED LOS has been associated with treatment delay in patients with pneumonia [41], poor outcomes in certain patient groups [42, 43], and even higher mortality in patients admitted to intensive care units from the ED [44]. Nippak et al [45] found the positive association between longer ED LOS and longer inpatient LOS on the basis of a hospital's data in Ontario, and thus suggested "continued efforts to further reduce ED LOS are crucial, because this has the potential to influence outcomes, efficiency of EDs and succession to inpatient status, which may affect costs to the healthcare system".

In addition to the potential for unfavorable outcomes, prolonged ED LOS is also associated with poor patient satisfaction. Taylor et al [46] found the perceived waiting time was one of the three most frequently identified service factors related to patient satisfaction in ED. After a decade, Parker et al [47] concluded that the correlation between LOS and patient satisfaction was negative. Recently, Chang et al [40] found that each additional hour of ED LOS was associated with 0.7% decrease (95% CI, 0.4 - 1.0; p<0.01) in proportion of patients giving a top satisfaction rating, as well as 0.7% decrease (95% CI, 0.4 - 0.9; p<0.01) in proportion of patients who would "definitely recommend" the hospital. The decline in satisfaction could result in poorer patient compliance with recommended treatments, poor health outcomes, and increased litigation [47].

1.4 ED OVERCROWDING AND ED LOS IN TAIWAN

In Taiwan, ED overcrowding is a national problem. After the initiation of National Health Insurance (NHI) in 1995[48, 49], the situation worsened. In 1999, Shih et al [50] stated significant overcrowding existed in EDs in Taiwan. Based on his analysis of the ED data from a single tertiary hospital, 3.6% of patients were held in the ED for more than 72 hours. The major reason was inpatient bed unavailability. In 2014, Hsu et al [7] investigated the association between patient characteristics and prolonged ED LOS. In their sample of 1,364 general medicine patients admitted from the ED, the mean ED LOS was 43.9 ± 41.0 hours! Among the subgroup of 207 patients with prolonged ED LOS, this number increased to 107.0 ± 63.5 hours.

Meanwhile, other researchers described this phenomenon in different ways. Tsai et al [51] used the 2002 Taiwan National Health Insurance Research Database to identify nonemergent emergency department conditions. They found nearly 15% of all emergency department visits were nonemergent, and an additional 20% visits could have been prevented by use of primary care. In 2015, Ko et al [52] conducted a population-based cross-sectional study. They leveraged data of 1 million people randomly selected from all beneficiaries of Taiwan's National Health Insurance claim database in 2010. They finally found 170,475 subjects of the 1 million beneficiaries used ED service in 2010, among which 8.2% and 0.3% had 4 to 12 and more than 12 ED visits, respectively. This might be explained partly by the unlimited and easy access to emergency services for all citizens under the NHI program [50].

In response, the Department of Health, along with the Bureau of National Health Insurance in 2012, launched a program and allocated NTD$320 million to increase ED quality and efficiency [53]. One of the three major goals of this program was to increase ED treatment efficiency. Hospitals will be rewarded financially if they achieve the 75th percentile among hospitals of the same level in any of the following:

1) less than 1% of ED patients stays are over 24 hours,

2) ED patients with major illness are transferred to the intensive care unit (ICU) within 6 hours after arrival,

3) ER patients on the first three levels of the emergency triage are hospitalized within 8 hours after arrival,

4) ER patients on level four or five of the emergency triage are discharged within 4 hours after arrival.

Besides financial incentives, the ratio of ED LOS over 48 hours has been added to the criteria in Hospital Accreditation Standards in 2015 [54]. The goal was set at either 0 or lower than the average among hospitals of the same level in 3 years.

1.5 Factors associated with ED LOS

Since ED LOS is such an important quality measure, several factors were identified in the medical literature to be associated with ED LOS.

Firstly, patient characteristics play a significant role in influencing ED LOS. Yoon et al [32] concluded that triage acuity, investigations, and consultations were important independent variables that influenced ED LOS. In 2016, Kreindler et al [55] conducted a systematic review of studies that assessed at least one patient-level predictor of ED LOS in an adult or mixed adult/pediatric population within an Organization for Economic Cooperation and Development country. They found that the most common factors associated with long ED LOS were need for admission, older age, receipt of diagnostic tests or consults, and ambulance arrival. They also suggested that specific patient complaints should be included.

Secondly, ED census was also mentioned as a contributor. After analyzing about 1,000,000 visits in 53 EDs in France, Capuano et al [56] concluded that ED LOS needs to be stratified by case mix and the total number of visits of the ED. Meanwhile, several American studies found similar results. Pine et al [30] found ED volume is one of the most significant variables associated with waiting times and lengths of visit based on their observations on 424 US hospitals. Furthermore, Handel et al [57], from 6-year observation on 445 EDs in the USA, found not only higher-volume EDs but also higher inpatient bed occupancies were associated with higher LOS.

Thirdly, the day of the week and ambulance diversion periods were found to impact ED LOS. Wiler et al [58] in their retrospective multicenter study found that ED LOS increased on days with higher percentage daily admissions, higher elopements, higher periods of ambulance diversion, and during weekdays, whereas LOS decreased on days with higher numbers of discharges and weekends. That also makes sense in regard to inpatient bed occupancy. The new information here is ED LOS might be different from day to day within a specific week. There might be underlying reasons for this phenomenon.

Besides these factors, there are some interesting factors identified in the studies recently conducted in Taiwan. Hsu et al in 2014 [7] found longer ED LOS was observed in general medical patients of high Charlson Comorbidity Index (CCI) score ($\geqq$ 3). For patients with do-not-resuscitate (DNR) consent, the odds ratio of prolonged ED stay was 1.60. Chaou et al [59] in 2017 investigated the influence of factors on the ED LOS of ED discharged patients. In their model, transferred patients, day shift arrival (8am-4pm), and increased ED daily census[1] were associated with prolonged ED LOS.

Although these ED LOS-related factors have been identified in the published literature, it is hard to tell which one(s) plays a bigger role than the others. Though some were regarded as important features in terms of ED design, operations, and policy decisions [60], the relationship between ED LOS and those features is not clear enough to fully predict ED LOS. Although some efforts have been made, there is no perfect predictive model to inform either operational officers, physicians or patients for their decision making. Kreindler et al [55] concluded that "...the available information is insufficiently precise to inform clinical or service-planning decisions; there is a need for a predictive model, …".

1.6 MACHINE LEARNING IN HEALTHCARE

Machine learning is the scientific discipline that focuses on how computers learn from data. Here the idea of learning is to be able to capture concepts by data without being explicitly programmed. Although it is not a new idea, it is well known in recent years thanks to speedy computation and efficient memory storage from computers. Different approaches, often described as algorithms, were invented to help computers learn from data and enable to predict the future trends.

There are two types of machine learning, namely supervised learning and unsupervised learning. Supervised learning starts with the goal of predicting a known outcome. In contrast, unsupervised

[1] In Chaou et al's model, the ED daily census was incorporated as a binary variable with a cutoff point of 558 patients, which was the 95th percentile of the study ED's daily census.

learning attempts to group the observations (clustering) by their natural pattern. According to learning tasks, supervised learning can be further defined into 2 categories: classification and regression. In classification problems, one tries to predict a discrete number of values. In contrast, regression is used to predict a continuous variable.

Machine learning has been applied in a broad range of fields, from financial services to outer space exploration, from personal preference prediction to presidential election prediction. A lot of efforts have been made in many areas of medicine, such as assisting in diagnosis [61], medical imaging reading [62], predicting patient outcomes [63], and hospital resource allocation [64]. With broad implementation of electronic health records (EHR), the future of machine learning application in healthcare seems more promising because EHR provides abundant digital health data serving as fuel for machine learning.

1.7 ED LOS AND MACHINE LEARNING

The idea of applying machine learning to forecast ED LOS is appealing because ED LOS is of great importance, valued by many stakeholders, and machine learning seems a powerful tool well suited to solve this kind of problems. In fact, being able to precisely predict patients' LOS might improve resource management both in an ED and a hospital. However, there are few studies available in the literature.

Azari et al in 2015 [65] attempted to tackle this problem as a classification task. The study focused on different approaches to identify the patient group with prolonged ED LOS (longer than 14 hours) using data available at presentation. They notified it was an imbalanced classification task, in which the prolonged stay group was the minority group. To address this concern, they proposed ensemble methods, which performed better than the traditional logistic regression models in terms of the overall classification performance measures (F1 and G-mean). However, the best F1 score they could achieve was still less than 0.7. Moreover, rather than providing individualized LOS forecast for each patient, the model could only tell whether a specific patient would stay longer than 14 hours or not.

Meanwhile, some scientist treated this LOS problem as a regression task. Wrenn et al [66] in 2005 developed and validated an artificial neural network by using about 16,700 ED patient data points with clinical and operational parameters available at presentation in Vanderbilt University Medical Center in Nashville, Tennessee. Patients were heterogeneous, including medical and surgical patients. Although the model's performance on the training set predicted LOS within an average of 2 hours, it declined to an average of 7.5 hours in the validation set. Even though they ran the model again on chief complaint-specific subsets of data, the result was still unsatisfying (within an average of 3.5 hours). In 2015 in Turkey, Gul et al [67] also leveraged an artificial neural network model on 1,500 patient data points with a total of 19 features (patient demographic features and a series of examination features). Their final result, reported as R squared, was 0.63.

Other scientists adopted linear regression methods. To better inform stakeholders, Ding et al [68] in 2009 used quantile regression of 9 features (both patient-level and system-level features) available at presentation to predict the 10^{th}, 50^{th} and 90^{th} percentile of ED LOS. They divided ED LOS into 3 components: waiting time, treatment time, and boarding time. They predicted the distribution of each component based on these features. In addition to successfully identifying patients' chief complaint and acuity level as the most important predictors of the three components, day and time of arrival were found as important predictors of wait time and boarding time. Although their model provided general information for patients and healthcare providers, they failed to forecast each patient's LOS. In 2014, Combes et al [69], a French team, presented a new methodological framework based on predictive data mining approach to estimate ED LOS. They proposed two linear regression models, both of which consisted of examination features (X-ray, laboratory tests, etc.). The better one can predict 28.24% of the test group correctly, and 61.83% for an error interval less than ± one hour. However, the models were generated from and validated on a pediatric emergency department in France, which limited its ability of generalization to an adult ED.

1.8 Predicting ED LOS at presentation

The key to accurate prediction is capturing a combination of patient characteristics and the current operating state of the ED [65]. However, the longer a patient stays in an ED, the more factors come to play that influence his LOS. And this statement is strongly supported by one of the French models described above, which is a linear regression function consisting only of 4 examination features [69]. On the other hand, being able to predict a patient's LOS early at presentation would be similarly effective in addressing barriers to expedited treatment and ED disposition. It would enable an ED manager to estimate the oncoming loading and better prepare by activating its backup resources. It also offers healthcare providers general information so that they could communicate better with either a patient or his family. Although the predicted bias is introduced by eliminating some features which could only be collected after presentation, the gained benefit from timely information and the resulting early response is much valuable and worthwhile. Moreover, prediction based on the information available at presentation might be achievable. In Ding's study [68], they concluded ED LOS varied significantly among patients, but in a predictable manner that is largely explained by information available at triage.

Chapter 2. METHOD

The proposed methodology framework aims at predictive modeling to approach the LOS of an arriving ED medical patient. The goal is to identify the most powerful predictors in order to predict LOS, with a useful model in the setting of an ED in Taiwan. The methodology adopted to achieve this objective is depicted in Figure 2.1. Each step is described in the next subsections. The UW Human Subjects Division determined that this work does not involve "human subjects" as defined by federal regulations and so does not require exempt status or IRB review.

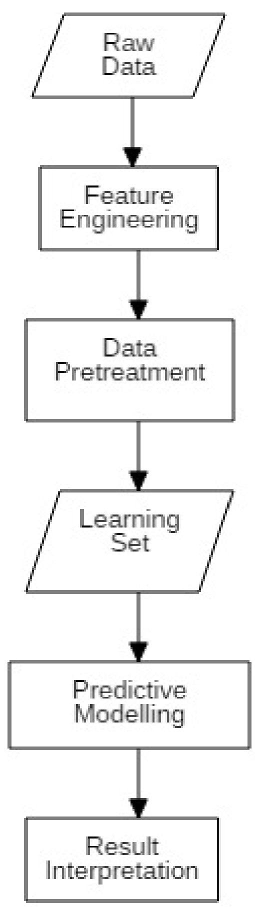

Figure 2.1. The proposed methodology framework. The parallelograms indicate data, and the rectangles indicate processes.

2.1 THE STUDY SETTING: FAR EASTERN MEMORIAL HOSPITAL, NEW TAIPEI CITY, TAIWAN

Far Eastern Memorial Hospital (FEMH), established in 1981, has been accredited as a medical center since 2006. FEMH has 1,174 acute-care beds and serves about 5,000 outpatient visits and 400 ED visits daily. It is in New Taipei City, where about 4 million inhabitants live.

The ED in FEMH is one of the 3 busiest EDs in Taiwan, with about 10,000 patient visits per month. Among these patients, about 1,000-1,200 cases are sent by ambulance. On holidays, the number of daily visits can reach 800 to 1000. There are 4 subunits in the ED, including medical, surgical, pediatric, and obstetric/gynecologic (OB/GYN) subunits. When a patient arrives in the ED, several questions are asked during an assessment by an experienced triage nurse. After the patient's vital signs are measured (except those in critical or unattainable conditions), a triage level[2] and subunit category are assigned to the patient according to the chief complaint and the judgment from the triage nurse. The triage nurse will enter the patient's demographic data, vital signs, and his chief complaints into the hospital information system. In 2015, the number of annual visits to the FEMH ED was 129,924, of which 66,887 (51.48%) were medical visits, 40,935 surgical visits, 21,281 pediatric visits, and 821 OB/GYN visits, respectively.

---

[2] Taiwan triage and acuity scale, TTAS

The five-level Taiwan Triage and Acuity Scale (TTAS) computerized system was officially launched nationwide in Taiwan in 2010. This system, based on the structures and the contents from the Canadian Triage and Acuity Scale system, had been validated to demonstrate its reliability and accuracy [70]. Besides the predefined standard terms of chief complaints, 3 first order modifiers were adopted in order to triage patients correctly, which included vital signs, pain severity and high-risk mechanisms of injury [71]. According to a patient's chief complaint and vital signs, a specific triage level would be assigned to this patient, from 1 to 5, indicating the acuity and the need for this patient (Figure 2.2 [72]). Triage 1 indicated the most critical patients, and triage 5 represented the least. Recently, the updated version of this system was declared by the Ministry of Health and Welfare and implemented in Jan 1, 2016 [71].

Except the OB/GYN subunit which is located on the 4th floor, all the other 3 subunits are in the main area of the ED, which is on the ground floor. After patients are triaged, they are led to the waiting area of each subunit. An assigned physician then visits him, and subsequently may order exams or prescribe medications depending on the situation. After a certain amount of time, the physician in charge will make a decision with the patient on the next step regarding either discharge or admission. If the decision of admission has been made, the patient is boarded, which means he is on the admission list. The timestamp will be recorded in the electronic system as "Time of Decision to be admitted". This boarded patient has to stay in the ED until there is an available bed in hospital ward or ICU for him. Once there is a bed, the patient and the nurse in charge will be notified. After hand-off with colleagues in ward by phone, the patient will be sent to the ward. And the timestamp here will be recorded as "Discharge Time from ED". On the other hand, if the decision is to discharge the patient instead of admission, the patient will receive medications and discharge advice. The final procedure for the patient is to go to the registry to pay copayment, and this timestamp is recorded as "Discharge Time from ED".

Figure 2.2. The Taiwan Triage and Acuity Scale, declared by the Ministry of Health and Welfare.

There are 42 observation beds in the ED, which are shared by medical, surgical and pediatric patients. In the medical subunit, there are at least 3 attending physicians, 1 nurse practitioner and 12 nurses working on the basis of 8-hour or 12-hour shifts. In some typical high-flow durations, additional providers will join the team to share the burden of care and improve the efficiency. Because FEMH is a teaching hospital, residents will sometimes join the team to take care of patients according to their training schedule. In addition, there is a fast-track system for non-urgent medical patients in the study ED, which is a common phenomenon of high volume EDs throughout over Taiwan. The target patient population is all adult non-trauma patients with triage level 3 or 4 or 5. Generally, they are either ambulatory or able to wait in a wheelchair. This system is used from 8 am to 11 pm every day.

The electronic medical record system in FEMH was initially implemented in June 2009. In the ED, administrative information, including registration, boarding and discharge, is manually entered into a database by clerks. Operational information, including medical orders and nursing orders, is stored in different databases, aligned with the outpatient systems. Laboratory results and image information are stored in other databases but could be extracted through links at the interface of the systems.

2.2 THE DATA

The data are extracted from the hospital information system in FEMH. However, they are from different databases and in different formats. These data points must be merged and cleaned. The collection period for this work is from Feb 1, 2016 to Sep 30, 2016. The approval from the hospital was obtained. All of the analyses were completed using the statistical package R, version 3.4.0 (http://www.r-project.org).

2.2.1 *Raw data*

There were 6 sets of files in the original data. The first set, "ED_Master" files, contains information about patients' administrative data (described in Table 2.1). The second set, "ED_first_order" files, contains information about patients' demographic data (described in Table 2.2). The third set, a single file of "Chief complaint data set of medical patients", contains information about patients' initial vital signs and chief complaints (described in Table 2.3). The fourth set, a single file of "Staffing data set", describes the number and the levels of healthcare provider hourly (described in Table 2.4). The fifth set, a single file of "Ambulance diversion data set", contains the exact starting time and the ending time of ambulance diversion from Jan 1, 2016 to Sep 30, 2016 (described in Table 2.5). The sixth, "Hospital bed availability data set", is a set of files; each file describes the daily report of the number of empty ward beds, which is reported at 11:59 pm every day (described in Table 2.6).

Variable	Type	Definition
regisDate	Date	Date of the patient arrival.
subdepart	Categorical	The subunit to which the patient is assigned by a triage nurse. (Medical, surgical, pediatric, or gynecologic)
Triage	Categorical	The patient's acuity assigned by a triage nurse. The range is from 1 to 5, where 1 is the most critical, and 5 the least critical.
byAmbu	Boolean	1=the patient is sent to ED by ambulance. 0=by other ways than ambulance.
regisTime	Hour/Minute	Time of the patient arrival.
EDdiscTime	Date/Time	Date/Time of the patient discharge.
EDdiscStatus	Categorical	The patient's discharge status, including admission, against advice discharge, discharge home, transfer, expired, escape, transfer, and for diagnosis certificate.
TimeAdmDecision	Date/Time	Time of a physician's decision to admit (only present in admitted patients).

Table 2.1. Raw data 1: Variable description of data in "ED_master" files.

Variable	Type	Definition
regisDate	Date	Date of the patient arrival.
subdepart	Categorical	The subunit to which the patient is assigned by a triage nurse. (Medical, surgical, pediatric, or gynecologic)
Triage	Categorical	The patient's acuity assigned by a triage nurse. The range is from 1 to 5, where 1 is the most critical, and 5 the least critical.
regisShift	Categorical	Shift when the patient arrives in the ED. 1=12am-8am 2=8am-4pm 3=4pm-12am
regisTime	Hour/Minute	Time of the patient arrival.
EDdiscTime	Date/Time	Date/Time of the patient discharge.
EDdiscStatus	Categorical	The patient's discharge status, including admission, against advice discharge, discharge home, transfer, expired, escape, transfer, and for diagnosis certificate.
gender	Categorical	Gender of the patient. 1=male 0=female
DOB	Date	Date of birth.

Table 2.2. Raw data 2: Variable description of data in "ED_first_order" files.

Variable	Type	Definition
regisDate	Date	Date of the patient arrival.
Triage	Categorical	The patient's acuity assigned by a triage nurse. The range is from 1 to 5, where 1 is the most critical, and 5 the least critical.
regisTime	Hour/Minute	Time of the patient arrival.
EDdiscTime	Date/Time	Date/Time of the patient discharge.
gender	Categorical	Gender of the patient. 1=male 0=female
DOB	Date	Date of birth.
SBP	Numerical	Systolic blood pressure at presentation, measured in mmHg.
DBP	Numerical	Diastolic blood pressure at presentation, measured in mmHg.
HR	Numerical	Heart rate at presentation, measured in beats per minute.
RR	Numerical	Respiratory rate at presentation, measured in counts per minute.
SPO2	Numerical	Oxygen saturation at presentation, measured by oximeter in percentage.
GCS	Numerical	Glasgow coma scale at presentation, assessed by a triage nurse.
Pain_scale	Numerical	Pain severity score at presentation, reported by the patient from 0 to 10.
Chief complaint	Character	Chief complaint at presentation, reported by the patient and recorded by a triage nurse with the computer-assisted system.

Table 2.3. Raw data 3: Variable description of data in "Chief complaint data set of medical patients" file.

Index	Type	Definition
Date	Date	Scheduled date of work.
Hour	Categorical	An hour of a specific date. A day is divided into 24 hours. Each row indicates that hour in the specific date.
Variable	**Type**	**Definition**
VS_ED	Numerical	The number of attending physicians taking care of oncoming medical patients in a specific time period.
VS_round	Numerical	The number of attending physicians taking care of the medical patients in the observation room of the ED in a specific time period.
VS_boarding	Numerical	The number of attending physicians taking care of the boarding medical patients in a specific time period.
PGY	Numerical	The number of post-graduate-year students in the medical subunit in a specific time period.
R1	Numerical	The number of first-year residents in the medical subunit in a specific time period.
R2	Numerical	The number of second-year residents in the medical subunit in a specific time period.
R3	Numerical	The number of third-year residents in the medical subunit in a specific time period.
R4	Numerical	The number of fourth-year residents in the medical subunit in a specific time period.
NP	Numerical	The number of nurse practitioners in the medical subunit in a specific time period.
NP_extra	Numerical	The number of extra nurse practitioners in the medical subunit in a specific time period.
RN	Numerical	The number of nurses in the medical subunit in a specific time period.
RN_extra	Numerical	The number of extra nurses in the medical subunit in a specific time period.

Table 2.4. Raw data 4: Variable description of data in “Staffing data set” file.

Variable	Type	Definition
Date	Date	Date of requested ambulance diversion.
StartTime	Hour/minute	The start time of ambulance diversion.
EndTime	Hour/minute	The end time of ambulance diversion.

Table 2.5. Raw data 5: Variable description of data in "Ambulance diversion data set" file.

Variable	Type	Definition
Date	Date	Date of report.
Ward	Categorical	The name of each ward in FEMH.
EmptyNum	Numerical	The number of empty beds in a specific ward.

Table 2.6. Raw data 6: Variable description of data in "Hospital bed availability data set" files.

2.2.2 *Data merging and feature engineering*

Due to various formats of data, efforts had been made to organize these data for analysis, which is the process of "representation". Furthermore, feature engineering was required to distill more information from these data. Figure 2.3 illustrated the data pipeline. These features were selected because they were either identified from the previous research or used in UW Medicine, or from my personal clinical experience. The goal was to create different attributes for each patient based on his presentation time and situation.

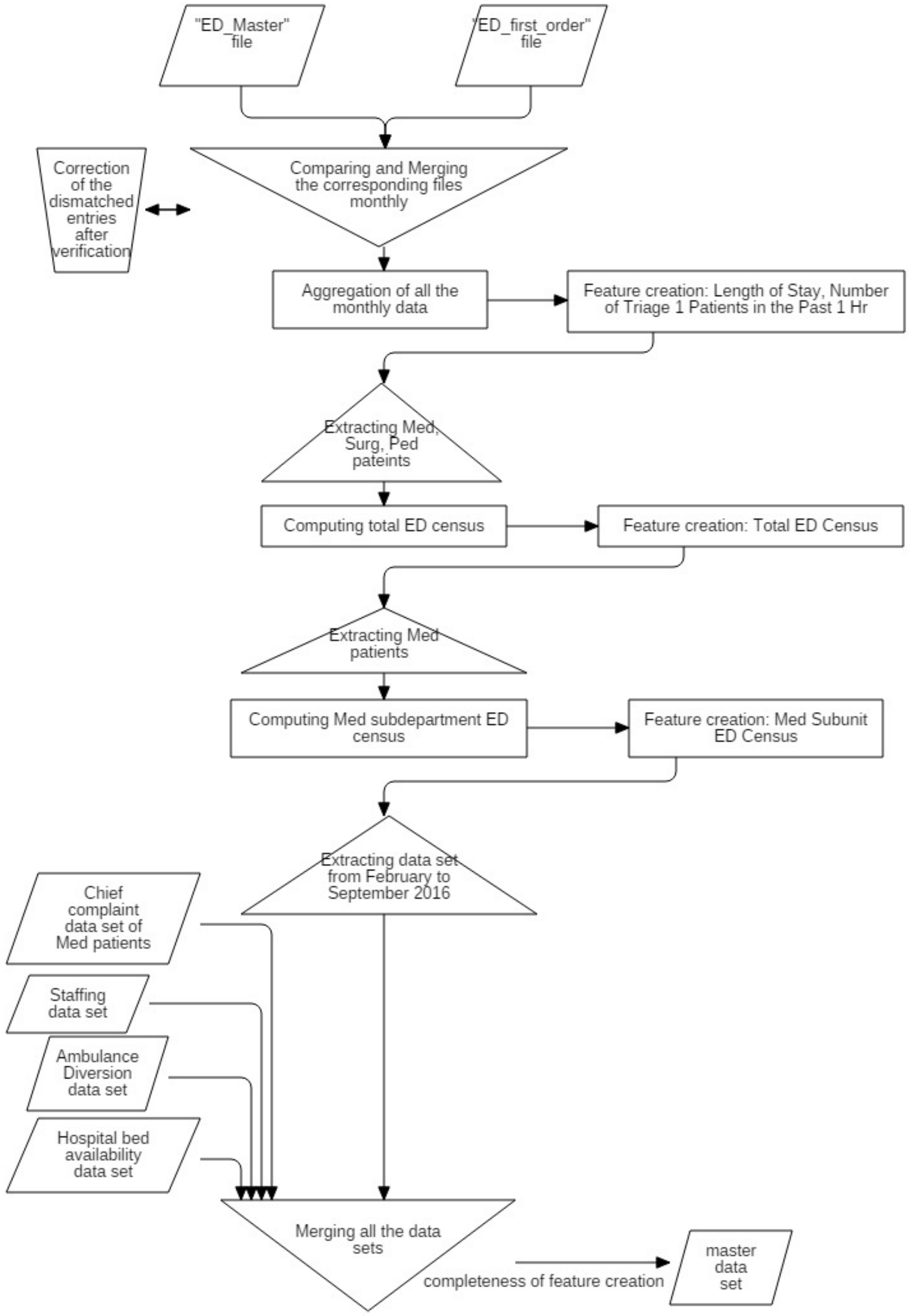

Figure 2.3. The data pipeline in the study. The parallelograms represent data. The trapezoid represents manual operation. The inverted triangles indicate merging the data. The triangles indicate extracting the data. The rectangles represent processes.

Besides calculating the associated ED census and medical subunit census for each patient, incorporating a patient's chief complaint needs to be addressed. Traditionally, chief complaints are free text, describing how a patient addresses his concern about this ED visit in his/her own words, which results in both difficulties in information extraction and requirement of new data mining skills such as natural language processing (NLP). Fortunately, the five-level triage computerized system was officially launched nationwide in Taiwan in 2010. Since then, two parts were used to describe a patient's chief complaint. The first part is using the predefined controlled terminology, and the second part is free text style as in the past. Because of this change, one can easily extract the information from chief complaints. The official, published, adult nontrauma controlled terminology is provided in Appendix A. Here a series of dummy variables based on these standard chief complaint terms were created to capture the information of a patient's chief complaint. For example, the variable "res1" would be 1 if "shortness of breathing" is present in a patient's chief complaint. Otherwise, this value would be 0.

As a result, a master data set was created, which contained 47,807 records and 190 columns. Each row indicated a single medical patient. And each column represented an attribute of this patient.

2.2.3 *Data pretreatment*

The phase of data pretreatment included outlier detection and exclusion, data transformation, few occurrence feature dropping, and missing data management, depicted in Figure 2.4. Each procedure is addressed in the next subsections.

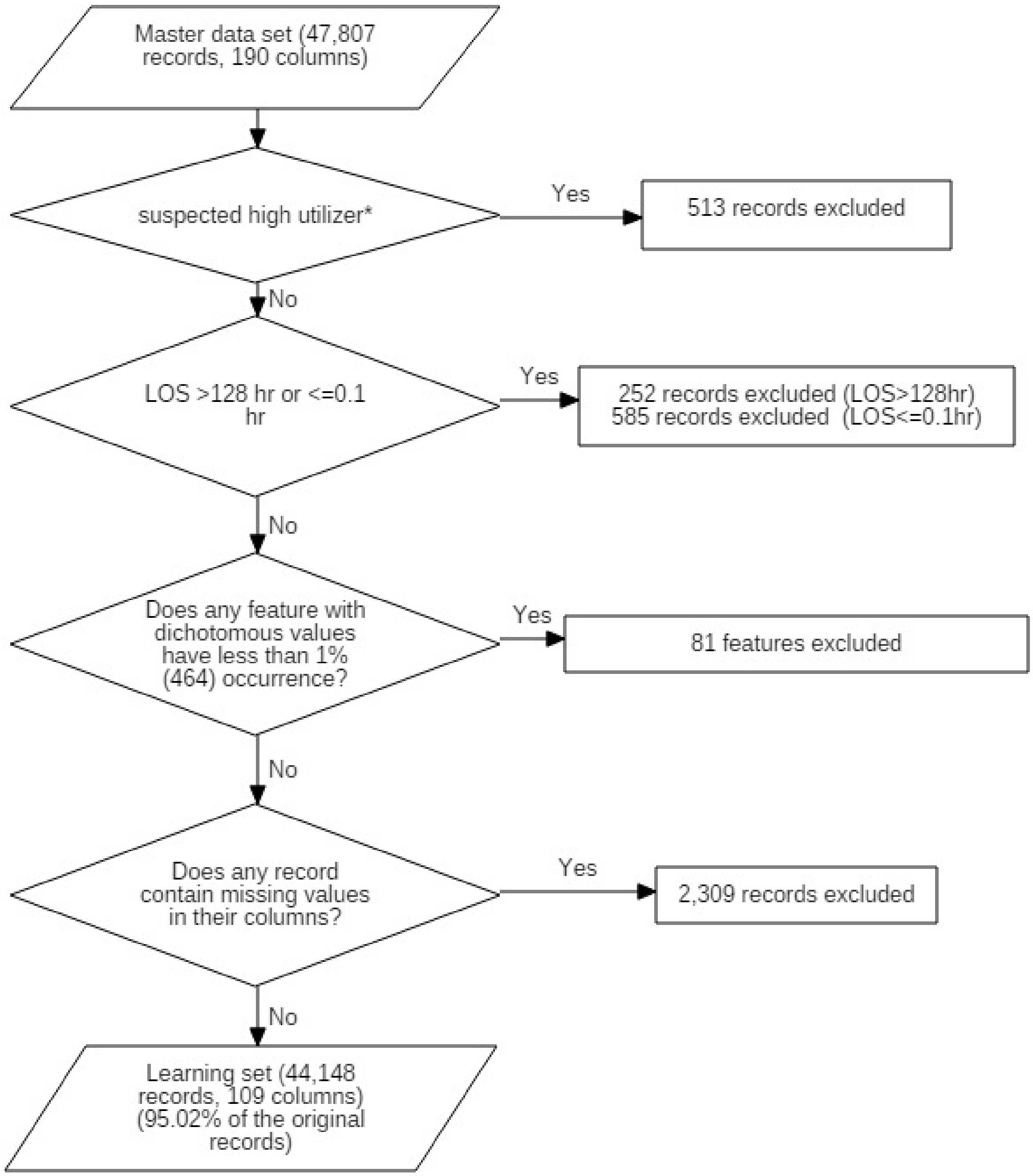

Figure 2.4. Data pretreatment flow. (*See text for definition.)

2.2.3.1 Outlier detection and exclusion

During the study period, there are 47,807 medical visits to the ED in FEMH. However, there are some outliers among these data points.

One problem is what we refer to as “high utilizers”. They tend to use ED services much more frequently than others. They might come to an ED because of medical problems, but more commonly they visit an ED because of social problems. The observations including high utilizers would greatly distort the data and hence the predictive model. Therefore, these records must be removed in advance of data exploration.

Without the information of medical record number, it is difficult to tell which record comes from those high utilizers. Here, a set of surrogate indicators was applied in an attempt to identify them, which was the attribute “gender” and the attribute “date of birth (DOB)”. By this method, all the patients were grouped by their gender and date of birth. The biggest number of visit is 513, which belongs to a group whose gender is female and the date of birth is some day in 1975. The number of the second frequently visiting group is 62. Furthermore, for this most frequently visiting group, the mean of ED LOS measured in hour is 0.287 and the variance is 5.51. Therefore, this group was excluded from the data set because this suggests a high utilizer.

There are some outliers in the extremely right-skewed distribution of the response variable, ED LOS measured in hour (encoded as EDLOS_hr). The mean of this variable is 8.57 and the median is 2.30. 585 observations have the value less than or equal to 0.1, which means they stay no longer than 6 minutes. Usually, they are for drug refill or a certificate, which doesn’t consume substantial ED resources. On the other hand, 252 observations have the value higher than 128, which means they stay more than 5 days in ED. I regard these data points as outliers and excluded them from my data set.

2.2.3.2 Data transformation

2.2.3.2.1 Transformation of the explanatory features

Kuhn et al in their book "Applied predictive modeling" [73] claimed that many machine learning algorithms work better if the features have symmetric and unimodal distributions. Therefore, 4 features with a skewed distribution were recognized and transformed in simple and easily interpretable ways to make the distribution more closely approximate a normal distribution, including square root transformation to 4 explanatory variables (hrtillnextbusihour, hrtillnextfullexam,surbedC, surbedNHI, explained in Appendix B).

2.2.3.2.2 Transformation of the responsible variable: from "EDLOS_hr" to "EDLOS_hr_log"

Although accuracy is the ultimate goal for all predictive models, practicability should be considered; that is, specific user needs should be considered in model design. For all time-sensitive predictive models, the priority is to precisely predict the imminent events. If an event is going to happen longer in the future, people usually can accept less precise prediction, such as weather prediction. And that is also the case in predicting ED LOS "at presentation". Assuming a patient's predicted LOS is 2 hours, the model's performance would be considered good if the final LOS were 4 hours (2 hours more) rather than 8 hours (6 hours more). In contrast, if the predicted LOS is 5 days, the criteria of evaluating the model's performance will be different because one might expect more variance and accept the final LOS even up to 1 week. Therefore, the different needs of accuracy on different LOS should be addressed.

Here, logarithmic 2 transformation was applied to the original response variable, "EDLOS_hr", to generate the new response variable, which is "EDLOS_hr_log".

$$EDLOS\_hr\_log = log_2(EDLOS\_hr)$$

For the rationale described above, this transformation also made the distribution closer to a normal distribution. Furthermore, most of the physicians and all the nurses work eight-hour shifts. If the predicted value of a specific patient was 2, the expected LOS would be half a shift, which is 4 hours. If the value was larger than 3, it was easy for providers to get that the patient would stay over a shift.

2.2.3.3 Few occurrence feature dropping

115 out of 190 features were deduced from patients' chief complaints, which were binary variables indicating whether or not a specific term is present in their chief complaint. However, some of them had only few occurrences. Such infrequent data make these features very susceptible to noise, which will lead to a flawed model. Therefore, 81 chief-complaint-related features were excluded by the criteria of less than 1% (464) occurrence. Appendix A also described the number of occurrence in each binary deduced variable and whether the variable was selected into the final model or not.

There are reasons to explain such low occurrence for specific features. Some are because the natural prevalence is low. For example, the variable CV1, indicating "cardiac arrest", only has 321 hits. Even when adding the number of hits of the variable res2, "respiratory arrest", the total amount is 364, which means there are only 364 patients sent to the ED with presentation of either cardiac or respiratory arrest during the study period. Some are because of traditional internal rules of subunit assignment in FEMH ED. A typical example is that patients with chief complaints concerning eyes or ears would be directed to the surgical subunit.

2.2.3.4 Dealing with missing values

Unavoidably, in the data set there were 2,309 records with missing values across 12 features. These 12 features could be divided into 3 categories, namely vital-sign related, bed availability related, and age.

There were 7 vital-sign related features with missing values, including systolic blood pressure, diastolic pressure, heart rate, oxygen saturation level, respiratory rate, Glasgow coma scale, and pain severity scale. These features were extracted directly from one file of the raw data, that is, "Chief complaint data set of medical patients" file. The numbers of missing values in each feature ranged from 5 (pain severity scale, denoted as "pain_scale_num") to 642 (oxygen saturation level, denoted as "SPO2_num"). The presence of missing values in these features occurred for several reasons. Firstly, vital signs were not assessed at triage because of either infeasibility or unnecessity. The former example included if a patient is in air hunger, the triage nurse would push the stretcher directly into the resuscitation area so that the patient could be managed immediately. For these patients, their first vital sign numbers could be found in their medical records, which was beyond the scope of this study. One of the latter examples was that a patient came to the ED simply for a certificate. There was no need to check the patient's vital signs. Secondly, the triage nurse might forget to type in the numbers when he/she was very busy. However, the resulting missing data were expected to be rare because some preventive measures such as internal audit systems had been rigorously conducted already by the nursing department of the ED.

The second source of the missing values is the empty bed reporting files. As mentioned previously, each file contains the numbers of empty beds in wards on each day. Among the study period of totally 183 days, however, 7 daily reporting files are missing, including Feb 17, Apr 23, Jul 17, Jul 27, Jul 29, Aug 8, and Aug 18. Therefore, the patients coming on the next day of these dates wouldn't have the associated values indicating the numbers of empty beds. The amount of this patient group is 1,299.

The final source of the missing values comes from ages. In 17 records, the values of "Date of Birth" are blank. The possible reasons why these cells are blank are including there is no identification document available while assessed at triage.

Although some attempts are made to impute these missing values[3], records with the missing values inside are excluded from the final data set because they account for less than 5% of the whole data set. Therefore, 44,148 records are preserved, which are 95.0% of the original data set.

---

[3] Attempts to impute the missing values

Except for the missing values of age, attempts had been made to impute the missing values from 2 other sources in a 2-step approach.

Step 1. Targeting missing values of empty bed number

Because 7 daily reports were missing, 1,299 records had no information in the features regarding the numbers of empty beds in the hospital, which were medbedC, medbedNHI, surbedC, and surbedNHI respectively. A data set was created from the original data set, including the covariant matrix (the numbers of empty beds on the previous day, the numbers of admission from the ED on the day, the numbers of admission from the ED on the next day, and the numbers of empty beds on the next day) and the response vectors (the numbers of empty beds on the day). Then for each feature, a linear regression model was fitted to the vector and the covariant matrix. The final model was determined by only the significant features with p value less than 0.05, and thereafter used in predicting the corresponding missing values in the vector. For example, the number of empty beds in medical wards on July 17 was missing. The values including the number of empty beds in medical wards on July 16, the

2.2.4 *Learning set*

After the preceding procedures, there were 44,148 records and 109 features in the final data set. Appendix B describes the names, the definitions, the attributes, the units, and the sources of these features. Besides of the responsible variable "EDLOS_hr_log", other 108 features could be further divided into 4 categories, including demography (4), chief complaint (34), vital sign (7), and registration status (63). Registration status features could be further divided into 4 subcategories, which were general indicators (4), hospital factors (7), ED factors (43) and ambulance-related indicators (9). There were 3 subgroups in ED factors, including the entire ED census (16), medical subunit census (15), and medical subunit staffing factors (12). Among 109 features, there are 59 numerical features and 50 categorical features. Among 50 categorical features, 45 are dichotomous and 3 are ordinal.

---

number of medical admission from the ED on July 17, the number of admission from the ED on July 18, and the number of empty beds in medical wards on July 18 were used as the predictors. Then these predictors were thrown into the fitted linear regression model so that the number of empty beds in medical wards on July 17 could be imputed.

Step 2. Targeting missing values in 7 vital-sign related features

Although the values of these features were missing, other features such as a patient's demographic data and chief complaints were still available in the data set, which were informative of a patient's vital sign. Here k nearest neighbor imputation method with k =10 was applied to the subset of the original data set including a patient's demographic features, chief complaint features, and vital-sign related features (see the description in the next section "2.2.4 Learning set"). The unrelevant registration status features were kept out in the subset. And so was the response variable "EDLOS_hr_log" in order to prevent data leakage.

The corresponding results were shown in the footnote in Chapter 3. "Results".

2.3 MODELING

2.3.1 *Algorithm selection*

The goal of this study is to predict LOS of the ED medical patients, as well as to inform which variables are most powerful. Firstly, the framing question about LOS is a regression problem rather than classification. Secondly, an algorithm able to provide feature importance would be the method of choice. Fortunately, such a function is provided in at least 2 algorithms, which are Least Absolute Shrinkage and Selection Operator Regression (simply called Lasso regression) and Random Forests.

Lasso regression is a kind of linear regression model. By adding a regularization term (alpha times L1 norm of the weight vector) to the cost function, all the corresponding weights of predictors are forced to be small rather than large in absolute value. In this way, the corresponding weight of those insignificant predictors would be 0 when the regularization term reaches to some optimal point. Therefore, Lasso regression outputs a model with few nonzero predictor weights and thus these "remaining" predictors are considered important. The importance of each predictor could be ranked by the absolute value of its weight. However, there are disadvantages to Lasso regression. Firstly, it is an automatic method which doesn't incorporate human judgement. It simply outputs the mathematically significant features that have greater than minimal effects on the outcome variable. But in some cases, a small effect might be substantially important. Secondly, the resulting model is linear, which might fail to capture the nonlinear relationship between the predictors and the outcome variable.

On the other hand, a Random Forest is an ensemble learning method from decision trees. To protect against overfitting, which is a common drawback of decision trees, the Random Forest algorithm introduces randomness in two ways. The first one is creating many bootstrap samples[4] and fitting a tree model to each bootstrap sample. The second one is instead of searching the best feature when splitting a node, it searches for the best feature among a random subset of features. In this way, the correlation between trees could be further decreased to yield an overall better model. Meanwhile, it can handle not only numerical variables but also categorical variables. The importance of each variable could then be estimated by computing the average depth at which it appears across all trees in the forest.

In summary, Random Forests were selected for modeling because this approach can capture the possible nonlinear relationship, as well as provide the relative importance of features.

2.3.2 *Testing and validation*

The final data set was randomly split into the training data set and the test data set in a ratio of 7:3. The training data set was used to train the model and the test data set served as unseen examples to evaluate the model. In addition, out-of-bag (OOB) errors[5] from the training data set were also recorded, which provided an additional validation and guided in tuning the parameters of the model.

---

[4] Bootstrap sample
A bootstrap sample is a sample created from a single original sample by bootstrapping, a type of random resampling with replacement.

[5] Out-of-bag (OOB) error
Random Forest is a specific form of bagging method, in which some instances could be sampled several times for any given learner while others could not be sampled at all. These unsampled instances are called out-of-bag (OOB) instances. Since a learner never assesses these OOB instances, it can be evaluated by these OOB instances in the form of OOB error. In Random Forest, the OOB error is to average out the OOB evaluations of each learner.

2.3.3 *Performance evaluation*

In this regression task, MSE (mean squared error) was chosen as the measure of performance for the model. An ideal regressor is characterized by a small value of MSE. The reasons were addressed as follows. Firstly, fitting a statistical model usually delivers forecasts optimal under a quadratic loss function, especially when the density forecast from statistical modeling is symmetric. Since the outcome variable has been transformed to make it closer to a normal distribution, MSE is a reasonable choice for this prediction. Secondly, MSE tends to put a higher weight on large deviations. In other words, the MSE-optimal model will have fewer large errors than an MAE (mean absolute error)-optimal model, though it may have much more small errors than a MAD one. In terms of predicting LOS at presentation, especially in a logarithmic scale, having many small errors are more desirable than having some few larger errors.

In addition to calculating the associated MSE in the test data set, the predicted logarithmic values were also transformed back to the values indicating the predicted length in hours so that these numbers could be side by side compared with the observed LOS in hours.

Chapter 3. RESULTS

3.1 DESCRIPTIVE STATISTICS

3.1.1 *Patients' demographic data*

During the 8-month study period (Feb 1, 2016 to Sep 30, 2016), 47,807 consecutive medical patients visited the ED. 44,148 patients were included based on the previously described pretreatment procedures. As mentioned above, these 44,148 entries were randomly split into the training data set and the test data set. The demographic details regarding this final data set and its two subsets were shown in Table 3.1.

	The final set	**The training set**	**The test set**
Number of patients	44,148	30,903	13,245
Age† (years)	51.1(19.9)	51.0(19.9)	51.2(19.9)
Male proportion	45.4%	45.2%	45.7%
Proportion of sent by Ambulance	10.3%	10.2%	10.5%
Number by triage level			
1	1,586(3.59%)	1,120(3.62%)	466(3.52%)
2	7,845(17.8%)	5,444(17.6%)	2,401(18.1%)
3	29,174(66.1%)	20,462(66.2%)	8,712(65.8%)
4	5,485(12.4%)	3,842(12.4%)	1,643(12.4%)
5	58(0.13%)	35(0.11%)	23(0.17%)
EDLOS_hr_log†	1.62(1.78)	1.62(1.77)	1.63(1.79)

†Displayed in the form of "mean (standard deviation)".

Table 3.1. Descriptive demographic statistics on the final data set and the two subsets.

3.1.2 *Statistics on ED LOS*

Among these 44,148 observations, the median of ED LOS measured in hours is 2.30, with an interquartile range of 4.30. The first quartile and the third quartile are 1.40 and 5.70, respectively. However, the mean is 7.73. The histogram was depicted in Figure 3.1, which makes apparent that this is a highly right-skewed distribution.

After logarithmic transformation (base 2) was applied, the density distribution of "EDLOS_hr_log" was closer to a normal distribution (shown in Figure 3.2). The mean and the standard deviation of this new outcome variable are 1.62 and 1.78, respectively. The related statistics in each data set were shown in Table 3.1, too.

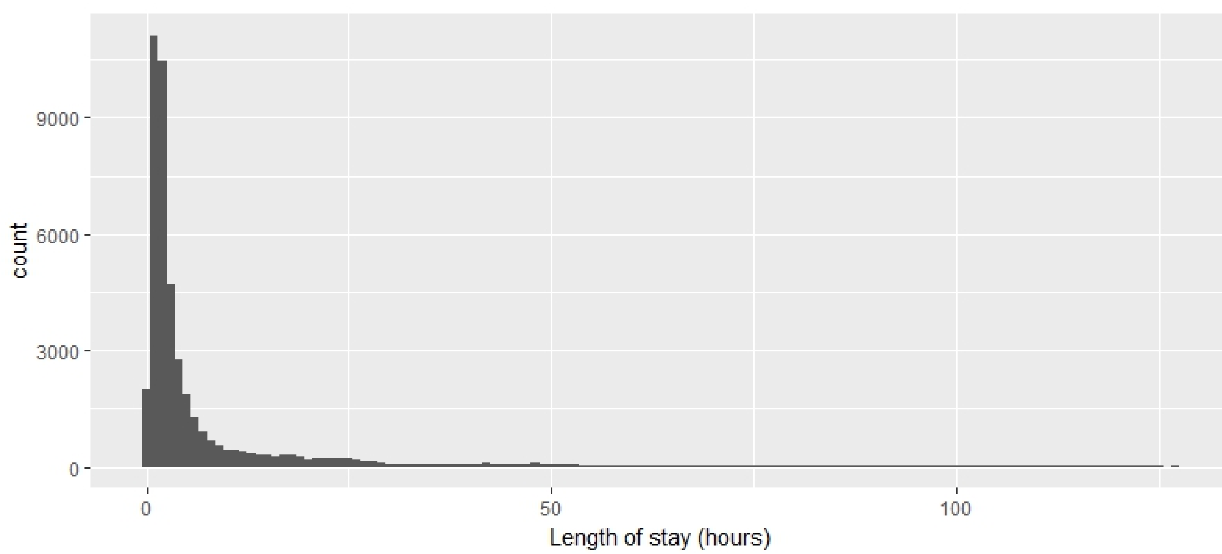

Figure 3.1. The histogram of medical patients' ED length of stay

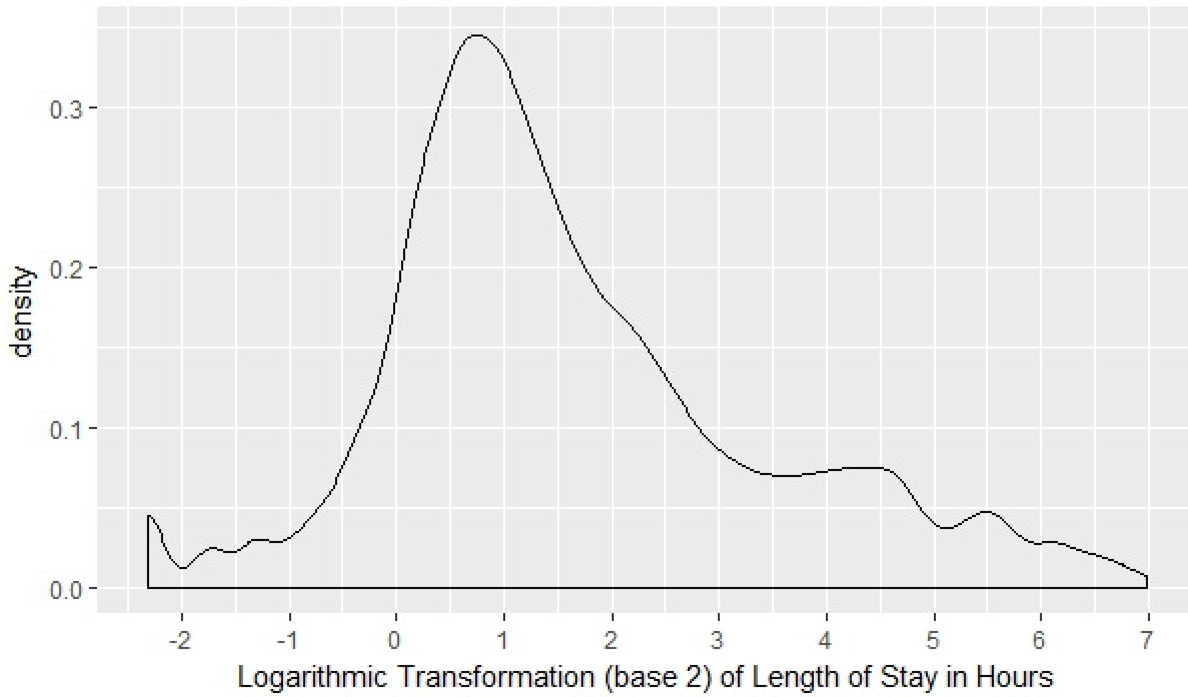

Figure 3.2. The density distribution of the new outcome variable, which is logarithmic transformation (base 2) of ED LOS.

3.2 Model Performance

3.2.1 *Model tuning*

The training data set was used to train the Random Forest model. Two parameters were required to tune the model so that the model wouldn't overfit or underfit the data. The model-specific OOB errors were used to find out the best tuning values for these 2 parameters.

The parameter "number of trees in the forest" (denoted as "ntree") was investigated first. Since Random Forest was an ensemble method, each tree was an individual weak learner. Data scientists need to decide how many trees are needed to achieve the prediction with low variance, that is with a more stable prediction. Figure 3.3 plotted the associated error of the model in different numbers of trees. From the figure, the decrease of the error achieved saturation after around 250 trees. Therefore, 400 was chosen as the value of the first parameter.

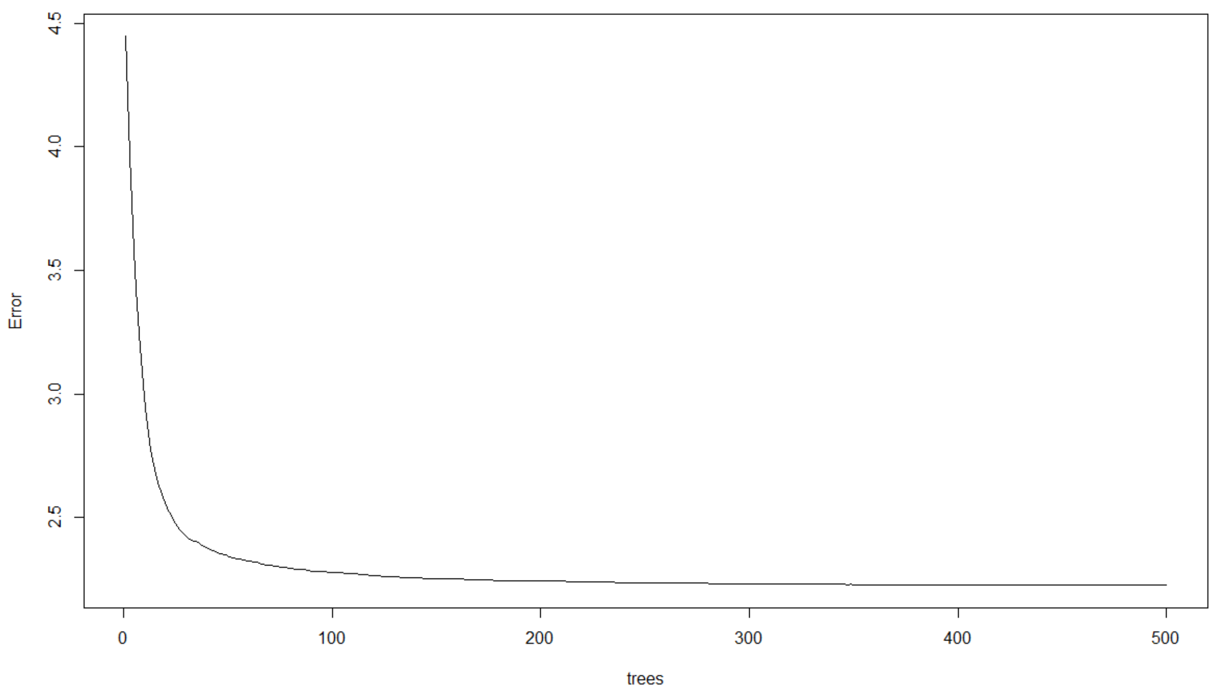

Figure 3.3. The OOB errors of the model by different numbers of trees.

In order to reduce the bias towards more powerful features, the Random Forest model only used a subset of features at each split within each tree. Therefore, the second parameter, "a number of features tried at each split" (denoted as "mtry"), should be assessed. Here the best value was determined with assistance of the model-specific OOB errors, depicted in Figure 3.4. The corresponding OOB error is smallest when mtry equals to 24. Therefore, 24 was selected to be the value of the second parameter.

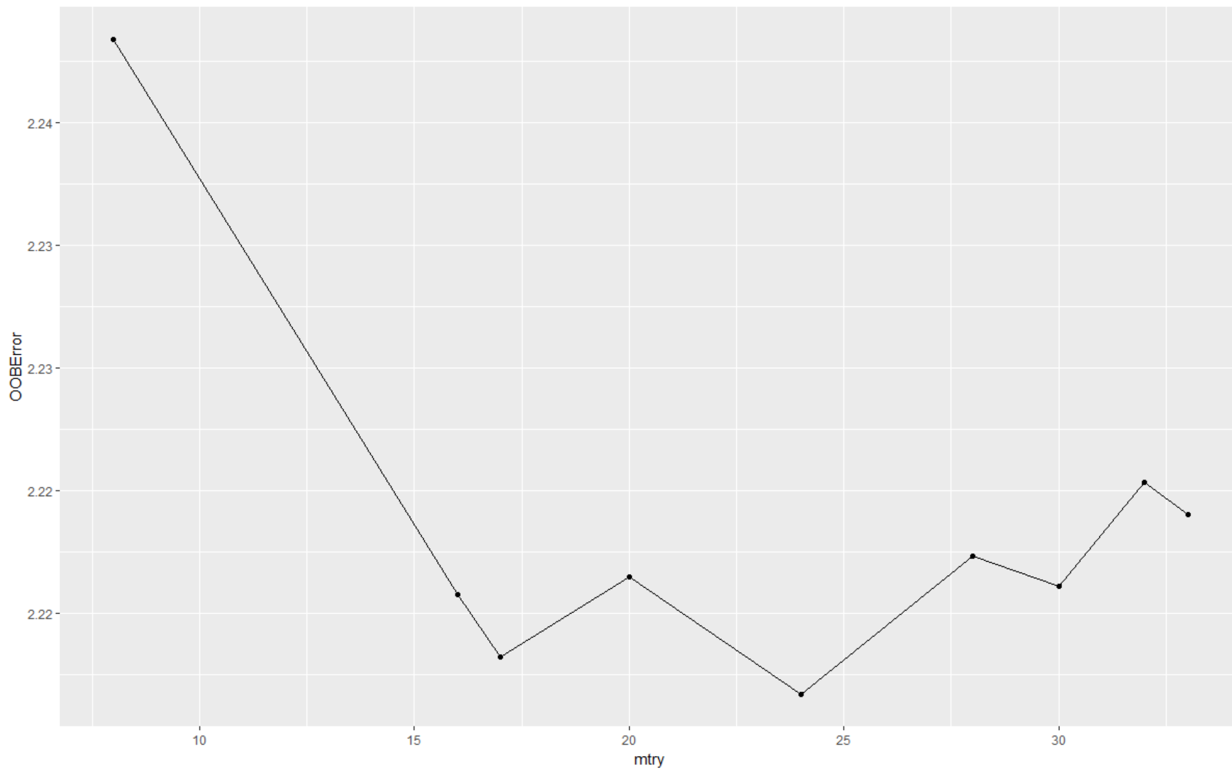

Figure 3.4. The corresponding OOB errors of the model in different values of “mtry”.

3.2.2 *OOB validation*

Since the values of the two parameters were decided, the corresponding Random Forest model was trained by the training data set. The OOB error in the form of MSE from this final model is 2.22.

3.2.3 *Performance on the test data set*

The test data set, preserved after the preceding split procedure, serves to evaluate the model. The resulting MSE in the test set is 2.26.[6] To better visualize the performance of the model output, the

---

[6] The results of the Random Forest models built on the imputed data sets.
As mentioned in section 2.3.3.4, the attempts of imputation were made in the two-step approach. The results were as follows.

scatterplot of the predicted values and the observed values was plotted in Figure 3.5. Given the error margin of ±1 (indicated as dashed lines in Figure 3.5), the model's prediction performed well when the observed values were between 0 and 3. However, the prediction was less precise when the observed values were at the extreme ends. While the observed values were large, the model tended to underestimate the expected length of stay. Meanwhile, when the observed values were small, the model was likely to overestimate the length of stay.

Besides of the comparison in the logarithmic scale, transforming the predicted log scale back to a normal scale (LOS measured in hours) would help to understand the model performance from a different view. The difference between the predicted LOS in hours and the observed LOS in hours, a.k.a. the residual, was calculated. The Residual was defined as follows:

$$Residual = Observed\ value - Predicted\ value$$

Figure 3.6 showed the residuals of the observations in the test data set. Although the maximum and the minimum of the residuals were 122.3 and -14.8 respectively, most of the observations were centered around 0, which meant the prediction was generally good. With an error interval of ± 4 hours (indicated as red vertical lines in Figure 3.6), the model correctly estimated 75.6% of the observations in the test set. If the error interval is strictly limited to ± 2 hours, only 59.5% of the

---

In the first step, 4 linear regression models were built respectively in term of imputing the missing values in 4 features regarding empty bed information. The performance metric, adjusted R-squared, was chosen to indicate the effectiveness of these 4 models. The values of adjusted R-squared for each model were 0.786 for "medbedC", 0.857 for "medbedNHI", 0.792 for "surbedC" and 0.808 for "surbedNHI", respectively. Therefore, 1,266 extra entries were imputed and incorporated into the final data set. The final 45,414-record data set was further randomly splitted into the training and the test data sets in a ratio of 7:3. The training data set with 31,789 records was used in training the model. The selected values of the tuning parameters were 400 for "ntree" and 16 for "mtry". The OOB estimate was 2.23. And MSE in the new test data set was also 2.23.

After the second step of imputing the missing values in 7 vital-sign related features, 986 extra records were incorporated into the previous 45,414-record data set. The final 46,400-record data set was split in the same, aforementioned fashion. The values of the tuning parameters were 400 for "ntree" and 17 for "mtry". The OOB estimate was 2.23 and the test set MSE was 2.28.

observations were correctly estimated. By the threshold of +4 hours, 19.1% of the observations were underestimated by the model. In this case, the predicted LOS was shorter than the observed one. Meanwhile, 5.3% of the observations were overestimated by the model at a cut point of -4 hours.

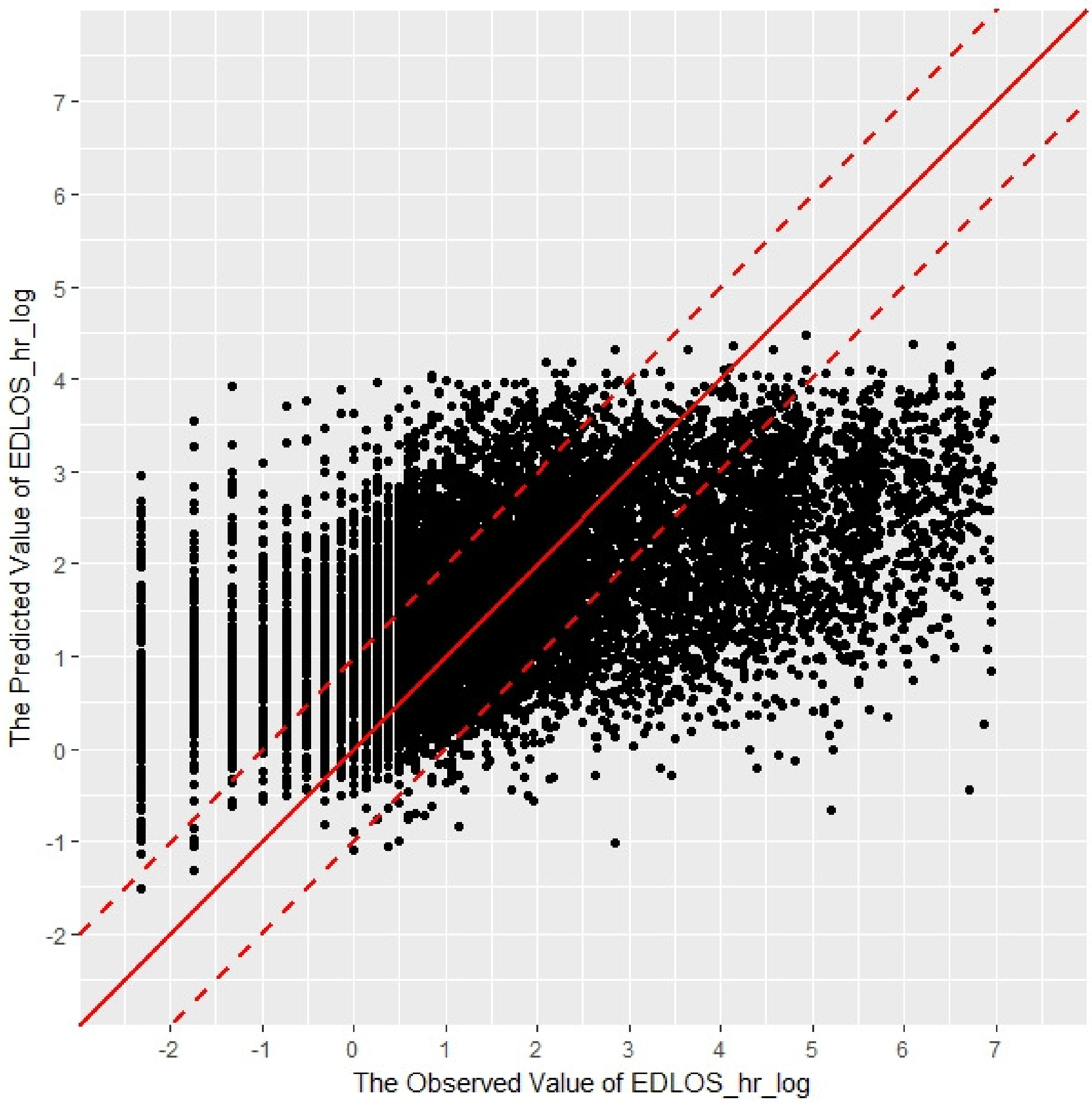

Figure 3.5. The scatterplot of the predicted values from the model and the observed values in the test set. The solid red line indicates the perfect prediction without any bias. The dashed red lines indicate the error margin of ±1.

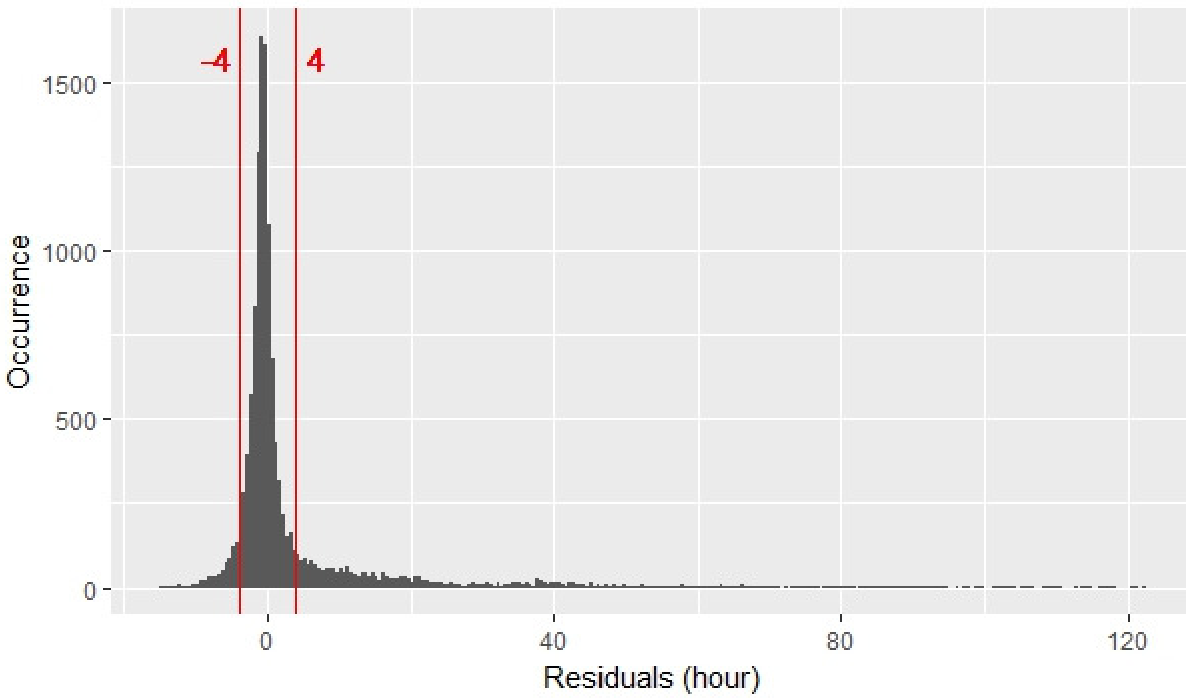

Figure 3.6. The difference between the observed EDLOS and the predicted EDLOS (both measured in hours), that is, the residuals in the test data set.

3.3 FEATURE IMPORTANCE

Figure 3.7 showed the importance plot output from the final model. The features were ordered by their computed values of "%IncMSE"[7]. This value could be regarded as the degree of the resulted MSE increase if values by random permutation were assigned to this feature. Therefore, the higher the "%IncMSE" is, the more important the feature is. The top 5 important features were age, triage,

[7] The measure "%IncMSE"
According to the document of 'randomForest' R package [74], the measure, "%IncMSE" was computed from permuting OOB data as follows: "For each tree, the prediction error on the out-of-bag portion of the data is recorded (error rate for classification, MSE for regression). Then the same is done after permuting each predictor variable. The difference between the two are then averaged over all trees, and normalized by the standard deviation of the differences."

mis5 (indicating "transferred patient"), hrtillnextbusihour_sqrt, and GI10 (indicating chief complaints with gastrointestinal tract bleeding). Furthermore, by feature category, the top 15 important features included 3 of 4 demographic features, 2 of 34 chief-complaint related features, 4 of 7 vital-sign related features, and 6 of 63 registration status features, which were depicted in Figure 3.8.

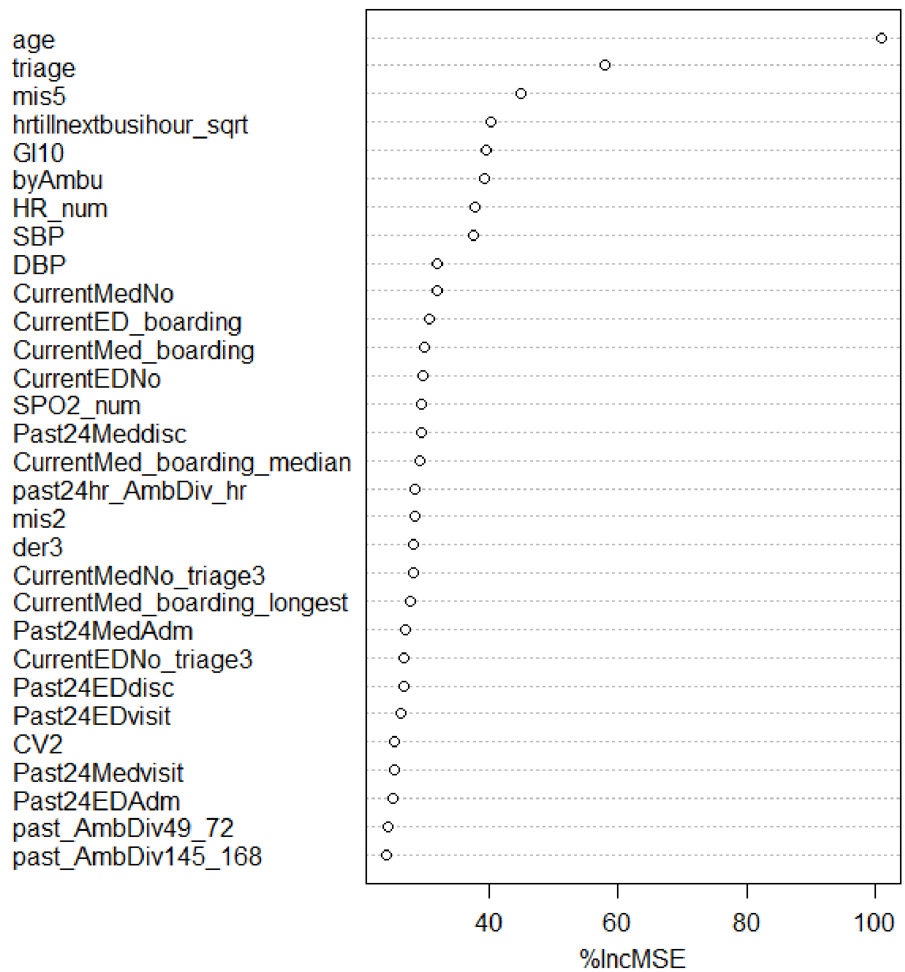

Figure 3.7. The feature importance plot, in which only the top 30 are included.

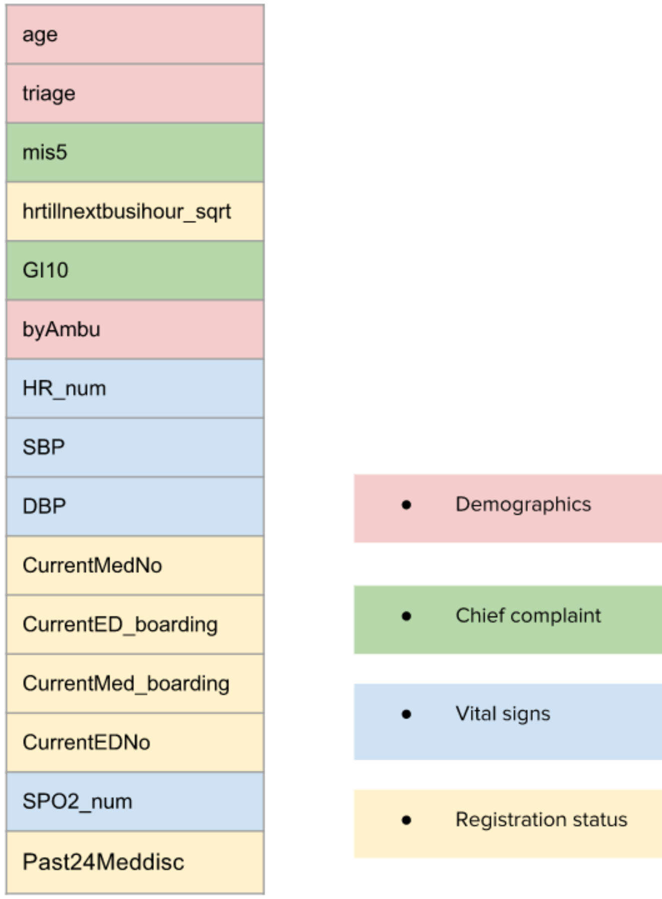

Figure 3.8. The infography of the top 15 important features and their categories. The order of this stack represents the ranking of each predictor, which means the higher it is, the more influential it is.

Chapter 4. DISCUSSION

4.1 MODEL PERFORMANCE

4.1.1 *Individual predictions*

4.1.1.1 Bias

The chosen metric of the final model (MSE) were 2.22 and 2.26 on the training data set and the test data set respectively. In machine learning, these numbers signal the model is high-bias and low-variance. Low variance is a good sign, which means the generalization ability is good. However, a high-bias model is unwelcome because it implies the predictions may be inaccurate. But, the magnitude of the bias is dependent on comparison with the currently available models. In the literature review, no such successful regression model has been created. Then how can one evaluate this model?

In machine learning, using a dummy regressor as a baseline model is often used to measure how much a model can improve the predictions. Here the dummy regressor adopted was using the mean of the outcome variable in the training data set, which was 1.62, as the predicted value for any future cases. The resulting MSE on the test set was 3.19. Therefore, the proposed model improves the predictions by 29.2%.

Besides, the MSE on the test set, 2.26, represented the variance of the prediction errors from the model. Since the distribution of the logarithmic outcome variable was close to a bell-shaped normal distribution, this number can be further conceptually interpreted as about 68% of the model's predictions fell within 1.50 from the actual value, and about 95% of the predictions fell within 3.01 from the actual value. Figure 3.5 further provided the information that the model predicted well when the observed values were between 0 and 3, which represented 1 and 8 of LOS measured in hours.

Unsurprisingly, the predictions were not bad when transformed back to the original outcome of interest, which is “EDLOS_hr”, as shown in Figure 3.6. Given the error range of ±4 hours (slightly larger than $2^{1.5}$), the model successfully predicted 75.6% of the cases. If we categorized the data points of the test data set into 3 groups based on their observed LOS (namely, less than 1 hour, 1 to 8 hours, and longer than 8 hours), and further colored the histogram in Figure 3.6, according to this categorization, it is easy to see that the majority of these successfully predicted cases is the group of the true LOS falling between 1 hour and 8 hours (as shown in Figure 4.1).

4.1.1.2 Sources of bias

On the other hand, focusing on where the biases originate may provide more insight into the model. In our case, most large errors made by the model came from predictions on those extreme values of LOS. The model tended to underestimate when the actual LOS was high. One reasonable explanation, and as stated in the previous section, was the model only took into consideration all the features available right before and right at presentation, which inevitably introduced the biases. A lot of oncoming factors that happened after registration could substantially influence LOS, including lab test, image study, consultation, admission, etc. This statement was further supported by the fact that the prediction model, derived by the French team in a pediatric ED [69], comprised only 4 examination predictors. Even in their second, more sophisticated model, only 2 out of 8 total predictors were the information available at registration.

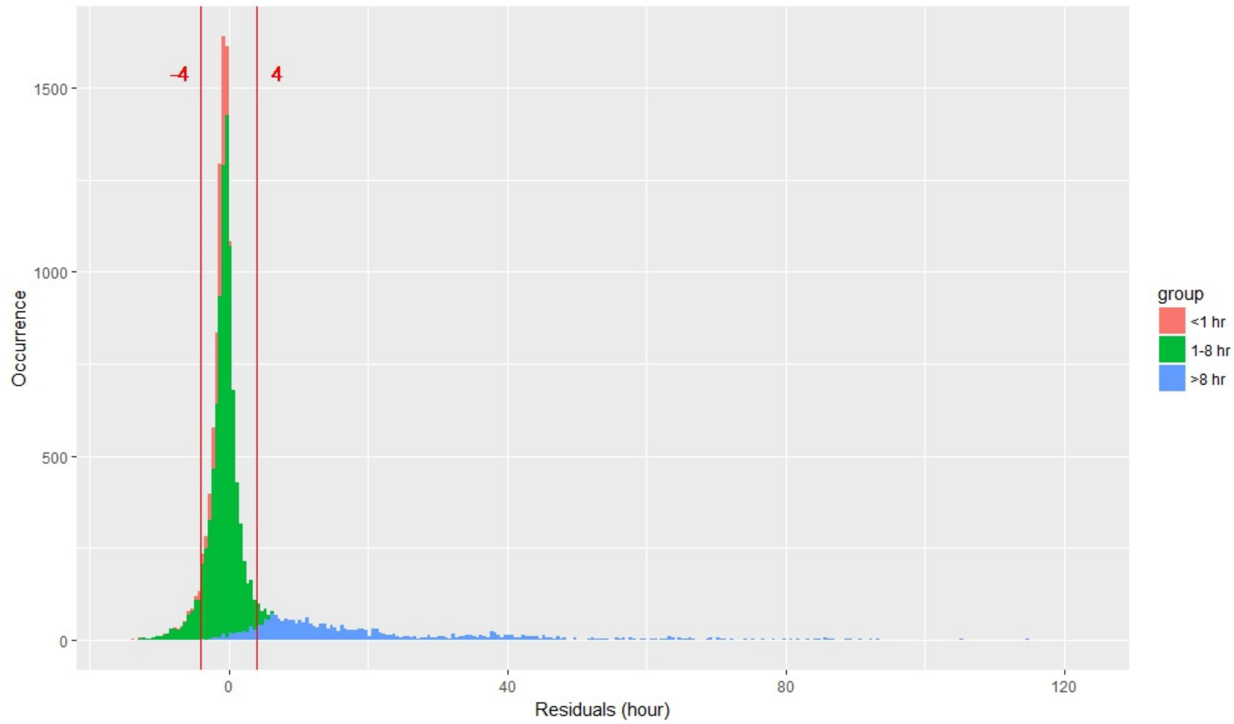

Figure 4.1. The colored histogram of Figure 3.6, which plots the residuals measured in hours in the test data set. Each color represents one of the 3 groups categorized by the observed LOS.

As for the overestimation on those extremely short LOS (mostly less than 1 hour), apparently some factors that could explain this phenomenon were missing. For these ultra-short-stay patients, most of them either received a plain X-ray study or medication such as an injection, or both. Therefore, the missing but critical piece of information able to answer this question would be a provider's evaluation right after assessment, which could be, for example, a tentative diagnosis by a physician, or secondary evaluation performed by a nurse. Again, this information couldn't be attained at registration.

Another source of bias is from intra-physician and inter-physician variability regarding either discharge or admission. Although most physicians have similar criteria when making this decision with a patient, there is a gray zone between these two decisions. In addition to medical issues, the patient's willingness and other socioeconomic factors must be considered. All these factors contribute to the physician's clinical judgement, which depends heavily on the physician's experience. And this clinical judgement might be further influenced by the current ED crowdedness perceived by the physician. Thus, given the same patient, a physician might suggest admission this time but longer observation or even other studies next time. Similarly, doctors might have different strategies even when they deal with the same patient.

In summary, our study demonstrated the prediction of EDLOS at presentation was achievable but imperfect. The model's general prediction was not numerically precise, mainly because of the inherent bias from unavailable information. It tended to overestimate when the true LOS was less than 1 hours, and underestimate when the true LOS was over 8 hours. However, when the true LOS fell between 1 hour and 8 hours, the model's performance was good enough to enable the foreseen picture of the ED in several hours later.

4.1.2 *Pattern recognition*

After exploring how the model performed in individual cases, let's look at the model performance in the whole data set. Despite some predictions made by the model in the test data set were quite inaccurate individually, the model in fact was able to capture the pattern of ED load, which further helped in human resource management. Here a moving average approach, an analytic tool that smoothed out irregularities to recognize trends, was applied to the whole final data set, which comprised 44,148 records. The detailed procedures were described as follows.

1) Firstly, each patient's predicted value (EDLOS_hr_log) was calculated from the proposed Random Forest model. To enable comparison, a simpler linear regression model which has only one and the most powerful one predictor "age" was created and applied to predict each patient's value of EDLOS_hr_log.

2) Secondly, with these two predicted vectors plus the original, observed EDLOS_hr_log vector, a window size of 48 hours was adopted to calculate each mean. For example, at the timestamp of Feb 8 00:00 am, I collected the observed values of EDLOS_hr_log for those who visited the ED from Feb 6 00:01 am to Feb 8 00:00 am and took the mean of these values. By the same token, the other 2 means were obtained from the two predicted vectors (Random Forest and linear regression) respectively.

3) Thirdly, a step of 30 minutes was applied, and calculation was repeated to get the next 3 means. Using the previously mentioned example, at the timestamp of Feb 8 00:30 am, the observed mean was computed by summing up the observed values of EDLOS_hr_log for those who arrived in the ED between Feb 6 00:31 am and Feb 8 00:30 am followed by taking the average, and so were the other 2 predicted means.

4) Repeat procedure 3 until the end of the study period, which was Sep 30 11:30 pm.

As a result, the calculated moving averages of these 3 vectors against the date were plotted in the same graph, as Figure 4.2. The orange line indicated the moving averages from the observed data. The blue line presented the moving averages from the Random Forest model. And the black line was the moving averages from the simpler, linear regression model that contained only "age" predictor. From the graph, one can tell the blue line, rather than the black line, was able to capture the peaks and valleys of the orange line, which meant the proposed model can detect the trends in the real world.

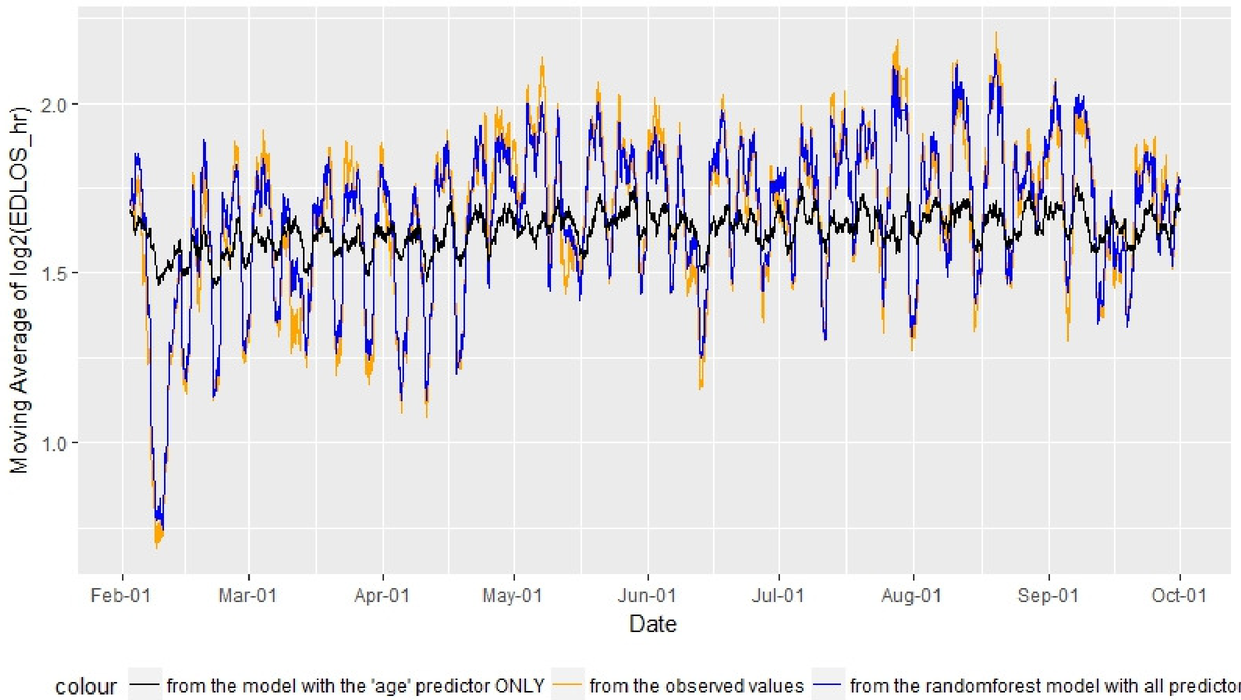

Figure 4.2. The moving averages calculated from the observed data and the two models against the date across the whole study period.

Now let's focus on the month of February, as in Figure 4.3. Here one can observe that the blue line catches up with the orange line fairly well. Furthermore, there are some valleys along the course. The three valleys on the right have something in common, that is, they all start with Monday and last for around 2 days. The left one, which is large, indicates the Chinese New Year, from Feb 7 (the Chinese New Year's Eve) to Feb 11 (Jan 4 of Chinese Lunar Calendar). By tradition, Taiwanese people tend to not to visit hospitals these days, which results in such a big and wide trough.

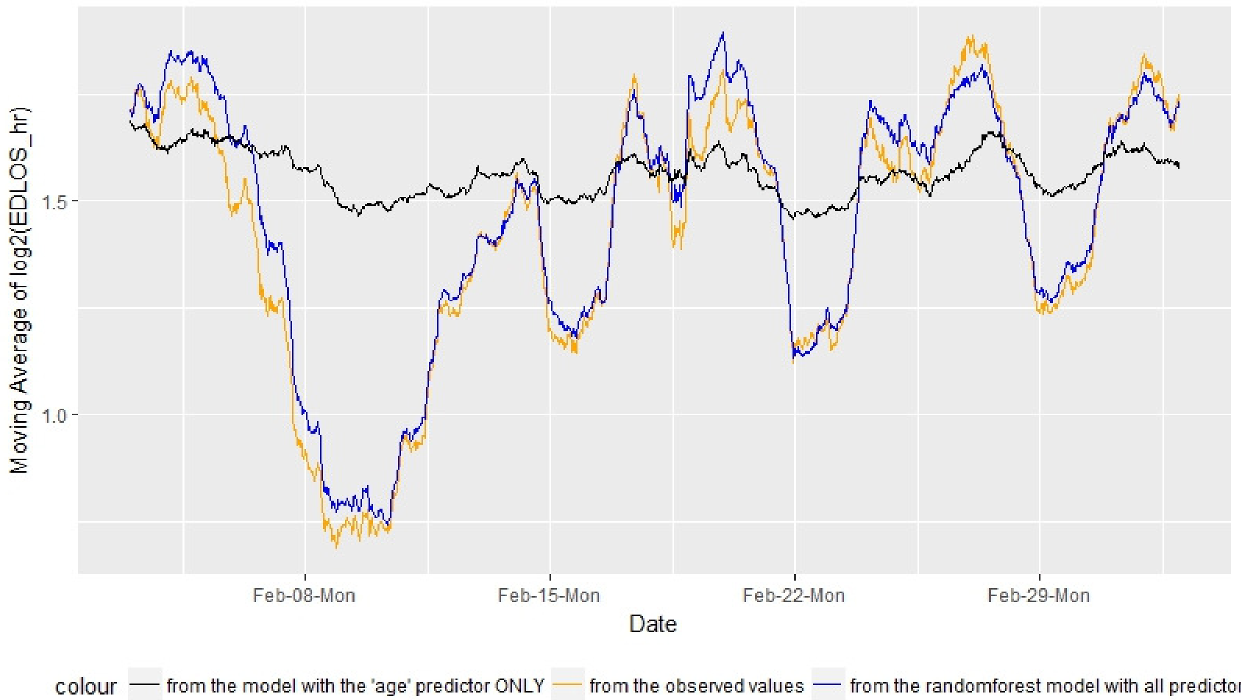

Figure 4.3. The moving averages calculated from the observed data and the two models against the date in February 2016.

In summary, the model can capture the patterns very well. This feature can benefit in the future manpower allocation.

4.2 FEATURE IMPORTANCE

One of the advantages provided by the Random Forest algorithm is that it helps modelers recognize the importance of each feature. Even though the algorithm doesn't provide a sign for a predictor, which means one cannot tell how the outcome variable changes by the predictor, it is still clinically insightful to look out what the algorithm suggests. The reasons are that all these features have their own clinical meaning and some of them are even already adopted by organizations to illustrate an ED's status.

Unsurprisingly, the two demographic features, age and triage, remain the first and the second places in rank of importance. It is easy to understand that age plays the most significant role in a patient's length of stay. With increased age, and consequently increased complexity of diseases, a doctor must consider more subtle but potential serious etiologies, arrange more tests, and sometimes need more time to observe the progress of a disease in order to recognize it. Therefore, the ED LOS will be affected by age. Triage, as the first indicator to allocate ED resource, surely should be able to reflect a patient's severity, which is related to his LOS. Another demographic feature, "byAmbu", was also selected by the model as the 6th most powerful predictor. This feature described whether a patient came to the ED by ambulance or not, which implied transported patients might stay longer because of severe symptoms. In Taiwan, such ambulance service is provided free mainly by the government, which introduces potential for misuse. However, even considering this, "byAmbu" is still powerful because it would introduce 39.4% of increase of the MSE if permuted.

The 3rd and the 5th most important variables are in the chief-complaint category, namely "transfer" and "gastrointestinal bleeding". Transferred patients are often from other hospitals or clinics, where they have been previously diagnosed and managed. Their treatment plans are usually much clearer, such as ward/ICU admission, other advanced image study required, or even surgical operation arrangement. Therefore, compared with the so-called "fresh" patients who need a thorough evaluation from the beginning, these patients' LOS is shorter given the same age and the same diagnosis. As for patients with “gastrointestinal bleeding”, usually they have to stay for a while, either for observation or lab testing. If there is any signal indicating acute blood loss, further diagnostic procedures such as panendoscopy or treatment interventions will be arranged depending on their specific treatment plans. Therefore, the LOS of patients with this feature will increase.

In addition to these 2 features, 3 more chief-complaint predictors are present among the top 30, namely “fever/chills” (mis2, ranked 18), “rash” (der3, ranked 19), and “chest pain” (CV2, ranked 26). For patients with fever/chills or chest pain, usually further studies are required, especially for high-risk groups. Interestingly, it is not clear how the LOS of patients would be affected by the symptom “rash”. If a patient’s chief complaint is solely “urticaria”, usually the patient can be discharged after an injection. However, if the rash is more bizarre or combined with other constitutional symptoms such as fever, malaise, and jaundice, it will take more time to obtain lab testing and perhaps dermatological consultation. In this case, the chief complaint “rash” might imply either shorter or longer LOS depending on the situation, which doesn’t behave like other mentioned chief complaints.

Now let’s move to the fourth place in the list, which is surprisingly “hrtillnextbusihour_sqrt”. To my knowledge, this feature has not been previously mentioned in the literature. The rationale behind it is associated with the uniqueness of Emergency Medicine. Although EDs provide thorough and generalized care in a 24/7 fashion, some services rely on the support from other specialties. For example, if a patient needs panendoscope either to identify a bleeding source in the upper gastrointestinal tract or to stop bleeding, a gastroenterologist would be consulted and then such a procedure could be arranged. To perform panendoscopy, not only a gastroenterologist but also medical assistants are required. In Taiwan, this procedure would be performed in a specific room outside the ED, called “endoscopy room” dedicated only to this kind of patient. Therefore, except for emergency conditions, these procedures would be held until normal business hours of a hospital. Given the same two patients with other variables controlled, but one arrives in an ED in Sunday evening, and the other arrives in Monday morning, these two patients would have different LOS for this reason. And there should be a variable to be able to describe the difference between these two patients. And the algorithm picked it up as the fourth powerful predictor in predicting ED LOS because the MSE would increase by 40.3% if one permutes this variable.

Following the above features, the next important predictors are so-called vital-sign features, which are a patient’s heart rate (“HR_num”, ranked 7), systolic blood pressure (“SBP”, ranked 8), diastolic blood pressure (“DBP”, ranked 9), and saturation (“SPO2_num”, ranked 14). This makes sense because it reflects a patient’s acuity at presentation. Surprisingly, the other 3 vital-sign features, namely respiratory rate, coma scale, and pain severity score, were not selected into the top 30 list. Even though permuting each of these 3 features still increase the MSE by 23.2%, 22.5%, and 18.4%, respectively. The fact emphasizes again a patient’s vital signs have great influence on his LOS, and the model provides the orders, a.k.a influential magnitude, of these vital signs.

Next comes the group of registration status features. Besides the most powerful registration status predictor, “hrtillnextbusihour_sqrt”, discussed previously, the other predictors in the top 15 list are all related to the numbers of patients, including the number of current medical patients in the ED (“CurrentMedNo”, ranked 10), the number of current all boarding patients in the ED (“CurrentED_boarding”, ranked 11), the number of current medical boarding patients

("CurrentMed_boarding", ranked 12), the number of all current patients in the ED ("CurrentEDNo", ranked 13), and the number of discharged medical patients in the past 24 hours ("past24Meddisc", ranked 15). Furthermore, not only the features related to numbers of patients but also the features regarding their actual boarding time were identified by the model as the powerful predictors. Respectively, the median and the maximum of the boarding time of current medical boarding patients served as the 16th and 21st powerful predictors in the model. This fact supports that both factors (numbers of patients and EDLOS of current boarding patients) affect LOS of an oncoming patient.

Interestingly, the feature describing ambulance diversion status at presentation, "AmbDivStatus" was not in the top 30 list. In fact, permuting this variable only increased 9.38% MSE. Therefore, when a patient came to the ED, whether the ED was on ambulance diversion or not didn't affect this patient's LOS very much. Instead, the patient's LOS was predicted by the cumulative ambulance diversion hours in the past 24 hours ("past24hr_AmbDiv_hr", ranked 17 with 28.6% of increased MSE if permuted). The more hours in the past 24 hours that the ED calls for ambulance diversion, the more crowded the ED is. Surprisingly, the number of hours for which the ED was on ambulance diversion between 48 hours ago and 72 hours ago ("past_AmbDiv_49_72", ranked 29) and the number of ambulance diversion hours between 144 hours ago and 168 hours ago ("past_AmbDiv_144_168", ranked 30) have some effects on an oncoming patient's LOS. This implies the ED crowdedness 3 days ago and 7 days ago both play a role here. Exploring the mechanisms behind this fact is interesting. Here is a proposed mechanism. The tsunami of patients arrives in the ED first, which is related to the feature "past24hr_AmbDiv_hr". It takes some time for the ED to "digest" these patients in addition to accepting newly arrived patients. About 3 days later, this tsunami has been passed into the admission wards and the ICUs so that crowdedness happens there. Again, for most admitted patients, it takes 7 days to be discharged from wards or transferred from ICUs. This phenomenon surely requires future study, especially the data created after ED services, to verify its accountability.

Surprisingly, none of the medical subunit staffing factors appeared in top 30 list. Among them, the feature, “RN”, scored the highest ranking (68), which resulted in 12.5% of increased MSE if permutated. The next one is “RN_extra”, which is in the 71st place. Why is this? To answer this question, we should examine these features more deeply. In the feature “RN”, there are only two values, 11 and 12 respectively. That means when a patient come to the ED, the number of nurses is either 11 or 12. Other staffing features have a common presentation, which is low variability. However, feature selection based on Random Forest algorithms is biased toward preferring variables with more categories or bigger range, which made the staffing factors less favorable. But that doesn’t mean these staffing factors are not influential. It just represents in our data set, the change between the numbers of nurses is too small to explain the variety of patients’ EDLOS.

4.3 GENERALIZABILITY

Generalizability, which refers to the external validity of predictions from a model, depends on the quality of the prediction model as developed for the development setting (internal validity), and on characteristics of the population where the model is applied [75]. The former component, the quality of this model, has been discussed in the previous sections. In addition, some characteristics of features must be considered in terms of generalizability, such as a clinical meaning of a certain feature and the easiness to access this feature. In fact, according to Appendix B, all the features are clinically meaningful despite some need efforts on feature engineering. These two characteristics are the advantages of this model.

Furthermore, model complexity is another critical dimension that must be addressed. Currently, the model takes over 100 features to make the predictions, which makes it hard to deploy because other sites may not have such rich data. In fact, Figure 3.7 could be regarded as feature selection performed by Random Forest algorithm. Based on their rankings, some features can be selected to create a new model. And one can still evaluate this newly-created model by its performance on unseen samples, which is the test data set. Figure 4.4 showed the relationship between the model complexity (the number of features selected into a new model) and the corresponding model performance (measured by MSE on the test data set). Here the models with the first 1 and the first 2 features were created by simple linear regression method. And for other models with more features, ridge regression was adopted. Finally, the results from the dummy regressor model (discussed in section 4.1.1.1) and the original Random Forest model were also plotted for comparison. The model with only 1 feature, "age", scored 2.83 of MSE[8]. By adding the second powerful feature, "triage", the resulting MSE further dropped to 2.68. From the plot, perhaps the model with 8 features has the best balance of complexity and performance.

[8] This model was the one used in comparison when the moving averages were calculated. (See section 4.1.2)

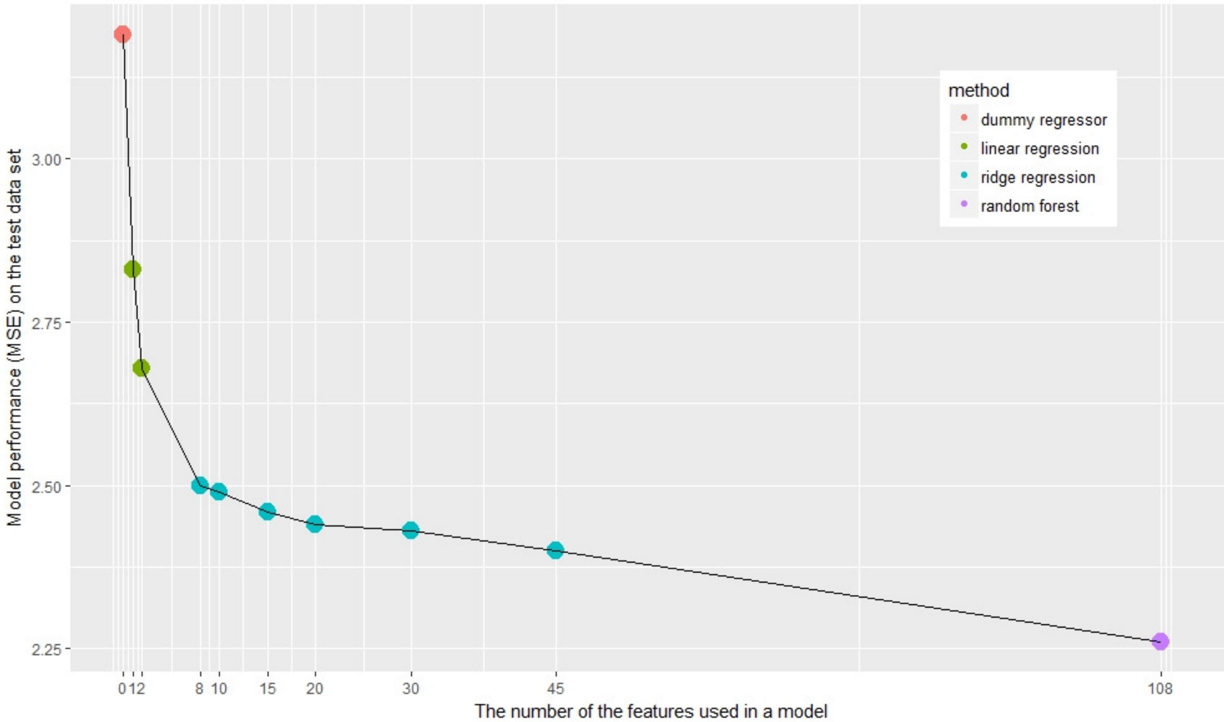

Figure 4.4. The relationship between the model complexity and the model performance.

On the other hand, the latter piece, which is the characteristics of the population where the model is applied, could be addressed in two different levels. The first level is aiming at the study hospital, FEMH. Since the study period is 8 months from February to September in 2016, the variation resulting from seasonal factors hasn't been considered. For example, there is usually higher incidence of cardiovascular diseases and cerebrovascular accidents in winter, which might consume more ICU and ward resources. Furthermore, as in other countries, Taiwan enters the peak season for flu in winter, resulting in higher burden in medical care. All these factors might contribute to different case-mix or age-mix when evaluating the model. Therefore, in terms of case-mix representation, the more data we have, the better the model is. However, it doesn't mean the longer study period would certainly result in a better model. There are some initiatives aiming at shortening EDLOS as discussed in section 1.4. For instance, FEMH has adopted several measures in 2017 to solve this problem, including assigning few wards as ED short stay wards and shortening the boarding time by accelerating the administrative process of admission. Although it is too early to conclude whether these are effective or not, these measures have already affected some patients, and thus their EDLOS. When generalizing the model, the ED, and even the whole hospital, must be treated as an evolving organism to minimize the bias.

The second level is targeting the tertiary care hospitals in Taiwan. Undoubtedly, case-mix would be the first challenge when generalizing our model. The case-mix in a specific ED would rely on the interaction among several factors, including the local population, patients' medical care-seeking behavior, and the locally available healthcare resources. Therefore, each hospital would have different distributions of case-mix. However, in contrast to the USA, most tertiary care hospitals in Taiwan, which are called medical centers, have shared very similar features because of Taiwan NHI[9]. For example, a tertiary care hospital, along with a level I trauma center and a teaching hospital itself, is commonplace in Taiwan. Therefore, FEMH ED could be regarded as a nutshell of all EDs of 19 medical centers in Taiwan. Furthermore, the interaction between an ED and its related hospital, and the cooperation between ED and other specialties, are similar among these hospitals. These homogeneous factors benefit in our model generalization.

[9] Taiwan NHI is a governmental single-payer system, characterized by its high coverage rate (around 99.8% in 2016). It has service contracts with 93% of Taiwan hospitals and clinics [76].

4.4 LIMITATIONS

There are several limitations in this study. The first is the study duration and the methodology. The study duration was 8 months rather than a common practice of 2 years (the data from the first year serves as a training set, and the one from the second year as a test set). The way to split the training data and the test data was by random rather than by consecutivity as described above. This fact might limit its generalizability to the ED of interest, FEMH ED. However, if I used the data from February to July 2016 as a training set and the data from August to September 2016 as a test set, seasonal factors might come to play and introduce bias. Considering this, splitting randomly was a reasonable approach.

The second limitation results from absence of available medical record numbers. Despite suspected high utilizers had been censored, some behaviors that might affect LOS could not be identified without medical record numbers. For example, a single patient could visit the ED twice or more in a very short time such as 3 days. The revisit reasons could be the same, such as unresolved problems or a newly developed symptom. In the former case, the LOS of this visit would be affected because some studies were done last time and physicians were already familiar with his condition. According to the statistics in the study ED during the study period, the numbers of this kinds of medical patients, no matter what revisit reasons are, are ranged from 239 to 297 monthly. The total number is up to 2122, roughly 4.8% of the final data set. This fact would introduce bias into the model.

Last but not least, certain features are related specifically to the ED of interest, FEMH ED. And this ED is a product of the interaction between the hospital, FEMH, and the associated healthcare ecosystem, which is Taiwan National Health Insurance System, a single payer system. Therefore, when it comes to generalizability, each feature should be examined in the context of the interplay of an ED, an associated hospital, and an associated healthcare system. Even so, the model sheds light on the interaction between a patient's ED LOS and those descriptive features which are from either the literature review, or the currently utilized indicators of UW Medicine, or my personal experience.

4.5 FUTURE IMPROVEMENT

4.5.1 *Improving bias*

4.5.1.1 Granularity of features

Since the model suffered from the bias problem rather than the variance problem, the way to improve it could be addressed in the two following ways. Firstly, better feature engineering should be attempted. Just like the example of features with ambulance diversion hours, one never knows how far to trace back until creating and testing them. Perhaps the granularity of features regarding the number of patients should be adjusted. Since most patients' LOS were around 2 hours, is it possible that a feature describing the total number of patients in the past 2 hours or 4 hours rather than the past 24 hours used in the study better predicts the outcome of interest?

4.5.1.2 Other available information at presentation

Besides of intra-physician and inter-physician variability discussed in section 4.1.2, some available information at presentation was not extracted into the model.

As mentioned before, the skill of text mining in this study for chief complaint data was at a nonexpert level, which only extracted the associated symptoms. The information about how long the symptom lasts and the occurring order of these symptoms remain in the data set. Because clinicians rely on these data to group patients and diagnose correctly, the predictions should be much improved if incorporating this information.

Besides, only some of these patients are "fresh" patients, which means this is their first time to visit the hospital. For other patients, before this visit to the ED, they might have been seen by physicians in outpatient services, or even been cared in inpatient services. These past medical records could provide clinicians a sense of a patient's health status. If these parts of information could be summarized and extracted, we could know more about this specific patient, even right at presentation.

Finally, a triage nurse's experience should be fully leveraged. Triage nurses are very experienced because of working in the ED for so many years. From a group of patients, they can easily recognize the simple but ominous signs (or conversely, reassure signs) of a single patient and assign him a proper triage level. For those not sent in by ambulance, triage nurses are also the first medical professionals to assess the patient. I believe their judgements would reliably contribute useful information about a patient. However, how to extract their judgements effectively and efficiently in a timely fashion remains a challenge because the basic concept of triage is not to spend too much time on the same patient.

4.5.1.3 Other algorithms

Searching other more powerful algorithms might be another path. That is, find out another more expressive model, such as artificial neural networks. With more data fed, neural networks could achieve better performance by its complexity of architecture, namely the hidden layers and the hidden units. One challenge would be how to take into consideration all the possible features including not only numerical but also categorical ones. The key relies on machine learning experts who have a solid grasp of knowledge about ED overcrowding.

4.5.2 *Model reconsideration*

From the previous paragraphs, we know that some features to address ED LOS are missing in this model, especially those occurring after presentation. To incorporate these factors and improve our accuracy, perhaps the ultimate model is making predictions on an ongoing basis. That is, when a patient shows up at triage, the model gives its first prediction. After 10 or 20 minutes, the model predicts again based on the tentative diagnosis made by a physician. And 1 or 2 hours later, the model predicts again based on the newly available data such as lab results, X-ray findings, etc. In this way, the model continues to update its prediction by newly fed data. To realize this idea, substantial efforts must be made in machine learning engineering because of the variety, the volume and the velocity of data.

4.5.3 *Rephrasing the question*

This study presented a regression task. As a result, the model performed less favorably when the observed LOS was either extremely short or extremely long. In fact, predicting such an ultra-short LOS was not clinically useful or beneficial to ED overcrowding. These patients, if many at a time, might result in a transient surge on ED load, but they are not the main reason of ED overcrowding. Moreover, to measure ED load, overestimation is preferable than underestimation. The major cost resulted from overestimation would be increased times of false alarms. In this circumstance, ED and the hospital are well prepared. And to correct these overestimations is not hard. When a patient is discharged from an ED, one can simply drop this discharged patient off from the system so that the effect of overestimation could be timely corrected. However, the cost that comes from underestimation could be very high. The ED would not be prepared, and the activation of mitigating measures would thus be slow. Overall, ED efficiency decreases, as well as patient safety is not guaranteed.

By the same token, in terms of illustrating the ED crowdedness several hours later, knowing they are going to stay short or long might be enough rather than knowing the exact LOS in these extreme groups. Thus, we could transform the original regression question into a new classification-regression problem. For example, for each patient at presentation, first we predict which group of LOS this patient will be, less than 1 hour, between 1 hour and 8 hours, or over 8 hours. If the patient is predicted to stay less than 1 hour, the predicted value of his LOS will be 1 hour by default (due to preferable overestimation). If the patient is predicted to stay between 1 hour and 8 hours, our model comes to play to predict his LOS. If the patient is predicted to stay over 8 hours, by default his LOS is set to 8 hours in the beginning. Afterwards, this longer-stayed patient's LOS will be updated regularly by the model described in 4.4.2, which captures the newest available data and gives estimation more accurately. In this way, the ED loading for the next few hours could be plotted by summing up these patient's LOS so that stakeholders could proactively act accordingly.

Chapter 5. CONCLUSION

The intersection of machine learning and medicine is emerging and evolving. Not only patients but also providers look forward to its potential. However, we should bear in mind that even with breakthroughs in this field, real impacts on patients and providers have seldom been demonstrated [77]. Clifton et al [78] investigated how health informatics systems based on machine learning methods had impacted the clinical management of patients, and concluded it was still in its infancy because there were few examples existed that affected the clinical management and changed the prognosis. They stated the one main reason is coming from the lack of multidisciplinary teams drawing on both clinical expertise and practitioners from the information sciences.

This study is an excellent example to illustrate the interaction. Despite several drawbacks, the model should still be regarded as a prototype of such models, serving as a future guide to feature engineering and a baseline model to enable the comparison. By providing better representation, a sounder model could be created from machine learning algorithms to predict and facilitate clinical management. We found solid evidence that ED LOS was predictive at presentation. In addition, the ranking of commonly mentioned features in literature or organizations is provided in this study to inform the future research. More importantly, it implies that strategies to tackle ED overcrowding and reduce ED LOS should not be limited in an ED itself. Rather, a higher-view strategy designed for a whole hospital and even a whole healthcare ecosystem should be implemented to attenuate the consequences of overcrowding and guarantee patient safety.

BIBLIOGRAPHY

1. Trzeciak, S. and E.P. Rivers, *Emergency department overcrowding in the United States: an emerging threat to patient safety and public health.* Emerg Med J, 2003. **20**(5): p. 402-5.
2. Ovens, H., *ED overcrowding: the Ontario approach.* Acad Emerg Med, 2011. **18**(12): p. 1242-5.
3. Jones, P.G. and S. Olsen, *Point prevalence of access block and overcrowding in New Zealand emergency departments in 2010 and their relationship to the 'Shorter Stays in ED' target.* Emerg Med Australas, 2011. **23**(5): p. 587-92.
4. Morris, Z.S., et al., *Emergency department crowding: towards an agenda for evidence-based intervention.* Emerg Med J, 2012. **29**(6): p. 460-6.
5. Herring, A.A., et al., *Increasing critical care admissions from U.S. emergency departments, 2001-2009.* Crit Care Med, 2013. **41**(5): p. 1197-204.
6. Erenler, A.K., et al., *Reasons for Overcrowding in the Emergency Department: Experiences and Suggestions of an Education and Research Hospital.* Turk J Emerg Med, 2014. **14**(2): p. 59-63.
7. Hsu, N.C., et al., *Why do general medical patients have a lengthy wait in the emergency department before admission?* J Formos Med Assoc, 2014. **113**(8): p. 557-61.
8. Baig, M.A., et al., *Overcrowding in the emergency departments: Challenges and opportunities for improvement.* J Pak Med Assoc, 2015. **65**(12): p. 1344-5.
9. Smith, B., S. Bouchoucha, and E. Watt, *'Care in a chair' - The impact of an overcrowded Emergency Department on the time to treatment and length of stay of self-presenting patients with abdominal pain.* Int Emerg Nurs, 2016.
10. Reznek, M.A., et al., *Door-to-Imaging Time for Acute Stroke Patients Is Adversely Affected by Emergency Department Crowding.* Stroke, 2017. **48**(1): p. 49-54.
11. Richardson, D.B., *Increase in patient mortality at 10 days associated with emergency department overcrowding.* Med J Aust, 2006. **184**(5): p. 213-6.
12. Geelhoed, G.C. and N.H. de Klerk, *Emergency department overcrowding, mortality and the 4-hour rule in Western Australia.* Med J Aust, 2012. **196**: p. 122-6.
13. de Araujo, P., M. Khraiche, and A. Tukan, *Does overcrowding and health insurance type impact patient outcomes in emergency departments?* Health Econ Rev, 2013. **3**(1): p. 25.
14. Rondeau, K.V. and L.H. Francescutti, *Emergency department overcrowding: the impact of resource scarcity on physician job satisfaction.* J Healthc Manag, 2005. **50**(5): p. 327-40; discussion 341-2.
15. Adriaenssens, J., V. De Gucht, and S. Maes, *Determinants and prevalence of burnout in emergency nurses: a systematic review of 25 years of research.* Int J Nurs Stud, 2015. **52**(2): p. 649-61.
16. Kawano, T., et al., *Direct relationship between aging and overcrowding in the ED, and a calculation formula for demand projection: a cross-sectional study.* Emerg Med J, 2014. **31**(1): p. 19-23.
17. Sharieff, G.Q., et al., *Improving emergency department time to provider, left-without-treatment rates, and average length of stay.* J Emerg Med, 2013. **45**(3): p. 426-32.
18. Moskop, J.C., et al., *Emergency department crowding, part 1--concept, causes, and moral consequences.* Ann Emerg Med, 2009. **53**(5): p. 605-11.

19. Ye, L., et al., *Prolonged length of stay in the emergency department in high-acuity patients at a Chinese tertiary hospital.* Emerg Med Australas, 2012. **24**(6): p. 634-40.
20. Khanna, S., et al., *Unravelling relationships: Hospital occupancy levels, discharge timing and emergency department access block.* Emerg Med Australas, 2012. **24**(5): p. 510-7.
21. Asplin, B.R., et al., *A conceptual model of emergency department crowding.* Ann Emerg Med, 2003. **42**(2): p. 173-80.
22. Hwang, U. and J. Concato, *Care in the emergency department: how crowded is overcrowded?* Acad Emerg Med, 2004. **11**(10): p. 1097-101.
23. Eitel, D.R., et al., *Improving service quality by understanding emergency department flow: a White Paper and position statement prepared for the American Academy of Emergency Medicine.* J Emerg Med, 2010. **38**(1): p. 70-9.
24. American College of Emergency Physicians. *Policy Statements*. Clinical and Practice Managements 2017 February 2013; Available from: https://www.acep.org/content.aspx?id=29156.
25. Pines, J.M. and M.L. McCarthy, *Executive summary: interventions to improve quality in the crowded emergency department.* Acad Emerg Med, 2011. **18**(12): p. 1229-33.
26. Weiss, S.J., et al., *Estimating the degree of emergency department overcrowding in academic medical centers: results of the National ED Overcrowding Study (NEDOCS).* Acad Emerg Med, 2004. **11**(1): p. 38-50.
27. Wang, H., et al., *Use of the SONET Score to Evaluate High Volume Emergency Department Overcrowding: A Prospective Derivation and Validation Study.* Emerg Med Int, 2015. **2015**: p. 401757.
28. Beniuk, K., A.A. Boyle, and P.J. Clarkson, *Emergency department crowding: prioritising quantified crowding measures using a Delphi study.* Emerg Med J, 2012. **29**(11): p. 868-71.
29. Boyle, A., et al., *Comparison of the International Crowding Measure in Emergency Departments (ICMED) and the National Emergency Department Overcrowding Score (NEDOCS) to measure emergency department crowding: pilot study.* Emerg Med J, 2016. **33**(5): p. 307-12.
30. Pines, J.M., S.L. Decker, and T. Hu, *Exogenous predictors of national performance measures for emergency department crowding.* Ann Emerg Med, 2012. **60**(3): p. 293-8.
31. Forster, A.J., et al., *The effect of hospital occupancy on emergency department length of stay and patient disposition.* Acad Emerg Med, 2003. **10**(2): p. 127-33.
32. Yoon, P., I. Steiner, and G. Reinhardt, *Analysis of factors influencing length of stay in the emergency department.* CJEM, 2003. **5**(3): p. 155-61.
33. Ay, D., M. Akkas, and B. Sivri, *Patient population and factors determining length of stay in adult ED of a Turkish University Medical Center.* Am J Emerg Med, 2010. **28**(3): p. 325-30.
34. Wong, H.J., et al., *Understanding hospital and emergency department congestion: an examination of inpatient admission trends and bed resources.* CJEM, 2010. **12**(1): p. 18-26.
35. Wang, H., et al., *The inaccuracy of determining overcrowding status by using the national ED overcrowding study tool.* Am J Emerg Med, 2014. **32**(10): p. 1230-6.
36. McCarthy, M.L., et al., *Crowding delays treatment and lengthens emergency department length of stay, even among high-acuity patients.* Ann Emerg Med, 2009. **54**(4): p. 492-503 e4.

37. Elder, E., A.N. Johnston, and J. Crilly, *Review article: systematic review of three key strategies designed to improve patient flow through the emergency department.* Emerg Med Australas, 2015. **27**(5): p. 394-404.
38. Mortimore, A. and S. Cooper, *The "4-hour target": emergency nurses' views.* Emerg Med J, 2007. **24**(6): p. 402-4.
39. Perera, M.L., et al., *The 'four-hour target' and the impact on Australian metropolitan acute surgical services.* ANZ J Surg, 2016. **86**(1-2): p. 74-8.
40. Chang, A.M., et al., *Associations of Emergency Department Length-of-Stay with Publicly Reported Quality-of-Care Measures.* Acad Emerg Med, 2016.
41. Fee, C., et al., *Effect of emergency department crowding on time to antibiotics in patients admitted with community-acquired pneumonia.* Ann Emerg Med, 2007. **50**(5): p. 501-9, 509 e1.
42. Hong, Y.C., et al., *The effect of prolonged ED stay on outcome in patients with necrotizing fasciitis.* Am J Emerg Med, 2009. **27**(4): p. 385-90.
43. Chen, H.C., et al., *The impacts of prolonged emergency department length of stay on clinical outcomes of patients with ST-segment elevation myocardial infarction after reperfusion.* Intern Emerg Med, 2016. **11**(1): p. 107-14.
44. Chalfin, D.B., et al., *Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit.* Crit Care Med, 2007. **35**(6): p. 1477-83.
45. Nippak, P.M., et al., *Is there a relation between emergency department and inpatient lengths of stay?* Can J Rural Med, 2014. **19**(1): p. 12-20.
46. Taylor, C. and J.R. Benger, *Patient satisfaction in emergency medicine.* Emerg Med J, 2004. **21**(5): p. 528-32.
47. Parker, B.T. and C. Marco, *Emergency department length of stay: accuracy of patient estimates.* West J Emerg Med, 2014. **15**(2): p. 170-5.
48. Lin, Y.L., et al., *The impact of National Health Insurance on the volume and severity of emergency department use.* Am J Emerg Med, 1998. **16**(1): p. 92-4.
49. Bullard, M.J., S.J. Liaw, and J.C. Chen, *Emergency medicine development in Taiwan.* Ann Emerg Med, 1996. **28**(5): p. 542-8.
50. Shih, F.Y., et al., *ED overcrowding in Taiwan: facts and strategies.* Am J Emerg Med, 1999. **17**(2): p. 198-202.
51. Tsai, J.C., W.Y. Chen, and Y.W. Liang, *Nonemergent emergency department visits under the National Health Insurance in Taiwan.* Health Policy, 2011. **100**(2-3): p. 189-95.
52. Ko, M., et al., *Prevalence of and Predictors for Frequent Utilization of Emergency Department: A Population-Based Study.* Medicine (Baltimore), 2015. **94**(29): p. e1205.
53. National Health Insurance Administration. *DOH allocates NT$320 million to increase ER quality and efficiency for major illness treatment.* 2012 [cited 2017 Feb 10]; Available from: http://www.nhi.gov.tw/english/HotNewsEnglish/HotnewsEnglish_Detail.aspx?News_ID=78&menu=1&menu_id=291.
54. Ministry of Health and Welfare. *多管齊下提升急診醫療照護適當性，盼各界攜手同努力.* 2016 [cited 2017 Feb 10]; Available from: http://www.mohw.gov.tw/news/571653951.
55. Kreindler, S.A., et al., *Patient characteristics associated with longer emergency department stay: a rapid review.* Emerg Med J, 2016. **33**(3): p. 194-9.

56. Capuano, F., et al., *Factors associated with the length of stay of patients discharged from emergency department in France.* Eur J Emerg Med, 2015. **22**(2): p. 92-8.
57. Handel, D.A., et al., *Association of emergency department and hospital characteristics with elopements and length of stay.* J Emerg Med, 2014. **46**(6): p. 839-46.
58. Wiler, J.L., et al., *Predictors of patient length of stay in 9 emergency departments.* Am J Emerg Med, 2012. **30**(9): p. 1860-4.
59. Chaou, C.H., et al., *Predicting Length of Stay among Patients Discharged from the Emergency Department-Using an Accelerated Failure Time Model.* PLoS One, 2017. **12**(1): p. e0165756.
60. Welch, S.J., et al., *Volume-related differences in emergency department performance.* Jt Comm J Qual Patient Saf, 2012. **38**(9): p. 395-402.
61. Nemoto, M., et al., *Machine Learning for Computer-aided Diagnosis.* Igaku Butsuri, 2016. **36**(1): p. 29-34.
62. Shen, D., et al., *Machine Learning and Data Mining in Medical Imaging.* IEEE J Biomed Health Inform, 2015. **19**(5): p. 1587-8.
63. Taylor, R.A., et al., *Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data-Driven, Machine Learning Approach.* Acad Emerg Med, 2016. **23**(3): p. 269-78.
64. Gartner, D. and R. Padman, *Improving Hospital-Wide Early Resource Allocation through Machine Learning.* Stud Health Technol Inform, 2015. **216**: p. 315-9.
65. Azari, A., V.P. Janeja, and S. Levin. *Imbalanced learning to predict long stay Emergency Department patients*. in *IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2015*. 2015. Institute of Electrical and Electronics Engineers Inc.
66. Wrenn, J., et al., *Estimating patient's length of stay in the Emergency Department with an artificial neural network.* AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium, 2005: p. 1155.
67. Gul, M. and A.F. Guneri, *FORECASTING PATIENT LENGTH OF STAY IN AN EMERGENCY DEPARTMENT BY ARTIFICIAL NEURAL NETWORKS.* Journal of Aeronautics & Space Technologies/Havacilik ve Uzay Teknolojileri Dergisi, 2015. **8**(2).
68. Ding, R., et al. *Predicting emergency department length of stay using quantile regression*. in *International Conference on Management and Service Science, MASS 2009*. 2009. Wuhan.
69. Combes, C., F. Kadri, and S. Chaabane. *Predicting Hospital Length Of Stay Using Regression Models: Application To Emergency Department*. in *10ème Conférence Francophone de Modélisation, Optimisation et Simulation-MOSIM'14*. 2014.
70. Ng, C.J., et al., *Validation of the Taiwan triage and acuity scale: a new computerised five-level triage system.* Emerg Med J, 2011. **28**(12): p. 1026-31.
71. Ministry of Health and Welfare. *公告修正「急診五級檢傷分類基準」，並自 105 年 1 月 1 日起施行*. 2015 [cited 2017 Feb 11]; Available from: http://www.mohw.gov.tw/cht/Ministry/DM2_P.aspx?f_list_no=9&fod_list_no=5349&doc_no=52649.
72. Ministry of Health and Welfare. *The Official Poster of Emergency Department 5-Level Triage System* 2016 [cited 2017 Feb 12]; Available from: http://www.mohw.gov.tw/MOHW_Upload/doc/ER%E6%80%A5%E8%A8%BA%E6%AA%A2%E5%82%B7%E5%88%86%E7%B4%9A-84x60cm_00590950000000.jpg.
73. Kuhn, M. and K. Johnson, *Applied predictive modeling*. Vol. 810. 2013: Springer.

74. RColorBrewer, S., et al., *Package 'randomForest'*. 2015.
75. Steyerberg E. (2009) Patterns of external validity. In: Clinical Prediction Models. Statistics for Biology and Health. Springer, New York, NY
76. Ministry of Health and Welfare. *National Health Insurance Annual Reporthe 2016-2017* [cited 2017 Nov 26]; Available from: https://www.nhi.gov.tw/Resource/webdata/13767_1_National%20Health%20Insurance%20in%20Taiwan%202016-2017(bilingual).pdf
77. Deo, R.C., *Machine Learning in Medicine.* Circulation, 2015. **132**(20): p. 1920-30.
78. Clifton, D.A., et al., *Health Informatics via Machine Learning for the Clinical Management of Patients.* Yearb Med Inform, 2015. **10**(1): p. 38-43.

APPENDIX A: CONTROLLED TERMINOLOGY

The adult nontrauma triage guideline and controlled terminology in Taiwan Triage and Acuity Scale

Category	Standard term of chief complaint	TTAS triage level*	Coding label in the model	Coding term in Chinese	Number of hit in the column of chief complaint**	Selected into the model***
Respiratory system	Shortness of breathing	1-4	res1	呼吸短促,喘,呼吸急促	4660	Yes
	Respiratory arrest	1	res2	呼吸停止	43	No
	Cough	2-5	res3	咳嗽	4803	Yes
	Hyperven-tilation	1-4	res4	換氣過度,手腳麻,四肢麻	307	No
	Hemopty-sis	1-4	res5	咳血	167	No
	Foreign body in airway	1-4	res6	呼吸道內異物,噎到	22	No

	Allergic reaction	1-5	res7	過敏反應,過敏,嘴唇腫	279	No
Cardiovas-cular system	Cardiac arrest	1	CV1	心跳停止	321	No
	Chest pain/chest tightness	1-5	CV2	胸痛,胸悶,胸+緊,胸+無力,胸+悶,胸+痛	4332	Yes
	Palpitation / Irregular heart beat	1-4	CV3	心悸,不規則心跳,心跳,HR	1520	Yes
	Hyperten-sion emergency	1-4	CV4	高血壓	809	Yes
	General malaise / weakness	1-4	CV5	全身虛弱,全身無力	1729	Yes
	Syncope	1-4	CV6	暈厥,暈倒,暈過去	413	No
	Genera-lized edema	1-4	CV7	全身性水腫	49	No

	Limb edema	2-5	CV8	肢體水腫,腳腫	293	No
	Cold, pulseless limb	1-2	CV9	冰冷無脈搏的肢體	1	No
	Unilateral limb redness and heat	2-4	CV10	單側肢體紅熱,腳+紅,腿+紅	454	No
Digestive system	Abdominal pain	1-4	GI1	腹痛,肚子痛,腹部不適,肚子不舒服,腹+痛	12062	Yes
	Anorexia	2-4	GI2	厭食,吃不下	308	No
	Constipa-tion	2-5	GI3	便祕,便秘,大便大不出	240	No
	Diarrhea	1-5	GI4	腹瀉,拉肚子,下瀉	4024	Yes
	Foreign body in rectum	2-5	GI5	直腸內異物	0	No

	Inguinal pain / mass	2-5	GI6	鼠蹊部疼痛,鼠蹊部腫塊,腹股溝	11	No
	Nausea / Vomiting	1-5	GI7	噁心,嘔吐,上吐,想吐	8445	Yes
	Rectoperi-neal pain	2-5	GI8	直腸會陰疼痛	4	No
	Hemato-emesis	1-3	GI9	吐血,血+口	401	No
	Bloody / Tarry stool	1-4	GI10	血便,黑便,血+腸	878	Yes
	Jaundice	2-4	GI11	黃疸	57	No
	Hiccup	3-5	GI12	打嗝	79	No
	Abdominal mass / distension	1-5	GI13	腹脹,腹部腫塊,腹水	1014	Yes
	Foreign body swallowing	1-4	GI14	吞食異物	7	No
Neurologic system	Conscious-ness change	1-4	neu1	意識程度改變,意識,	1566	Yes

				沒記憶,嗜睡		
	Confusion	2-4	neu2	混亂	205	No
	Vertigo / Dizziness	1-4	neu3	眩暈,頭暈	6640	Yes
	Headache	1-4	neu4	頭痛	3325	Yes
	Convulsion	1-3	neu5	抽搐	443	No
	Unsteady gait / Ataxia	1-4	neu6	步態不穩,運動失調,步態失調	81	No
	Tremor	2-4	neu7	震顫,抖動,抖	366	No
	Limb weakness	1-3	neu8	肢體無力,無力	3193	Yes
	Loss of sensation / Paresthesia	3-4	neu9	知覺喪失,感覺異常,麻	1129	Yes
	Symptoms of stroke (sudden onset of slurred speech,	1-3	neu10	中風症狀	865	Yes

	unilateral paresthesia, sudden onset of vision abnorma-lity (anomalo-pia))					
Skeletal system	Backpain	1-4	ske1	背痛,背+痛	685	Yes
	Upper limb pain	1-5	ske2	上肢疼痛,手+痛,指+痛	879	Yes
	Lower limb pain	1-5	ske3	下肢疼痛,腳+痛,足+痛,趾+痛,膝+痛,腿+痛	1300	Yes
	Swelling of a joint	2-5	ske4	關節腫脹,膝+腫,踝+腫	171	No
Urological system	Flank pain	1-4	uro1	腰痛	1189	Yes
	Hematuria	1-5	uro2	血尿	471	Yes

	Discharge / Lesion in the genital organ	3-5	uro3	生殖器官分泌物	1	No
	Swelling of penis	2-5	uro4	陰莖腫脹	0	No
	Scrotal pain / swelling	2-4	uro5	陰囊疼痛	2	No
	Urine retention	2-4	uro6	尿滯留	91	No
	Symptoms associated with urinary tract infection	2-5	uro7	頻尿,解尿疼痛,解尿困難	916	Yes
	Oliguria	1-4	uro8	少尿	56	No
	Polyuria	1-4	uro9	多尿	4	No
	Inguinal pain / mass	2-5	uro10	鼠蹊部疼痛,鼠蹊部腫塊,腹股溝	11	No

ENT system	Otalgia	3-5	ENT1	耳朵疼痛,耳+痛	170	No
	Foreign body in ears	3-5	ENT2	耳內異物	0	No
	Change in hearing ability	3-5	ENT3	聽力改變	0	No
	Tinnitus	4	ENT4	耳鳴	98	No
	Ear discharge	3-5	ENT5	耳朵分泌物	0	No
	Tooth / Gum problems	2-5	ENT6	牙齒,牙齦	32	No
	Throat pain	1-5	ENT7	喉嚨痛	2139	Yes
	Neck swelling / pain	1-5	ENT8	頸部腫脹,頸部疼痛,頸+痛,落枕,頸+痠,脖子,頸+緊	444	No
	Swallow-ing difficulty	1-3	ENT9	吞嚥困難	85	No

	Facial pain (nontrau-matic / Not involved with tooth problems)	1-5	ENT10	顏面疼痛,臉+痛	237	No
	Epistaxis	1-5	ENT11	流鼻血,鼻+血	67	No
	Nasal stuffness due to allergy or nonspecific factor	5	ENT12	過敏或非特定因素引起的鼻塞	0	No
	Foreign body in nose	2-5	ENT13	鼻內異物	0	No
	Symptoms associated with upper respiratory tract infection	2-5	ENT14	上呼吸道感染,鼻塞,流鼻水,咳嗽,喉嚨痛	5781	Yes
Ophthalmo logic system	Eye discharge	3-5	oph1	眼睛分泌物	0	No

	Foreign body in eyes	2-5	oph2	眼睛內異物	0	No
	Visual disorder	2,4	oph3	視覺障礙,視力模糊,眼+模糊	88	No
	Eye pain	2-4	oph4	眼睛疼痛	13	No
	Red eye / itching	3-5	oph5	眼睛紅/癢	0	No
	Photophobia / light injury	2-4	oph6	畏光	2	No
	Diplopia / double vision	3-4	oph7	複視,雙影	10	No
	Eye rim swelling	2-5	oph8	眼眶腫脹,眼+腫	213	No
	Eye recheck	2-5	oph9	眼睛複檢	0	No
	Chemical exposure to eye	2	oph10	化學物質暴露眼睛	0	No

Dermatological system	Blood or body fluid exposure	2-3	der1	血液體液暴露,針扎	21	No
	Pruritus	3-5	der2	搔癢,癢	778	Yes
	Rash	2-5	der3	紅疹,疹	1100	Yes
	Local redness / swelling	1-5	der4	局部紅腫	644	Yes
	Mass / corn	2-5	der5	腫塊,長繭	432	No
	Breast redness / swelling	2-5	der6	乳房紅腫,乳房+腫,乳房+痛	8	No
	Suspected contagious skin lesion	3-5	der7	疑似傳染性皮膚病	13	No
	Cyanosis	1-3	der8	發紺	8	No
	Spontaneous bruising	1-3	der9	自發性瘀斑,瘀青,凝血異常	15	No
	Foreign body in the skin	3-5	der10	皮膚內異物	3	No

	Other dermato-logic condition	3-5	der11	其他皮膚情況	154	No
Gynecologic / Obstetric system	Menstrua-tion related	1-4	GYN1	月經問題,月經,MC	170	No
	Foreign body in vagina	1-5	GYN2	陰道內異物	2	No
	Vaginal discharge	2-5	GYN3	陰道分泌物	6	No
	Confirmed / suspected sexual assault	1	GYN4	性侵	2	No
	Vaginal bleeding	1-4	GYN5	陰道出血	14	No
	Labia swelling	2-5	GYN6	陰唇腫脹,陰唇+腫	7	No
	Vaginal pain / itching	3-5	GYN7	陰道疼痛/搔癢,陰道腫痛	2	No

	Pregnancy related problem, over 20 weeks	1-4	GYN8	懷孕問題大於 20 周	12	No
	Pregnancy related problem, less than 20 weeks	1-4	GYN9	懷孕問題小於 20 周	27	No
	Postpartum bleeding	1-3	GYN10	產後出血	0	No
Mental system	Depression / suicide	2-4	psy1	憂鬱,自殺	347	No
	Anxiety / Agitation	2-4	psy2	焦慮,激動,情緒	433	No
	Illusion / delusion	2-5	psy3	幻覺,妄想	186	No
	Insomnia	4-5	psy4	失眠,睡不著	226	No
	Violent behavior / self-harm / harm	1-3	psy5	暴力行為,自傷,傷害他人	139	No

	Social problem	3-5	psy6	社會問題,社交問題	3	No
	Bizarre behavior	1-5	psy7	怪異行為	6	No
Substance abuse / misuse	Substance misuse / intoxica-tion	1-4	sub1	物質誤用,物質,中毒,毒氣	108	No
	Substance withdrawal	1-4	sub2	物質戒斷	0	No
	Drug overdose	1-4	sub3	藥物過量	186	No
Miscellaneous	Exposure to infectious disease	5	mis1	暴露於傳染性疾病	5	No
	Fever / chills	1-4	mis2	畏寒,chillness,發燒(exclude"無發燒")	7821	Yes
	High blood sugar	1-3	mis3	高血糖,血糖高	280	No

	Low blood sugar	1-3	mis4	低血糖,血糖低,D50W	324	No
	Referral / Transferred	1-5	mis5	轉診,OPD,門診,診所	2124	Yes
	Change dressing	3-5	mis6	換藥	2	No
	Stitch removal	3-5	mis7	拆線	3	No
	Image check	5	mis8	影像檢查	1	No
	Medical device related	1-4	mis9	醫療裝置問題，尿管 ,PTCD,節律器,洗腎管,體內電擊器	234	No
	Certificate application	5	mis10	診斷書	256	No
	Drug refill / prescrip-tion	5	mis11	開藥	313	No
	Ring removal	2-5	mis12	移除戒指	0	No

	Abnormal lab/ exam result	2-4	mis13	檢查結果異常	166	No
	Post-operative complica-tion	2-5	mis14	手術後併發症	2	No
	Pale / Anemia	1-4	mis15	蒼白,貧血	300	No

*Besides standard terms of chief complaints serving as the primary modulator, TTAS triage level would also be influenced by some so-called secondary modulators, such as respiratory distress, hemodynamically stable, central versus peripheral pain, pain severity, and so on.

**The numbers here indicated how many patients out of my study population have the specific term in their chief complaint. For example, 2932 patients had “shortness of breathing” in their chief complaint.

***Some standard terms won’t be used in this model. The major reason is that as coding terms, they serve so poorly that they either catch few chief complaints or catch none.

APPENDIX B: CODEBOOK

Variable Name	**Definition**	**Attribute**	**Unit**	**Source**
Response variable				
EDLOS_hr_log	The log value to the base 2 of a patient's ED length of stay measured in unit of hours	Numerical		Computed (if the patient is discharged from ED, the ED LOS would be the time difference between the time of discharge and the time of registration. If the patient is admitted to the ward or ICU, the ED LOS would be the time difference between the time of admittance and the time of registration)
Explanatory variables				
1.A patient's demographic data				

age	A patient's age	Numerical	Year old, to 1 digit	Computed (the difference between the date of registration and the date of birth)
gender	A patient's gender	Categorical (Dichotomous) (1=male)		Raw data
triage	The triage level of a patient	Categorical (Ordinal)		Raw data
byAmbu	A patient is sent by ambulance	Categorical (Dichotomous)		Raw data
2. Chief complaint data				
res1	"Shortness of breath" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
res3	"Cough" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
CV2	"Chest pain" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced

CV3	"Palpitation" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
CV4	"Hypertension" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
CV5	"Generalized malaise" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
GI1	"Abdominal pain" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
GI4	"Diarrhea" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
GI7	"Nausea" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
GI10	"Bloody stool" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
GI13	"Abdominal distension/mass"	Categorical (Dichotomous)		Deduced

	is in a patient's chief complaint			
neu1	"Consciousness change" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
neu3	"Vertigo" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
neu4	"Headache" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
neu8	"Limb weakness" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
neu9	"Loss/Abnormality of sensation" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
neu10	"Symptoms of stroke" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced

ske1	"Back pain" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
ske2	"Upper limb pain" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
ske3	"Lower limb pain" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
uro1	"Flank pain" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
uro2	"Hematuria" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
uro7	"Frequency" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
ENT7	"Throat pain" is in a patient's chief complaint	Categorical (Dichotomous)		Deduced
ENT14	Any one of ase phrases ("Upper respiratory	Categorical (Dichotomous)		Deduced

	airway infection”, “nasal congestion”,”rhinorrhea”,”cough”,”throat pain”) is in a patient’s chief complaint			
der2	“itching” is in a patient’s chief complaint	Categorical (Dichotomous)		Deduced
der3	“Rash” is in a patient’s chief complaint	Categorical (Dichotomous)		Deduced
der4	“Local redness or swelling” is in a patient’s chief complaint	Categorical (Dichotomous)		Deduced
mis2	“Fever or chills” is in a patient’s chief complaint	Categorical (Dichotomous)		Deduced
mis5	“Transfer (from other facility)” is in a patient’s chief complaint	Categorical (Dichotomous)		Deduced
gendis	“Generalized” and “discomfort”	Categorical (Dichotomous)		Deduced

	are both in a patient's chief complaint			
DM	"Diabetes"-related terms are mentioned in a patient's chief complaint	Categorical (Dichotomous)		Deduced
CA	"Cancer"-related terms are mentioned in a patient's chief complaint	Categorical (Dichotomous)		Deduced
ESRD	"End-stage renal disease"-related terms are mentioned in a patient's chief complaint	Categorical (Dichotomous)		Deduced
3. A patient's vital signs at triage				
SBP	A patient's systolic blood pressure	Numerical	mmHg	Raw data
DBP	A patient's diastolic blood pressure	Numerical	mmHg	Raw data

HR_num	A patient's heart beat per minute	Numerical	Count	Raw data
SPO2_num	A patient's oxygen saturation level	Numerical	%	Raw data
RR_num	A patient's respiratory rate per minute	Numerical	Count	Raw data
GCS_num	A patient's consciousness level, measured in Glasgow Coma Scale	Categorical (Ordinal)		Raw data
pain_scale_num	A patient's self-assessed pain severity	Categorical (Ordinal)		Raw data
4. Registration status features				
4.1 General indicators				
regisShift	The shift on which a patient arrives in the ED	Categorical (1->00:00~07:59 2->08:00~15:59 3->16:00~23:59)		Raw data
whatday	The day on which a patient arrives in the ED	Categorical(1->Mon, 2->Tue, 3->Wed,4->Thu,		Computed from the date of registration

		5->Fri, 6->Sat, 7->Sun)		
longweekend	The day of registration is among a long weekend (more than 2 days)	Categorical (Dichotomous)		Computed from the date of registration
holiday	The day of registration is a holiday	Categorical (Dichotomous)		Computed from the date of registration
4.2 Hospital factors				
regularOPD	The availability of regular outpatient services while registration	Categorical (Dichotomous)		Computed from the date and the time of registration
hrtillnextbusihour_sqrt	The square root of the number of hours remaining until the next business hours of the hospital	Numerical	Square root of hour	Computed from the date and time of registration and the hospital working hour
hrtillnextfullexam_sqrt	The square root of the number of hours remaining until the next time that full	Numerical	The square root of hour	Computed from the date and time of registration and the hospital working hour

	health exams are available			
medbedC	The daily reported number of medical ward beds with copayment in the hospital	Numerical	Count	Computed
medbedNHI	The number of medical ward beds without copayment in the hospital	Numerical	Count	Computed
surbedC_sqrt	The square root of the number of surgical ward beds with copayment in the hospital	Numerical	The square root of count	Computed
surbedNHI_sqrt	The square root of the number of surgical ward beds without copayment in the hospital	Numerical	The square root of count	Computed
4.3 ED factors				
4.3.1 Whole-ED census				

past1hr_triage1	The presence of any patients triaged 1 in the past 1 hour before registration	Categorical (Dichotomous)		Computed
past1hr_triage1_med	The presence of any medical patients triaged 1 in the past 1 hour before registration	Categorical (Dichotomous)		Computed
past1hr_triage1_surg	The presence of any surgical patients triaged 1 in the past 1 hour before registration	Categorical (Dichotomous)		Computed
past1hr_triage_ped	The presence of any pediatric patients triaged 1 in the past 1 hour before registration	Categorical (Dichotomous)		Computed
past1hr_triage_gyn	The presence of any gynecologic/obstetric patients	Categorical (Dichotomous)		Computed

	triaged 1 in the past 1 hour before registration			
CurrentEDNo	The total number of patients in ED while registration	Numerical	Count	Computed
CurrentEDNo_triage1	The total number of patients with triage 1 in ED while registration	Numerical	Count	Computed
CurrentEDNo_triage2	The total number of patients with triage 2 in ED while registration	Numerical	Count	Computed
CurrentEDNo_triage3	The total number of patients with triage 3 in ED while registration	Numerical	Count	Computed
CurrentEDNo_triage4	The total number of patients with triage 4 in ED	Numerical	Count	Computed

	while registration			
CurrentEDNo_triage5	The total number of patients with triage 5 in ED while registration	Numerical	Count	Computed
CurrentED_boarding	The total number of patients that determined to be admitted but still staying in ED due to any reason	Numerical	Count	Computed
Past24EDvisit	In the past 24 hours before registration, the total number of patients visiting the ED	Numerical	Count	Computed
Past24EDdisc	In the past 24 hours before registration, the total number of patients discharged from ED	Numerical	Count	Computed

Past24EDAdmDecision	In the past 24 hours, the total number of patients in ED that have been decided to be admitted	Numerical	Count	Computed
Past24EDAdm	In the past 24 hours, the total number of patients admitted to ward or ICU from the ED	Numerical	Count	Computed
4.3.2 Medical subunit census				
CurrentMedNo	The total number of patients in the medical subdepartment while registration	Numerical	Count	Computed
CurrentMedNo_triage1	The total number of patients with triage 1 in the medical subdepartment while registration	Numerical	Count	Computed

CurrentMedNo_triage2	The total number of patients with triage 2 in the medical subdepartment while registration	Numerical	Count	Computed
CurrentMedNo_triage3	The total number of patients with triage 3 in the medical subdepartment while registration	Numerical	Count	Computed
CurrentMedNo_triage4	The total number of patients with triage 4 in the medical subdepartment while registration	Numerical	Count	Computed
CurrentMedNo_triage5	The total number of patients with triage 5 in the medical subdepartment while registration	Numerical	Count	Computed

CurrentMed_boarding	The total number of patients that determined to be admitted but still staying in the medical subdepartment due to any reason	Numerical	Count	Computed
CurrentMed_boarding_longest	Among those medical boarding patients, the longest ED length of stay while registration	Numerical	Hour	Computed
CurrentMed_boarding_median	Among those medical boarding patients, the median ED length of stay while registration	Numerical	Hour	Computed
Past24Medvisit	In the past 24 hours before registration, the	Numerical	Count	Computed

	total number of patients visiting the medical subdepartment			
Past24Meddisc	In the past 24 hours before registration, the total number of patients discharged from the medical subdepartment	Numerical	Count	Computed
Past24Meddisc_LOS_median	Among those discharged from the medical subdepartment in the past 24 hours, the median ED length of stay	Numerical	Hour	Computed
Past24MedAdmDecision	In the past 24 hours, the total number of patients in the medical subdepartment that have been	Numerical	Count	Computed

	decided to be admitted			
Past24MedAdm	In the past 24 hours, the total number of patients admitted to ward or ICU from the medical subdepartment	Numerical	Count	Computed
Past24MedAdm _LOS_median	Among those medical patients admitted in the past 24 hours, the median ED length of stay	Numerical	Hour	Computed
4.3.3 Medical subunit staffing factors while registration				
VS_ED	The number of attending physicians taking care of oncoming medical patients while registration	Numerical	Count	Raw data
VS_round	The number of attending physicians	Numerical	Count	Raw data

	taking care of the medical patients in the observation room while registration			
VS_boarding	The number of attending physicians taking care of the boarding medical patients while registration	Numerical	Count	Raw data
PGY	The number of post-graduate-year students in the medical subdepartment	Numerical	Count	Raw data
R1	The number of first-year residents in the medical subdepartment	Numerical	Count	Raw data
R2	The number of second-year residents in the	Numerical	Count	Raw data

	medical subdepartment			
R3	The number of third-year residents in the medical subdepartment	Numerical	Count	Raw data
R4	The number of fourth-year residents in the medical subdepartment	Numerical	Count	Raw data
NP	The number of nurse practitioners in the medical subdepartment	Numerical	Count	Raw data
NP_extra	The number of extra nurse practitioners in the medical subdepartment	Numerical	Count	Raw data
RN	The number of nurses in the medical subdepartment	Numerical	Count	Raw data

RN_extra	The number of extra nurses in the medical subdepartment	Numerical	Count	Raw data
4.4 Ambulance-related indicators while registration				
AmbDivStatus	At registration, is the ED on "ambulance diversion" or not	Categorical (Dichotomous)		Deduced
past24hr_AmbDiv_hr	How many hours in the past 24 hours was the ED on "ambulance diversion"	Numerical	Hours, to 1 digit	Deduced
past_AmbDiv25_48	How many hours between the past 24 hours and the past 48 hours was the ED on "ambulance diversion"	Numerical	Hours, to 1 digit	Deduced
past_AmbDiv49_72	How many hours between the past 48 hours and the past 72 hours was the ED on	Numerical	Hours, to 1 digit	Deduced

	"ambulance diversion"			
past_AmbDiv73_96	How many hours between the past 72 hours and the past 96 hours was the ED on "ambulance diversion"	Numerical	Hours, to 1 digit	Deduced
past_AmbDiv97_120	How many hours between the past 97 hours and the past 120 hours was the ED on "ambulance diversion"	Numerical	Hours, to 1 digit	Deduced
past_AmbDiv121_144	How many hours between the past 120 hours and the past 144 hours was the ED on "ambulance diversion"	Numerical	Hours, to 1 digit	Deduced
past_AmbDiv145_168	How many hours between the past 144 hours and the past 168	Numerical	Hours, to 1 digit	Deduced

	hours was the ED on “ambulance diversion”			
past_AmbDiv169_240	How many hours between the past 168 hours and the past 240 hours was the ED on “ambulance diversion”	Numerical	Hours, to 1 digit	Deduced