# Risk Factors for Potentially Preventable Readmissions – A Case Study of Diseases of the Circulatory System

Chengyang Wu
Master of Research

Supervisor team:
Professor Stefan Trück
Dr Oscar Perez Concha
Dr Bavani Arunasalam

Department of Applied Finance and Actuarial Studies
Macquarie University

6 November 2016

# Contents

# List of Tables

# List of Figures

# Abstract

Potentially preventable hospital readmission rates are considered an important indicator of quality of health care. These re-hospitalisations are likely to be harmful, costly and potentially avoidable. This research identifies the risk factors that cause potentially preventable readmissions.

We analysed admissions to hospital of 7,044 patients with a circulatory system disease as principal diagnosis using a health insurance claim database collected between 2010 and 2016. We developed an algorithm that identifies preventable readmissions from the insurance claim records and subsequently a logistic regression model that allows us to identify the risk factors of these potentially preventable readmissions. The analysis gives a specific focus on examining whether cost-based measures can help explain the risk for patients readmit to hospital with preventable reasons.

Our findings suggest that patients with circulatory system diseases were more likely to have a potentially preventable hospital readmission if they had one or more of the following factors at the time of the initial admission: being male, more complications (comorbidities) apart from the main diagnosis, stroke conditions, and having procedures of digital subtraction angiography of aorta and lower limb. Importantly, the more doctors charged over the scheduled fees for the medical service associated with a patient's initial admission, the less likely the patient would be readmitted for a preventable reason.

# Statement

I **CHENGYANG WU**, certify that the work in this thesis entitled **'Risk Factors for Potentially Preventable Readmissions – A Case Study of Diseases of Circulatory System'** has not been submitted for a higher degree to any university or institution other than Macquarie University.

I also certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my research have been acknowledged.

In addition, I certify that in the thesis the sources of information and literature used are indicated.

# Acknowledgments

I wish to acknowledge the financial support and the data analytics platform provided by the Capital Markets Cooperative Research Centre under the Health Market Quality Program and our health industry partner's.

I would like to express my appreciation to all those who helped me complete this research.

I would first like to thank my principal supervisor Professor Stefan Trueck whose suggestions and advice helped improve this thesis dramatically. Your advice on research has been invaluable and also allowed me to develop as an academic researcher. Special thanks to my co-supervisor, Dr Oscar Perez Concha, for his kindness and gentle guidance during the processes of developing algorithms, model evaluation and editing the thesis.

I would also like to thank Dr Bavani Arunasalam for your support and unending encouragement through this year. It has truly been a joy to work with you, and also your support during hardship is much appreciated. Thank you for reviewing my thesis as well.

To Dr Uma Srinivasan, thank you for reviewing my thesis.

To my partner, thank you for your company and looking after me when I was sick during this year.

# 1 Introduction

## 1.1 Overview

Circulatory system diseases were the leading cause of death in the last decade in Australia and affect more than 3.72 million people (Heart Foundation 2016). The most common circulatory system diseases are cardiovascular diseases, which include ischaemic heart diseases, cerebrovascular diseases, heart failure and others (Australian Bureau of Statistics 2015). Cardiovascular disease affects one in six Australians and kills one person every 12 minutes (Heart Foundation 2016). This problem will get more prevalent in a society with an aging population, insufficient physical activity and poor diet (Heart Foundation 2016).

Circulatory system diseases are the main cause for hospitalisations, and one of the major causes of readmissions to hospitals (Anika et al. 2014; Li et al. 2015). Readmissions to hospital can be scheduled as part of the treatment (planned readmission), or as a result of a new and unexpected disease (unplanned and unpreventable readmissions). But unfortunately, many of the readmissions to hospital are potentially preventable, that is, they are clinically related to the initial admission (index admission) (Halfon et al. 2006; Kripalani et al. 2007) and might occur due to inadequacies in provision of care in the initial hospitalisation, discharge planning, post-discharge follow-up, or coordination between inpatient and outpatient health care teams (Goldfield et al. 2008).

Given the high risk of patients with circulatory system diseases being readmitted to hospital due to circulatory system diseases, it is critical to investigate the preventability of these readmissions and the risk factors that can explain these potentially preventable readmissions (PPRs).

PRRs are very harmful to patients. Generally, they can lead to prolonged inactivity, delayed return to work, further medical procedures and increased cost, in addition to the emotional strain for the patient (Private Healthcare Australia 2015). Furthermore, PPRs are potentially life-threatening to patients. Patients may be readmitted because of hospital errors or omissions in care during the initial hospitalisation, adverse events, or inadequate post-discharge care (Ganguly et al. 2013), which can be potentially prevented by pre-discharge assessment, medication reconciliation, communication between the hospital care team and

aftercare providers, and patient education (Benbassat and Taragin 2000; Agency for Healthcare Research and Quality 2016; Edelman 2016). PPRs also include readmissions due to adverse events (Australian Institute of Health and Welfare 2016) defined as "incidents in which harm resulted to a person receiving health care" including "infections, falls resulting in injuries, and problems with medication and medical devices" (Australian Institute of Health and Welfare 2016). Adverse events are treated as an important indicator for hospital performance and the safety of patient care (Health NSW 2015; Australian Institute of Health and Welfare 2016). According to Australian hospital statistics for 2014-2015, the average rate of preventable readmissions caused by adverse events is around 22 per 1,000 separations for 'selective procedures'. . Note that the term 'selective procedures' refers to procedures which are medically necessary but can be delayed for at least 24 hours. (Australian Institute of Health and Welfare 2016). This rate is similar to the rate five years ago (Australian Institute of Health and Welfare 2016), which indicates there is room for improvement in the safety and effectiveness of care nationwide.

PPRs are associated with lower quality of care, and are an indication of poor resolution of the main cause of hospitalisation (Benbassat and Taragin 2000; Australian Institute of Health and Welfare 2016). Hence, solving the issue of PPRs continues to be a priority for both policy makers and health care providers. For example, in the financial year 2013, CMS (Medicare and Medicaid Services) in the United States introduced the Hospital Readmissions Reduction Program to reduce PPRs. The program included three conditions that cause the highest number of readmissions at a national level: acute myocardial infarction, heart failure and pneumonia. If the participating hospitals have an excessive number of readmissions after 30 days of discharge of patients with these conditions, and in addition, these readmissions are proved to be preventable, the hospitals will be penalised. Additionally, the preventable readmission rate for each hospital may have an impact on hospitals' reputations and patients' choice of hospitals if the rate is made public (Boulding et al. 2011). Hence, many hospitals chose to invest in improving their quality of care in order to reduce PPRs (Carroll, Edwards and Lashbrook 2011). These strategies brought organisational changes that had effective results on improving quality of care, such as standardised process for implementing new protocols, education of clinical staff on safety measures, reduction of errors, infection prevention or expansion of hospital services (Carroll, Edwards and Lashbrook 2011).

Besides the quality and safety of care, federal regulatory bodies in America are increasingly concerned about financial funding for health care. The PricewaterhouseCoopers' Health Research Institute estimated the total cost of preventable hospital readmissions in the United States was $25 billion annually in 2006 (Ganguly et al. 2013). As Medicare (in the US) alone currently spends $15 billion a year on hospital readmissions and 18% of Medicare patients discharged from a hospital are readmitted within 30 days (Katterl et al. 2012), another purpose of the Hospital Readmissions Reduction Program is to reduce total Medicare payments to hospitals with excessive 30-day preventable readmissions. The program has already had a significant financial impact (Ganguly et al. 2013).

Reducing PPRs is also beneficial to health insurance companies. They can save unnecessary expenses, keep health costs under control, and ensure that health insurance premiums remain affordable (Private Healthcare Australia 2015). However, a reduction in PPRs requires cooperation from hospitals and staff.

A number of studies also found that smooth care transitions, which include improvement in discharge planning, better cooperation between hospitals and out-patient care providers, and adequate instructions and training to patients so they can take better care of themselves after discharge, could reduce PPRs (Benbassat and Taragin 2000; Boutwell and Hwu 2009; Edelman 2016). In the United States, many transition programs have been implemented for quality improvement (Distel, Casey, and Prasad 2016). For example, the program of Partnership for Patients: Better Care Lower Costs was funded by the Affordable Care Act and run by the Department of Health and Human Services to connect hospitals and community-based organisations. This program does not only work to improve the safety of patients in hospital, but also to improve care transitions including post-discharge follow-up (American Hospital Association 2011).

As an indicator of hospital performance and the quality and safety of health care, PPRs have been investigated by a number of studies. Current research in the US reports a range of clinical factors that influence PPRs, including comorbidities, severity class, patient age, patient health status and previous use of the healthcare system (Vest et al. 2010). Besides these factors, some studies also examined social determinants of health such as insurance status, marital status and access to care. However, the findings are not conclusive. For example, some studies found that gender did not have significant predictive power to explain

or predict PPRs (Makris et al. 2010; Li et al. 2015), while another study concluded that male patients were more likely to be readmitted to hospital with preventable clinical conditions (Weeks et al. 2009).

In addition, studies identifying risk factors for PPRs in the Australian context are limited; and research investigating the relationship between the costs related to the initial admission and the risk of PPRs is not seen in the literature. Therefore, this study was conducted on claims data from an Australian private insurance fund with the aim of identifying and examining the risk factors of PPRs; and focusing on whether the costs associated with a patient's initial admission have some predictive power to explain whether the patient will encounter a PPR in the 30 days after the initial discharge.

This study will also provide an approach to identify PPRs. It is developed on the basis of the algorithms of the Australian Institute of Health and Welfare (AIHW) for unplanned readmissions and SQLape's algorithm for avoidable readmissions. SQLape is a software tool that is widely used in Switzerland to measure the quality of hospital discharges. It uses diagnosis codes and procedure codes of both admissions and readmissions to identify PPRs. This algorithm is proposed because it enables non-experts to identify PPRs with claims data that usually contain diagnosis codes and procedure codes without requiring medical experts to review medical records from hospitals.

Besides the algorithm for identifying PPRs, the conducted analysis provides an approach to explore the predictive relationship between identified risk factors and PPRs. Following prior studies, logistic regression is used to build a model for identifying risk factors for PPRs. As an interesting contribution to the literature, this study examines not only demographic and clinical factors but also cost-based factors as explanatory variables. Diagnostic tests and model evaluation techniques, including the ROC curve (AUC), McFadden's R-square, and Hosmer-Lemeshow tests were then used to measure the performance of the applied risk models.

## 1.2 Research aim

The overall aim of this study is to identify the risk factors that cause PPRs for patients initially admitted to hospital with a diagnosis of circulatory system diseases. The objectives are:

1. to identify patient-level, hospital-level and clinical factors for PPRs such as age, length of stay (LOS), gender or secondary diagnoses
2. to identify additional cost-related risk factors that help to explain the causes for PPRs.

The specific tasks are:

1. to investigate the relationship between cost-based variables of PPRs among patients with circulatory system diseases, including scheduled fee, hospital fees, medical fees and total schedule fees
2. to examine demographic factors of PPRs among patients with circulatory system diseases, including age and gender
3. to examine clinical factors of PPRs among patients with circulatory system diseases, including comorbidity index, LOS, principal diagnosis, secondary diagnosis, and principal procedure
4. to test whether cost-related risk factors will improve the goodness-of-fit for the models that only include demographic and clinical risk factors.

## 1.3 Significance of study

Most prior studies investigating the risk factors of PPRs were conducted in the United States, with few studies in Australia. This study uses insurance claim data from an Australian private health insurance fund and provides valuable information about whether the factors that have been examined in prior studies have a consistent impact on PPRs among patients in Australian hospitals. Findings from this study will be a good reference for policy makers and health care providers to measure hospital performance. The risk factors are also helpful for healthcare providers to identify patients who require more attention and additional interventions after initial discharge, if the patients have similar characteristics.

Prior studies for PPRs have mostly reported on clinical, demographic and/or socioeconomic factors, with little literature on the relationships between cost-based factors and the risk of PPRs. This study fills this gap and provides evidence for this problem. In addition, the algorithm developed in this study provides an approach to identify PPRs without requiring medical experts. It automatically identifies PPRs with diagnosis and procedure codes (see Section 4.2.1), which enables hospital management panels, policy makers, private health insurance funds, and other non-experts to measure PPRs quickly and easily. The use of claims data also provides an insight to insurance companies of the risk factors of PPRs.

## 1.4 Outline of thesis

This thesis includes six chapters. Chapter 1 is an introduction to the study, which includes an overview of the issues of PPRs and circulatory system diseases, research aims, and significance of the study.

Chapter 2 provides the literature review relating to the study area. It has five sections. The first section describes the application of electronic health insurance claim data in medical research. The second section presents the various definitions for index admissions and readmissions in prior studies. The third section reviews the risk factors that have been found and discussed in previous studies. The fourth section describes research on PPRs in Australia. The last section summarises the main approaches to modelling in the literature.

Chapter 3 addresses the methods applied in this study, including the algorithm for identifying PPRs, the approaches to modelling and the methods for model evaluation.

Chapter 4 describes the samples for modelling, the definitions of key variables, the methods used for processing data, and the hypotheses for the associations between cost-based variables and PPRs.

Chapter 5 presents the results of this study. It describes the preliminary model, testing models and the final model, then the estimation results of modelling, model evaluation and interpretation of the results are presented. It then discusses the most significant findings from this study with a comparison to the findings from prior studies.

Chapter 6 summarises the research aims and findings and presents the strengths and limitations of this study. Finally, the conclusions for the study are presented.

# 2 Literature review

## 2.1 Application of electronic health insurance claim data for health services research

Claims databases are electronic records of millions of transactions that have occurred between patients and healthcare providers (Ferver et al. 2009). This data typically describe the billable interactions (insurance claims) between insured patients and the healthcare delivery system. There are four general categories: inpatient, outpatient, pharmacy and enrolment interactions (UW Data Resources in the Health Sciences 2016). They record information on diagnosis, treatments and procedures, providers as well as financial measures such as billed amounts, reimbursed amounts, and patient cost sharing (Tyree, Bonnie and Lafferty 2006).

There are many advantages of claims data, and it has long been used for health services research because it is anonymous, plentiful, inexpensive and widely available in electronic format (Hicks 2003). In addition, it is an ideal replacement or complement to medical records (Ferver et al. 2009). Medical records are defined as a chronological written account of a patient that includes information on a patient's examination, treatment and medical history; and also records the physician's physical findings, the results of diagnostic tests and procedures, and medications and therapeutic procedures (Dictionary.com 2016). However, there are some disadvantages to using medical data in research studies. Firstly, medical records are expensive and difficult to obtain due to privacy issues. Secondly, they are usually not fully available in electronic format (Ferver et al. 2009). In order to obtain the information that is relevant to the research and transfer it into electronic format, researchers apply a process which is known as abstracting (Panacek 2007). The process of abstracting requires hiring experts who are usually doctors or advanced medical students to interpret the information in medical records (Ferver et al. 2009). The use of claims data can reduce the cost of research by avoiding the process of abstracting. The use of claims data also solves some problems of studies that collect data by surveys or interviews, for example, claims data does not require a patient's authorisation and is free from non-response and dropout (Baron and Weiderpass 2000).

Besides the advantages of claims data that are beneficial in research, there are other advantages for specific research. Firstly, claims databases are useful for finding sizable groups of patients with rare conditions, such as quadriplegia or aplastic anaemia, who might be difficult to locate by other means (Couris et al. 2003). Secondly and very important for our study, claims data is convenient for researchers to establish the cost for certain diagnoses or perform cost-effectiveness analyses because they contain information on fee schedules, reimbursement amounts, and other financial items (Morris et al. 2003).

Claims data is widely used in areas such as access to health care, prevention and detection of diseases, quality assessment of healthcare services, analysis of morbidity, mortality and adverse events, and analysis of interventions, therapies, and treatments (Hicks 2003).

According to a study by reviewing 1,956 original research studies that were published during 2000-2005 in five health care journals (Ferver et al. 2009), it was concluded that, claims data was primarily used in studies that focused on access to health care (49% of claims-based studies) followed by quality issues (23.8% of claims-based studies); and they were less likely to be used in studies of morbidity issues (9.1% of claims-based studies) or studies of prevention (5.6% of claims-based studies).

The aim of this research is to identify the risk factors that cause PPRs. We chose to use claims data because they have been widely used in past research in quality of care; and also they have many advantages that are useful for our research. For example, they provide the electronic records of a patient's diagnosis codes for each hospital admission, which is convenient to identify the admissions of patients with circulatory system diseases without patient privacy issues. The information on diagnoses and treatments is also useful to group the potential preventable readmissions by our algorithm, which is discussed later. The most important reason is that it provides billed amounts and reimbursed amounts for each hospital admission, which enables us to develop cost measure variables to examine whether these variables can help to explain preventable hospital readmissions.

## 2.2 Defining index admission and readmission

A readmission is defined as a return hospitalisation that follows a prior admission (the so called index admission) within the readmission time interval, usually 30 days (Goldfield et al.

2008). Generally, the prior admissions, or initial admissions, are defined as index admissions, excluding admissions in which the patients die in the hospital; admissions in which the patients transfer to another healthcare facility; and admissions in which the patients leave against medical advice (Agency for Healthcare Research and Quality 2012). The index admission is also the starting point for analysing subsequent hospitalisations.

The definitions differ depending on the study. Some studies restricted patients' age range for the index admissions (Krumholz et al. 2011; Lichtman et al. 2013); others excluded the prior admission as the index admission if the readmission was planned (Yam et al. 2010; Lichtman et al. 2013); and some studies focused on readmissions after the initial discharge of treatment of a particular disease (Garcia et al. 2003; Kumbhani et al. 2009).

On defining readmission in the literature, Lavenberg et al. (2014) summarised that readmissions might be counted differently depending on whether they are to the same hospital or to any hospital, whether they are for the same (or a related) condition or for any condition, whether a patient is allowed to be counted only once during the follow-up period, and whether observation stays are considered. The time interval from initial discharge to readmission ranges from 7 days to 365 days (Bottle et al. 2014) depending on the type of stay. Three different opinions of whether the readmission time interval could impact the readmission rate have been seen in previous research:

1) One view is that longer readmission time intervals decrease the likelihood that a readmission was related to the clinical care or discharge planning in the initial admission (Hannan et al. 2003).
2) On the other hand, Bottle et al. (2014) conducted research to identify the effect of the readmission primary diagnosis and time interval in heart failure patients. They analysed readmission at 7, 30, 90, 182 and 365 days after the index discharge and concluded that the time since discharge made little difference to the readmissions.
3) The third opinion is that a 30-day readmission time interval is optimal to identify readmissions, which has been proved mathematically in two studies (Halfon et al. 2002; Heggestad 2002).

## 2.3 Risk factors for potentially preventable readmissions

There are numerous studies on the factors that are associated with risk of hospital readmissions, while relatively little research focuses on risk factors for PPRs specifically.

This section focuses on the risk factors that are associated with risk of potentially preventable (avoidable) readmissions. A number of factors have been investigated in previous studies and they are grouped as demographic factors, clinical factors and other factors.

### 2.3.1 Demographic factors

This section explores the relationships between PPRs and demographic factors including patient's age, gender, race, marital status, insurance status and socioeconomic status. A large number of studies found increasing age significantly influences how likely a patient readmission to hospital is (Vest et al. 2010). However, there are some differences in the research findings on the influence of patient's age on PPRs. Two studies found that preventable readmissions have been associated with age: one found that patients of older age (over 65) were more likely to be readmitted to hospital with preventable conditions (Lichtman et al. 2013), and the other study found that extremes of age had an association with PPRs (Saunders et al. 2015). However, two other studies found age did not have any significant influence on PPRs (Donze et al. 2013; Donze, Lipsitz and Schnipper 2014).

A patient's gender was found to have an inconsistent effect on PPRs. A few studies found gender was not a significant risk factor for PPRs (Yam et al. 2010; Donze et al. 2013; Donze, Lipsitz and Schnipper 2014). However, Shams et al. (2015) found that male patients with heart failure, acute myocardial infarction, pneumonia or obstructive pulmonary disease were more likely to have an avoidable readmission. On the other hand, female patients with stroke or cardiovascular diseases had an increased likelihood of returning to hospital for preventable conditions (Lichtman et al. 2013).

On the effect of patient's race on likelihood of PPRs, two studies In the US found that race was a factor associated with PPRs, but they did not clearly note which race of patients had a higher likelihood of being readmitted for preventable conditions (Donze et al. 2013; Shams, Ajorlou and Yang 2015).

Marital status of medical patients has been examined by Donze et al. (2013) who categorised patients into three groups: current spouse or partner, single/never married, and separated/divorced/widowed/no answer. The study found marital status was not associated with the risk of PPRs. Two other studies supported this finding, based on the data of patients with heart failure, acute myocardial infarction, pneumonia or chronic obstructive pulmonary disease, and the data of readmissions due to end-of-life issues respectively (Donze, Lipsitz and Schnipper 2014; Shams, Ajorlou and Yang 2015). However, Moore, Gao and Shulan (2013) reported that unmarried patients experienced significantly more readmissions.

There are limited studies in the literature that examined whether insurance status is associated with PPRs. A retrospective cohort study based on 2011-2012 Veteran Health Administration data in the United States found insurance status (Medicare, Medicaid, private, none) did not significantly affect avoidable readmissions (Shams, Ajorlou and Yang 2015). In contrast, another study (Hasan et al. 2010) reported that insurance status was a significant predictor of early readmission in research on identifying predictors of hospital readmission in general medicine patients. However, Hasan et al. (2010) replicated a similar analysis in patients discharged to sub-acute and they found insurance status were much less predictive of readmission. In both analyses, they examined four insurance statuses: Medicare, Medicaid (a social healthcare program for families and individuals with limited resources), self-pay and private insurance. Medicare, Medicaid and self-pay were found to be more significant predictors of the risk of PPRs than private insurance

### 2.3.2 Administrative factors

Length of hospital stay is defined as the period of admitted patient care between a formal admission and a formal separation (Australian Institute of Health and Welfare 2016). It is widely considered a key indicator of hospital performance for costing and management (Kulinskaya, Kornbrot and Gao 2005; Australian Institute of Health and Welfare 2016). Liu, Phillips, and Codde (2001) found that LOS could be influenced by factors such as diagnoses, age, payment classification, source of referral, specialty of doctor, and ethnic group.

There is good evidence that LOS during the index admission is an important indicator for risk of PPRs (Farraris et al. 2001; Halfon et al. 2002; Hasan et al. 2010; Yam et al. 2010; Donze et al. 2013; Lichtman et al. 2013; Shams, Ajorlou and Yang 2015). Strong evidence in the

literature indicates that long LOS in the index admission increases the risk of the patient being readmitted to hospital with potentially preventable conditions (Halfon et al. 2002; Yam et al. 2010; Donze et al. 2013; Lichtman et al. 2013). A study focusing on readmissions of patients with heart failure, acute myocardial infarction, pneumonia or chronic obstructive pulmonary disease showed two different outcomes for effect of LOS on PPRs. It found the LOS was significantly associated with PPRs after the initial hospitalisations caused by heart failure and pneumonia, while it was insignificant for cases of acute myocardial infarction and chronic obstructive pulmonary disease (Shams, Ajorlou and Yang 2015).

### 2.3.3 Clinical factors

Comorbidity is defined as the co-occurrence of two or more physical diseases, physical disorders or mental disorders in the same person simultaneously or sequentially (National Institute on Drug Abuse 2010). The effect of comorbidity on PPRs is seen consistently in the literature: the higher the number of comorbidities, the higher the risk of PPRs (Halfon et al. 2002; Lichtman et al. 2013; Donze, Lipsitz and Schnipper 2014; Donze et al. 2016). For example, Lichtman et al. (2013) found that cardiovascular-related comorbid conditions were one of the strong predictors of preventable readmission within 30 days after the initial discharge with a primary diagnosis of ischemic stroke among Medicare beneficiaries aged over 65 years old in the US. Another study (Shams, Ajorlou and Yang 2015). It also found comorbid conditions of chronic bronchitis, malignant neoplasm, mental disorder and substance abuse were associated with PPRs among patients with heart failure, acute myocardial infarction, pneumonia, or chronic obstructive pulmonary disease. Additionally, comorbid conditions of vascular disease, cardiorespiratory, atrial fibrillation and anemia were found to be associated with PPRs among patients with heart failure and acute myocardial infarction (Shams, Ajorlou and Yang 2015).

There are several methods for measuring comorbidity. The most common methods calculate scores based on the different diseases that a patient could have. One of the most commonly used methods for hospital administrative data is the Charlson index (Elixhauser et al. 1998). It categorises comorbidities of patients based on the International Classification of Diseases diagnosis codes at four levels (scores of 1, 2, 3 and 6, depending on the severity of the clinical conditions) to indicate the risk of 1-year mortality for the patient. The higher the score, the higher the risk of mortality for the patient (Manitoba Centre for Health Policy

2016). For example, a score of 6 is assigned for the clinical conditions of malignant tumor, metastasis, and AIDS. Halfon et al. (2002) found that patients with a Charlson score of 3 or higher were more likely to have a readmission to hospital.

### 2.3.4 Other factors

As well as the demographic and clinical factors, there are some other factors that have been found to be associated with risk of preventable readmissions. A study based on readmissions among patients with stroke and other cerebrovascular disease indicated that the most significant factor contributing to PPRs was delayed elective procedures without medical or surgical indication, followed by inadequate outpatient care, incomplete initial evaluations, delayed palliative care consultation and inadequate discharge instructions (Nahab et al. 2012).

Another study based on readmissions after acute medical treatment concluded that 71% of readmissions were judged to be preventable if there was more judicious care provided to patients in the index admissions. Shalchi et al. (2009) found factors leading to preventable readmissions were inadequate therapy, poor discharge planning, hospital-acquired infection and drug-related factors.

## 2.4 Potentially preventable readmissions in Australia

In the United States, there have been numerous studies addressing preventable or avoidable readmissions. However, there are fewer studies in the Australian context in the existing literature. Studies on PPRs in Australia mainly focused on unplanned readmissions and readmissions caused by preventable adverse events (Miles and Lowe 1999; McLean, Mendis and Canalese 2008; Kilkenny et al. 2013).

An early study investigated the preventability of unplanned readmissions to hospital among Australian patients (Miles and Lowe 1999) by examining 3,081 admissions by medical review. It was found that 24 of those admissions were unplanned readmissions caused by inappropriate medical management, and 16 of these unplanned readmissions were due to highly preventable adverse events.

Another study addressing the factors associated with readmissions to hospital after stroke reported that data on factors related to readmissions for patients with stroke in Australia was

limited (Kilkenny et al. 2013). It found that severe complications during their initial admissions and the discharge experience had much impact on the likelihood of patients being readmitted. It also provided some recommendations to hospitals to prevent readmissions among patients with stroke including greater vigilance and monitoring to avoid preventable serious adverse events, such as urinary tract infections or falls; and additional attention on patients with a moderate to severe disability to prevent their unplanned readmissions.

A retrospective study on unplanned hospital readmissions to a regional Australian hospital over the period 1996 to 2005 (McLean, Mendis and Canalese 2008) reported that between 9% and 48% of all readmissions were preventable according to the judgements made by different clinicians and administrators. It also found that patients aged over 75 years were more likely to encounter unplanned readmissions; 50% of unplanned readmissions occurred within the first week after the initial discharge; and the top five causes of unplanned readmissions were chronic obstructive pulmonary disease, complications of procedures, heart failure and pneumonia, angina and acute bronchiolitis, of which four were circulatory system diseases. This finding also motivated us to conduct the study addressing PPRs among patients with circulatory system diseases in Australia.

## 2.5 Modelling of readmissions

Many studies on hospital readmissions focused on building risk or prediction models and identifying risk factors and predictors associated with readmissions (Allaudeenet al. 2011; Renton et al. 2011; Nahab et al. 2012; Donze et al. 2013; Yam et al. 2010). The different approaches applied in these studies are discussed below as regression approaches and other approaches.

### 2.5.1 Regression approaches

In prior studies, Logistic regression (LR) was the most popular method to predict the risk of readmissions (Seidensticker et al. 2014). For example, in research on potentially avoidable 30-day hospital readmissions in medical patients, Donze et al. (2013) applied a multivariable LR to predict a score of a patient's risk for readmission. The risk score was developed by using a regression coefficient–based scoring method. The higher the risk score, the more likely the patient would be readmitted after discharge within 30 days of the index admission. Based on the backward multivariable LR analysis, they identified seven significant

independent predictors: haemoglobin at discharge, sodium level at discharge, discharge from an oncology service, procedure during the index admission, index type of admission (non-elective vs elective), number of admissions during the past 12 months, and LOS. It is worthy to note that the study included sodium level as a predictor of readmission, which had not been seen in other studies (Donze et al. 2013).

Some studies developed the risk or prediction model on the basis of a LR model combined with other methods. For example, Lichtman et al. (2013) applied a random-effects logistic regression with Markov Chain Monte Carlo simulations to determine patient-level factors associated with preventable readmissions due to ischemic stroke, including age, gender, race, LOS, discharge disposition, comorbid conditions and medical history. A Wilcoxon rank-sum test was used to compare patient characteristics between the readmitted and admission-free groups. The Kaplan–Meier method was used to estimate the observed all-cause and preventable 30-day readmission rates. This study found 11.9% of readmissions due to ischemic stroke were preventable, and patients that are an older age, female, and with a history of comorbid cardiovascular conditions were more frequently readmitted to hospital for preventable conditions (Lichtman et al. 2013).

A few studies also used survival analysis (or hazard models) to estimate the time duration between consecutive patient readmissions. The most common survival model used in the literature is the Cox proportional hazard model (Bardhan et al. 2012). In the research on identifying preventable readmission risk factors for patients with chronic conditions, Rico et al. (2016) applied a multivariate logistic regression model as the baseline model to estimate the 30-day readmission risk, and also a Cox proportional regression model to assess the risk over time with the proportional hazards assumption. The study included five chronic conditions: CHF (congestive heart failure, COPD (chronic obstructive pulmonary disease), AMI (acute myocardial infarction), Pneumonia, and Type 2 diabetes. The study examined patient factors (age, language, marital status, race and gender), case severity factors (behavioural health, severity of illness, LOS, Charlson comorbidity), hospital factors (the presence of a hospitalist, discharge day of week, admission type, payer class, discharge disposition, admission type). They found the patient-level factors and case severity factors showed inconsistent influences among different disease groups. For example, age was found to be a significant risk factor only in the Type 2 diabetes group, but not for other disease groups. For hospital factors, the presence of a hospitalist and the discharge day of week were

not found statistically significant in any of the models, while LOS was significant in the models across all disease groups, except for AMI (Rico et al. 2016).

The Probit model is another commonly used regression model. Erickson et al. (2014) conducted a study for 30-day hospital readmissions in patients receiving haemodialysis to investigate whether outpatient provider practices influence the risk of readmissions and death after 30 days after the initial discharge. This study had two stages. In the first stage, they used linear regression to predict visit frequency and the average visits to prevalent haemodialysis patients in a patient's facility during the calendar year. In the second stage, they developed a linear probability model to predict the probability of readmissions which was a function of the predicted visit frequency from the first stage, controlling for demographic characteristics, comorbid conditions, and facility characteristics. The readmissions within patients were accounted for by block bootstrap SEMs (Structural Equation Modellings), with 10,000 simulations. They found the patients who died or were readmitted to hospital less frequently visited outpatient providers face-to-face after the initial discharge. Those patients were also more likely to have comorbidities, and longer LOS in the index admissions (Erickson et al. 2014).

### 2.5.2 Other approaches

Besides the classic models, machine learning techniques have also been applied in a number of studies on readmissions in recent years.

Lin (2008) employed random forest algorithms to develop the prediction model for 30-day hospital readmissions with 10, 50 or 100 trees. They classified the patients of 10 subpopulations into different groups according to ranking of the predicted readmissions probabilities by the random forest model. The top 10% of patients with the highest predicted readmission risk were classified as "readmitted" and the bottom 10% patients with the lowest predicted readmission risk were classified as "not readmitted". The 10 subpopulations were patients with cancer, chronic kidney disease, chronic obstructive pulmonary disease, diabetes, heart failure, acute myocardial infarction, pulmonary hypertension, sickle cell anaemia, stroke, and history of transplant. They found the performance of the model varied among different subpopulations; it was most effective for stroke patients with baseline readmission

rate of 12%, and least effective for transplant patients with baseline readmission rate of 23% (Lin 2008).

As a fast and simple classification model, Naïve Bayes (NB) was applied in a study of predicting risk of readmission for congestive heart failure patients (Zolfaghar et al. 2013). This study aimed to develop a multi-layer classifier to predict the risk a patient would be readmitted to hospital within 30 days after the initial discharge. Two multi-layer classifiers were constructed at the beginning and each classifier had three layers for the classifier: "Predicting if patient will be ever readmitted" (layer 1), "Predicting if patient will be readmitted within 60 days" (layer 2), and "Predicting if patient will be readmitted within 30 days" (layer 3). NB was used to solve the problems of layer 1 and layer 2 for both of the multi-layer classifiers. For the first multi-layer classifier, NB was applied to layer 3; while for the second multi-layer classifier, a Support Vector Machine (SVM) classifier was used for layer 3. Additionally, they included another two baseline models for comparing with the two multi-layer classifier models. For the first baseline model, the features were selected by Chi-square and NB; for the second baseline model, the features were selected by Chi-square and SVM. After comparing the confusion matrix results, such as true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), they found the two multi-layer classifiers outperformed the two baseline models, and their performances just differed slightly. The baseline model that was applied SVM performed worst on predicting the risk of readmissions, because SVM could not handle the imbalanced data (Zolfaghar et al. 2013).

A few studies have argued that regression approaches were not the best performers for modelling the risk of readmissions (Lee 2012; Futoma et al. 2015). For instance, a comparison of prediction models for risk of early readmissions was made by Futoma et al. (2015). They constructed different models based on the same data to predict early hospital readmissions, including standard logistic regression, logistic regression with multi-step variable selection, penalised logistic regression, random forest and support vector machine. They found random forests, penalised logistic regressions and deep neural networks have significantly better predictive performance than other methods. They also concluded that the more complicated the models, the higher the overall predictive accuracy; but the complicated models are usually also difficult to tune and interpret.

Another study comparing the prediction models for risk of readmissions supported the finding of Lee (2012). The researcher also constructed three models using data from academic hospital patients: a logistic regression, a decision tree and a neural network. Lee (2012) concluded that the model applying a decision tree had the best predictive power by assessing and comparing their misclassification rate, root asymptotic standard error, lift chart, and ROC curve (Lee 2012).

Regarding risk models or prediction models for PPRs, techniques that have typically been applied in prior studies include regression models such as logistic regression, survival models (e.g. Cox proportional hazard regression models) and probit models, as well as machine learning techniques. Hereby, logistic regression models are probably the most popular approach applied in modelling the risk of readmissions (Seidensticker et al. 2014). The approach has the following favourable features: 1) LR models can include multiple continuous or categorical explanatory variables; 2) the model is easy to interpret with respect to the association between each explanatory variable and the outcome variable, using the estimated coefficients or odds ratios. However, there are also some drawbacks: the models do not necessarily select all relevant variables automatically and can typically not handle imbalanced data effectively (Geng 2001).

Machine learning techniques, such as random forest and deep neural networks, also have been applied in some previous studies dealing with the prediction or risk of readmissions. There is some evidence suggesting that these models have the ability to outperform LR models with regards to their predictive power. They usually can automatically handle missing values and imbalanced data effectively (except SVM). Besides, they can select important variables (Geng 2001, Futoma et al. 2015, Zolfaghar et al. 2013). At the same time, results from machine learning techniques are often far more difficult to interpret and the models have been criticized for over-fitting the data what may lead to rather poor out-of-sample prediction results.

# 3 Applied methods

This chapter describes the methods used in this study to examine risk factors for PPRs for patients with circulatory system diseases. It includes Section 3.1 on the method used to identify PPRs, Section 3.2 on the model definition, and Section 3.3 on the model evaluation.

## 3.1 Identifying potentially preventable readmissions

The proportion of PPRs among readmissions varied widely in the literature. The lowest observed rate was 21% (Dawes et al. 2014) while the highest rate was 76% (Ganguly et al. 2013). Besides the different data sets, this disparity in rates can be due to the different definitions of PPRs and the methods to identify them (Jackson et al. 2013).

Two main methods of identifying PPRs are seen in the literature: identification of PPRs manually by an expert panel in which experts review patients' medical records and determine whether the given readmission is potentially preventable or unpreventable (Goldfield et al. 2008; Yam et al. 2010; Nahab et al. 2012; Saunders et al. 2015); and identification of PPRs automatically (Dawes et al. 2013; Lichtman et al. 2013; Lavenberg et al. 2014).

The most obvious advantage of forming an expert panel is that experts can identify PPRs more accurately. For example, they can check the medical errors in the given data, and judge if a readmission is preventable or not based on their experience with criteria such as the inadequate treatment of an infection, a missed diagnosis, medication errors, errors in discharge planning, or inadequate education and instructions given to the patient (Saunders et al. 2015).

The criteria and approaches for classifying a readmission as potentially preventable or unpreventable are inconsistent in the literature. In a 2010 study in Hong Kong, a quality assessment checklist was developed to record the reasons for patients' readmissions and the preventability of these readmissions considering system, clinician, patient and social factors. These factors were summarised based on international literature and they include a "classification scheme for assessing readmissions, a categorization of the causes of readmission, a checklist for assessing preventability and correlation of the principal and associated factors for readmission" (Yam et al. 2010). A panel of eight experienced

physicians used the checklist to identify the preventable readmissions and each record was reviewed by two physicians independently. If a pair of physicians had different opinions on an assessment, they were required to discuss it together and come to agreement. Otherwise, the other experts would take over the case and make a decision on whether it was preventable or not (Yam et al. 2010).

In an earlier study, Goldfield et al. (2008) also formed a clinical panel to identify PPRs by assessing whether the APR-DRG (The All Patient Refined Diagnosis-related Group) of readmission is clinically related to the APR-DRG of index admission. If so, the readmission is potentially preventable. APR-DRG is one version of DRG (Diagnosis Related Groups). "The DRGs are a patient classification scheme which provides a means of relating the type of patients a hospital treats to the costs incurred by the hospital" (Averill et al. 2003). Based on DRG, The APR-DRG also addresses patient severity of illness and risk of mortality as well as resource intensity (Averill et al. 2003). Goldfield et al. (2008) also defined that a readmission was considered to be clinically related to the initial admission if a patient had a readmission for these reasons:

1) a continued, recurrent or a closely related condition to the index admission
2) an acute decompensation of a chronic problem that was not the reason for the index admission, but was plausibly related to care either during or immediately after the initial admission
3) an acute medical complication plausibly related to care during the initial admission
4) a surgical procedure to address a continuation or a recurrence of the problem.

A systematic review of proportional or preventable readmissions also proved that the criteria used to classify PPRs in current studies were subjective and there are large variations in the application of criteria (Walraven et al. 2011).

In Australia, the AIHW (Australia Institute of Health and Welfare) reports some hospital quality measures. One of them is based on unplanned readmissions, which are treated in this thesis as PPRs. According to the AIHW, the readmissions with some adverse events were directly identified as PPRs. Adverse events are defined as incidents in which harm resulted to a person receiving health care. They include infections, falls resulting in injuries, and problems with medication and medical devices. Some of these adverse events may be preventable (AIHW 2016). Hospital readmissions with a principal diagnosis that indicates an

adverse event are also defined as unplanned readmissions (AIHW 2016). Some unplanned readmissions could be preventable, if the principal diagnosis of the readmission indicates conditions such as complications following infusion, transfusion and therapeutic injection, complications of cardiac and vascular prosthetic devices, complications of internal orthopaedic prosthetic devices, implants and grafts (AIHW 2016; Appendix 1: AIHW unplanned readmissions).

In this study, we develop a hybrid approach adopting both the AIHW unplanned readmissions and rationales of SQLape avoidable readmissions to classify PPRs. SQLape is a tool that is widely used to measure the quality of hospital discharges. It uses diagnosis codes and procedure codes of both admissions and readmissions to identify PPRs. It has recently been tested in the US in a single-centre study, and a multi-hospital study is underway (Lavenberg et al. 2014).

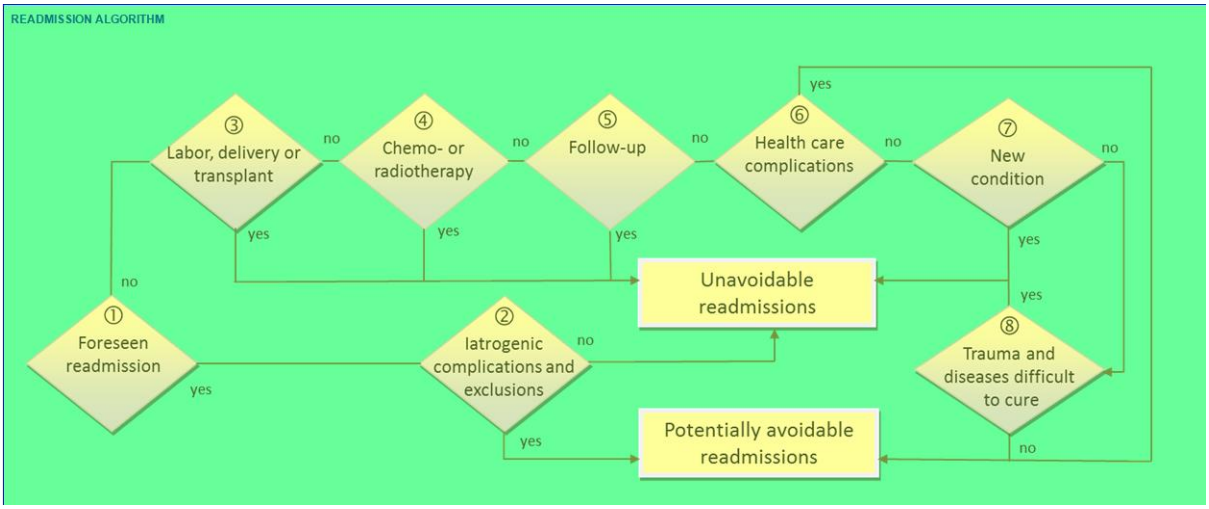


**Figure 1 SQLape PPR algorithm (sqlape.com 2016).**

From the diagram of the SQLape algorithm above, we can see that there are eight steps in total to identify PPRs. In step one and two, the foreseen readmissions with iatrogenic complications and exclusions are identified as potentially preventable readmissions (SQLape.com 2016).

The foreseen readmissions include 1) procedure of material removal or replacement, 2) procedure of temporary stoma closure, 3) diagnosis of postoperative aftercare, 4) main readmission diagnosis is clinical related to at least one diagnosis of the previous stay if the readmission is: programmed surgical readmission after a surgical, or obstetrical stay, or

programmed surgical readmission after a non-surgical/non-obstetrical stay, or programmed non-surgical intervention after a non-surgical/non-obstetrical stay (SQLape.com 2016).

In step two, it excludes unplanned readmissions or surgery complications for ungrouped cases as unavoidable. And the exclusions are unplanned readmissions or surgery complications for ungrouped cases (SQLape.com 2016).

In step three, four and five, it groups labour, delivery, transplant, chemo, radiotherapy and follow-up into unavoidable readmissions (SQLape.com 2016).

In step six, it includes the readmissions that have iatrogenic complications, preventable diseases (e.g. deep vein thrombosis, pulmonary embolism, and decubitus ulcer) and other health care complications as potentially preventable readmissions (SQLape.com 2016).

In step seven, it defines new medical condition as the readmission damaged system (determined by the main diagnosis) is not equal to any damaged system of the index hospitalization (determined by all diagnoses of the previous stay). It includes the readmission that have the new medical conditions in systems of blood, circulatory, cutaneous, digestive, endocrine, ENT, female, hepatic, locomotion, nervous, new-born, mental, ocular, respiratory and urinary as unpreventable readmissions(SQLape.com 2016).

In the last step, it excludes readmissions of trauma or diseases difficult to cure as unpreventable readmissions; the left over readmissions are potentially preventable readmissions (SQLape.com 2016).

No matter which tool is applied in research, there is a limitation that it may not reflect all potentially preventable conditions related to the index admission. A study of comparing manual and automated methods for identifying all-cause PPRs concludes that manual reviews cannot be replaced by automated methods because the concordance between the two methods was not high enough (Jackson et al. 2013). However, there is a good evidence that compared to the manual methods, SQL is an algorithm that has high sensitivity and specification (true positive rate is 96% and false positive rate is 4%) for classifying PPRs within 30 days after initial discharge (Halfon et al. 2006).

A PPR is defined as a hospital readmission within 30 days after the initial discharge that has:

1) an adverse event as the main diagnosis for readmission as defined by AIHW as an unplanned readmission (AIHW 2016)); or
2) a condition related to the index admission that is not part of a follow-up visit to the hospital or
3) any other complication or preventable disease that can be due to lack of quality of care in the index admission and it is not part of a follow-up visit to the hospital.

The diagram of the PPR algorithm is shown in Figure 2.

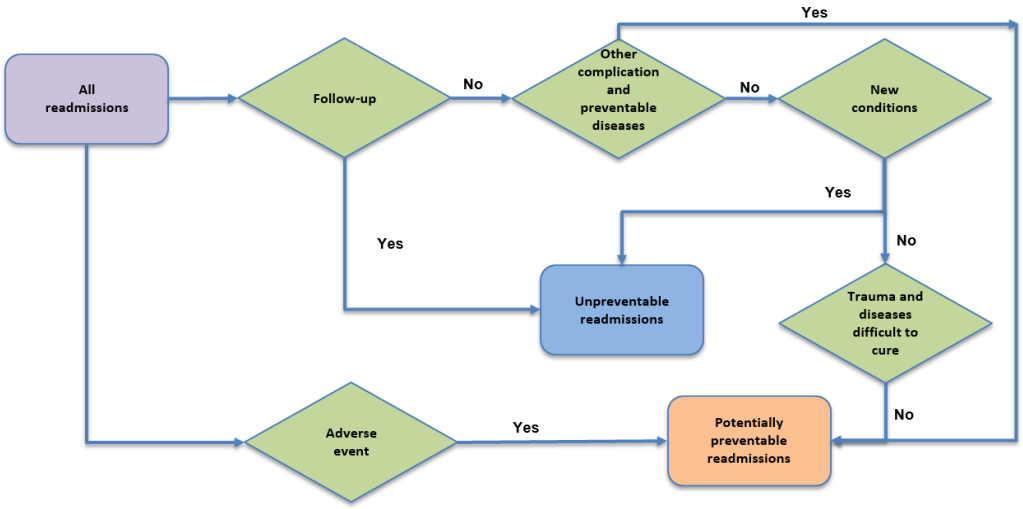


**Figure 2 Algorithm developed for classifying PPRs within 30 days after the initial discharge.**

As Figure 2 shows, the algorithm to classify PPRs in this study (called "PPR algorithm") is stated as follows:

Step1: Apply the AIHW method (unplanned readmissions) to classify the readmissions with some adverse events as PPRs (see Appendix 1: Unplanned readmissions).

Step 2: Exclude the readmissions with follow-up treatments and classify them as unpreventable readmissions. Follow-up care includes the treatment follow-up, rehabilitation, and procedures not carried out in the main index admission.

Step 3: For the readmissions that neither have follow-ups nor adverse events, classify those with some complications that are not classified as adverse events (called "other health condition" in this study) and preventable diseases as PPRs. The other health complications include conditions such as cardiac arrest with successful resuscitation, cardiogenic shock, and septic shock (see Appendix 2: Other health complications). Preventable diseases include deep vein thrombosis, pulmonary embolism, and decubitus ulcer (see Appendix 3: Preventable diseases).

Step 4: For the readmissions that are excluded by steps 1 to 3 (the readmissions that do not have follow-ups, adverse events, other health complications or preventable diseases), the algorithm judges whether a readmission has new conditions, and if so, the readmission is classified as an unpreventable readmission. A readmission that has new conditions is defined as a readmission whose principal diagnosis is not related to any diagnosis already known during the index admission (SQLape 2016). Conversely, a readmission without a new condition is that in which its principal diagnosis is related to at least one diagnosis already known during the index admission.

Step 5: For the readmissions that are excluded by steps 1 to 4 (the readmissions that do not have follow-ups, adverse events, other health complications or preventable diseases, or new conditions), the algorithm classifies a readmission that is not related to a trauma or diseases that are difficult to cure as a PPR (see Appendix 4: Trauma and diseases that are difficult to cure). The remaining readmissions are classified as unpreventable readmissions.

### 3.2 Modelling for identifying risk factors of potentially preventable readmissions

Much of the research in hospital readmissions has been focused on identifying risk factors and building prediction models (Yam et al. 2010; Allaudeen et al. 2011; Renton et al. 2011; Nahab et al. 2012; Donze et al. 2013). Logistic regression (LR) is the most commonly used method to predict the risk of readmissions in prior research (Zeng et al. 2014). There are a number of reasons for the popularity of this method. First, logistic regression is perfectly designed to deal with discrete choice models, where the dependent variable can be coded either as zero or one. Further, logistic regression can deal with multiple explanatory variables which are either dichotomous, ordinal, continuous variables and easily allows for the inclusion of interaction terms. LR also doesn't require error terms to be normally distributed and can handle nonlinear effects. Finally, the associations between each explanatory variable and the outcome variable can easily be interpreted through the estimated coefficients and the odds ratio. Given these advantages, we follow prior research and apply a LR model to examine the relationship between the index admissions with PPRs and the potential risk factors. A Wald z-statistic was used to test the statistical significance of each coefficient ($\boldsymbol{\beta}$) in the model. The level of significance was set at 0.05 with a confidence interval of 95%.

LR weights the independent predictor variables and assigns a $Y$ score to each patient in a form of probability of having a PPR within 30 days after the index admission. In the LR model, the incidence of a PPR for patient i, $y_i$ , is assumed as a binary outcome, and follows a Binomial distribution.

$Y_i \sim Binomial\ (p_i)$,

where $p_i$ is the probability of readmission of patient i.

For example, let $Y = 1$ denote the patient i encounters a PPR, and $Y = 0$ denote the patient does not encounter a PPR or a readmission. Then the probability that patient i encounters a readmission is denoted as

$$P(Y = 1|x_1, \dots, x_k) = f\ (x_1, \dots, x_k) \tag{1}$$

The function $f$ denotes the logistic distribution function as below:

$$P(Y = 1|x_1, \dots, x_k) = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)} \tag{2}$$

From (1) to (2), the regression is transformed into the interval (0, 1), and the $logit(x_i)$ is further defined as:

$$logit(x) = \log(\frac{x}{1 - x}) \tag{3}$$

The model can be rewritten as:

$$logit(P(Y = 1|x_1, \dots, x_k)) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \tag{4}$$

where $\beta_0, \beta_1, \dots, \beta_k$ are real constants, and $\beta_k$ represents the effects of predictor $x_k$ (Trueck and Rachev 2009).

In addition, $logit(x)$ is also called log odds for predictor variable x. We apply the exponential function to (4) and set $x_1 = x$, and $x_1 = x$ +1, and then derive (5) and (6) as below:

$$P(Y = 1|x_1 = x, \dots, x_k) = \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) \tag{5}$$

$$P(Y = 1|x_1 = x + 1, \dots, x_k) = \exp[\beta_0 + \beta_1 (x_1 + 1) + \cdots + \beta_k x_k] \tag{6}$$

The effect of the predictor $x_1$ on the probability of $Y$ has different values depending on the value of $x_1$. Figure 3 illustrate an example for how the probability of having a PPR changes as the patient's age.

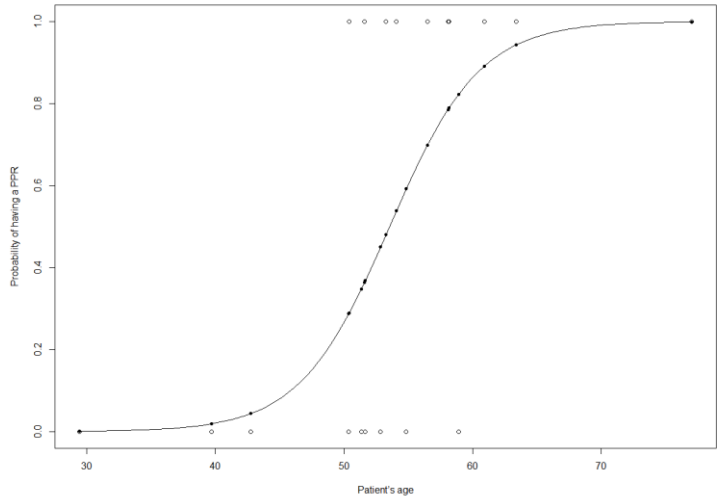


**Figure 3 A logistic regression curve showing probability of having a PPR versus patient's age.**

In order to measure the constant effect of the predictor $x_1$, on the likelihood that a patient will have a PPR, odds ratio is applied in this study.

After (6) is divided by (5), we get the odds ratio for a unit increase in $x_1$, which is also exponential of the coefficient for predictor $x_1$ in logistic regression.

$$\text{Odds ratio for unit increase in } x_1 = \exp(\beta_1) \tag{7}$$

Similarly, for factor $x_k$,

$$\text{Odds ratio for unit increase in } x_k = \exp(\beta_k) \tag{8}$$

There are some advantages for applying LR models. Firstly, the output of LR is always between 0 and 1; the variables can be numeric and/or categorical (Geng 2001). Secondly, given the odds ratio for each variable in the LR model, it is easier to interpret the results. Thirdly, the effect of interaction terms can be assessed in a LR model, which greatly expands understanding of the relationships among the variables (Chungrong and Edward 2003). Finally, LR does not require transformation of continuous variables (Geng 2001). Although a LR model has some attractive advantages, there are some disadvantages: it could suffer from over-fitting; it does not select important variables automatically; and it does not handle imbalanced data effectively (Geng 2001).

### 3.3 Model evaluation

There are various methods of model evaluation for different approaches to modelling. McFadden's R-square was recommended in a study addressing measures of fit for LR (Allisonet 2014).

McFadden's R-square is defined as:

$$R^2_{McF} = 1 - \ln(L_M) / \ln(L_0), \quad (9)$$

where the $L_M$ is the likelihood for the model being estimated, $L_0$ is the likelihood for the null model (the model only including intercept), ln() is the natural logarithm, and n is the sample size (Allisonet 2014).

McFadden's R-square suggests how much improvement is offered by the full model over the null model. The value of McFadden's R-square falls between 0 and 1, and a higher value indicates a better model fit.

However, not every study measured how well LR fits the data; some studies were found to apply Hosmer-Lemeshow (HL) to compute a goodness-of-fit statistic (Ferraris et al. 2001; Silverstein et al. 2008; Bradley et al. 2013). The test assesses whether or not the observed rates of readmission matched the expected rates of readmission in subgroups of the model population. Its statistic is obtained by calculating the Pearson chi-square and large values of Hosmer-Lemeshow statistic (and small p-values) indicate a lack of fit of the model. A p-value below 0.05 suggests that the model needs some interactions or non-linearities to improve the fitness (Allisonet 2014).

Some studies used k-fold cross validation to assess the performance of the model (Lin 2008; Hosseinzadeh et al. 2013; Zolfaghar et al. 2013). The researchers divided the data into k equal-size subsets, and built the models k times. They left out one subset from the training set (the subsets for training the model), and used it as the testing set each time. As this method averages the data set over k different subsets, it reduces the variance caused by different partitioning of the data.

Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) were commonly used for model evaluation in the studies of readmission prediction, regardless of

the regression, machine learning or other approaches applied (Hosseinzadeh et al. 2013; Robin et al. 2013; Shams et al. 2014; Donze et al. 2016).

ROC curve presents the model's true positive rate (also known as sensitivity) on Y-axis and false positive rate (also known as 1- specificity) on X-axis visually for varying cut-off rates. True positive rate (TPR) was used to measure the proportion of the event (PPRs) that are correctly identified as such by the model; false positive rate (FPR) was used to measure the proportion of the non-events (non-PPRs or no readmissions) that were identified as events (PPRs). The overall accuracy was applied to measure the proportion of both the events (PPRs) and non-events (non-PPRs or no readmissions) that are correctly identified as such by the model.

AUC assesses the overall predictive accuracy of model. AUC of 1 (maximum) represents a model is perfect in differentiating the events (PPRs) and non-events (non-PPRs or no readmissions). The larger the AUC, the higher is overall predictive accuracy of model, and also the better is overall performance of model. AUC higher than 0.9 represents an excellent model (see Figure 4 model A); and AUC between 0.8 and 0.9 represents a good model (see Figure 4 model B); and AUC between 0.7 and 0.8 represents a fair model (see Figure 4 model C); and AUC under 0.7 represents a poor model (see Figure 4 model D) (Karimollah 2013).

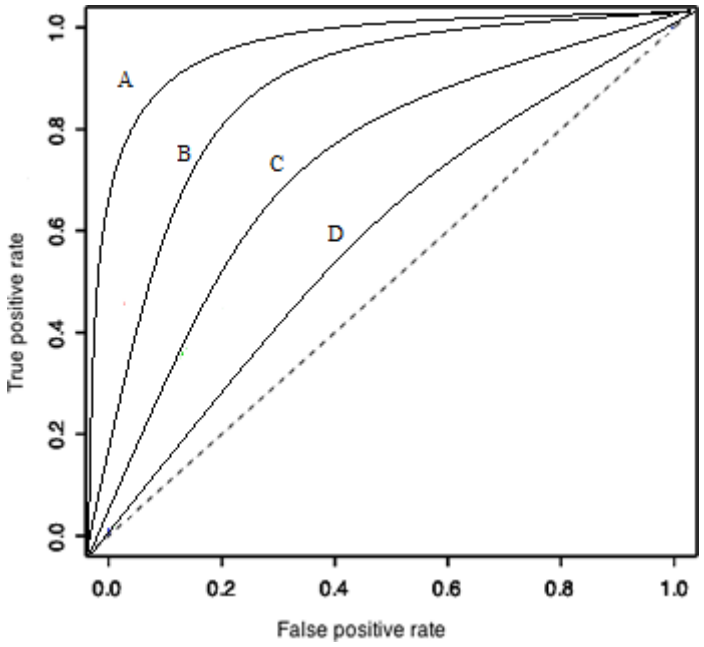


**Figure 4 Plot for ROC curves indicating excellent model (A) , good model (B), fair model (C) and model (D).**

In this study, we applied ROC curve (AUC) to represent the model's TPR (sensitivity) and FPR (1- specificity) visually and understand the overall predictive power of the model. For the diagnoses of testing models (see Section 5.2.3), besides AUC, McFadden's R-square and Hosmer-Lemeshow were also applied to compare how well each testing model fitted the data.

# 4 Data and hypothesis development

In this chapter, Section 4.1 presents description of the data, Section 4.2 defines key variables, Section 4.3 describes data processing and Section 4.4 presents the hypotheses development.

## 4.1 The data

This retrospective cohort study used a database of six years of claims data (from calendar years 2010 to 2016) on inpatients from a private health insurance fund in Australia. Identifiable information about patients and hospitals had been removed, and instead there are unique identifiers to distinguish different hospitals (hospital ID), patients (member ID) and admissions (admission number). The data includes clinical, demographic and cost-related information for privately insured admitted patient services, and the variables were arranged into four groups in this study: hospital-level, patient-level, clinical variables and cost-related variables.

The hospital-level variables are shown in Table 1.

**Table 1 Hospital-level variables in the data set**

| **Hospital-level variables** |
|---|
| Hospital identifier number |
| Hospital type (private or public hospital) |
| Hospital contract status (the index for contracted or non-contracted hospitals with the health insurance fund) |
| Payment model between the given hospital and fund |

Due to confidentiality agreement, the actual payment models between the given hospital and the fund (the reimbursement model agreed between hospital and fund), as well as relating values such as frequency and proportions, cannot be listed in the thesis. So payment model A, B, C and D were used to represent the actual payment models in the database.

The patient-level variables are shown in Table 2.

**Table 2 Patient-level variables in the data set**

| **Patient-level variables** |
|---|
| Patient identifier number |
| Patient's age |
| Gender |

The clinical variables are shown in Table 3.

**Table 3 Clinical variables in the data set**

| **Clinical-level variables** |
|---|
| Admission identifier number |
| Admission date |
| Separation date |
| Diagnosis codes |
| Procedure codes |
| MBS codes |
| DRG codes |
| LOS |
| ICU days |
| Charlson index |
| Minute of operating in theatre |
| Mode of separation |

The cost-related variables include are shown in Table 4.

**Table 4 Cost-related variables in the data set**

| **Cost-related variables** |
|---|
| Hospital benefits |
| Hospital fees |
| Medical benefits |
| Medical fees |
| Prosthesis benefits |
| Prosthesis fees |
| Hospital out-of-pockets |
| Medical out-of-pockets |
| Total schedule fees |
| Principal MBS Schedule fees |

Due to confidentiality agreement, the actual values, and the mean values for the cost-related variables cannot be presented in the thesis.

## 4.2 Definition of key variables

This section describes the key variables used in the analysis, including ICD-10 (diagnosis codes and procedure codes), MBS codes, DRG codes, Charlson index and cost-related variables.

### 4.2.1 Diagnoses codes (ICD-10-AM) and procedure sentences (ACHI)

In the database ICD-10-AM codes are used as diagnosis codes for reporting each patient's diseases and ICD-10-ACHI codes are used as procedure codes for reporting each patient's procedures.

The International Classification of Diseases (ICD) was developed by the World Health Organization as a diagnostic classification standard for reporting diseases and health conditions for all clinical and research purposes (WHO, 2016). Some countries have their own particular versions. In Australia the classification system is ICD-10-AM (Australian Modification), ACHI (Australian Classification of Health Interventions) and ACS (Australian Coding Standards). ICD-10-AM is very similar to the ICD 10th version with some modifications. The codes were developed by the National Centre for Classification in Health for Australian clinical practice (Australian Consortium for Classification Development 2016). "ICD-10-AM uses an alphanumeric coding scheme for diseases and external causes of injury. It is structured by body system and aetiology, and comprises three, four and five character categories" (Australian Consortium for Classification Development 2016).

The first letter of an ICD-10-AM code represents the disease classification. For example, in the code "I01.0", "I" indicates "Diseases of the circulatory system", and the first letter and the following two digits together, "I01", indicate the sub-classification of disease "Rheumatic fever with heart involvement"; the whole code "I01.0" indicates "Rheumatic fever with heart involvement with acute rheumatic pericarditis" (Australian Institute of Health and Welfare 2016).

ICD-10-ACHI codes are the Australian Classification of Health Interventions developed by National Centre for Classification in Health based on the Medicare Benefits Schedule (MBS). They are used in conjunction with ICD-10-AM as the national standard in Australian hospitals for reporting patients' procedures. They have seven digits including the five digits and a two-digit extension. The first five digits are the same as MBS codes but not applied to every case. The first five digits also indicate the principal axis and the two-digit extension represents specific interventions included in that item. For example, in the ICD-10-AHCI code "41629-00", the first five digits "41629" indicate "middle ear", the two-digit extension "00" indicates "examination", and the whole code "41629-00" indicates "exploration of middle ear" (Cumerlato 2012; Australian Consortium for Classification Development 2016).

In the analysis of this research, the diagnosis codes (ICD-10-AM) were grouped as the first letter and the following two digits to simplify the classification of a patient's diseases, as well as reduce the number of levels of principal or additional diagnosis code as a categorical variable. For example, we grouped codes I500-I509 as 'I50', which refers to heart failure for principal diagnosis.

Similarly, the procedure codes (ICD-10-AHCI) are grouped as the first five digits to indicate the principal axis of the given procedure. For example, we grouped codes 3209000-32090002 as "32090", which refers to Fibreoptic colonoscopy to caecum for principal procedure.

### 4.2.2 Principal Medicare Benefits Schedule

"Principal MBS is a Medicare Benefits Schedule (MBS) item number and selected on the basis of: (a) the MBS item number with the highest schedule; and (b) the patient's first visit to a theatre or procedure room/coronary angiography suite. It does not relate to the medical item billed by the doctor. It may not necessarily correlate to the Principal Procedure Code" (Hospital to Department data specifications 2016).

MBS is a listing of the Medicare services subsidised by the Australian Government. The MBS item number indicates the relevant Medicare service and they are usually used by hospitals or doctors to bill to Medicare for their relevant service. They are numeric basis with five digits. An MBS item number is also assigned with its MBS category and MBS subgroup. For example, MBS item number "49561", which indicates a knee, arthroscopic surgery, belongs to MBS Category "3" (Therapeutic procedures), and MBS subgroup "T8" (Surgical operations) (Department of Health 2016). Thus based on the MBS item number, the corresponding MBS category and MBS subgroup are created as two new variables for each admission in the database.

### 4.2.3 Diagnosis Related Groups codes

The Diagnosis Related Groups (DRG) in our database are the Australian version, AR-DRG codes (Australian Refined Diagnosis Related Groups), which. AR-DRG is an inpatient classification system that provides a clinically meaningful way of relating the number and types of patients treated in a hospital to the resources required by the hospital. It takes into

account the different mix of cases treated in hospitals, which provides a basis for service planning and financing.

DRG is assigned by grouper software to a patient considering information on patient's age and gender, LOS, same day status, admission weight for infants aged less than 365 days, mental health legal status, mode of separation, principal diagnosis, other primary diagnoses, comorbidities relevant to the admission, complications, and procedures performed (both therapeutic and diagnostic).

An AR-DRG code is structured with one letter, two digits in the middle and a letter at the end, for example, "E65A". The first letter indicates the major diagnostic category; the two digits in the middle indicate the partition to which the DRG belongs, such as 1-39 for surgical, 40-59 for other and 60-99 for medical; the last letter ranks the resource consumption associated with the patient's treatment, which has five levels: "A" for "highest consumption", "B" for "second highest", "C" for "third highest", "D" for "fourth highest" and "Z" for "no split" (WA Department of Health Clinical Casemix Handbook 2014).

Given the different meaning of the three parts of an AR-DRG code, a DRG for the given admission in the database is split as DRG group 1 and group 2. DRG group 1 is the first letter with two digits in the middle from the original code, which together indicate the patient's major diagnostic category and the associated partition; and DRG group 2 is the last letter indicating the resource consumption associated with the patient's treatment. Splitting DRG into two separate groups avoids the problem that the model may not handle excessive levels for a categorical variable, but also simplifies integrating DRG codes that contain multiple information.

#### 4.2.4 Charlson index

Charlson index is one of the most commonly used comorbidity measures for hospital administrative data (Elixhauser et al. 1998). It categorises comorbidities of patients based on the ICD (International Classification of Diseases) diagnosis codes at four levels (1, 2, 3 and 6) to indicate the risk of 1-year mortality for a patient. Zero means the given patient has no comorbidities, and the higher score indicates higher risk of mortality or higher resource use (Manitoba Centre for Health Policy 2016).

### 4.2.5 Cost-related variables

There are four cost-related items associated with a patient or member's admission: benefits, fees, schedule fees and out-of-pockets. Benefits and fees are arranged into three high level variables in the database: hospital benefits (fees); medical benefits (fees); and prosthesis benefits (fees).

Hospital fees are the charges billed by the hospital for a given admission, which include cost of accommodation and theatre; while hospital benefits are the amount paid by the health fund to the hospital for the admission. The fees are not always fully covered by the health fund, which means the hospital benefits are equal to or less than hospital fees.

Medical fees are the charges billed by doctors for their service associated with a patient's admission; and medical benefits are the payments from the health insurance fund to the doctor for the admission.

MBS (Medicare Benefits Schedule) codes are used by doctors to bill to Medicare and the schedule fee for each code is determined by Medicare and is also publicly available. There could be more than one MBS code applied for a given admission. Medical benefits, medical fees and total schedule fees are the aggregated benefits, fees, and schedule fees for all MBS codes used by doctors for the admission.

Medical fees are usually higher than the total schedule fees for a given admission, as doctors are able to charge for their services. Medicare only covers 75% of the schedule fee for each MBS code. The health fund will cover all or part of the remaining fee if the health insurance fund has Gap Cover Doctors agreements with the doctor. Otherwise the remaining portion of medical fees has to be paid out of a patient's own pocket (see Equation (11)).

Prosthesis benefits are the payments from the health fund for prostheses. Prosthesis fees reflect the gross maximum charge raised for a prosthesis. For each procedure, how much the health insurance fund will pay for a particular prosthesis and whether a patient will have any 'gap' to pay depends on the individual policy.

There are two types of out-of-pockets costs: hospital out-of-pockets; and medical out-of-pockets. The functions for calculating out-of-pockets are:

$$\text{Hospital out-of-pockets} = \text{Hospital fees} - \text{Hospital benefits} \quad (10)$$

$$\text{Medical out-of-pockets} = \text{Medical fees} - 75\%\text{Total schedule fees} - \text{Medical benefits} \quad (11)$$

$$\begin{aligned}\text{Total out-of-pockets} &= \text{Hospital out-of-pockets} + \text{Medical out-of-pockets} \quad (12)\\ &= (\text{Hospital fees} + \text{Medical fees}) - \\ &\quad (\text{Hospital benefit} + \text{Medical benefit} + 75\%\ \text{schedule fee})\\ &= \text{Total fees} - \text{Total benefits}\end{aligned}$$

The total schedule fees are the sum of schedule fee of each MBS code associated with a given admission. The principal MBS schedule fee is the schedule fee for the principal MBS only.

To sum up, for every admission (all index admissions and readmissions), the cost-related variables are in dollars and non-negative. They include: hospital benefits, medical benefits, prosthesis benefits, hospital fees, medical fees, prosthesis fees, hospital out-of-pockets, medical out-of-pockets, total schedule fees and principal MBS fee.

### 4.3 Data processing

This section describes data processing for the analysis. Figure 5 presents the process.

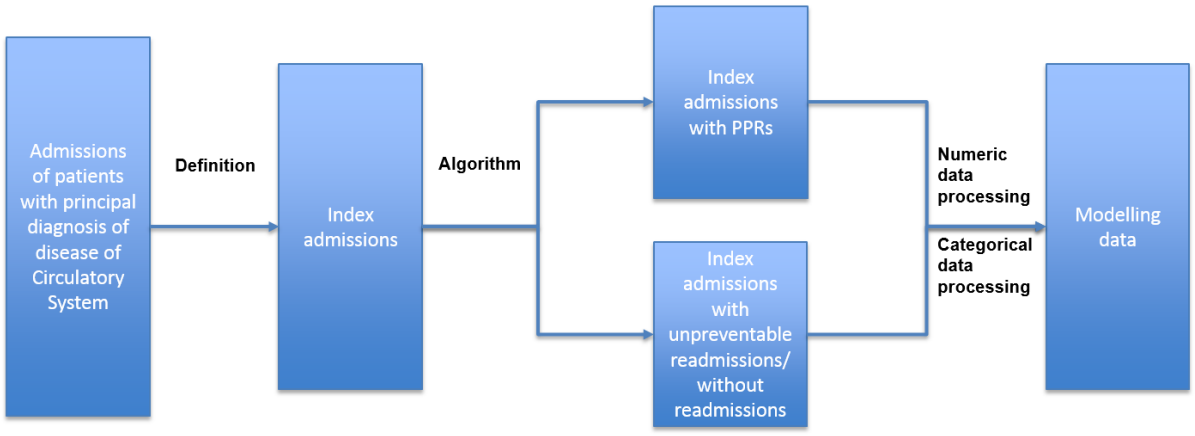


**Figure 5 Diagram for the steps of processing the original data set to the modelling data set.**

As Figure 5 shows, firstly, eligible admissions and readmissions were selected based on our definitions for them (see Section 4.3.1); secondly, PPRs were identified by PPR algorithm (see Section 4.3.1); and thirdly, numeric data and categorical data were processed separately (see Sections 4.3.2 and 4.3.3) for analysis and modelling in the next stage.

### 4.3.1 Selecting eligible admissions and readmissions

The first step of data processing was extracting the data associated with patients who had diseases of the circulatory system in the initial discharge for admissions to hospital between January 2010 and April 2016. The data set contained 21,009 records that correspond to 12,804 individual patients with principal discharge diagnoses of diseases of the circulatory system. These patients were identified by principal diagnosis code (ICD-10-AM) starting with "I". Following the method of Agency for Healthcare Research and Quality, we excluded patients who died, transferred to another institute, or left against medical advice in the initial hospital stay. We defined a readmission as a return to hospital within 30 days after discharge of index admission. The readmissions were all-cause hospitalisations, and each readmission could only have one corresponding index admission.

After the first step, there were 9,627 index admissions and 2,021 readmissions that correspond to 7,044 individual patients with principal discharge diagnoses of diseases of the circulatory system.

Next, PPR algorithm was applied to the processed data from the last step, and a new categorical variable for indicating PPR is constructed, labelled 1 for PPR, and 0 for non-PPR or no readmission following a given index admission. We found 850 PPRs out of 2,021 readmissions (42.06% among readmissions, and 8.83% among all admissions).

Variables used for the analysis in this study are shown in Table 5.

**Table 5 Selected variables for analysis**

| |
|---|
| Age |
| Gender |
| Principal diagnosis group |
| Secondary diagnosis group |
| MBS category |

| |
|---|
| MBS group |
| LOS |
| ICU days |
| Minutes in operating theatre |
| Charlson index |
| DRG group 1 |
| DRG group 2 |
| Procedure group |
| Hospital fees |
| Medical fees |
| Prosthesis fees |
| Hospital benefits |
| Medical benefits |
| Prosthesis benefits |
| Hospital out-of-pockets |
| Medical out-of-pockets |
| Total out-of-pockets |
| Total fees |
| Total benefits |
| Total schedule fees |
| Principal MBS schedule fee |

### 4.3.2 Processing numeric variables

Missing values for the age, LOS and Charlson index were imputed with the median of the given variable. The extreme values of ICU days over 30 days and LOS over 90 days were removed. In addition, the observations with implausible numeric variables were dropped from the data set; for example, negative age, negative hospital fees or medical fees.

### 4.3.3 Processing categorical variables

The model may not handle too many levels for some categorical variables. Some were regrouped as the top 50 categories with highest frequencies (at least 10 observations) and the rest as “others” for each variable including principal diagnosis group (the first letter and the following two digits from the original code), secondary diagnosis group (the first letter and the following two digits from the original code), principal procedure group (the first five digits from the original code), DRG group 1 (the first letter with two digits in the middle from the original code) and DRG group 2 (the last letter from the original code).

## 4.4 Hypotheses development

To explore the relationships between cost-based variables and the risk of PPRs, the cost-related variables from the original data set were examined. Some cost-based variables were constructed for testing hypothesised relationships, and they were referred to as cost measure variables in this study. This section describes the newly developed cost-measure variables and the hypotheses relating to them.

### Hypothesis 1

Medical fees associated with a patient's admission are the charges billed by doctors and are usually higher than total schedule fees. The gap between medical fees and total schedule fees indicates how much the charge for the medical treatment was above the standard rate, which might also indicate additional treatments for the patient. In order to examine whether the gap influences the risk of a patient being readmitted to hospital within 30 days after the initial discharge, three cost measure variables were developed as follows:

*Cost-measure variable 1.1*

It is the z-score of ratio of medical fees over total schedule fees (see Equation (13)) for a given index admission. The distribution of the ratios was observed left-skewed, and standardising puts them on the same scale [0, 1].

$$\text{Ratio of medical fees over total schedule fee} = \frac{\text{Medical fees}}{\text{Total schedule fees}} \tag{13}$$

*Cost-measure variable 1.2*

It is a dummy variable where '1' indicates Equation (13) for a given index admission is higher than the median among all index admissions with the same DRG; while '0' indicates it is lower than the median. DRG group 1 is selected as the basis to split the hospital fees because they are assigned to each admission by patient's diagnosis, procedures, age and LOS.

*Hypothesis 1*

The more medical fees exceeds the total schedule fees, indicating additional treatments, the less likely it is for the patient to encounter a PPR.

### Hypothesis 2

Hospital fees are the charges billed by the hospital for a given admission. Hospital fees are associated with the cost for accommodation and theatre, which may reflect the additional services (accommodation) provided and the level of complexity of procedure. Hospital fees

are mainly determined by the contract between a hospital and the insurance fund for certain treatments, but it is still worthy to examine whether the hospital fees have an impact on the risk of PPRs for the same diagnosis group (DRG group 1).

*Cost-measure variable 2*

It is a dummy variable where '1' indicates the hospital fees for a given index admission are higher than the median among all index admission with the same DRG; while '0' indicates the hospital fees are lower than the median.

*Hypothesis 2*

If the hospital fees are charged in the higher half among the index admissions with the same diagnosis group (cost-measure variable 2 = "1"), it is less likely for the patient to encounter a PPR.

**Hypothesis 3**

The principal MBS Schedule fee is the schedule fee for the principal MBS. Usually higher schedule fee for principal MBS indicates the higher level of complexity the treatment, or more additional treatments. Cost-measure variable 3 was developed to test the relationship between the principal MBS Schedule Fee and the risk of PPRs.

*Cost-measure variable 3*

It is a dummy variable where '1' indicates the principal MBS schedule fee of a given index admission is higher than the median value for all index admissions, while '0' indicates the principal MBS schedule fee is lower than the median.

*Hypothesis 3*

If the principal MBS Schedule fee is charged in the higher half among the index admissions (cost-measure variable 3 = "1" is significant), it is more likely for the patient to encounter a PPR.

**Hypothesis 4**

Given the principal MBS does not relate to the medical item billed by the doctor, it is selected on the basis of: (a) the patient's first visit to a theatre or procedure room/coronary angiography suite; and (b) the MBS item number with the highest benefit amount (Hospital to Department data specifications 2016). It also reflects the level of complexity for the given treatment. The effect of principal MBS schedule fee on the risk of PPRs may differ for different values of LOS. Regarding this problem, cost-measure variable 4 was developed.

*Cost-measure variable 4*

It is a dummy for interaction of (LOS * principal MBS schedule fee) with value '1' indicating the interaction of LOS and principal MBS schedule fee for a given index admission is higher than median among all index admissions, while '0' indicates the interaction is lower than the median.

*Hypothesis 4*

If the interaction of principal MBS schedule fee times LOS in the higher half among the index admissions (cost-measure variable 4 = "1" is significant), it is more likely for the patient to encounter a PPR.

**Hypothesis 5**

A study found patients with private health insurance are less likely to be readmitted to hospital compared to patients who are only covered by a public care system (Hasan et al. 2010). Following a similar approach, we explore whether patients with better (more comprehensive) health cover are also less likely to have PPRs.

Usually cheaper (less comprehensive) health insurance policies come with exclusions and benefit limitations, and may cause higher out-of-pocket costs to patients; while higher level policies, which are usually more expensive, provide more comprehensive cover and lower out-of-pocket costs. We assumed total out-of-pocket (the gap between total fees and total benefits, see Equation (12)) was a measure of the level of health insurance policy; and the smaller the gap, the higher the level of the policy.

The out-of-pocket is the amount paid by patients either for medical or hospital charges over and above what the patient gets back from Medicare (medical out-of-pocket) or private health insurance (hospital out-of-pocket). Cost measure variables 5.1, 5.2 and 5.3 were developed to examine the effect of total out-of-pocket on risk of PPRs.

*Cost measure variable 5.1*

It is the z-score of the ratio of total fees over total benefits (see equation (14)) for a given index admission. The distribution of the ratios is observed left-skewed, and standardising puts them on the same scale [0, 1].

$$\text{Ratio of total fees over total benefits} = \frac{\text{Total fees}}{\text{Total benefits}} \tag{14}$$

*Cost-measure variable 5.2*

It is a dummy variable where '1' indicates the ratio of total fees over total benefits for a given index admission is higher than the median among all index admission with the same DRG; while '0' indicates the ratio is lower than the median.

*Hypothesis 5*

The higher the level of policy the patient holds, the less likely the patient encounters a PPR.

For the newly constructed cost-measure variables, such as interaction of LOS and principal MBS schedule fees, ratio of medical fee to schedule fee, we cut off the records with 1% highest and 1% lowest value for each variable.

# 5 Empirical Analysis

The purpose of this chapter is to present and discuss the results of the study. It includes four sections. Section 5.1 will provide descriptive statistics for the data as well as results for some preliminary data analysis for the sample. In section 5.2, we will discuss the applied models for identifying risk factors of PPRs and provide results for the estimated univariate and multiple logistic regression models. Section 5.3 contains model evaluation results for the proposed model that has administrative, clinical as well as cost-related explanatory variables. Section 5.4 will then provide a discussion of the results, and compares them to findings from prior studies in the literature.

## 5.1 Preliminary data analysis

In the following we provide descriptive statistics and results for a conducted preliminary data analysis that is used in this study. For quantitative variables, we report the minimum, first and third quantiles, median, mean, maximum and the number of missing values. For categorical variables, we report the frequency and percentages for the values that the variable can take on..

Pearson's correlation and the Goodman-Kruskal gamma statistic (Goodman and Kruskal 1972) were applied to measure the associations for the continuous and categorical cost-related/cost-measure variables respectively.

In order to investigate the relationship between PPRs and each variable, a Chi-square test, two-sample t-tests and univariate regression with a Wald test were applied. Tests were conducted at a 5% level of significance. The Chi-square test was used to examine the association between categorical variables and PPRs; for continuous variables, a t-test was used to examine the difference between patients that had PPRs and patients that were classified as not having a PPR or did not have a readmission.

After processing the data, there were 7,044 unique patients with 9,627 admissions caused by circulatory system diseases from January 2010 to April 2016. Of those, 2,021 index admissions were followed by a readmission, and 7,606 were not. Among the 2,021 readmissions, 850 PPRs were identified by the PPR algorithm (42.06% out of the

readmissions, and 11.18% out of all admissions). The proportion of PPRs among the readmissions was in a range similar to the ones reported by other studies, where the rate of PPRs ranges from 21% (Dawes et al. 2014) up to 76% (Ganguly et al. 2013).

The overall average LOS was 2.99 days and average age of the patients was 66.54 years. There were 2,909 females with an average age of 66.63 years, and 4,135 males with an average age of 66.49 years. Given that the claim data was provided by a private health insurance fund, most of the index admissions occurred in private hospitals (n=9,233, 96.0%), whereas the rest occurred either in public hospitals (n=123, 1.28%) or same-day hospitals (n=271,2.82%).

The top 5 principal diagnosis with the highest frequencies in index admissions were:

1) *Chronic ischaemic heart disease* (n=1681, 17.46%),
2) *Unspecified Atrial fibrillation and atrial flutter* (n=1219, 12.66%),
3) *Internal thrombosed haemorrhoids* (n=1077, 11.19%),
4) *Unstable angina* (n=987, 10.25%),
   and
5) *Varicose veins of lower extremities* (n= 772, 8.02%).

A per secondary diagnosis, there were 2,270 index admissions whose record does not contain a secondary diagnosis. The top 5 most frequent ones were:

1) *Chronic ischaemic heart disease* (n=1468, 15.25%),
2) *Personal history of certain other diseases* (n=1015, 10.54%),
3) *Essential (primary) hypertension* (n=654, 6.80%),
4) *Presence of cardiac and vascular implants and grafts* (n= 350, 3.64%)
   and
5) *Type 2 diabetes mellitus* (n=323, 3.36%).

In order to compare the difference in the patient-level, clinical, cost-related variables between the index admissions with PPRs and the index admissions followed with non-PPRs or no readmissions were compared. A summary of the results is provided in Tables 6-9. The tables provide statistics on the mean, standard deviation (SD) and p-values for the conducted t-tests for continuous variables; while frequency (Freq), proportion and p-values of conducted Chi-square tests for categorical variables are presented.

Table 6 illustrates the description of patient-level factors for the index admission with PPRs and the index admissions with non-PPRs or no readmissions. There were significant differences between the two groups for age (0.0000) and gender (p-value = 0.0051). The average age of the patients who encountered PPRs (mean=66.19) was around 4 years older than the age of patients who did not have a PPR (mean=70.21); and males were more frequently to have a PPR.

**Table 6 Description for patient-level variables in the index admissions with PPRs and the index admissions with non-PPRs or no readmissions**

| Variables | | Non-PPRs or No readmissions | | | | PPRs | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | n= 8777 | | | | n = 850 | | | | |
| | | Mean | SD | Freq | Proportion | Mean | SD | Freq | Proportion | P-value |
| **Age** | | 66.19 | 14.05 | | | 70.21 | 11.39 | | | 0.0000 |
| **Gender** | Male | | | 5334 | 61% | | | 558 | 66% | 0.0051 |
| | Female | | | 3443 | 39% | | | 292 | 34% | |

Table 7 presents the descriptive statistics for the clinical variables among the index admissions with PPRs and the index admissions followed with non-PPRs or no readmissions. It was noted that the patients who encounter PPRs typically had longer LOS (in days) (mean = 3.63, P-value = 0.0007), a shorter time in the operating theatre (in minutes) (mean = 61.34, P-value = 0.0001), and a higher Charlson index (mean = 4.07, P-value = 0.0000). Additionally, MBS category *therapeutic procedures* and MBS group surgical operation showed larger proportion in the group of index admissions without PPRs or readmissions. *Coronary angiography with heart catheterisation* was noted as the most frequent procedure among the index admission with PPRs, which also showed 14% higher than the proportion for index admission without PPRs or readmissions. The proportion of *chronic ischaemic heart diseases* for the index admissions with PPRs was found slightly higher in both principal (proportion = 22%) and secondary diagnosis (proportion = 21%).

There were significant differences between non-PPRs or no readmissions group and the PPRs group for LOS (p-value =0.0007 ), ICU days(p-value = 0.01115), Minutes in operating theatre(p-value = 0.0001), Charlson index (p-value = 0.0000), payment model (p-value = 0.0000), MBS category (p-value = 0.0000), MBS group (p-value = 0.0000), resources taken to the treatment (p-value =0.0000), DRG group 1 (p-value = 0.0000), principal diagnosis

group (p-value =0.0000), secondary diagnosis group (p-value = 0.0000), and principal procedure group (p-value = 0.0000). No significant differences were noted with regards to the index of ICU (p-value = 0.0640).

**Table 7 Description for clinical variables in the index admissions with PPRs and the index admissions with non-PPRs or no readmissions**

| Variables | Non-PPRs or No readmissions | | | | PPRs | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Freq | Proportion | Mean | SD | Freq | Proportion | P-value |
| **LOS** | 2.92 | 3.99 | | | 3.63 | 5.62 | | | 0.0007 |
| **ICU days** | 0.17 | 0.71 | | | 0.18 | 1.23 | | | 0.0115 |
| **Minutes in operating theatre** | 71.17 | 95.45 | | | 61.34 | 75.70 | | | 0.0001 |
| **Charlson index** | 3.40 | 1.80 | | | 4.07 | 2.00 | | | 0.0000 |
| **Payment model** | | | | | | | | | 0.0000 |
| **Index of ICU** | | | | | | | | | 0.0640 |
| Patient did NOT have ICU in index admission | | | 7986 | 91% | | | 790 | 93% | |
| Patient have ICU in index admission | | | 791 | 9% | | | 60 | 7% | |
| **MBS category in index admission** | | | | | | | | | |
| Therapeutic Procedures | | | 7235 | 82% | | | 576 | 68% | 0.0000 |
| Diagnostic Imaging Services | | | 961 | 11% | | | 183 | 22% | |
| Professional Attendances | | | 507 | 6% | | | 74 | 9% | |
| Pathology Services | | | 56 | 1% | | | 13 | 2% | |
| Diagnostic Procedures | | | 18 | 0% | | | 4 | 0% | |
| **MBS group** | | | | | | | | | |
| Surgical Operations | | | 6756 | 77% | | | 539 | 63% | 0.0000 |
| Diagnostic Radiology | | | 397 | 5% | | | 110 | 13% | |
| Consultant Physician Attendances | | | 444 | 5% | | | 66 | 8% | |
| Ultrasound | | | 358 | 4% | | | 36 | 4% | |
| Computed Tomography | | | 130 | 1% | | | 29 | 3% | |
| Miscellaneous Therapeutic Procedures | | | 309 | 4% | | | 25 | 3% | |
| Relative Value Guide For Anaesthesia - Medicare Benefits Are Only Payable For Anaesthesia Performed In Association With An Eligible Service | | | 156 | 2% | | | 11 | 1% | |
| Others | | | 227 | 3% | | | 34 | 4% | |
| **Resource taken in the initial treatment** | | | | | | | | | |
| Highest Consumption | | | 1343 | 15% | | | 146 | 17% | 0.0000 |
| Second Highest Consumption | | | 3980 | 45% | | | 489 | 58% | |
| Third Highest Consumption | | | 1077 | 12% | | | 77 | 9% | |
| Fourth Highest Consumption | | | 2 | 0% | | | 0 | 0% | |
| No split | | | 2375 | 27% | | | 138 | 16% | |

The actual payment model cannot be listed due to confidentiality agreement.

(table 7 continued)

| Variables | Non-PPRs or No readmissions | | PPRs | | |
|---|---|---|---|---|---|
| | Freq | Proportion | Freq | Proportion | P-value |
| **DRG Group 1** | 2341 | 27% | 334 | 39% | 0.0000 |
| Circulatory System - other partitions | 3295 | 38% | 214 | 25% | |
| Circulatory System - surgical partitions | 1539 | 18% | 203 | 24% | |
| Circulatory System - medical partitions | 139 | 2% | 55 | 6% | |
| Nervous System - medical partitions | 152 | 2% | 12 | 1% | |
| Nervous System - surgical partitions | 582 | 7% | 5 | 1% | |
| Digestive System - other partitions | 523 | 6% | 6 | 1% | |
| Digestive System - surgical partitions | 206 | 2% | 21 | 2% | |
| Others | | | | | |
| | | | | | |
| **Principal diagnosis in index admission** | | | | | |
| Chronic ischaemic heart disease | 1495 | 17% | 186 | 22% | 0.0000 |
| Angina pectoris or Unstable angina | 885 | 10% | 102 | 12% | |
| Atrial fibrillation and flutter | 1132 | 13% | 87 | 10% | |
| Atherosclerosis | 533 | 6% | 79 | 9% | |
| Nonrheumatic aortic valve disorders | 198 | 2% | 61 | 7% | |
| Others | 4534 | 52% | 335 | 39% | |
| | | | | | |
| **Secondary diagnosis in index admission** | | | | | |
| Chronic ischaemic heart disease | 1289 | 15% | 179 | 21% | 0.0000 |
| No Secondary Diagnosis | 2137 | 24% | 133 | 16% | |
| Personal history of certain other diseases | 936 | 11% | 79 | 9% | |
| Essential (primary) hypertension | 587 | 7% | 67 | 8% | |
| Presence of cardiac and vascular implants and grafts | 317 | 4% | 33 | 4% | |
| Others | 3511 | 40% | 24 | 42% | |
| | | | | | |
| **Princpal procedure in index admission** | | | | | |
| Coronary angiography with left or/and right heart catheterisation | 2057 | 23% | 314 | 37% | 0.0000 |
| Allied health intervention | 658 | 7% | 101 | 12% | |
| Percutaneous/open insertion of 1 or >2 transluminal stent into single/multiple coronary artery | 779 | 9% | 56 | 7% | |
| Digital subtraction angiography of aorta and lower limb, >= 10 data acquisition runs, unilateral/bilateral | 61 | 1% | 29 | 3% | |
| Lung ventilation study | 524 | 6% | 28 | 3% | |
| Others | 4698 | 54% | 322 | 38% | |

Due to confidentiality agreements with the health fund, the mean and SD for the cost-related variables (the variables associated with fees and benefits originally from the data set) cannot be presented here. Therefore, Table 8 only provides the test statistic and p-values for the conducted t-test for these variables.

As Table 8 represents, there were significant differences between non-PPRs or no readmissions group and the PPRs group for total schedule fees (p-value = 0.0588), ICU charge (p-value = 0.0080), medical fees (p-value = 0.0178), prosthesis fees (p-value = 0.0006), medical benefits (p-value = 0.0046), prosthesis benefits (p-value =0.0006), hospital out-of-pockets (p-value = 0.0010), and medical out-of-pockets (p-value = 0.0001). No

significant differences were noted with regards to principal MBS schedule fee (p-value = 0.9235), hospital fees (p-value = 0.6056), and hospital benefits (p-value =0.7073).

**Table 8 Results for conducted t-tests for cost-based measures. We test for a significant difference between patients with a PPRs and those who were not classified as having a PPR or did not have a readmission.**

| Variables | p-value |
|---|---|
| Principal MBS ScheduleFee | 0.9235 |
| Total Schedule Fees | 0.0588 |
| ICU charge | 0.0080 |
| Hospital Fees | 0.6056 |
| Medical Fees | 0.0178 |
| Prosthesis Fees | 0.0006 |
| Hospital Benefits | 0.7073 |
| Medical Benefits | 0.0046 |
| Prosthesis Benefits | 0.0006 |
| Hospital Out-of-Pockets | 0.0010 |
| Medical Out-of-Pockets | 0.0001 |

Table 9 summarised the cost-measure variables with their mean, SD and p-value for t-test. There were significant differences between non-PPRs or no readmissions group and the PPRs group for *cost-measure variable 1.1* (p-value = 0.0001), *3* (p-value = 0.0000) and *5.1* (p-value =.0.0069); while no significant differences were observed for *cost-measure variable 1.2* (p-value = 0.0709) *and 4* (p-value = 0.8452).

**Table 9 Descriptive for cost-measure variables in the index admissions with PPRs and the index admissions with non-PPRs or no readmissions**

| Variables | Non-PPRs or No readmissions | | | | PPRs | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Freq | Proportion | Mean | SD | Freq | Proportion | P-value |
| Cost-measure variable 1.1 | 0.00 | 1.00 | | | -0.23 | 0.70 | | | 0.0001 |
| Cost-measure variable 1.2 | | | 3637 | 51% | | | 334 | 47% | 0.0709 |
| Cost-measure variable 2 | | | 3529 | 50% | | | 347 | 49% | 0.9178 |
| Cost-measure variable 3 | | | 3909 | 55% | | | 339 | 48% | 0.0000 |
| Cost-measure variable 4 | | | 3921 | 55% | | | 391 | 56% | 0.8452 |
| Cost-measure variable 5.1 | 0.03 | 0.97 | | | -0.08 | 1.06 | | | 0.0069 |
| Cost-measure variable 5.2 | | | 3684 | 52% | | | 348 | 49% | 0.2568 |

The Pearson's correlation (a measure of the linear dependence between two variables) between the cost-related/cost-measure variables were checked and outlined in the table below.

**Table 10 Correlation table for cost-related/cost-measure variables**

| | ICU charge | Principal MBS Schedule Fee | Hospital Fees | Hospital Benefits | Medical Fees | Medical Benefits | Prosthesis Fees | Prosthesis Benefits | Hospital Out-of-Pocket | Medical Out-of-Pocket | Schedule Fees | Cost-measure variable 1.1 | Cost-measure variable 5.1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ICU charge** | 1.00 | 0.36 | 0.48 | 0.48 | 0.61 | 0.62 | 0.04 | 0.04 | 0.00 | 0.27 | 0.60 | 0.02 | -0.05 |
| **Principal MBS Schedule Fee** | | 1.00 | 0.44 | 0.44 | 0.61 | 0.62 | 0.14 | 0.14 | 0.08 | 0.25 | 0.62 | -0.11 | -0.03 |
| **Hospital Feess** | | | 1.00 | 1.00 | 0.78 | 0.78 | 0.24 | 0.24 | 0.04 | 0.31 | 0.78 | -0.10 | -0.41 |
| **Hospital Benefits** | | | | 1.00 | 0.78 | 0.78 | 0.24 | 0.24 | 0.00 | 0.30 | 0.78 | -0.10 | -0.42 |
| **Medical Fees** | | | | | 1.00 | 0.98 | 0.22 | 0.22 | 0.04 | 0.51 | 0.98 | 0.01 | -0.10 |
| **Medical Benefits** | | | | | | 1.00 | 0.21 | 0.21 | 0.04 | 0.36 | 0.98 | -0.07 | -0.13 |
| **Prosthesis Fees** | | | | | | | 1.00 | 1.00 | 0.00 | 0.06 | 0.23 | -0.08 | -0.10 |
| **Prosthesis Benefits** | | | | | | | | 1.00 | 0.00 | 0.06 | 0.23 | -0.08 | -0.10 |
| **Hospital Out-of-Pockets** | | | | | | | | | 1.00 | 0.04 | 0.04 | 0.03 | 0.21 |
| **Medical Out-of-Pockets** | | | | | | | | | | 1.00 | 0.37 | 0.55 | 0.10 |
| **Schedule Fees** | | | | | | | | | | | 1.00 | -0.11 | -0.12 |
| **Cost-measure variable 1.1** | | | | | | | | | | | | 1.00 | 0.25 |
| **Cost-measure variable 5.1** | | | | | | | | | | | | | 1.00 |

Table 10 shows that the hospital fees, medical fees, prosthesis benefits were highly correlated with correlations between these variables being greater 0.9. This does not come as a surprise, since one would expect that the amounts paid by the private health fund to hospitals/doctors (benefits) can be expected to be highly related to how much hospitals/doctors charged for their services (fees). Medical fees/benefits and total schedule fees were also highly correlated (correlation > 0.9). Again, this could be expected since schedule fees are "standard fees" for a group of treatments and managed by Medicare. Doctors usually charge a higher amount than suggested by schedule fees for their services (see Section 4.2.5). Hospital fees /benefits also showed high correlations with medical fees /benefits and schedule fees (correlation >0.7). The reason for this is that hospital fees /benefits and medical fees /benefits typically change with the complexity and quantity of treatments associated with a patient's hospitalisation.

The ICU charge and principal MBS schedule fees also exhibited significant correlations with Medical fees, Medical benefits and Schedule Fees (although slightly lower, with a coefficient of correlation > 0.5). A reason for this might be that if a patient received additional and/or complex treatments, more services would be required from hospital and doctors. The medical out-of-pockets (correlation = 0.5079) and cost-measure 1.1 (correlation = 0.5594) also appeared to exhibit significant correlations with medical fees, as medical fees were included in the formulas for calculating these two variables (see Equation (13)).

For the categorical cost-measure variables, the concordance between them were checked by Goodman-Kruskal gamma (GK gamma) statistics and the results are reported in Table 11. GK gamma falls in range [-1, 1]; and if the two variables are highly associated, the estimate of GK gamma should be close to 1 (Gokatas and Isci 2011).

**Table 11 GK gamma for cost-measure variables**

| | **Cost-measure variable 1.2** | **Cost-measure variable 2** | **Cost-measure variable 3** | **Cost-measure variable 4** | **Cost-measure variable 5.2** |
|---|---|---|---|---|---|
| **Cost-measure variable 1.2** | 1.00 | 0.09 | -0.10 | -0.01 | 0.33 |
| **Cost-measure variable 2** | | 1.00 | 0.12 | 0.57 | -0.59 |
| **Cost-measure variable 3** | | | 1.00 | 0.64 | 0.00 |
| **Cost-measure variable 4** | | | | 1.00 | -0.09 |
| **Cost-measure variable 5.2** | | | | | 1.00 |

As the Table 10 shows, the cost-measure variable 3 and 4 had a stronger association (GK gamma = 0.64), which might be caused by that they both included principal MBS schedule fee in their formulas (see Section 4.2.5). *Cost-measure variable 2* appeared to have a slight positive association with *cost-measure variable 4* (GK gamma = 0.57) and a slight negative association with *cost-measure variable 5.2* (GK gamma = -0.59), which might be caused by that the LOS is one of determinates for hospital fees (see Section 4.2.5).

## 5.2 Model estimation and results

### 5.2.1 Applied models

Data were managed and analysed using R (Studio version 3.2.4); the R-packages used in this analysis were RPostgreSQL, data.table, Hmisc, doBy, caret, and pROC.

Simple LR regression was applied to develop both the univariate and multiple regression models in order to examine the relationships between the potential risk factors in index admissions and PPRs. In a LR model, a positive coefficient indicates there is a positive relationship between the given variable (factor) and the occurrence of a PPR, while a negative coefficient indicates a negative relationship between the explanatory and dependent variable. The odds ratio (OR) measures the effect of a unit increase of a predictor, to the risk of having a PPR. The Wald z-statistic was used to test the statistical significance of coefficients in the model.

We are also particularly interested in whether the cost-related/cost-measure risk factors can improve the predictive power of a model that only includes patient-level, hospital-level and clinical risk factors. To do this, we first built a baseline multiple LR model that excludes all cost-related/cost-measure risk factors. Then we add these additional variables to the baseline model (which we will refer to as 'test model' in the following). The diagnostic tests, including ROC curve (AUC), McFadden R-square, and HL test were then used to compare the performance of the different models.

Regarding the baseline model, all patient-level, hospital-level and clinical variables were used in the preliminary model, including age, gender, hospital type, payment model between hospital and fund, the index of hospital contract status with the fund, LOS, Charlson index,

principal diagnosis, secondary diagnosis, principal procedure, MBS category, MBS group, ICU day, index of ICU, minutes of operating in theatre, DRG group 1 (DRG diagnosis and partition group), and DRG group 2 (level of consumption resources).

Starting with this full model, using stepwise backward regression one factor was removed each time and AIC statistic (Akaike Information Criterion) was used to find the model that best fits the data. This backward variable selection was stopped when the model with the smallest AIC was found, and this model was chosen as the baseline. The baseline model includes the following risk factors, 1) patient's age, 2) Charlson index, 3) principal diagnosis, 4) procedure group, 5) MBS group, 6) LOS (length of stay), 7) minutes of operating in theatre, 8) the level of the resources consumption for the treatment, 9) hospital contract status 10) ICU check.

On the basis of equation (4), the baseline model is presented as below:

$$logit\big(P(Y=1|x_1,\dots,x_k)\big) \tag{15}$$
$$= \beta_0 + \beta_1\, age + \beta_2 CI + \beta_3 LOS + \beta_2 MOT + \sum_{i=1\ or\ 0} \beta_3\, D_i\, LRC$$
$$+ \sum_{i=1\ or\ 0} \beta_4\, D_i HCS + \sum_{i=1\ or\ 0} \beta_5\, D_i ICU + \sum_{i=1}^{45} D_i \gamma_i PDG$$
$$+ \sum_{i=1}^{50} D_i \delta_i PROC + \sum_{i=1}^{20} D_i \varepsilon_i MBS$$

Hereby, $Y = 1$ indicates the incidence of a PPR within 30 days after the index admission, and $D_i$ indicates the associated variable is a dummy, and belongs to category $i$ ; $D_i = 1$ indicates incidence of the associated variable, while $D_i = 0$ indicates no incidence.

$\boldsymbol{\beta, \gamma, \delta\ \varepsilon}$ are coefficients for the associated with risk factors in the LR model.

*CI*: Charlson index,

*MOT:* minutes on operating in theatre,

*LRC:* the level of the resources consumption for the treatment,

*HCS:* hospital contract status,

*ICU:* ICU check,

*PDG:* principal diagnosis group,

*PROC:* procedure group,

*MBS:* MBS group.

The final model was then developed as a multiple LR model, including all variables in the baseline model and the significant cost-related/cost-measure variables. Thus, the final model takes the following form:

$$
\begin{aligned}
logit\big(P(Y=1|x_1,\dots,x_k)\big) \qquad (16)\\
= \beta_0 + \beta_1\, age + \beta_2 CI + \beta_3 LOS + \beta_2 MOT\\
+ \sum\nolimits_{i=1\ or\ 0} \beta_3\, D_i\, LRC + \sum\nolimits_{i=1\ or\ 0} \beta_4\, D_i HCS\\
+ \sum\nolimits_{i=1\ or\ 0} \beta_5\, D_i ICU + \sum_{i=1}^{45} D_i \gamma_i PDG + \sum_{i=1}^{50} D_i \delta_i PROC\\
+ \sum_{i=1}^{20} D_i \varepsilon_i MBS + \sum_{i=1}^{k} \theta_i CBV_i
\end{aligned}
$$

, where $Y = 1$ indicates incidence of a PPR in 30 days after the index admission, and

$D_i$ indicates the associated variable is a dummy, and belong to category $i$ ; $D_i = 1$ indicates incidence of the associated variable, while $D_i = 0$ indicates no incidence.

$\boldsymbol{\beta, \gamma, \delta\ \varepsilon}$ are coefficients for the associated with risk factors in the LR model.

*CI*: Charlson index,

*MOT:* minutes of operating in theatre,

*LRC:* the level of the resources consumption for the treatment,

*HCS:* hospital contract status,

*ICU:* ICU check,

*PDG:* principal diagnosis group,

*PROC:* procedure group,

*MBS:* MBS group.

*CBV:* up to *k* cost-based variables including cost-related variables originally from the data set and constructed cost-measure variables.

### 5.2.2 Univariate regression results

This section reports the results for the conducted univariate LR in the order of patient-level variables, hospital-level variables, clinical variables and cost-related/cost-measure variables.

For each variable, the odds ratio (OR), 95% Confidence Interval for the odds ratio (95% CI) and the p-value of the Wald z-statistic (p-value) are presented.

**Patient-level variables**

The univariate logistic regression revealed that the odds of having PPRs were 1.23 times higher for females than for males (OR= 1.23, 95% CI [1.06, 1.43], p-value = 0.0054).

Although a patient's age did significantly affect the odds of having a PPR within 30 days after initial discharge (OR= 1.02, 95% CI [1.02, 1.03], p-value = 0.000), the effect was not very high, i.e. the odds of having a PPR increased approximately 2% per one year increase of age.

**Hospital-level variables**

Regarding the hospital-level factors, the hospital types (public, private or same-day hospital) were not found to have a significant relationship with PPRs (p-values <0.05).

However, the payment models agreed between the fund and hospitals were found to be significantly associated with PPRs. Compared to payment model A, payment model B, payment model C and payment model D were found to significantly increase the risk of a PPR. The OR, 95% CI and p-value for the different payment models are presented in Table 11 (note that a more detailed description of the actual payment models cannot be provided here due to a confidentiality agreement with the health fund.)

**Table 12 Results of univariate LR for payment models**

| **Payment Model** | OR | 95% CI for OR | P-value |
|---|---|---|---|
| **Base level** Payment model A | | | |
| Payment model B | 4.23 | [1.73 , 10.34] | 0.0016 |
| Payment model C | 3.56 | [2.27 , 5.58] | 0.0000 |
| Payment model D | 3.97 | [2.55 , 6.18] | 0.0000 |

**Clinical variables**

We also find that patients that had a longer LOS (OR = 1.03, 95% CI [1.02, 1.05], p-value = 0.0000), a higher Charlson index (OR = 1.19, 95% CI [1.15, 1.24], p-value = 0.000), or/and a shorter time for operating in the theatre (OR = 1.00, 95% CI [1.00, 1.00], p-value = 0.0032) had higher odds of having a PPR.

There were also some disease characteristics in the index admissions that were found associated with higher risks of having a PPR. Results for these variables are summarised in Tables 13-17.

*Principal diagnosis*

The univariate LR revealed that the odds of having PPRs for patients with principal diagnosis of *essential (primary) hypertension* was one third of odds of having PPRs for the patients with principal diagnosis of *rheumatic mitral valve diseases* (OR= 0.31, 95% CI [0.01, 0.59], p-value = 0.0054).

Additionally, the odds of having PPRs for patients with principal diagnosis of 1) *cardiomyopathy,* 2) *atrioventricular and left bundle-branch block,* 3) *conduction disorders,* 4) *paroxysmal tachycardia,* 5) *atrial fibrillation and flutter,* 6) *cardiac arrhythmias,* 7) *complications and ill-defined descriptions of heart disease,* 8) *phlebitis and thrombophlebitis of other sites,* 9) *varicose veins of lower extremities,* 10) *haemorrhoids,* 11) *oesophageal varices,* 12) *gastric varices, 13) disorders of veins,* 14) *hypotension,* and 15) *other circulatory diseases* were much smaller than for patients with principal diagnosis of *rheumatic mitral valve diseases* (OR <0.2, P-values < 0.05).

**Table 13 Results of univariate LR for principal diagnosis**

| **Principal diagnosis group** | OR | 95% CI for OR | p-value |
|---|---|---|---|
| **Base level** Rheumatic mitral valve diseases | | | |
| Essential (primary) hypertension | 0.31 | [0.01 , 0.59] | 0.0167 |
| Cardiomyopathy | 0.08 | [0.02 , 0.42] | 0.0029 |
| Atrioventricular and left bundle-branch block | 0.13 | [0.03 , 0.62] | 0.0112 |
| Other conduction disorders | 0.08 | [0.01 , 0.83] | 0.0347 |
| Paroxysmal tachycardia | 0.13 | [0.03 , 0.57] | 0.0065 |
| Atrial fibrillation and flutter | 0.20 | [0.05 , 0.79] | 0.0209 |
| Other cardiac arrhythmias | 0.05 | [0.01 , 0.31] | 0.0011 |
| Complications and ill-defined descriptions of heart disease | 0.05 | [0.01 , 0.59] | 0.0167 |
| Phlebitis and thrombophlebitis of other | 0.18 | [0.03 , 0.94] | 0.0425 |

| | | | |
|---|---|---|---|
| sites | | | |
| Varicose veins of lower extremities | 0.04 | [0.01 , 0.17] | 0.0000 |
| Haemorrhoids | 0.02 | [0.01 , 0.1] | 0.0000 |
| Oesophageal varices | 0.15 | [0.03 , 0.88] | 0.0357 |
| Gastric varices | 0.06 | [0.01 , 0.64] | 0.0202 |
| Other disorders of veins | 0.09 | [0.01 , 0.97] | 0.0475 |
| Hypotension | 0.11 | [0.02 , 0.65] | 0.0150 |
| Others | 0.17 | [0.04 , 0.75] | 0.0198 |

*Secondary diagnosis*

According to the results of conducted univariate LRs, patients with secondary diagnosis of 1) *spastic hemiplegia*, 2) *unstable angina,* 3) *other pulmonary heart diseases, atherosclerosis,* 4) *acute kidney failure,* and Chronic conditions such as 5) *hypertension,* 6) *ischaemic heart disease,* 7) *coronary artery disease,* 8) *chronic heart failure,* 9) *chronic congestive heart disease/failure* were over 5 times more likely to have a PPR (OR >5, p-values > 0.05), in comparison to patients with secondary diagnosis of Rheumatic mitral valve diseases.

**Table 14 Results of univariate LR for secondary diagnosis**

| **Secondary diagnosis group** | OR | 95% CI for OR | P-value |
|---|---|---|---|
| **Base level** Rheumatic mitral valve diseases | | | |
| Spastic hemiplegia | 10.00 | [2.03 , 49.32] | 0.0047 |
| Unstable angina | 5.40 | [1.0 , 24.29] | 0.0280 |
| Other pulmonary heart diseases | 6.00 | [1.00 , 35.92] | 0.0497 |
| Atherosclerosis | 6.46 | [1.40 , 29.93] | 0.0170 |
| Acute kidney failure | 10.29 | [1.87 , 56.74] | 0.0075 |
| Chronic conditions, such as hypertension, ischaemic heart disease, coronary artery disease, chronic heart failure, chronic congestive heart disease/failure | 5.18 | [1.09 , 24.57] | 0.0385 |

*Principal MBS Category*

The univariate LR revealed that the odds of having PPRs for patients that belonged to MBS Category 3 – *Therapeutic Procedures* (OR =0.55, 95% CI [0.42, 0.71], p-value= 0.0000) were about half of those for patients belonging to MBS category 1 – *Professional attendances*.

*Principal MBS group*

The univariate LR revealed that there were no specific principal MBS groups that significantly increase/decrease the likelihood of a patient being readmitted to hospital for preventable conditions (P-values < 0.05).

*Procedures*

The odds of having PPRs for patients that had procedures of 1) *fibreoptic colonoscopy to caecum with polypectomy,* 2) *rubber band ligation of haemorrhoids,* 3) *laser haemorrhoidectomy,* and 4) *Interruption of sapheno-femoral junction varicose veins* were slightly lower than those for patients that had a procedure of *testing of other cardiac pacemaker* in the index admission (OR <0.2, p-values < 0.05); while the odds of having PPRs for patients that had procedures of *Digital subtraction angiography of head and neck* were eight times greater than those for patients that had procedure of testing of other cardiac pacemaker (OR >8, p-values < 0.05).

**Table 15 Results of univariate LR for procedures**

| **procedure group** | OR | 95% CI for OR | P-value |
|---|---|---|---|
| **Base level** Testing of other cardiac pacemaker | | | |
| Fibreoptic colonoscopy to caecum, with polypectomy | 0.15 | [0.02 , 0.93] | 0.0417 |
| Rubber band ligation of haemorrhoids | 0.13 | [0.03 , 0.64] | 0.0117 |
| Laser haemorrhoidectomy | 0.15 | [0.03 , 0.79] | 0.0251 |
| Interruption of sapheno-femoral junction varicose veins | 0.10 | [0.02 , 0.42] | 0.0017 |
| Digital subtraction angiography of head and neck | 8.73 | [2.29 , 33.3] | 0.0015 |

*DRG*

The odds of PPRs for patients that were assigned with DRG of *Nervous system medical partitions* (OR = 2.94, 95% CI [1.26, 6.87], p-value = 0.0128) were almost 3 times larger than those for patients of *Transplant surgical partitions*; while patients of *Digestive system medical* (OR = 0.06, 95% CI [0.02, 0.21], p-value = 0.0000) or other partitions (OR = 0.09, 95% CI [0.03, 0.26], p-value = 0.0000) had a lower chance of having a PPR compared to the patients of *Transplant surgical partitions.*

In addition, the odds of having PPRs for patients who had the third highest resource consumption in the initial treatment (OR = 0.66, 95% CI [0.49, 0.88], p-value = 0.0043) showed approximately half less than that for the patients who consumed the highest resources.

*Cost-related variables*

Regarding the cost-related variables, it was found that the univariate LR models showed that ICU charge, principal MBS schedule fee, hospital fees/benefits, medical fees/benefit, and total schedule fees for index admission did not have any significant relationships with the risk of having a PPR (p-values > 0.05).

However, hospital out-of-pockets, medical out-of-pockets, and prosthesis fees/benefits were found significantly related with the risk of having PPRs.

The OR, 95% CI, and p-values for cost-related variables are reported in Table 16.

**Table 16 Results of univariate LR for cost-related variables**

| **Variables** | **OR** | **95% CI for OR** | **P-value** |
|---|---|---|---|
| ICD charge | 0.98 | [0.91 , 1.05] | 0.5450 |
| Principal MBS schedule fee | 1.00 | [0.94 , 1.07] | 0.9370 |
| Hospital fees | 1.01 | [0.94 , 1.08] | 0.7820 |
| Hospital benefits | 1.01 | [0.95 , 1.08] | 0.6890 |
| Medical fees | 0.96 | [0.89 , 1.03] | 0.2710 |
| Medical benefits | 0.95 | [0.88 , 1.02] | 0.1570 |
| Total schedule fees | 0.98 | [0.91 , 1.05] | 0.4930 |
| Hospital Out-of-Pockets | 0.79 | [0.67 , 0.92] | 0.0026 |
| Medical Out-of-Pockets | 0.89 | [0.81 , 0.97] | 0.0111 |

| | | | |
|---|---|---|---|
| Prosthesis Fees | 0.87 | [0.79 , 0.96] | 0.0056 |
| Prosthesis Benefits | 0.87 | [0.80 , 0.96] | 0.0057 |

The results are based on using the z-score of the cost-related/measure variables.

As Table 16 shows, values for the variables hospital out-of-pockets (OR = 0.70, 95% CI [0.67, 0.92], p-value = 0.0026), medical out-of-pockets (OR = 0.89, 95% CI [0.81, 0.97], p-value = 0.0111), prosthesis fees (OR = 0.87, 95% CI [0.79, 0.96], p-value = 0.0056) and prosthesis benefits (OR = 0.87, 95% CI [0.80, 0.96], p-value = 0.0058) indicate a significant negative relationships with the risk of having a PPR.

*Cost-measure variables*

As Table 17 shows, the cost-measure variable 1.1 (OR = 0.77, 95% CI [0.70, 0.84, p-value = 0.0000) showed a significant negative relationship with the risk of PPRs, while cost-measure variable 3 (OR = 1.33, 95% CI [1.15, 1.55, p-value = 0.0002) showed a significant positive relationship with the risk of having a PPR. The other cost-measure variables were found insignificant in the univariate LR model.

**Table 17 Results of univariate LR for cost-measure variables**

| **Variables** | **OR** | **95% CI for OR** | **P-value** |
|---|---|---|---|
| Cost-measure variable 1.1 | 0.77 | [0.70 , 0.84] | 0.0000 |
| Cost-measure variable 1.2 | 0.89 | [0.76 , 1.03] | 0.1070 |
| Cost-measure variable 2 | 0.98 | [0.85 , 1.13] | 0.7540 |
| Cost-measure variable 3 | 1.33 | [1.15 , 1.55] | 0.0002 |
| Cost-measure variable 4 | 1.09 | [0.94 , 1.26] | 0.2460 |
| Cost-measure variable 5.1 | 0.97 | [0.90 , 1.05] | 0.4950 |
| Cost-measure variable 5.2 | 0.93 | [0.80 , 1.07] | 0.3040 |

### 5.2.3 Results for multiple logistic regression

**Preliminary model**

The preliminary model was developed by multivariate logistic regression including all patient-level and clinical variables. The AIC backward variable selection was applied to find the baseline model (see Section 5.2.1), and the value of the AIC for the preliminary model

was 5250.22 and deceased to 5202.13 for the baseline model. The baseline model included these risk factors,

1) patient's age,
2) Charlson index,
3) principal diagnosis,
4) procedure group,
5) MBS group,
6) LOS (length of stay),
7) minutes of operating in theatre,
8) the level of the resources consumption for the treatment,
9) hospital contract status,
10) ICU check.

Then cost-related variables that were either (i) directly available from the data provided by the health fund, or (ii) constructed as outlined in Section 4.4 were added to the baseline model to examine whether the variable would improve the predictive power of the model.

Unfortunately, the cost-related variables originally from the data set such as hospital fees/benefits, medical fees, hospital out-of-pockets, total schedule fees, ICU charge, prosthesis fees/benefits, and the newly constructed cost-measure variable 1.2, cost-measure variable 2, cost-measure variable 3, cost-measure variable 5.1, and cost-measure variable 5.2 were found to be insignificant whenbeing added to the baseline model. So for these variables, we reject the hypotheis of providing additional explanatory or predictive power for distinguishing between PPRs and non-PPRs/no readmissions.

In the other word, the hospital fees (Hypotheses 2) and/or the principal MBS schedule fee (Hypotheses 3) charged in the higher half among the index admissions with the same diagnosis group (DRG group 1) ) were found not to be associated with the risk for a patient to encounter a PPR.

Additionally, it was also found that holding a high level of private health insurance policy (Hypotheses 5) could not help to reduce the risk of having a PPR for the patients in our sample.

The cost-measure variable 1.1 (OR =0.88, 95%CI [0.79, 0.98], p-value = 0.0186) suggests that the more the medical fees exceeds the total schedule fees, the less likely it is for the given patient to encounter a PPR. However, the cost-measure 1.2 failed to verify Hypotheses 1 in multivariate LR model (P-value >0.05).

The cost-measure variable 3 was found significant in the univariate LR (OR =1.33, 95%, CI [1.15, 1.55], p-value = 0.0002) but insignificant in the testing model (P-value = 0.7832).

The cost-measure variable 4 was found insignificant in the univariate LR (OR =1.09, 95%CI [0.94, 1.26], p-value = 0.2460) but significant in testing model (OR =0.67, 95%CI [0.54, 0.83], p-value = 0.0002). So the effect of principal MBS schedule fees on the risk of having PPRs is significantly different for different values of LOS (p-value < 0.05). However, the OR for the cost-measure 4 was less than 1, which presented a negative relationship with the risk of having PPRs. So the results suggest that if the interaction of (principal MBS schedule fee * LOS) is above the median for a specific index admissions (cost-measure variable 4 = "1" is significant), it is more likely for the patient to encounter a PPR.

Medical fees (OR = 0.83, 95%CI [0.71, 0.83], p-value = 0.0232), medical out-of-pockets (OR =0.85, 95%CI [0.75, 0.95], p-value = 0.0034) and principal MBS schedule fee (OR =0.87, 95%CI [0.76, 1.00], p-value = 0.0473) were also found significant in the relating testing models, and presented negative relationships with the risk of having PRPs.

In addition, the diagnostic tests (AUC, McFadden R-square and HL Test) indicated that the testing models out-performed the baseline model. Regarding the significant cost-related/cost-measure variables, the coefficients and p-values for variables included in the corresponding testing models are reported in Table 18.

**Table 18 Results for testing models including coefficients and P-values for each variable.**

| | Baseline model | | Testing Model 1 | | Testing Model 2 | | Testing Model 3 | | Testing Model 4 | | Testing Model 5 | | Testing Model 6 | | Testing Model 7 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coef. | P-value | Coef. | P-value | Coef. | P-value | Coef. | P-value | Coef. | P-value | Coef. | P-value | Coef. | P-value | Coef. | P-value |
| Medical Fees | | | -0.18 | 0.0232 | | | | | | | | | | | | |
| Medical Out-of-Pocket | | | | | -0.17 | 0.0034 | | | | | | | | | | |
| Principal MBS Schedule Fee | | | | | | | -0.14 | 0.0473 | | | | | | | | |
| Cost-measure variable 1.1 | | | | | | | | | -0.1273 | 0.0186 | | | | | -0.14 | 0.0148 |
| Cost-measure variable 3 (dummy = 1) | | | | | | | | | | | -0.0318 | 0.7832 | | | | |
| Cost-measure variable 4 (dummy = 1) | | | | | | | | | | | | | -0.39 | 0.0002 | -0.39 | 0.0006 |
| **Variables** | | | | | | | | | | | | | | | | |
| Age | -0.02 | 0.0008 | -0.02 | 0.0005 | -0.02 | 0.0005 | -0.02 | 0.0015 | -0.02 | 0.0002 | -0.02 | 0.0007 | -0.02 | 0.0009 | -0.02 | 0.0009 |
| Principal Diagnosis | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| Principal MBS group | No | | No | | No | | Yes | | Yes | | No | | No | | Yes | |
| Length of Stay | 0.02 | 0.0644 | 0.04 | 0.0034 | 0.02 | 0.0142 | 0.02 | 0.0647 | 0.03 | 0.0075 | 0.02 | 0.0545 | 0.04 | 0.0008 | 0.04 | 0.0008 |
| Patient received ICU | -0.34 | 0.0745 | -0.18 | 0.3673 | -0.28 | 0.1436 | -0.29 | 0.1291 | -0.34 | 0.0958 | -0.30 | 0.1251 | -0.32 | 0.1303 | -0.33 | 0.1251 |
| Minutes for operating in theatre | 0.00 | 0.1764 | 0.00 | 0.3584 | 0.00 | 0.2049 | 0.00 | 0.2898 | 0.00 | 0.0497 | 0.00 | 0.2037 | 0.00 | 0.0738 | 0.00 | 0.0746 |
| Charlson index | 0.17 | 0.0000 | 0.17 | 0.0000 | 0.17 | 0.0000 | 0.17 | 0.0000 | 0.18 | 0.0000 | 0.17 | 0.0000 | 0.17 | 0.0000 | 0.17 | 0.0000 |
| Principal Procedure | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| level of resource consumption | No | | No | | No | | No | | Yes | | Yes | | No | | No | |
| Hospital is contracted with fund | 0.70 | 0.0061 | 0.17 | 0.0000 | 0.70 | 0.0068 | 0.76 | 0.0037 | 0.94 | 0.0037 | 0.75 | 0.0056 | 0.93 | 0.0058 | 0.95 | 0.0051 |
| **Diagnosis for models** | | | | | | | | | | | | | | | | |
| AUC | 0.7581 | | 0.7593 | | 0.7604 | | 0.7593 | | 0.7610 | | 0.7568 | | 0.7629 | | 0.7647 | |
| McFadden R-square | 0.1320 | | 0.1366 | | 0.1376 | | 0.1566 | | 0.1384 | | 0.1321 | | 0.1342 | | 0.1415 | |
| Hosmer-Lemeshow | 0.3746 | | 0.3308 | | 0.4870 | | 0.7943 | | 0.7918 | | 0.4485 | | 0.6921 | | 0.6712 | |

The results for medical fees, medical out-of-pockets and Principal MBS schedule fee are based on using the z-score of them in the multivariate LR model.

"Yes" indicates the variables is a dummy and significant (P-value <0.05) in the model.

"No" indicates the variables is a dummy and insignificant in the model (P-value >0.05).

As Table 18 shows, the effects of some clinical-level variables such as principal MBS group, LOS, minutes of operating in theatre and level of resource consumption in index admissions, to the risk of having PPRs were found various in different testing models. It was also found that the variable principal MBS group was only statistically significant (P-values <0.5) in Model 3 and Model 4. We observe that except for the baseline model and Model 3, LOS was statistically significant in all other models. Minutes of operating in theatre was only statistically significant in Model 4; and the level of resource consumption was only statistically significant in Model 4 and 5.

The AUC and McFadden R-square both indicated that the extended models showed slightly better predictive power compared to the baseline model; while conducted HL tests suggested that Model 3, 4, 6 and 7 possibly fitted the data better than the baseline model and the other models (based on results for the observed p-values).

**Final benchmark**

The final benchmark model was determined as the multiple LR model including the variables from the baseline model, as well as statistically significant cost-based. Starting from Model 6 (the model with the most significant additional cost-based variable), we included additional cost-based variables that were statically significant, and then test whether inclusion of these additional variables would help to improve the model. As Table 18 shows, the significant cost-based variables in the tested models are medical fees, medical out-of-pockets, principal MBS schedule fee, cost-measure 1.1 and cost-measure 4.

When adding cost-measure 1.1 to Model 6, cost-measure 1.1 and cost-measure 4 were both significant in the model (see Table 18-Testing model 7). However, adding any of the other cost-based variables did not result in significant coefficients for all cost-based variables in the model.

Finally, based on these results, Model 7 was selected as the final model, including the following explanatory variables: 1) patient's age, 2) Charlson index, 3) principal diagnosis, 4) procedure group, 5) MBS group, 6) LOS , 7) minutes of operating in theatre, 8) the level of the resources consumption for the treatment, 9) hospital contract status ,10) ICU check, 11) cost-measure variable 1.1, and 12) cost-measure variable 4.

All the significant variables in the final model were summarised in table below.

**Table 19 Significant risk factors in the final model and their OR, 95% CI and P-values**

| Significant variables | OR | 95% CI | P-value |
|---|---|---|---|
| Age | 0.98 | [0.97 , 0.99] | 0.0010 |
| Principal Diagnosis of Intracranial haemorrhage (nontraumatic) | 10.78 | [1.77 , 65.57] | 0.0098 |
| Principal Diagnosis of Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction | 5.82 | [1.02 , 33.27] | 0.0478 |
| Principal Diagnosis of Varicose veins of lower extremities | 0.09 | [0.01 , 0.97] | 0.0476 |
| Principal Diagnosis internal/external haemorrhoids | 0.03 | [0.00, 0.28] | 0.0020 |
| Minutes of operating in theatre | 1.00 | [1.00 , 1.00] | 0.0379 |
| Length of stay | 1.04 | [1.02 , 1.07] | 0.0012 |
| Charlson index | 1.18 | [1.09 , 1.28] | 0.0000 |
| Principal procedure of Initial psychiatric interview, of person other than patient, > 45 minutes duration | 0.08 | [0.01 , 0.56] | 0.0105 |
| Principal procedure of Digital subtraction angiography of aorta and lower limb, >= 10 data acquisition runs, unilateral | 4.71 | [1 , 22.08] | 0.0496 |
| Hospital contract status - contracted with the health fund | 2.57 | [1.33 , 4.99] | 0.0052 |
| Cost-measure 1.1 | 0.87 | [0.78 , 0.97] | 0.0148 |
| Cost-measure 4 | 0.67 | [0.54 , 0.84] | 0.0005 |

The results of the final model revealed that the odds of having PPRs within 30 days after the initial discharge for a patient was slightly influenced by the patient's age and LOS. It decreased by 2% as the patient's age increased by 1 year, increased by 4% as the patient had one additional day in hospital, and increased 18% as the patient's Charlson index increased by a value of one.

The odds of having PPRs within 30 days after the initial discharge for patients with principal diagnosis of *Intracranial haemorrhage (nontraumatic)*, and *Occlusion and stenosis of precerebral arteries (not resulting in cerebral infarction)* were about 11 times and 6 times higher than the patients with a principal diagnosis of *multiple valve diseases* (the base level for principal diagnosis group in the final model) respectively.

We further found that the odds of having PPRs for patients with a principal procedure of *digital subtraction angiography of aorta and lower limb* were almost 5 times higher than the patents had principal procedure of *testing of other cardiac pacemaker* (the base level for principal procedure group in the final model); while the patients with principal procedure of *initial psychiatric interview* were slightly less likely to have PPRs.

However, the patients that had a principal diagnosis of *varicose veins of lower extremities* and *internal/external haemorrhoids* were much less likely to have PPRs compared to patients with a principal diagnosis of *multiple valve diseases* (OR <= 0.10).

Regarding the developed cost-measure variables, the odds of having PPRs within 30 days was decreased by 32% per 1% increase of the cost-measure variable 1.1 (z-score for ratio of medical fees/total schedule fees). The odds of having PPRs for the patients who had interaction of LOS and principal MBS schedule fee that was higher than median among all index admissions (the cost-measure variable 4 = "1"), was about 0.67 lower than that for the patients who did not.

## 5.3 Model evaluation results

**Tested models**

As Table 18 shows, Model 6 had the best classifier performance with the highest AUC (0.7629) compared to the testing models only including one cost-based variable. However, all tested models had small values (close to 0.10) for McFadden R-square, which indicated the predictive power of them was relatively low. Goodness-of-fit was measured by conducted HL tests. It was interesting to see that Model 3 (HL p-value = 0.7943) and 4 (HL p-value = 0.7918) increased the HL p-value more than two times in comparison to the baseline model (HL p-value = 0.3746), a possible indication for a better fit of these two models in comparison to the other models.

**Final model**

The final model (Model 7) was tested by ROC (AUC = 0.7647), McFadden R-square (0.1415) and HL test (0.6721). The AUC larger than 0.70 and smaller than 0.08 suggested that the final model had moderate predictive accuracy. The McFadden R-square statistic and AUC did not indicate that final model presented a substantial improvement in goodness-of-fit compared to the baseline model. However, p-value of HL test was a little bit higher than that in the baseline model (HL= 0.3746). Figure 6 provides a plot of the ROC curve for Model 7.

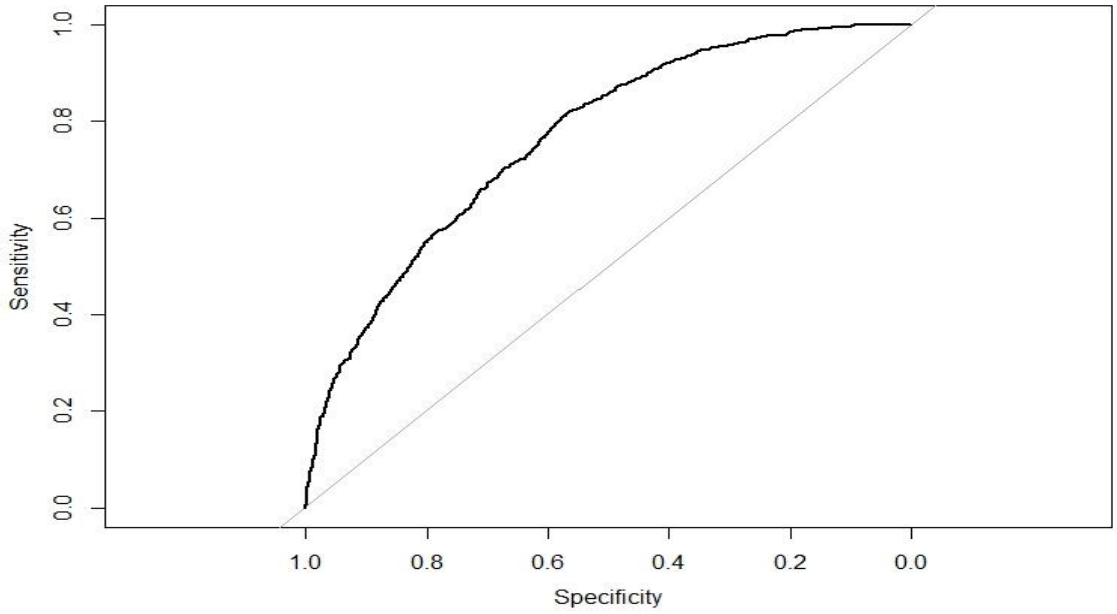


**Figure 6 ROC curve for the final model.**

In order to measure how well the final model can predict the PPS, the confusion matrix with the cut-off rate 0.50 for the final model is presented in Table 20.

**Table 20 Confusion matrix for final model at cut-off rate 0.50**

| | | **Model prediction** | |
|---|---|---|---|
| | | PPRs | Non-PPRs or no readmissions |
| **Observed** | PPRs | 29 | 704 |
| | Non-PPRs or no readmissions | 21 | 7445 |

The TPR was 0.04, FPR was 0.003 and the accuracy rate was 0.91. Obviously, a cut-off rate 0.50 was not necessarily the optimal threshold, since it only yielded a relatively low TPR. Considering the benefits for identifying PPRs were more important than the cost of misclassification, the cut-off was varied in such a way that is provided a higher TPR (see Table 21). In order to get the highest TPR, the cut-off rate was chosen at 0.100; while the cut-off rate 0.125 was also acceptable because the model still can classify more than half of the PPRs with a relative low FPR and high accuracy.

**Table 21 Cut-off rates vs. TPR, FPR and Accuracy for LR final model**

| Cut-off rate | TPR | FPR | Accuracy |
|---|---|---|---|
| 0.100 | 0.70 | 0.33 | 0.67 |
| 0.125 | 0.57 | 0.23 | 0.75 |
| 0.150 | 0.43 | 0.13 | 0.83 |
| 0.175 | 0.31 | 0.07 | 0.88 |
| 0.200 | 0.28 | 0.05 | 0.89 |
| 0.225 | 0.25 | 0.04 | 0.89 |
| 0.250 | 0.22 | 0.04 | 0.90 |
| 0.275 | 0.20 | 0.03 | 0.90 |
| 0.300 | 0.18 | 0.02 | 0.90 |
| 0.325 | 0.16 | 0.02 | 0.91 |
| 0.350 | 0.10 | 0.01 | 0.91 |
| 0.375 | 0.08 | 0.01 | 0.91 |
| 0.400 | 0.07 | 0.01 | 0.91 |
| 0.425 | 0.05 | 0.01 | 0.91 |
| 0.450 | 0.05 | 0.00 | 0.91 |
| 0.475 | 0.04 | 0.00 | 0.91 |
| 0.500 | 0.04 | 0.00 | 0.91 |

As Figure 7 shows, as expected the TPR rates decreases when cut-off rates are increased. When the cut-off rate was 0.75, the final model could not identify PPRs successfully (TPR=0.00). However, the accuracy was still high as 0.91, and it might be caused by the final model classified most of non-PPRs or no readmissions correctly.

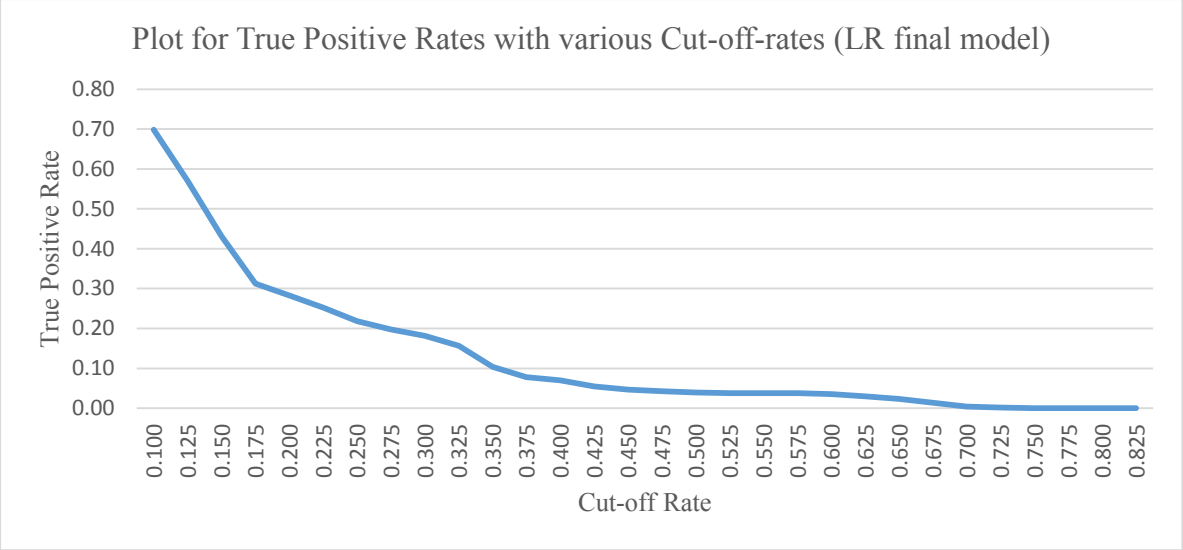


**Figure 7 Cut-off rates vs. TPR rates for LR final model.**

## 5.4 Discussion of results

**The rates of PPRs**

This study estimated that the proportion of PPRs amongst readmissions was 42.1%, and 8.8% among index admissions; these numbers were close to those rates reported in the study conducted in the US by Donze et al. (2013). They reported that PPRs were 36.5% amongst readmissions and 8.5% amongst index admissions.

**Patient-level factors**

In this study, it was found that patients' age did not strongly influence the risk of having PPRs. This finding was supported by other two studies conducted by Donze et al. in (2013) and (2014). In contrast, Lichtman et al. (2013) found that patients with older age had a higher likelihood of being readmitted to hospital due to preventable reasons. Nevertheless, their study only considered patients 65 years old and older, whereas the age range in our study was much broader (15 to 104 years old).

In our study, male patients are more likely to be readmitted to hospital with PPRs than females; this was consistent with the findings of a study conducted in patients with heart failure, acute myocardial infarction, pneumonia, or COPD by Shams, Ajorlou and Yang in 2015.

**Hospital-level factors**

Regarding the hospital-level factors, the hospital type, payment model between hospitals and the fund, and hospital contract status were examined. It was found that hospital type was not a significant factor in either univariate or multivariate models. With respect to payment models, the univariate LR analysis indicated that some payment models had significant positive relationships with the risks of PPRs. However, payment model was not selected for the final model by AIC backward selection.

**Clinical-level factors**

In this study, the five principal diagnoses most commonly seen among patients with PPRs were:

*Chronic ischaemic heart disease* (22%), *angina pectoris or unstable angina* (12%), *arial fibrillation and flutter* (10%), *atherosclerosis* (9%), and *no rheumatic aortic valve disorders* (7%). However, none of them were found significantly associated with the risk of having PPRs in the final model.

It was found the patients that were diagnosed with stroke conditions, such as *intracranial haemorrhage (non-traumatic)*, and/or *occlusion and stenosis of precerebral arteries* in the index admission, were much more likely to have a PPR within 30 days after the initial discharge.

Evidence showed that stroke was a main cause of readmissions, and patients with stroke usually returned to hospital because of infection, recurrent stroke and cardiovascular disease (Bjerkreim et al. 2016). Some of the causes can be avoided by coordinated timing of elective procedures and earlier outpatient follow-up (Nahab et al. 2012).

The univariate analysis indicated that some secondary diagnosis appeared to be associated with the risk of having PPRs. Some of these diagnose were, for example, *spastic hemiplegia*, *unstable angina*, *other pulmonary heart diseases*, *atherosclerosis*, *acute kidney failure* and other chronic conditions. However, the variable selection algorithm did not include the secondary diagnosis as an important risk factor in the final model. It might be because the principal diagnosis or other clinical-related variables had much stronger predictive power than the secondary diagnosis did.

Although a lot of studies found that LOS for the index admission was a significant risk factor (Halfon et al. 2002; Yam et al. 2010; Donze et al. 2013; Lichtman et al 2013), this study

found that LOS just slightly increased the risk of having PPRs. One study that supports this finding was the study from Shams, Ajorlou and Yang 2015 that revealed that LOS was insignificantly associated with the risk of readmissions due to acute myocardial infarction and COPD.

In our study, comorbidities were measured by Charlson index. The higher the score, the greater risk of one-year mortality. We found the Charlson index was strongly associated with the risk of having PPRs, which was consistent with findings from other studies.

**Hypotheses and Cost-measure factors**

As far as we are concerned, we could not find any articles in which the relationships between cost-based factors and the risk of PPRs were analysed. Based on the results of the univariate models, the original cost-related variables from the data set did not have any significant relationships with the risk of PPRs.

That is why we constructed seven cost-measure variables along with five relative hypotheses that were tested in the study. A few cost-measure variables were found significantly associated with the risk of having PPRs for patients in the univariate LR model. These were *cost-measure variable 1.1* (z-score of the ratio of (medical fee / total schedule fee)) and *cost-measure variable 3* (a dummy variable that indicates whether the principal MBS schedule fee of a given index admission is higher than median amongst all index admissions or not).

It is interesting to see that *cost-measure variable 4* (a dummy variable that 'indicates the interaction of LOS and Principal MBS schedule fee for a given index admission is higher than median among all index admissions) was insignificant in the univariate LR model. Nevertheless, it was found highly significant in the multivariate LR model; whereas the *cost-measure variable 3* behaved in the opposite way.

Finally, *cost-measure 1.1* and *4* were selected in the final model, because any of the other cost-based variables did not result in significant coefficients for all cost-based variables in the model. According to the results of model diagnostic tests (McFadden R-square, HL test, and AUC), they also increased model's predictive power, goodness-of-fit, and overall accuracy compared to the baseline model.

Regarding the Hypothesised relationships between cost-measure variables and the risk of PPRs, the study found that the higher the payment of the medical treatment (Medical fees) above the standard rate (Total schedule fees), the less likely the patient will have a PPR.

In addition, the effect of principal MBS schedule fee (the level of complexity of the treatment) on PPRs was significantly different for different values of LOS. The high value of the interaction of principal MBS schedule fee and LOS may lead to less risk of PPR.

# 6. **Conclusion**

This chapter provides summary of the background and aims of this thesis, a discussion of the strength and limitations, and the conclusions based on the conducted study.

## 6.1 Background and aims

Hospital readmissions due to circulatory system diseases are a leading topic of practice reform and healthcare policy. They are common, costly and in many cases, potentially preventable (Vest et al. 2010). As the Australian population has been aging rapidly, circulatory system diseases place a significant burden of morbidities and mortality on individuals, and impose heavy costs on public and private healthcare system. In addition, it was found that diseases of the circulatory system have the highest number of admissions and lead to significant hospital costs. Partially, these costs are also a result of potentially preventable conditions among adults and children (Jiang, Russo, and Barrett 2006).

PPRs have been explored in a number of studies in the United States. Most of these studies have focused on identifying risk factors associated with PPRs and/or the prediction of PPRs. A variety of risk factors have been found in these studies, such as patient's age, gender, LOS, comorbidities, race, marital status, and socioeconomic status etc. However, these findings were not conclusive. Additionally, only a few studies on PPRs have been conducted in the Australias context so far.

The aims of this study were:

1. to identify patient-level, hospital-level and clinical factors for PPRs such as age, length of stay (LOS), gender or secondary diagnoses.
2. to identify additional cost-related risk factors that help to explain the causes for PPRs

To our knowledge, this is the first study exploring the relationships between cost-based factors and the risk of having PPRs in Australia. To fill this gap in the literature, this study examined cost-based variables based on an insurance claims data set provided by a private health fund. Further, we constructed additional cost measures such as, e.g., ratios between medical fees and schedule fees, in order to test their contribution to explaining PPRs.

## 6.2 Summary of major findings

This study estimated that the proportion of PPRs among readmission in the considered database was 42.1%, while it was 8.8% among index admissions. It found that the most common principal diagnosis among admitted patients with circulatory system diseases who had PPRs within 30 days after discharge were *chronic ischaemic heart disease*, *angina pectoris or unstable angina*, *arial fibrillation and flutter*, *atherosclerosis*, and *no rheumatic aortic valve disorders*. The average age, LOS, and Charlson index of the patients who encountered a PPR were higher than those among the patients who were not readmitted to hospital or did not have a PPR.

It was found that patients with circulatory system diseases were more likely to have a PPR if they were exposed to one or more of the following factors in the index admission: being male, more complications (comorbidities) apart from the main diagnosis, stroke conditions, and having procedures of digital subtraction angiography of aorta and lower limb. Furthermore, the factors that slightly influenced the risk of PPRs were, age, minutes of operating in theatre and LOS.

Next to these standard patient level, clinical and hospital level variables, we also examined the impact of cost-based variables on PPRs. While most of the considered variables did not help to provide additional explanatory power for PPRs, in particular two measures seemed to further improve the applied multiple logistic regression model. We found that the more the doctors charged above the schedule fees for medical services associated with a patient's index admission, the less likely the patient was to have a PPR. Since the gap between medical fees and total schedule fees is also an indication of possible additional treatments for a patient, we interpret these results the following way: additional costs or treatments beyond those incorporated into standard schedule fees may have the potential to reduce the likelihood of PPRs. Another finding was that LOS had significantly different effects on the risk of having PPRs for different principal MBS schedule fees. Our results suggest that a combination of higher principal MBS schedule fees and LOS may lead to less risk for a patient to have a PPR. Since the coefficient for the variable LOS is positive in the estimated model, we interpret these results in the following manner: the estimated negative coefficient of the interaction effect (LOS * Principal MBS Schedule fee) suggests that the higher risk of having a PPR for

patients with longer stays in the hospital is probably most pronounced for those patients with a lower principal MBS schedule fee, i.e. those with typically less complex procedures.

## 6.3 Strengths and limitations of the study

This thesis contributes to the existing knowledge in relation to the risk factors of PPRs among Australian admitted patients with circulatory system diseases. First, this study is the first to empirically investigate the relationships between costs associated with index admissions and PPRs among admitted patients with circulatory system diseases. Second, this study provides valuable information about the factors influencing PPRs due to circulatory system diseases in the Australian context for health researchers, health care providers, and policy makers. Third, this study was conducted using a private health insurance claim data set that covers all states in Australia. Therefore, the findings in this study provide a comprehensive picture on possible causes for PPRs. Fourth, the conducted study developed an algorithm that is useful with regards to identifying PPRs and non-PPRs. Thus, it provides a valuable method for non-experts, such as medical researchers, policy makers, hospital management panels and private insurance health companies etc. to classify PPRs in a fast and straightforward manner. Last but not least, the techniques applied for examining risk factors and testing hypotheses in the study provide a reliable approach for modelling the risk of having PPRs.

Despite the strengths of this study, some limitations should be taken into consideration. The algorithm developed in this study for classifying PPRs may not reflect all potentially preventable clinical conditions related to the index admission. The SQLape Algorithm, which we used as the key reference for developing the algorithm in the study, has not been validated in Australia. Therefore, this leads to a limitation in identifying PPRs among Australian admitted patients. Another limitation is that some factors which may influence the risk of having PPRs, such as race, marital status, income status and blood pressure etc. could not be examined in this study because they were unavailable in the database. In addition, this study used a sample of patients with circulatory system disease, so findings cannot be generalised to patients with other diseases. Further, there was a significant ratio of missing values for the considered cost-based variables such that the sample of patients in this study was significantly reduced in comparison to the entire dataset.

Overall, the empirical results of this study suggest that some of the constructed cost-based measures have the potential to better explain and potentially predict the risk of PPRs. Thus, we recommend to further examine the relationship between PPRs and cost-based variables in future studies.

# Appendix

## Appendix 1: Unplanned readmissions

Unplanned readmissions are the admissions where the principal diagnosis indicates an adverse event, including:

1. Complications following infusion, transfusion and therapeutic injection
2. Complications of procedures, not elsewhere classified
3. Complications of cardiac and vascular prosthetic devices, implants and grafts
4. Complications of genitourinary prosthetic devices, implants and grafts
5. Complications of internal orthopaedic prosthetic devices, implants and grafts
6. Complications of other internal prosthetic devices, implants and grafts
7. Failure and rejection of transplanted organs and tissues
8. Complications peculiar to reattachment and amputation
9. Other complications of surgical and medical care, not elsewhere classified
10. Sequelae of complications of surgical and medical care, not elsewhere classified
11. Postprocedural endocrine and metabolic disorders, not elsewhere classified
12. Postprocedural disorders of nervous system, not elsewhere classified
13. Postprocedural disorders of eye and adnexa, not elsewhere classified
14. Postprocedural disorders of ear and mastoid process, not elsewhere classified
15. Postprocedural disorders of circulatory system, not elsewhere classified
16. Postprocedural respiratory disorders, not elsewhere classified
17. Postprocedural disorders of digestive system, not elsewhere classified
18. Postprocedural musculoskeletal disorders, not elsewhere classified
19. Postprocedural disorders of genitourinary system, not elsewhere classified

## Appendix 2: Other health complications

Other health complications that indicate PPRs:

1. Volume depletion
2. Cardiac arrest with successful resuscitation
3. Sudden cardiac death, so described
4. Cardiac arrest, unspecified
5. Fistula of stomach and duodenum
6. Diverticulum of appendix
7. Anal fistula
8. Rectal fistula
9. Fistula of intestine
10. Haemoperitoneum
11. Fistula of gallbladder
12. Perforation of bile duct
13. Fistula of bile duct
14. Gastrointestinal haemorrhage, unspecified
15. Vesicointestinal fistula
16. Vesical fistula, not elsewhere classified
17. Vesicovaginal fistula
18. Other female urinary-genital tract fistulae
19. Fistula of vagina to small intestine
20. Fistula of vagina to large intestine
21. Other female intestinal-genital tract fistulae
22. Female genital tract-skin fistulae
23. Other female genital tract fistulae
24. Female genital tract fistula, unspecified
25. Other intrapartum haemorrhage
26. Intrapartum haemorrhage, unspecified
27. Obstetric death of unspecified cause
28. Death from direct obstetric cause
29. Death from indirect obstetric cause
30. Death from obstetric cause, unspecified
31. Death from sequelae of direct obstetric cause
32. Death from sequelae of indirect obstetric cause
33. Death from sequelae of obstetric cause, unspecified
34. Haemorrhage from other sites in respiratory passages
35. Haemorrhage from respiratory passages, unspecified
36. Cardiogenic shock
37. Hypovolaemic shock
38. Septic shock
39. Other shock
40. Shock, unspecified
41. Haemorrhage, not elsewhere classified

42. Instantaneous death
43. Death occurring less than 24 hours from onset of symptoms, not otherwise explained
44. Unattended death
45. Other ill-defined and unspecified causes of mortality
46. Post traumatic wound infection, not elsewhere classified

## Appendix 3: Preventable diseases

Preventable diseases that indicate PPRs, including:

1. Pulmonary embolism with mention of acute corpulmonale
2. Pulmonary embolism without mention of acute corpulmonale
3. Phlebitis and thrombophlebitis of femoral vein
4. Phlebitis and thrombophlebitis of other deep vessels of lower extremities
5. Phlebitis and thrombophlebitis of lower extremities, unspecified
6. Phlebitis and thrombophlebitis of other sites
7. Phlebitis and thrombophlebitis of unspecified site
8. Budd-Chiari syndrome
9. Thrombophlebitis migrans
10. Embolism and thrombosis of vena cava
11. Embolism and thrombosis of renal vein
12. Embolism and thrombosis of other specified veins
13. Embolism and thrombosis of unspecified vein
14. Stage I decubitus ulcer and pressure area
15. Stage II decubitus ulcer and pressure area
16. Stage III decubitus ulcer and pressure area
17. Stage IV decubitus ulcer and pressure area
18. Decubitus ulcer and pressure area, unspecified

## Appendix 4: Trauma and diseases that are difficult to cure

Trauma and diseases that are difficult to cure and indicate PPRs, including:

1. Idiopathic thrombocytopenic purpura
2. Multiple sclerosis
3. Alcoholic fatty liver
4. Alcoholic cirrhosis of liver
5. Toxic liver disease with fibrosis and cirrhosis of liver
6. Other and unspecified cirrhosis of liver
7. Calculus of kidney
8. Calculus of ureter
9. Calculus of kidney with calculus of ureter
10. Urinary calculus, unspecified
11. Calculus in bladder
12. Other lower urinary tract calculus
13. Calculus of lower urinary tract, unspecified
14. Urinary calculus in schistosomiasis
15. Calculus of urinary tract in other diseases classified elsewhere
16. Unspecified renal colic
17. Ascites
18. Cheek and lip biting
19. Sunburn, erythema
20. Sunburn, partial thickness
21. Sunburn, full thickness
22. Sunburn, unspecified
23. Traumatic arthropathy
24. Traumatic spondylopathy
25. Muscle strain
26. Rupture of popliteal cyst
27. Stress fracture, not elsewhere classified
28. Subluxation stenosis of neural canal
29. Chilblains
30. Exposure to radiation
31. Exposure to other pollution

# References


Agency for Healthcare Research and Quality, 2012, 'Methods: calculating readmissions for HCUPnet', Available from http://hcupnet.ahrq.gov/HCUPnet.app/Methods-HCUPnet%20readmissions.pdf [5 May 2016]

Australian Institute of Health and Welfare 2016. Available from http://www.aihw.gov.au [5 May 2016].

Allaudeen, N, Schnipper, J, Orav, E, Wachter, R, & Vidyarthi, R 2011, 'Inability of Providers to Predict Unplanned Readmissions', *Journal of General Internal Medicine*, Vol. 26, No.7, Pp771-776.

Allisonet. P 2014, 'Measures of Fit for Logistic Regression'. Available from http://statisticalhorizons.com/wp-content/uploads/GOFForLogisticRegression-Paper.pdf [15 May 2016].

American Hospital Association 2011, 'Examining the drivers of readmissions and reducing unnecessary readmissions for better patient care'. Available from http://www.aha.org/research/reports/tw/11sep-tw-readmissions.pdf [15 May 2016].

Anika, L, Marguerite, L, Barrett, M, Jiang, J & Claudia, A 2014, 'Conditions with the largest number of adult hospital readmissions by payer, 2011', *Healthcare Cost and Utilization Project*. Available from http://www.hcup-us.ahrq.gov/reports/statbriefs/sb172-Conditions-Readmissions-Payer.pdf [5 May 2016]
[5 May 2016]

Australian Bureau of Statistics 2015, 'Australian's leading causes of death, 2015'. Available from
http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/3303.0~2015~Main%20Features~Australia's%20leading%20causes%20of%20death,%202015~3 [5 May 2016].

Australian Consortium for Classification Development 2016. Available from https://www.accd.net.au/ [5 May 2016].

Averill, R, Goldfield, N, Hughes, J, Bonazelli, J, McCullough, E, Steinbeck, B, Mullin, R & Tang, A 2003, 'APR-DRGs - Methodology Overview'. Available from https://www.hcup-us.ahrq.gov/db/nation/nis/APR-DRGsV20MethodologyOverviewandBibliography.pdf [2 May 2016].

Baron. J and Weiderpass. E, 2000, 'An introduction to epidemiological research with medical databases', *Annals of Epidemiology*, Vol.10, NO.4, PP200- 4. Available from http://www.annalsofepidemiology.org/article/S1047-2797(00)00039-9/abstract pdf [14 March 2016].

Benbassat. J and Taragin. M, 2000, 'Hospital readmissions as a measure of quality of health care: advantages and limitations', *Arch Intern Med*, Vol.160, No.8, PP1074-1081.

Bottle et al. 2014, *'Effect of the readmission primary diagnosis and time interval in heart failure patients: analysis of English administrative data',* European Journal of Heart Failure, Vol.16, No.8, PP846-853.

Bhuvan, S, Kumar, A, Zafar, A, & Kishore, V 2016, "Identifying Diabetic Patients with High Risk of Readmission". Available from https://arxiv.org/pdf/1602.04257.pdf [14 March 2016].

Boulding, W, Glickman, S, Manary, M, Schulman, K, & Staelin, R 2011, 'Relationship between patient satisfaction with inpatient care and hospital readmission within 30 days', *American Journal of Managed Care*, Vol.17, NO.1 PP41-48.

Boutwell, A & Hwu, S 2009, *'Effective interventions to reduce rehospitalizations: a survey of the published evidence',* Institute for Healthcare Improvement. Available from http://www.ihi.org/resources/Pages/Publications/EffectiveInterventionsReduceRehospitalizationsASurveyPublishedEvidence.aspx [20 Arial 2016].

Chungrong, A, & Edward, C 2003, 'Interaction terms in logit and probit models', *Economics Letters*, Vol.80, No1, PP123-129.

Couris,C, Schott, A, Ecochard, R, Morgon, E & Colin, C 2003, 'A literature review to assess the use of claims databases in identifying incident cancer cases', *Health Services and Outcomes Research Methodology*, Vol.4, NO.1, PP49- 63.

Cumerlato, M 2012, 'ICD-10-AM/ACHI as the building blocks for Activity Based Funding (ABF)'. Available from https://sydney.edu.au/health-sciences/ncch/docs/mcumerlato301112.pdf [14 May 2016].

Dawes, A, Sacks, G, Russell, M, Lin, A, Maggard-Gibbons, M, Winograd, D, Chung, HR, Tillou, A, Hiatt, JR, & Ko, C 2014, 'Preventable readmissions to surgical services: lessons learned and targets for improvement', *American College of Surgeons,* Vol. 219, No.3.

Dictionary.com 2006. Available from http://www.dictionary.com/browse/medical-record [14 May 2016].

Distel, E, Casey, M, & Prasad, S 2016, 'Reducing potentially-preventable readmissions in critical access hospitals', *Flex Monitoring Team Policy*, Vol. 43. Available from http://www.flexmonitoring.org/wp-content/uploads/2016/03/PB43-readmissions.pdf [20 Arial 2016].

Donze, J, Aujesky, D, Williams, D & Schnipper, J 2013, 'Potentially avoidable 30-Day hospital readmissions in medical Patients', *JAMA,* Vol.173, No.8.

Donze, J, Lipsitz, S, & Schnipper, J. 2014, 'Risk factors for potentially avoidable readmissions due to end-of-life care issues', *Journal of Hospital Medicine,* Vol. 9, No. 5, PP310-314.

Edelman, T, 2016, 'Reducing Hospital Readmissions by Addressing the Causes', *Medical Advocacy.* Available from http://www.medicareadvocacy.org/reducing-hospital-readmissions-by-addressing-the-causes/ [20 Arial 2016].

Elixhauser, A, Steiner, C, Harris, D & Coffey, R 1998, 'Comorbidity Measures for Use with Administrative Data', *Medical Care,* Vol.36, No.1, PP8-27.

Farraris, V, Ferraris, S, Harmon, R & Evans, B 2001, 'Risk factors for early hospital readmission after cardiac operations', *Surgery for Acquired Cardiovascular Disease*, Vol.122, No.2, PP278-286.

Ferver,K, Burton,B, and Jesilow,P 2009, *'The use of claims data in healthcare research',* The Open Public Health Journal*,* Vol. 2, P11-24. Available from *http://benthamopen.com/contents/pdf/TOPHJ/TOPHJ-2-11.pdf* [14 March 2016].

Garcia, J, Farrero, E, Félez, M, Izquierdo, J, Marrades, R & Antó, J 2003, 'Risk factors of readmission to hospital for a COPD exacerbation: a prospective study', *Thorax*, Vol.58, PP100-105.

Ganguly, I, Ausfahl, C, Dworkin, J & Lakhanpal, R 2013, 'Solving Preventable Readmissions: Challenges, Strategies and the Need for a Clinical Analytics Solution'. Available from http://njhimss.org/images/documents/whitepapers/Solving-Preventable-Readmissions-White_Paper.pdf [20 Arial 2016].

Geng, M 2001, 'A comparison of logistic regression to random forests for exploring differences in risk factors associated with stage at diagnosis between black and white colon cancer patients'. Available from http://d-scholarship.pitt.edu/7034/1/realfinalplus_ETD2006.pdf [20 Arial 2016].

Goldfield, N, McCullough, E, Hughes, J, Tang, A, Eastman, B, Rawlins, L & Averill, R 2008*,* 'Identifying Potentially Preventable Readmissions', *Health Care Finance Review*, Vol.30, No.1, PP75-91.

Gokatas, A & Isci, O 2011, 'A comparison of the most commonly used measures of association for doubly ordered square contingency tables via simulation', *Metodoloski Zecezki*, Vol.8, No.1, PP17-37. Available from http://www.stat-d.si/mz/mz8.1/goktas.pdf [20 Sep 2016].

Goodman,L & Kruskal, W 1972, 'Measures of association for cross classifications, IV: simplification of asymptotic variances', *Journal of the American Statistical Association*, Vol.67, No.338, PP415–421.

Halfon, P, Eggli, Y, Melle, G, Chevalier, J, Wasserfallen, J & Burnand, B 2002, 'Measuring potentially avoidable hospital readmissions', *Journal of Clinical Epidemiology*, Vol.55, No.6, PP573–587.

Halfon, P, Eggli, Y, Melle, G, Chevalier, J, Wasserfallen, J & Burnand, B 2006, 'Validation of the potentially avoidable hospital readmission rate as a routine indicator of the quality of hospital care', *Medical Care*, Vol.44, No.11, PP.972-981.

Hannan. E, Racz, M, Walford, G, Ryan, T, Isom, O, Bennett, E, & Jones, R 2003, 'Predictors of readmission for complications of coronary artery bypass graft surgery', *JAMA*, Vol.290, No.6, pp.773-780.

Hain. P, Hain,P , H, Gay, J, Berutti, T, Whitney, G, Wang, W, & Saville, B 2013, 'Preventability of Early Readmissions at a Children's Hospital', *Prediatrics*, Vol.131, No1.

Health NSW, 2015. Available from http://www.health.nsw.gov.au/hospitals/going_to_hospital/pages/default.aspx [7 May 2016]

Heart Foundation, 2016, 'Cardiovascular disease, heart disease and heart attack'. Available from https://heartfoundation.org.au/about-us/what-we-do/heart-disease-in-australia [7 May 2016].

Heggestad, T, 2002, 'Do hospital length of stay and staffing ratio affect elderly patients' risk of readmission? A nation-wide study of Norwegian hospitals', *Health Services Research*, Vol.37, No.3, PP647-665.

Hicks, J, 2003, 'The potential of claims data to support the measurement of healthcare quality'. Available from http://www.rand.org/pubs/rgs_dissertations/RGSD171.html [20 Arial 2016]

Hosseinzadeh, A, Izadi, M, Verma, A, Precup, D & Buckeridge, D 2013, 'Assessing the predictability of hospital readmission using machine learning', *Association for the Advancement of Artificial Intelligence,* PP1532-1538. Available from http://www.aaai.org/ocs/index.php/IAAI/IAAI13/paper/viewFile/6475/6431[20 Arial 2016].

Jackson, A, Fireman, E, Feigenbaum, P, Neuwirth, E, Kipnis, P & Bellows, J 2013, 'Manual and automated methods for identifying potentially preventable readmissions: a comparison in a large healthcare system', *BMC Medical Informatics and Decision Making*, Vol.24, No.28.

Jiang. H, Russo, & Barrett, M 2006, 'Nationwide Frequency and Costs of Potentially Preventable Hospitalizations, 2006', *H.CUP*. Available from http://www.hcup-us.ahrq.gov/reports/statbriefs/sb72.jsp [20 Arial 2016].

Katterl, R, Anikeeva, O, Butler, C, Brown, L, Smith, B & Bywood, P 2012, 'Potentially avoidable hospitalisations in Australia: Causes for hospitalisations and primary health care interventions', *Primary Health Care Research & Information Service*. Available from http://www.phcris.org.au/phplib/filedownload.php?file=/elib/lib/downloaded_files/publications/pdfs/phcris_pub_8388.pdf [20 Arial 2016]

Kilkenny, M, Longworth, M, Pollack, M, Levi, C & Cadilhac, D 2013, 'Factors associated with 28-day hospital readmission after stroke in Australia', *American Heart Association*, Vol. 44, No.8, PP 2260–2268.

Krumholz, H, Lin, Z, Drye, E, Desai, M, Han, L, Rapp, M, Mattera, J, & Normand, S 2011, 'An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction', *Circulation: Cardiovascular Quality and Outcomes,* Vol.4, PP243-252.

Kripalani. S, LeFevre, F, Phillips, C, Williams, M, Basaviah, P, & Baker, D 2007, 'Deficits in communication and information transfer between hospital-based and primary care physicians', *JAMA*, Vol.297, No. 28, PP831–841.

Kulinskaya, E, Kornbrot, D & Gao, H 2005, 'Length of stay as a performance indicator: robust statistical methodology', *Journal of Management Mathematics*, Vol.16, PP369-381.

Kumbhani. D, Healey, N, Thatte, H, Birjiniuk, V, Crittenden, M, Treanor, P & Khuri, S 2009, 'Intraoperative myocardial acidosis as a risk for hospital readmission after cardiac surgery', *Am J Surg*, Vol.198, PP373-380.

Lavenberg, L, Leas, B, Umscheid, CA, Williams, K, Goldmann, DR & Kripalani, S 2014, 'Assessing preventability in the quest to reduce hospital readmissions', *J Hosp Med.,* Vol.9, No.9, PP598-603.

Li. J, Yong, T, Hakendorf, P, Ben-Tovim, D & Thompson, C 2015, 'Identifying risk factors and patterns for unplanned readmission to a general medical service', *Australian Health Review,* Vol.39, PP56–62.

Lichtman. J, Leifheit-Limson, E, Jones, S, Wang, Y & Goldstein, L 2013, 'Preventable readmissions within 30 days of ischemic stroke among Medicare beneficiaries', *STROKEAHA*, VOL.44, No.12, PP3429-3435.

Lin. X, 2012, 'A predictive random forest model on hospital 30-day readmission using electronic health records'. Available from file:///C:/Users/jwu.LORICA/Downloads/lin_dissertation.pdf [21 Jun 2016].

Liu. Y, Phillips, M & Codde, J 2001, 'Factors influencing patients 'length of stay Australian Health Review', Vol.24, No.2, PP63-70.

Makris, N, Dulhunty, J, Paratz, J, Bandeshe & H, Gowardman, J 2010, 'Unplanned early readmission to the intensive care unit: a case-control study of patient, intensive care and ward-related factors', *Anaesth Intensive Care*, Vol.38, Pp723-731.

Manitoba Centre for Health Policy 2016, 'Concept: Charlson Comorbidity Index'. Available from http://mchp-appserv.cpe.umanitoba.ca/viewConcept.php?conceptID=1098 [21 Mar 2016]

McLean, R, Mendis, K & Canalese, J 2008, 'A ten-year retrospective study of unplanned hospital readmissions to a regional Australian hospital', *Aust Health Rev*, Vo. 32, No.3, PP537–547. Available from
http://www.publish.csiro.au/ah/pdf/AH080537 1098 [22 Mar 2016]

Moore, C, Gao, K & Shulan, M 2013, 'Racial, income, and marital status disparities in hospital readmissions within a veterans-integrated health care network', *Eval Health Prof*, Vol.38, PP491-507.

Morris, A, Flowers, C, Morris, K, Schmidt, W, Pommier, R &Vetto, J 2003, 'Comparing the cost-effectiveness of the triple test score to traditional methods for evaluating palpable breast masses', *Med Care*, Vol.41, NO.8, PP962-71.

Nahab, F, Takesaka, I, Mailyan, E, Judd, L, Culler, L, Webb, A, Frankel, M, Choi, D, & Helmers, S 2012, 'Avoidable 30-day readmissions among patients with stroke and other cerebrovascular disease', *Neurohospitalist*, Vol.2, No.1, PP7-11.

National Institute on Drug Abuse, 2010. Available from https://www.drugabuse.gov/publications/research-reports/comorbidity-addiction-other-mental-illnesses/what-comorbidity [21 May 2016]

Hasan, O, Meltzer, D, Shaykevich, S, Bell, C, Kaboli, P, Auerbach, A, Wetterneck, T, Arora, V, Zhang, J & Schnipper, J 2010, 'Hospital readmission in general medicine patients: a prediction model', *Journal of General Internal Medicine*, Vol.25, NO.3.

Panacek, E 2007 'Performing chart review studies', *Air Med J*, Vol.26, NO.5, PP206-10.

QualityNet, 2016, 'Hospital Readmissions Reduction Program'. Available from https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1228772412458 [20 May 2016].

Renton. J, Pilcher, D, Santamaria, J, Stow, P, Bailey, M, Hart, G & Duke, G 2011, 'Factors associated with increased risk of readmission to intensive care in Australia', *Intensive Care Med*, Vol.37, No.11, PP1800-1808.

Private Healthcare Australia, 2015. Available from http://www.privatehealthcareaustralia.org.au/wp-content/uploads/PHA-Media-Release-No-More-Excuses-24Aug15.pdf [14 March 2016]

Rich, M, Vinson, J, Sperry, J, Shah, A, Spinner, L, Chuang, M & Roman, V 1993, 'Prevention of readmission in elderly patients with congestive heart failure: results of a prospective, randomized pilot study', *J Gen Intern Med*, Vol.8, No.11, PP585-90.

Carroll, S, Edwards, J & Lashbrook, A 2011, 'Reducing hospital readmissions: lessons from top-performing hospitals', *The Commonwealth Fund*, Vol.5, PP1-20.

Rosen, A, Chen, Q, Shin, M, O'Brien, W, Shwartz, M, Mull, H, Cevasco, M & Borzecki, A 2014, 'Medical and surgical readmissions in the Veterans Health Administration: what proportion are related to the index hospitalization?', *Med Care.,* Vol. 52, No.3, PP243-249.

Shalchi, Z, Saso, S, Li, H, Rowlandson, E & Tennant, R 2009, 'Factors influencing hospital readmission rates after acute medical treatment', *Clin Med (Lond),* Vol. 9 No.5, PP426-30.

Shams, I, Ajorlou, S & Yang, K 2015, 'A predictive analytics approach to reducing 30-day avoidable readmissions among patients with heart failure, acute myocardial infarction, pneumonia, or COPD', *Health Care Management Science*, Vol.18, No.1, Pp19-34.

Saunders, N, Nichols, S, Antiporda, M, Johnson, K, Walker, K, Nilsson, R, Graham, L, Old, M, Klisovic, R, Penza, S & Schmidt, C 2015, 'Examination of Unplanned 30-Day Readmissions to a Comprehensive Cancer Hospital', *Journal of Oncology Practice*, Vol.11, No.2, PP177-181.

SQLape 2016, Available from www.sqlape.com [20 Feb 2016].

Miles, T & Lowe, J 1999, 'Are unplanned readmissions to hospital really preventable?', *Journal of Quality in Clinical Practice,* Vol. 19, No.4, PP 211–214.

Trueck. S & Rachev.S.T 2009, 'Rating based modeling of credit risk'*, Academic Press*, 2009, PP21-23.

Tyree, P, Bonnie, K & Lafferty, W 2006, 'Challenges of using medical insurance claims data for utilization analysis'*, American Journal of Medical Quality,* 2006 August, Vol. 21, No. 4 269-275.

UW Data Resources in the Health Sciences, 2016. Available from http://guides.lib.uw.edu/c.php?g=99209&p=642709#12207218 [20 May 2016].

Vest. J, Gamm, L, Oxford, B Gonzales, M & Slawson, K 2010, 'Determinants of preventable readmissions in the United States: a systematic review', *Implementation Science*, 2010, Vol.88. Available from http://implementationscience.biomedcentral.com/articles/10.1186/1748-5908-5-88 [14 March 2016]

WA Department of Health Clinical Casemix Handbook 2014. Available from http://www.health.wa.gov.au/activity/docs/ABF-ABM_Clinical_Casemix_Handbook_2012-2014.pdf [14 Mar 2016].

Walraven, C, Bennett, C, Jennings, A, Austin, PC & Forster, A 2011, 'Proportion of hospital readmissions deemed avoidable: a systematic review', *CMAJ*, Vol.183, No.7, PP391-402.

Weeks, W, Lee, R, Wallace, A, West, A & Bagian, J 2009, 'Do older rural and urban veterans experience different rates of unplanned readmission to VA and Non-VA hospitals'*, The Journal of Rural Health*, Vol. 25, No. 1, PP62-69.

World Health Organisation (WHO) 2016. Available from http://www.who.int/en/ [14 July 2016].

Yam, C, Wong, E, Chan, F, Leung, M, Wong, F, Cheung, AW & Yeoh, E 2010, 'Avoidable readmission in Hong Kong - system, clinician, patient or social factor?', *BMC Health Services Research*, Vol.10, No. 311.

Zeng, L, Smriti, N, Jamie, R, Seidensticker, S, Clark , C, Sonstein, L, Trevino, R & Sharma, G 2014, 'Statistical Models for Hospital Readmission Prediction with
Application to Chronic Obstructive Pulmonary Disease (COPD)
Patients'. Available from http://ieomsociety.org/ieom2014/pdfs/37.pdf [14 July 2016].

Zolfaghar, K, Verbiest, N, Agarwal, J, Meadem, N, Chin, S, Roy, S, Teredesai, A, Hazel, D, Amoroso, P & Reed, L 2013, 'Predicting risk-of-readmission for congestive heart failure patients: a multi-layer approach', Available from https://arxiv.org/abs/1306.2094 88 [14 July 2016].