The dissertation of Yifeng Yu was reviewed and approved* by the following:

Harriet B. Nembhard
Professor of Industrial Engineering
Dissertation Advisor
Chair of Committee

Conrad Tucker
Assistant Professor of Industrial Engineering

Ling Rothrock
Professor of Industrial Engineering

Naleef Fareed
Assistant Professor of Health Policy and Administration

Andrea Yevchak-Sillner
Assistant Professor of Nursing
Special Member

Robert Voigt
Professor of Industrial Engineering
Chair of Graduate Program

*Signatures are on file in the Graduate School
ABSTRACT

Total hip arthroplasty (THA) is the most effective surgical intervention for patients with severe hip arthritis or hip injury. It has been shown to significantly improve patients’ quality of life with respect to physical function, pain relief, and overall health.

In recent years, unplanned readmission after THA has become an increasingly serious problem in the United States. Unplanned readmission significantly reduces the quality of life of THA patients. Moreover, in fiscal year 2015, the Centers for Medicare and Medicaid Services (CMS) started to penalize hospitals for high 30-day readmission rates after THA surgery. Relevant studies indicated that readmission after THA surgery is deemed a healthcare quality indicator and that a considerable portion of 30-day unplanned readmission after THA can be effectively prevented. Therefore, it is necessary to mitigate the potential problems and improve THA care quality in order to reduce unplanned readmission.

Moreover, the demand for THA surgery is expected to increase from 326,100 in 2010 to 572,000 in 2030 in the United States. However, the available capacity of THA may be insufficient for such a massively increasing demand. In this landscape, effectively using THA surgery to optimize patient outcome becomes an important topic.

With the advancement in healthcare information technology, healthcare data has become exponentially available. Data science and engineering is needed in the healthcare ecosystem to support medical decision making, facilitate quality improvement initiatives, and improve patient outcomes. However, data science and engineering has not been adequately used in improving THA patient outcomes. The objective of this doctoral dissertation is threefold: 1) to advance data-driven healthcare for THA, 2) to provide insights into informed medical decision making, and 3) to support and guide efforts in THA care quality improvement.

Since 30-day unplanned readmission has become an increasingly serious problem and hospitals now bear the risk of CMS penalties, it is very important for hospitals to understand their THA surgical performance in real time. We obtained identified THA patient-level data records from an academic medical center in central Pennsylvania, United States. We proposed using machine learning algorithms to conduct patient risk stratification, and combining them with statistical process control to perform surgical outcome evaluation. The results indicate that random forest outperforms the most commonly-used risk-adjustment method, logistic regression, in identifying high-risk patients. Therefore, the control chart based on random forest provides more convincing results on surgical outcome evaluation. With our
proposed risk-adjusted monitoring framework, the medical team can better target interventions on future high-risk patients, and diagnose potential care quality problems in a timely manner.

Due to the possible shortage of THA surgery in the next 15 years, it is important to take cost-effectiveness into account when determining the appropriate treatment option for each patient. We proposed a method for clustering hip arthritis patients and analyzing the cost-effectiveness of THA surgery for each patient subgroup. The results indicate that THA surgery is more cost-effective for relatively young patients with few co-morbidities, while it is not cost-effective for the oldest patient group in this study. With our proposed method, coupled with the characteristics of each patient subgroup, the medical team can better determine the optimal treatment option for each hip arthritis patient in order to improve patient outcomes.

Finally, when implementing medical or surgical interventions on THA patients, convincing evidence must show that a certain intervention really benefits the patients. The problem, however, is that relevant studies show mixed results on the effectiveness of a certain intervention. To address this issue, we used meta-analysis to deliver a valid and comprehensive understanding of the effectiveness of an intervention and we analyzed pre-operative exercise as an example. The results indicate that while pre-operative exercise provides moderate benefits with respect to pain relief, physical function, and activity of daily living before surgery, these benefits diminish and become insignificant post-operatively. Therefore, our meta-analysis provides important insights into what medical or surgical interventions have proven effectiveness in improving THA care quality and patient outcomes, based on the mixed results of existing literature.

In conclusion, this data-driven approach to THA advances the translation of data science and engineering in healthcare, by using the increasingly available healthcare data to deliver insightful solutions. This approach can be used to facilitate efforts in THA care quality improvement, with the purpose of managing care quality and improving THA patient outcomes.
# TABLE OF CONTENTS

List of Figures ........................................................................................................... viii
List of Tables ................................................................................................................ x
List of Abbreviations ................................................................................................... xii
Acknowledgements ....................................................................................................... xiii

Chapter 1 Introduction ................................................................................................. 1

1.1 Total hip arthroplasty (THA) background .............................................................. 1
1.2 Unplanned readmission after THA ......................................................................... 3
1.3 Insufficient capacity of THA .................................................................................. 5
1.4 Pre-operative exercise intervention for THA patients .............................................. 6
1.5 Healthcare quality, as well as data science and engineering .................................. 7
1.6 Research objective and framework ......................................................................... 8
1.7 Outline ..................................................................................................................... 10

Chapter 2 Risk-Adjusted Control Charts for Monitoring THA Outcomes ................. 12

2.1 Introduction ............................................................................................................. 12
2.2 Literature review .................................................................................................... 13
  2.2.1 Risk factors of readmission after THA ............................................................. 13
  2.2.2 Risk-adjusted SPC ........................................................................................ 15
2.3 Research objective ................................................................................................. 16
2.4 Risk-adjustment methods ....................................................................................... 17
  2.4.1 Logistic regression .......................................................................................... 18
  2.4.2 Decision tree .................................................................................................. 19
  2.4.3 Random forest ............................................................................................... 19
2.5 Evaluation criteria of risk-adjustment methods ...................................................... 20
2.6 Cumulative sum (CUSUM) control chart .............................................................. 22
2.7 Study setting .......................................................................................................... 23
2.8 Results .................................................................................................................... 25
  2.8.1 Risk adjustment by the three machine learning algorithms ......................... 25
  2.8.2 Risk factors identified by the three risk-adjustment methods ....................... 29
  2.8.3 CUSUM charts based on risk-adjustment .................................................... 30
2.9 Discussion .............................................................................................................. 33
2.10 Conclusion ............................................................................................................. 35

Chapter 3 Cost-Effectiveness of THA Surgery for Patient Subgroups ......................... 36

3.1 Introduction .......................................................................................................... 36
3.2 Literature review .................................................................................................... 37
  3.2.1 Studies on the cost-effectiveness of THA ....................................................... 37
  3.2.2 Studies on cluster analysis in healthcare ....................................................... 40
3.3 Research objective ............................................................................................... 41
3.4 Methods ................................................................................................................. 41
3.4.1 Supervised machine learning algorithms ................................................. 42
3.4.2 Unsupervised machine learning algorithms ........................................... 42
3.4.3 Markov simulation model ................................................................. 45
3.4.4 Cost parameters .................................................................................. 48
3.4.5 Quality of life parameters .................................................................... 52
3.5 Results ...................................................................................................... 53
  3.5.1 Risk factors identified by supervised learning algorithms ................. 53
  3.5.2 Cluster analysis of patients ................................................................. 54
  3.5.3 Cost-effectiveness analysis of each patient cluster ............................. 56
3.6 Sensitivity analysis .................................................................................. 58
3.7 Discussion ............................................................................................... 59
3.8 Conclusion .............................................................................................. 62

Chapter 4 Meta-Analysis of the Effectiveness of Pre-operative Exercise ............... 64

4.1 Introduction ............................................................................................. 64
4.2 Literature review ..................................................................................... 65
4.3 Research objective ................................................................................... 70
4.4 Methods ..................................................................................................... 70
  4.4.1 Search methodology and eligibility criteria ......................................... 70
  4.4.2 PEDro scale ....................................................................................... 71
  4.4.3 Meta-analysis and quantifying heterogeneity ...................................... 75
4.5 Effect measures of pre-operative exercise ................................................ 77
  4.5.1 WOMAC ............................................................................................ 77
  4.5.2 Harris Hip Score ................................................................................ 78
  4.5.3 HOOS ................................................................................................ 78
4.6 Results ..................................................................................................... 79
  4.6.1 Methodological quality of included studies ......................................... 79
  4.6.2 Meta-analysis on WOMAC ................................................................. 82
  4.6.3 Meta-analysis on Harris Hip Score ...................................................... 87
  4.6.4 Meta-analysis on HOOS ..................................................................... 90
  4.6.5 Meta-analysis on LOS ....................................................................... 99
4.7 Discussion ............................................................................................... 101
4.8 Conclusion .............................................................................................. 103

Chapter 5 Conclusion ..................................................................................... 105

5.1 Monitoring THA surgical outcomes ......................................................... 107
5.2 Cost-effectiveness of THA for patient subgroups ..................................... 108
5.3 Meta-analysis of the effectiveness of pre-operative exercise .................... 109
5.4 Limitations .............................................................................................. 110
  5.4.1 Data .................................................................................................... 110
  5.4.2 Assumptions ....................................................................................... 111
  5.4.3 Quality factors ................................................................................... 112
  5.4.4 Generalizability ................................................................................ 112
5.5 Future work ............................................................................................ 113
  5.5.1 Patient risk stratification and surgical outcome evaluation ................ 113
  5.5.2 Cost-effectiveness of THA for each patient subgroup ......................... 114
  5.5.3 Meta-analysis of the effectiveness of pre-operative exercise .............. 114
Appendix A  R code for logistic regression, decision tree, and random forest ...............116
Appendix B  R code for determining ARL₀ of CUSUM control charts .....................120
Appendix C  R code for CUSUM control charts based on logistic regression and random forest ..........................................................................................................................127
Appendix D  R code for K-means clustering and hierarchical clustering .....................131
Appendix E  Comparison of results of K-means clustering and hierarchical clustering...143
Appendix F  R code for Markov simulation model of cluster 1 patients ......................148
Appendix G  R code for meta-analysis ..........................................................................156
References ..................................................................................................................162
LIST OF FIGURES

Figure 1-1. The artificial hip joint and its position in the patient.................................2
Figure 1-2. The framework of this dissertation research...............................................9
Figure 2-1. 30-day unplanned readmission rates during 2011-2015.................................24
Figure 2-2. Readmission risks of all the patients estimated by logistic regression.............28
Figure 2-3. Readmission risks of all the patients estimated by random forest.................28
Figure 2-4. CUSUM chart (top for deterioration and bottom for improvement) by logistic regression ..................................................................................................................31
Figure 2-5. CUSUM chart (top for deterioration and bottom for improvement) by random forest ..................................................................................................................32
Figure 3-1. Markov simulation model of the surgery option.............................................45
Figure 3-2. Total within-cluster variance in K-means clustering....................................55
Figure 3-3. Cost-effectiveness acceptability curves for each patient cluster ..................59
Figure 4-1. The PRISMA diagram of study selection.......................................................80
Figure 4-2. WOMAC pain score before surgery..............................................................83
Figure 4-3. WOMAC pain score at three months after surgery.......................................84
Figure 4-4. WOMAC pain score at three months after surgery in jackknife analysis ........85
Figure 4-5. WOMAC function score before surgery .......................................................86
Figure 4-6. WOMAC function score at three months after surgery ...............................86
Figure 4-7. Harris Hip Score before surgery.................................................................88
Figure 4-8. Harris Hip Score at discharge.....................................................................89
Figure 4-9. Harris Hip Score at three months after surgery ...........................................89
Figure 4-10. HOOS activity of daily living score before surgery....................................91
Figure 4-11. HOOS activity of daily living score at six weeks after surgery ....................92
Figure 4-12. HOOS pain score before surgery .............................................................93
Figure 4-13. HOOS pain score at six weeks after surgery .............................................93
Figure 4-14. HOOS symptoms score before surgery .........................................................94
Figure 4-15. HOOS symptoms score at six weeks after surgery ........................................95
Figure 4-16. HOOS sport and recreation score before surgery ........................................96
Figure 4-17. HOOS sport and recreation score before surgery after jackknife analysis ..........96
Figure 4-18. HOOS sport and recreation score at six weeks after surgery ...........................97
Figure 4-19. HOOS hip-related quality of life score before surgery ....................................98
Figure 4-20. HOOS hip-related quality of life score at six weeks after surgery .....................98
Figure 4-21. LOS of the intervention group and the control group .....................................100
LIST OF TABLES

Table 2-1. A summary of risk factors of THA readmission ......................................... 14
Table 2-2. A summary of the studies on SPC in healthcare ........................................ 16
Table 2-3. An example of the confusion matrix ............................................................ 21
Table 2-4. Comparison of patients on continuous variables ........................................... 24
Table 2-5. Comparison of patients on categorical variables .......................................... 25
Table 2-6. AUC of the three risk-adjustment methods with different training dataset sizes ... 27
Table 2-7. Brier score of the three risk-adjustment methods with different training dataset sizes.. 27
Table 2-8. The confusion matrix with probability 0.4 as threshold by random forest ........ 29
Table 2-9. The top-10 high-frequency predictors of readmission risk ............................ 30
Table 3-1. A summary of studies on cost-effectiveness of THA .................................... 38
Table 3-2. A summary of studies on cluster analysis in healthcare ................................. 40
Table 3-3. A summary of different linkages in hierarchical clustering ............................. 44
Table 3-4. Transition probabilities between health states .............................................. 47
Table 3-5. A summary of medical resources needed by patients .................................... 49
Table 3-6. A summary of annual medical costs for surgery option and non-surgery option ... 51
Table 3-7. A summary of the HRQoL parameters .......................................................... 53
Table 3-8. K-means clustering results with six clusters .................................................. 56
Table 3-9. Hierarchical clustering results with six clusters ............................................. 56
Table 3-10. Cost-effectiveness ratios of THA surgery for each cluster ............................. 57
Table 4-1. Studies on the effect of pre-operative exercise on THA patients ...................... 66
Table 4-2. Methodological quality of the 11 included studies ......................................... 81
Table 4-3. The details of WOMAC in the three included studies ..................................... 82
Table 4-4. Mean and standard deviation of WOMAC pain scores in these 3 studies ........ 83
Table 4-5. Mean and standard deviation of WOMAC function scores in these 3 studies .... 85
Table 4-6. The details of Harris Hip Score in the four included studies .................................87
Table 4-7. Mean and standard deviation of Harris Hip Score in these 4 studies ......................88
Table 4-8. The details of HOOS in the four included studies ..................................................90
Table 4-9. Mean and standard deviation of HOOS activity of daily living scores in these 4 studies ...........................................................................................................................................91
Table 4-10. Mean and standard deviation of HOOS pain scores in these 4 studies .............92
Table 4-11. Mean and standard deviation of HOOS symptoms scores in these 4 studies ......94
Table 4-12. Mean and standard deviation of HOOS sport and recreation scores in these 4 studies .................................................................................................................................95
Table 4-13. Mean and standard deviation of HOOS hip-related quality of life scores in these 4 studies ..................................................................................................................................................97
Table 4-14. Mean and standard deviation of LOS in these 4 studies .....................................99
LIST OF ABBREVIATIONS

AAHKS: American Association of Hip and Knee Surgeons

ARLo: average in-control run length

AUC: area under the curve

BMI: body mass index

CMS: Centers for Medicare and Medicaid Services

CUSUM: cumulative sum

EQ-5D: EuroQol five dimensions questionnaire

HOOS: Hip disability and Osteoarthritis Outcome Score

HRQoL: health-related quality of life

LOS: length of stay

PEDro: Physiotherapy Evidence Database

QALY: quality-adjusted life year

SD: standard deviation

SNF: skilled nursing facility

SPC: statistical process control

THA: total hip arthroplasty

TKA: total knee arthroplasty

WOMAC: Western Ontario and McMaster Universities Arthritis Index
ACKNOWLEDGEMENTS

It has been a long journey to complete this dissertation, but I have not travelled alone. Many individuals have constantly supported and helped me.

First of all, I would like to express my special thanks to my advisor, Dr. Harriet Nembhard, for her guidance and support during my PhD study. Her vision and leadership was the underpinning of this work. Her encouragements strengthened my determination to accomplish this dissertation.

I also want to express my gratitude to my committee members: Dr. Conrad Tucker, Dr. Ling Rothrock, Dr. Naleef Fareed and Dr. Andrea Yevchak-Sillner. They have all offered very insightful advice on my dissertation and helped me greatly during my PhD study. I also would like to thank Dr. Mehmet Kilinc for his suggestions.

I would like to thank the National Science Foundation – Center for Health Organization Transformation (NSF-CHOT). Participating in CHOT conferences enabled me to obtain valuable experience in healthcare by interacting with researchers from multiple universities and members of the Industry Advisory Board. I also want to express my appreciation to the Clinical and Translational Science Institute (CTSI) of the Pennsylvania State University for providing data records on total hip arthroplasty patients.

Last but not least, I am extremely grateful to my family and friends for their support and encouragement. Since I was young, my parents have shaped me in a way that formed my positive, enthusiastic, and open-minded personality. Their unconditional love has guided my life. I am also grateful to my friends at the State College Chinese Alliance Church. Their fellowship and prayers provided me with great peace and joy during my PhD study.
Chapter 1 Introduction

1.1 Total hip arthroplasty (THA) background

Orthopedic surgery is the branch of surgery concerned with conditions involving the musculoskeletal system. The modern THA was pioneered by Sir John Chamley, expert in tribology at Wrightington Hospital, England in the 1960s. He found that joint surfaces could be replaced by implants cemented to the bone. His design consisted of a stainless steel one-piece femoral stem and head, as well as a polyethylene, acetabular component. Both of these components were fixed to the bone using bone cement. For over two decades, the Chamley Low Friction Arthroplasty and its derivative designs were the most frequently-used systems in the world (Wroblewski, 2002). This formed the basis for modern THA. Thus, so far THA has a history of more than 50 years. Its procedures are relatively standardized and mature in comparison with other types of surgery.

Modern orthopedic surgery and musculoskeletal research have sought to make surgery less invasive and to make implanted components better and more durable. However, current THA fundamentals have not dramatically changed for 50 years. Basically, damaged bone and cartilage are removed and replaced by prosthetic components. Figure 1-1 shows the artificial hip joint and its position in the patient (OrthoInfo, 2015).
THA is a very effective surgery for most of the patients with severe arthritis or hip injury, and it has been shown to significantly improve patients’ quality of life with respect to physical function, pain relief, and overall health (Laupacis et al., 1993; Jones, Voaklander, Johnston, & Suarez-Almazor, 2001). In 2010, 326,100 THA surgeries were performed among inpatients of all ages in the United States. Inpatients aged 45 and over accounted for approximately 95% of all these THAs. The annual number of THA among inpatients aged 45 and over more than doubled between 2000 and 2010, increasing from 138,700 in 2000 to 310,800 in 2010 (Wolford, Palso, & Bercovitz, 2015). It is forecasted that by 2030, 572,000 THA surgeries will be conducted annually in the United States (Kurtz, Ong, Lau, Mowat, & Halpern, 2007). It should also be noted that during 2000-2010, the increasing trends in THA among patients in age groups 45-54 (160%) and 55-64 (85%) were greater than the trends in age groups 65-74 (62%) as well as 75 and over (68%). In 2000, the percentages of THA performed on patients aged 45-54, 55-64, 65-74, as well as 75 and over were 12%, 24%, 34%, and 30%, respectively. But in 2010 these percentages changed to 17%, 29%, 28%, and 26%, respectively (Wolford et al., 2015). These statistics indicate that gradually a higher percentage of THA surgery is being performed on relatively young patients.
1.2 Unplanned readmission after THA

During the post-discharge period, the THA patients sometimes develop surgical complications, leading to unplanned hospital readmission. The common surgical complications include dislocation, surgical site infection, hematoma, non-infected draining wound, and deep vein thrombosis (Centers for Medicare and Medicaid Services (CMS), 2016a; Mahomed et al., 2003; Husted, Otte, Kristensen, Ørsnes, & Kehlet, 2010; SooHoo, Farng, Lieberman, Chambers, & Zingmond, 2010; Schairer, Sing, Vail, & Bozic, 2014). Recently, 30-day readmission rates of 4% to 11% after THA have been reported in the United States (not necessarily Medicare patients) (Clement et al., 2013; Halawi et al., 2015; Healy et al., 2016; Keeney et al., 2015; Ramkumar et al., 2015; Bohl, Shen, Kayupov, & Della Valle, 2016). This is unacceptable because THA is among the most effective surgical interventions that consistently provide function improvement and pain relief (Laupacis et al., 1993; Jones et al., 2001). Post-operative complications and readmissions significantly reduce the quality of life of THA patients (Towheed & Hochberg, 1996; Enocson et al., 2009). Moreover, additional costs are incurred for treating readmitted patients. The whole cost of THA surgery varies among providers: the Medicare expenditure for surgery, hospitalization, and recovery ranges from $16,500 to $33,000 across geographic areas (CMS, 2016b). CMS has not disclosed the cost of treating THA readmission, but other agencies estimated that the cost of treating a THA readmission averages $12,300 (Qasim & Andrews, 2012; Becker’s Infection Control & Clinical Quality, 2013). Thus, the most conservative estimate is that $170 million were spent in treating THA-relevant readmission in 2015 in the United States. These expenditures are expected to double by 2030, as the number of THA surgeries keeps increasing.

In order to effectively reduce readmission after THA surgery, in fiscal year 2015 CMS started to penalize hospitals for high 30-day readmission rates after THA. CMS reduces payments to underperforming hospitals with higher rates of 30-day readmission, compared with the risk-adjusted national average rate. In fiscal year 2015, 2559 of 3479 hospitals in the CMS Hospital Readmission Reduction Program had a hip/knee arthroplasty excess readmission ratio (CMS, 2016c). The total amount of financial penalty for excess hip/knee hospital readmission is the greatest driver of the increase in the overall CMS penalty by almost $200 million in fiscal year
2015 (Becker’s Hospital CFO, 2014; Kaiser Health News, 2014). Therefore, now hospitals bear considerable financial risks for readmission after THA. Apparently, unplanned readmission after THA harms both patients and hospitals.

Although researchers have different opinions on the exact percentage, they all agree that a considerable portion of the 30-day unplanned readmission after THA can be effectively prevented (Pellegrini, Donaldson, Farber, Lehman, & Evarts, 2005; Keeney et al., 2015; Weinberg, Kraay, Fitzgerald, Sidagam, & Wera, 2016). For instance, relevant studies indicated that patients with previous surgery history near the hip (not necessarily THA surgery) have increased risk of dislocation after THA surgery, especially within the 30-day post-discharge period (Lewinnek, Lewis, Tarr, Compere, & Zimmerman, 1978; Woo & Morrey, 1982). Moreover, patients with a family history of venous thromboembolism have significantly elevated risk of such complication after THA surgery (Mont et al., 2004; Markovic-Denic et al., 2012). This indicates the need for a patient risk stratification tool to effectively distinguish these high-risk patients from the relatively low-risk patients. During the post-discharge period, these high-risk patients identified by the risk stratification tool could be more closely followed and monitored to enable early detection of such surgical complications.

Relevant studies on readmission indicated that insufficient care quality is a significant contributing factor to unplanned readmission after surgery (James, 2013). Readmission after THA surgery is deemed a healthcare quality indicator suggesting incomplete management of relevant issues in patient care (Learmonth, Young, & Rorabeck, 2007). Strategies to improve care quality and reduce preventable readmission include providing better and safer care during the inpatient stay, paying more careful attention to patient medications, discharge planning and improved communication with patients and caregivers regarding follow-up care, as well as improving care transition through better communication and collaboration with other community providers (James, 2013). These strategies should be helpful in the prevention and early detection of post-operative THA surgical complications, so that more severe complications and readmissions can be effectively reduced. Moreover, a real-time surgical outcome evaluation framework is needed in this situation. With such a framework, poor patient outcomes can be
detected in real time, thus problem diagnosis and quality improvement efforts can be initiated in a timely manner.

1.3 Insufficient capacity of THA

With the aging of population and prevalence of obesity, the demand for THA in the United States is expected to increase from 326,100 in 2010 to 572,000 in 2030. In order to address such a massive increase in demand, a combination of increased economic resources, operative efficiency, technical capacity (i.e., additional surgeons), and implant longevity are needed (Kurtz et al., 2007).

Moreover, under the Medicare payment policy in the last decade, per capita payment for THA surgery was used to pay the hospitals. Billing for first assistant fees for THA was not allowed when the first assistant is an Accreditation Council for Graduate Medical Education (ACGME)-approved fellow. Thus, an unintended consequence of this Medicare payment policy is to discourage ACGME accreditation of fellowships in surgical specialties such as THA, where Medicare patients constitute a large percentage of the caseload. Therefore, in the last decade many orthopedic residents did not choose adult reconstruction fellowships. The workforce of adult reconstruction was predicted to be significantly inadequate to meet the demand for THA by 2020 (Iorio et al., 2008).

With a massively increasing demand and an insufficient supply for THA surgery, understanding the probable outcome that each patient can obtain from surgery becomes an important issue. We need to analyze the cost-effectiveness of THA surgery for different patient subgroups. By doing this, we can effectively determine the most appropriate treatment option for each patient to improve patient outcomes.
1.4 Pre-operative exercise intervention for THA patients

In the last two decades, several medical interventions and surgical interventions have been proposed for patients undergoing THA surgery. These interventions include pre-operative exercise program (Gilbey, Ackland, Tapper, & Wang, 2003; Crowe & Henderson, 2003; Rooks et al., 2006), minimal incision surgery protocol (Nuelle & Mann, 2007; Dorr, Maheshwari, Long, Wan, & Sirianni, 2007), integrated pain management (Ranawat & Ranawat, 2007), and enhanced recovery program after surgery (Larsen, Sorensen, Hansen, Thomsen, & Soballe, 2008; Malviya et al., 2011).

Exercise and physical therapy is a cornerstone of rehabilitation following THA surgery (Kuster, 2002). Relevant studies indicated that pre-operative fitness is predictive of post-operative outcomes of THA patients (Fortin et al., 1999). Moreover, patients scheduled for THA have to wait for at least one month before actually being operated on (Wallis & Taylor, 2011). Therefore, in the last two decades a pre-operative exercise program has been proposed for patients awaiting THA surgery. Most of the pre-operative exercise programs are adapted from post-operative exercise and physical therapy. These pre-operative exercise programs include some forms of strengthening, flexibility, and aerobic exercise, with the purpose of improving function and reducing pain at the pre-surgery period. The pre-operative exercise program also aims at positively influencing the post-operative patient outcomes. However, the studies on pre-operative exercise delivered mixed results regarding its effectiveness. While some studies indicated that pre-operative exercise program is effective in improving THA patient outcomes, others showed that this program has no benefit. Therefore, advanced statistical techniques are needed to combine these results and deliver a comprehensive perspective on the effectiveness of pre-operative exercise program. With such an analysis, we can have a better understanding of whether pre-operative exercise program is worthy of being carried out to improve THA care quality.
1.5 Healthcare quality, as well as data science and engineering

According to the healthcare quality framework of Donabedian, one can assess structural, process, and outcome quality (Donabedian, 1966). Structural quality refers to the use of metrics such as nurse-to-bed ratios. An example of a process-quality variable would be the percentage of surgical patients receiving antibiotics within a prescribed time period before surgery. Outcome-quality variables, on the other hand, reflect the patients’ results. For THA, an outcome variable would be whether or not a patient is readmitted to the hospital within 30 days of discharge. In addition, gained quality of life after THA surgery is a very important measure of patient outcome (Laupacis et al., 1993; Jones et al., 2001). Ko also stressed the importance of using outcome quality as the quality measure in healthcare (Ko, 2009). In general, the proper use of outcome variables requires considerably more effort in collecting patient-level data records, but it is the most informative approach to reflect healthcare quality (Woodall, Fogel, & Steiner, 2015).

In the most recent decade, there has been a great deal of interest in improving the quality of healthcare, with particular emphasis on surgical quality (Woodall et al., 2015). However, according to Peter Drucker, healthcare is the most difficult, chaotic, and complex industry to manage today (Drucker, 2012). Healthcare leaders are dealing with a multitude of quality improvement pressures. They need accurate, timely, and readily available information to make decisions. With the advancement of healthcare information technology, such as electronic health records and radiology information systems, healthcare data becomes exponentially available. It is now possible, however, that quality improvement efforts could be hindered by having so much data available, but without the necessary experience and tools to analyze it and put it to good use.

Data science and engineering has become increasingly used in healthcare to facilitate quality improvement initiatives and improve patient outcomes. This is an interdisciplinary field pertains to scientific methods, processes, and systems to extract knowledge or insights from data in various forms, either structured or unstructured. Data science and engineering is a combination and continuation of data analysis fields such as statistics, classification, clustering, machine learning, data mining, and predictive analytics (Dhar, 2013; Leek, 2013). According to Dan Mote, the president of National Academy of Engineering, the purpose of engineering is to create
solutions serving the needs of society and the welfare of humanity. In healthcare, data science and engineering should deliver data-driven solutions that help healthcare organizations gain insights into current performance and guide future actions, by discerning patterns and relationships in data and using that understanding to support decision making (Strome, 2013). Data science and engineering in healthcare exists for the purpose of improving the quality, safety, efficiency, and effectiveness of healthcare delivery. Now healthcare organizations are drawing from diverse, non-traditional professions to form quality improvement and innovation teams. In addition to nurses, physicians, and administrators, it is not uncommon to see engineers, computer scientists, and other specialist roles working within healthcare. This multi-disciplinary team brings incredible creativity, diversity, and flexibility. By using data-driven approaches to dig into the wealth of captured data, healthcare providers will be able to gain the critical knowledge that they need to answer many questions on quality improvement.

### 1.6 Research objective and framework

As mentioned in Section 1.2, the complications and unplanned readmissions that occur after THA surgery indicate suboptimal care quality, and thus quality improvement initiatives are needed for THA patient care. Moreover, with the rapid increase in THA demand over the next 15 years, it is crucial to effectively use the scarce medical resources in order to achieve the best possible patient outcomes. Since data science and engineering tools are effective methods in supporting informed medical decision making, they should be utilized to deliver data-driven solutions to facilitate improving THA care quality and patient outcomes.

Although data science and engineering tools have great potential in supporting and facilitating the quality improvement in THA care delivery, so far they have not been adequately used in improving THA patient outcomes. The objective of this doctoral dissertation is threefold: 1) to advance data-driven healthcare for THA, 2) to provide insights into informed medical decision making, and 3) to support and guide efforts in THA care quality improvement. The framework for achieving these objectives is shown in Figure 1-2 and can be viewed as follows.
Since 30-day unplanned readmission has become an increasingly serious problem and now hospitals bear the financial risk of CMS penalty, it is very important for hospitals to understand their THA surgical outcomes in real time. However, due to the inherent difference in patient risk profiles, the surgical outcomes should be risk adjusted. Therefore, we propose a framework for effectively conducting patient risk stratification and real-time surgical outcome evaluation. By doing this, high-risk patients are more likely to be accurately identified and better targeted. Furthermore, it becomes possible that individualized care plans can be implemented to enhance patient recovery and closely monitor high-risk patients at the post-discharge stage. Once deterioration in THA surgical outcome is detected, the medical team members can conduct root cause analysis and take measures to resolve the problem in a timely manner.

Figure 1-2. The framework of this dissertation research
Moreover, in the next 15 years there will probably be a shortage of available medical resources for THA surgery. Thus, it is very important to identify subgroups of hip arthritis patients who can benefit the most from THA surgery. In such a process, distinguishing patients using a single factor, such as age or gender, may not lead to the optimal partition of patient subgroups. Therefore, we aim at proposing a better methodology for clustering hip arthritis patients and analyzing the cost-effectiveness of THA surgery for each patient cluster. With our proposed methodology, THA medical resources can be more effectively used to improve patient outcomes.

Finally, when delivering medical or surgical interventions to THA patients, convincing evidence must show that a certain intervention really benefits the patients in terms of pain relief, physical function, activity of daily living, length of stay (LOS), etc. But the problem is, relevant studies may show mixed results on the effectiveness of a certain intervention. Therefore, we use meta-analysis, a statistical method that compares and contrasts the results of multiple studies on the same topic, to deliver a valid and comprehensive understanding of the effectiveness of an intervention. With such a validation process, interventions with proven benefits should be implemented, while those without proven benefits should not be carried out. By doing this, the effective interventions can be better identified and implemented to improve THA care quality and patient outcomes.

Achievement of these objectives may ultimately lead to considerable improvements in care quality for THA patients. When patient outcomes are improved and unplanned readmissions are reduced, medical expenditures for THA patients can be reduced accordingly. These contributions are very meaningful and crucial, because of the massively increasing demand for THA surgery and the quality of life that patients can obtain after the surgery.

1.7 Outline

The remainder of this dissertation is organized as follows. Chapter 2 is on developing an advanced risk-adjustment framework for conducting patient risk stratification and real-time surgical outcome evaluation. Chapter 3 is on clustering hip arthritis patients based on multiple
risk factors, and analyzing the cost-effectiveness of THA surgery for each patient cluster. Chapter 4 is on using meta-analysis to understand the comprehensive impact of pre-operative exercise on THA patient outcomes. Finally, Chapter 5 summarizes the main contributions of this dissertation and discusses possibilities for future work.
Chapter 2 Risk-Adjusted Control Charts for Monitoring THA Outcomes

2.1 Introduction

Statistical process control (SPC) is a method of quality control that uses statistical methods to monitor and control a process. Monitoring and controlling the process ensures that it operates at its full potential. At its full potential, the process can make as much conforming products as possible with a minimum (if not an elimination) of waste (i.e., rework or scrap). SPC can be applied to any process where the conforming product can be measured. Control charts and design of experiments are two of the primary approaches for SPC.

SPC has been widely and successfully used in the manufacturing settings to improve product quality. In manufacturing, quality is defined as conformance to specification. SPC uses statistical tools to observe the performance of the production process in order to detect significant variations before they result in a sub-standard product. Any source of variation at any point of time in a process will fall into one of two classes.

1) “Common Causes” - sometimes referred to as non-assignable, normal sources of variation. It refers to many sources of variation that consistently acts on process. These types of causes produce a stable and repeatable distribution over time.
2) “Special Causes” - sometimes referred to as assignable sources of variation. It refers to any factor causing variation that affects only some of the process output. They are often intermittent and unpredictable.

Most processes have many sources of variation, but most of them are minor and may be ignored. If the dominant sources of variation are identified, resources for change can be focused on them. If the dominant assignable sources of variation are detected, potentially they can be identified and removed. Once removed, the process is said to be “stable”. When a process is stable, its variation should remain within a known set of limits. That is, at least, until another assignable source of variation occurs (Neave & Wheeler, 1996).
SPC was first applied to the healthcare settings in 1971 (Riddick & Giddings, 1971). In healthcare, however, SPC is often risk adjusted, because patients with diverse demographic information and medical conditions have very different risks of a certain medical outcome (i.e., death or other adverse events). In this study, risk-adjusted SPC is used to monitor the 30-day unplanned readmission after THA surgery.

2.2 Literature review

2.2.1 Risk factors of readmission after THA

Since THA patients have diverse demographic information and medical conditions, the readmission risk after THA varies from patient to patient. As a result, a lot of studies have been conducted on identifying the risk factors of readmission after THA. These studies used different analytical tools, and each study came up with a set of significant risk factors. A summary of the significant risk factors identified in these studies is shown in Table 2-1. A “+” sign represents that the study in the corresponding column indicated that the risk factor in the corresponding row has a positive impact on readmission risk, whereas a “−” sign represents a negative impact on readmission risk.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Mahomed et al., 2003</th>
<th>Zhan, Kaczmarek, Loyo-Berrios, &amp; Bright, 2007</th>
<th>SooHoo et al., 2010</th>
<th>Huddleston, Wang, Uquila, Herndon, &amp; Maloney, 2012</th>
<th>Pugely, Callaghan, Martin, Cram, &amp; Gao, 2013</th>
<th>Zmistowski et al., 2013</th>
<th>Schairer et al., 2014</th>
<th>Mesko et al., 2014</th>
<th>Saucedo et al., 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black race</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Bleeding disorder</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiac disease</td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decreased distance between home and hospital</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dependent functional status</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Discharge to inpatient rehabilitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discharge to skilled nursing facility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>High ASA class</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Increased age</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Increased LOS</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Low income</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low-volume hospital</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Medical co-morbidity</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Obesity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steroid use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substance abuse</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Underweight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Unilateral replacement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>

The frequently-identified risk factors that increase readmission risk include male, black race, medical co-morbidity, obesity, diabetes, and increased LOS. Some of these studies obtained
different effects with respect to certain risk factors. Mahomed et al. (2003); Zhan et al. (2007); SooHoo et al. (2010); and Huddleston et al. (2012) all indicated that increased age has a positive effect on readmission risk, while Zmistowski et al. (2013) showed that relatively young patients are more likely to be readmitted after THA. Moreover, some studies investigated the impact of LOS on readmission risk after THA, after adjusting for the influence of other factors. Husted et al. (2010) studied whether fast-track hip and knee arthroplasty increases readmission rate. The results indicated that readmission rate does not increase as LOS decreases. Vorhies, Wang, Herndon, Maloney, & Huddleston (2011) identified that LOS after THA significantly reduced during 2002-2007, but there was no significant difference in readmission rate during this period. Cram et al. (2011) studied the THA surgeries conducted on Medicare patients during 1991-2008. However, this study indicated that during this period there was a decrease in LOS, but an increase in readmission rate. It is probable that the mixed results on certain risk factors can be attributed to the difference in patient data or the statistical methods used in these studies.

These studies provide insights into the risk factors relevant to readmission after THA. However, these studies are not comparable to some extent, because the patient data and the set of prospective risk factors used in these studies are different. Besides, in most of these studies, only one statistical method was used and the method was not tested on a new patient dataset. Therefore, we do not know the robustness of the results when the patient dataset changes.

2.2.2 Risk-adjusted SPC

In recent years, risk-adjusted SPC has been widely used to monitor infection rate (Gustafson, 2000; Morton et al., 2001), cardiac-surgery outcome (Steiner, Cook, Farewell, & Treasure, 2000; Steiner & Mackay, 2014), arterial-surgery outcome (Steiner, Cook, & Farewell, 2001; Beiles & Morton, 2004), coronary artery bypass surgery (Caputo, Reeves, Rogers, Ascione, & Angelini, 2004; Novick, Fox, Stitt, Forbes, & Steiner, 2006), gastroesophageal-surgery outcome (Collins, Jibawi, & McCulloch, 2011), organ-transplant outcome (Rogers, Ganesh, Banner, & Bonser, 2005; Collett, Sibanda, Pioli, Bradley, & Rudge, 2009), and so on. A summary of the studies on SPC in healthcare and the statistical methods used for risk adjustment is shown in Table 2-2.
Table 2.2. A summary of the studies on SPC in healthcare

<table>
<thead>
<tr>
<th>Study</th>
<th>Outcome being monitored</th>
<th>Method for risk adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gustafson, 2000</td>
<td>Infection rate</td>
<td>Standardized infection ratio</td>
</tr>
<tr>
<td>Morton et al., 2001</td>
<td>Infection rate</td>
<td>None</td>
</tr>
<tr>
<td>Steiner et al., 2000</td>
<td>Cardiac surgery</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Steiner &amp; Mackay, 2014</td>
<td>Cardiac surgery</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Steiner et al., 2001</td>
<td>Arterial surgery</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Beiles &amp; Morton, 2004</td>
<td>Arterial surgery</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Caputo et al., 2004</td>
<td>Coronary artery bypass surgery</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Novick et al., 2006</td>
<td>Coronary artery bypass surgery</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Collins et al., 2011</td>
<td>Stomach or esophagus surgery</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Rogers et al., 2005</td>
<td>Heart or lung transplant</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Collett et al., 2009</td>
<td>Kidney transplant</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Cook, Coory, &amp; Webster, 2011</td>
<td>Acute myocardial infarction</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Harris et al., 2005</td>
<td>Ruptured abdominal aortic aneurysms</td>
<td>V-POSSUM scoring system</td>
</tr>
</tbody>
</table>

These studies provide important insights into effectively monitoring surgical outcomes. However, there exist certain limitations in these studies. As is shown in Table 2-2, most of these studies use logistic regression for risk adjustment. Logistic regression has some apparent disadvantages in risk adjustment, because the functional forms of all the variables and their interactions have to be specified correctly in the model. This is difficult to do, especially when there are many variables and the modeler lacks domain knowledge of which variables are likely to be significant (Liu, Traskin, Lorch, George, & Small, 2015). Therefore, risk-adjustment frameworks based on logistic regression may not accurately evaluate the actual surgical performance of a medical team.

2.3 Research objective

In fiscal year 2015, hospitals with high readmission rates after THA began to be penalized by CMS. Thus, it is very important for medical teams to understand their performance on THA surgery. Moreover, reducing variation in surgical practice is very important, because high variation in surgical practice may lead to negative patient outcomes. In the healthcare settings, patient outcomes are regarded as the most informative and vital feedback on care quality and medical team performance (Porter & Teisberg, 2007; Department of Health, 2010). Therefore, it
is very important for the medical team to understand the surgical outcomes of THA patients and diagnose the potential problems in its medical practice. To the best of our knowledge, however, no study has been conducted on using risk-adjusted control charts to monitor a hospital’s THA surgery outcomes in real time.

The objectives of this study are to propose an effective framework for real-time monitoring of THA surgery outcomes and to provide a visualization tool for the medical team to understand its surgical outcome performance. First, we used advanced risk-adjustment techniques to accurately estimate the readmission risk of each patient, and identify the risk factors of 30-day unplanned readmission. Second, we investigated an effective way to monitor the medical team’s surgical outcome performance in real time. With such a real-time monitoring framework, the medical team becomes more informed of its performance on THA surgery. As a result, the team should be able to diagnose potential problems in a timely manner, given that inferior patient outcomes have been detected by the control chart. Therefore, fewer patients will be jeopardized from inferior care quality. The real-time monitoring framework can support informed medical decision making, and thus benefit both hospitals and THA patients.

2.4 Risk-adjustment methods

All the three risk-adjustment methods used in this study are machine learning algorithms. Machine learning is a subfield of computer science that was evolved from the study of pattern recognition and computational learning theory in artificial intelligence. In 1959, Arthur Samuel defined machine learning as a “field of study that gives computers the ability to learn without being explicitly programmed”. Machine learning involves the construction of algorithms that can learn from and make predictions on data. Such algorithms operate by building a model from example inputs in order to make data-driven predictions or decisions expressed as outputs, rather than following strictly static program instructions (Simon, 2013).

Machine learning tools are very powerful for exploring and understanding data. These tools can be classified as supervised and unsupervised. Broadly speaking, supervised machine learning involves building a statistical model for predicting or estimating an output based on one or more
inputs. With unsupervised machine learning, there are inputs but no supervising output. Nevertheless, we can learn relationship and structure from such data (James, Witten, Hastie, & Tibshirani, 2013). The frequently-used supervised machine learning algorithms include logistic regression, naïve Bayesian network, neural network, discriminant analysis, tree-based methods, and support vector machine. Tree-based methods include decision tree, bootstrap aggregation (bagging), random forest, and boosting.

The three machine learning algorithms used in our study are logistic regression, decision tree, and random forest. Logistic regression is a regression-based algorithm, and it is the most commonly-used risk adjustment method in the healthcare settings. Decision tree is the simplest tree-based machine learning algorithm, and the results produced by decision tree are very easy to interpret, especially for a person without any domain knowledge of machine learning. Random forest is a non-parametric ensemble of trees method that has been shown to have desirable performance in some complicated prediction problems.

### 2.4.1 Logistic regression

Binary logistic regression model quantifies the impact of covariates on a dichotomous response variable. The model is:

\[
p_t = \frac{1}{1+\exp(-[\beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_p x_{pt}])}
\]

where \( p_t \) is the probability of unplanned readmission of the \( t \)th patient; \( X_{1t}, X_{2t}, \ldots, X_{pt} \) are the values of the \( p \) covariates of the \( t \)th patient; \( \beta_1, \beta_2, \ldots, \beta_p \) are the coefficients for each of the \( p \) covariates, respectively.

Logistic regression is very easy to interpret and use. But it has serious limitations, because the functional forms of all the variables and their interactions have to be specified correctly in the model. However, in practice, the functional form of the covariates’ effects may be complex and unknown to the analyst, making it difficult to correctly specify these terms (Liu et al., 2015). One
solution to this general probability estimation problem is to treat it as a non-parametric regression problem, for which tree-based methods are applicable.

### 2.4.2 Decision tree

In a decision tree, the predictor space is segmented into smaller regions through a series of rules that define a split at each level of the tree. Recursive binary splitting is used to grow the tree with one of three common criteria: classification error rate, Gini index, or cross-entropy, which are defined below.

#### Classification error rate

\[ E = 1 - \max \hat{p}_{mk} \]  

where \( \hat{p}_{mk} \) is the proportion of observations in the \( m \)th group that are from the \( k \)th class (\( k=2 \) for binary classification).

#### Gini index

\[ G = \sum_{k=1}^{2} \hat{p}_{mk}(1 - \hat{p}_{mk}) \]  

#### Cross-entropy

\[ C = -\sum_{k=1}^{2} \hat{p}_{mk} \log(\hat{p}_{mk}) \]

At each step, all the predictors and possible values of cut points are considered, and then the optimal predictor and cut point are selected to minimize the evaluation criterion at this step. By doing so, the predictor space is gradually split into smaller regions. Finally, for all the observations that fall into the same region, the same prediction is made. The prediction is simply the mean of the response values of the observations in this region (James et al, 2013).

### 2.4.3 Random forest

Random forest is a tree-based algorithm developed by Leo Breiman. It is efficient and suitable for complicated non-linear classification and regression problems. Let the number of training cases be \( N \) and the number of variables be \( K \), then the algorithm is shown below (Breiman, 2001):
1) Choose a bootstrap sample of the training set (cases sampled with replacement from the
entire training dataset) and use the cases left out by bootstrapping (i.e., out-of-bag data)
to estimate the error of the tree.

2) A tree is grown by recursively splitting one node into two distinct subsets. $k < K$ variables
are used to determine the best split of a node which minimizes some impurity measures
in the training data. Common choices of impurity measures include Gini index for
classification and mean squared error for regression.

3) Each tree is fully grown and no pruning is performed. For regularization purposes, the
splitting may be stopped when reaching certain pre-set criterion (e.g., node size or tree
depth).

4) Repeat the above process to generate $ntree$ number of trees.

5) For prediction, the new sample is pushed down in each tree until it reaches the terminal
node. The final prediction for this new sample is the average value of these predictions by
each tree.

Random forest has been shown to have high accuracy in a variety of difficult regression and
classification problems.

### 2.5 Evaluation criteria of risk-adjustment methods

We used *area under the curve (AUC)* and *Brier score* as evaluation criteria for logistic
regression, decision tree, and random forest in risk adjustment.

Before introducing AUC, we need to know *sensitivity* and *specificity*. Table 2-3 shows an
example of the *confusion matrix* (Tucker et al., 2015). In a two-class classification problem, the
matrix is 3*3 with the leftmost 2*2 matrix referring to class predictions. The columns of this 2*2
matrix represent the actual readmission status (0 or 1), while the rows represent the predicted
readmission status. The column averages show how often an actual readmission status is
correctly predicted. The row averages show how often a predicted readmission status is correctly
assigned to an actual readmission status. The corner average shows the overall prediction
accuracy.
Table 2-3. An example of the confusion matrix

<table>
<thead>
<tr>
<th>Actual readmission status</th>
<th>0</th>
<th>1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted readmission status</td>
<td>607 (True negative, TN)</td>
<td>21 (False positive, FP)</td>
<td>628 (97%)</td>
</tr>
<tr>
<td>1</td>
<td>36 (False negative, FN)</td>
<td>10 (True positive, TP)</td>
<td>46 (22%)</td>
</tr>
<tr>
<td>Total</td>
<td>643 (94%)</td>
<td>31 (32%)</td>
<td>674 (92%)</td>
</tr>
</tbody>
</table>

The terms sensitivity and specificity characterize the performance of a classification method. Sensitivity, also called recall, is the true positives divided by the total number of elements predicted as positive, which is 22% in this example. Specificity is the true negatives divided by the total number of elements predicted as negative, which is 97% in this case. Equations (2-5) and (2-6) define sensitivity and specificity, respectively.

\[
\text{sensitivity} = \frac{TP}{TP+FN} \quad (2-5)
\]

\[
\text{specificity} = \frac{TN}{TN+FP} \quad (2-6)
\]

The receiver operating characteristic (ROC) curve is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting sensitivity against (1-specificity) at various threshold settings. The overall performance of a classification algorithm, summarized over all possible discrimination thresholds, is given by the AUC. An ideal predictor, or one that can perfectly separate the classes of a response variable, has an AUC of 1.0, whereas a classifier that is no better than random guessing would have an AUC around 0.5. An ideal ROC curve hugs the top left corner, so the larger the AUC the better the classifier. Normally, the AUC of an effective classifier should be greater than 0.7 (James et al., 2013).

The Brier score is used to evaluate the estimated probabilities of readmission. The Brier score is calculated as:

\[
\text{Brier score} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{P}(y_i = 1|x_i))^2 \quad (2-7)
\]
for a sample of \( n \) data points (Liu et al., 2015). Clearly, it measures the deviation of the predicted probability from the actual outcome.

### 2.6 Cumulative sum (CUSUM) control chart

The CUSUM chart is one of the most widely-used control charts in SPC. It has been shown to be ideally suited to detecting small persistent process changes (Montgomery, 2007). The unadjusted CUSUM statistic is:

\[
C_t = \max(0, C_{t-1} + W_t), \quad t = 1, 2, 3, \ldots
\]  

(2-8)

where \( C_0 = 0 \), and \( W_t \) is the sample weight or score assigned to the \( t \)th individual or subgroup. Through a judicious choice of \( W_t \), CUSUM can be designed to detect increases or decreases in the key parameter of interest, \( \theta \). The CUSUM statistic sequentially tests the null hypothesis \( H_0 : \theta = \theta_0 \) versus the alternative hypothesis \( H_A : \theta = \theta_A \). The value of \( \theta_0 \) is typically determined by the in-control process performance, whereas \( \theta_A \) represents the out-of-control performance that needs to be detected. The process is assumed to be in control as long as \( C_t < h \), and is deemed to have shifted to be out of control if \( C_t \geq h \) at any time \( t \). \( h \) is a pre-specified constant, called the control limit. In the healthcare settings, Equation (2-8) is often used to detect deterioration in the key parameter \( \theta \), whereas Equation (2-9) is used to detect improvement in \( \theta \).

\[
C_t' = \min(0, C_{t-1}' - W_t'), \quad t = 1, 2, 3, \ldots
\]  

(2-9)

Even if the process is consistently in control (i.e., \( H_0 \) is true), the CUSUM chart will eventually signal due to the inherent randomness in the process. This is called a *false alarm*. The in-control run length of CUSUM is defined as the time (or the number of observations) required before the CUSUM statistic \( C_t \) first exceeds the control limit \( h \), given that the process is in control. The expected or average in-control run length is denoted as \( ARL_0 \).

It has been shown that the optimal choice for the risk-adjusted CUSUM weight \( W_t \) is based on the log-likelihood ratio (Moustakides, 1986). The CUSUM weight statistic is:
\[ W_t = \begin{cases} \log \left( \frac{1-p_t+R_0p_t}{1-p_t+R_a p_t} \right) & \text{if } y_t = 0 \\ \log \left( \frac{(1-p_t+R_a p_t)R_a}{(1-p_t+R_a p_t)R_0} \right) & \text{if } y_t = 1 \end{cases} \] (2-10)

where \( y_t \) is the actual outcome of the \( t \)th patient, with \( y_t = 1 \) indicating that the \( t \)th patient has an adverse event and \( y_t = 0 \) otherwise. \( p_t \) is the probability of adverse event for the \( t \)th patient estimated by risk-adjustment models, such as logistic regression, decision tree, and random forest. \( R_0 \) and \( R_a \) represent the odds ratios under the null hypothesis \( H_0 \) and the alternative hypothesis \( H_A \), respectively. \( R_0 \) is typically set to 1.0 and \( R_a \) is based on the abnormality level to be detected. This risk-adjusted CUSUM chart is the most widely-accepted surveillance approach for monitoring risk-adjusted binary surgical outcomes (Woodall et al., 2015).

### 2.7 Study setting

Under Institutional Review Board (IRB) approval, we obtained identified THA patient-level data from the electronic health records of an academic medical center located in central Pennsylvania, United States. From January, 2011 to August, 2015, there were 1347 cases of THA conducted at this medical center. The data for these 1347 observations address over 60 attributes including but not limited to admission date, age, gender, ethnicity, zip code, LOS, allergy, diagnosis, discharge disposition, diagnosis-related group (DRG), pain assessment, and procedure. In total, 706 unique ICD-9 codes are used to represent the documented disease of these THA patients. We collapsed subtypes of the same disease, such as various subtypes of diabetes, into one disease category. To account for common diseases across these patients, we recoded each type of disease as a binary variable.

Among these 1347 records, 77 cases of unplanned readmission occurred within 30 days of discharge. Figure 2-1 shows the 30-day unplanned readmission rates for each year of 2011-2015.
The annual 30-day unplanned readmission rates show an increasing trend in these years, increasing from 2.43% in 2011 to 10.18% in 2015. While it is possible that the increase is partially attributable to a more diligent accounting for readmission within the organization, it is also clear that the outcome is worse. With the CMS penalty program in place for THA in fiscal year 2015, there is additional motivation to investigate the causes of this increasing unplanned readmission rate. Comparisons of patients without unplanned readmission and those with unplanned readmission are shown in Table 2-4 and Table 2-5. $P$-value is the significance in testing the difference in the mean or percentage of the two groups of patients.

Table 2-4. Comparison of patients on continuous variables

<table>
<thead>
<tr>
<th>Explanatory variables</th>
<th>Without unplanned readmission</th>
<th>With unplanned readmission</th>
<th>Statistical results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Standard deviation</td>
<td>Range</td>
</tr>
<tr>
<td>Age (year)</td>
<td>63.85</td>
<td>15.54</td>
<td>38-103</td>
</tr>
<tr>
<td>LOS (day)</td>
<td>3.50</td>
<td>3.13</td>
<td>1-20</td>
</tr>
</tbody>
</table>
Table 2-5. Comparison of patients on categorical variables

<table>
<thead>
<tr>
<th>Explanatory variables</th>
<th>Without unplanned readmission</th>
<th>With unplanned readmission</th>
<th>Statistical results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>43.78</td>
<td>49.35</td>
<td>0.35</td>
</tr>
<tr>
<td>Discharge to rehabilitation facility</td>
<td>27.64</td>
<td>62.34</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Discharge to home health</td>
<td>3.15</td>
<td>3.90</td>
<td>0.74</td>
</tr>
<tr>
<td>Body mass index (BMI) over 40</td>
<td>8.90</td>
<td>18.18</td>
<td>0.04</td>
</tr>
<tr>
<td>Cardiac disease</td>
<td>9.61</td>
<td>9.09</td>
<td>0.88</td>
</tr>
<tr>
<td>Coronary atherosclerosis</td>
<td>13.94</td>
<td>16.88</td>
<td>0.51</td>
</tr>
<tr>
<td>Diabetes</td>
<td>16.69</td>
<td>22.08</td>
<td>0.27</td>
</tr>
<tr>
<td>Heart failure</td>
<td>5.91</td>
<td>14.29</td>
<td>0.04</td>
</tr>
<tr>
<td>Hypertension</td>
<td>54.88</td>
<td>67.53</td>
<td>0.02</td>
</tr>
<tr>
<td>Kidney disease</td>
<td>6.54</td>
<td>18.18</td>
<td>0.01</td>
</tr>
<tr>
<td>Liver disease</td>
<td>2.91</td>
<td>3.90</td>
<td>0.67</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>5.51</td>
<td>7.79</td>
<td>0.47</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>1.73</td>
<td>3.90</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Table 2-4 and Table 2-5 indicate that at the 0.05 significance level, readmitted patients are relatively older, have a longer LOS, and have a higher percentage of being discharged to rehabilitation facility or skilled nursing facility (SNF). With respect to co-morbidities, a higher percentage of readmitted patients have BMI over 40, heart failure, hypertension, and kidney disease. These risk factors are similar to those identified in the relevant studies (Mahomed et al., 2003; Zhan et al., 2007; SooHoo et al., 2010; Huddleston et al., 2012; Pugely et al., 2013; Zmistowski et al., 2013; Mesko et al., 2014; Saucedo et al., 2014; Schairer et al., 2014).

2.8 Results

2.8.1 Risk adjustment by the three machine learning algorithms

We used logistic regression, decision tree, and random forest to estimate the 30-day unplanned readmission risk of the 1347 THA patients. Given the large number of available variables, we first experimented with dimension reduction techniques in order to find the right combination of predictors that could accurately represent the key relationships between patient characteristics
and the 30-day unplanned readmission risk. Among all the disease types, only 10 of them occur with a frequency greater than or equal to 27 (i.e., 2%) among all the THA patients. Thus, we excluded all the remaining diseases that occur with a frequency less than 27 to reduce the dimension of the patient data. The final dataset includes 1347 observations and 16 explanatory variables. These explanatory variables are age, gender, ethnicity, LOS, BMI over 40, coronary atherosclerosis, diabetes, heart failure, cardiac disease, myocardial infarction, hypertension, pneumonia, liver disease, kidney disease, discharge to rehabilitation facility or SNF, and discharge to home health. The response variable is the 30-day unplanned readmission status of each patient (1 or 0). We also referred to how CMS estimates the expected readmissions for each hospital. CMS uses 33 explanatory variables for THA patients. Only two of these variables are demographic factors (i.e., age and gender), whereas all the rest are variables on disease or co-morbid conditions (CMS, 2016d).

In order to compare the effectiveness of logistic regression, decision tree, and random forest in risk adjustment, we separated the THA patient dataset into a training dataset and a test dataset. We fitted these risk-adjustment methods on the training set, and then compared the results on both the training set and the test set. As in regression modeling, creating training validation and test data is an important step to prevent over-fitting. Moreover, we varied the size of training dataset from 50% to 90% of the whole patient dataset (with an increment of 10%), and we used the remaining data in each partition as the test set. The whole dataset size is always 1347 in each partition. To achieve more robust results, 10 different partitions were conducted at each percentage level and the results were averaged. Among the 77 unplanned readmissions, there is a patient who was readmitted twice. To our knowledge, there is no well-accepted approach for dealing with multiple entries of the same patient. We separated these two entries, with one entry in the training dataset and the other entry in the test dataset. The rationale is that if both of them are in the training set or test set, they will bias the results more seriously.

The analyses were conducted in the statistical package R 3.1.0 (R Development Core Team, 2008). Table 2-6 shows the average AUC of logistic regression, decision tree, and random forest with different training dataset sizes. Table 2-7 shows the average Brier score of these three risk-adjustment methods. The numbers in the parenthesis are standard deviation.
Table 2-6. AUC of the three risk-adjustment methods with different training dataset sizes

<table>
<thead>
<tr>
<th>Training dataset size (%)</th>
<th>Logistic regression</th>
<th>Decision tree</th>
<th>Random forest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training set</td>
<td>Test set</td>
<td>Training set</td>
</tr>
<tr>
<td>50</td>
<td>0.8654</td>
<td>0.7701</td>
<td>0.7899</td>
</tr>
<tr>
<td></td>
<td>(0.0530)</td>
<td>(0.0625)</td>
<td>(0.0647)</td>
</tr>
<tr>
<td>60</td>
<td>0.8745</td>
<td>0.7835</td>
<td>0.8308</td>
</tr>
<tr>
<td></td>
<td>(0.0509)</td>
<td>(0.0518)</td>
<td>(0.0552)</td>
</tr>
<tr>
<td>70</td>
<td>0.8649</td>
<td>0.8173</td>
<td>0.8277</td>
</tr>
<tr>
<td></td>
<td>(0.0447)</td>
<td>(0.0506)</td>
<td>(0.0492)</td>
</tr>
<tr>
<td>80</td>
<td>0.8707</td>
<td>0.7950</td>
<td>0.8117</td>
</tr>
<tr>
<td></td>
<td>(0.0430)</td>
<td>(0.0544)</td>
<td>(0.0483)</td>
</tr>
<tr>
<td>90</td>
<td>0.8632</td>
<td>0.7927</td>
<td>0.8073</td>
</tr>
<tr>
<td></td>
<td>(0.0445)</td>
<td>(0.0485)</td>
<td>(0.0460)</td>
</tr>
</tbody>
</table>

Table 2-7. Brier score of the three risk-adjustment methods with different training dataset sizes

<table>
<thead>
<tr>
<th>Training dataset size (%)</th>
<th>Logistic regression</th>
<th>Decision tree</th>
<th>Random forest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training set</td>
<td>Test set</td>
<td>Training set</td>
</tr>
<tr>
<td>50</td>
<td>0.0490</td>
<td>0.0538</td>
<td>0.0472</td>
</tr>
<tr>
<td></td>
<td>(0.0028)</td>
<td>(0.0026)</td>
<td>(0.0054)</td>
</tr>
<tr>
<td>60</td>
<td>0.0497</td>
<td>0.0526</td>
<td>0.0468</td>
</tr>
<tr>
<td></td>
<td>(0.0035)</td>
<td>(0.0029)</td>
<td>(0.0040)</td>
</tr>
<tr>
<td>70</td>
<td>0.0506</td>
<td>0.0501</td>
<td>0.0468</td>
</tr>
<tr>
<td></td>
<td>(0.0031)</td>
<td>(0.0032)</td>
<td>(0.0039)</td>
</tr>
<tr>
<td>80</td>
<td>0.0502</td>
<td>0.0491</td>
<td>0.0487</td>
</tr>
<tr>
<td></td>
<td>(0.0011)</td>
<td>(0.0028)</td>
<td>(0.0015)</td>
</tr>
<tr>
<td>90</td>
<td>0.0499</td>
<td>0.0502</td>
<td>0.0488</td>
</tr>
<tr>
<td></td>
<td>(0.0014)</td>
<td>(0.0031)</td>
<td>(0.0017)</td>
</tr>
</tbody>
</table>

Table 2-6 and Table 2-7 indicate that random forest outperforms logistic regression and decision tree with respect to both AUC and the Brier score. The only exception is that decision tree outperforms random forest with respect to AUC in the last scenario. In addition, the superiority of random forest over the other two methods is more significant when the training dataset size is relatively small (i.e., a lower percentage of the whole patient dataset). This indicates that the performances of logistic regression and decision tree are more sensitive to training dataset size, while the performance of random forest is more robust to training sample size. A relevant study also indicated that logistic regression is sensitive to training sample size
Moreover, the standard deviation of random forest is smaller than those of the other two methods.

Since logistic regression is the most commonly-used risk-adjustment method and random forest performs the best in this study, we compared the difference in readmission probabilities estimated by these two methods. These two methods both achieved their highest test-set AUC when using 70% of the whole patient dataset as the training set, thus we used the readmission probabilities estimated in this scenario. Figure 2-2 shows the readmission risks of all 1347 THA patients estimated by logistic regression. The horizontal axis is the time sequence of patient discharge date from the medical center. The vertical axis is the readmission risk, or probability of readmission. The readmission risks of all the patients estimated by random forest are shown in Figure 2-3.

![Figure 2-2. Readmission risks of all the patients estimated by logistic regression](image)

![Figure 2-3. Readmission risks of all the patients estimated by random forest](image)
Comparing Figure 2-2 with Figure 2-3, it is apparent that random forest more significantly distinguishes the high-risk patients from the relatively low-risk patients. The high-risk patients have readmission probabilities greater than 0.4, while low-risk patients’ readmission probabilities are less than 0.2. We reviewed the actual readmission status of all the patients with estimated probabilities greater than 0.4 by random forest. Among the 43 patients in Figure 2-3 with high risk for readmission, 42 of them were actually readmitted within 30 days of discharge. The confusion matrix with probability 0.4 as the threshold by random forest is shown in Table 2-8.

<table>
<thead>
<tr>
<th>Actual readmission status</th>
<th>0</th>
<th>1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted readmission status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1269 (True negative, TN)</td>
<td>35 (False positive, FP)</td>
<td>1304 (97%)</td>
</tr>
<tr>
<td>1</td>
<td>1 (False negative, FN)</td>
<td>42 (True positive, TP)</td>
<td>43 (98%)</td>
</tr>
<tr>
<td>Total</td>
<td>1270 (100%)</td>
<td>77 (55%)</td>
<td>1347 (92%)</td>
</tr>
</tbody>
</table>

In contrast, there is no such significant difference in the estimation by logistic regression, as is shown in Figure 2-2. In logistic regression, only 2 patients are estimated with readmission probabilities greater than 0.4. The high-risk patients do not significantly “stand out” from the relatively low-risk patients in logistic regression.

### 2.8.2 Risk factors identified by the three risk-adjustment methods

In addition to accurately distinguishing high-risk patients from the relatively low-risk patients, it is important to identify the risk factors of 30-day readmission. By doing this, the medical team can effectively take advantage of the information in electronic health records to identify high-risk patients. Table 2-9 shows the top-10 high-frequency predictors of readmission risk identified by each of the three risk-adjustment methods. High frequency indicates that the variable is present in many of the risk-adjustment models fitted with different training datasets.
Table 2-9. The top-10 high-frequency predictors of readmission risk

<table>
<thead>
<tr>
<th>Rank</th>
<th>Logistic regression</th>
<th>Decision tree</th>
<th>Random forest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BMI over 40</td>
<td>Age</td>
<td>Age</td>
</tr>
<tr>
<td>2</td>
<td>Discharge to rehabilitation facility or SNF</td>
<td>LOS</td>
<td>BMI over 40</td>
</tr>
<tr>
<td>3</td>
<td>Kidney disease</td>
<td>Discharge to rehabilitation facility or SNF</td>
<td>LOS</td>
</tr>
<tr>
<td>4</td>
<td>Discharge to home health</td>
<td>Male</td>
<td>Hypertension</td>
</tr>
<tr>
<td>5</td>
<td>Age</td>
<td>BMI over 40</td>
<td>Male</td>
</tr>
<tr>
<td>6</td>
<td>Hypertension</td>
<td>Kidney disease</td>
<td>Diabetes</td>
</tr>
<tr>
<td>7</td>
<td>Male</td>
<td>Hypertension</td>
<td>Discharge to rehabilitation facility or SNF</td>
</tr>
<tr>
<td>8</td>
<td>Coronary atherosclerosis</td>
<td>Diabetes</td>
<td>Kidney disease</td>
</tr>
<tr>
<td>9</td>
<td>Black ethnicity</td>
<td>Myocardial infarction</td>
<td>Discharge to home health</td>
</tr>
<tr>
<td>10</td>
<td>Heart failure</td>
<td>Coronary atherosclerosis</td>
<td>Heart failure</td>
</tr>
</tbody>
</table>

Table 2-9 indicates that the three risk-adjustment methods identified a set of common risk factors, including BMI over 40, discharge to rehabilitation facility or SNF, kidney disease, age, hypertension, and male. All these factors are positively correlated with 30-day readmission risk. Since there is high consistency in the risk factors determined by the three risk-adjustment methods, these factors are probably important in patient risk stratification.

2.8.3 CUSUM charts based on risk-adjustment

We integrated the risk-adjustment results in Figure 2-2 and Figure 2-3 into the CUSUM control chart. We used these readmission probabilities in Equation (2-10) to calculate the risk-adjusted CUSUM weight $W_t$. The odds ratios $R_d$ were set to 2.0 for detecting deterioration and 0.5 for detecting improvement in THA surgical outcomes. The control limits $h$ were determined to have an $ARL_0$ of 4000 observations. The rationale is that since there are 1347 patient-level data records, the CUSUM chart with an $ARL_0$ of 4000 would have a low probability of false alarm. We built Monte-Carlo simulation models to determine the control limits with such an $ARL_0$.

For logistic regression, the upper control limit is 4.58 and the lower control limit is -4.24. Then we conducted a SPC analysis of the THA surgical outcomes at the medical center, using
the 30-day unplanned readmission as the adverse event. The CUSUM chart with risk-adjustment from logistic regression is shown in Figure 2-4. The first control chart is used for detecting deterioration in surgical outcomes and the second control chart is used for detecting improvement in surgical outcomes. The black lines are the control statistics and the red lines are the control limits.

Figure 2-4. CUSUM chart (top for deterioration and bottom for improvement) by logistic regression

Similarly, the upper control limit and the lower control limit of random forest were determined with the same $ARL_0$ by Monte-Carlo simulation. They are 4.19 and -3.94, respectively. The CUSUM chart with risk-adjustment results from random forest is shown in Figure 2-5.
The differences in risk adjustment by logistic regression and random forest lead to the differences in CUSUM charts. For the CUSUM chart based on logistic regression, there exists a signal for improvement in THA surgical outcomes at the 638\textsuperscript{th} patient, as is indicated in Figure 2-4. But no such signal exists in the CUSUM chart based on random forest. Both CUSUM charts signal for deterioration, but at different times. The chart based on logistic regression signals for deterioration at the 1044\textsuperscript{th} patient, whereas the counterpart based on random forest signals at the 1283\textsuperscript{th} patient.

In early years, THA outcomes at the medical center are good, thus the CUSUM chart based on logistic regression signals for improvement at the 638\textsuperscript{th} patient. The CUSUM chart based on random forest also indicates a trend of improvement around the 638\textsuperscript{th} patient, but the trend is not statistically significant for a signal. However, after the 950\textsuperscript{th} patient, the THA surgical outcomes gradually deteriorate and more unplanned readmissions occur. Thus, the control statistic gradually moves towards the upper control limit and finally exceeds the limit. Therefore, we are able to state that the increase in annual readmission rate shown in Figure 2-1 can hardly be explained by chance, because it reaches the statistical significance level and results in a deterioration signal.
In this study, the CUSUM chart based on logistic regression signals changes in surgical outcomes, especially in the chart for detecting improvement. The CUSUM chart based on random forest also indicates an improvement followed by deterioration in surgical outcomes, but the signal is weaker. One possible explanation is that the patient risk profile estimated by logistic regression is more homogeneous than that estimated by random forest, as are shown in Figure 2-2 and Figure 2-3. If a patient is readmitted within 30 days of discharge, the CUSUM weight $W_t$ based on logistic regression is penalized more heavily using Equation (2-10). In contrast, random forest significantly distinguishes high-risk patients from relatively low-risk patients. Thus, if a high-risk patient is readmitted, it does not penalize $W_t$ as heavily as logistic regression does. This may explain why the CUSUM chart based on random forest signals for deterioration later than the counterpart based on logistic regression.

Similarly, it can be seen from Figure 2-2 and Figure 2-3 that the readmission risks of low-risk patients estimated by random forest are even lower than those estimated by logistic regression. Thus, using Equation (2-10), the CUSUM weight $W_t$ based on random forest is less “rewarded” if a low-risk patient is not readmitted. This may explain why the CUSUM chart based on random forest does not signal for improvement at the 638th patient, while the counterpart based on logistic regression signals.

Since both control charts signal for deterioration after the 1000th patient, there is strong evidence for developing and implementing out-of-control action plans as a part of the broader quality improvement strategy. Such an approach would help identify potential problems in patient care in order to effectively reduce unplanned readmission after THA.

2.9 Discussion

Risk adjustment is important for correctly comparing and evaluating medical or surgical outcomes, because the risk profile of patients treated by different medical teams are different. We do not want to unfairly criticize or penalize a medical team simply because the team treats sicker patients. In this study, we obtained 1,347 THA patient-level data records from an academic medical center in central Pennsylvania, United States. We compared logistic
regression, decision tree, and random forest in estimating the 30-day unplanned readmission risk of THA patients. The results indicate that random forest outperforms the other two methods, especially when the training dataset size is relatively small. After that, we integrated the risk-adjustment results of logistic regression and random forest into the CUSUM chart, respectively. Since random forest more accurately accounts for the risk profile of different patients, the CUSUM chart based on random forest provides more reliable and convincing real-time monitoring results than the counterpart based on logistic regression. Thus, integrating random forest algorithm into control chart provides a more accurate and valuable framework for monitoring and evaluating the surgical outcome performance of a medical team in real time.

In an acute care setting, it is crucial to correctly identify high-risk patients. This is the first step towards providing patient-centered care as well as targeting medical and surgical interventions. Our study shows that random forest has a higher accuracy in identifying high-risk patients than logistic regression and decision tree. Acute care facilities could benefit from using the proposed random forest model. By doing this, the medical team could accurately identify THA patients with high readmission risk and thus make informed medical decisions on treating these patients. For instance, medical staff at a hospital and a rehabilitation facility may work together to ensure that high-risk patients receive appropriate care transitions. In addition, more intensive post-discharge monitoring, such as frequent follow-ups and using tele-health devices, can be carried out for high-risk patients. Such enhanced care plan can probably improve care quality and reduce unplanned readmission.

Moreover, combining machine learning algorithms with control charts provide new opportunities to apply both methods in medical decision support. Based on the proposed framework, the medical team may be able to identify care quality problems in a timely manner, as the risk-adjusted outcomes can be visualized from the control chart. Moreover, effective monitoring of surgical performance helps reduce variation in medical practice, which is an important intervention to improve patient outcomes. Therefore, our proposed framework could be beneficial for improving care quality, reducing unplanned readmission, and avoiding CMS penalty.
This study has several limitations. Our patient-level data records did not include factors such as post-operative care quality and patient adherence after discharge. These factors, however, might influence unplanned readmission risk of THA patients. Moreover, this study was conducted utilizing data from one academic medical center in central Pennsylvania, United States, and may not be directly generalizable to other hospitals or the whole healthcare system.

2.10 Conclusion

Unplanned readmission after THA has become an increasingly serious problem in the United States, because of its negative effects on both patients and hospitals. Accurately identifying patients with high readmission risk is the first step towards targeting medical intervention and providing patient-centered care. Moreover, it is important to monitor and evaluate THA surgical outcomes in real time, so that the medical team can obtain a better understanding of its actual performance while taking into account the inherent difference in patient risk profile.

Our proposed framework, which combines machine learning with SPC, provides an effective mechanism for stratifying patient risks and evaluating surgical outcomes based on 30-day readmission. These are important steps in improving THA patient outcomes overall. In this study, this framework outperforms the most commonly-used risk-adjustment framework, which is based on logistic regression. We suggest that there are numerous opportunities for medical teams to make informed decisions on treating THA patients. Furthermore, with risk stratification, patients who are interested in improving their surgical outcomes can better understand their own risks, so that they may actively learn how to prevent common complications after THA.

Finally, our proposed framework provides the possibility of effectively utilizing the patient data in the electronic health records to make informed medical decisions and positively influence patient outcomes. It becomes increasingly possible that the readmission problem after THA can be effectively tackled and the cost of treating readmitted patients can be significantly reduced.
3.1 Introduction

Health policy influences the way that nations, states, and other population groups distribute resources among different competing interventions and populations, based on anticipated benefits. Since there is insufficient public health funding to implement all recommended healthcare interventions, an important question is how to best allocate scarce resources across these interventions. There is motivation, therefore, to select the best set of interventions and implement them on the targeted population in order to maximize the benefits of the available health resources.

Cost-effectiveness analysis is a form of economic analysis that compares the relative costs and outcomes (i.e., effects) of different courses of resources allocation. Typically, cost-effectiveness is expressed in terms of a ratio, where the denominator is a gain in health from an outcome measure called quality-adjusted life year (QALY) and the numerator is the cost associated with such health gain (Gold, Siegel, Russell, & Weinstein, 1996). QALY takes into account morbidity and mortality, and it calibrates people’s life expectancy based on their quality of life (Weinstein, Torrance, & McGuire, 2009). A relatively small cost per QALY indicates that a healthcare intervention is desirable. Healthcare interventions can be listed in ascending order of cost-effectiveness ratio, and interventions are selected starting from the top of the list until the available funding is depleted.

In some other situations, thresholds are used to determine whether an intervention provides acceptable values, given the costs involved (Torrance et al., 1996). In the United Kingdom, a cost-effectiveness threshold of £20,000-30,000 (approximately US $30,000-50,000) per QALY is typically used (Devlin & Parkin, 2004; McCabe, Claxton, & Culyer, 2008), while in the United States the threshold is typically US $50,000 (Owens, Qaseem, Chou, & Shekelle, 2011). Cost-effectiveness analysis has been widely used, because it integrates measurements of change in health pre- and post-intervention, life expectancy and quality, as well as the cost of intervention (Lubowitz & Appleby, 2011). The integration of these elements has been found to
provide an appropriate baseline for comparison in the resource allocation process in healthcare (Weinstein, Siegel, Gold, Kamlet, & Russell, 1996).

THA surgery is very effective for patients with serious hip arthritis or hip injury, and it has been shown to significantly improve patients’ quality of life with respect to physical function, pain relief, and overall health (Laupacis et al., 1993; Jones et al., 2001). However, this does not necessarily indicate that THA is a cost-effective treatment option for each THA patient, because THA patients have diverse demographic, socio-economic, and co-morbid conditions. Moreover, the demand for THA in the United States is expected to increase from 326,100 in 2010 to 572,000 in 2030. The capacity of orthopedic surgeon, however, may not be sufficient to address such a massively increasing demand (Kurtz et al., 2007; Iorio et al., 2008). We are, therefore, motivated to consider how to maximize patient benefits in terms of gained QALY, with the purpose of informing the broader efforts to improve patient outcomes under the constraint of limited medical resources.

3.2 Literature review

3.2.1 Studies on the cost-effectiveness of THA

Our review of the literature shows eight studies on the cost-effectiveness of THA surgery. A summary of these studies is shown in Table 3-1.
Table 3-1. A summary of studies on cost-effectiveness of THA

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample size</th>
<th>Year of enrollment</th>
<th>Study period</th>
<th>Tools used to measure quality of life</th>
<th>Results</th>
<th>Limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rissanen et al., 1997</td>
<td>276</td>
<td>1991-1992</td>
<td>From admission to 24 months post-operatively</td>
<td>15D</td>
<td>The total cost for each patient was FIM 45,000 (US $10,500) during the two years. THA is more cost-effective for patients younger than 70. But THA is cost-effective for all age groups</td>
<td>Only tracked the costs and QALY within 24 months post-operatively. No benchmark treatment</td>
</tr>
<tr>
<td>Räsänen et al., 2007</td>
<td>96</td>
<td>2002</td>
<td>From admission to 12 months post-operatively</td>
<td>15D</td>
<td>Patients with a favorable result in utility score were significantly younger than those with a less optimal result. The average cost per QALY gained was FIM 6,710</td>
<td>Only tracked the costs within 12 months post-operatively</td>
</tr>
<tr>
<td>Jenkins et al., 2013</td>
<td>348</td>
<td>2010</td>
<td>From admission to 12 months post-operatively</td>
<td>EQ-5D</td>
<td>The mean procedure cost was UK £8956. If future QALY was discounted at 3.5%, cost per QALY was £2852</td>
<td>No medical cost after THA surgery was considered. No benchmark treatment</td>
</tr>
<tr>
<td>Boettcher, 1992</td>
<td>42</td>
<td>1980-1986</td>
<td>From admission to 5 years post-operatively</td>
<td>ASA classification</td>
<td>THA surgery was more cost effective than maintaining patients in a nursing home. But no QALY was evaluated</td>
<td>No QALY was considered. Only compared THA surgery with maintaining patients in nursing home</td>
</tr>
<tr>
<td>Liang et al., 1986</td>
<td>23</td>
<td>1982-1983</td>
<td>From admission to 6 months post-operatively</td>
<td>Bush Index of Well Being (BUIWB)</td>
<td>The cost-effectiveness of THA was US $2,703/0.01 unit of BUIWB (in the study it was represented as 0.0037 BUIWB/$1,000). It was more cost-effective for patients who initially had the poorest health</td>
<td>Only tracked the costs and QALY within 6 months post-operatively. Only compared with pre-surgery situation</td>
</tr>
<tr>
<td>Study</td>
<td>Sample size</td>
<td>Year of enrollment</td>
<td>Study period</td>
<td>Tools used to measure quality of life</td>
<td>Results</td>
<td>Limitation</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------------</td>
<td>--------------------</td>
<td>--------------</td>
<td>---------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Chang, Pellissier, &amp; Hazen, 1996</td>
<td>Stochastic modeling study, no actual patient enrollment</td>
<td>N/A</td>
<td>From admission to death</td>
<td>American College of Rheumatology Functional Classification, and a risk utility assessment tool</td>
<td>For 60-year-old women, the average lifetime cost of THA was $117,000 less than the non-surgery treatment, and the additional QALY gain was 6.9. For a 85-year-old men, THA was $9,100 more than the non-surgery treatment and the additional QALY gain was 2.0</td>
<td>Only analyzed the cost-effectiveness for very specific patient types, based on gender, age and ethnicity</td>
</tr>
<tr>
<td>Rorabeck et al., 1994</td>
<td>164</td>
<td>1987-1992</td>
<td>From admission to 24 months post-operatively</td>
<td>Time trade-off utility</td>
<td>THA costed less than CAD $20,000 per additional QALY. Cement THA and cement-less THA had very similar cost-effectiveness</td>
<td>Only tracked the costs and QALY within 24 months post-operatively. Did not identify patient subgroups. No benchmark treatment</td>
</tr>
<tr>
<td>Pennington et al., 2013</td>
<td>Markov modeling study, no actual patient enrollment</td>
<td>N/A</td>
<td>From admission to death</td>
<td>EQ-5D</td>
<td>Compared cemented, cement-less, and hybrid prostheses. The results indicated that hybrid prostheses have the highest probability of being the most cost-effective in all subgroups, except in women aged 80</td>
<td>No medical cost other than primary and revision surgery was considered. Only analyzed the cost-effectiveness for very specific patient subgroups based on age. No comparison with non-surgery treatment</td>
</tr>
</tbody>
</table>

Table 3-1 indicates that most of studies on the cost-effectiveness of THA were conducted by using questionnaires to survey THA patients before surgery and at the post-operative follow-ups. While these studies offered some insights into the cost-effectiveness of THA surgery, there are some important limitations. First, none of these studies differentiated patients, except two studies
that used demographic factors as the differentiating factors. This is not desirable because THA may have different cost-effectiveness for different patient subgroups, and factors other than demographics, such as co-morbid conditions, may influence the cost-effectiveness ratio as well. Second, most of these studies only considered the cost-effectiveness ratio of THA surgery, and did not compare it with any alternative treatments, e.g., non-surgery treatment of hip arthritis. In an environment where medical resources are limited, as we mentioned above, it is important to compare this medical or surgical intervention with at least one benchmark/alternative treatment. Third, most of these studies had a follow-up period of two years or less. Thus, only the cost and quality of life during the immediate period following THA surgery were considered. However, for surgeries like THA, which are intended to have long-term benefits, the relevant medical cost and quality of life for a much longer period of time, potentially throughout a patient’s lifetime, should be taken into account and appropriate methodology needs to be used to account for this.

3.2.2 Studies on cluster analysis in healthcare

Cluster analysis has been widely used in the healthcare settings to separate patients into different subgroups and then analyze the specific characteristics of each subgroup. A considerable number of studies have been conducted in this area. A summary of the studies most relevant to our work is shown in Table 3-2.

<table>
<thead>
<tr>
<th>Study</th>
<th>Type of patient</th>
<th>Clustering method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haldar et al., 2008</td>
<td>Asthma</td>
<td>K-means</td>
</tr>
<tr>
<td>Michel et al., 2000</td>
<td>Hip fracture</td>
<td>Hierarchical</td>
</tr>
<tr>
<td>Eslick, Howell, Hammer, &amp; Talley, 2004</td>
<td>Gastrointestinal disorder</td>
<td>K-means</td>
</tr>
<tr>
<td>Cox, Enns, &amp; Larsen, 2001</td>
<td>Depression</td>
<td>K-means</td>
</tr>
<tr>
<td>Sugar, Lenert, &amp; Olshen, 1999</td>
<td>Depression</td>
<td>K-means</td>
</tr>
<tr>
<td>Gröppel, Kapitany, &amp; Baumgartner, 2000</td>
<td>Non-epileptic seizure</td>
<td>Hierarchical</td>
</tr>
<tr>
<td>Whitwell et al., 2009</td>
<td>Fronto-temporal dementia</td>
<td>Hierarchical</td>
</tr>
<tr>
<td>Jamison, Rock, &amp; Parris, 1988</td>
<td>Chronic pain</td>
<td>K-means and Ward’s</td>
</tr>
<tr>
<td>Sanders &amp; Brena, 1993</td>
<td>Chronic pain</td>
<td>Hierarchical</td>
</tr>
<tr>
<td>Olson, Dey, Kumar, Monsen, &amp; Westra, 2016</td>
<td>Medicare home health patients</td>
<td>Hierarchical</td>
</tr>
</tbody>
</table>
Table 3-2 indicates that cluster analysis has been used to study a variety of medical conditions. The most frequently-used clustering algorithms are K-means clustering and hierarchical clustering.

3.3 Research objective

Given the likely shortage of medical resources to address the rapidly increasing demand for THA surgery in the next 15 years, it is very important to effectively use the limited medical resources to improve patient outcomes. Thus, cost-effectiveness is an important factor that should be considered in determining the optimal treatment option for each patient. The objective of this study is to propose a medical decision support framework for identifying THA patient clusters and analyzing the cost-effectiveness of THA for each cluster.

More specifically, we combine machine learning algorithms with Markov simulation models, thus allowing more accurate cost-effectiveness analysis at the patient subgroup level. There are two main components in this proposed framework. First, we segment THA patients into different clusters based on their demographics and co-morbid conditions. By using multiple factors to cluster patients, the patients in the same cluster have more homogeneous characteristics, hence leading to more accurate results. Second, we use cluster-specific Markov simulation models to analyze the cost-effectiveness of THA surgery for each patient cluster, using the non-surgery treatment as the benchmark. Therefore, the most appropriate treatment option can be better determined for each patient. With such a decision support framework, the limited medical resources of THA can be more effectively used to improve patient outcomes.

3.4 Methods

We used 1,347 observations on THA patients to conduct this study. These records were described in Chapter 2. We used supervised machine learning algorithms to identify the risk factors. Then we used unsupervised machine learning algorithms to partition hip arthritis patients into different clusters, based on the identified risk factors. The differences between supervised
and unsupervised machine learning algorithms were described in Section 2.4. The rationale of using risk factors for cluster analysis is to more accurately estimate the quality of life of different patients. Finally, we built cluster-specific Markov simulation models to analyze the cost-effectiveness of THA surgery for each patient cluster.

3.4.1 Supervised machine learning algorithms

We used logistic regression, decision tree, and random forest for identifying the risk factors of cluster analysis. The details of these three supervised machine learning algorithms were described in Sections 2.4.1, 2.4.2, and 2.4.3, respectively.

3.4.2 Unsupervised machine learning algorithms

In this study, we used K-means clustering algorithm and hierarchical clustering algorithm for partitioning hip arthritis patients into subgroups.

K-means clustering is a simple and elegant approach for partitioning a dataset of $n$ observations into $K$ distinct, non-overlapping clusters. To perform K-means clustering, we must first specify the desired number of clusters $K$; then the K-means clustering algorithm will assign each observation to exactly one of the $K$ clusters. The idea behind K-means clustering is that a good clustering result is one for which the within-cluster variation is as small as possible. The within-cluster variation is a measure of the amount by which the observations within a cluster differ from each other. Therefore, in K-means clustering we want to partition the observations into $K$ clusters such that the total within-cluster variation, summed over all the $K$ clusters, is as small as possible (James et al., 2013).

We also need to define the within-cluster variation before attempting to minimize it. The most common definition of this variation is squared Euclidean distance. Let $C_1, \ldots, C_k$ denote sets containing the indices of the observations in each cluster, and let $p$ denote the dimension of each observation. That is, we define
\[ W(C_k) = \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2 \]  

(3-1)

where \( |C_k| \) denotes the number of observations in the \( k \)th cluster.

However, when the variables are categorical or logical (binary) or a mix of different types, using a standard distance measure such as Euclidean distance is not appropriate. Since our THA patient-level data records were a mix of continuous variables and binary variables, we used Gower’s distance to measure the similarity between two different patients (Gower, 1971). Gower’s distance measure can be applied to continuous, categorical, and binary data. In Gower’s distance, the distance between two observations \( i \) and \( i' \) on the \( j \)th dimension is calculated as:

\[ S_{ii'j} = \frac{|x_{ij} - x_{i'j}|}{R_j} \]  

(3-2)

where \( R_j \) is the range of \( j \)th dimension among all the observations. The overall Gower’s distance is:

\[ S_{ii'} = \frac{\sum_{j=1}^{p} S_{ii'j}}{p} \]  

(3-3)

Thus, in our study the optimization problem that defines K-means clustering is:

\[
\text{minimize}_{C_1, \ldots, C_K} \left\{ \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} S_{ii'j} \right\}
\]

(3-4)

With this objective function, the K-means clustering algorithm is:

1) Randomly assign a number, from 1 to \( K \), to each of the observations. These serve as initial cluster assignments for the observations.

2) For each of the \( K \) clusters, compute the cluster centroid. The \( k \)th cluster centroid is the vector of the \( p \) feature means of the observations in the \( k \)th cluster.

3) Assign each observation to the cluster whose centroid is the closet, where closet is defined by the smallest Gower’s distance.

4) Iterate Step 2 and Step 3, until the cluster assignments stop changing.

When the assignment results no longer change, the optimal clustering has been found.
The hierarchical clustering algorithm also proceeds iteratively. At the beginning, each of the \( n \) observations is treated as its own cluster. The two clusters that are most similar to each other are then fused, so that now there are \( n-1 \) clusters. In our study, similarity between two observations is measured by Gower’s distance in Equation (3-3). Next, the two clusters that are most similar to each other are fused again, so that there now are \( n-2 \) clusters. The algorithm proceeds in this fashion until all of the observations belong to one single cluster. The output from hierarchical clustering is shown as a dendrogram, which represents the nested grouping of patterns and similarity levels at which groupings change (Jain, Murty, & Flynn, 1999).

One important issue in hierarchical clustering is how to determine the two clusters that should be fused, given that one or both of these clusters contain multiple observations. This is defined as linkage, which is the dissimilarity between two groups of observations. The three most common types of linkage – complete, single, and average – are briefly described in Table 3-3.

<table>
<thead>
<tr>
<th>Linkage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete</td>
<td>Maximal inter-cluster dissimilarity. Compute all pairwise dissimilarities between the observations in cluster A and the observations in cluster B, and record the largest of these dissimilarities</td>
</tr>
<tr>
<td>Single</td>
<td>Minimal inter-cluster dissimilarity. Compute all pairwise dissimilarities between the observations in cluster A and the observations in cluster B, and record the smallest of these dissimilarities</td>
</tr>
<tr>
<td>Average</td>
<td>Mean inter-cluster dissimilarity. Compute all pairwise dissimilarities between the observations in cluster A and the observations in cluster B, and record the average of these dissimilarities</td>
</tr>
</tbody>
</table>

Average linkage and complete linkage are generally preferred over single linkage, as they tend to yield more balanced dendrograms (James et al., 2013). The linkages in our study are based on Gower’s distance between observations. At each step, the two clusters with the smallest linkage value are fused.
3.4.3 Markov simulation model

Markov simulation model is a stochastic model used to study randomly changing systems. State transitions depend on the current state and the transition probabilities. It captures the transitions between stages of life and is particularly suitable for modeling the long-term benefits of THA as we aim for. We used Markov simulation model to estimate the costs and quality of life components of hip arthritis patients in each cluster. In addition, the model parameters are cluster-specific.

Relevant studies indicate that readmission after THA surgery mainly occurs within the first 12 months after surgery (Enocson et al. 2009; Zmistowski et al. 2013; Schairer et al. 2014). Thus, if a patient is readmitted to the hospital at least once within the year after surgery, he/she enters the “Readmission” state in the Markov simulation model. Moreover, relevant studies indicate that most of the THA patients need at most one re-operation on the artificial hip joint, and very few of them need more than two re-operations (Daly & Morrey, 1992; Schulte, Callaghan, Kelley, & Johnston, 1993; Mulroy, Estok, & Harris, 1995; Madey, Callaghan, Olejniczak, Goetz, & Johnston, 1997). In addition, with the advancement in implant material, artificial hip joint can be used for an increasingly longer time period. Thus, we assumed that a patient would have at most two re-operations during his/her remaining life after THA surgery. The Markov simulation model of the surgery option includes the following states: THA, Readmission, Fair, 1st re-operation, Fair after 1st re-operation, 2nd re-operation, Fair after 2nd re-operation, and Death. This Markov simulation model is shown in Figure 3-1.

![Figure 3-1. Markov simulation model of the surgery option](image-url)
Each patient enters the Markov simulation model when he/she receives THA operation (the state “THA”), and the patient is tracked as he/she moves through different health states until death. Transitions between states are assumed to occur annually and depend on state-specific transition probabilities. The transition probabilities are estimated from relevant literature on THA as well as the identified THA patient data that we obtained. Age-dependent death rates (i.e., not relevant to THA) are obtained from the 2011 United States life table (Arias 2011). These transition probabilities are summarized in Table 3-4.
Table 3-4. Transition probabilities between health states

<table>
<thead>
<tr>
<th>Starting state</th>
<th>Ending state</th>
<th>Transition probability</th>
<th>Range</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>THA</td>
<td>Readmission</td>
<td>Cluster-specific</td>
<td>0.0232-0.1818</td>
<td>THA patient-level data records</td>
</tr>
<tr>
<td>THA</td>
<td>Death</td>
<td>One-year death rate after THA</td>
<td>0.008702-0.040322</td>
<td>Murray, Britton, &amp; Bulstrode, 1996; Dearborn &amp; Harris, 1998; Wood et al., 2002; Mahomed et al., 2003; Arias 2011</td>
</tr>
<tr>
<td>THA</td>
<td>Fair</td>
<td>1- the previous two rows</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Fair</td>
<td>Death</td>
<td>Age-dependent</td>
<td>N/A</td>
<td>Arias 2011</td>
</tr>
<tr>
<td>Fair</td>
<td>1st re-operation</td>
<td>0.009</td>
<td>0.008-0.009</td>
<td>Kavanagh, Dewitz, Ilstrup, Stauffer, &amp; Coventry, 1989</td>
</tr>
<tr>
<td>Fair</td>
<td>Fair</td>
<td>1- the previous two rows</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Readmission</td>
<td>1st re-operation</td>
<td>0.039</td>
<td>N/A</td>
<td>Nolan, Fitzgerald, Beckenbaugh, &amp; Coventry, 1975</td>
</tr>
<tr>
<td>Readmission</td>
<td>Fair</td>
<td>0.961 (i.e., 1-0.039)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>1st re-operation</td>
<td>Death</td>
<td>0.026</td>
<td>N/A</td>
<td>Mahomed et al., 2003</td>
</tr>
<tr>
<td>1st re-operation</td>
<td>2nd re-operation</td>
<td>0.06</td>
<td>0.015-0.265</td>
<td>Chang et al., 1996</td>
</tr>
<tr>
<td>1st re-operation</td>
<td>Fair after 1st re-operation</td>
<td>0.914 (i.e., 1-0.026-0.06)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Fair after 1st re-operation</td>
<td>Death</td>
<td>Age-dependent</td>
<td>N/A</td>
<td>Arias 2011</td>
</tr>
<tr>
<td>Fair after 1st re-operation</td>
<td>2nd re-operation</td>
<td>0.04</td>
<td>0.01-0.22</td>
<td>Chang et al., 1996</td>
</tr>
<tr>
<td>Fair after 1st re-operation</td>
<td>Fair after 1st re-operation</td>
<td>1- the previous two rows</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>2nd re-operation</td>
<td>Death</td>
<td>0.05</td>
<td>N/A</td>
<td>Assumption, no relevant studies found</td>
</tr>
<tr>
<td>2nd re-operation</td>
<td>Fair after 2nd re-operation</td>
<td>0.95 (i.e., 1-0.05)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Fair after 2nd re-operation</td>
<td>Death</td>
<td>Age-dependent</td>
<td>N/A</td>
<td>Arias 2011</td>
</tr>
<tr>
<td>Fair after 2nd re-operation</td>
<td>Fair after 2nd re-operation</td>
<td>1- previous row</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

The Markov simulation model calculates the expected lifetime medical cost and quality of life of patients in each cluster in the surgery option. Future cost and quality of life are discounted at
3%, which are in line with the criterion of Global Burden of Disease (GBD) study (Sassi 2006). The discounted qualify of life is QALY. Finally, the cost and QALY of the surgery option are compared with those of the non-surgery option to calculate the cost-effectiveness ratio. The cost and quality of life parameters used in this study are introduced in the next two sections.

### 3.4.4 Cost parameters

In this cost-effectiveness analysis, we considered the total cost involved in caring for a patient (i.e., from a healthcare provider’s perspective), not just the patient’s out-of-pocket cost. The cost of THA surgery varies greatly among providers, and the Medicare expenditure for surgery, hospitalization, and recovery ranges from $16,500 to $33,000 across geographic areas (CMS, 2016b). This range coincides well with the total fair price of THA surgery on Healthcare Bluebook, namely $27,849 (Healthcare Bluebook, 2016a), which is the value used in our analysis. The cost of a revision hip replacement is estimated to be $37,787 (Healthcare Bluebook, 2016b).

After THA surgery, a patient needs multiple types of medical resources to maintain the gain from surgery. A patient with conservative treatment needs similar types and even more medical resources for dealing with hip arthritis (Rissanen et al., 1997; American Association of Hip and Knee Surgeons (AAHKS), 2016a; Teeny, York, Mesko, & Rea, 2003; American Nurse Today, 2016; Arthritis Foundation, 2016). These medical resources are categorized in Table 3-5.
Table 3-5. A summary of medical resources needed by patients

<table>
<thead>
<tr>
<th>Medical resource</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medications</td>
<td>Used in both surgery and conservative treatments. Pain relievers, such as acetaminophen (Tylenol), are needed for reducing pain. Non-steroidal, anti-inflammatory medicines such as aspirin, ibuprofen (Motrin or Advil), and naproxen (Aleve), can help reduce pain and swelling in the joint. More potent types of pain relievers are prescription-strength, non-steroidal, anti-inflammatory drugs that can be prescribed by the doctor.</td>
</tr>
<tr>
<td>Physical therapist</td>
<td>Patients in both groups need to visit the physical therapist. Physical therapy to strengthen the muscles around the hip joint may help absorb some of the shock imparted to the joint. Physical therapy can help reduce pain, swelling, and stiffness of arthritis, and it can help improve joint function. It can also make it easier for the patient to walk, bend, kneel, squat, and sit.</td>
</tr>
<tr>
<td>Orthopedist/rheumatologist</td>
<td>The patients in surgery option need to visit the orthopedist for follow-up evaluation and radiograph interpretation of their artificial hip joints. The patients in conservative treatment should visit the rheumatologist to check the status and progression of hip arthritis, as well as for treatment.</td>
</tr>
<tr>
<td>Primary care physician</td>
<td>The patients in the non-surgery group need to visit their primary care physicians for diagnosis of hip arthritis and for referral to rheumatologists. After that, primary care physicians continue to handle most of the care for these patients. But for patients in the surgery group, they visit primary care physicians at a much lower frequency, given that the artificial hip joints function as expected.</td>
</tr>
<tr>
<td>Nurse</td>
<td>Nurses assist orthopedist/rheumatologist and primary care physician during patient visits.</td>
</tr>
</tbody>
</table>

The costs in Table 3-5 may differ between the surgery option and non-surgery option. Medication cost includes the costs of prescription drugs and non-prescription drugs. Rissanen et al. (1997) indicated that the consumption of prescription drugs after THA surgery remained at the pre-operative level, while the use of other drugs slightly increased. Therefore, the medication costs are assumed to be the same in the surgery option and non-surgery option.

Physical therapy is needed for both treatment options. The number of prescribed physical therapy sessions for hip arthritis patients varies across different studies. Some of these sessions can be 10 weeks long (Hermann, Holsgaard-Larsen, Zerahn, Mejdahl, & Overgaard, 2016), while other sessions may only take 2 to 6 weeks (Bitterli, Sieber, Hartmann, & De Bruin, 2011; McGregor, Rylands, Owen, Doré, & Hughes, 2004; Hoogeboom, Dronkers, van den Ende, Oosting, & van Meeteren, 2010; Oosting et al., 2012). Typically, the patient visits the physical
therapist between 1 and 3 times a week, and each training session lasts for about 1 hour. After the training session, the patient repeats the exercise on his/her own. Depending on the patient’s pain and function, the physical therapist may change the content of therapy after a certain time period. Rissanen et al. (1997) indicated that THA surgery significantly reduces the number of patient visits to physical therapist. Thus, by averaging the results of previous studies on physical therapy, we estimated that a THA surgery patient visits the physical therapist 9 times/year in each subsequent year after surgery, while a patient with conservative treatment visits the physical therapist 18 times/year. The cost of a visit to physical therapist was estimated to be $100 (GuideDoc, 2016).

The next component is the cost of visiting orthopedist or rheumatologist. The survey on members of AAHKS indicated that most surgeons advocate biennial or more frequent follow-up evaluation and radiograph interpretation of each THA patient (Teeny et al., 2003). Thus, we assumed that a THA patient needs to visit the orthopedist annually. Relevant studies indicate that on average an arthritis patient visits rheumatologist three times per year (Katz et al., 1998; Põlluste, Kallikorm, Meiesaar, & Lember, 2011). The cost of visiting an orthopedist or rheumatologist was estimated to be $300 (HealthTap, 2016).

As indicated in Table 3-5, a non-surgery patient needs to visit primary care physician for relieving and treating his/her hip arthritis (Solomon et al., 2002; Buszewicz et al., 2006). We referred to relevant studies, and these studies indicate that on average a non-surgery patient needs to visit primary care physician four times per year (Põlluste et al., 2011). For a THA patient, he/she may meet primary care physician much less often, given that the artificial hip joint functions well. To the best of our knowledge, we did not find any study on how often a THA patient needs to visit primary care physician on his/her artificial hip joint. Thus, we assumed that on average a THA patient needs to meet primary care physician specifically regarding his/her artificial hip joint once per year. The cost of a visit to primary care physician was estimated to be $128 (Johns Hopkins Bloomberg School of Public Health, 2015).

Relevant studies indicate that at least one nursing staff member was listed as a provider at 50.6% of all arthritis visits to primary care physician or rheumatologist in the United States (Hootman, Helmick, & Schappert, 2002). Thus, based on the frequency that THA patients and
non-surgery patients visit orthopedist/rheumatologist and primary care physician, we can obtain the frequency that they visit nursing staff member. By calculating in this manner, on average a THA patient visits nursing staff member once per year, while a non-surgery patient visits nursing staff member four times per year. The cost of a nurse visit was estimated to be $30 (Care.com, 2016).

A summary of these annual medical costs for the surgery option (i.e., in the state “Fair” after THA surgery) and non-surgery option are shown in Table 3-6.

<table>
<thead>
<tr>
<th>Type of cost</th>
<th>Annual cost for surgery option ($)</th>
<th>Annual cost for non-surgery option ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical therapist</td>
<td>900</td>
<td>1,800</td>
</tr>
<tr>
<td>Orthopedist/rheumatologist</td>
<td>300</td>
<td>900</td>
</tr>
<tr>
<td>Primary care physician</td>
<td>128</td>
<td>512</td>
</tr>
<tr>
<td>Nursing staff</td>
<td>30</td>
<td>120</td>
</tr>
<tr>
<td>Total annual cost</td>
<td>1,358</td>
<td>3,332</td>
</tr>
</tbody>
</table>

When considering the surgery-related cost components of THA surgery, we also need to take into account the expected costs of using rehabilitation facility and home health. This is because some THA patients are discharged to these facilities or use these services after discharge. The average costs of using rehabilitation facility and home health were estimated to be $7,564 and $4,627 per THA patient, respectively (Sigurdsson et al., 2008). Thus, the total cost of the state “THA” is the sum of surgery cost, expected cost of using rehabilitation facility, and expected cost of using home health. These expected costs are cluster specific, because the percentage of patients using these services differs among clusters. Similarly, the total cost of the states “1st re-operation” and “2nd re-operation” are cluster-specific. We assumed that the annual cost of fair outcome after re-operation is the same as that of fair outcome after THA surgery. Finally, we should consider the expected cost of treating unplanned readmission after THA surgery. The average cost of treating a readmitted THA patient was estimated to be $15,333 (Qasim & Andrews, 2012; Becker’s Infection Control & Clinical Quality, 2013). Since the probability of being readmitted differs among clusters, the expected cost of the state “Readmission” is also
cluster specific. All the aforementioned costs have been inflated to 2016 US dollars using the medical care component of the Consumer Price Index.

3.4.5 Quality of life parameters

In our study, the health-related quality of life (HRQoL) is measured by EuroQol five dimensions questionnaire (EQ-5D) scale. EQ-5D is a standardized instrument for measuring generic health status. The EQ-5D includes five dimensions: mobility, self-care, usual activities, pain or discomfort, and anxiety or depression. These are each divided into three categories of severity corresponding to no, moderate, or severe/extreme problems. The EQ-5D thereby defines 243 different health states that can be converted into a single index score representing health utilities. The EQ-5D score ranges from 0 (being dead) to 1 (perfect health). The health status measured with EQ-5D is used for estimating quality of life of that health status. Then the quality of life can be combined with time to compute QALY. EQ-5D is one of the most commonly-used HRQoL measures, and its good validity and reliability have been reported in various health conditions (Garratt, Schmidt, Mackintosh, & Fitzpatrick, 2002).

Many of the patients in our patient-level data records have multiple chronic conditions. Since old age and co-morbid conditions decrease a patient’s quality of life, we looked for relevant information from literature to quantify their impacts on HRQoL. Saarni et al. (2006) studied the impacts of old age and 29 chronic conditions on HRQoL. The chronic conditions in this study include cancer, diabetes, heart failure, hip or knee arthritis, etc. Tajima et al. (2010) studied the impacts of different stages of chronic kidney disease on HRQoL. Jia & Lubetkin (2005) as well as Sach et al. (2007) studied the effects of overweight and obesity on HRQoL. Jia & Lubetkin (2005) also quantified the influence of hypertension on HRQoL. Enocson et al (2009) investigated the impact of dislocation on HRQoL of THA patients. More importantly, all these studies used EQ-5D as the HRQoL measure. Thus, we used these results to quantify the quality of life of patients in our study. A summary of the HRQoL parameters used in our cost-effectiveness analysis is shown in Table 3-7.
Table 3-7. A summary of the HRQoL parameters

<table>
<thead>
<tr>
<th>Condition</th>
<th>Impact on HRQoL in EQ-5D (mean and standard deviation (SD) or P-value)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 54-74</td>
<td>-0.043 (SD=0.010)</td>
<td>Saarni et al., 2006</td>
</tr>
<tr>
<td>Age 75-79</td>
<td>-0.077 (SD=0.013)</td>
<td>Saarni et al., 2006</td>
</tr>
<tr>
<td>Hip arthritis</td>
<td>-0.101 (SD=0.007)</td>
<td>Saarni et al., 2006</td>
</tr>
<tr>
<td>Kidney disease</td>
<td>-0.117 (SD=0.013)</td>
<td>Tajima et al., 2010</td>
</tr>
<tr>
<td>Hypertension</td>
<td>-0.053 (P&lt;0.0001)</td>
<td>Jia &amp; Lubetkin, 2005</td>
</tr>
<tr>
<td>Overweight (BMI 25.00-29.99)</td>
<td>-0.013 (P=0.0115)</td>
<td>Jia &amp; Lubetkin, 2005</td>
</tr>
<tr>
<td>Class I obesity (BMI 30.00-34.99)</td>
<td>-0.033 (P&lt;0.0001)</td>
<td>Jia &amp; Lubetkin, 2005</td>
</tr>
<tr>
<td>Complication or readmission</td>
<td>-0.275 (SD=0.110)</td>
<td>Enocson et al., 2009</td>
</tr>
</tbody>
</table>

To the best of our knowledge, the existing studies did not compare the quality of life of patients after re-operation with that after THA surgery. Thus, we assumed that the quality of life immediately after re-operation is similar to that immediately after THA surgery. In addition, we assumed that if a patient achieves fair outcome after 1st re-operation or 2nd re-operation, the patient’s quality of life is comparable to that of fair outcome after THA surgery.

### 3.5 Results

#### 3.5.1 Risk factors identified by supervised learning algorithms

In order to conduct an effective cluster analysis, we need to identify the factors that can be used to classify hip arthritis patients into different subgroups. Hip arthritis patients have diverse demographic, co-morbid, and surgery-related conditions. In addition, they have very different risks of post-operative readmission, given that they select to receive THA surgery. The readmission significantly influences a patient’s post-operative quality of life (Enocson et al., 2009; Knutsson, & Engberg, 1999).

We used three supervised machine learning algorithms to identify the risk factors for clustering hip arthritis patients: logistic regression, decision tree, and random forest. In Chapter 2, Table 2-9 indicates that these three algorithms identified a set of common risk factors, including BMI, discharge to rehabilitation facility or SNF, kidney disease, age, hypertension, and
male. All these factors are positively correlated with the 30-day unplanned readmission risk, which significantly influences quality of life parameters.

It should be noted that the discharge disposition of a patient is only generated in the surgery option. If discharge disposition were used as a factor for clustering patients, it would indicate that all future patients will receive THA surgery. Since this might not be true in the cost-effectiveness analysis, we excluded discharge to rehabilitation facility or SNF from the risk factor set. Therefore, the risk factors for clustering patients include age, male, hypertension, kidney disease, and BMI.

3.5.2 Cluster analysis of patients

We used two algorithms, K-means clustering algorithm and hierarchical clustering algorithm, for clustering hip arthritis patients into subgroups. All the analyses were conducted in statistical software R 3.1.0 using package “kmeans” and “hclust” (R Development Core Team, 2008). Gower’s distance was used as the distance measure between data points in both clustering algorithms. In hierarchical clustering algorithm, complete linkage was used to measure the distance between clusters. For K-means clustering, we need to determine the appropriate value of k, which is the number of clusters. There is no single right answer to this question. In practice, we need to try several different choices, and look for the one with the most useful or interpretable solution. Within-cluster homogeneity is considered as a very important criterion in K-means clustering (Eslick et al. 2004; James et al. 2013). Figure 3-2 shows the total within-cluster variance as a function of k.
In practice, several different choices of k should be considered, and the one with the most useful or interpretable solution should be used (James et al., 2013). Figure 3-2 shows that after k=8, the total within-cluster variance decreases at a much slower rate than that before. This indicates that further increasing k will no longer provide significantly distinguishable patient clusters. However, in order to effectively determine the cost-effectiveness of THA for each patient subgroup, the value of k should not be too small. Therefore, the options that could be considered are k=5, 6, 7, and 8. In addition, it is very important to ensure that different clusters are highly distinguishable. If a new solution splits large clusters into more sub-clusters, but these sub-clusters are not highly distinguishable, we still prefer the original solution. This is our second criterion. With these two criteria, we determined to use six clusters. The clustering results with k=6 are shown in Table 3-8.

Moreover, we compared the results of K-means clustering algorithm with the results of hierarchical clustering algorithm. The rationale is that if these two algorithms generate very similar clustering results at a certain k value, the patient clusters in this scenario are probably distinguishable and interpretable. The details of comparing the clustering results at k=5, 6, 7, and 8 are shown in Appendix E. There was no scenario in which these two algorithms generated exactly the same clustering results. But at k=6, K-means clustering and hierarchical clustering produced the most similar results, and the clusters are distinguishable and interpretable. The results of hierarchical clustering are shown in Table 3-9.
Table 3-8. K-means clustering results with six clusters

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Size</th>
<th>Average age (year)</th>
<th>Percentage of male</th>
<th>Percentage of hypertension</th>
<th>Percentage of kidney disease</th>
<th>Average BMI (kg/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>325</td>
<td>64.58</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>30.97</td>
</tr>
<tr>
<td>2</td>
<td>221</td>
<td>54.18</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>30.05</td>
</tr>
<tr>
<td>3</td>
<td>302</td>
<td>60.12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30.16</td>
</tr>
<tr>
<td>4</td>
<td>402</td>
<td>70.29</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>30.19</td>
</tr>
<tr>
<td>5</td>
<td>75</td>
<td>75.47</td>
<td>52</td>
<td>0</td>
<td>100</td>
<td>30.55</td>
</tr>
<tr>
<td>6</td>
<td>22</td>
<td>69.82</td>
<td>40.91</td>
<td>100</td>
<td>100</td>
<td>33.09</td>
</tr>
</tbody>
</table>

Table 3-9. Hierarchical clustering results with six clusters

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Size</th>
<th>Average age (year)</th>
<th>Percentage of male</th>
<th>Percentage of hypertension</th>
<th>Percentage of kidney disease</th>
<th>Average BMI (kg/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>325</td>
<td>64.58</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>30.97</td>
</tr>
<tr>
<td>2</td>
<td>221</td>
<td>54.18</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>30.05</td>
</tr>
<tr>
<td>3</td>
<td>302</td>
<td>60.12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30.16</td>
</tr>
<tr>
<td>4</td>
<td>402</td>
<td>70.29</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>30.19</td>
</tr>
<tr>
<td>5</td>
<td>48</td>
<td>71.42</td>
<td>100</td>
<td>18.75</td>
<td>100</td>
<td>30.27</td>
</tr>
<tr>
<td>6</td>
<td>49</td>
<td>76.90</td>
<td>0</td>
<td>26.53</td>
<td>100</td>
<td>31.96</td>
</tr>
</tbody>
</table>

Table 3-8 and Table 3-9 indicate that the clusters 1, 2, 3 and 4 generated by these two clustering algorithms are exactly the same, respectively. The differences are in clusters 5 and 6. K-means clustering separates these 97 THA patients into a cluster of 75 patients and the other of 22 patients, while hierarchical clustering divides them into a cluster of 48 patients and the other of 49 patients. Since four out of the six clusters are identical, these two algorithms reached very similar results at k=6. Table 3-8 and Table 3-9 also indicate that age, gender, hypertension, and kidney disease tend to define the unique characteristics of each patient cluster. For example, cluster 1 is comprised of male patients around 65 years old, all with hypertension, but no kidney disease. Cluster 3 consists of female patients around 60 years old, with no hypertension and no kidney disease. In contrast, BMI does not differ much among different clusters.

3.5.3 Cost-effectiveness analysis of each patient cluster

The clusters 5 and 6 in K-means clustering are more distinguishable, in terms of co-morbidities, than the corresponding clusters in hierarchical clustering. Co-morbidities have
significant influences on the quality of life of patients. Therefore, we used cluster-specific Markov simulation model to calculate the cost-effectiveness of THA surgery for each patient cluster in K-means clustering. We used the non-surgery option as the benchmark treatment. Statistical software R 3.1.0 was used for programming the Markov simulation models (R Development Core Team, 2008). The cost-effectiveness ratios of THA surgery for each of the six patient clusters are shown in Table 3-10. The results are based on 500 runs of the Markov simulation model for each cluster.

Table 3-10. Cost-effectiveness ratios of THA surgery for each cluster

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Cost-effectiveness ratio ($/QALY)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7,964.24</td>
</tr>
<tr>
<td>2</td>
<td>2,528.69</td>
</tr>
<tr>
<td>3</td>
<td>16,332.78</td>
</tr>
<tr>
<td>4</td>
<td>22,017.56</td>
</tr>
<tr>
<td>5</td>
<td>64,125.93</td>
</tr>
<tr>
<td>6</td>
<td>27,314.51</td>
</tr>
</tbody>
</table>

The most commonly-used cost-effectiveness threshold is $50,000/QALY (Owens et al., 2011). Thus, THA surgery is cost-effective for all the patient clusters, except cluster 5. Table 3-8 indicates that cluster 5 corresponds to the oldest group of patients, with an average age of 75.47. These patients do not have hypertension, but they have kidney disease. Since this patient group has the shortest expected remaining lifespan, there are fewer downstream gains in quality of life in the surgery option. In addition, a high percentage of cluster 5 patients incur the costs of rehabilitation facility, home health, and readmission. In contrast, the expected remaining lifespan is longer in other patient groups, thus the savings in subsequent years can more easily justify the upfront surgery cost. Moreover, a lower percentage of patients in these clusters incur the costs of rehabilitation facility, home health, and readmission, except for cluster 6. Among all the six patient clusters, cluster 2 has the smallest cost-effectiveness ratio, followed by cluster 1 and cluster 3. Table 3-8 indicates that these three patient groups are the youngest among all the six clusters, and a lower percentage of them have hypertension or kidney disease.

If no cluster analysis were conducted for the 1,347 patients in this study, the cost-effectiveness ratio of THA surgery would be $14,252/QALY, compared with the non-surgery treatment. Thus,
we may think that THA is cost-effective for all the 1,347 patients. However, we showed that actually it is not cost-effective for patients in cluster 5.

3.6 Sensitivity analysis

One of the main problems in cost-effectiveness analysis is that some parameters are subject to change. Thus, sensitivity analysis on these parameters is needed to provide a better understanding of how robust the results are. In particular, sensitivity analysis of model parameters in multi-state Markov simulation models is a critical component in evaluating the model findings.

In this study, probabilistic sensitivity analysis was performed on all the model parameters. Specifically, the transition probabilities were assumed to be uniformly distributed in the range provided in Table 3-4. A 95% confidence interval was used for each entry of the quality of life parameters. In addition, the THA surgery cost, re-operation cost, and the annual cost of fair outcomes were assumed to be uniformly distributed in the range of zero to twice of the original cost. The probabilistic sensitivity analysis was based on Monte-Carlo simulation with 500 runs, and the parameters varied randomly according to their respective distributions. This approach examines the effect of joint uncertainty in the model’s variables. The results of probabilistic sensitivity analysis are shown via the cost-effectiveness acceptability curves in Figure 3-3. The curves are interpreted as the probability that the true cost-effectiveness ratio falls below a certain value.
Figure 3-3 indicates that with the threshold of $50,000/QALY, THA surgery is cost-effective with probabilities 0.80, 0.90, 0.77, 0.84, 0.40, and 0.78 for each cluster, respectively. The results of probabilistic sensitivity analysis are quite in accordance with the results of base case analysis, indicating that the cost-effectiveness ratio is probably above $50,000/QALY for cluster 5.

3.7 Discussion

Clinical studies indicated that THA significantly improves patients’ quality of life with respect to pain relief, physical function, and overall health (Laupacis et al., 1993; Jones et al., 2001).
However, due to the aging population and prevalence of obesity in the United States, there may be insufficient economic resources and technical capacities (i.e., additional surgeons) to meet the ever-increasing demand for THA in the next 15 years (Kurtz et al., 2007; Iorio et al., 2008). In addition, research shows that THA patients in the United States are less functionally limited before surgery than THA patients in Canada (Laupacis et al., 1993; Katz, Phillips, Fossel, & Liang, 1994). This suggests THA surgery may not be the necessary treatment option for all hip arthritis patients in the United States. Therefore, performing THA surgery on the most suitable patients becomes a crucial consideration in resource allocation. Some European countries, such as the United Kingdom, are facing similar challenges of insufficient supply of THA surgery. Health policy makers in these countries are also thinking about how to effectively use THA surgery (Jenkins et al., 2013).

In this study, we clustered patients into different subgroups, using the patient-level data records from an academic medical center in central Pennsylvania, United States. We examined the clustering results to make sure that the obtained patient clusters are highly distinguishable and interpretable. After that, we analyzed whether THA surgery is a cost-effective treatment option for each patient cluster, compared with conservative treatment. The results indicate that THA is cost-effective for all the patient clusters, except the cluster of the oldest patients with co-morbidities. The findings are robust over a wide range of sensitivity analysis.

Our findings help answer questions about whether THA is cost-effective for the oldest age category (Chang et al., 1996). Our results are in accordance with a previous short-term clinical study (only focused on two years after surgery), indicating that THA surgery is more cost-effective for patients younger than 70 (Rissanen et al., 1997). Moreover, the clinical appropriateness of THA in older age group has already been debated in the relevant literature (Jacobsson, Rehnberg, & Djerf, 1991; Boettcher 1992). Combined with the findings in our study, alternative treatments are probably the preferable options for hip arthritis patients older than 75 but without very severe pain and function problems, from the perspectives of both cost-effectiveness and THA surgery risk. However, it should be noted that although cost-effectiveness is one important criterion in such medical decision making process, it is not the only criterion for determining the appropriate treatment option.
Our study combines machine learning algorithms with cluster-specific Markov simulation models, allowing specific and detailed cost-effectiveness analysis to be conducted at the patient subgroup level. If the cost-effectiveness analysis were conducted at the entire population level, we would have mistakenly concluded that THA is cost-effective for all the patients. In addition, multiple criteria, including demographic factors and co-morbid conditions, are used to determine the subgroup to which a patient belongs. This approach is more accurate than using only a single factor, such as age or gender, to classify patients. The patient subgroups obtained with our clustering methods probably have more homogeneous characteristics. Therefore, the treatment option determined for each patient cluster may lead to less variability in patient outcomes.

Moreover, this study offers important insights into effectively utilizing the limited medical resources of THA. For a future patient who comes into a hospital with hip arthritis, our clustering algorithm can be used to determine the cluster to which the patient belongs, based on the Gower’s distance of this patient and the six cluster centroids. With his/her cluster assignment, the medical team can better determine whether THA surgery can significantly benefit this patient. Therefore, our proposed framework can be used to optimally utilize THA surgery by identifying the patients who can benefit the most from surgery, in terms of gained quality of life. Since patient outcome is one of the most important measures of healthcare quality (Donabedian, 1966; Ko, 2009), increasing the quality of life of hip arthritis patients is a crucial element in the THA quality improvement framework.

Our study also has important health policy implications. First, with the likely shortage of THA medical resources in the next 15 years, health policies need to be carried out to determine the criteria for receiving THA surgery, from the perspectives of patients’ health condition, cost-effectiveness, surgery risk, and patients’ willingness. Second, health policy makers may want to re-consider any policies that may potentially limit medical resources to THA surgery as a whole: THA is cost-effective for the majority of the patients, e.g., 94.43% in this study. On the other hand, preventions and interventions should be carried out to alleviate the rapidly increasing demand for THA surgery. For example, we need to more closely examine how weight control plans and physical exercises plans are implemented on patients with slight to moderate hip
arthritis. These two interventions are very effective in preventing hip arthritis from getting worse.

There are some limitations in this study. First, more detailed sensitivity analysis may be helpful. For example, there is no confirmed range for some of the transition probabilities (e.g., the age-dependent death rate) in the Markov simulation model, so we are unable to incorporate these ranges into the sensitivity analysis. Second, we obtained the impacts of old age and co-morbidities on quality of life from relevant literature. A more accurate way to estimate these impacts is to conduct a questionnaire survey on the THA patients, which, however, would require significant resources that we do not have now. Third, our cost-effectiveness analysis is based on the THA patients in an academic medical center in central Pennsylvania, United States. While we believe this sample is representative, conducting similar analysis on bigger samples from other sources would potentially help to generalize our results.

### 3.8 Conclusion

In this study, we used machine learning algorithms to partition hip arthritis patients into different clusters, based on which we used cluster-specific Markov simulation models to determine whether THA surgery is cost-effective for each patient cluster. The results indicate that patients younger than 65 and with few co-morbidities tend to benefit the most from THA surgery, while THA surgery is not cost-effective for patients older than 75 with co-morbidities.

With the rapidly increasing demand for THA in the next 15 years, it is important to effectively use the limited medical resources of THA. This study indicates that it is very important to analyze the cost-effectiveness of THA surgery at the patient subgroup level. Otherwise, we would have mistakenly concluded that THA is cost-effective for all the patients. Such wrong conclusion may mislead THA medical resources utilization. Moreover, health policies on increasing THA capacity and preventing serious hip arthritis should be considered and implemented to alleviate the potential shortage of THA in the future.
By combining machine learning algorithms with Markov simulation models, we proposed a data-driven approach that can be used to effectively use THA surgery to achieve the optimal patient utility and improve patient outcome. Since patient outcome, especially gained quality of life, is an important measure of THA care quality, our data-driven approach fits into the framework of THA care quality improvement and facilitates such efforts.
Chapter 4 Meta-Analysis of the Effectiveness of Pre-Operative Exercise

4.1 Introduction

Individuals with advanced hip arthritis are more likely to have a low level of muscle strength and exercise tolerance, compared with healthy individuals (Murray, Brewer, & Zuege, 1972; Vaz, Kramer, Rorabeck, & Bourne, 1993; Shih, Du, Lin, & Wu, 1994; Neumann, 1999). Increased pain and deterioration of the hip also lead to decreased mobility and reduced independence in some activities of daily living (Bunning & Materson, 1991). To mitigate this problem, appropriate exercise has been recommended for patients with hip arthritis. Appropriate exercise results in stronger, better conditioned peri-articular muscles, tendons, and ligaments, attenuating joint forces during movement (Felson et al., 2000).

Exercise is also a cornerstone of rehabilitation following THA and other surgical procedures (Kuster, 2002). In addition, some researchers indicated that pre-operative fitness is predictive of post-operative outcome of THA (Fortin et al., 1999). Thus, in the last two decades, pre-operative exercise and physical therapy have been suggested for patients waiting for THA surgery. Typically, patients scheduled for THA have to wait for more than one month before actually being operated on (Wallis & Taylor, 2011). Therefore, some physical therapists around the world have carried out pre-operative exercise on these patients during this waiting period. The exercise and physical therapy include some forms of strengthening, flexibility, and aerobic exercise, with the purpose of improving muscle strength and mobility at the pre-surgery period. The pre-operative exercise program also aims at positively influencing the post-operative outcomes.

There is a considerable number of studies on the effect of pre-operative exercise on patients with lower-extremity surgery, especially for hip patients and knee patients. However, these studies have different design, sample size, exercise duration, exercise frequency, exercise type, etc. Moreover, these studies show mixed results on the effectiveness of pre-operative exercise.

Meta-analysis is a statistical analysis that combines the results of multiple scientific studies. The basic tenet behind meta-analysis is that there is a common truth behind all conceptually
similar scientific studies, but the common truth has been measured with a certain error within the individual studies. The aim is to use approaches from statistics to derive a pooled estimate closest to the unknown common truth, based on how this error is perceived. In essence, all existing methods yield a weighted average from the results of the individual studies. What differs is the manner in which these weights are allocated and the manner in which the uncertainty is computed around the generated point estimate. In addition to providing an estimate of the unknown common truth, meta-analysis has the capacity to contrast results from different studies and identify patterns among study results, sources of disagreement among those results, or other interesting relationships that may come to light in the context of multiple studies (Rothman, Greenland, & Lash, 2008). Meta-analysis can be thought of as “conducting research on previous research”. Meta-analysis can only proceed if we are able to identify a common statistical measure that is shared among studies. In addition, the common measure should have a standard error, so that we can proceed with computing a weighted average of this common measure. Such weighting usually takes into consideration the sample sizes of the individual studies, although it can also include other factors, such as study quality.

A key benefit of this approach is the aggregation of information, leading to a higher statistical power and more robust point estimate than those in any individual studies (Walker, Hernandez, & Kattan, 2008). Meta-analysis is often an important component of systematic review procedures. For instance, a meta-analysis may be conducted on several clinical trials of a medical treatment, in an effort to obtain a better understanding of the treatment effect.

In this chapter, we conducted a meta-analysis on the existing studies to investigate the effect of pre-operative exercise on THA patient outcomes.

4.2 Literature review

The studies on the effectiveness of pre-operative exercise on THA patients vary in the patient demographics and severity of hip arthritis, exercise program type, duration of the program (i.e., how many weeks), frequency of the exercise class (i.e., how many times a week), duration of each class, exercise intensity, and so on. Moreover, some studies focused more on the
effectiveness at the pre-operative phase (i.e., right after the patient completes the exercise program), while others concentrated more on the impact at the post-operative period. Finally, these studies used different evaluation criteria for assessing the effectiveness of pre-operative exercise. Therefore, these studies showed mixed results on the effectiveness of pre-operative exercise on THA patients. Table 4-1 shows a summary of these studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample size (intervention group and control group)</th>
<th>Intervention type</th>
<th>Intervention duration</th>
<th>Time of evaluation</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferrara et al., 2008</td>
<td>11 and 12</td>
<td>Physiotherapy including exercise, strength and flexibility training</td>
<td>One month</td>
<td>Baseline; the day before surgery; fifteen days, four weeks, and three months after surgery</td>
<td>Intervention group has better hip external rotation and less pain at each evaluation</td>
</tr>
<tr>
<td>Gocen, Sen, Unver, Karatosun, &amp; Gunal, 2004</td>
<td>29 and 30</td>
<td>Straight leg raising, stretching of hamstring and hip flexors, and strengthening of upper extremity exercise</td>
<td>Eight weeks</td>
<td>Baseline; immediately before surgery; at discharge; three months and two years after surgery</td>
<td>Intervention group has less pain and better function at discharge and at three months, but not at two years</td>
</tr>
<tr>
<td>Gilbey et al, 2003</td>
<td>32 and 25</td>
<td>Exercise program that improves lower limb strength and range of hip flexion motion</td>
<td>Eight weeks</td>
<td>Baseline; one week before surgery; three weeks, twelve weeks, and twenty-four weeks after surgery</td>
<td>Intervention group has greater improvements in combined hip strength and range of motion at all the evaluation time points</td>
</tr>
<tr>
<td>Vukomanović, Popović, Đurović, &amp; Krstić, 2008</td>
<td>18 and 18</td>
<td>Exercise from the post-operative rehabilitation program</td>
<td>Not given</td>
<td>On admission; third day after surgery; at discharge; and fifteen months after surgery</td>
<td>No differences between the two groups at discharge. But the intervention group was significantly more independent in performing basic activities</td>
</tr>
<tr>
<td>Study</td>
<td>Sample size (intervention group and control group)</td>
<td>Intervention type</td>
<td>Time of evaluation</td>
<td>Results</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Rooks et al., 2006</td>
<td>25 and 24</td>
<td>Water and land-based exercise involving single-joint movements, strength training, and flexibility exercise</td>
<td>Six weeks; Baseline; within one week before surgery; during hospitalization; eight weeks and twenty-six weeks after surgery</td>
<td>Intervention group has less pain and better leg press strength prior to surgery. But the two groups do not differ in outcomes eight and twenty-six weeks post-operatively</td>
<td></td>
</tr>
<tr>
<td>Hermann et al., 2016</td>
<td>38 and 39</td>
<td>Progressive explosive-type resistance training focusing on hip and knee extension, and leg press</td>
<td>Ten weeks; Baseline; and within one week before surgery</td>
<td>Intervention group has greater improvements in activity of daily living, function, and pain</td>
<td></td>
</tr>
<tr>
<td>Bitterli et al., 2011</td>
<td>30 and 32</td>
<td>Sensorimotor training program to improve strength and mobility</td>
<td>Two to six weeks; The day before surgery; at discharge; four months and twelve months after surgery</td>
<td>Intervention group has greater improvements in pain and quality of life before surgery. But these effects are lost after surgery</td>
<td></td>
</tr>
<tr>
<td>Villadsen, Overgaard, Holsgaard-Larsen, Christensen, &amp; Roos, 2014</td>
<td>43 and 41</td>
<td>Exercise program focusing on lower extremity muscle strength and functional exercise</td>
<td>Eight weeks; Baseline; within one week before surgery; six weeks and three months after surgery</td>
<td>Intervention group has significantly more improvements in activity of daily living and pain until six weeks after surgery. But no difference is found three months post-operatively</td>
<td></td>
</tr>
<tr>
<td>Hoogeboom et al., 2010</td>
<td>10 and 10</td>
<td>Exercise program focusing on lower extremity and functional physical activities</td>
<td>Three to six weeks; Baseline; within one week before surgery; and during each day of hospital stay (post-operatively)</td>
<td>There is no difference both pre-operatively and post-operatively. LOS does not differ between the two groups</td>
<td></td>
</tr>
</tbody>
</table>
### Table: Summary of Studies on Pre-operative Exercise for THA and TKA

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample size (intervention group and control group)</th>
<th>Intervention type</th>
<th>Intervention duration</th>
<th>Time of evaluation</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>McGregor et al., 2004</td>
<td>15 and 20</td>
<td>Exercise for post-operative rehabilitation and physical function for daily living</td>
<td>Two to four weeks</td>
<td>Baseline; within one week before surgery; at discharge; and three months after surgery</td>
<td>Intervention group has higher quality of life at discharge and three months. This group also has a shorter LOS</td>
</tr>
<tr>
<td>Oosting et al., 2012</td>
<td>14 and 12</td>
<td>Exercise program focusing on functional activity and walking capacity</td>
<td>Three to six weeks</td>
<td>Baseline; two to four days before admission; at discharge; and six weeks after discharge</td>
<td>No significant differences are seen in pain, function, and LOS</td>
</tr>
</tbody>
</table>

Several literature reviews and meta-analyses have been conducted on the effectiveness of pre-operative exercise on THA and total knee arthroplasty (TKA) patients. Coudeyre et al. (2007) conducted a systematic literature review on whether pre-operative rehabilitation modifies post-operative outcomes of THA and TKA. This literature review indicated that rehabilitation before THA and TKA surgery contributes to reduced hospital LOS and improves discharge conditions. In addition, it concluded that a pre-operative rehabilitation program, comprising at least physical therapy and education, is recommended before THA and TKA surgery. Occupational therapy could also be proposed before THA. But no quantitative analysis was conducted in this literature review to support these conclusions. Wallis & Taylor (2011) focused on whether pre-operative interventions for hip and knee osteoarthritis provide benefits before and after joint replacement. Their meta-analysis provided low to moderate quality evidence that exercise interventions for hip osteoarthritis reduce pain and improve activity prior to THA surgery. It also indicated that pre-operative exercise with education program may improve function at three weeks after THA surgery, but hospital LOS is not reduced. Hoogeboom et al. (2012) studied the association between pre-operative therapeutic exercise and its effectiveness in terms of post-operative functional recovery for THA and TKA patients. Based on the results of literature review and
meta-analysis, it indicated that therapeutic exercise is not associated with functional recovery during the hospital stay and within three months of surgery. Thus, the authors concluded that pre-operative therapeutic exercise for total joint arthroplasty does not demonstrate beneficial effects on post-operative functional recovery. Gill & McBurney (2013) conducted a meta-analysis on the pre-operative effects of exercise-based interventions on pain and physical function for patients awaiting THA and TKA surgery. The results indicated that exercise has a medium-sized effect on both pain and function at the pre-operative stage. Thus, the authors concluded that patients awaiting THA surgery can benefit from pre-operative exercise programs.

Wang et al. (2016) used meta-analysis to assess the clinical impact of pre-habilitation before hip and knee surgery on post-operative outcomes. The results indicated that pre-habilitation slightly reduces pain within four weeks post-operatively, but no difference was found beyond four weeks. Pre-habilitation slightly improved function at six to eight weeks and at twelve weeks, but it had no impact on LOS. The effects of pre-habilitation on pain and function were too small to be considered clinically important and were not robust over time.

Although these studies provide important insights into the effects of pre-operative exercise on THA patients, there are certain limitations in these systematic literature review and meta-analysis studies. Firstly, some of these meta-analyses mixed THA outcomes with TKA outcomes and analyzed the results together. This may negatively influence the validity of the results, because pre-operative exercise may have different effects on THA patients and TKA patients. Secondly, some of these studies focused exclusively on the outcomes at the pre-operative stage, while others only analyzed the outcomes at the post-operative stage. This may prevent readers from obtaining a comprehensive understanding of the effectiveness of pre-operative exercise throughout the perioperative period. Thirdly, some of these meta-analyses included non-randomized controlled studies. Randomized controlled trial is important because it allows investigators to minimize bias by ensuring that most prognostic factors such as age, gender, and weight are similar in the intervention group and control group (Rorabeck et al., 1994). Due to these limitations, the results of these meta-analyses should be understood with caution.
4.3 Research objective

Pre-operative exercise has become increasingly prevalent for patients awaiting THA surgery, and considerable medical resources have been devoted to it in some hospitals. However, so far there has been no convincing results on its effectiveness on THA patients throughout the perioperative care period. In this chapter, we aim at delivering a rigor and comprehensive perspective on the effectiveness of pre-operative exercise on THA patients. Such a perspective will help medical teams better understand the relevant benefits of this intervention. Therefore, our study will provide important insights into what medical interventions should be delivered to THA patients in order to improve care quality and patient outcomes.

In response to the limitations in previous meta-analyses, we conducted a meta-analysis exclusively on pre-operative exercise for THA patients. In addition, we only included randomized controlled studies in our meta-analysis. Finally, we investigated the effectiveness of pre-operative exercise both pre-operatively and post-operatively.

4.4 Methods

In this study, we searched three databases to find relevant literature on pre-operative exercise for THA patients. Then we used the Physiotherapy Evidence Database (PEDro) scale, a rating system for evaluating the validity of trials for clinical decision making, to assess the methodological quality of each included study. Finally, we utilized random effects model to quantify the effects of pre-operative exercise on pain, function, activity of daily living, and LOS of THA patients. We conducted analyses on the same effect measures both pre-operatively (i.e., right after the exercise and before surgery) and post-operatively.

4.4.1 Search methodology and eligibility criteria

We systematically searched three databases, including PubMed, Embase, and Cochrane Library, up to September, 30, 2016. Typical search terms included hip, joint replacement,
arthroplasty, pre-operative (preoperative), physical therapy, exercise, physiotherapy, hydrotherapy, and rehabilitation. Titles and abstracts identified by the literature searches were screened, using pre-determined eligibility criteria to determine whether the study should be included in our review. In addition, reference lists of the studies included in our review were screened.

Studies were eligible for being included into our review and meta-analysis only if all of the following criteria are satisfied:

1) The study has to be randomized controlled trials.
2) The study has to be designed for patients awaiting THA surgery. The study may also include patients awaiting TKA surgery, but these two patient groups should be listed separately in their demographics and results.
3) The study has to compare pre-operative exercise or physiotherapy program (i.e., prescribed exercise or physiotherapy without co-interventions such as nutritional counselling, acupuncture, transcutaneous electrical nerve stimulation, etc) with no formal pre-operative rehabilitation program. Studies that only include education on pre-operative exercise but without formal exercise were excluded.
4) The study has to report at least one THA-relevant outcome (pain, function, activity of daily living, etc) at both the pre-operative stage and the post-operative stage.
5) The study needs to compare the intervention group with the control group on the THA-relevant outcomes.

4.4.2 PEDro scale

PEDro is a database of randomized trials and systematic reviews in physiotherapy (physical therapy). The database, developed and maintained by the Centre for Evidence-Based Physiotherapy in Sydney, Australia, is freely available on the web. PEDro was developed to give physiotherapists (physical therapists) and consumers of physiotherapy services rapid access to the best available evidence of the effects of physiotherapy interventions (Sherrington, Herbert, Maher, & Moseley, 2000).
Each clinical trial in the PEDro database is rated on the basis of its methodological quality using the PEDro scale. The PEDro scale is based on the Delphi list developed by Verhagen and colleagues at the Department of Epidemiology, University of Maastricht, Netherlands (Verhagen et al., 1998). The PEDro scale is designed to help the users of the PEDro database rapidly identify which of the clinical trials indexed on the PEDro database are likely to be internally valid, and could have sufficient statistical information to make their results interpretable. Several studies indicated that PEDro scale is a rating system with high reliability and validity (Maher, Sherrington, Herbert, Moseley, & Elkins, 2003; de Morton, 2009). The PEDro scale has 11 items, and they are explained in detail below (PEDro Physiotherapy Evidence Database, 2016).

1) Eligibility criteria were specified

This criterion is satisfied if the report describes the source of subjects and a list of criteria used to determine who was eligible to participate in the study.

This criterion influences external validity, but not the internal or statistical validity of the trial. It has been included in the PEDro scale so that all items of the Delphi scale are represented on the PEDro scale. But this item is not used to calculate the PEDro score.

2) Subjects were randomly allocated to groups

A study is considered to have used random allocation if the report states that allocation was random. The precise method of randomization need not be specified. Procedures such as coin-tossing and dice-rolling should be considered random. Quasi-randomized allocation procedures such as allocation by hospital record number or birth date, or alternation, do not satisfy this criterion.

3) Allocation was concealed

Concealed allocation means that the person who determined if a subject was eligible for inclusion in the trial was unaware, when this decision was made, of which group the subject would be allocated to. A point is awarded for this criterion, even if it is not stated that allocation was concealed, when the report states that allocation was by sealed opaque envelopes or that allocation involved contacting the holder of the allocation schedule who was “off-site”.

72
4) The groups were similar at baseline regarding the most important prognostic indicators.

At a minimum, in studies of therapeutic interventions, the report must describe at least one measure of the severity of the condition being treated and at least one different key outcome measure at baseline. The rater must be satisfied that the groups’ outcomes would not be expected to differ, on the basis of baseline differences in prognostic variables alone, by a clinically significant amount. This criterion is satisfied even if only baseline data of study completers are presented.

5) There was blinding of all subjects

Blinding means the person in question (subject, therapist or assessor) did not know which group the subject had been allocated to. In addition, subjects and therapists are only considered to be “blind” if it could be expected that they would have been unable to distinguish between the treatments applied to different groups. In trials in which key outcomes are self-reported, the assessor is considered to be blind if the subject was blind.

Blinding of subjects involves ensuring that subjects were unable to discriminate whether they had or had not received the treatment. When subjects have been blinded, the reader can be sure that the treatment effect was not due to placebo effects or Hawthorne effects.

6) There was blinding of all therapists who administered the therapy

Blinding of therapists involves ensuring that therapists were unable to discriminate whether individual subjects had or had not received the treatment. When therapists have been blinded, the reader can be sure that the apparent effect (or lack of effect) of treatment was not due to the therapists’ enthusiasm or lack of enthusiasm for the treatment or control conditions.

7) There was blinding of all assessors who measured at least one key outcome

Blinding of assessors involves ensuring that assessors were unable to discriminate whether individual subjects had or had not received the treatment. When assessors have been blinded, the reader can be sure that the apparent effect (or lack of effect) of treatment was not due to the assessors’ biases impinging on their measures of outcomes.
8) Measures of at least one key outcome were obtained from more than 85% of the subjects initially allocated to groups.

This criterion is only satisfied if the report explicitly states both the number of subjects initially allocated to groups and the number of subjects from whom key outcome measures were obtained. In trials in which outcomes are measured at several points in time, a key outcome must have been measured in more than 85% of subjects at one of these points in time.

9) All subjects for whom outcome measures were available received the treatment or control condition as allocated or, where this was not the case, data for at least one key outcome were analyzed by “intention to treat”.

An intention to treat analysis means that even if subjects did not receive the treatment condition (or the control condition) as allocated, the analysis was performed as if subjects received the treatment condition (or the control condition) they were allocated to. The criterion is satisfied, even if there is no mention of analysis by intention to treat, if the report explicitly states that all subjects received treatment or control condition as allocated.

10) The results of between-group statistical comparisons are reported for at least one key outcome.

A between-group statistical comparison involves statistical comparison of one group with another. Depending on the design of the study, this may involve comparison of two or more treatments, or comparison of treatment with a control condition. The analysis may be a simple comparison of outcomes measured after the treatment was administered, or a comparison of the change in one group with the change in another. The comparison may be in the form of hypothesis testing or in the form of an estimate and its confidence interval.

11) The study provides both point measures and measures of variability for at least one key outcome.

A point measure is a measure of the size of the treatment effect. The treatment effect may be described as a difference in group outcomes, or as the outcome in all groups. Measures of variability include standard deviations, standard errors, confidence intervals, inter-quartile
ranges, and ranges. Point measures and measures of variability may be provided graphically as long as it is clear what is being graphed. Where outcomes are categorical, this criterion is considered to have been met if the number of subjects in each category is given for each group.

In PEDro scale, a point is only awarded when a criterion is clearly satisfied. If on a literal reading of the trial report it is possible that a criterion is not satisfied, a point should not be awarded for that criterion. Since criterion 1 measures external validity, the point for criterion 1 does not count into the total PEDro score. Therefore, the highest possible PEDro score is 10 points.

4.4.3 Meta-analysis and quantifying heterogeneity

We used the random-effects meta-analysis model proposed by Higgins and Thompson to quantify the heterogeneity in meta-analysis (Higgins & Thompson, 2002). This method is the most commonly-used method in meta-analysis, and it is independent of the number of studies and the treatment effect metric.

Let \( y_i \) denote an estimate of parameter \( \theta_i \) from study \( i \) \((i = 1, \ldots, k)\), and its precision, defined as the reciprocal of the estimate’s variance, is \( w_i \). We make the conventional assumption that the precisions are known, although in reality these are estimated from the data in each study. In a traditional fixed-effects meta-analysis, \( \theta_i \)'s are assumed identical and a summary estimate, \( \hat{\mu}_F \), is calculated as a weighted average of the study estimates, using the precisions as weights:

\[
\hat{\mu}_F = \frac{\sum w_i y_i / \sum w_i}{75}
\]

The variance of \( \hat{\mu}_F \) under the fixed-effects assumption is:

\[
v_F = 1 / \sum w_i
\]

A basic random-effects meta-analysis may be achieved by incorporating an estimate of the between-study heterogeneity, \( \tau^2 \), into the weights to produce a summary estimate:

\[
\hat{\mu}_R = \frac{\sum w_i^* y_i / \sum w_i^*}{75}
\]
where \( w_i^* = (w_i^{-1} + \tilde{\tau}^2)^{-1} \). An approximate variance of \( \hat{\mu}_R \) under the random-effects assumption is:

\[
v_R = 1 / \sum w_i^* \tag{4-4}\]

A test of homogeneity of the \( \theta_i \)'s from a given set of studies is provided by referring the statistic:

\[
Q = \sum w_i (y_i - \hat{\mu}_F)^2 \tag{4-5}
\]

to a \( \chi^2 \) distribution with \( k - 1 \) degrees of freedom. A moment-based estimate of \( \tau^2 \) may be obtained by equating the observed value of \( Q \) with its expectation:

\[
E[Q] = \tau^2 \left( \sum w_i - \frac{\sum w_i^*}{\sum w_i} \right) + k - 1 \tag{4-6}
\]

yielding

\[
\hat{\tau}^2 = \frac{Q - (k - 1)}{\sum w_i^{-1} \frac{\sum w_i^*}{\sum w_i}} \tag{4-7}
\]

By convention this is replaced with zero if \( Q < k - 1 \). To quantify the heterogeneity in a meta-analysis, we also need to quantify a number, \( \hat{\sigma}^2 \), that describes the typical within-study variance. \( \hat{\sigma}^2 \) is estimated by \( s^2 \), which is:

\[
\hat{\sigma}^2 = s^2 = \frac{\sum w_i(k - 1)}{(Q, w_i)^2 - \sum w_i^*} \tag{4-8}
\]

Therefore, the statistic used to quantify heterogeneity in a meta-analysis is:

\[
I^2 = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{\sigma}^2} \tag{4-9}
\]

The interpretation of \( I^2 \) is intuitive – the percentage of total variation across studies due to between-study heterogeneity. It is simple to calculate and can usually be derived from published studies. In addition, it does not inherently depend on the number of studies in the meta-analysis. Finally, it may be interpreted similarly irrespective of the type of outcome data (e.g.,
dichotomous, quantitative, or time to event) and choice of effect measure (e.g., treatment effect, odds ratio, or hazard ratio). The value of $I^2$ is always between 0% and 100%. A naive categorization of values for $I^2$ would not be appropriate for all circumstance, although we would tentatively assign adjectives of low, moderate, and high to $I^2$ values of 25%, 50%, and 75%, respectively (Higgins, Thompson, Deeks, & Altman, 2003).

4.5 Effect measures of pre-operative exercise

In this study, we used the Western Ontario and McMaster Universities Arthritis Index (WOMAC), Harris Hip Score, as well as Hip disability and Osteoarthritis Outcome Score (HOOS) as effect measures of pre-operative exercise. These three are all highly valid and reliable tools, and they are frequently used to study the clinical outcomes, such as pain and function, of hip arthritis patients (Söderman & Malchau, 2001; Nilsdotter, Lohmander, Klässbo, & Roos, 2003).

4.5.1 WOMAC

The WOMAC is a widely used, proprietary set of standardized questionnaires used by health professionals to evaluate the condition of patients with osteoarthritis of the knee or hip, including pain, stiffness, and physical functioning of the joints. It can be self-administered and was developed at Western Ontario and McMaster Universities in Canada in 1982 (WOMAC Osteoarthritis Index, 2016).

The WOMAC measures five items for pain (score range 0-20), two for stiffness (score range 0-8), and seventeen for functional limitation (score range 0-68) (Quintana et al., 2006). Smaller scores indicate better health status. Physical functioning questions cover everyday activities such as stair use, standing up from a sitting or lying position, standing, bending, walking, getting in and out of a car, shopping, putting on or taking off socks, lying in bed, getting in or out of a bath, sitting, as well as heavy and light household duties. The WOMAC has been extensively used in both observational studies and studies on examining changes following treatments of
pharmacology, arthroplasty, exercise, and physical therapy (American College of Rheumatology, 2016).

The WOMAC is among the most widely-used assessments in arthritis research, and it has been translated into more than 65 languages (WOMAC Osteoarthritis Index, 2016).

4.5.2 Harris Hip Score

The Harris Hip Score is a clinician-based outcome tool that is frequently used for the evaluation of patients following THA surgery. It was developed in 1969 and since then has undergone multiple revisions. Today, the scope of the Harris Hip Score has increased and many clinicians now use it for the assessment of femoral neck fractures and osteoarthritis (Nilsdotter & Bremander, 2011).

The Harris Hip Score is composed of four sub-scales: pain (44 points), function (47 points), absence of deformity (4 points), and range of motion (5 points). The pain domain measures pain severity, its effect on activities, and the need for pain medication. The function domain consists of daily activities and gait, while the deformity sub-scale takes into account hip flexion, adduction, internal rotation, and leg length discrepancy.

The survey has 10 question items and the score ranges from 0-100, with a higher score representing less dysfunction and better outcomes.

4.5.3 HOOS

The HOOS is a patient-reported outcome tool that is used to evaluate hip arthritis patients. The HOOS questionnaire was built upon the WOMAC index. Questions from the WOMAC were used as the basis for the HOOS survey, and two additional dimensions were added: sport and recreation, as well as hip-related quality of life.
The HOOS is composed of 40 questions and attempts to assess patient outcomes by looking at five sub-scales: pain (10 items), symptom (5 items), activity of daily living (17 items), sport and recreation (4 items), and hip-related quality of life (4 items). A HOOS score is calculated by using a simple formula to produce a score that ranges from 0-100 for each sub-scale, with a higher score representing better status. The HOOS survey is easy to administer and relatively easy to score. In addition, the HOOS survey has been found to be more responsive than the WOMAC survey (Nilsdotter et al., 2003).

4.6 Results

4.6.1 Methodological quality of included studies

The abstracts of studies retrieved from PubMed, Embase, and Cochrane Library were reviewed. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of study selection process is shown in Figure 4-1. Finally, 11 studies met all the five inclusion criteria.
Figure 4-1. The PRISMA diagram of study selection

The methodological quality of each of these 11 studies was rated by the PEDro scale. The results of rating are shown in Table 4-2.
Table 4-2. Methodological quality of the 11 included studies

<table>
<thead>
<tr>
<th>Study</th>
<th>PEDro criteria*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Ferrara et al., 2008</td>
<td>1</td>
</tr>
<tr>
<td>Gocen et al., 2004</td>
<td>1</td>
</tr>
<tr>
<td>Gilbey et al., 2003</td>
<td>1</td>
</tr>
<tr>
<td>Vukomanović et al., 2008</td>
<td>1</td>
</tr>
<tr>
<td>Rooks et al., 2006</td>
<td>1</td>
</tr>
<tr>
<td>Hermann et al., 2016</td>
<td>1</td>
</tr>
<tr>
<td>Bitterli et al., 2011</td>
<td>1</td>
</tr>
<tr>
<td>Villadsen et al., 2014</td>
<td>1</td>
</tr>
<tr>
<td>Hoogeboom et al., 2010</td>
<td>1</td>
</tr>
<tr>
<td>McGregor et al., 2004</td>
<td>1</td>
</tr>
<tr>
<td>Oosting et al., 2012</td>
<td>1</td>
</tr>
</tbody>
</table>


Item scoring: 1 = present, 0 = absent; criterion 1 is not included in the total score.

Among the 11 included studies, 8 of them are considered as high-quality trials (PEDro score ≥ 6). The other 3 trials receive a PEDro score of 5. The average PEDro score of these 11 studies is 6.36. For the pre-operative exercise intervention, it is impossible to blind the participant or the therapist from the intervention. Thus, all these studies score 0 on these two criteria (i.e., criteria 5 and criteria 6), reducing the maximum possible PEDro score to 8. All the 11 studies satisfy the criteria of random allocation, baseline similarity between groups, between-group statistical comparison, as well as reporting point measures and variability measures. 4 studies used concealed allocation, and 5 studies blinded the assessors in evaluation. 9 studies had a follow-up rate greater than 85% on at least one key outcome measures. For the details of these 11 studies, please refer to Table 4-1.
We conducted meta-analysis on WOMAC scores (pain and function), Harris Hip Score, HOOS scores (activity of daily living, pain, symptom, sport and recreation, as well as hip-related quality of life), and LOS for the exercise group and the non-exercise group of THA patients.

### 4.6.2 Meta-analysis on WOMAC

In order to conduct an effective meta-analysis, the same THA-relevant outcome should be measured at the same time (or at very close time) across different studies. Otherwise, the outcomes are not comparable and bias may arise. Among the 11 included studies, 5 studies reported WOMAC score of the intervention group and the control group (Ferrara et al., 2008; Gilbey et al., 2003; Rooks et al., 2006; Bitterli et al., 2011; McGregor et al., 2004). But one study did not report the standard deviation of WOMAC (Bitterli et al., 2011), and one study did not report any sub-scales of WOMAC (Gilbey et al., 2003). Therefore, 3 studies can be used in the meta-analysis on the sub-scales of WOMAC. The details of WOMAC and time of measurement in these 3 studies are shown in Table 4-3.

<table>
<thead>
<tr>
<th>Study</th>
<th>Included sub-scales of WOMAC</th>
<th>Time of measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferrara et al., 2008</td>
<td>Pain, function, and stiffness</td>
<td>Baseline; the day before surgery; fifteen days, four weeks, and three months after surgery</td>
</tr>
<tr>
<td>Rooks et al., 2006</td>
<td>Pain and function</td>
<td>Baseline; within one week before surgery; during hospitalization; eight weeks and twenty-six weeks after surgery</td>
</tr>
<tr>
<td>McGregor et al., 2004</td>
<td>Pain, function, and stiffness</td>
<td>Baseline; within one week before surgery; at discharge; and three months after surgery</td>
</tr>
</tbody>
</table>

From Table 4-3, it is clear that we can conduct meta-analysis on the pain and function sub-scales of WOMAC, because Rooks et al. (2006) does not provide stiffness sub-scale. The common time of measurements across these three studies are baseline, the day before surgery (or within one week before surgery), and three months after surgery (or eight weeks after surgery).
We obtained the corresponding WOMAC pain scores at these three time points from these studies, and they are listed in Table 4-4.

<table>
<thead>
<tr>
<th>Study</th>
<th>Group</th>
<th>Baseline Mean</th>
<th>Standard deviation</th>
<th>Before surgery Mean</th>
<th>Standard deviation</th>
<th>Three months after surgery Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferrara et al., 2008</td>
<td>Intervention</td>
<td>10.55</td>
<td>3.88</td>
<td>8.00</td>
<td>3.80</td>
<td>1.70</td>
<td>2.35</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>11.17</td>
<td>3.46</td>
<td>11.00</td>
<td>3.60</td>
<td>2.20</td>
<td>1.75</td>
</tr>
<tr>
<td>Rooks et al., 2006</td>
<td>Intervention</td>
<td>8.00</td>
<td>3.70</td>
<td>7.80</td>
<td>4.10</td>
<td>2.60</td>
<td>2.60</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>8.80</td>
<td>3.20</td>
<td>9.90</td>
<td>2.90</td>
<td>2.70</td>
<td>2.00</td>
</tr>
<tr>
<td>McGregor et al., 2004</td>
<td>Intervention</td>
<td>11.20</td>
<td>2.20</td>
<td>10.20</td>
<td>2.70</td>
<td>2.70</td>
<td>2.10</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>11.40</td>
<td>4.10</td>
<td>10.30</td>
<td>4.10</td>
<td>0.05</td>
<td>4.40</td>
</tr>
</tbody>
</table>

Using the information in Table 4-4 and the sample sizes (i.e., the number of patients in the intervention group and the control group shown in Table 4-1) of these three studies, we utilized random-effects model to calculate $I^2$ in our meta-analysis. In addition, we constructed forest plots to judge if there is any heterogeneity in the effects of pre-operative exercise across these three studies. All the analyses were conducted in statistical software R 3.1.0 using packages “metacont” and “forest” (R Development Core Team, 2008). Table 4-2 indicates that baseline similarities exist in these studies, thus we just need to analyze the situation before surgery and at three months after surgery. The forest plot of the WOMAC pain score before surgery across these three studies is shown in Figure 4-2.

The mean difference in Figure 4-2 is the weighted mean difference in WOMAC pain score between the intervention group and the control group. Since smaller WOMAC score indicates
better health status, a negative weighted mean difference shows better outcome in the intervention group than in the control group. The random-effects model shows that the difference in WOMAC pain score between the two groups is marginally significant (3 studies, 107 patients, weighted mean difference -1.6, 95% confidence interval -3.20 to 0.00, P-value 0.0504). The $I^2$ value is 27.5%, indicating that there is only low heterogeneity across the 3 studies. Since baseline similarities are shown in Table 4-2, the results indicate that pre-operative exercise is effective in reducing WOMAC hip pain in the intervention group before surgery (i.e., right after the exercise program), compared with the control group without any exercise.

After that, we need to see whether the difference in WOMAC pain score is still significant at the post-operative stage. The forest plot of WOMAC pain score at three months after surgery is shown in Figure 4-3.

![Figure 4-3. WOMAC pain score at three months after surgery](image)

Figure 4-3 indicates that the weighted mean difference is not statistically significant (3 studies, 107 patients, weighted mean difference 0.51, 95% confidence interval -1.13 to 2.15, P-value 0.5442). However, the $I^2$ value is 64.2%, indicating that the heterogeneity among these three studies is more than moderate. Relatively high heterogeneity degrades the validity of results in meta-analysis. In such a situation, a reasonable approach is to conduct a jackknife sensitivity analysis, in which we sequentially remove each study to detect the individual study responsible for the heterogeneity. For outcomes in which the removal of a single study result in a complete resolution of heterogeneity, the overall estimate after removal of the outlier is presented (Barochia et al., 2010).
In Figure 4-3, it is apparent that the third study (McGregor et al., 2004) mainly causes the heterogeneity in weighted mean difference. Thus, we removed this study and conducted the meta-analysis again. The result is shown in Figure 4-4.

After removing the third study, there is no heterogeneity between the two remaining studies. The weighted mean difference is less than 0, but it is not statistically significant (2 studies, 72 patients, weighted mean difference -0.25, 95% confidence interval -1.28 to 0.79, P-value 0.6398). The results indicate that based on the included studies, the impact of pre-operative exercise on WOMAC pain is no longer significant at three months post-operatively. Therefore, this meta-analysis shows that with respect to WOMAC pain scores, the effect of pre-operative exercise is only significant before surgery, namely right after the exercise program.

Similarly, we conducted meta-analysis of pre-operative exercise on WOMAC function scores, both before surgery and at three months after surgery, using the three studies in Table 4-3. The corresponding WOMAC function scores are shown in Table 4-5.

Table 4-5. Mean and standard deviation of WOMAC function scores in these 3 studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Group</th>
<th>Baseline</th>
<th>Before surgery</th>
<th>Three months after surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Standard deviation</td>
<td>Mean</td>
</tr>
<tr>
<td>Ferrara et al., 2008</td>
<td>Intervention</td>
<td>39.64</td>
<td>13.69</td>
<td>33.70</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>44.92</td>
<td>6.00</td>
<td>43.50</td>
</tr>
<tr>
<td>Rooks et al., 2006</td>
<td>Intervention</td>
<td>29.10</td>
<td>12.90</td>
<td>26.90</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>29.80</td>
<td>11.20</td>
<td>33.70</td>
</tr>
<tr>
<td>McGregor et al., 2004</td>
<td>Intervention</td>
<td>39.90</td>
<td>11.20</td>
<td>35.80</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>41.70</td>
<td>11.00</td>
<td>41.00</td>
</tr>
</tbody>
</table>
The forest plot of the WOMAC function score before surgery across these three studies is shown in Figure 4-5.

The difference in WOMAC function score between the two groups is statistically significant (3 studies, 107 patients, weighted mean difference -6.86, 95% confidence interval -11.21 to -2.50, P-value 0.0020). The $I^2$ value is 0%, showing very good homogeneity across the three studies. The results indicate that pre-operative exercise is effective in improving WOMAC hip function in the intervention group before surgery. This group shows better functional status before surgery than the control group.

Then we need to judge whether the difference in WOMAC function score is still significant at the post-operative stage. The forest plot of WOMAC function score at three months after surgery is shown in Figure 4-6.

Figure 4-6 indicates that the weighted mean difference between the two groups is not statistically significant (3 studies, 107 patients, weighted mean difference -3.21, 95% confidence
interval -8.73 to 2.31, \( P\)-value 0.2543). The \( I^2 \) value is 44.9\%, which is less than moderate heterogeneity. The results on WOMAC function scores indicate that based on the three included studies, the pre-operative benefit of exercise program is non-existent at three months post-operatively.

Therefore, the results on WOMAC scores indicate that pre-operative exercise offers significant benefits to the intervention group before surgery with respect to both WOMAC pain and WOMAC function. But these benefits become insignificant at three months post-operatively.

4.6.3 Meta-analysis on Harris Hip Score

Among the 11 included studies, 4 studies reported Harris Hip Score of the intervention group and the control group (Ferrara et al., 2008; Gocen et al., 2004; Vukomanović et al., 2008; McGregor et al., 2004). The details of Harris Hip Score and time of measurement in these 4 studies are shown in Table 4-6.

<table>
<thead>
<tr>
<th>Study</th>
<th>Included sub-scales of Harris Hip Score</th>
<th>Time of measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferrara et al., 2008</td>
<td>None</td>
<td>Baseline; the day before surgery; fifteen days, four weeks, and three months after surgery</td>
</tr>
<tr>
<td>Gocen et al., 2004</td>
<td>None</td>
<td>Immediately before surgery; at discharge; three months and two years after surgery</td>
</tr>
<tr>
<td>Vukomanović et al., 2008</td>
<td>None</td>
<td>On admission; third day after surgery; at discharge; and fifteen months after surgery</td>
</tr>
<tr>
<td>McGregor et al., 2004</td>
<td>None</td>
<td>Baseline; within one week before surgery; at discharge; and three months after surgery</td>
</tr>
</tbody>
</table>

Since none of these 4 studies provided any sub-scales of Harris Hip Score, we performed the meta-analysis on the whole Harris Hip Score. The only common time of measurement across all these 4 studies is the day before surgery (or within one week before surgery). However, three studies provided measurements at discharge (Gocen et al., 2004; Vukomanović et al., 2008; McGregor et al., 2004) and at three months post-operatively (Ferrara et al., 2008; Gocen et al.,
We obtained the corresponding Harris Hip Scores at these three time points from these studies, and they are listed in Table 4-7.

<table>
<thead>
<tr>
<th>Study</th>
<th>Group</th>
<th>Before surgery</th>
<th>At discharge</th>
<th>Three months after surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Standard deviation</td>
<td>Mean</td>
</tr>
<tr>
<td>Ferrara et al., 2008</td>
<td>Intervention</td>
<td>43.60</td>
<td>15.70</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>34.90</td>
<td>15.50</td>
<td>N/A</td>
</tr>
<tr>
<td>Gocen et al., 2004</td>
<td>Intervention</td>
<td>51.48</td>
<td>18.32</td>
<td>64.46</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>45.30</td>
<td>12.98</td>
<td>59.36</td>
</tr>
<tr>
<td>Vukomanović et al., 2008</td>
<td>Intervention</td>
<td>44.00</td>
<td>7.25</td>
<td>51.25</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>45.75</td>
<td>11.82</td>
<td>50.10</td>
</tr>
<tr>
<td>McGregor et al., 2004</td>
<td>Intervention</td>
<td>45.40</td>
<td>11.50</td>
<td>62.60</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>43.20</td>
<td>16.20</td>
<td>53.80</td>
</tr>
</tbody>
</table>

The forest plot of Harris Hip Score before surgery across these 4 studies is shown in Figure 4-7.

Since higher Harris Hip Score indicates better pain and function status, a positive weighted mean difference favors the intervention group over the control group. Although the weighted mean difference is positive, it is not statistically significant (4 studies, 157 patients, weighted mean difference 2.37, 95% confidence interval -2.17 to 6.90, P-value 0.3060). The $I^2$ value is 15.2%, showing quite small heterogeneity across the 4 studies. The results indicate that pre-operative exercise offers no advantage in terms of Harris Hip Score before surgery.
Although the weighted mean difference in Harris Hip Score is not statistically significant right before surgery, we still need to judge whether the difference is significant at the post-operative stage. The forest plot of Harris Hip Score at discharge is shown in Figure 4-8.

Figure 4-8 indicates that the weighted mean difference is statistically significant (3 studies, 134 patients, weighted mean difference 4.54, 95% confidence interval 0.78 to 8.30, P-value 0.0179). The $I^2$ value is 48.4%, which is less than moderate heterogeneity. However, it should be noted that one study (Ferrara et al., 2008) did not provide Harris Hip Score at discharge.

The forest plot of Harris Hip Score at three months post-operatively is shown in Figure 4-9.

Figure 4-9 indicates that the weighted mean difference is statistically significant (3 studies, 117 patients, weighted mean difference 5.91, 95% confidence interval 1.68 to 10.14, P-value 0.0062). The $I^2$ value is 0%, indicating very good homogeneity among these three studies. However, it should be noted that one study (Vukomanović et al., 2008) did not provide Harris Hip Score at three months post-operatively.
Therefore, the included studies indicate that pre-operative exercise may offer benefits to THA patients at discharge and at three months post-operatively, provided that their health status is measured by Harris Hip Score. But the pre-operative benefit is not statistically significant in the 4 included studies.

4.6.4 Meta-analysis on HOOS

Among the 11 included studies, 4 studies reported HOOS scores of the intervention group and the control group (Hermann et al., 2016; Villadsen et al., 2014; Hoogeboom et al., 2010; Oosting et al., 2012). The details of HOOS and time of measurement in these 4 studies are shown in Table 4-8.

Table 4-8. The details of HOOS in the four included studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Included sub-scales of HOOS</th>
<th>Time of measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hermann et al., 2016</td>
<td>Activity of daily living, pain, symptoms, sport &amp; recreation, hip-related quality of life</td>
<td>Baseline; within one week before surgery</td>
</tr>
<tr>
<td>Villadsen et al., 2014</td>
<td>Activity of daily living, pain, symptoms, sport &amp; recreation, hip-related quality of life</td>
<td>Baseline; within one week before surgery; six weeks and three months after surgery</td>
</tr>
<tr>
<td>Hoogeboom et al., 2010</td>
<td>Activity of daily living, pain, symptoms, sport &amp; recreation, hip-related quality of life</td>
<td>Baseline; within one week before surgery</td>
</tr>
<tr>
<td>Oosting et al., 2012</td>
<td>Activity of daily living, pain, symptoms, sport &amp; recreation, hip-related quality of life</td>
<td>Baseline; two to four days before admission; and six weeks after discharge</td>
</tr>
</tbody>
</table>

All the 4 studies provided all the 5 sub-scales of HOOS: activity of daily living, pain, symptoms, sport and recreation, as well as hip-related quality of life. The common time of measurement across all these 4 studies are baseline and within one week before surgery (or two to four days before admission). However, two studies provided measurements at six weeks post-operatively (Villadsen et al., 2014; Oosting et al., 2012). We obtained the corresponding HOOS activity of daily living scores at these three time points from these studies, and they are listed in Table 4-9.
Table 4-9. Mean and standard deviation of HOOS activity of daily living scores in these 4 studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Group</th>
<th>Baseline</th>
<th>Before surgery</th>
<th>Six weeks after surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Standard deviation</td>
<td>Mean</td>
</tr>
<tr>
<td>Hermann et al., 2016</td>
<td>Intervention</td>
<td>49.20</td>
<td>12.50</td>
<td>59.90</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>48.10</td>
<td>13.80</td>
<td>48.70</td>
</tr>
<tr>
<td>Villadsen et al., 2014</td>
<td>Intervention</td>
<td>48.00</td>
<td>15.00</td>
<td>54.20</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>47.30</td>
<td>19.40</td>
<td>46.10</td>
</tr>
<tr>
<td>Hoogeboom et al., 2010</td>
<td>Intervention</td>
<td>51.50</td>
<td>16.70</td>
<td>51.10</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>54.80</td>
<td>19.70</td>
<td>52.30</td>
</tr>
<tr>
<td>Oosting et al., 2012</td>
<td>Intervention</td>
<td>49.60</td>
<td>16.50</td>
<td>49.80</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>49.80</td>
<td>15.60</td>
<td>46.30</td>
</tr>
</tbody>
</table>

The forest plot of HOOS activity of daily living score before surgery across these 4 studies is shown in Figure 4-10.

![Forest plot of HOOS activity of daily living score before surgery across 4 studies](image)

Since higher HOOS score indicates better health status, a positive weighted mean difference favors the intervention group over the control group. Figure 4-10 shows that the weighted mean difference is positive and statistically significant (4 studies, 210 patients, weighted mean difference 7.69, 95% confidence interval 3.62 to 11.85, P-value 0.0003). The $I^2$ value is 3.4%, showing very small heterogeneity across these 4 studies. The results indicate that pre-operative exercise offers significant benefits in activity of daily living before surgery.

Then we need to examine if the benefits of pre-operative exercise program still persist at six weeks post-operatively. The forest plot at this time of measurement is shown in Figure 4-11.
Figure 4-11. HOOS activity of daily living score at six weeks after surgery

Figure 4-11 indicates that the weighted mean difference is positive, but not statistically significant (2 studies, 110 patients, weighted mean difference 3.62, 95% confidence interval -2.30 to 9.54, $P$-value 0.2306). The $I^2$ value is 0%. However, it should be noted that the weighted mean difference and confidence interval are only based on two studies. But the weighted mean difference at six weeks post-operatively is still smaller in magnitude than that before surgery, which is 7.69 in Figure 4-10. This indicates that the benefits of pre-operative exercise are reduced after THA operation.

Similarly, the corresponding HOOS pain scores are shown in Table 4-10.

<table>
<thead>
<tr>
<th>Study</th>
<th>Group</th>
<th>Baseline</th>
<th>Before surgery</th>
<th>Six weeks after surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Standard deviation</td>
<td>Mean</td>
</tr>
<tr>
<td>Hermann et al., 2016</td>
<td>Intervention</td>
<td>48.00</td>
<td>12.70</td>
<td>55.40</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>46.30</td>
<td>14.40</td>
<td>45.90</td>
</tr>
<tr>
<td>Villadsen et al., 2014</td>
<td>Intervention</td>
<td>45.80</td>
<td>13.90</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>45.60</td>
<td>15.60</td>
<td>N/A</td>
</tr>
<tr>
<td>Hoogeboom et al., 2010</td>
<td>Intervention</td>
<td>48.00</td>
<td>16.70</td>
<td>55.30</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>53.00</td>
<td>17.70</td>
<td>49.30</td>
</tr>
<tr>
<td>Oosting et al., 2012</td>
<td>Intervention</td>
<td>45.40</td>
<td>14.70</td>
<td>48.20</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>51.00</td>
<td>11.40</td>
<td>47.20</td>
</tr>
</tbody>
</table>

The forest plot of the HOOS pain score before surgery across the 3 studies (Hermann et al., 2016; Hoogeboom et al., 2010; Oosting et al., 2012) is shown in Figure 4-12.
Figure 4-12. HOOS pain score before surgery

Figure 4-12 shows that the weighted mean difference is positive and statistically significant (3 studies, 126 patients, weighted mean difference 5.96, 95% confidence interval 0.50 to 11.42, \(P\)-value 0.0323). The \(I^2\) value is 15.4%, showing quite small heterogeneity across these 3 studies. The results indicate that pre-operative exercise offers significant benefits in HOOS pain before surgery. This coincides with the results on WOMAC pain score before surgery.

Then we checked whether the pre-operative reduction in pain is still significant after THA surgery. The forest plot of HOOS pain score at six weeks after surgery is shown in Figure 4-13.

Similar to the results of HOOS activity of daily living, Figure 4-13 indicates that the weighted mean difference is positive, but it is not statistically significant (2 studies, 110 patients, weighted mean difference 3.02, 95% confidence interval -2.46 to 8.50, \(P\)-value 0.2803). In addition, the weighted mean difference at six weeks post-operatively is smaller in magnitude than that before surgery (i.e., 5.96 in Figure 4-12), indicating that the effect of pre-operative exercise on pain might have been reduced after surgery.
Then we conducted meta-analysis on HOOS symptoms scores. The HOOS symptoms sub-scale focuses on whether the patient can spread legs apart and stride out. It also asks for the stiffness in the hip joint. The corresponding HOOS symptoms scores in the four included studies are shown in Table 4-11.

Table 4-11. Mean and standard deviation of HOOS symptoms scores in these 4 studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Group</th>
<th>Baseline</th>
<th></th>
<th></th>
<th>Before surgery</th>
<th></th>
<th></th>
<th>Six weeks after surgery</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Hermann et al., 2016</td>
<td>Intervention</td>
<td>44.50</td>
<td>16.40</td>
<td>56.60</td>
<td>19.80</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>43.10</td>
<td>18.50</td>
<td>45.40</td>
<td>16.70</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Villadsen et al., 2014</td>
<td>Intervention</td>
<td>41.40</td>
<td>14.60</td>
<td>N/A</td>
<td>N/A</td>
<td>72.40</td>
<td>16.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40.00</td>
<td>17.40</td>
<td>N/A</td>
<td>N/A</td>
<td>71.40</td>
<td>15.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoogeboom et al., 2010</td>
<td>Intervention</td>
<td>57.50</td>
<td>13.00</td>
<td>56.50</td>
<td>14.20</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>59.50</td>
<td>14.60</td>
<td>59.00</td>
<td>15.60</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oosting et al., 2012</td>
<td>Intervention</td>
<td>51.00</td>
<td>16.60</td>
<td>52.10</td>
<td>12.20</td>
<td>70.10</td>
<td>15.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>55.00</td>
<td>17.30</td>
<td>51.00</td>
<td>18.30</td>
<td>72.30</td>
<td>10.80</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The forest plot of HOOS symptoms score before surgery across the 3 studies is shown in Figure 4-14.

Figure 4-14. HOOS symptoms score before surgery

Figure 4-14 shows that the weighted mean difference is not statistically significant (3 studies, 126 patients, weighted mean difference 4.52, 95% confidence interval -4.22 to 13.25, P-value 0.3109). The $I^2$ value is 48.2%, showing almost moderate heterogeneity across these 3 studies. The results indicate that pre-operative exercise does not significantly improve HOOS symptoms score before surgery.

The forest plot of HOOS symptoms score at six weeks after surgery is shown in Figure 4-15.
Figure 4-15. HOOS symptoms score at six weeks after surgery

Figure 4-15 indicates that the weighted mean difference is close to zero and not statistically significant (2 studies, 110 patients, weighted mean difference 0.03, 95% confidence interval -5.54 to 5.60, P-value 0.9914). Based on the four included studies, pre-operative exercise offers no improvement in HOOS symptoms score in the intervention group, both at the pre-operative stage and at the post-operative stage.

After that, we conducted meta-analysis on HOOS sport and recreation scores. The corresponding HOOS sport and recreation scores in the four included studies are shown in Table 4-12.

Table 4-12. Mean and standard deviation of HOOS sport and recreation scores in these 4 studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Group</th>
<th>Baseline</th>
<th>Before surgery</th>
<th>Six weeks after surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Standard deviation</td>
<td>Mean</td>
</tr>
<tr>
<td>Hermann et al., 2016</td>
<td>Intervention</td>
<td>28.10</td>
<td>15.20</td>
<td>38.50</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>27.80</td>
<td>17.70</td>
<td>28.60</td>
</tr>
<tr>
<td>Villadsen et al., 2014</td>
<td>Intervention</td>
<td>29.00</td>
<td>14.80</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>28.50</td>
<td>16.70</td>
<td>N/A</td>
</tr>
<tr>
<td>Hoogeboom et al., 2010</td>
<td>Intervention</td>
<td>22.50</td>
<td>15.60</td>
<td>25.00</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>27.80</td>
<td>20.80</td>
<td>32.50</td>
</tr>
<tr>
<td>Oosting et al., 2012</td>
<td>Intervention</td>
<td>25.50</td>
<td>13.30</td>
<td>28.20</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>29.10</td>
<td>13.10</td>
<td>31.80</td>
</tr>
</tbody>
</table>

The forest plot of HOOS sport and recreation score before surgery across the 3 studies is shown in Figure 4-16.
Apparently, there is more than moderate heterogeneity in the results ($I^2 = 67.5\%$). The mean difference in the first study (Hermann et al., 2016) deviates from those in the other two studies (Hoogeboom et al., 2010; Oosting et al., 2012). Thus, we conducted a jackknife sensitivity analysis by removing the first study. The forest plot of the remaining two studies is shown in Figure 4-17.

Figure 4-16. HOOS sport and recreation score before surgery

Figure 4-17. HOOS sport and recreation score before surgery after jackknife analysis

Figure 4-17 shows that after jackknife sensitivity analysis, the heterogeneity is totally removed ($I^2 = 0\%$). But the mean difference is still not statistically significant (2 studies, 46 patients, weighted mean difference -5.01, 95% confidence interval -14.26 to 4.24, $P$-value 0.2882). In addition, the mean differences in these two studies are even negative (i.e., -7.50 and -3.60), favoring the control group.

Then we looked at HOOS sport and recreation score at six weeks after surgery. The forest plot of the 2 studies (Villadsen et al., 2014; Oosting et al., 2012) is shown in Figure 4-18.
Figure 4-18. HOOS sport and recreation score at six weeks after surgery

Figure 4-18 indicates that the weighted mean difference is not statistically significant (2 studies, 110 patients, weighted mean difference -1.12, 95% confidence interval -8.26 to 6.01, P-value 0.7578). Based on the four included studies, pre-operative exercise does not provide significant benefits in HOOS sport and recreation score, both before surgery and at six weeks after surgery.

The last sub-scale of HOOS is hip-related quality of life. The corresponding HOOS hip-related quality of life scores in the four included studies are shown in Table 4-13.

Table 4-13. Mean and standard deviation of HOOS hip-related quality of life scores in these 4 studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Group</th>
<th>Baseline Mean</th>
<th>Baseline Standard deviation</th>
<th>Before surgery Mean</th>
<th>Before surgery Standard deviation</th>
<th>Six weeks after surgery Mean</th>
<th>Six weeks after surgery Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hermann et al., 2016</td>
<td>Intervention</td>
<td>32.10</td>
<td>14.40</td>
<td>38.80</td>
<td>17.20</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>29.20</td>
<td>15.60</td>
<td>N/A</td>
<td>60.20</td>
<td>13.60</td>
<td>N/A</td>
</tr>
<tr>
<td>Villadsen et al., 2014</td>
<td>Intervention</td>
<td>31.50</td>
<td>11.20</td>
<td>N/A</td>
<td>N/A</td>
<td>60.20</td>
<td>13.60</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>30.60</td>
<td>15.10</td>
<td>N/A</td>
<td>N/A</td>
<td>54.20</td>
<td>13.90</td>
</tr>
<tr>
<td>Hoogeboom et al., 2010</td>
<td>Intervention</td>
<td>30.70</td>
<td>13.60</td>
<td>36.30</td>
<td>15.80</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>41.00</td>
<td>18.30</td>
<td>43.30</td>
<td>15.40</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Oosting et al., 2012</td>
<td>Intervention</td>
<td>39.00</td>
<td>11.40</td>
<td>38.00</td>
<td>16.10</td>
<td>61.80</td>
<td>16.40</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>31.90</td>
<td>9.90</td>
<td>34.60</td>
<td>11.10</td>
<td>65.80</td>
<td>16.10</td>
</tr>
</tbody>
</table>

The forest plot of HOOS hip-related quality of life score before surgery across the 3 studies is shown in Figure 4-19.
Figure 4-19. HOOS hip-related quality of life score before surgery

Figure 4-19 shows that the weighted mean difference is not statistically significant (3 studies, 126 patients, weighted mean difference 3.23, 95% confidence interval -3.72 to 10.18, P-value 0.3625). The $I^2$ value is 43.8%, which is less than moderate heterogeneity.

The forest plot of HOOS hip-related quality of life score at six weeks after surgery is shown in Figure 4-20.

Figure 4-20. HOOS hip-related quality of life score at six weeks after surgery

Figure 4-20 indicates that there is moderate heterogeneity in the results ($I^2 = 50.2\%$). However, since only two studies provided HOOS hip-related quality of life score at six weeks after surgery, we were unable to perform jackknife sensitivity analysis. Figure 4-19 indicates that pre-operative exercise does not significantly improve hip-related quality of life before surgery. Therefore, it is very unlikely that the intervention group achieved significantly better outcomes in hip-related quality of life than the control group at six weeks post-operatively.

To summarize, pre-operative exercise program provides significant benefits to the intervention group before surgery (i.e., right after the exercise) with respect to HOOS activity of daily living...
and HOOS pain. But these benefits become insignificant at six weeks after THA surgery. Moreover, pre-operative exercise does not offer advantage with respect to HOOS symptoms, HOOS sport and recreation, as well as HOOS hip-related quality of life, both pre-operatively and post-operatively.

4.6.5 Meta-analysis on LOS

Studies indicated that LOS is a significant cost driver of THA surgery (Antoniou et al., 2004; Klouche, Sariali, & Mamoudy, 2010). Thus, we analyzed whether pre-operative exercise has any effect on LOS of THA patients. Among the 11 included studies, 4 studies reported LOS of the intervention group and the control group (Vukomanović et al., 2008; Hoogeboom et al., 2010; McGregor et al., 2004; Oosting et al., 2012). The details of LOS in these 4 studies are shown in Table 4-14.

Table 4-14. Mean and standard deviation of LOS in these 4 studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Group</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vukomanović et al., 2008</td>
<td>Intervention</td>
<td>9.8</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>10.2</td>
<td>1.7</td>
</tr>
<tr>
<td>Hoogeboom et al., 2010</td>
<td>Intervention</td>
<td>6.7</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>6.9</td>
<td>2.2</td>
</tr>
<tr>
<td>McGregor et al., 2004</td>
<td>Intervention</td>
<td>15.3</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>17.8</td>
<td>3.6</td>
</tr>
<tr>
<td>Oosting et al., 2012</td>
<td>Intervention</td>
<td>5.1</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>5.4</td>
<td>2.1</td>
</tr>
</tbody>
</table>

The forest plot of LOS in these 4 studies is shown in Figure 4-21.
Since shorter LOS indicates that a patient achieves discharge criteria earlier, negative weighted mean difference favors the intervention group. Figure 4-21 shows that the weighted mean difference is negative, but not statistically significant (4 studies, 124 patients, weighted mean difference \(-0.58\), 95% confidence interval \([-1.36\) to 0.21], \(P\)-value 0.1516). The \(I^2\) value is 9.8%, indicating very small heterogeneity across these 4 studies.

Figure 4-21 also indicates that the third study (McGregor et al., 2004) has a quite large weighted mean difference (-2.50 days) in LOS between the two patient groups. This study is the earliest among all the 4 studies, and it was published in 2004. In the last two decades, LOS of THA surgery has been showing a decreasing trend (Husted et al., 2010; Vorhies et al., 2011; Cram et al., 2011). This is probably why the other three studies (Vukomanović et al., 2008; Hoogeboom et al., 2010; Oosting et al., 2012) have a shorter LOS than that in McGregor et al. (2004). However, the weighted mean differences in these three studies are all less than 1 day, and they are not statistically significant. Therefore, pre-operative exercise does not have a significant impact on LOS of THA patients. It can slightly reduce LOS at most.

WOMAC score, Harris Hip Score, and HOOS score are highly valid and reliable measures of clinical outcomes of hip arthritis patients (Söderman & Malchau, 2001; Nilsson et al., 2003). Since all these measures are used both pre-operatively and post-operatively to assess the outcomes of the exercise group and the non-exercise group, our meta-analysis delivers a comprehensive picture on the perioperative effectiveness of pre-operative exercise program. To summarize, pre-operative exercise provides moderate benefits to THA patients before surgery with respect to pain, function, and activity of daily living. But such benefits become very low at...
the post-operative stage, especially at more than three months after surgery. In addition, pre-operative exercise has no significant influence on LOS, which is a main cost driver of THA surgery.

4.7 Discussion

Pre-operative exercise program is an increasingly prevalent intervention for patients undergoing THA or TKA surgery. A number of studies have been conducted on investigating whether pre-operative exercise is effective in improving outcomes of THA patients. However, these systematic literature reviews and meta-analyses have certain limitations, such as not quantifying the effects (Coudeyre et al., 2007), mixing hip patients with knee patients (Hoogeboom et al., 2012; Wang et al., 2016), studying the effects only at the pre-operative stage (Gill & McBurney, 2013) or only at the post-operative stage (Wang et al., 2016), and not tracking the same evaluation criteria (WOMAC or HOOS) throughout the perioperative period (Wallis & Taylor, 2011).

In this study, we conducted a meta-analysis on the effects of pre-operative exercise program only on THA patients. In addition, we tracked the effects throughout the perioperative care period using the same evaluation criteria. After a comprehensive literature review, 11 studies using randomized controlled trials were included into our meta-analysis. The results of our meta-analysis indicate that the intervention group has significantly or almost significantly better WOMAC pain and function scores before surgery, compared with those in the control group. But these advantages shrink after surgery and are no longer significant at three months post-operatively. The same situation occurs in HOOS activity of daily living and pain scores. These results suggest that pre-operative benefits might not translate into post-operative benefits. There are no benefits in terms of HOOS symptoms, sport and recreation, as well as hip-related quality of life, both at the pre-operative stage and at the post-operative stage. Moreover, pre-operative exercise does not result in a significant reduction in LOS.

It is interesting to note that the benefits of pre-operative exercise on Harris Hip Score are not significant before surgery, but the benefits become significant at discharge and at three months
after surgery. Moreover, the magnitude of the weighted mean difference increases from 2.37 before surgery to 4.54 at discharge, and then to 5.91 at three months after surgery. This is counter-intuitive, because normally the impacts of THA surgery dampen the benefits produced by pre-operative exercise (Gill & McBurney, 2013). However, the 4 studies on Harris Hip Score did not provide any sub-scale. Thus, we are unable to know whether the improvements over time are due to pain, function, absence of deformity, or range of motion.

It may be argued that the improved outcomes in activity of daily living, pain, and function while waiting for THA surgery are important for patients, because the waiting time can exceed one year in some cases and is typically about three months in the current literature review (Wallis & Taylor, 2011). In addition, this waiting time will probably show an increasing trend in the next 15 years, due to the ever-increasing demand and the insufficient capacity for THA surgery in the United States. However, the value of pre-operative exercise program is reduced if it makes less significant difference post-operatively. In our meta-analysis, pre-operative exercise program renders post-operative weighted mean differences in WOMAC pain, WOMAC function, HOOS activity of daily living, and HOOS pain by 0.25, 3.21, 3.62 and 3.02, respectively. In addition, none of these post-operative weighted mean differences are statistically significant. Relevant studies indicate that the minimal clinically important improvement in WOMAC pain at the post-operative stage is 9.7, and that for WOMAC function is 7.9 (Ehrich et al., 2000; Quintana et al., 2005; Tubach et al., 2005). To the best of our knowledge, there is no published study on the minimal clinically important improvement in HOOS scores for hip arthritis patients. Since HOOS sub-scales are in the range of 0-100, it is very unlikely that mean differences of 3.62 and 3.02 are considered as clinically important at the post-operative stage.

Given the moderate effect of pre-operative exercise at the pre-operative stage and its low benefit at the post-operative stage, the medical team needs to consider whether this intervention should be implemented on THA patients while they are waiting for surgery. If the medical team believes that significant improvements in post-operative outcomes should be emphasized, pre-operative exercise probably should not be incorporated into the interventions intended for THA care quality improvement. Therefore, our meta-analysis provides important insights into what
medical or surgical interventions have proven effectiveness in improving THA care quality and patient outcome, based on the mixed results of existing literature.

There are several limitations in our literature review and meta-analysis. Firstly, the definition of pre-operative exercise was heterogeneous among these 11 included studies, with variations in exercise type, duration, frequency, and intensity. No two studies had exactly the same exercise program. Moreover, sub-level meta-analysis of different exercise programs was not possible, because of the limited number of studies meeting the inclusion criteria. Finally, some of these studies had relatively small sample sizes of both the intervention group and the control group, thus increasing potential bias and reducing validity of the results. In the future, studies on pre-operative exercise program should carry out well-defined exercise and enroll a relatively large number of THA patients.

4.8 Conclusion

We used a rigorous methodology to conduct a comprehensive literature review and meta-analysis on the effect of pre-operative exercise program on THA patients. In addition, we used the same evaluation criteria throughout the perioperative care stage. 11 published studies met our inclusion criteria. The results indicate that while pre-operative exercise offers moderate benefits before surgery with respect to activity of daily living, pain, and function, it provides very low benefits post-operatively on these aspects. It is possible that marked improvements in pain and function that come from replacing the painful hip joint during surgery far outweigh the modest contribution from pre-operative exercise. Moreover, pre-operative exercise has no impact on hip-related quality of life and LOS. By aggregating the 11 studies, this meta-analysis provides medical teams with more convincing evidence and a more complete perspective on the effectiveness of pre-operative exercise program. With such a comprehensive perspective, medical teams are able to better determine whether pre-operative exercise program should be carried out to relieve pain and improve functional status of THA patients, compared with the benefits of other THA-relevant interventions.
By combining the mixed results of relevant studies and conducting a meta-analysis, we are able to obtain insights into what medical or surgical interventions should be provided to the THA patients, based on the proven benefits of each intervention. Therefore, our data-driven approach can be used in the THA care quality improvement framework, with the purpose of managing care quality and improving THA patient outcomes.
Chapter 5 Conclusion

THA plays a very important role in treating patients with progressively worsening severe arthritis in the hip joint. With the aging of population and the prevalence of obesity in the United States, the demand for THA surgery is increasing rapidly. The demand in 2030 is expected to be 572,000, which is twice the number of the cases in 2010. Therefore, a combination of increased economic resources, operative efficiency, technical capacity (i.e., additional surgeons), and implant longevity are needed (Kurtz et al., 2007). Moreover, THA has been shown to significantly improve patients’ quality of life with respect to physical function, pain relief, and overall health (Laupacis et al., 1993; Jones et al., 2001).

However, unplanned readmission after THA has become an increasingly serious problem in the United States. The most conservative estimate is that $170 million was spent in treating THA readmission in 2015 in the United States. This number is expected to grow as the volume of THA surgery increases.

In fiscal year 2015, CMS started to penalize hospitals for high 30-day readmission rate after THA. To compound matters, the penalty cap is expected to increase in the next few years. As to the patients, post-operative complication and readmission significantly reduce their quality of life (Towheed & Hochberg, 1996; Enocson et al., 2009). Since patient outcome is a very important indicator of care quality, unplanned readmission demonstrates a need to resolve many issues in THA patient care (Learmonth et al., 2007).

With the advancement in health information technology, the amount of healthcare data is increasing exponentially. The explosion of available healthcare data and the need for real-time medical decision support necessitate the use of data science and engineering tools as a powerful facilitator of healthcare quality improvement.

One major goal of this research is to use data science and engineering tools to advance quality in THA patient care. We leverage patient-level data records and data from relevant literature, aiming at delivering data-driven insights into improving THA care quality and patient outcomes. More specifically, Chapter 2 is on a care quality monitoring framework, using patient outcome
(i.e., readmission) as the measure. Chapter 3 is on identifying patient subgroups and quantifying the quality of life that each subgroup could obtain after THA surgery. Chapter 4 is on analyzing whether pre-operative exercise intervention improves the quality of life of THA patients throughout the perioperative period. These are all very important components in improving THA care quality and patient outcomes. Therefore, our data-driven solutions have the potential to be used by hospital quality management.

The data science and engineering tools used in this research yield straightforward and applicable results. Thus, our multi-disciplinary approach coupled with the effective use of data analytics methods contribute to facilitating the implementation and translation of data science and engineering in healthcare. Such translation is very important and meaningful in the big data era of healthcare.

To summarize, this research contributes to the following themes:

- Better understanding of the current gaps in THA care quality improvement
- Better identification of potential innovative solutions
- Improved insights into medical decision support
- Improved quality in healthcare

The combined benefits with regard to these themes will provide important guidance on THA care quality improvement, with the purpose of improving patient outcome and reducing unplanned readmission.

The remainder of this chapter is organized as follows. Major contributions of each of the three main chapters are discussed in Section 5.1 through Section 5.3. The limitations of our research are summarized in Section 5.4. Finally, possible extensions and future work are discussed in Section 5.5.
5.1 Monitoring THA surgical outcomes

The objective for monitoring THA surgical outcomes is to propose an accurate framework for patient risk stratification and real-time surgical outcome evaluation, using unplanned readmission as the patient outcome measure.

This research provides a better understanding of how to evaluate a medical team’s surgical outcome performance, while taking into account the inherent difference in patient risk profile. Since we do not want to unfairly penalize a medical team simply because the team treats sicker patients, patient risk stratification plays an important role in evaluating a medical team’s performance in an objective and unbiased manner.

Moreover, we combined supervised machine learning algorithms with SPC, and proposed an accurate framework for real-time surgical outcome evaluation. Our proposed framework outperforms the conventional risk-adjustment framework, which is based on logistic regression.

This research provides important implications to healthcare practitioners. With an accurate patient risk stratification tool, the THA patients with high 30-day readmission risk can be effectively identified. Thus, medical interventions, such as frequent follow-ups and using tele-health devices, can be better targeted on these high-risk patients, especially at the post-discharge stage. This data-driven approach may enable potential medical or surgical complications to be detected at an earlier stage, and thus to reduce unplanned readmissions.

Moreover, our proposed framework provides the medical team with a better understanding of the team’s surgical outcome performance in real time. With such a framework, the team members can visualize deterioration in surgical outcomes in a timely manner. Therefore, these members are able to diagnose why deterioration occurs and then take measures to mitigate the problem, using methods such as root cause analysis.

Therefore, our proposed framework effectively uses patient-level data from electronic health records to facilitate making evidence-based medical decisions, improving care quality, and reducing unplanned readmissions.
5.2 Cost-effectiveness of THA for patient subgroups

With the probable shortage of THA surgery capacity in the United States, it is very important to understand the cost-effectiveness of THA surgery for different hip arthritis patients. The objectives of this study are to conduct a more accurate and comprehensive evaluation of the cost-effectiveness of THA surgery, as well as to improve patient outcomes in terms of gained quality of life. In this study, we propose a medical decision support method that can be used to effectively segment patient population and quantify the cost-effectiveness of THA surgery.

We combine unsupervised machine learning algorithms with cluster-specific Markov simulation models, thus allowing cost-effectiveness analysis to be conducted at the patient cluster level. In most of the existing literature, cost-effectiveness analysis is only conducted at the population level. If our cost-effectiveness analysis were conducted at the entire population level, we would have mistakenly concluded that THA is cost-effective for all the patients. Moreover, multiple criteria, including demographic factors and co-morbid conditions, are used to determine the subgroup that a patient belongs to. Thus, the identified patient subgroups are likely to have more homogeneous characteristics.

Our methodology has important implications on medical resources allocation in the society. Health policy makers may need to re-consider any policies that may potentially limit medical resources to THA surgery as a whole: THA is cost-effective for the majority of the patients, e.g., 94.43% in this study. By doing this, a higher portion of the demand for THA can be satisfied in the next 15 years.

This study also provides important insights into determining the appropriate treatment option for each patient. For instance, since relatively young hip arthritis patients with few co-morbidities have low risk of developing post-operative complications and they tend to have a longer remaining lifespan after surgery, these patients benefit the most from THA surgery. In contrast, the medical team may need to convince patients older than 75 with co-morbidities to receive alternative treatments, based on cost-effectiveness and THA surgery risk. Since gained quality of life is an important measure of THA surgical outcomes, this data-driven solution fits into the framework of THA care quality improvement and guides such efforts.
5.3 Meta-analysis of the effectiveness of pre-operative exercise

Mixed results on a medical or surgical intervention make it difficult for healthcare practitioners to interpret whether this intervention is actually effective or not. The objective of this study is to use meta-analysis to provide a rigor and comprehensive understanding of the effectiveness of pre-operative exercise on THA patients throughout the perioperative care period.

This study has important contributions to the medical team on what medical or surgical interventions should be implemented on THA patients. Our meta-analysis shows that pre-operative exercise offers modest benefits while a patient is waiting for THA surgery. This is somehow important, because the average waiting time for THA surgery is three months and the longest waiting time is more than one year (Wallis & Taylor, 2011). Waiting time will probably keep increasing, as the demand for THA surgery is increasing rapidly in the United States.

However, the value of pre-operative exercise declines if it does not make a significant difference post-operatively, because normally post-operative outcomes, such as pain and function, are emphasized more than the pre-operative outcomes (i.e., the outcomes right after pre-operative exercise). Moreover, pre-operative exercise does not significantly reduce LOS.

This study suggests that a medical team can obtain more accurate and comprehensive insights into the effectiveness of a medical or surgical intervention by using meta-analysis. With such insights, interventions with proven effectiveness should be incorporated into THA patient care and implemented on patients, while those without significant benefits should be ruled out. Therefore, an improved quality engineering approach to patient care becomes possible, and this approach has important implications to both medical team and quality management team in a hospital. This is an important step towards managing care quality and improving patient outcomes.
5.4 Limitations

There are several limitations in this doctoral dissertation. These limitations can be classified into four categories: data, assumptions, quality factors, and generalizability.

5.4.1 Data

This dissertation shows how data science and engineering tools, such as predictive modeling, machine learning, SPC, cluster analysis, as well as statistics and meta-analysis, can be used to address healthcare topics and deliver data-driven insights. Identified THA patient-level data records from an academic medical center in central Pennsylvania, United States were used in Chapter 2 and Chapter 3 to illustrate the value of our proposed methods, while data from relevant literature were used in Chapter 4 for the meta-analysis.

But the lack of complete data leads to some limitations in these studies. For instance, the identified THA patient-level data records only included 30-day readmissions to the same medical center. But it is possible that a patient had THA surgery at this medical center, but was readmitted to another hospital within 30 days of discharge. In this case, these readmissions were not captured in our data records. In addition, the data records did not incorporate potential post-discharge risk factors, such as post-discharge care quality and patient adherence to medications.

In the cost-effectiveness analysis in Chapter 3, the cost parameters and quality of life parameters were obtained from relevant literature and other online sources. These parameters may come from a single study or a single institution, and thus may not be representative of the relevant healthcare cost parameters and quality of life parameters across the United States. In Chapter 4, although we searched three databases, there may still exist studies that meet our inclusion criteria, but is not included in our search results. Thus, the incomplete data may somehow reduce the power of our proposed methods.

However, the results of each chapter are still convincing because they were validated through expert opinions, alternative models, or statistical tests. In particular, we showed the risk factors
of 30-day unplanned readmission in Chapter 2 to orthopedic surgeons at the academic medical center and asked for their opinions. In Chapter 3, the results of K-means clustering algorithm were validated by hierarchical clustering algorithm. In Chapter 4, statistical tests were performed to understand whether the weighted mean differences are statistically significant. Moreover, some degree of estimation error may be acceptable, because our proposed data science and engineering models focus on facilitating THA care quality improvement initiatives, rather than reporting the exact numbers of future healthcare savings. Finally, our studies indicate what additional data should be collected in order to build more accurate models and provide more insightful healthcare implications.

5.4.2 Assumptions

In this dissertation, one major assumption is that the data from the academic medical center, relevant literature, and open sources are correct. Errors, however, may exist in these data, especially in the patient-level data records from the academic medical center. Although the data were cleaned to exclude missing data and outliers that could potentially bias the results, the possibility of error still remains.

In Chapter 2, one key assumption in CUSUM control chart is that the expected performance of a medical team remains at a constant level. But this may not be true in reality. To illustrate, orthopedic surgeons may join or leave the academic medical center during 2011 and 2015. Actually, we know that one orthopedic surgeon with expertise in THA left the academic medical center in 2012. In addition, as a healthcare practitioner becomes increasingly experienced in providing care to patients, whether he or she is a surgeon, anesthesiologist, or nurse, the practitioner’s expected performance tends to show an improving trend. But since the CUSUM control chart based on random forest detected deterioration after the 1000th patient, we should ask whether there were less experienced members joining the medical team after 2014 and thus resulted in the lower performance level.
5.4.3 Quality factors

Although some cost-associated factors and effectiveness-associated factors are discussed, this research focuses on quality-associated factors. Data science and engineering tools are used to support medical decision making in THA care quality improvement. The CUSUM control charts in Chapter 2 are used to measure risk-adjusted readmissions, which is a performance quality measure of patient outcomes. In Chapter 3, the quality of life that each patient subgroup could obtain after THA surgery is also a quality measure. In Chapter 4, the effectiveness of pre-operative exercise is also measured by aspects related to quality of life, such as pain, function, and activity of daily living.

But quality-associated factors such as mortality rate and patient satisfaction were not provided in the patient-level data records. Including these quality-associated factors into the surgical outcome evaluation in Chapter 2 and the cost-effectiveness analysis in Chapter 3 would probably make our study results more convincing.

5.4.4 Generalizability

Our proposed methods and frameworks can be applied to other hospitals, or even to other healthcare systems. In particular, hospitals may use our proposed framework in Chapter 2 to effectively conduct patient risk stratification and real-time surgical outcome evaluation. The medical decision makers may use our proposed methods in Chapter 3 to cluster hip arthritis patients and quantify the cost-effectiveness of THA surgery for each patient subgroup.

However, all the specific analyses should be re-conducted, because our analyses are based on the patient-level data records from an academic medical center. The patients at this academic medical center may not be representative of the patient population who need THA surgery. Thus, the risk factors of 30-day readmission may vary from hospital to hospital, because the THA patients treated by these hospitals may have different demographics, co-morbid conditions, and surgery-related factors. Moreover, the exact number of patient clusters and the cost-effectiveness
ratios may change when our methodology is applied to large-size and more representative patient samples.

5.5 Future work

In this section, we discuss possible future extensions of the three main chapters.

5.5.1 Patient risk stratification and surgical outcome evaluation

The risk factors of 30-day unplanned readmission after THA were investigated by logistic regression, decision tree, and random forest. The results indicate that these three risk-adjustment methods identified a set of common risk factors. These risk factors include BMI over 40, discharge to rehabilitation facility, kidney disease, age, hypertension, and male.

One possible extension of this patient risk stratification is to incorporate patients’ medical history and family history information into our analysis. Relevant studies indicated that THA patients with previous surgery history near the hip have increased risk of dislocation, especially during the 30-day post-discharge period (Lewinnek et al., 1978; Woo & Morrey, 1982). Moreover, THA patients with a family history of venous thromboembolism have significantly higher risk of such complication at the post-operative stage (Mont et al., 2004; Markovic-Denic et al., 2012). But medical history and family history were not provided in the patient-level data records that we have. These factors can be used as independent variables in the risk-adjustment algorithms, and thus the power of these models will probably increase. Therefore, the future study may achieve even higher accuracy in patient risk stratification. Besides, THA patients’ medical history and family history provide additional opportunities to better manage care transitions, especially for high-risk patients. In addition, future research will look at further validation of these risk-adjustment models on larger, more representative patient samples.

Moreover, in the future we may work closely with the medical team at the academic medical center on surgical outcome evaluation, based on our proposed control charts. In addition, we
might work with them on using lean and six sigma methods, a commonly-used methodology for quality improvement in healthcare, to diagnose potential problems, improve care quality, and reduce unplanned readmission after THA.

5.5.2 Cost-effectiveness of THA for each patient subgroup

We used unsupervised machine learning algorithms to cluster hip arthritis patients into different subgroups, and then we used cluster-specific Markov simulation models to analyze the cost-effectiveness of THA surgery for each patient subgroup. The results indicate that THA is more cost-effective for relatively young patients with few co-morbidities, while it is not cost-effective for the oldest patient group (average age 75.47).

In the future, we will conduct cluster analysis and cost-effectiveness analysis using large-size and more representative hip arthritis patient-level data records. Preferably, these data records will contain the relevant medical cost and quality of life information.

Moreover, we may model the decision process of whether or not to have THA surgery from a hip arthritis patient’s perspective. The patient may only consider his/her out-of-pocket costs of both the surgery option and the non-surgery option. With such an analysis, we will obtain understanding of the decision making process of both healthcare provider and hip arthritis patient. Therefore, improved data-driven solutions might be delivered by taking into account patients’ opinions.

5.5.3 Meta-analysis of the effectiveness of pre-operative exercise

Meta-analysis was conducted for understanding the effectiveness of pre-operative exercise on THA patients, using the mixed results of relevant literature. The results indicate that while pre-operative exercise program offers moderate benefits with respect to pain relief, physical function, and activity of daily living at the pre-operative stage, the benefits diminish after THA surgery and become insignificant at three months post-operatively.
In the future, we may use the same meta-analysis method to study other THA-relevant interventions, such as enhanced recovery program after THA surgery and using tele-health device for patient recovery, based on the results of relevant literature. By conducting these meta-analyses, the medical team will obtain a better understanding of the effectiveness of each medical or surgical intervention. Therefore, the medical team will be able to incorporate effective interventions into THA care delivery to improve patient outcomes.
Appendix A

R code for logistic regression, decision tree, and random forest

readPro = read.csv("C:/PhD Comprehensive Exam/Statistical learning/Finalized dataset_all cause.csv", header=T)
attach(readPro)
install.packages("verification")
library(verification)
install.packages("ROSE")
library(ROSE)
install.packages("MASS")
library(MASS)
install.packages("pROC")
library(pROC)

#Partition the whole dataset into training dataset and test dataset
set.seed(1)
n = dim(readPro)[1]
test = sample(n, n/10)
readPro.train = readPro[-test,]
readPro.test = readPro[test,]

#Logistic regression using AIC as criteria on the training dataset
install.packages("leaps")
library(leaps)
# Logistic regression backward selection
glm.full.train = glm(Unplanned~Age+log(Distance)+White+Black+Male+log(LOS)+TotalHip+BMIOver40+CoronaryAtherosclerosis+Diabetes+HeartFailure+CardiacDisease+MyocardialInfarction+Hypertension+Pneumonia+LiverDisease+KidneyDisease+DischargeToRehabOrSnf+DischargeToHomeHealth, family=binomial(), data=readPro.train)
summary(glm.full.train)
backAIC.train = step(glm.full.train, direction="backward", data=readPro.train)
summary(backAIC.train)

#Logistic regression forward selection
mint.train = glm(Unplanned~1, family=binomial(), data=readPro.train)
forwardAIC.train = step(mint.train, scope=list(lower=~1, upper=~Age+log(Distance)+White+Black+Male+log(LOS)+TotalHip+BMIOver40+CoronaryAtherosclerosis+Diabetes+HeartFailure+CardiacDisease+MyocardialInfarction+Hypertension+Pneumonia+LiverDisease+KidneyDisease+DischargeToRehabOrSnf+DischargeToHomeHealth), direction="forward", family=binomial(), data=readPro.train)
summary(forwardAIC.train)

#Accuracy of predicting training dataset of the whole dataset by backward selection
logistic.backward.train.whole = predict(backAIC.train, readPro.train, type="response")

#Accuracy of predicting training dataset of the whole dataset by forward selection
logistic.forward.train.whole = predict(forwardAIC.train, readPro.train, type="response")
auc(readPro.train$Unplanned, logistic.forward.train.whole)
#Accuracy of predicting test dataset of the whole dataset by backward selection
logistic.backward.test.whole = predict(backAIC.train, readPro.test, type="response")
auc(readPro.test$Unplanned, logistic.backward.test.whole)
#Accuracy of predicting test dataset of the whole dataset by forward selection
logistic.forward.test.whole = predict(forwardAIC.train, readPro.test, type="response")
auc(readPro.test$Unplanned, logistic.forward.test.whole)
#Brier score on logistic regression on training dataset
unplannedWholeTrain = readPro.train$Unplanned
sumSquaWholeTrain = 0
for(i in 1:length(unplannedWholeTrain)){
squa = (unplannedWholeTrain[i]-logistic.backward.train.whole[i])*(unplannedWholeTrain[i]-logistic.backward.train.whole[i]);
sumSquaWholeTrain = sumSquaWholeTrain + squa;
}
print(sumSquaWholeTrain)
#Brier score on logistic regression on test dataset
unplannedWholeTest = readPro.test$Unplanned
sumSquaWholeTest = 0
for(i in 1:length(unplannedWholeTest)){
squa = (unplannedWholeTest[i]-logistic.backward.test.whole[i])*(unplannedWholeTest[i]-logistic.backward.test.whole[i]);
sumSquaWholeTest = sumSquaWholeTest + squa;
}
print(sumSquaWholeTest)

#Decision tree
install.packages("tree")
library(tree)
#Decision tree on the training dataset
tree.train.whole =
tree(Unplanned~Age+log(Distance)+White+Black+Male+log(LOS)+TotalHip+BMIOver40+CoronaryAtherosclerosis+Diabetes+HeartFailure+CardiacDisease+MyocardialInfarction+Hypertension+Pneumonia+LiverDisease+KidneyDisease+DischargeToRehabOrSnf+DischargeToHomeHealth, data=readPro.train)
summary(tree.train.whole)
tree.train.whole
dev.new(width=25, height=15)
plot(tree.train.whole)
text(tree.train.whole, pretty=0)
#Make prediction on the training dataset
tree.train.pred.whole = predict(tree.train.whole, readPro.train, type="vector")
auc(readPro.train$Unplanned, tree.train.pred.whole)
#Make prediction on the test dataset
tree.test.pred.whole = predict(tree.train.whole, readPro.test, type="vector")
auc(readPro.test$Unplanned, tree.test.pred.whole)
#Brier score on decision tree on training dataset
unplannedWholeTrain = readPro.train$Unplanned
sumSquaWholeTrain = 0
for(i in 1:length(unplannedWholeTrain)){

117
squa = (unplannedWholeTrain[i]-tree.train.pred.whole[i])*(unplannedWholeTrain[i]-tree.train.pred.whole[i]);
sumSquaWholeTrain = sumSquaWholeTrain + squa;
}
print(sumSquaWholeTrain)
#Brier score on decision tree on test dataset
unplannedWholeTest = readPro.test$Unplanned
sumSquaWholeTest = 0
for(i in 1: length(unplannedWholeTest)){
squa = (unplannedWholeTest[i]-tree.test.pred.whole[i])*(unplannedWholeTest[i]-tree.test.pred.whole[i]);
sumSquaWholeTest = sumSquaWholeTest + squa;
}
print(sumSquaWholeTest)

#Random forest
install.packages("randomForest")
library(randomForest)
Readmission = ifelse(readPro$Unplanned==1, "Yes", "No")
readProSta = data.frame(readPro, Readmission)
readProSta.train = readProSta[-test,]
readProSta.test = readProSta[test,]
set.seed(1)
rf.train.whole = randomForest(Readmission~., data=readProSta.train, mtry=5, importance=TRUE)
rf.train.whole
#Make prediction on the training dataset
rf.train.pred.whole = predict(rf.train.whole, newdata=readProSta.train, type="prob")
auc(readProSta.train$Unplanned, rf.train.pred.whole[,2])
#Make prediction on the test dataset
rf.test.pred.whole = predict(rf.train.whole, newdata=readProSta.test, type="prob")
auc(readProSta.test$Unplanned, rf.test.pred.whole[,2])
#Brier score on random forest on training dataset
unplannedWholeTrain = readPro.train$Unplanned
sumSquaWholeTrain = 0
for(i in 1: length(unplannedWholeTrain)){
squa = (unplannedWholeTrain[i]-rf.train.pred.whole[i,2])*(unplannedWholeTrain[i]-rf.train.pred.whole[i,2]);
sumSquaWholeTrain = sumSquaWholeTrain + squa;
}
print(sumSquaWholeTrain)
#Brier score on random forest on test dataset
unplannedWholeTest = readPro.test$Unplanned
sumSquaWholeTest = 0
for(i in 1: length(unplannedWholeTest)){
squa = (unplannedWholeTest[i]-rf.test.pred.whole[i,2])*(unplannedWholeTest[i]-rf.test.pred.whole[i,2]);
sumSquaWholeTest = sumSquaWholeTest + squa;
}
print(sumSquaWholeTest)
#Draw Figure 2-2 and Figure 2-3
logistic.whole = predict(backAIC.train, readPro, type="response")
dev.new(width=25, height=8)
plot(logistic.whole, xlab = "Time Sequence", ylab = "Readmission Risk", col="black", type = 'l', ylim = c(0, 1.0))
print(subset(logistic.whole, logistic.whole>=0.4))
rf.whole = predict(rf.train.whole, newdata=readProSta, type="prob")
dev.new(width=25, height=8)
plot(rf.whole[,2], xlab = "Time Sequence", ylab = "Readmission Risk", col="black", type = 'l', ylim = c(0, 1.0))
print(subset(rf.whole[,2], rf.whole[,2]>=0.4))
Appendix B

R code for determining ARL₀ of CUSUM control charts

# Determine upper control limit of CUSUM based on logistic regression
readPro = read.csv("C:/PhD Comprehensive Exam/Statistical learning/Finalized dataset_all cause.csv", header=T);
attach(readPro);
install.packages("verification");
library(verification);
install.packages("ROSE");
library(ROSE);
install.packages("MASS");
library(MASS);

set.seed(1);
n = dim(readPro)[1];
test = sample(n, n/(3.3333));
readPro.train = readPro[-test,];
readPro.test = readPro[test,];
logistic.train =
  glm(Unplanned~Male+BMIOver40+Hypertension+KidneyDisease+DischargeToRehabOrSnf+DischargeToHomeHealth, data=readPro.train, family=binomial());
summary(logistic.train);
logistic.pred.all = predict(logistic.train, newdata=readPro, type="response");
yRisk = Unplanned;
conLimitDeter = 4.58;
zAftDeter = 0.0;
zPreDeter = 0.0;
oddsRatioDeter = 2;
run = 5000;
arlDeter = seq(0.0, 0.0, length=run);
exhaust = 100000;
range = length(yRisk);

for(j in 1: run){
  for(i in 1: exhaust){
    randomNum1 = runif(1,0,1);
    index = ceiling(range*randomNum1);
    randomNum2 = runif(1,0,1);
    if(randomNum2 <= logistic.pred.all[index]){ y = 1; }
    else { y = 0; }
  }
  if(y==1){
    
  }

120
\[ w_1 = \log\left(\frac{(1 - \text{logistic.pred.all[index]} + 1 \times \text{logistic.pred.all[index]} \times \text{oddsRatioDeter})}{(1 - \text{logistic.pred.all[index]} + \text{oddsRatioDeter} \times \text{logistic.pred.all[index]} \times 1)}\right); \]

else{
\[ w_1 = \log\left(\frac{(1 - \text{logistic.pred.all[index]} + 1 \times \text{logistic.pred.all[index]})}{(1 - \text{logistic.pred.all[index]} + \text{oddsRatioDeter} \times \text{logistic.pred.all[index]})}\right); \]
}\n\[
\text{zAftDeter} = \max((\text{zPreDeter} + w_1), 0);
\]

if(zAftDeter <= \text{conLimitDeter}){
\[
\text{zPreDeter} = \text{zAftDeter};
\]
}

\[
\text{arlDeter}[j] = i;
\]
\[
\text{zPreDeter} = 0.0;
\]
\[
\text{zAftDeter} = 0.0;
\]
\[
\text{break};
\]
if((i == \text{exhaust}) && (\text{zAftDeter} <= \text{conLimitDeter})){
\[
\text{print}(i);
\]
\[
\text{arlDeter}[j] = i;
\]
\[
\text{zPreDeter} = 0.0;
\]
\[
\text{zAftDeter} = 0.0;
\]
}
}
\[
\text{avg} = \text{mean(arlDeter)};
\]
\[
\text{print(avg)};
\]
\[
\text{std} = \text{sd(arlDeter)};
\]
\[
\text{print(std)};
\]
detach(readPro);

#Determine lower control limit of CUSUM based on logistic regression
\[
\text{readPro} = \text{read.csv("C:/PhD Comprehensive Exam/Statistical learning/Finalized dataset_all cause.csv", header=T)};
\]
\[
\text{attach(readPro)};
\]
\[
\text{install.packages("verification")};
\]
\[
\text{library(verification)};
\]
\[
\text{install.packages("ROSE")};
\]
\[
\text{library(ROSE)};
\]
\[
\text{install.packages("MASS")};
\]
\[
\text{library(MASS)};
\]
\[
\text{set.seed(1)};
\]
\[
\text{n} = \text{dim(readPro)}[1];
\]
\[
\text{test} = \text{sample(n, n/(3.3333))};
\]
readPro.train = readPro[-test,];
readPro.test = readPro[test,];
logistic.train =
glm(Unplanned~Male+BMIOver40+Hypertension+KidneyDisease+DischargeToRehabOrSnf+DischargeToHomeHealth, data=readPro.train, family=binomial());
summary(logistic.train);
logistic.pred.all = predict(logistic.train, newdata=readPro, type="response");
yRisk = Unplanned;
conLimitImpro = -4.24;
zAftImpro = 0.0;
zPreImpro = 0.0;
oddsRatioImpro = 0.5;
run = 5000;
arlImpro = seq(0.0, 0.0, length=run);
exhaust = 100000;
range = length(yRisk);

for(j in 1: run){
  for(i in 1: exhaust){
    randomNum1 = runif(1,0,1);
    index = ceiling(range*randomNum1);
    randomNum2 = runif(1,0,1);
    if(randomNum2 <= logistic.pred.all[index]){y = 1;}
    else {y = 0;}

    if(y==1){
      w2 = log(((1-logistic.pred.all[index]+1*logistic.pred.all[index])*oddsRatioImpro)/((1-logistic.pred.all[index]+oddsRatioImpro*logistic.pred.all[index])*1));
    }
    else{
      w2 = log((1-logistic.pred.all[index]+1*logistic.pred.all[index])/(1-logistic.pred.all[index]+oddsRatioImpro*logistic.pred.all[index]));
    }
    zAftImpro = min((zPreImpro-w2), 0);

    if(zAftImpro >= conLimitImpro){zPreImpro = zAftImpro;}
    else{
      print(i);
      arlImpro[j] = i;
      zPreImpro = 0.0;
      zAftImpro = 0.0;
      break;
    }
  }
}

122
if((i == exhaust) && (zAftImpro >= conLimitImpro)) {
    print(i);
    arlImpro[j] = i;
    zPreImpro = 0.0;
    zAftImpro = 0.0;
}
}

avg = mean(arlImpro);
print(avg);
std = sd(arlImpro);
print(std);
detach(readPro);

# Determine upper control limit of CUSUM based on random forest
readPro = read.csv("C:/PhD Comprehensive Exam/Statistical learning/Finalized dataset_all cause.csv", header=T);
attach(readPro);
install.packages("verification");
library(verification);
install.packages("ROSE");
library(ROSE);
install.packages("MASS");
library(MASS);
set.seed(1);
n = dim(readPro)[1];
test = sample(n, n/(3.3333));
readPro.train = readPro[-test,];
readPro.test = readPro[test,];
Readmission = ifelse(readPro$Unplanned==1, "Yes", "No");
readm = data.frame(readPro, Readmission);
readm.train = readm[-test,];
readm.test = readm[test,];
install.packages("randomForest");
library(randomForest);
set.seed(1);
rf.train = randomForest(Readmission~., data=readm.train, mtry=5, importance=TRUE);
rf.train;
yhat.rf.pred.rep = predict(rf.train, newdata=readm, type="prob");
random.pred.all = yhat.rf.pred.rep[,2];
yRisk = Unplanned;
conLimitDeter = 4.19;
zAftDeter = 0.0;
zPreDeter = 0.0;
oddsRatioDeter = 2;
run = 5000;
arlDeter = seq(0.0, 0.0, length=run);
exhaust = 100000;
range = length(yRisk);

for(j in 1: run){
    for(i in 1: exhaust){
        randomNum1 = runif(1,0,1);
        index = ceiling(range*randomNum1);
        randomNum2 = runif(1,0,1);
        if(randomNum2 <= random.pred.all[index]){
            y = 1;
        } else {
            y = 0;
        }
        if(y==1){
            w1 = log(((1-random.pred.all[index])+1*random.pred.all[index])*oddsRatioDeter)/((1-
            random.pred.all[index]+oddsRatioDeter*random.pred.all[index])*1));
        } else{
            w1 = log((1-random.pred.all[index])+1*random.pred.all[index])/(1-
            random.pred.all[index]+oddsRatioDeter*random.pred.all[index]));
        }
        zAftDeter = max((zPreDeter+w1), 0);
        if(zAftDeter <= conLimitDeter){
            zPreDeter = zAftDeter;
        } else{
            print(i);
            arlDeter[j] = i;
            zPreDeter = 0.0;
            zAftDeter = 0.0;
            break;
        }
        if((i == exhaust) && (zAftDeter <= conLimitDeter)){
            print(i);
            arlDeter[j] = i;
            zPreDeter = 0.0;
            zAftDeter = 0.0;
        }
    }
}

avg = mean(arlDeter);
print(avg);
std = sd(arlDeter);
print(std);
detach(readPro);

# Determine lower control limit of CUSUM based on random forest
readPro = read.csv("C:/PhD Comprehensive Exam/Statistical learning/Finalized dataset_all cause.csv", header=T);
attach(readPro);
install.packages("verification");
library(verification);
install.packages("ROSE");
library(ROSE);
install.packages("MASS");
library(MASS);

set.seed(1);
n = dim(readPro)[1];
test = sample(n, n/(3.3333));
readPro.train = readPro[-test,];
readPro.test = readPro[test,];
Readmission = ifelse(readPro$Unplanned==1, "Yes", "No");
readm = data.frame(readPro, Readmission);
readm.train = readm[-test,];
readm.test = readm[test,];
install.packages("randomForest");
library(randomForest);
set.seed(1);
rf.train = randomForest(Readmission~., data=readm.train, mtry=5, importance=TRUE);
rf.train;
yhat.rf.pred.rep = predict(rf.train, newdata=readm, type="prob");
random.pred.all = yhat.rf.pred.rep[,2];
yRisk = Unplanned;
conLimitImpro = -3.94;
zAftImpro = 0.0;
zPreImpro = 0.0;
oddsRatioImpro = 0.5;
run = 5000;
arlImpro = seq(0.0, 0.0, length=run);
exhaust = 100000;
range = length(yRisk);

for(j in 1: run){
  for(i in 1: exhaust){
    randomNum1 = runif(1,0,1);    
    index = ceiling(range*randomNum1);   
    randomNum2 = runif(1,0,1);
    if(randomNum2 <= random.pred.all[index]){  
      y = 1;
    }
    else {
      y = 0;
    }
  }
}

125
if(y==1){
    w2 = log(((1-random.pred.all[index]+1*random.pred.all[index])*oddsRatioImpro)/((1-
    random.pred.all[index]+oddsRatioImpro*random.pred.all[index])*1));
} else{
    w2 = log((1-
    random.pred.all[index]+1*random.pred.all[index])/(1
    random.pred.all[index]+oddsRatioImpro*random.pred.all[index]));
}

zAftImpro = min((zPreImpro-w2), 0)

if(zAftImpro >= conLimitImpro){
    zPreImpro = zAftImpro;
} else{
    print(i);
    arlImpro[j] = i;
    zPreImpro = 0.0;
    zAftImpro = 0.0;
    break;
}

if((i == exhaust) && (zAftImpro >= conLimitImpro)){
    print(i);
    arlImpro[j] = i;
    zPreImpro = 0.0;
    zAftImpro = 0.0;
}
}

avg = mean(arlImpro);
print(avg);
std = sd(arlImpro);
print(std);
detach(readPro);
Appendix C

R code for CUSUM control charts based on logistics regression and random forest

#CUSUM chart based on logistic regression
readPro = read.csv("C:/PhD Comprehensive Exam/Statistical learning/Finalized dataset_all cause.csv", header=T);
attach(readPro);
install.packages("verification");
library(verification);
install.packages("ROSE");
library(ROSE);
install.packages("MASS");
library(MASS);

set.seed(1);
n = dim(readPro)[1];
test = sample(n, n/(3.3333));
readPro.train = readPro[-test,];
readPro.test = readPro[test,];
logistic.train = glm(Unplanned~Male+BMIOver40+Hypertension+KidneyDisease+DischargetoRehabOrSnf+Discharge
toHomeHealth, data=readPro.train, family=binomial());
summary(logistic.train);
logistic.pred.all = predict(logistic.train, newdata=readPro, type="response");
yRisk = Unplanned;
conLimitDeter = 4.58;
conLimitImprove = -4.24;
conLimitDeterArray = seq(conLimitDeter, conLimitDeter, length=length(yRisk));
conLimitImproveArray = seq(conLimitImprove, conLimitImprove, length=length(yRisk));
zAftDeter = 0.0;
zPreDeter = 0.0;
zAftImprove = 0.0;
zPreImprove = 0.0;
conStaDeter = seq(0.0, 0.0, length=length(yRisk));
conStaImprove = seq(0.0, 0.0, length=length(yRisk));
oddsRatioDeter = 2;
oddsRatioImprove = 0.5;
range = length(yRisk);

for (i in 1:length(yRisk)){
  if(yRisk[i] == 1){
    w1 = log(((1-logistic.pred.all[i]+1*logistic.pred.all[i])*oddsRatioDeter)/((1-
    logistic.pred.all[i]+oddsRatioDeter*logistic.pred.all[i])*1));
    w2 = log(((1-logistic.pred.all[i]+1*logistic.pred.all[i])*oddsRatioImprove)/((1-
    logistic.pred.all[i]+oddsRatioImprove*logistic.pred.all[i])*1));
  } else{
w1 = log((1-logistic.pred.all[i]+1*logistic.pred.all[i])/(1-logistic.pred.all[i]+oddsRatioDeter*logistic.pred.all[i]));

w2 = log((1-logistic.pred.all[i]+1*logistic.pred.all[i])/(1-logistic.pred.all[i]+oddsRatioImpro*logistic.pred.all[i]));

} 

zAftDeter = max((zPreDeter+w1), 0);
zAftImpro = min((zPreImpro-w2), 0);
conStaDeter[i] = zAftDeter;
conStaImpro[i] = zAftImpro;

if((zAftDeter <= conLimitDeter)&&(zAftImpro >= conLimitImpro)){
  zPreDeter = zAftDeter;
  zPreImpro = zAftImpro;
}
else{
  print(i);
  zPreDeter=0.0;
  zPreImpro=0.0;
}

dev.new(width=25, height=12);
par(mfrow=c(2,1));
plot(conLimitDeterArray, xlab = "Time Sequence", ylab = "", col="red", type = 'l', ylim = c(0, 5.0));
par(new = TRUE);
plot(conStaDeter, xlab = " ", ylab = " ", col="black", type = 'l', ylim = c(0, 5.0));
par(new = FALSE);
plot(conLimitImproArray, xlab = "Time Sequence", ylab = " ", col="red", type = 'l', ylim = c(-5.0, 0));
par(new = TRUE);
plot(conStaImpro, xlab = " ", ylab = " ", col="black", type = 'l', ylim = c(-5.0, 0));
detach(readPro);

#CUSUM chart based on random forest
readPro = read.csv("C:/PhD Comprehensive Exam/Statistical learning/Finalized dataset_all cause.csv", header=T);
attach(readPro);
install.packages("verification");
library(verification);
install.packages("ROSE");
library(ROSE);
install.packages("MASS");
library(MASS);

set.seed(1);
n = dim(readPro)[1];
test = sample(n, n/(3.3333));
readPro.train = readPro[-test,];
readPro.test = readPro[test,];
Readmission = ifelse(readPro$Unplanned==1, "Yes", "No");
readm = data.frame(readPro, Readmission);

128
readm.train = readm[-test,];
readm.test = readm[test,];
install.packages("randomForest");
library(randomForest);
set.seed(1);
rf.train = randomForest(Readmission~-Unplanned, data=readm.train, mtry=5, importance=TRUE);
rf.train;
yhat.rf.pred.rep = predict(rf.train, newdata=readm, type="prob");
random.pred.all = yhat.rf.pred.rep[,2];
yRisk = Unplanned;
conLimitDeter = 4.19;
conLimitImpro = -3.94;
conLimitDeterArray = seq(conLimitDeter, conLimitDeter, length=length(yRisk));
conLimitImproArray = seq(conLimitImpro, conLimitImpro, length=length(yRisk));
zAftDeter = 0.0;
zPreDeter = 0.0;
zAftImpro = 0.0;
zPreImpro = 0.0;
conStaDeter = seq(0.0, 0.0, length=length(yRisk));
conStaImpro = seq(0.0, 0.0, length=length(yRisk));
oddsRatioDeter = 2;
oddsRatioImpro = 0.5;
range = length(yRisk);

for (i in 1: length(yRisk)){
  if(yRisk[i] == 1){
    w1 = log(((1-random.pred.all[i]+1*random.pred.all[i])*oddsRatioDeter)/((1-
random.pred.all[i]+oddsRatioDeter*random.pred.all[i])*1));
    w2 = log(((1-random.pred.all[i]+1*random.pred.all[i])*oddsRatioImpro)/((1-
random.pred.all[i]+oddsRatioImpro*random.pred.all[i])*1));
  }
  else{
    w1 = log((1-random.pred.all[i]+1*random.pred.all[i])/(1-
random.pred.all[i]+oddsRatioDeter*random.pred.all[i]));
    w2 = log((1-random.pred.all[i]+1*random.pred.all[i])/(1-
random.pred.all[i]+oddsRatioImpro*random.pred.all[i]));
  }
  zAftDeter = max((zPreDeter+w1), 0);
  zAftImpro = min((zPreImpro-w2), 0);
  conStaDeter[i] = zAftDeter;
  conStaImpro[i] = zAftImpro;
  if((zAftDeter <= conLimitDeter)&&(zAftImpro >= conLimitImpro)){
    zPreDeter = zAftDeter;
    zPreImpro = zAftImpro;
  }
  else{
    print(i);
    zPreDeter=0.0;
zPreImpro=0.0;
}
}
dev.new(width=25, height=12);
par(mfrow=c(2,1));
plot(conLimitDeterArray, xlab = "Time Sequence", ylab = " ", col="red", type = 'l', ylim = c(0, 5.0));
par(new = TRUE);
plot(conStaDeter, xlab = " ", ylab = " ", col="black", type = 'l', ylim = c(0, 5.0));
par(new = FALSE);
plot(conLimitImproArray, xlab = "Time Sequence", ylab = " ", col="red", type = 'l', ylim = c(-5.0, 0));
par(new = TRUE);
plot(conStaImpro, xlab = " ", ylab = " ", col="black", type = 'l', ylim = c(-5.0, 0));
detach(readPro);
Appendix D

R code for K-means clustering and hierarchical clustering

readPro = read.csv("C:/PhD Comprehensive Exam/Statistical learning/Finalized dataset_all cause.csv", header=T)
attach(readPro)
install.packages("cluster")
library(cluster)

#Extract the six risk factors used for clustering
readProPart = readPro[, c(1,2,5,15,18,21)]
dim(readProPart)
#Transform the distance measure to Gower’s distance
dist = daisy(readProPart, metric="gower", stand=FALSE, type=list(logratio=c(2)))

#K-means clustering
#Use 1 cluster
set.seed(1)
km.out1 = kmeans(dist, 1, nstart=20)
km.out1$tot.withinss
#Use 2 clusters
set.seed(1)
km.out2 = kmeans(dist, 2, nstart=20)
km.out2$tot.withinss
#Use 3 clusters
set.seed(1)
km.out3 = kmeans(dist, 3, nstart=20)
km.out3$tot.withinss
#Use 4 clusters
set.seed(1)
km.out4 = kmeans(dist, 4, nstart=20)
km.out4$tot.withinss
#Use 5 clusters
set.seed(1)
km.out5 = kmeans(dist, 5, nstart=20)
km.out5$tot.withinss
#Use 6 clusters
set.seed(1)
km.out6 = kmeans(dist, 6, nstart=20)
km.out6$tot.withinss
#Use 7 clusters
set.seed(1)
km.out7 = kmeans(dist, 7, nstart=20)
km.out7$tot.withinss
#Use 8 clusters
set.seed(1)
km.out8 = kmeans(dist, 8, nstart=20)
km.out8$tot.withinss
#Use 9 clusters
set.seed(1)
km.out9 = kmeans(dist, 9, nstart=20)
km.out9$tot.withinss
#Use 10 clusters
set.seed(1)
km.out10 = kmeans(dist, 10, nstart=20)
km.out10$tot.withinss
#Use 11 clusters
set.seed(1)
km.out11 = kmeans(dist, 11, nstart=20)
km.out11$tot.withinss
#Use 12 clusters
set.seed(1)
km.out12 = kmeans(dist, 12, nstart=20)
km.out12$tot.withinss
#Use 13 clusters
set.seed(1)
km.out13 = kmeans(dist, 13, nstart=20)
km.out13$tot.withinss
#Use 14 clusters
set.seed(1)
km.out14 = kmeans(dist, 14, nstart=20)
km.out14$tot.withinss
#Use 15 clusters
set.seed(1)
km.out15 = kmeans(dist, 15, nstart=20)
km.out15$tot.withinss
#Use 16 clusters
set.seed(1)
km.out16 = kmeans(dist, 16, nstart=20)
km.out16$tot.withinss
#Use 17 clusters
set.seed(1)
km.out17 = kmeans(dist, 17, nstart=20)
km.out17$tot.withinss
#Use 18 clusters
set.seed(1)
km.out18 = kmeans(dist, 18, nstart=20)
km.out18$tot.withinss
#Use 19 clusters
set.seed(1)
km.out19 = kmeans(dist, 19, nstart=20)
km.out19$tot.withinss
#Use 20 clusters
set.seed(1)
km.out20 = kmeans(dist, 20, nstart=20)
km.out20$tot.withinss
# If we use 5 clusters
clusterAssignKM5 = km.out5$cluster
readProPartComKM5 = data.frame(readProPart, clusterAssignKM5)

# Get the average statistics of cluster 1
dim(subset(readProPartComKM5, clusterAssignKM5==1))[1]
mean(subset(readProPartComKM5, clusterAssignKM5==1)$Age)
mean(subset(readProPartComKM5, clusterAssignKM5==1)$Distance)
mean(subset(readProPartComKM5, clusterAssignKM5==1)$Male)
mean(subset(readProPartComKM5, clusterAssignKM5==1)$Hypertension)
mean(subset(readProPartComKM5, clusterAssignKM5==1)$KidneyDisease)
mean(subset(readProPartComKM5, clusterAssignKM5==1)$BMI)

# Get the average statistics of cluster 2
dim(subset(readProPartComKM5, clusterAssignKM5==2))[1]
mean(subset(readProPartComKM5, clusterAssignKM5==2)$Age)
mean(subset(readProPartComKM5, clusterAssignKM5==2)$Distance)
mean(subset(readProPartComKM5, clusterAssignKM5==2)$Male)
mean(subset(readProPartComKM5, clusterAssignKM5==2)$Hypertension)
mean(subset(readProPartComKM5, clusterAssignKM5==2)$KidneyDisease)
mean(subset(readProPartComKM5, clusterAssignKM5==2)$BMI)

# Get the average statistics of cluster 3
dim(subset(readProPartComKM5, clusterAssignKM5==3))[1]
mean(subset(readProPartComKM5, clusterAssignKM5==3)$Age)
mean(subset(readProPartComKM5, clusterAssignKM5==3)$Distance)
mean(subset(readProPartComKM5, clusterAssignKM5==3)$Male)
mean(subset(readProPartComKM5, clusterAssignKM5==3)$Hypertension)
mean(subset(readProPartComKM5, clusterAssignKM5==3)$KidneyDisease)
mean(subset(readProPartComKM5, clusterAssignKM5==3)$BMI)

# Get the average statistics of cluster 4
dim(subset(readProPartComKM5, clusterAssignKM5==4))[1]
mean(subset(readProPartComKM5, clusterAssignKM5==4)$Age)
mean(subset(readProPartComKM5, clusterAssignKM5==4)$Distance)
mean(subset(readProPartComKM5, clusterAssignKM5==4)$Male)
mean(subset(readProPartComKM5, clusterAssignKM5==4)$Hypertension)
mean(subset(readProPartComKM5, clusterAssignKM5==4)$KidneyDisease)
mean(subset(readProPartComKM5, clusterAssignKM5==4)$BMI)

# Get the average statistics of cluster 5
dim(subset(readProPartComKM5, clusterAssignKM5==5))[1]
mean(subset(readProPartComKM5, clusterAssignKM5==5)$Age)
mean(subset(readProPartComKM5, clusterAssignKM5==5)$Distance)
mean(subset(readProPartComKM5, clusterAssignKM5==5)$Male)
mean(subset(readProPartComKM5, clusterAssignKM5==5)$Hypertension)
mean(subset(readProPartComKM5, clusterAssignKM5==5)$KidneyDisease)
mean(subset(readProPartComKM5, clusterAssignKM5==5)$BMI)

# If we use 6 clusters
clusterAssignKM6 = km.out6$cluster
readProPartComKM6 = data.frame(readProPart, clusterAssignKM6)

# Get the average statistics of cluster 1
dim(subset(readProPartComKM6, clusterAssignKM6==1))[1]
mean(subset(readProPartComKM6, clusterAssignKM6==1)$Age)
mean(subset(readProPartComKM6, clusterAssignKM6==1)$Distance)
mean(subset(readProPartComKM6, clusterAssignKM6==1)$Male)
mean(subset(readProPartComKM6, clusterAssignKM6==1)$Hypertension)
mean(subset(readProPartComKM6, clusterAssignKM6==1)$KidneyDisease)
mean(subset(readProPartComKM6, clusterAssignKM6==1)$BMI)

#Get the average statistics of cluster 2
dim(subset(readProPartComKM6, clusterAssignKM6==2))[[1]]
mean(subset(readProPartComKM6, clusterAssignKM6==2)$Age)
mean(subset(readProPartComKM6, clusterAssignKM6==2)$Distance)
mean(subset(readProPartComKM6, clusterAssignKM6==2)$Male)
mean(subset(readProPartComKM6, clusterAssignKM6==2)$Hypertension)
mean(subset(readProPartComKM6, clusterAssignKM6==2)$KidneyDisease)
mean(subset(readProPartComKM6, clusterAssignKM6==2)$BMI)

#Get the average statistics of cluster 3
dim(subset(readProPartComKM6, clusterAssignKM6==3))[[1]]
mean(subset(readProPartComKM6, clusterAssignKM6==3)$Age)
mean(subset(readProPartComKM6, clusterAssignKM6==3)$Distance)
mean(subset(readProPartComKM6, clusterAssignKM6==3)$Male)
mean(subset(readProPartComKM6, clusterAssignKM6==3)$Hypertension)
mean(subset(readProPartComKM6, clusterAssignKM6==3)$KidneyDisease)
mean(subset(readProPartComKM6, clusterAssignKM6==3)$BMI)

#Get the average statistics of cluster 4
dim(subset(readProPartComKM6, clusterAssignKM6==4))[[1]]
mean(subset(readProPartComKM6, clusterAssignKM6==4)$Age)
mean(subset(readProPartComKM6, clusterAssignKM6==4)$Distance)
mean(subset(readProPartComKM6, clusterAssignKM6==4)$Male)
mean(subset(readProPartComKM6, clusterAssignKM6==4)$Hypertension)
mean(subset(readProPartComKM6, clusterAssignKM6==4)$KidneyDisease)
mean(subset(readProPartComKM6, clusterAssignKM6==4)$BMI)

#Get the average statistics of cluster 5
dim(subset(readProPartComKM6, clusterAssignKM6==5))[[1]]
mean(subset(readProPartComKM6, clusterAssignKM6==5)$Age)
mean(subset(readProPartComKM6, clusterAssignKM6==5)$Distance)
mean(subset(readProPartComKM6, clusterAssignKM6==5)$Male)
mean(subset(readProPartComKM6, clusterAssignKM6==5)$Hypertension)
mean(subset(readProPartComKM6, clusterAssignKM6==5)$KidneyDisease)
mean(subset(readProPartComKM6, clusterAssignKM6==5)$BMI)

#Get the average statistics of cluster 6
dim(subset(readProPartComKM6, clusterAssignKM6==6))[[1]]
mean(subset(readProPartComKM6, clusterAssignKM6==6)$Age)
mean(subset(readProPartComKM6, clusterAssignKM6==6)$Distance)
mean(subset(readProPartComKM6, clusterAssignKM6==6)$Male)
mean(subset(readProPartComKM6, clusterAssignKM6==6)$Hypertension)
mean(subset(readProPartComKM6, clusterAssignKM6==6)$KidneyDisease)
mean(subset(readProPartComKM6, clusterAssignKM6==6)$BMI)

#If we use 7 clusters
clusterAssignKM7 = km.out7$cluster
readProPartComKM7 = data.frame(readProPart, clusterAssignKM7)
# Get the average statistics of cluster 1

```r
dim(subset(readProPartComKM7, clusterAssignKM7==1))[[1]]
mean(subset(readProPartComKM7, clusterAssignKM7==1)$Age)
mean(subset(readProPartComKM7, clusterAssignKM7==1)$Distance)
mean(subset(readProPartComKM7, clusterAssignKM7==1)$Male)
mean(subset(readProPartComKM7, clusterAssignKM7==1)$Hypertension)
mean(subset(readProPartComKM7, clusterAssignKM7==1)$KidneyDisease)
mean(subset(readProPartComKM7, clusterAssignKM7==1)$BMI)
```

# Get the average statistics of cluster 2

```r
dim(subset(readProPartComKM7, clusterAssignKM7==2))[[1]]
mean(subset(readProPartComKM7, clusterAssignKM7==2)$Age)
mean(subset(readProPartComKM7, clusterAssignKM7==2)$Distance)
mean(subset(readProPartComKM7, clusterAssignKM7==2)$Male)
mean(subset(readProPartComKM7, clusterAssignKM7==2)$Hypertension)
mean(subset(readProPartComKM7, clusterAssignKM7==2)$KidneyDisease)
mean(subset(readProPartComKM7, clusterAssignKM7==2)$BMI)
```

# Get the average statistics of cluster 3

```r
dim(subset(readProPartComKM7, clusterAssignKM7==3))[[1]]
mean(subset(readProPartComKM7, clusterAssignKM7==3)$Age)
mean(subset(readProPartComKM7, clusterAssignKM7==3)$Distance)
mean(subset(readProPartComKM7, clusterAssignKM7==3)$Male)
mean(subset(readProPartComKM7, clusterAssignKM7==3)$Hypertension)
mean(subset(readProPartComKM7, clusterAssignKM7==3)$KidneyDisease)
mean(subset(readProPartComKM7, clusterAssignKM7==3)$BMI)
```

# Get the average statistics of cluster 4

```r
dim(subset(readProPartComKM7, clusterAssignKM7==4))[[1]]
mean(subset(readProPartComKM7, clusterAssignKM7==4)$Age)
mean(subset(readProPartComKM7, clusterAssignKM7==4)$Distance)
mean(subset(readProPartComKM7, clusterAssignKM7==4)$Male)
mean(subset(readProPartComKM7, clusterAssignKM7==4)$Hypertension)
mean(subset(readProPartComKM7, clusterAssignKM7==4)$KidneyDisease)
mean(subset(readProPartComKM7, clusterAssignKM7==4)$BMI)
```

# Get the average statistics of cluster 5

```r
dim(subset(readProPartComKM7, clusterAssignKM7==5))[[1]]
mean(subset(readProPartComKM7, clusterAssignKM7==5)$Age)
mean(subset(readProPartComKM7, clusterAssignKM7==5)$Distance)
mean(subset(readProPartComKM7, clusterAssignKM7==5)$Male)
mean(subset(readProPartComKM7, clusterAssignKM7==5)$Hypertension)
mean(subset(readProPartComKM7, clusterAssignKM7==5)$KidneyDisease)
mean(subset(readProPartComKM7, clusterAssignKM7==5)$BMI)
```

# Get the average statistics of cluster 6

```r
dim(subset(readProPartComKM7, clusterAssignKM7==6))[[1]]
mean(subset(readProPartComKM7, clusterAssignKM7==6)$Age)
mean(subset(readProPartComKM7, clusterAssignKM7==6)$Distance)
mean(subset(readProPartComKM7, clusterAssignKM7==6)$Male)
mean(subset(readProPartComKM7, clusterAssignKM7==6)$Hypertension)
mean(subset(readProPartComKM7, clusterAssignKM7==6)$KidneyDisease)
mean(subset(readProPartComKM7, clusterAssignKM7==6)$BMI)
```
# Get the average statistics of cluster 7

dim(subset(readProPartComKM7, clusterAssignKM7==7))[1]
mean(subset(readProPartComKM7, clusterAssignKM7==7)$Age)
mean(subset(readProPartComKM7, clusterAssignKM7==7)$Distance)
mean(subset(readProPartComKM7, clusterAssignKM7==7)$Male)
mean(subset(readProPartComKM7, clusterAssignKM7==7)$Hypertension)
mean(subset(readProPartComKM7, clusterAssignKM7==7)$KidneyDisease)
mean(subset(readProPartComKM7, clusterAssignKM7==7)$BMI)

# If we use 8 clusters
clusterAssignKM8 = km.out8$cluster
readProPartComKM8 = data.frame(readProPart, clusterAssignKM8)

# Get the average statistics of cluster 1

dim(subset(readProPartComKM8, clusterAssignKM8==1))[1]
mean(subset(readProPartComKM8, clusterAssignKM8==1)$Age)
mean(subset(readProPartComKM8, clusterAssignKM8==1)$Distance)
mean(subset(readProPartComKM8, clusterAssignKM8==1)$Male)
mean(subset(readProPartComKM8, clusterAssignKM8==1)$Hypertension)
mean(subset(readProPartComKM8, clusterAssignKM8==1)$KidneyDisease)
mean(subset(readProPartComKM8, clusterAssignKM8==1)$BMI)

# Get the average statistics of cluster 2

dim(subset(readProPartComKM8, clusterAssignKM8==2))[1]
mean(subset(readProPartComKM8, clusterAssignKM8==2)$Age)
mean(subset(readProPartComKM8, clusterAssignKM8==2)$Distance)
mean(subset(readProPartComKM8, clusterAssignKM8==2)$Male)
mean(subset(readProPartComKM8, clusterAssignKM8==2)$Hypertension)
mean(subset(readProPartComKM8, clusterAssignKM8==2)$KidneyDisease)
mean(subset(readProPartComKM8, clusterAssignKM8==2)$BMI)

# Get the average statistics of cluster 3

dim(subset(readProPartComKM8, clusterAssignKM8==3))[1]
mean(subset(readProPartComKM8, clusterAssignKM8==3)$Age)
mean(subset(readProPartComKM8, clusterAssignKM8==3)$Distance)
mean(subset(readProPartComKM8, clusterAssignKM8==3)$Male)
mean(subset(readProPartComKM8, clusterAssignKM8==3)$Hypertension)
mean(subset(readProPartComKM8, clusterAssignKM8==3)$KidneyDisease)
mean(subset(readProPartComKM8, clusterAssignKM8==3)$BMI)

# Get the average statistics of cluster 4

dim(subset(readProPartComKM8, clusterAssignKM8==4))[1]
mean(subset(readProPartComKM8, clusterAssignKM8==4)$Age)
mean(subset(readProPartComKM8, clusterAssignKM8==4)$Distance)
mean(subset(readProPartComKM8, clusterAssignKM8==4)$Male)
mean(subset(readProPartComKM8, clusterAssignKM8==4)$Hypertension)
mean(subset(readProPartComKM8, clusterAssignKM8==4)$KidneyDisease)
mean(subset(readProPartComKM8, clusterAssignKM8==4)$BMI)

# Get the average statistics of cluster 5

dim(subset(readProPartComKM8, clusterAssignKM8==5))[1]
mean(subset(readProPartComKM8, clusterAssignKM8==5)$Age)
mean(subset(readProPartComKM8, clusterAssignKM8==5)$Distance)
mean(subset(readProPartComKM8, clusterAssignKM8==5)$Male)
mean(subset(readProPartComKM8, clusterAssignKM8==5)$Hypertension)
mean(subset(readProPartComKM8, clusterAssignKM8==5)$KidneyDisease)
mean(subset(readProPartComKM8, clusterAssignKM8==5)$BMI)

#Get the average statistics of cluster 6
dim(subset(readProPartComKM8, clusterAssignKM8==6))[1]
mean(subset(readProPartComKM8, clusterAssignKM8==6)$Age)
mean(subset(readProPartComKM8, clusterAssignKM8==6)$Distance)
mean(subset(readProPartComKM8, clusterAssignKM8==6)$Male)
mean(subset(readProPartComKM8, clusterAssignKM8==6)$Hypertension)
mean(subset(readProPartComKM8, clusterAssignKM8==6)$KidneyDisease)
mean(subset(readProPartComKM8, clusterAssignKM8==6)$BMI)

#Get the average statistics of cluster 7
dim(subset(readProPartComKM8, clusterAssignKM8==7))[1]
mean(subset(readProPartComKM8, clusterAssignKM8==7)$Age)
mean(subset(readProPartComKM8, clusterAssignKM8==7)$Distance)
mean(subset(readProPartComKM8, clusterAssignKM8==7)$Male)
mean(subset(readProPartComKM8, clusterAssignKM8==7)$Hypertension)
mean(subset(readProPartComKM8, clusterAssignKM8==7)$KidneyDisease)
mean(subset(readProPartComKM8, clusterAssignKM8==7)$BMI)

#Get the average statistics of cluster 8
dim(subset(readProPartComKM8, clusterAssignKM8==8))[1]
mean(subset(readProPartComKM8, clusterAssignKM8==8)$Age)
mean(subset(readProPartComKM8, clusterAssignKM8==8)$Distance)
mean(subset(readProPartComKM8, clusterAssignKM8==8)$Male)
mean(subset(readProPartComKM8, clusterAssignKM8==8)$Hypertension)
mean(subset(readProPartComKM8, clusterAssignKM8==8)$KidneyDisease)
mean(subset(readProPartComKM8, clusterAssignKM8==8)$BMI)

#Hierarchical clustering with complete linkage
hc.complete = hclust(dist, method="complete")
dev.new(width=20, height=10)
plot(hc.complete, main="Complete Linkage", xlab=" ", sub=" ", cex=1.2)

# If we use 5 clusters
hiClusterAssignCom5 = cutree(hc.complete, 5)
readProPartComHi5 = data.frame(readProPart, hiClusterAssignCom5)

#Get the average statistics of cluster 1
dim(subset(readProPartComHi5, hiClusterAssignCom5==1))[1]
mean(subset(readProPartComHi5, hiClusterAssignCom5==1)$Age)
mean(subset(readProPartComHi5, hiClusterAssignCom5==1)$Distance)
mean(subset(readProPartComHi5, hiClusterAssignCom5==1)$Male)
mean(subset(readProPartComHi5, hiClusterAssignCom5==1)$Hypertension)
mean(subset(readProPartComHi5, hiClusterAssignCom5==1)$KidneyDisease)
mean(subset(readProPartComHi5, hiClusterAssignCom5==1)$BMI)

#Get the average statistics of cluster 2
dim(subset(readProPartComHi5, hiClusterAssignCom5==2))[1]
mean(subset(readProPartComHi5, hiClusterAssignCom5==2)$Age)
mean(subset(readProPartComHi5, hiClusterAssignCom5==2)$Distance)
mean(subset(readProPartComHi5, hiClusterAssignCom5==2)$Male)
mean(subset(readProPartComHi5, hiClusterAssignCom5==2)$Hypertension)
mean(subset(readProPartComHi5, hiClusterAssignCom5==2)$KidneyDisease)
mean(subset(readProPartComHi5, hiClusterAssignCom5==2)$BMI)
#Get the average statistics of cluster 3
dim(subset(readProPartComHi5, hiClusterAssignCom5==3))[1]
mean(subset(readProPartComHi5, hiClusterAssignCom5==3)$Age)
mean(subset(readProPartComHi5, hiClusterAssignCom5==3)$Distance)
mean(subset(readProPartComHi5, hiClusterAssignCom5==3)$Male)
mean(subset(readProPartComHi5, hiClusterAssignCom5==3)$Hypertension)
mean(subset(readProPartComHi5, hiClusterAssignCom5==3)$KidneyDisease)
mean(subset(readProPartComHi5, hiClusterAssignCom5==3)$BMI)
#Get the average statistics of cluster 4
dim(subset(readProPartComHi5, hiClusterAssignCom5==4))[1]
mean(subset(readProPartComHi5, hiClusterAssignCom5==4)$Age)
mean(subset(readProPartComHi5, hiClusterAssignCom5==4)$Distance)
mean(subset(readProPartComHi5, hiClusterAssignCom5==4)$Male)
mean(subset(readProPartComHi5, hiClusterAssignCom5==4)$Hypertension)
mean(subset(readProPartComHi5, hiClusterAssignCom5==4)$KidneyDisease)
mean(subset(readProPartComHi5, hiClusterAssignCom5==4)$BMI)
#Get the average statistics of cluster 5
dim(subset(readProPartComHi5, hiClusterAssignCom5==5))[1]
mean(subset(readProPartComHi5, hiClusterAssignCom5==5)$Age)
mean(subset(readProPartComHi5, hiClusterAssignCom5==5)$Distance)
mean(subset(readProPartComHi5, hiClusterAssignCom5==5)$Male)
mean(subset(readProPartComHi5, hiClusterAssignCom5==5)$Hypertension)
mean(subset(readProPartComHi5, hiClusterAssignCom5==5)$KidneyDisease)
mean(subset(readProPartComHi5, hiClusterAssignCom5==5)$BMI)
#If we use 6 clusters
hiClusterAssignCom6 = cutree(hc.complete, 6)
readProPartComHi6 = data.frame(readProPart, hiClusterAssignCom6)
#Get the average statistics of cluster 1
dim(subset(readProPartComHi6, hiClusterAssignCom6==1))[1]
mean(subset(readProPartComHi6, hiClusterAssignCom6==1)$Age)
mean(subset(readProPartComHi6, hiClusterAssignCom6==1)$Distance)
mean(subset(readProPartComHi6, hiClusterAssignCom6==1)$Male)
mean(subset(readProPartComHi6, hiClusterAssignCom6==1)$Hypertension)
mean(subset(readProPartComHi6, hiClusterAssignCom6==1)$KidneyDisease)
mean(subset(readProPartComHi6, hiClusterAssignCom6==1)$BMI)
#Get the average statistics of cluster 2
dim(subset(readProPartComHi6, hiClusterAssignCom6==2))[1]
mean(subset(readProPartComHi6, hiClusterAssignCom6==2)$Age)
mean(subset(readProPartComHi6, hiClusterAssignCom6==2)$Distance)
mean(subset(readProPartComHi6, hiClusterAssignCom6==2)$Male)
mean(subset(readProPartComHi6, hiClusterAssignCom6==2)$Hypertension)
mean(subset(readProPartComHi6, hiClusterAssignCom6==2)$KidneyDisease)
mean(subset(readProPartComHi6, hiClusterAssignCom6==2)$BMI)
#Get the average statistics of cluster 3
dim(subset(readProPartComHi6, hiClusterAssignCom6==3))[1]
mean(subset(readProPartComHi6, hiClusterAssignCom6==3)$Age)
mean(subset(readProPartComHi6, hiClusterAssignCom6==3)$Distance)
mean(subset(readProPartComHi6, hiClusterAssignCom6==3)$Male)
mean(subset(readProPartComHi6, hiClusterAssignCom6==3)$Hypertension)
mean(subset(readProPartComHi6, hiClusterAssignCom6==3)$KidneyDisease)
mean(subset(readProPartComHi6, hiClusterAssignCom6==3)$BMI)

# Get the average statistics of cluster 4
(dim(subset(readProPartComHi6, hiClusterAssignCom6==4))[1]
mean(subset(readProPartComHi6, hiClusterAssignCom6==4)$Age)
mean(subset(readProPartComHi6, hiClusterAssignCom6==4)$Distance)
mean(subset(readProPartComHi6, hiClusterAssignCom6==4)$Male)
mean(subset(readProPartComHi6, hiClusterAssignCom6==4)$Hypertension)
mean(subset(readProPartComHi6, hiClusterAssignCom6==4)$KidneyDisease)
mean(subset(readProPartComHi6, hiClusterAssignCom6==4)$BMI)

# Get the average statistics of cluster 5
(dim(subset(readProPartComHi6, hiClusterAssignCom6==5))[1]
mean(subset(readProPartComHi6, hiClusterAssignCom6==5)$Age)
mean(subset(readProPartComHi6, hiClusterAssignCom6==5)$Distance)
mean(subset(readProPartComHi6, hiClusterAssignCom6==5)$Male)
mean(subset(readProPartComHi6, hiClusterAssignCom6==5)$Hypertension)
mean(subset(readProPartComHi6, hiClusterAssignCom6==5)$KidneyDisease)
mean(subset(readProPartComHi6, hiClusterAssignCom6==5)$BMI)

# Get the average statistics of cluster 6
(dim(subset(readProPartComHi6, hiClusterAssignCom6==6))[1]
mean(subset(readProPartComHi6, hiClusterAssignCom6==6)$Age)
mean(subset(readProPartComHi6, hiClusterAssignCom6==6)$Distance)
mean(subset(readProPartComHi6, hiClusterAssignCom6==6)$Male)
mean(subset(readProPartComHi6, hiClusterAssignCom6==6)$Hypertension)
mean(subset(readProPartComHi6, hiClusterAssignCom6==6)$KidneyDisease)
mean(subset(readProPartComHi6, hiClusterAssignCom6==6)$BMI)

# If we use 7 clusters
hiClusterAssignCom7 = cutree(hc.complete, 7)
readProPartComHi7 = data.frame(readProPart, hiClusterAssignCom7)

# Get the average statistics of cluster 1
(dim(subset(readProPartComHi7, hiClusterAssignCom7==1))[1]
mean(subset(readProPartComHi7, hiClusterAssignCom7==1)$Age)
mean(subset(readProPartComHi7, hiClusterAssignCom7==1)$Distance)
mean(subset(readProPartComHi7, hiClusterAssignCom7==1)$Male)
mean(subset(readProPartComHi7, hiClusterAssignCom7==1)$Hypertension)
mean(subset(readProPartComHi7, hiClusterAssignCom7==1)$KidneyDisease)
mean(subset(readProPartComHi7, hiClusterAssignCom7==1)$BMI)

# Get the average statistics of cluster 2
(dim(subset(readProPartComHi7, hiClusterAssignCom7==2))[1]
mean(subset(readProPartComHi7, hiClusterAssignCom7==2)$Age)
mean(subset(readProPartComHi7, hiClusterAssignCom7==2)$Distance)
mean(subset(readProPartComHi7, hiClusterAssignCom7==2)$Male)
mean(subset(readProPartComHi7, hiClusterAssignCom7==2)$Hypertension)
mean(subset(readProPartComHi7, hiClusterAssignCom7==2)$KidneyDisease)
mean(subset(readProPartComHi7, hiClusterAssignCom7==2)$BMI)

# Get the average statistics of cluster 3
(dim(subset(readProPartComHi7, hiClusterAssignCom7==3))[1]
mean(subset(readProPartComHi7, hiClusterAssignCom7==3)$Age)
mean(subset(readProPartComHi7, hiClusterAssignCom7==3)$Distance)
mean(subset(readProPartComHi7, hiClusterAssignCom7==3)$Male)
mean(subset(readProPartComHi7, hiClusterAssignCom7==3)$Hypertension)
mean(subset(readProPartComHi7, hiClusterAssignCom7==3)$KidneyDisease)
mean(subset(readProPartComHi7, hiClusterAssignCom7==3)$BMI)

#Get the average statistics of cluster 4
dim(subset(readProPartComHi7, hiClusterAssignCom7==4))[[1]]
mean(subset(readProPartComHi7, hiClusterAssignCom7==4)$Age)
mean(subset(readProPartComHi7, hiClusterAssignCom7==4)$Distance)
mean(subset(readProPartComHi7, hiClusterAssignCom7==4)$Male)
mean(subset(readProPartComHi7, hiClusterAssignCom7==4)$Hypertension)
mean(subset(readProPartComHi7, hiClusterAssignCom7==4)$KidneyDisease)
mean(subset(readProPartComHi7, hiClusterAssignCom7==4)$BMI)

#Get the average statistics of cluster 5
dim(subset(readProPartComHi7, hiClusterAssignCom7==5))[[1]]
mean(subset(readProPartComHi7, hiClusterAssignCom7==5)$Age)
mean(subset(readProPartComHi7, hiClusterAssignCom7==5)$Distance)
mean(subset(readProPartComHi7, hiClusterAssignCom7==5)$Male)
mean(subset(readProPartComHi7, hiClusterAssignCom7==5)$Hypertension)
mean(subset(readProPartComHi7, hiClusterAssignCom7==5)$KidneyDisease)
mean(subset(readProPartComHi7, hiClusterAssignCom7==5)$BMI)

#Get the average statistics of cluster 6
dim(subset(readProPartComHi7, hiClusterAssignCom7==6))[[1]]
mean(subset(readProPartComHi7, hiClusterAssignCom7==6)$Age)
mean(subset(readProPartComHi7, hiClusterAssignCom7==6)$Distance)
mean(subset(readProPartComHi7, hiClusterAssignCom7==6)$Male)
mean(subset(readProPartComHi7, hiClusterAssignCom7==6)$Hypertension)
mean(subset(readProPartComHi7, hiClusterAssignCom7==6)$KidneyDisease)
mean(subset(readProPartComHi7, hiClusterAssignCom7==6)$BMI)

#Get the average statistics of cluster 7
dim(subset(readProPartComHi7, hiClusterAssignCom7==7))[[1]]
mean(subset(readProPartComHi7, hiClusterAssignCom7==7)$Age)
mean(subset(readProPartComHi7, hiClusterAssignCom7==7)$Distance)
mean(subset(readProPartComHi7, hiClusterAssignCom7==7)$Male)
mean(subset(readProPartComHi7, hiClusterAssignCom7==7)$Hypertension)
mean(subset(readProPartComHi7, hiClusterAssignCom7==7)$KidneyDisease)
mean(subset(readProPartComHi7, hiClusterAssignCom7==7)$BMI)

#If we use 8 clusters
hiClusterAssignCom8 = cutree(hc.complete, 8)
readProPartComHi8 = data.frame(readProPart, hiClusterAssignCom8)

#Get the average statistics of cluster 1
dim(subset(readProPartComHi8, hiClusterAssignCom8==1))[[1]]
mean(subset(readProPartComHi8, hiClusterAssignCom8==1)$Age)
mean(subset(readProPartComHi8, hiClusterAssignCom8==1)$Distance)
mean(subset(readProPartComHi8, hiClusterAssignCom8==1)$Male)
mean(subset(readProPartComHi8, hiClusterAssignCom8==1)$Hypertension)
mean(subset(readProPartComHi8, hiClusterAssignCom8==1)$KidneyDisease)
mean(subset(readProPartComHi8, hiClusterAssignCom8==1)$BMI)
# Get the average statistics of cluster 2
```r
dim(subset(readProPartComHi8, hiClusterAssignCom8==2))[1]
mean(subset(readProPartComHi8, hiClusterAssignCom8==2)$Age)
mean(subset(readProPartComHi8, hiClusterAssignCom8==2)$Distance)
mean(subset(readProPartComHi8, hiClusterAssignCom8==2)$Male)
mean(subset(readProPartComHi8, hiClusterAssignCom8==2)$Hypertension)
mean(subset(readProPartComHi8, hiClusterAssignCom8==2)$KidneyDisease)
mean(subset(readProPartComHi8, hiClusterAssignCom8==2)$BMI)
```

# Get the average statistics of cluster 3
```r
dim(subset(readProPartComHi8, hiClusterAssignCom8==3))[1]
mean(subset(readProPartComHi8, hiClusterAssignCom8==3)$Age)
mean(subset(readProPartComHi8, hiClusterAssignCom8==3)$Distance)
mean(subset(readProPartComHi8, hiClusterAssignCom8==3)$Male)
mean(subset(readProPartComHi8, hiClusterAssignCom8==3)$Hypertension)
mean(subset(readProPartComHi8, hiClusterAssignCom8==3)$KidneyDisease)
mean(subset(readProPartComHi8, hiClusterAssignCom8==3)$BMI)
```

# Get the average statistics of cluster 4
```r
dim(subset(readProPartComHi8, hiClusterAssignCom8==4))[1]
mean(subset(readProPartComHi8, hiClusterAssignCom8==4)$Age)
mean(subset(readProPartComHi8, hiClusterAssignCom8==4)$Distance)
mean(subset(readProPartComHi8, hiClusterAssignCom8==4)$Male)
mean(subset(readProPartComHi8, hiClusterAssignCom8==4)$Hypertension)
mean(subset(readProPartComHi8, hiClusterAssignCom8==4)$KidneyDisease)
mean(subset(readProPartComHi8, hiClusterAssignCom8==4)$BMI)
```

# Get the average statistics of cluster 5
```r
dim(subset(readProPartComHi8, hiClusterAssignCom8==5))[1]
mean(subset(readProPartComHi8, hiClusterAssignCom8==5)$Age)
mean(subset(readProPartComHi8, hiClusterAssignCom8==5)$Distance)
mean(subset(readProPartComHi8, hiClusterAssignCom8==5)$Male)
mean(subset(readProPartComHi8, hiClusterAssignCom8==5)$Hypertension)
mean(subset(readProPartComHi8, hiClusterAssignCom8==5)$KidneyDisease)
mean(subset(readProPartComHi8, hiClusterAssignCom8==5)$BMI)
```

# Get the average statistics of cluster 6
```r
dim(subset(readProPartComHi8, hiClusterAssignCom8==6))[1]
mean(subset(readProPartComHi8, hiClusterAssignCom8==6)$Age)
mean(subset(readProPartComHi8, hiClusterAssignCom8==6)$Distance)
mean(subset(readProPartComHi8, hiClusterAssignCom8==6)$Male)
mean(subset(readProPartComHi8, hiClusterAssignCom8==6)$Hypertension)
mean(subset(readProPartComHi8, hiClusterAssignCom8==6)$KidneyDisease)
mean(subset(readProPartComHi8, hiClusterAssignCom8==6)$BMI)
```

# Get the average statistics of cluster 7
```r
dim(subset(readProPartComHi8, hiClusterAssignCom8==7))[1]
mean(subset(readProPartComHi8, hiClusterAssignCom8==7)$Age)
mean(subset(readProPartComHi8, hiClusterAssignCom8==7)$Distance)
mean(subset(readProPartComHi8, hiClusterAssignCom8==7)$Male)
mean(subset(readProPartComHi8, hiClusterAssignCom8==7)$Hypertension)
mean(subset(readProPartComHi8, hiClusterAssignCom8==7)$KidneyDisease)
mean(subset(readProPartComHi8, hiClusterAssignCom8==7)$BMI)
```
# Get the average statistics of cluster 8

dim(subset(readProPartComHi8, hiClusterAssignCom8==8))[[1]]
mean(subset(readProPartComHi8, hiClusterAssignCom8==8)$Age)
mean(subset(readProPartComHi8, hiClusterAssignCom8==8)$Distance)
mean(subset(readProPartComHi8, hiClusterAssignCom8==8)$Male)
mean(subset(readProPartComHi8, hiClusterAssignCom8==8)$Hypertension)
mean(subset(readProPartComHi8, hiClusterAssignCom8==8)$KidneyDisease)
mean(subset(readProPartComHi8, hiClusterAssignCom8==8)$BMI)

# Split the whole data set into the 6 clusters determined by k-means clustering
readWhole = data.frame(readPro, clusterAssignKM6)
cluster1with6 = subset(readWhole, clusterAssignKM6==1)
cluster2with6 = subset(readWhole, clusterAssignKM6==2)
cluster3with6 = subset(readWhole, clusterAssignKM6==3)
cluster4with6 = subset(readWhole, clusterAssignKM6==4)
cluster5with6 = subset(readWhole, clusterAssignKM6==5)
cluster6with6 = subset(readWhole, clusterAssignKM6==6)
dim(subset(cluster1with6, Unplanned==1))[[1]]
dim(subset(cluster2with6, Unplanned==1))[[1]]
dim(subset(cluster3with6, Unplanned==1))[[1]]
dim(subset(cluster4with6, Unplanned==1))[[1]]
dim(subset(cluster5with6, Unplanned==1))[[1]]
dim(subset(cluster6with6, Unplanned==1))[[1]]
dim(subset(cluster1with6, DischargeToRehabOrSnf==1))[[1]]
dim(subset(cluster2with6, DischargeToRehabOrSnf==1))[[1]]
dim(subset(cluster3with6, DischargeToRehabOrSnf==1))[[1]]
dim(subset(cluster4with6, DischargeToRehabOrSnf==1))[[1]]
dim(subset(cluster5with6, DischargeToRehabOrSnf==1))[[1]]
dim(subset(cluster6with6, DischargeToRehabOrSnf==1))[[1]]
dim(subset(cluster1with6, DischargeToHomeHealth==1))[[1]]
dim(subset(cluster2with6, DischargeToHomeHealth==1))[[1]]
dim(subset(cluster3with6, DischargeToHomeHealth==1))[[1]]
dim(subset(cluster4with6, DischargeToHomeHealth==1))[[1]]
dim(subset(cluster5with6, DischargeToHomeHealth==1))[[1]]
dim(subset(cluster6with6, DischargeToHomeHealth==1))[[1]]
Appendix E

Comparison of results of K-means clustering and hierarchical clustering

Scenario 1: k=5

Table E-1. K-means clustering results with 5 clusters

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Size</th>
<th>Average age (year)</th>
<th>Percentage of male</th>
<th>Percentage of hypertension</th>
<th>Percentage of kidney disease</th>
<th>Average BMI (kg/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>302</td>
<td>60.12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30.16</td>
</tr>
<tr>
<td>2</td>
<td>221</td>
<td>54.18</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>30.05</td>
</tr>
<tr>
<td>3</td>
<td>333</td>
<td>64.55</td>
<td>100</td>
<td>100</td>
<td>2.40</td>
<td>30.95</td>
</tr>
<tr>
<td>4</td>
<td>414</td>
<td>70.41</td>
<td>0</td>
<td>100</td>
<td>2.90</td>
<td>30.28</td>
</tr>
<tr>
<td>5</td>
<td>77</td>
<td>75.30</td>
<td>51.95</td>
<td>2.60</td>
<td>100</td>
<td>30.95</td>
</tr>
</tbody>
</table>

Table E-2. Hierarchical clustering results with 5 clusters

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Size</th>
<th>Average age (year)</th>
<th>Percentage of male</th>
<th>Percentage of hypertension</th>
<th>Percentage of kidney disease</th>
<th>Average BMI (kg/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>523</td>
<td>57.61</td>
<td>42.26</td>
<td>0</td>
<td>0</td>
<td>30.11</td>
</tr>
<tr>
<td>2</td>
<td>402</td>
<td>70.29</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>30.19</td>
</tr>
<tr>
<td>3</td>
<td>325</td>
<td>64.58</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>30.97</td>
</tr>
<tr>
<td>4</td>
<td>48</td>
<td>71.42</td>
<td>100</td>
<td>18.75</td>
<td>100</td>
<td>30.27</td>
</tr>
<tr>
<td>5</td>
<td>49</td>
<td>76.90</td>
<td>0</td>
<td>26.53</td>
<td>100</td>
<td>31.96</td>
</tr>
</tbody>
</table>

With 5 clusters, K-means clustering algorithm and hierarchical clustering algorithm produced no identical cluster.
Scenario 2: k=6

Table E-3. K-means clustering results with 6 clusters

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Size</th>
<th>Average age (year)</th>
<th>Percentage of male</th>
<th>Percentage of hypertension</th>
<th>Percentage of kidney disease</th>
<th>Average BMI (kg/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>325</td>
<td>64.58</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>30.97</td>
</tr>
<tr>
<td>2</td>
<td>221</td>
<td>54.18</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>30.05</td>
</tr>
<tr>
<td>3</td>
<td>302</td>
<td>60.12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30.16</td>
</tr>
<tr>
<td>4</td>
<td>402</td>
<td>70.29</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>30.19</td>
</tr>
<tr>
<td>5</td>
<td>75</td>
<td>75.47</td>
<td>52</td>
<td>0</td>
<td>100</td>
<td>30.55</td>
</tr>
<tr>
<td>6</td>
<td>22</td>
<td>69.82</td>
<td>40.91</td>
<td>100</td>
<td>100</td>
<td>33.09</td>
</tr>
</tbody>
</table>

Table E-4. Hierarchical clustering results with 6 clusters

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Size</th>
<th>Average age (year)</th>
<th>Percentage of male</th>
<th>Percentage of hypertension</th>
<th>Percentage of kidney disease</th>
<th>Average BMI (kg/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>325</td>
<td>64.58</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>30.97</td>
</tr>
<tr>
<td>2</td>
<td>221</td>
<td>54.18</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>30.05</td>
</tr>
<tr>
<td>3</td>
<td>302</td>
<td>60.12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30.16</td>
</tr>
<tr>
<td>4</td>
<td>402</td>
<td>70.29</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>30.19</td>
</tr>
<tr>
<td>5</td>
<td>48</td>
<td>71.42</td>
<td>100</td>
<td>18.75</td>
<td>100</td>
<td>30.27</td>
</tr>
<tr>
<td>6</td>
<td>49</td>
<td>76.90</td>
<td>0</td>
<td>26.53</td>
<td>100</td>
<td>31.96</td>
</tr>
</tbody>
</table>

With 6 clusters, cluster 1, 2, 3 and 4 produced by these two algorithms are identical, respectively.
Scenario 3: k=7

Table E-5. K-means clustering results with 7 clusters

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Size</th>
<th>Average age (year)</th>
<th>Percentage of male</th>
<th>Percentage of hypertension</th>
<th>Percentage of kidney disease</th>
<th>Average BMI (kg/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>325</td>
<td>64.58</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>30.97</td>
</tr>
<tr>
<td>2</td>
<td>221</td>
<td>54.18</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>30.05</td>
</tr>
<tr>
<td>3</td>
<td>302</td>
<td>60.12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30.16</td>
</tr>
<tr>
<td>4</td>
<td>226</td>
<td>72.79</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>30.57</td>
</tr>
<tr>
<td>5</td>
<td>176</td>
<td>67.07</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>29.71</td>
</tr>
<tr>
<td>6</td>
<td>75</td>
<td>75.47</td>
<td>52</td>
<td>0</td>
<td>100</td>
<td>30.55</td>
</tr>
<tr>
<td>7</td>
<td>22</td>
<td>69.82</td>
<td>40.91</td>
<td>100</td>
<td>100</td>
<td>33.09</td>
</tr>
</tbody>
</table>

Table E-6. Hierarchical clustering results with 7 clusters

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Size</th>
<th>Average age (year)</th>
<th>Percentage of male</th>
<th>Percentage of hypertension</th>
<th>Percentage of kidney disease</th>
<th>Average BMI (kg/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>325</td>
<td>64.58</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>30.97</td>
</tr>
<tr>
<td>2</td>
<td>221</td>
<td>54.18</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>30.05</td>
</tr>
<tr>
<td>3</td>
<td>302</td>
<td>60.12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30.16</td>
</tr>
<tr>
<td>4</td>
<td>402</td>
<td>70.29</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>30.19</td>
</tr>
<tr>
<td>5</td>
<td>49</td>
<td>76.90</td>
<td>0</td>
<td>26.53</td>
<td>100</td>
<td>31.96</td>
</tr>
<tr>
<td>6</td>
<td>39</td>
<td>73.82</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>29.63</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>61.00</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>33.06</td>
</tr>
</tbody>
</table>

With 7 clusters, cluster 1, 2, and 3 produced by these two algorithms are identical, respectively.
Scenario 4: k=8

Table E-7. K-means clustering results with 8 clusters

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Size</th>
<th>Average age (year)</th>
<th>Percentage of male</th>
<th>Percentage of hypertension</th>
<th>Percentage of kidney disease</th>
<th>Average BMI (kg/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>325</td>
<td>64.58</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>30.97</td>
</tr>
<tr>
<td>2</td>
<td>221</td>
<td>54.18</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>30.05</td>
</tr>
<tr>
<td>3</td>
<td>302</td>
<td>60.12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30.16</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>77.25</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>31.55</td>
</tr>
<tr>
<td>5</td>
<td>176</td>
<td>67.07</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>29.71</td>
</tr>
<tr>
<td>6</td>
<td>225</td>
<td>72.73</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>30.40</td>
</tr>
<tr>
<td>7</td>
<td>48</td>
<td>71.42</td>
<td>100</td>
<td>18.75</td>
<td>100</td>
<td>30.27</td>
</tr>
<tr>
<td>8</td>
<td>14</td>
<td>76.71</td>
<td>0</td>
<td>100</td>
<td>92.86</td>
<td>35.68</td>
</tr>
</tbody>
</table>

Table E-8. Hierarchical clustering results with 8 clusters

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Size</th>
<th>Average age (year)</th>
<th>Percentage of male</th>
<th>Percentage of hypertension</th>
<th>Percentage of kidney disease</th>
<th>Average BMI (kg/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>325</td>
<td>64.58</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>30.97</td>
</tr>
<tr>
<td>2</td>
<td>221</td>
<td>54.18</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>30.05</td>
</tr>
<tr>
<td>3</td>
<td>302</td>
<td>60.12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30.16</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>77.25</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>31.55</td>
</tr>
<tr>
<td>5</td>
<td>402</td>
<td>70.29</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>30.19</td>
</tr>
<tr>
<td>6</td>
<td>39</td>
<td>73.82</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>29.63</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>61.00</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>33.06</td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>75.92</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>33.12</td>
</tr>
</tbody>
</table>

With 8 clusters, cluster 1, 2, 3, and 4 produced by these two algorithms are identical, respectively.
With \( k=5, 6, 7 \) and 8, the number of identical clusters produced by these two algorithms are shown in Table E-9.

<table>
<thead>
<tr>
<th>( K )</th>
<th>Number of identical clusters</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>66.67</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>42.86</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>50</td>
</tr>
</tbody>
</table>

Table E-9 indicates that \( k=6 \) leads to the highest consistency in the clustering results by \( K \)-means clustering algorithm and hierarchical clustering algorithm. Therefore, we used the clustering results with \( k=6 \) in the subsequent cost-effectiveness analysis in Chapter 3.
Appendix F

R code for Markov simulation model of cluster 1 patients

**Surgery option:**

```r
set.seed(1);
complete = 100;
ndr = c(0.012586, 0.013763, 0.015057, 0.016380, 0.017756, 0.019299, 0.021039, 0.022997, 0.025182,
        0.027634, 0.030322, 0.033309, 0.036740, 0.040688, 0.045172, 0.050072, 0.055306, 0.061241, 0.067893,
        0.075594, 0.084649, 0.094437, 0.105152, 0.116835, 0.129516, 0.143215, 0.157937, 0.173671, 0.190385,
        0.208029, 0.226531, 0.245796, 0.265711, 0.286142, 0.306941, 1.000000);
totalCost = 30025.92;
totalQoL = 0.871;
cost = 0.0;
qoL = 0.0;
k=0;
curState = 0;
nextState = 0;
for(i in 64: complete){
    if(curState == 0){
        randomNum0 = runif(1,0,1);

        if(randomNum0 <= 0.014568){
            nextState = 7;
            cost = 0.0;
            qoL = 0.0;
            k=i;
            break;
        }
        else if((randomNum0 > 0.014568) && (randomNum0 <= 0.085368)){
            nextState = 2;
            cost = 15333.00;
            qoL = 0.596;
            totalCost = totalCost + cost;
            totalQoL = totalQoL + qoL;
        }
        else{
            nextState = 1;
            cost = 1358.00;
            qoL = 0.871;
            totalCost = totalCost + cost;
            totalQoL = totalQoL + qoL;
        }
    }
}
```

148
currentState = nextState;
}
else if(currentState == 1){
    randomNum1 = runif(1,0,1);

    if(randomNum1 <= ndr[i-64]){ 
        nextState = 7;
        cost = 0.0;
        qoL = 0.0;
        k=i;
        break;
    }
    else if((randomNum1 > ndr[i-64]) && (randomNum1 <= (ndr[i-64]+0.009))){
        nextState = 3;
        if(i <= 74){
            cost = 39963.92;
            qoL = 0.596;
            totalCost = totalCost + cost;
            totalQoL = totalQoL + qoL;
        }
        else{
            cost = 39963.92;
            qoL = 0.562;
            totalCost = totalCost + cost;
            totalQoL = totalQoL + qoL;
        }
    }
    else{
        nextState = 1;
        if(i <= 74){
            cost = 1358.00;
            qoL = 0.871;
            totalCost = totalCost + cost;
            totalQoL = totalQoL + qoL;
        }
        else{
            cost = 1358.00;
            qoL = 0.837;
            totalCost = totalCost + cost;
            totalQoL = totalQoL + qoL;
        }
    }
currentState = nextState;

else if(currentState == 2){
    randomNum2 = runif(1,0,1);

    if(randomNum2 <= 0.039){
        nextState = 3;

        if(i <= 74){
            cost = 39963.92;
            qoL = 0.596;
            totalCost = totalCost + cost;
            totalQoL = totalQoL + qoL;

        }
        else{
            cost = 39963.92;
            qoL = 0.562;
            totalCost = totalCost + cost;
            totalQoL = totalQoL + qoL;

        }
    }
    else{
        nextState = 1;

        if(i <= 74){
            cost = 1358.00;
            qoL = 0.871;
            totalCost = totalCost + cost;
            totalQoL = totalQoL + qoL;

        }
        else{
            cost = 1358.00;
            qoL = 0.837;
            totalCost = totalCost + cost;
            totalQoL = totalQoL + qoL;

        }
    }
}

currentState = nextState;
else if(currentState == 3) {
    randomNum3 = runif(1,0,1);

    if(randomNum3 <= 0.026) {
        nextState = 7;
        cost = 0.0;
        qoL = 0.0;
        k=i;
        break;
    }

    else if((randomNum3 > 0.026) && (randomNum3 <= 0.086)) {
        nextState = 5;
        if(i <= 74) {
            cost = 39963.92;
            qoL = 0.596;
            totalCost = totalCost + cost;
            totalQoL = totalQoL + qoL;
        } else {
            cost = 39963.92;
            qoL = 0.562;
            totalCost = totalCost + cost;
            totalQoL = totalQoL + qoL;
        }
    }

    else {
        nextState = 4;
        if(i <= 74) {
            cost = 1358.00;
            qoL = 0.871;
            totalCost = totalCost + cost;
            totalQoL = totalQoL + qoL;
        } else {
            cost = 1358.00;
            qoL = 0.837;
            totalCost = totalCost + cost;
            totalQoL = totalQoL + qoL;
        }
    }
}

}
currentState = nextState;
}
else if(currentState == 4){
    randomNum4 = runif(1,0,1);

    if(randomNum4 <= ndr[i-64]){  
        nextState = 7;
        cost = 0.0;
        qoL = 0.0;
        k=i;
        break;
    }
}
else if((randomNum4 > ndr[i-64]) && (randomNum4 <= (ndr[i-64]+0.04))){
    nextState = 5;

    if(i <= 74){
        cost = 39963.92;
        qoL = 0.596;
        totalCost = totalCost + cost;
        totalQoL = totalQoL + qoL;
    }
    else{
        cost = 39963.92;
        qoL = 0.562;
        totalCost = totalCost + cost;
        totalQoL = totalQoL + qoL;
    }
}
else{
    nextState = 4;

    if(i <= 74){
        cost = 1358.00;
        qoL = 0.871;
        totalCost = totalCost + cost;
        totalQoL = totalQoL + qoL;
    }
    else{
        cost = 1358.00;
        qoL = 0.837;
        totalCost = totalCost + cost;
        totalQoL = totalQoL + qoL;
    }
}
else if(currentState == 5){
    randomNum5 = runif(1,0,1);
    if(randomNum5 <= 0.05){
        nextState = 7;
        cost = 0.0;
        qoL = 0.0;
        k=i;
        break;
    }
    else {
        nextState = 6;
        if(i <= 74){
            cost = 1358.00;
            qoL = 0.871;
            totalCost = totalCost + cost;
            totalQoL = totalQoL + qoL;
        }
        else{
            cost = 1358.00;
            qoL = 0.837;
            totalCost = totalCost + cost;
            totalQoL = totalQoL + qoL;
        }
    }
}
currentState = nextState;
}
else if(currentState == 6){
    randomNum6 = runif(1,0,1);
    if(randomNum6 <= ndr[i-64]){  
        nextState = 7;
        cost = 0.0;
        qoL = 0.0;
        k=i;
    }
}
break;

}
else{
nextState = 6;

if(i <= 74){
cost = 1358.00;
qoL = 0.871;
totalCost = totalCost + cost;
totalQoL = totalQoL + qoL;
}
else{
cost = 1358.00;
qoL = 0.837;
totalCost = totalCost + cost;
totalQoL = totalQoL + qoL;
}
}

currentState = nextState;
}

print(totalCost);
print(totalQoL);
print(k);

**Non-surgery option:**
set.seed(1);
complete = 100;
ndr = c(0.011568, 0.012586, 0.013763, 0.015057, 0.016380, 0.017756, 0.019299, 0.021039, 0.022997,
0.025182, 0.027634, 0.030322, 0.033309, 0.036740, 0.040688, 0.045172, 0.050072, 0.055306, 0.061241,
0.067893, 0.075594, 0.084649, 0.094437, 0.105152, 0.116835, 0.129516, 0.143215, 0.157937, 0.173671,
0.190385, 0.208029, 0.226531, 0.245796, 0.265711, 0.286142, 0.306941, 1.000000);
totalCost = 3332.00;
totalQoL = 0.770;
cost = 0.0;
qoL = 0.0;
k=0;

currentState = 0;
nextState = 0;
for(i in 64: complete){
    if(currentState == 0){
        randomNum0 = runif(1,0,1);
        if(randomNum0 <= ndr[i-63]){  
            nextState = 1;
            cost = 0.0;
            qoL = 0.0;
            k=i;
            break;
        }  
        else{
            nextState = 0;
            if(i <= 74){
                cost = 3332.00;
                qoL = 0.770;
                totalCost = totalCost + cost;
                totalQoL = totalQoL + qoL;
            }  
            else{
                cost = 3332.00;
                qoL = 0.736;
                totalCost = totalCost + cost;
                totalQoL = totalQoL + qoL;
            }
        }
    }
    currentState = nextState;
}
print(totalCost);
print(totalQoL);
print(k);
Appendix G

R code for meta-analysis

install.packages("meta")
library(meta)

#WOMAC pain before surgery
n.e.womac.pain.befoSurgery = c(11, 25, 15)
mean.e.womac.pain.befoSurgery = c(8.0, 7.8, 10.2)
sd.e.womac.pain.befoSurgery = c(3.8, 4.1, 2.7)
n.c.womac.pain.befoSurgery = c(12, 24, 20)
mean.c.womac.pain.befoSurgery = c(11.0, 9.9, 10.3)
sd.c.womac.pain.befoSurgery = c(3.6, 2.9, 4.1)
meta.womac.pain.befoSurgery = metacont(n.e.womac.pain.befoSurgery,
mean.e.womac.pain.befoSurgery, sd.e.womac.pain.befoSurgery, n.c.womac.pain.befoSurgery,
mean.c.womac.pain.befoSurgery, sd.c.womac.pain.befoSurgery, sm="MD")
summary(meta.womac.pain.befoSurgery)
dev.new(width=20, height=10)
forest(meta.womac.pain.befoSurgery, comb.fixed=FALSE)

#WOMAC pain at three months after surgery
n.e.womac.pain.twoOrThreeMonAft = c(11, 25, 15)
mean.e.womac.pain.twoOrThreeMonAft = c(1.7, 2.6, 2.7)
sd.e.womac.pain.twoOrThreeMonAft = c(2.35, 2.6, 2.1)
n.c.womac.pain.twoOrThreeMonAft = c(12, 24, 20)
mean.c.womac.pain.twoOrThreeMonAft = c(2.2, 2.7, 0.05)
sd.c.womac.pain.twoOrThreeMonAft = c(1.75, 2.0, 4.4)
meta.womac.pain.twoOrThreeMonAft = metacont(n.e.womac.pain.twoOrThreeMonAft,
mean.e.womac.pain.twoOrThreeMonAft, sd.e.womac.pain.twoOrThreeMonAft,
n.c.womac.pain.twoOrThreeMonAft, mean.c.womac.pain.twoOrThreeMonAft,
sd.c.womac.pain.twoOrThreeMonAft, sm="MD")
summary(meta.womac.pain.twoOrThreeMonAft)
dev.new(width=20, height=10)
forest(meta.womac.pain.twoOrThreeMonAft, comb.fixed=FALSE)
#Jackknife sensivity analysis after removing the third study
n.e.womac.pain.twoOrThreeMonAft.jack = c(11, 25)
mean.e.womac.pain.twoOrThreeMonAft.jack = c(1.7, 2.6)
sd.e.womac.pain.twoOrThreeMonAft.jack = c(2.35, 2.6)
n.c.womac.pain.twoOrThreeMonAft.jack = c(12, 24)
mean.c.womac.pain.twoOrThreeMonAft.jack = c(2.2, 2.7)
sd.c.womac.pain.twoOrThreeMonAft.jack = c(1.75, 2.0)
meta.womac.pain.twoOrThreeMonAft.jack = metacont(n.e.womac.pain.twoOrThreeMonAft.jack,
mean.e.womac.pain.twoOrThreeMonAft.jack, sd.e.womac.pain.twoOrThreeMonAft.jack,
n.c.womac.pain.twoOrThreeMonAft.jack, mean.c.womac.pain.twoOrThreeMonAft.jack,
sd.c.womac.pain.twoOrThreeMonAft.jack, sm="MD")
summary(meta.womac.pain.twoOrThreeMonAft.jack)
dev.new(width=20, height=10)
forest(meta.womac.pain.twoOrThreeMonAft, comb.fixed=FALSE)

#WOMAC function before surgery
n.e.womac.function.befoSurgery = c(11, 25, 15)
mean.e.womac.function.befoSurgery = c(33.7, 26.9, 35.8)
sd.e.womac.function.befoSurgery = c(13.8, 11.9, 12.0)
n.c.womac.function.befoSurgery = c(12, 24, 20)
mean.c.womac.function.befoSurgery = c(43.5, 33.7, 41.0)
sd.c.womac.function.befoSurgery = c(9.5, 10.9, 10.0)
meta.womac.function.befoSurgery = metacont(n.e.womac.function.befoSurgery, mean.e.womac.function.befoSurgery, sd.e.womac.function.befoSurgery, n.c.womac.function.befoSurgery, mean.c.womac.function.befoSurgery, sd.c.womac.function.befoSurgery, sm="MD")
summary(meta.womac.function.befoSurgery)
dev.new(width=20, height=10)
forest(meta.womac.function.befoSurgery, comb.fixed=FALSE)

#WOMAC function at three months after surgery
n.e.womac.function.twoOrThreeMonAft = c(11, 25, 15)
mean.e.womac.function.twoOrThreeMonAft = c(18.3, 12.8, 15.9)
sd.e.womac.function.twoOrThreeMonAft = c(12.36, 9.0, 10.3)
n.c.womac.function.twoOrThreeMonAft = c(12, 24, 20)
mean.c.womac.function.twoOrThreeMonAft = c(28.5, 12.9, 18.4)
sd.c.womac.function.twoOrThreeMonAft = c(10.01, 8.0, 13.8)
meta.womac.function.twoOrThreeMonAft = metacont(n.e.womac.function.twoOrThreeMonAft, mean.e.womac.function.twoOrThreeMonAft, sd.e.womac.function.twoOrThreeMonAft, n.c.womac.function.twoOrThreeMonAft, mean.c.womac.function.twoOrThreeMonAft, sd.c.womac.function.twoOrThreeMonAft, sm="MD")
summary(meta.womac.function.twoOrThreeMonAft)
dev.new(width=20, height=10)
forest(meta.womac.function.twoOrThreeMonAft, comb.fixed=FALSE)

#Harris Hip Score before surgery
n.e.harris.befoSurgery = c(11, 29, 20, 15)
mean.e.harris.befoSurgery = c(43.6, 51.48, 44.0, 45.4)
sd.e.harris.befoSurgery = c(15.7, 18.32, 7.25, 11.5)
n.c.harris.befoSurgery = c(12, 30, 20, 20)
mean.c.harris.befoSurgery = c(34.9, 45.3, 45.75, 43.2)
sd.c.harris.befoSurgery = c(15.5, 12.98, 11.82, 16.2)
meta.harris.befoSurgery = metacont(n.e.harris.befoSurgery, mean.e.harris.befoSurgery, sd.e.harris.befoSurgery, n.c.harris.befoSurgery, mean.c.harris.befoSurgery, sd.c.harris.befoSurgery, sm="MD")
summary(meta.harris.befoSurgery)
dev.new(width=20, height=10)
forest(meta.harris.befoSurgery, comb.fixed=FALSE)

#Harris Hip Score at discharge
n.e.harris.atDischarge = c(29, 20, 15)
mean.e.harris.atDischarge = c(64.46, 51.25, 62.6)
sd.e.harris.atDischarge = c(6.92, 8.17, 7.9)
n.c.harris.atDischarge = c(30, 20, 20)
mean.c.harris.atDischarge = c(59.36, 50.1, 53.8)
sd.c.harris.atDischarge = c(6.82, 6.17, 12.0)
meta.harris.atDischarge = metacont(n.e.harris.atDischarge, mean.e.harris.atDischarge, 
sd.e.harris.atDischarge, n.c.harris.atDischarge, mean.c.harris.atDischarge, sd.c.harris.atDischarge, 
sm="MD")
summary(meta.harris.atDischarge)
dev.new(width=20, height=10)
forest(meta.harris.atDischarge, comb.fixed=FALSE)

# Harris Hip Score at three months after surgery
n.e.harris.threeMonAft = c(11, 29, 15)
mean.e.harris.threeMonAft = c(69.47, 85.3, 74.2)
sd.e.harris.threeMonAft = c(7.49, 11.78, 11.7)
n.c.harris.threeMonAft = c(12, 30, 20)
mean.c.harris.threeMonAft = c(65.2, 78.7, 68.8)
sd.c.harris.threeMonAft = c(15.4, 9.41, 16.2)
meta.harris.threeMonAft = metacont(n.e.harris.threeMonAft, mean.e.harris.threeMonAft, 
sd.e.harris.threeMonAft, n.c.harris.threeMonAft, mean.c.harris.threeMonAft, sd.c.harris.threeMonAft, 
sm="MD")
summary(meta.harris.threeMonAft)
dev.new(width=20, height=10)
forest(meta.harris.threeMonAft, comb.fixed=FALSE)

#HOOS activity of daily living before surgery
n.e.hoos.ADL.befoSurgery = c(40, 43, 10, 14)
mean.e.hoos.ADL.befoSurgery = c(59.9, 54.2, 51.1, 49.8)
sd.e.hoos.ADL.befoSurgery = c(17.1, 13.4, 10.5, 10.6)
n.c.hoos.ADL.befoSurgery = c(40, 41, 10, 12)
mean.c.hoos.ADL.befoSurgery = c(48.7, 46.1, 52.3, 46.3)
sd.e.hoos.ADL.befoSurgery = c(13.9, 16.2, 21.2, 14.9)
meta.hoos.ADL.befoSurgery = metacont(n.e.hoos.ADL.befoSurgery, mean.e.hoos.ADL.befoSurgery, 
sd.e.hoos.ADL.befoSurgery, n.c.hoos.ADL.befoSurgery, mean.c.hoos.ADL.befoSurgery, 
sd.c.hoos.ADL.befoSurgery, sm="MD")
summary(meta.hoos.ADL.befoSurgery)
dev.new(width=20, height=10)
forest(meta.hoos.ADL.befoSurgery, comb.fixed=FALSE)

#HOOS activity of daily living at six weeks after surgery
n.e.hoos.ADL.sixWeeksAft = c(43, 14)
mean.e.hoos.ADL.sixWeeksAft = c(77.7, 78.5)
sd.e.hoos.ADL.sixWeeksAft = c(18.2, 10.8)
n.c.hoos.ADL.sixWeeksAft = c(41, 12)
mean.c.hoos.ADL.sixWeeksAft = c(74.0, 75.0)
sd.e.hoos.ADL.sixWeeksAft = c(17.4, 13.3)
meta.hoos.ADL.sixWeeksAft = metacont(n.e.hoos.ADL.sixWeeksAft, mean.e.hoos.ADL.sixWeeksAft, 
sd.e.hoos.ADL.sixWeeksAft, n.c.hoos.ADL.sixWeeksAft, mean.c.hoos.ADL.sixWeeksAft, sd.c.hoos.ADL.sixWeeksAft, 
sm="MD")
summary(meta.hoos.ADL.sixWeeksAft)
dev.new(width=20, height=10)
forest(meta.hoos.ADL.sixWeeksAft, comb.fixed=FALSE)

#HOOS pain before surgery
n.e.hoos.pain.befoSurgery = c(40, 10, 14)
mean.e.hoos.pain.befoSurgery = c(55.4, 55.3, 48.2)
sd.e.hoos.pain.befoSurgery = c(16.9, 12.0, 9.2)
n.c.hoos.pain.befoSurgery = c(40, 10, 12)
mean.c.hoos.pain.befoSurgery = c(45.9, 49.3, 47.2)
sd.c.hoos.pain.befoSurgery = c(14.1, 17.0, 12.2)
meta.hoos.pain.befoSurgery = metacont(n.e.hoos.pain.befoSurgery, mean.e.hoos.pain.befoSurgery, 
sd.e.hoos.pain.befoSurgery, n.c.hoos.pain.befoSurgery, mean.c.hoos.pain.befoSurgery, 
sd.c.hoos.pain.befoSurgery, sm="MD")
summary(meta.hoos.pain.befoSurgery)
dev.new(width=20, height=10)
forest(meta.hoos.pain.befoSurgery, comb.fixed=FALSE)

#HOOS pain at six weeks after surgery
n.e.hoos.pain.sixWeeksAft = c(43, 14)
mean.e.hoos.pain.sixWeeksAft = c(82.2, 80.5)
sd.e.hoos.pain.sixWeeksAft = c(14.8, 8.6)
n.c.hoos.pain.sixWeeksAft = c(41, 12)
mean.c.hoos.pain.sixWeeksAft = c(79.1, 77.7)
sd.c.hoos.pain.sixWeeksAft = c(15.2, 16.8)
meta.hoos.pain.sixWeeksAft = metacont(n.e.hoos.pain.sixWeeksAft, mean.e.hoos.pain.sixWeeksAft, 
sd.e.hoos.pain.sixWeeksAft, n.c.hoos.pain.sixWeeksAft, mean.c.hoos.pain.sixWeeksAft, 
sd.c.hoos.pain.sixWeeksAft, sm="MD")
summary(meta.hoos.pain.sixWeeksAft)
dev.new(width=20, height=10)
forest(meta.hoos.pain.sixWeeksAft, comb.fixed=FALSE)

#HOOS symptoms before surgery
n.e.hoos.symptom.befoSurgery = c(40, 10, 14)
mean.e.hoos.symptom.befoSurgery = c(56.6, 56.5, 52.1)
sd.e.hoos.symptom.befoSurgery = c(19.8, 14.2, 12.2)
n.c.hoos.symptom.befoSurgery = c(40, 10, 12)
mean.c.hoos.symptom.befoSurgery = c(45.4, 59.0, 51.0)
sd.c.hoos.symptom.befoSurgery = c(16.7, 15.6, 18.3)
meta.hoos.symptom.befoSurgery = metacont(n.e.hoos.symptom.befoSurgery, 
mean.e.hoos.symptom.befoSurgery, sd.e.hoos.symptom.befoSurgery, n.c.hoos.symptom.befoSurgery, 
mean.c.hoos.symptom.befoSurgery, sd.c.hoos.symptom.befoSurgery, sm="MD")
summary(meta.hoos.symptom.befoSurgery)
dev.new(width=20, height=10)
forest(meta.hoos.symptom.befoSurgery, comb.fixed=FALSE)

#HOOS symptoms at six weeks after surgery
n.e.hoos.symptom.sixWeeksAft = c(43, 14)
mean.e.hoos.symptom.sixWeeksAft = c(72.4, 70.1)
sd.e.hoos.symptom.sixWeeksAft = c(16.0, 15.4)
n.c.hoos.symptom.sixWeeksAft = c(41, 12)
mean.c.hoos.symptom.sixWeeksAft = c(71.4, 72.3)
sd.c.hoos.symptom.sixWeeksAft = c(15.2, 10.8)
meta.hoos.symptom.sixWeeksAft = metacont(n.e.hoos.symptom.sixWeeksAft, mean.e.hoos.symptom.sixWeeksAft, sd.e.hoos.symptom.sixWeeksAft, n.c.hoos.symptom.sixWeeksAft, mean.c.hoos.symptom.sixWeeksAft, sd.c.hoos.symptom.sixWeeksAft, sm="MD")
summary(meta.hoos.symptom.sixWeeksAft)
dev.new(width=20, height=10)
forest(meta.hoos.symptom.sixWeeksAft, comb.fixed=FALSE)

#HOOS sports and recreation before surgery
n.e.hoos.sportsAndRecre.befoSurgery = c(40, 10, 14)
mean.e.hoos.sportsAndRecre.befoSurgery = c(38.5, 25.0, 28.2)
sd.e.hoos.sportsAndRecre.befoSurgery = c(18.9, 14.4, 16.6)
n.c.hoos.sportsAndRecre.befoSurgery = c(40, 10, 12)
mean.c.hoos.sportsAndRecre.befoSurgery = c(28.6, 32.5, 31.8)
sd.c.hoos.sportsAndRecre.befoSurgery = c(15.4, 20.2, 13.5)
meta.hoos.sportsAndRecre.befoSurgery = metacont(n.e.hoos.sportsAndRecre.befoSurgery, mean.e.hoos.sportsAndRecre.befoSurgery, sd.e.hoos.sportsAndRecre.befoSurgery, n.c.hoos.sportsAndRecre.befoSurgery, mean.c.hoos.sportsAndRecre.befoSurgery, sd.c.hoos.sportsAndRecre.befoSurgery, sm="MD")
summary(meta.hoos.sportsAndRecre.befoSurgery)
dev.new(width=20, height=10)
forest(meta.hoos.sportsAndRecre.befoSurgery, comb.fixed=FALSE)

#HOOS sports and recreation before surgery jackknife analysis
n.e.hoos.sportsAndRecre.befoSurgery.jackknife = c(10, 14)
mean.e.hoos.sportsAndRecre.befoSurgery.jackknife = c(25.0, 28.2)
sd.e.hoos.sportsAndRecre.befoSurgery.jackknife = c(14.4, 16.6)
n.c.hoos.sportsAndRecre.befoSurgery.jackknife = c(10, 12)
mean.c.hoos.sportsAndRecre.befoSurgery.jackknife = c(32.5, 31.8)
sd.c.hoos.sportsAndRecre.befoSurgery.jackknife = c(20.2, 13.5)
meta.hoos.sportsAndRecre.befoSurgery.jackknife = metacont(n.e.hoos.sportsAndRecre.befoSurgery.jackknife, mean.e.hoos.sportsAndRecre.befoSurgery.jackknife, sd.e.hoos.sportsAndRecre.befoSurgery.jackknife, n.c.hoos.sportsAndRecre.befoSurgery.jackknife, mean.c.hoos.sportsAndRecre.befoSurgery.jackknife, sd.c.hoos.sportsAndRecre.befoSurgery.jackknife, sm="MD")
summary(meta.hoos.sportsAndRecre.befoSurgery.jackknife)
dev.new(width=20, height=10)
forest(meta.hoos.sportsAndRecre.befoSurgery.jackknife, comb.fixed=FALSE)

#HOOS sports and recreation at six weeks after surgery
n.e.hoos.sportsAndRecre.sixWeeksAft = c(43, 14)
mean.e.hoos.sportsAndRecre.sixWeeksAft = c(48.9, 62.5)
sd.e.hoos.sportsAndRecre.sixWeeksAft = c(18.6, 18.7)
n.c.hoos.sportsAndRecre.sixWeeksAft = c(41, 12)
mean.c.hoos.sportsAndRecre.sixWeeksAft = c(49.1, 67.0)
sd.c.hoos.sportsAndRecre.sixWeeksAft = c(19.0, 21.0)
meta.hoos.sportsAndRecre.sixWeeksAft = metacont(n.e.hoos.sportsAndRecre.sixWeeksAft, mean.e.hoos.sportsAndRecre.sixWeeksAft, sd.e.hoos.sportsAndRecre.sixWeeksAft,
n.e.hoos.sportsAndRecre.sixWeeksAft, mean.c.hoos.sportsAndRecre.sixWeeksAft,  
sd.c.hoos.sportsAndRecre.sixWeeksAft, sm=“MD”)  
summary(meta.hoos.sportsAndRecre.sixWeeksAft)  
dev.new(width=20, height=10)  
forest(meta.hoos.sportsAndRecre.sixWeeksAft, comb.fixed=FALSE)

#HOOS hip-related quality of life before surgery
n.e.hoos.qualityOfLife.befoSurgery = c(40, 10, 14)  
mean.e.hoos.qualityOfLife.befoSurgery = c(38.8, 36.3, 38.0)  
sd.e.hoos.qualityOfLife.befoSurgery = c(17.2, 15.8, 10.1)  
n.c.hoos.qualityOfLife.befoSurgery = c(40, 10, 12)  
mean.c.hoos.qualityOfLife.befoSurgery = c(31.2, 43.3, 34.6)  
sd.c.hoos.qualityOfLife.befoSurgery = c(13.9, 15.4, 11.1)  
meta.hoos.qualityOfLife.befoSurgery = metacont(n.e.hoos.qualityOfLife.befoSurgery,  
mean.e.hoos.qualityOfLife.befoSurgery, sd.e.hoos.qualityOfLife.befoSurgery,  
n.c.hoos.qualityOfLife.befoSurgery, mean.c.hoos.qualityOfLife.befoSurgery,  
sd.c.hoos.qualityOfLife.befoSurgery, sm=“MD”)  
summary(meta.hoos.qualityOfLife.befoSurgery)  
dev.new(width=20, height=10)  
forest(meta.hoos.qualityOfLife.befoSurgery, comb.fixed=FALSE)

#HOOS hip-related quality of life at six weeks after surgery
n.e.hoos.qualityOfLife.sixWeeksAft = c(43, 14)  
mean.e.hoos.qualityOfLife.sixWeeksAft = c(60.2, 61.8)  
sd.e.hoos.qualityOfLife.sixWeeksAft = c(13.6, 16.4)  
n.c.hoos.qualityOfLife.sixWeeksAft = c(41, 12)  
mean.c.hoos.qualityOfLife.sixWeeksAft = c(54.2, 65.8)  
sd.c.hoos.qualityOfLife.sixWeeksAft = c(13.9, 16.1)  
meta.hoos.qualityOfLife.sixWeeksAft = metacont(n.e.hoos.qualityOfLife.sixWeeksAft,  
mean.e.hoos.qualityOfLife.sixWeeksAft, sd.e.hoos.qualityOfLife.sixWeeksAft,  
n.c.hoos.qualityOfLife.sixWeeksAft, mean.c.hoos.qualityOfLife.sixWeeksAft,  
sd.c.hoos.qualityOfLife.sixWeeksAft, sm=“MD”)  
summary(meta.hoos.qualityOfLife.sixWeeksAft)  
dev.new(width=20, height=10)  
forest(meta.hoos.qualityOfLife.sixWeeksAft, comb.fixed=FALSE)

LOS:

n.e.los = c(20, 10, 15, 14)  
mean.e.los = c(9.8, 6.7, 15.3, 5.1)  
sd.e.los = c(2.4, 1.8, 3.1, 1.0)  
n.c.los = c(20, 10, 20, 15)  
mean.c.los = c(10.2, 6.9, 17.8, 5.4)  
sd.c.los = c(1.7, 2.2, 3.6, 2.1)  
meta.los = metacont(n.e.los, mean.e.los, sd.e.los, n.c.los, mean.c.los, sd.c.los, sm=“MD”)  
summary(meta.los)  
dev.new(width=20, height=10)  
forest(meta.los, comb.fixed=FALSE)
References


James, J. (2013). Medicare Hospital Readmissions Reduction Program: To Improve Care and Lower Costs, Medicare Imposes a Financial Penalty on Hospitals with Excess Readmissions. Project HOPE.
randomized clinical trial involving 87 patients with 3 months of follow-up. *Acta Orthopaedica, 79*(2), 149-159.


VITA

EDUCATION

The Pennsylvania State University, University Park, Pennsylvania
Doctor of Philosophy in Industrial Engineering
Expected graduation date: August 2017

Georgia Institute of Technology, Atlanta, Georgia
Master of Science in Industrial Engineering
Graduation date: December 2011

Southeast University, Nanjing, China
Bachelor of Engineering in Transportation and Logistics
Graduation date: June 2010

JOURNAL PUBLICATIONS


PEER REVIEWED CONFERENCE PROCEEDINGS