1-31-2021 10:45 AM

Predicting Dissatisfaction with Total Knee Arthroplasty

Joseph S. Munn, *The University of Western Ontario*

Supervisor: Dr. Bert Chesworth, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Health and Rehabilitation Sciences

Abstract

Total knee arthroplasty (TKA) is a common and effective treatment for end-stage osteoarthritis. However, approximately 20% of patients are consistently found to be dissatisfied or unsure of their satisfaction with their TKA at 1-year post-surgery. Two strategies for reducing dissatisfaction have appeared in the literature. The first aims to reduce the number of dissatisfied patients who will undergo surgery by identifying patients likely to be dissatisfied pre-surgery. The second aims to implement interventions that improve patient satisfaction by generating realistic expectations for outcomes and by association improving met expectations scores. It may be possible to improve the performance of these interventions by identifying dissatisfied patients early in the recovery process. This thesis aimed to determine if dissatisfied patients can be identified pre-surgery or early in the recovery process as well as to better understand the met expectations variable.

Four studies were completed for this thesis. Study 1 attempted to predict 1-year post-surgery dissatisfaction using logistic regression and machine learning methods with pre-surgery and surgical variables. Study 2 aimed to identify patients dissatisfied at 1-year post-surgery using logistic regression and classification trees with pre and 3-month post-surgery data. Study 3 expanded on the findings of study 2 by creating a prediction tool to identify dissatisfied patients that can be easily administered in a clinical setting at 3-months post-surgery. This was done by using a pooled index that included five individual questionnaire items drawn from the 3-month post-surgery Knee injury and Osteoarthritis Outcome Score and the Knee Society Knee Scoring System questionnaires. Study 4 investigated the 1-year post-surgery met expectations variable and its relationship with satisfaction.

The results of study 1 indicated that pre-surgery and surgical variables were not sufficient to discriminate between satisfied and dissatisfied patients at 1-year post-surgery. The results of study 2 indicated that it is feasible to accurately identify patients who will be dissatisfied at 1-year post-surgery using 3-months post-surgery data. Study 3 found that the pooled index was able to accurately discriminate between satisfied and dissatisfied patients. The results of study 4 indicate that met expectations moderate the relationship between pain and satisfaction. This means that as met expectations scores increase, pain becomes less important for improving satisfaction.

Overall, this thesis found that patients dissatisfied at 1-year post-surgery cannot accurately be identified using pre-surgery and surgical variables. However, these patients can be identified at 3-months post-surgery using a simple prediction tool that can be easily administered in a clinical setting. Lastly, this thesis found that met expectations represent an important subjective threshold for patients and it may be reasonable to target unmet expectations to improve satisfaction.

Keywords: Prediction, Patient satisfaction, Patient expectations, Total knee arthroplasty, TKA

Summary for Lay Audience

Total knee arthroplasty (TKA) or total knee replacement is a popular and effective treatment for severe osteoarthritis. However, approximately 20% of patients are found to be dissatisfied with their TKA 1-year post-surgery. There have been two general strategies to reduce dissatisfaction rates presented in the literature. The first aims to reduce the number of dissatisfied patients who will undergo surgery by identifying patients likely to be dissatisfied pre-surgery. The second strategy is to introduce interventions that reduce dissatisfaction. When interventions have been attempted, they have aimed to generate realistic expectations for patients to improve met expectations scores and by association patient satisfaction. These interventions have had mixed results. Identifying dissatisfied patients early in the recovery process and better understanding the met expectations variable could help make interventions more effective. This thesis investigated if dissatisfied patients can be identified pre-surgery or early in the recovery process and aimed to better understand patients' met expectations.

Four studies were completed for this thesis. The first attempted to identify patients who would be dissatisfied at 1-year post-surgery using only pre-surgery data. This study found that dissatisfied patients at 1-year post-surgery could not be identified pre-surgery. The second and third studies attempted to identify dissatisfied patients early in the recovery process. These studies found that dissatisfied patients can be identified early post-surgery. The third study also found that this can be done using a simple tool that can be easily administered in a clinical setting. The fourth study investigated the met expectations variable to try and better understand its relationship with satisfaction. This study found that as met expectations increased, pain became less important for predicting dissatisfaction. This means that met expectations may represent an important subjective performance threshold for patients. Overall, this thesis found that 1-year post-surgery dissatisfaction cannot be predicted before surgery; but it can be identified early in the recovery process. Lastly, this thesis found that met expectations represent an important threshold for performance and it may be reasonable to target unmet expectations to improve patient satisfaction.

Co-Authorship Statement

This thesis includes manuscripts that are currently published in peer-reviewed journals, under review, or being prepared for submission.

Chapter 4: *Logistic Regression and Machine Learning Models Cannot Discriminate Between Satisfied and Dissatisfied Total Knee Arthroplasty Patients*. This manuscript has been published in the *Journal of Arthroplasty*. This manuscript was co-authored by Joseph S. Munn, Brent A. Lanting, Steven J. MacDonald, Lyndsay E. Somerville, Jacquelyn D. Marsh, Dianne M. Bryant, and Bert M. Chesworth. Joseph S. Munn was primarily responsible for study design, data analysis, and writing/preparing the manuscript. Bert M. Chesworth, supervised study design, data analysis, and writing/preparing the manuscript. Brent A. Lanting, Lyndsay E. Somerville, and Jacquelyn D. Marsh collected data and assisted with study design and manuscript preparation. Steven J. MacDonald and Dianne M. Bryant assisted with study design and manuscript preparation.

Chapter 5: *Can 1-Year Dissatisfaction with TKA be Predicted Pre-Surgery or Early in the Recovery Process?* This manuscript is being prepared for submission. This manuscript was co-authored by Joseph S. Munn, Sharon E. Culliton, Dianne M. Bryant, Lauren E. Cipriano, Steven J. MacDonald, and Bert M. Chesworth. Joseph S. Munn was primarily responsible for study design, data analysis, and writing/preparing the manuscript. Bert M. Chesworth, assisted with data collection, supervised study design, data analysis, and writing/preparing the manuscript. Sharon E. Culliton collected data and assisted with manuscript preparation. Dianne M. Bryant, Lauren E. Cipriano, and Steven J. MacDonald assisted with study design and manuscript preparation.

Chapter 6: *Can a Pooled Index Predict Dissatisfaction With 1-Year Total Knee Arthroplasty Early in the Recovery Process?* This manuscript is being prepared for submission. This manuscript was co-authored by Joseph S. Munn, Sharon E. Culliton, Dianne M. Bryant, Steven J. MacDonald, and Bert M. Chesworth. Joseph S. Munn was primarily responsible for study design, data analysis, and writing/preparing the manuscript. Bert M. Chesworth, assisted with data collection, supervised study design, data analysis, and writing/preparing the manuscript. Sharon E. Culliton collected data and assisted with manuscript preparation. Dianne M. Bryant and Steven J. MacDonald assisted with study design and manuscript preparation.

Chapter 7: *Can met expectations moderate the relationship between pain/function and satisfaction in total knee arthroplasty?* This manuscript has been published in the *Journal of Arthroplasty*. This manuscript was co-authored by Joseph S. Munn, Sharon E. Culliton, Dianne M. Bryant, Steven J. MacDonald, and Bert M. Chesworth. Joseph S. Munn was primarily responsible for study design, data analysis, and writing/preparing the manuscript. Bert M. Chesworth, assisted with data collection, supervised study design, data analysis, and writing/preparing the manuscript. Sharon E. Culliton collected data and assisted with manuscript preparation. Dianne M. Bryant and Steven J. MacDonald assisted with study design and manuscript preparation.

Acknowledgements

I would like to express my sincere thanks and appreciation to the individuals who have helped me complete this PhD. First and foremost, I would like to thank my supervisor Dr. Bert Chesworth, whose constant attention, support, and mentorship have made this PhD possible. You have been a role model and a tremendous source of knowledge. Your calm demeanor and steady hand have made you a joy to work with and learn from.

I would like to thank my committee members Dr. Steven MacDonald and Dr. Dianne Bryant for their insights, wisdom, and support throughout this process. I would also like to thank Dr. Sharon E. Culliton, Dr. Brent Lanting, Dr. Lyndsay Somerville, and Dr. Bryn Zomar without whom this work would not have been possible.

Lastly, I would like to thank my parents, Karen and Alan, my sister and brother-in-law, Amy and Adam, and my partner Ally for their love and support throughout this process.

Contents

List of Tables

List of Figures

List of Appendices

List of Acronyms

ADL	Activities of Daily Living
ASA	American Society of Anesthesiologists
AUC	Area Under the Curve
BMI	Body Mass Index
CART	Classification and Regression Trees
HADS	Hospital Anxiety and Depression Scale
KOOS	Knee Injury and Osteoarthritis Outcome Score
KSCRS	Knee Society Clinical Rating System
KSS	Knee Society Knee Scoring System
LR	Logistic Regression
LR-	Negative Likelihood Ratio
LR+	Positive Likelihood Ratio
NLP	Natural Language Processing
NPV	Negative Predictive Value
OA	Osteoarthritis
OKS	Oxford Knee Score
p	P-Value
PCS	Pain Catastrophizing Scale
PPV	Positive Predictive Value
PROM	Patient Reported Outcome Measure
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic
SCQ	Self-Administered Comorbidity Questionnaire
SD	Standard Deviation
SF-12	12-Item Short Form Health Survey Version 2
SF-12 MCS	12-Item Short Form Health Survey Version 2 Mental Component Score
SF-12 PCS	12-Item Short Form Health Survey Version 2 Physical Component Score
SVM	Support Vector Machine
TJA	Total Joint Arthroplasty
TKA	Total Knee Arthroplasty

UCLA	University of California Los Angeles
VIF	Variance Inflation Factor
WOMAC	Western Ontario and McMaster Universities Osteoarthritis Index
XG	Extreme Gradient

1 Introduction

1.1 Osteoarthritis and Total Knee Arthroplasty

Osteoarthritis (OA) is an extremely prevalent debilitating disease that affects approximately 1 in 8 Canadians[1]. Typical OA symptoms include joint pain, stiffness, and swelling. These symptoms progressively worsen over time, are the result of joint damage that cannot be repaired by the body, and result in reduced quality of life[1–3]. The joints most commonly affected by OA are the hands, knees, hips, spine, and feet[1–3]. The Arthritis Alliance of Canada has outlined three key strategies for prevention and treatment of arthritis, these include: obesity reduction, effective pain management strategies, and total joint arthroplasty (TJA)[1].

TJA is a surgical procedure where the diseased joint is removed and replaced with a prosthetic. This treatment is typically reserved for end-stage OA. TJA has been found to be a very effective treatment for patients, often relieving pain and improving joint function. Along with benefits to the patient, it is estimated that between 2010 and 2040 enhanced access to TJA will result in a cumulative cost savings of $17 billion for Canadians[1].

Total knee arthroplasty (TKA) is an extremely common type of TJA, with 60,209 knees being replaced in Canada in 2019-2020[4]. In recent years the number of TKAs has risen by roughly 5% a year, with this trend expected to continue. At an average cost of $10,500 per procedure this represented an approximate $632 million cost to the health care system in 2019-2020 alone[4].

Despite the evidence that TKA is an effective and cost-effective treatment, a significant number of patients are dissatisfied with their surgery. Approximately 20% of patients are consistently found to be dissatisfied following TKA, with studies reporting dissatisfaction rates as high as 35%[5,6]. Considering the prevalence, growth, and the cost of TKA, the problem of dissatisfied patients is a significant one.

1.2 Patient Satisfaction

Patient satisfaction is a concept that has not been concretely and consistently defined and there does not appear to be an agreed upon definition[7]. Patient satisfaction is considered a multidimensional[8] and subjective concept which can have different meanings for patients based on circumstances and past life experiences[7]. Definitions of patient satisfaction provided in the literature tend to fall into two different categories. The first set of definitions describes satisfaction as a patient's attitudes or emotions towards different elements of care[7,9,10]. The second set of definitions compare health care service or outcomes received to a set of subjective standards or expectations held by the patients[7,8,11]. Expectation confirmation theory provides a framework that describes how outcomes and expectations relate to satisfaction[12]. This theory states that the difference between expectations and actual outcomes determine satisfaction. Although not explicitly defined or stated, satisfaction is often discussed within the context of expectation confirmation theory in the TKA literature[13].

1.3 Conceptual Framework of TKA Satisfaction for This Dissertation

The conceptual framework of TKA satisfaction employed for this thesis is included in figure 1.1. This framework has been adapted from expectation confirmation theory[12]. Oliver's[12] expectation confirmation theory framework has been included in blue, with adaptations to the framework included in green.

The traditional expectation confirmation theory model includes three factors that directly or indirectly impact satisfaction. The first factor is perceived performance (knee pain and knee function), which directly influences satisfaction as well as influencing confirmation (post-surgery met expectations). The second factor is expectation (pre-surgery expectations). A patient expectation has been defined as the anticipation that an outcome is likely to occur as a result of a medical intervention[14]. This primarily affects satisfaction by being the reference point or the desired level of performance that the outcome of the surgery or actual knee performance is compared against. It has also been suggested that perceived performance and expectations can influence each other[13]. For example, a patient with poorer pre-surgery performance may expect more dramatic improvement post-surgery. Confirmation (post-surgery met expectations) is the patient's evaluation of the disparity between expected and actual performance. If actual performance does not meet or exceed expected performance, then it is theorized that a patient is more likely to be dissatisfied.

The adaptations to the expectation confirmation theory model represent additional variables that have been found or hypothesized to influence TKA satisfaction. Mental health, general health, surgical factors, and demographic factors are thought to both directly and indirectly affect satisfaction[5,6,15].

Lastly a line connecting pre-surgery expectations and satisfaction has been included. It has been suggested that if patients expect a better outcome, they are more likely to achieve a better outcome[13]. This indicates that pre-surgery expectations may directly affect satisfaction.

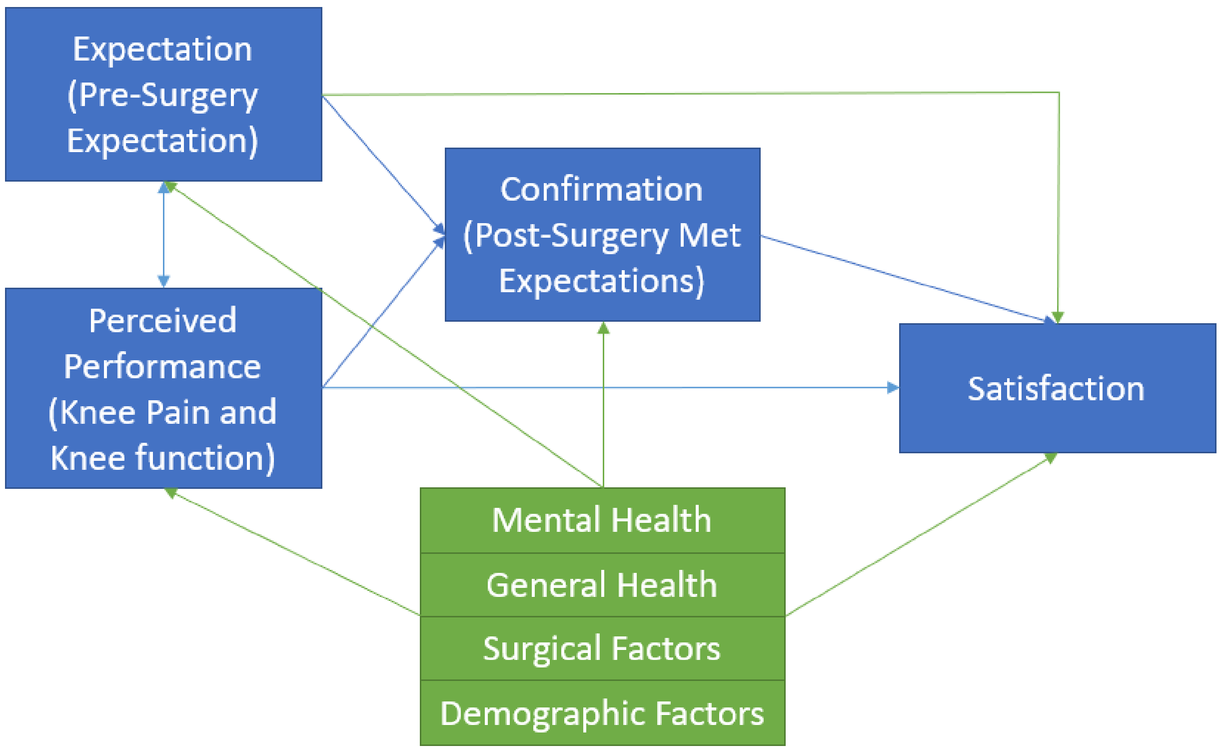


Figure 1.1. The conceptual framework of TKA satisfaction adapted from the expectation confirmation theory model of Oliver[12].
TKA, Total Knee Arthroplasty

1.4 Reducing TKA Dissatisfaction Through Prediction

There are two non-surgical strategies for improving TKA patient satisfaction rates that have been explored in the literature. The first is to decrease the number of patients who will be dissatisfied who receive a TKA. Reducing the number of dissatisfied patients' pre-surgery would require that it be possible to predict which patients will be dissatisfied before they undergo surgery. TKA satisfaction prediction models have previously been attempted in the literature, although it is not clear how useful they can be[16–19]. This is explored in greater detail in the literature review (section 2.5).

The second strategy involves the implementation of interventions to improve satisfaction. These interventions have frequently attempted to improve satisfaction by exploiting the expectations and confirmation components of the expectation confirmation theory framework by generating realistic expectations for patients[20–22]. These interventions have had mixed results and are explored in greater detail in the literature review (section 2.4). It has been suggested that it may be possible to identify patients as candidates for intervention early in the post-surgical recovery period (e.g. at 3-months post-surgery) although, there has been limited investigation in this area. Additional information is provided in the literature review (section 2.5).

This thesis aimed to determine if dissatisfied patients can be identified pre-surgery or early in the recovery process as well as to better understand the met expectations variable. Informed by the conceptual framework of TKA satisfaction created for this dissertation, four studies were completed.

1.5 Thesis Summary and Structure

This dissertation is presented as an integrated article that includes four studies (chapters 4-7). Chapter 1 introduces the dissertation subject and conceptual framework. Chapter 2 is a review of the TKA satisfaction literature. Chapter 3 outlines the thesis objectives. Chapter 4 includes a study that attempts to predict 1-year post-surgery dissatisfaction using logistic regression and machine learning methods with pre-surgery and surgical variables. Chapter 5 includes a study that attempts to identify 1-year post-surgery dissatisfaction with pre-surgery and 3-month post-surgery variables using logistic regression and classification trees. Chapter 6 includes a study that created a pooled index to predict 1-year post-surgery dissatisfaction using 3-month post-surgery data that can be easily transported to a clinical setting. Chapter 7 includes a study that evaluates how met expectations moderate the relationship between pain/function and satisfaction. Chapter 8 includes an overall discussion of the thesis findings, implications, limitations, and future research directions. Sample size and data quality information for all studies has been included in Appendices A and B.

1.6 References

1. Bombardier C, Hawker G, Mosher D. The impact of arthritis in Canada: today and over the next 30 years. 2011.
2. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis : estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis. 2014;73(7):1323–30.
3. Jones G. What's new in osteoarthritis pathogenesis? Intern Med J. 2016;46(2):229–36.
4. Hip and Knee Replacements in Canada: CJRR Annual Statistics Summary, 2019–2020. Ottawa, ON; 2021.
5. Kahlenberg CA, Nwachukwu BU, Mclawhorn AS, Cross MB, Cornell CN, Padgett DE. Patient satisfaction after total knee replacement: a systematic review. HSS J. 2018;14(2):192–201.
6. Nakano N, Shoman H, Olavarria F, Matsumoto T, Kuroda R, Khanduja V. Why are patients dissatisfied following a total knee replacement? A systematic review. Int Orthop. 2020;44(10):1971–2007.
7. Batbaatar E, Dorjdagva J, Luvsannyam A, Pierto A. Conceptualisation of patient satisfaction: a systematic narrative literature review. Perspect Public Health. 2015;135(5):243–50.
8. Graham B, Green A, James M, Katz J, Swiontkowski M. Measuring patient satisfaction in orthopaedic surgery. JBJS. 2015;97(1):80–4.
9. Jenkinson C, Coulter A, Bruster S, Richards N, Chandola T. Patients' experiences and satisfaction with health care: results of a questionnaire study of specific aspects of care. Qual Saf Heal care. 2002;11(4):355–339.
10. Baker R. Viewpoint Pragmatic model of patient satisfaction in general practice: progress towards a theory. BMJ Qual Saf. 1997;6(4):201–4.
11. Hills R, Kitchen S. Development of a model of patient satisfaction with physiotherapy. Physiother Theory Pract. 2007;23(5):255–71.
12. Oliver RL. Satisfaction: A behavioural perspective on the consumer. Second Edi. New York, NY: M.E. Sharpe; 2010. 61–89 p.
13. Hafkamp FJ, Gosens T, de Vries J, den Oudsten BL. Do dissatisfied patients have unrealistic expectations? A systematic review and best-evidence synthesis in knee and hip arthroplasty patients. EFORT Open Rev. 2020;5(4):226–40.
14. Bowling A, Rowe G, Lambert N, Waddington M, Mahtani KR, Kenten C, et al. The measurement of

patients' expectations for health care: a review and psychometric testing of a measure of patients' expectations. Health Technol Assess (Rockv). 2012;16(30).

15. Gunaratne R, Pratt DN, Banda J, Fick DP, Khan RJK, Robertson BW. Patient dissatisfaction following total knee arthroplasty: a systematic review of the literaturere. J Arthroplasty. 2017;32(12):3854–60.

16. Kunze KN, Polce EM, Sadauskas AJ, Levine BR. Development of machine learning algorithms to predict patient dissatisfaction after primary total knee arthroplasty. J Arthroplasty. 2020;35(11):3117–22.

17. Kunze KN, Akram F, Fuller BC, Zabawa L, Sporer SM, Levine BR. Internal validation of a predictive model for satisfaction after primary total knee arthroplasty. J Arthroplasty. 2019;34(4):663–70.

18. Farooq H, Deckard ER, Ziemba-Davis M, Madsen A, Meneghini RM. Predictors of patient satisfaction following primary total knee arthroplasty: results from a traditional statistical model and a machine learning algorithm. J Arthroplasty. 2020;35(11):3123–30.

19. Garriga C, Sanchez-Santos MT, Judge A, Perneger T, Hannouche D, Lübbeke A, et al. Development of a model predicting non-satisfaction 1 year after primary total knee replacement in the UK and transportation to Switzerland. Sci Rep. 2018;8(1):1–8.

20. Culliton SE, Bryant DM, MacDonald SJ, Hibbert KM, Chesworth BM. Effect of an e-learning tool on expectations and satisfaction following total knee arthroplasty: a randomized controlled trial. J Arthroplasty. 2018;33(7):2153–8.

21. Tolk JJ, Janssen RPA, Haanstra TM, van der Steen MC, Bierma-Zeinstra SMA, Reijman M. The influence of expectation modification in knee arthroplasty on satisfaction of patients: a randomized controlled trial. Bone Joint J. 2021;103(4):619–26.

22. Gautreau SJ, Gould ON, Allanach WW, Clark AE, Massoeurs SJ, Forsythe ME. Total knee arthroplasty communication checklist increases patient satisfaction. J Arthroplasty. 2019;34(3):456–61.

2 Literature Review

2.1 Introduction

This literature review is divided into seven sections. Section one introduces the literature review. Section two provides some additional background total knee arthroplasty (TKA) background. Section three discusses the measurement and analysis of TKA satisfaction. Section four discusses factors that have been found to be associated with TKA satisfaction. Section five discusses interventions that have been implemented to improve TKA satisfaction. Section six examines the pre-surgery and early post-surgery prediction of dissatisfied patients. Lastly, section seven provides a statement of the problem, identifies gaps in the literature, and provides a rationale for this dissertation.

2.2 Additional TKA Background

2.2.1 Indicators for and Contraindicators for TKA

TKA is typically reserved for patients with moderate to severe radiological and clinical osteoarthritis symptoms[1]. However, physicians and surgeons are inconsistent in their identification of physical and clinical indicators for TKA and clear symptom thresholds have not been established[1,2]. Physical symptoms generally include joint space narrowing, the presence and size of osteophytes and bone cysts, squaring of the condyles, and bone sclerosis[3]. Clinical symptoms typically include pain during activities, pain at rest or occurring at night, difficulties performing activities of daily living, and decreased mobility[4]. Pain and disability should be adversely affecting a patient's quality of life on a daily basis[4]. Additionally, it is generally agreed that TKA should only be considered after the patient has failed to respond to more conservative non-operative treatments[1,3].

With the exception of severe psychiatric disorders, such as advanced dementia, there are essentially no agreed upon contraindications for TKA[1]. However, there are contextual factors that that may affect the outcome or benefit that a patient is likely to experience from surgery. These include, but are not limited to, age[5], sex[6] , body mass index (BMI)[7], mental health[1,8], comorbidities[8,9], other troublesome joints[9], health-related quality of life[10], arthritis severity[11], pre-surgery functional ability[8,12], pre-surgery pain[5,8], employment status[13,14], worker compensation[14], wait times[15], smoking status[16], opioid use[17,18], and readiness to undergo surgery[19].

2.2.2 TKA Recovery

Patients typically take 1-year to fully recover from TKA[20]. Within this 1-year period, most patients follow a similar recovery pattern. Immediately following surgery, patients will experience increased pain and decreased functional ability[21]. From the date of surgery to 3-months post-surgery, patients will experience the majority of their progress with rapid improvements in pain and function[20,21]. Between 3 and 6-months post-surgery, patients will continue to improve, although at a slower pace[20]. From 6-months to 1-year post-surgery, change in the performance of the knee is minimal, however, small improvements are still noted[20]. This recovery pattern is observed in range of motion, pain relief, quadricep strength, and general function[20,21]. It is possible for improvements to be observed beyond 1-year post-surgery, however, patient reported outcome measures (PROMs) scores have generally been found to be the same in patients at 1 and 2-years post-surgery[22].

2.2.3 Evaluating the Outcome of a TKA

2.2.3.1 Determinants of a Good Outcome

When determining if a patient has achieved a good outcome or the surgery was successful, there are multiple factors that must be considered. Ideally the surgery should be complication free and surgical intervention or revision should not be required[23,24]. Potential complications and common modes of failure that arise during and following surgery include fractures, infections, loosening of the implant, improper alignment of the implant, and post-surgery complications unrelated to the knee that require an extended hospital stay[23,24]. Additionally, it is desirable that the replaced knee have no or limited flexion contracture or extension lag, good mediolateral stability with $<5^{o}$ of movement in the joint, good anteroposterior stability with <5mm of movement, neutral alignment, good range of motion with $>120^{o}$ of flexion, and minimal stiffness, swelling or clicking in the knee[25]. Lastly, good PROMs including pain[26,27], functional ability[26,27], and health-related quality of life[10] are important for a patient to achieve a good outcome.

2.2.4 Defining a Good TKA Outcome

The outcome of TKA is typically defined using patient satisfaction[26–28]. Patient satisfaction generally requires that the patient have their expectations met[5,29–32] and experience substantial improvements in pain[26,27], functional ability[26,27], and health-related quality of life[10]. In addition to satisfaction, other methods of defining a good outcome have been used. These include using pain and function PROMs alone, as well as employing composite scores that pool outcomes.

When using pain and function scores alone, success is typically defined using minimally clinically important differences or by establishing patient reported performance thresholds that must be met for a patient to be satisfied[9,33–36]. However, these thresholds are typically established using some type of satisfaction or quality of life measure.

Composite scores that have been used to assess surgical success typically pool PROMs outcomes into a single score that is then used to determine if a good outcome was achieved. These scores generally include an assessment of pain, function, satisfaction, and willingness to undergo surgery again[14,33,34]. Additionally, the OARSI-OMERACT responder criteria in combination with patient satisfaction has been employed to define a good outcome[19,37].

At present there does not appear to be an agreed upon or method of defining a good outcome in the TKA literature. However, all methods generally require that patients experience significant pain relief and functional improvement.

2.3 TKA Patient Satisfaction

2.3.1 Measurement of Patient Satisfaction for TKA

TKA patient satisfaction has been measured in a variety of different ways, however, it is most frequently measured using a single item general satisfaction Likert style question[5,38–48]. This question typically has five[5,41,45–47,49,50] possible response options ranging from very dissatisfied to very satisfied, although some researchers have employed 11[48], 10[51], 4[38,42,52,53], 3[39,43,54], and 2[40] possible responses. A neutral, neither satisfied or dissatisfied, or an unsure response option is typically included in these questionnaires. A single item visual analogue scale has also been used[55]. Other studies have aggregated scores of multiple items, either by summing or taking the mean of the response values[38,42]. Even when continuous scores are used, they are typically dichotomized. These satisfaction measures are frequently unvalidated.

2.3.2 Analysis of Patient Satisfaction for TKA

The most common methods of analyzing patient satisfaction are ordinal logistic regression[10], binary logistic regression[5,38,39,43–46,48,50,52,56], linear regression[42,46–48], and tree-based machine learning methods[57,58].

Ordinal logistic regression is an analysis method that is employed to analyze an ordinal dependent variable as predicted by one or more independent variable[59].

Logistic Regression is an analysis method that is employed to analyze a binary dependent variable as predicted by one or more independent variables[60].

Linear regression is an analysis method that is employed to analyze a continuous dependent variable as predicted by one or more independent variables[60].

Tree-based machine learning methods are a set of algorithms that use variations of a classification tree or combinations of classification trees to predict a categorical dependent variable using one or more independent variables[61].

2.4 Factors Affecting TKA Patient Satisfaction

2.4.1 Demographic Factors

2.4.1.1 Age

Age appears to have a complex relationship with satisfaction. Bourne[5] found that advanced age is a predictor of dissatisfaction, however, Huijberts[40] found that as age increases patients become more likely to be satisfied. Additionally, Clement[62], and Goh[63] have found that patients over the age of 80 were more and less satisfied, respectively, than their younger counterparts. Patient satisfaction has also been reported to be lower among patients under 55[64,65], with greater levels of residual pain and limitation in function following surgery[66]. Conversely, it has also been reported that patients under the age of 60 are more satisfied with their surgery and more likely to have their expectations met[52]. Based on the available literature, it is unclear precisely what effect age has on patient satisfaction.

2.4.1.2 Sex

Female sex has been associated with dissatisfaction[67,68]. However, this finding has not been consistently observed.

2.4.1.3 Living Status

Living status has been found to be a significant predictor of satisfaction, with patients who live alone being more likely to be dissatisfied with their knee replacement[5].

2.4.1.4 Body Mass Index

Both lower[56] and higher[27,69,70] BMI has been associated with dissatisfaction. However, BMI has generally not been linked to TKA satisfaction in the literature and it rarely appears in multivariable models that use dissatisfaction as the dependent variable. At present it is unclear what effect BMI has on satisfaction.

2.4.1.5 Medical Comorbidities

Back pain[71,72] and pain in other joints were significant predictors of dissatisfaction along with the total number of comorbidities that a patient experienced[41].

2.4.1.6 Arthritis Severity and Duration

Severe pre-operative arthritis has been associated with higher satisfaction scores following surgery[46,68]. Robertsson[73] found that patients with longer disease duration were more likely to be satisfied with their surgery. Additionally, severe medial cartilage damage has been associated with satisfaction[51]. At the other end of the spectrum, mid and low grade OA prior to surgery has been associated with dissatisfaction[46,74].

2.4.1.7 Radiographic Indicators

Several radiographic indicators of satisfaction have been observed. Specifically, increased size of the patella and the presence and increased size of lateral osteophytes were positively associated with patient satisfaction[75]. High radiographic joint narrowing was associated with satisfaction[69] and patients with complete pre-surgery joint space collapse were found to be more satisfied at 1-year post-surgery[76]. Lastly, patients with an intact anterior cruciate ligament were more likely to be dissatisfied[77].

2.4.1.8 Socioeconomic Status

The effect of socioeconomic status (SES) on TKA satisfaction is difficult to determine. Many patients undergoing TKA are older adults who may be retired. SES is difficult to assess in the retired population as income will not accurately reflect the patient's financial situation[78]. SES has been found to have no effects on satisfaction or surgical outcomes in studies performed in the Netherlands[47] and Australia[79] and low SES has been found to significantly predict dissatisfaction in a study performed in the United States[80]. As health care systems and social safety nets differ across countries it is difficult to compare or apply the results of studies performed on different populations.

2.4.1.9 Workers' Compensation and Return to Work

Patients who are receiving workers' compensation have been found to be less likely to be satisfied and more likely to have poorer outcomes post-surgery[14]. However, the study did not explicitly evaluate patient satisfaction, but included an unvalidated questionnaire that claimed to measure satisfaction[14]. Additionally, patients who are able to return to work have been found to have greater met expectations scores than those who cannot[81]. A number of factors have been observed to affect return to work including pre-surgery length of sick leave, age, type of work (physical or sedentary), strain placed on the knee by work, and if a patient is self-employed[13,81,82]. Psychosocial factors including motivation to return to work, job satisfaction, recognition received at work, and development opportunities have also been observed to impact return to work[82,83]. Although not established explicitly in the TKA literature, job satisfaction in orthopaedic trauma patients has also been linked to patient satisfaction[84].

2.4.2 Psychological Factors

2.4.2.1 TKA Patient Expectations

Patient expectations in the TKA literature are discussed in relation to met expectations and pre-surgery expectations. Expectations are often established pre-surgery and patients are asked to rate how much pain and functional ability they expect to have post-surgery. Post-surgery patients are then asked if their expectations have been met.

2.4.2.1.1 Pre-Surgery Expectations

Pre-surgery expectations are theorized to affect satisfaction directly by increasing patient optimism and belief in recovery outcomes[29,85]. It has been found across a wide variety of clinical conditions that patients who believe they are going to do better will have a better recovery[86]. The effect of TKA pre-surgery expectations on satisfaction has been examined in a number of studies[39,42,49,50,87–90], however, a significant relationship between pre-surgery expectations and satisfaction is rarely found[39,88]. Kiran[39] found that higher pre-surgery expectations required that a patient also have higher pain relief to be satisfied. Becker[88] found that higher pre-surgery expectations were associated with improved pain and function scores, met expectations, and satisfaction. Higher pre-surgery expectations have also been found to be independently associated with greater pain relief[91]. However, expectation confirmation theory would suggest that unrealistically high pre-surgery expectations are likely to lead to dissatisfaction as these high expectations are unlikely to be met[42].

2.4.2.1.2 Met Expectations

Met expectations as a predictor of satisfaction is a consistent finding in the literature[5,29–31,41,49,87,88,92]. Met expectations in the TKA literature are frequently discussed in relation to satisfaction through the expectation confirmation theory framework[93].

2.4.2.2 Personality

Personality has been found to be a significant predictor of satisfaction and patient outcomes following surgery. Extraversion[94] and greater life satisfaction[95] have been associated with improved patient outcomes and satisfaction. Whereas, neuroticism, anxious and introverted personalities[94], somatic stress, and emotional instability[95] have been associated with dissatisfaction. Ramaesh[96] found no direct effect of personality on satisfaction, although it was concluded that personality may interact with disease processes to impact satisfaction. The precise effect that personality has on satisfaction is presently unclear.

2.4.2.3 Mental Health

Pre and post-surgery depression, anxiety, and depressive symptoms have been found to have a significant negative effect on satisfaction[27,41,87,97–101]. Although consistently found in univariate analyses, this does not always translate to multivariable models[102]. In addition to anxiety and depression, pain catastrophizing[103], somatization disorder[104], and lower internal locus of control[102] have been observed to have a negative effect on satisfaction. Pain catastrophizing refers to a patient's rumination, magnification, and feelings of helplessness in response to pain[105]. Somatization is physical discomfort that is experienced despite having no apparent physical cause[104]. Locus of control refers to patient's beliefs about their role and ability to affect their own recovery process[106].

2.4.2.4 Patient Perception

Patient perception has also been found to have an effect on satisfaction[107]. It was found that patients who perceived their knee as being misaligned despite good radiographic alignment had poorer perception of range of motion and greater dissatisfaction than their counterparts.

2.4.3 Surgical Factors

2.4.3.1 Type of Implant

There have been conflicting findings with regards to the effect of implant type on satisfaction. A number of studies have observed minimal or no effect on PROMs and satisfaction[5,43,56,108–111], whereas others have reported significant differences between different types of implants[112–115]. Mobile bearing implants in different studies have been found to improve[114,115] and have no effect on satisfaction[5,116]. Similarly, studies have found that patients with single radius designs have greater[117] and no different[111] satisfaction than patients with multi radius implants. However, medial pivot implants have been found to provide greater satisfaction than non medial pivot designs[118]. High flexion versus standard implants have not been observed to have an effect on satisfaction[114,119,120]. No differences in satisfaction between fixation method were observed with cemented, uncemented, and hybrid fixation[40,121]. When evaluating more modern implants Chua[109] found that current implants were the equivalent of their earlier counterparts in satisfaction levels. Reiman[122] found that patient-specific implants and cutting guides have been associated with improved satisfaction. However, the population that received the patient-specific implants were significantly younger.

2.4.3.2 Surgical Technique or Procedure

Surgical technique for the most part has not been observed to impact satisfaction. Minimally invasive surgery has been reported to improve recovery speed, however, there have been no studies that have indicated that this has an effect on long term patient satisfaction[123]. Custom cutting guidelines[124], lateral release[43], the number of structures released[48], if a tourniquet is used[48,125], if unilateral or bilateral surgery is performed[56], and the surgeon performing the procedure[56] have also been found to have no or minimal impact on satisfaction. Patellar resurfacing has been found to have both no effect[43,73,126–129] and a significant effect on patient satisfaction[130,131], although this effect is typically small. Collateral ligament laxity in bilateral knees on the patients' favored side has been found to have an effect on patient satisfaction, with increased laxity being associated with greater satisfaction[132]. Internal rotation of the femoral component has been associated with slightly decreased satisfaction scores[133]. Restricted kinematic alignment[134] and inverse kinematic alignment[135], processes that aim to match a patients native joint alignment more closely, have been found to provide greater satisfaction than mechanical and adjusted mechanical alignment. Robotic assisted surgery has also been associated with greater expectation fulfillment and satisfaction[136].

2.4.3.3 Complications

Post-surgery complications that required a readmission and overnight stay in the hospital because of the replaced joint, for example, an infection or a blood clot, have been found to be associated with dissatisfaction[5,69].

2.4.4 Patient Reported and Physical Post-Operative Outcomes

2.4.4.1 Pain

Pain relief is one of the most consistent and influential predictors of satisfaction in the literature[5,10,38–41,44,45,50,56,87,90,137–140]. Pain relief or post-surgery residual pain has been consistently found to predict satisfaction. The knee society score considers pain while lying and sitting to be extremely important in overall patient satisfaction[92]. Bourne[5] found that patients with severe pre-surgery pain while lying or sitting were more likely to be dissatisfied. In contrast Maratt[141] found that patients with severe pre-surgery pain, but higher health-related quality of life were more likely to be satisfied. Contralateral knee pain has also been associated with dissatisfaction[142].

2.4.4.2 Function

Patient reported functional performance is also a consistent and influential predictor of patient satisfaction[10,38–41,44,45,49,50,54,56,90,137–139,143,144]. Typically, function measurements are split into two different subcategories, daily living and recreational or leisurely function. Multi item satisfaction questionnaires typically address both daily living and recreational activity[137,145]. Daily living function typically includes activities like, walking, going up stairs, getting in and out of bed, and performing light household duties. Recreational or leisure activity is typically considered to be a more advanced activity like running, squatting, swimming, dancing, or other sporting related activities[90,137]. PROMs of ability to perform daily living activities are typically referenced as being important to satisfaction[10,38–41,44,45,49,50,54,56,90,137–139,146]. Leisure activities, physical activity and the desire of the patient to perform sporting activities have also been found to be important[147,148], especially for younger patients[149] who have been found to have high expectations about their ability to perform leisure activities post-surgery[90].

In conjunction with PROMs, physical measures including, quadriceps strength, range of motion, and walking distance have been associated with patient satisfaction. Change in range of motion from pre-surgery was found to be significantly associated with satisfaction[114]. Loss of range of motion[56], improved flexion[139], pre and post-surgery varus alignment[5,150], pre-surgery valgus alignment[5], and pre and post-surgery flexion contracture[151] have been found to be predictors of patient dissatisfaction. However, overall degree of flexion has not been found to be significantly correlated with satisfaction[152–154]. Quadricep strength, ability to walk longer distances, and climb more stairs were positively associated with improved levels of satisfaction[138,139]. These measures were also found to coincide with higher PROM scores.

2.4.4.3 Health-Related Quality of Life

The 12 and 36 item Short Form Health Survey (SF-12 and SF-36) mental component score and physical component score, as well as the EuroQol 5D (EQ-5D) have been found to be positively associated with TKA patient satisfaction. In general, better health-related quality of life scores are associated with greater levels of satisfaction post-surgery[10,155–157].

2.4.5 Other Factors

There are additional factors that have been found to affect patient satisfaction. The amount of time a patient has to wait to undergo their knee arthroplasty has a negative impact on their satisfaction[15]. Pre-surgery narcotic use has also been found to be associated with greater dissatisfaction and pain post-surgery[44]. Retrospective perception of hospital stay at 6-months post-surgery was found to be associated with satisfaction[158]. Ability of the patient to choose the hospital at which they will undergo surgery has been associated with greater satisfaction[159]. The number of allergies a patient has and drug allergies have been associated with greater dissatisfaction[57]. Patients who have a second TKA are more likely to be more satisfied with their second operation[160,161]. Lastly, smoking status has been found to be associated with dissatisfaction[16].

2.5 Satisfaction Interventions

2.5.1 Met Expectations Interventions

As met expectations have been found to be significant predictors of satisfaction, interventions modifying patient expectations have been considered as a potential avenue to improve TKA satisfaction. Patients typically have higher expectations for the outcomes of their surgeries than surgeons do[162]. This is especially true of younger patients as they tend to expect to be able to perform higher intensity daily and leisure activities[149].

Several interventions have been attempted in order to influence patient expectations to improve satisfaction (table 2.1)[163–165]. These interventions have aimed to exploit expectation confirmation theory by generating realistic expectations for patients to improve met expectations levels and by association improve patient satisfaction. These interventions have been implemented in different ways and had mixed results.

Table 2.1. Studies that have implemented interventions to improve patient satisfaction.

Study	Intervention type	Outcome (1-year post-surgery)
Culliton 2018	Pre and post-surgery e-learning tool	No effect
Gautreau 2019	Post-surgery surgery surgeon checklist	Improved satisfaction
Tolk 2021	Pre-surgery patient education package	Improved satisfaction

2.5.1.1 Culliton 2018

Culliton[164] used an e-learning tool to provide an educational package meant to improve patient satisfaction. It was found that this intervention was not effective and did not improve patient satisfaction at 1-year post-surgery. Although, it was observed that patients did not continue to use the e-learning tool over time.

2.5.1.2 Gautreau 2019

Gautreau[166] found that a post-surgery surgeon-administered checklist to generate realistic expectations for patients was able to improve satisfaction at 6-months post-surgery. However, surgeons were observed to spend more time with patients who received the checklist than those not included in the intervention. Time spent with patients has been found to be associated with patient satisfaction. Additionally, this study did not include an overall satisfaction measure and a number of the satisfaction domains measured focused on satisfaction with the experience and the surgeon, not the outcome of surgery. Patients were found to be more satisfied with higher met expectations, although no significant differences between satisfaction with pain and function outcomes were observed between patients who did and did not receive the checklist intervention.

2.5.1.3 Tolk 2021

Tolk[165] found that a pre-surgery education intervention that aimed to generate realistic expectations was able to improve satisfaction at 1-year post-surgery. The study found that the education program significantly lowered patient expectations and patients had significantly higher met expectations scores and by association greater patient satisfaction.

2.5.1.4 Summary

It appears that interventions to improve satisfaction can be effective, however, mode of delivery and the timing of the intervention appear to be important.

2.5.1.5 Problems With the Met Expectations Construct

It is assumed that met expectations are actually measuring how closely a patient's expectations match up with their actual condition. This interpretation of a single met expectation question has been criticized as it has been argued that this is simply a different type of performance question, that may not actually be measuring how closely expectations match up with actual outcomes, but instead simply provide a different scale to measure functional outcomes[167]. Additional work needs to be done to better understand the met expectations variable and what precisely it is measuring and how it is explicitly related to satisfaction.

2.6 Predicting Dissatisfaction

2.6.1 Pre-surgery Prediction and Early Post-Surgery Identification of Dissatisfied Patients

There have been a number of attempts to predict post-surgery satisfaction using pre-surgery and surgical variables. More recently the identification of dissatisfied patients early in the recovery process has also been investigated. All studies found by the author that predict dissatisfaction pre-surgery or early in the recovery process have been included in table 2.2.

Studies that aim to predict if a patient will be dissatisfied pre-surgery or early in the recovery process have only recently been attempted. To the best of the author's knowledge the earliest study explicitly aimed to identify post-surgery dissatisfied patients pre-surgery was completed in 2016[168], with the majority of these studies having been completed between 2018 and 2021[16,57,58,169–171]. The data that have been used to develop these models have come from a number of different countries and to the best of the author's knowledge there have been no studies that have examined a Canadian population. The majority of these papers have attempted to identify dissatisfied patients using pre-surgery variables, however, Goh[169] did use early post-surgery variables to predict dissatisfaction. The satisfaction end-points used in these studies ranged from 3-months to 3-years post-surgery, with the most commonly used endpoints being 1 and 2-years post-surgery.

Dissatisfaction has most commonly been predicted by dichotomizing the satisfaction variable into dissatisfied/unsure/neutral and satisfied patients. Two methodological approaches have been employed to predict dissatisfaction using pre-surgery and surgical variables. The first uses traditional statistical methods, primarily linear and logistic regression. The second method employs machine learning algorithms. Machine learning methods are a set of algorithms that can learn and adapt without specific instructions being provided to them[61]. The use of machine learning methods in orthopaedic medicine is a fairly recent development with both studies that have employed machine learning to predict dissatisfaction being published in 2020[57,58]. The prediction models that have been created have made use of a wide variety of metrics to assess the performance of their models, with the most common metrics used being area under the curve (AUC), calibration metrics, sensitivity, and specificity.

The final conclusion of the authors on the ability of their models to predict dissatisfaction is mixed, with several authors concluding that satisfaction can be predicted and others finding that it cannot be predicted or presenting contradictory findings. The two studies that have used machine learning methods have both concluded that dissatisfaction can be predicted post-surgery using pre-surgery variables. However, it is unclear precisely how clinically useful any of these models would be.

Table 2.2. Studies that have attempted to predict dissatisfaction pre-surgery or early in the recovery process.

Logistic Regression (LR)							
Study	Data Source (Sample Size/Dissatisfaction Rate)	Time-Point of Prediction	Post-Surgery End-Point	Prediction Method	Significant Independent Variables	Metrics Reported	Authors' Conclusion
Van Onsem 2016	Netherlands (113/0.12)	Pre-surgery	3-months	Linear regression, binomial LR	10	R^2, sensitivity, specificity, PPV, NPV	Can predict dissatisfaction
Garriga 2018	United Kingdom (450/0.142) and Switzerland (791/0.199)	Pre-surgery	1-year	Binomial LR	6	AUC, calibration plot	Contradictory findings
Zabawa 2019*	USA (203/0.35)	Pre-surgery	3-years	Linear regression, binomial LR	10	R^2	Cannot predict dissatisfaction
Calkins 2019*	USA (145/0.083)	Pre-surgery	3-months	Linear regression, binomial LR	10	Sensitivity, specificity, NPV	Cannot predict dissatisfaction
Kunze 2019	USA (484/not reported)	Pre-surgery	1-year	Binomial LR	1	AUC, sensitivity, specificity, NPV	Can predict dissatisfaction
Goh 2021	Singapore (4359/0.089)	6-months post-surgery	2-years	Binomial LR	1	AUC	Can predict dissatisfaction
Machine Learning							
Study		Time-Point of Prediction	Post-Surgery End-Point	Prediction Method		Metrics	Conclusion
Kunze 2020	USA (430/0.09)	Pre-surgery	2-years	LR, stochastic gradient boosting, random forest, support vector machine, neural network, elastic-net LR	5	AUC, Brier score, calibration plot, calibration intercept, and calibration slope	Can predict dissatisfaction
Farooq 2020	USA (897/0.147)	Pre-surgery	1-year	LR, gradient boosting machine	9	AUC, sensitivity, and specificity	Can predict dissatisfaction

*Validation study of Van Onsem 2016; LR, logistic regression; USA, United States of America; PPV, Positive predictive value; NPV, negative predictive value; AUC, area under the curve

2.6.1.1 Linear and Logistic Regression

Linear regression is a type of prediction model that uses a continuous dependent variable. Logistic regression is a set of prediction models that use a categorical variable as the dependent variable. Typically, this is a dichotomous or binomial variable, however ordinal, and multinomial variables can also be used. Van Onsem[168], Zabawa[170], and Calkins[171] used a combination of linear and logistic regression models to predict dissatisfaction. Garriga[16] and Kunze[17] used logistic regression to predict or help develop a checklist to predict patient dissatisfaction.

2.6.1.1.1 Van Onsem 2016

Van Onsem[168] used linear and logistic regression to predict 3-month post-surgery dissatisfaction using pre-surgery variables only. This study included a sample size of 113 patients with a dissatisfaction rate of 0.12 (n=14). In this study individual items from PROMs questionnaires were used as predictors of dissatisfaction. A total of 10 items were included, these questions evaluated gender, age, pain, stiffness in the morning, grinding and clicking, being aware of the knee, anxiety/depression, and two questions evaluating pain catastrophizing. The questions included in the prediction model were selected using linear regression with a continuous satisfaction variable used as the dependent variable. This continuous satisfaction variable was then dichotomized, and logistic regression was used to predict if a patient was satisfied or not. However, it was not clearly explained how logistic regression was used to generate a prediction for this model. Van Onsem[168] did not report AUC or accuracy for the study, but did report a sensitivity of 0.97 (correctly predicted as satisfied), specificity of 0.50 (correctly predicted as dissatisfied), a positive predictive value (PPV) of 0.93, and a negative predictive value (NPV) of 0.66. Using the satisfaction rate (0.88), along with sensitivity and specificity scores, an overall prediction accuracy of 0.91 can be calculated. This indicates that this model had an accuracy that was marginally (0.03) higher than using no prediction model at all. Problematically, this study included no method to control or adjust for model optimism. This suggests that any reported performance metrics are likely to be inflated. Additionally, this study only included 14 dissatisfied patients and It is unlikely that this would be sufficient to create a reliable prediction model[172].

2.6.1.1.2 Garriga 2018

Garriga[16] used logistic regression to predict 1-year post-surgery satisfaction using pre-surgery variables only. This study included a sample size of 450 patients from the United Kingdom and a dissatisfaction rate of 0.142 (n=64) in their training population and a sample size of 791 patients from Switzerland and a dissatisfaction rate of 0.199 (n=158) in their validation population. Variables included in the model were age, sex, prior treatment for anxiety, current smoker, injection of corticosteroids, and standardised oxford knee score (OKS)/Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). Garriga[16] reported an optimism adjusted AUC of 0.65 and a calibration plot was provided. This model was not found to transport well to a new population with an AUC of 0.55 being found using the validation population. An AUC score of >0.7 is typically considered as an acceptable score and an AUC of 0.5 is a model that is not predictive or does not perform better than chance. This indicates this model does not perform particularly well when discriminating between satisfied and dissatisfied patients. The authors attributed part of this drop off to the differences between the countries in the training and validation population. Despite the poor discrimination performance, this model appeared to have good calibration when completing a visual inspection of the calibration plot. However, additional calibration metrics such as a Brier score, Cox/calibration intercept, and Cox/calibration slope would help make this evaluation more rigorous.

2.6.1.1.3 Zabawa 2019 and Calkins 2019

Zabawa[170] and Calkins[171] used the Van Onsem[168] model to attempt to predict dissatisfaction at 3-years and 3-months post-surgery respectively using pre-surgery variables. Zabawa[170] created linear and logistic regression models, and performance was assessed and compared to the original Van Onsem[168] paper. This paper did not report any metrics for the logistic regression model. However, it was stated that the accuracy of the model was poor and that it did not support the claims made by Van Onsem[168] that the model could predict dissatisfaction. However, the Zabawa[170] study used an end-point of 3-years post-surgery and Van Onsem[168] used an end-point of 3-months. It is possible that this contributed to the poor performance of the model in the Zabawa[170] paper. Calkins[171] used the 3-month post-surgery time point and reported a model sensitivity of 0.939 (correctly predicted as satisfied) and a specificity of 0 (correctly predicted as dissatisfied). This indicates that the model was not able to identify any of the dissatisfied patients included in the sample.

2.6.1.1.4 Kunze 2019

Kunze[17] used a combination of 11 items to create a single checklist score ranging from -55 to 110 (worst to best) to predict post-surgery dissatisfaction. This score included BMI, drug allergies, osteophytes, patella thickness to soft tissue shadow skin thickness, flexion contracture, diabetes, opioid use, comorbidities, previous knee surgery, diagnosis for surgery, and smoking status/smoking history. The patient's checklist score, age, BMI, and gender were included in a linear and logistic regression model. Both linear and logistic regression models found that the checklist score was a significant predictor of dissatisfaction. No metrics were reported for either model. Using the checklist score a receiver operating characteristic (ROC) curve was created to evaluate the ability of the checklist to distinguish between satisfied and dissatisfied patients and an AUC 0.72 was reported. The authors claimed that a score of 96.5 on the checklist, provided a sensitivity of 0.975 (correctly predicted as dissatisfied), a specificity of 0.93 (correctly predicted as satisfied), and a NPV of 0.957. However, when examining the ROC curve included in the study, it is unclear how these sensitivity and specificity values were calculated or achieved. The ROC curve indicates a sensitivity of approximately 0.9 would correspond with a specificity of approximately 0.2. This is vastly different than the performance that the authors have reported. Additionally, it is unclear how satisfaction was measured or how a patient was determined to be satisfied or dissatisfied. Lastly there was no indication of how many dissatisfied patients were included in the study.

2.6.1.1.5 Summary

It appears that dissatisfaction cannot be consistently predicted pre-surgery using linear regression, logistic regression, or a checklist. Although Van Onsem[168], Garriga[16], and Kunze[17] claimed that their models were successful, the Van Onsem[168] and Garriga[16] models did not perform well when tested on new data. Additionally, both the Van Onsem[168] and Garriga[16] models appear to suffer from low accuracy and may not or only marginally outperform the no information rate. Because satisfaction rate was not reported, and it is unclear how the claimed specificity and sensitivity was achieved it is not possible to make this determination using the Kunze[17] model. When establishing a decision threshold, it is important to consider the priorities of the decision maker. Overall accuracy may be less important than high sensitivity or specificity. Although, it is still likely desirable that the model outperform the no information rate and achieve better accuracy than simply predicting that all patients are satisfied.

2.6.1.2 Machine Learning

Machine learning is a branch of artificial intelligence that employs algorithms that can learn from data. This allows the model building process to be partially automated. Machine learning methods are also typically more flexible than traditional statistical methods, however, they are often more difficult to interpret. Recently there has been a proliferation of the use of machine learning techniques in orthopaedic medicine. This has included attempts by Kunze[57] and Farooq[58] to use machine learning methods to predict dissatisfaction. These methods have been found to outperform standard logistic regression and have been reported to have good discriminatory capacity.

2.6.1.2.1 Kunze 2020

Kunze[57] compared the performance of a logistic regression model with a number of different machine learning methods. This study included 430 patients of which 0.09 (n=40) were dissatisfied. Patients were split into a training and test set of 0.70 and 0.30 respectively. The machine learning and statistical methods that were used were logistic regression, stochastic gradient boosting, random forest, support vector machine, neural network, and elastic-net logistic regression. All outcomes were reported based on test set performance. The random forest was found to be the optimal model with an AUC of 0.77, Brier score of 0.082, a calibration intercept of 0.093, and a calibration slope of 0.74. Using a relative influence plot, important predictors in the model were found to be age, surgeon completed Knee Society Knee Scoring System (KSS), pre-operative patient health state, and comorbidities. Kunze[57] did not provide a cut point or any information about sensitivity and specificity of their model. Based on a visual inspection of the ROC curve for the best performing model, it appears that despite high AUC values, their model may not outperform the no information rate or only do so marginally. However, the authors did complete a decision curve analysis which indicated that the model could be beneficial in helping to determine which patients should undergo surgery and which should not.

Problematically, this study only included 40 dissatisfied patients. If the training and test sets were created using a stratified split on the satisfaction variable, this would mean that models were trained using 28 and tested on 12 dissatisfied patients. This study may not have the necessary sample size or number of dissatisfied patients to make strong conclusions about the ability of a model to predict dissatisfaction.

2.6.1.2.2 Farooq 2020

Farooq[58] also compared the performance of logistic regression with a number of different models created using gradient boosted trees. This study included 897 patients with a dissatisfaction rate of

0.147 (n=132). The optimal model was found to be a gradient boosted tree which had an AUC of 0.81. The most important predictors of dissatisfaction in this study were grouped into high, middle, and low tiers of influence. The highest tier included the generation that the patient belonged to (baby boomer or generation x) and if the TKA was cruciate retaining, condylar stabilizing, or posterior stabilized. The middle tier included inflammatory conditions, scheduled narcotic use, the surgeon who completed the surgery, and depression. The bottom tier included, lumbar spine pain, sex, and PCL released or preserved. No calibration information was included for this model, however, a sensitivity and specificity of 0.73 (correctly predicted as dissatisfied) and 0.746 (correctly predicted as satisfied) respectively was reported. Using the dissatisfaction rate, sensitivity, and specificity scores included in this study, an overall accuracy of 0.743 was calculated. This means that if this model was used to predict patient dissatisfaction, it would perform 0.11 worse than the no information rate or simply predicting that all patients will be satisfied with their surgery. It may also be possible to improve the accuracy of the model by adjusting the cut point that distinguishes between satisfied and dissatisfied patients. Although, that is unclear based on the information provided by the study.

2.6.1.3 Summary

Machine learning models appear to show promise in predicting 1 and 2-year post-surgery dissatisfaction using pre-surgery and surgical variables. Both Kunze[57] and Farooq[58] have reported that their models successfully predict dissatisfaction. However, it appears that the models generated suffer from a similar problem as the logistic regression models. These models appear not to, or only marginally, outperform the no information rate. It remains unclear if these models can perform well enough to be useful in a clinical setting.

2.6.2 Post-surgery Early Identification of Dissatisfied Patients

Worse outcomes early in the recovery process have been associated with poorer outcomes at 1-year post-surgery. Williams[173] found that the OKS and flexion at 3-months were associated with 1-year post-surgery satisfaction. Bryan[174] found that mental and physical health at 6-months were associated with dissatisfaction at 1-year post-surgery. Young-Shand[175] indicated that differences in pain, health-related quality of life, and function between satisfied and dissatisfied patients may be observable as early as 6-weeks post-surgery. These studies did not evaluate the discriminatory or predictive ability of any of the identified variables.

Goh[169] evaluated the ability of the 6-month post-surgery OKS, surgeon completed KSS, and 36-Item Short Form Health Survey questionnaires to discriminate between satisfied and dissatisfied patients at 2-years post-surgery. The OKS and KSS knee score were found to discriminate best with AUCs of 0.76 and 0.704 respectively. This study did not create a multivariable model and only evaluated the discriminatory capacity of questionnaires. While this study appears to show that these variables can successfully discriminate between satisfied and dissatisfied patients it is unclear how useful they would be in a clinical setting.

2.6.2.1 Summary

Despite the apparent success of the variables included in the Goh[169] study, more work needs to be done to evaluate how well patient dissatisfaction can be predicted early in the recovery process. It appears that differences between satisfied and dissatisfied patients can be identified as early as 3-months post-surgery. It is yet to be determined if this can be an effective approach to identifying dissatisfied patients.

2.7 Statement of the Problem

2.7.1 Pre-Surgery Prediction of Satisfaction

Based on the current literature, it is unclear if post-surgery dissatisfaction can be effectively predicted using pre-surgery and surgical variables. When predicting dissatisfaction, machine learning models appear to outperform standard logistic regression. However, it is difficult to determine how effective and clinically useful these models are. There are a number of avenues that could potentially improve the performance of pre-surgery prediction models that remain unexplored. Specifically, PROMs variables have not been used in conjunction with machine learning methods. Additionally, none of the current research has been performed in Canada. Using Canadian data to determine what works and what does not in predicting 1-year post-surgery dissatisfaction using pre-surgery and surgical variables could allow for a reduction in the number of dissatisfied patients.

2.7.2 Early Post-Surgery Identification of Dissatisfied Patients

A single study has evaluated the ability of prediction models to identify patients likely to be dissatisfied early in the recovery process. However, this study did not use multivariable models, included a limited number of PROMs, was not completed in Canada, and used 6-month post-surgery data to predict 2-year post-surgery dissatisfaction. The literature suggests that factors affecting dissatisfaction at 1-year post-surgery can potentially be identified at 6-weeks or 3-months post-surgery. Additionally, this paper did not consider how easily their model could be transported to a clinical setting. If patients likely to be dissatisfied can be identified early in the recovery process, it may be possible to implement interventions that improve satisfaction.

2.7.3 The Met Expectations Variable and Satisfaction Interventions

Met expectations have been criticized as simply being another performance variable (see section 2.4.1.5). As met expectations are central to the interventions being performed to improve patient satisfaction, it is important to understand what the met expectations variable is measuring. Met expectations have not been well defined in the TKA literature and very little work has been done to determine if the met expectations variable is measuring what it claims to measure. There have been a limited number of attempts to improve satisfaction through interventions that aim to generate realistic expectations for patients and improve their met expectations scores. Despite the mixed results of the interventions this appears to be a potentially fruitful avenue to improve satisfaction. Generating a better understanding of the relationship between met expectations and satisfaction could aid in the development of successful interventions.

2.8 References

1. Adie S, Harris I, Chuan A, Lewis P, Naylor JM. Selecting and optimising patients for total knee arthroplasty. Med J Aust. 2019;210(3):135–41.
2. Wright JGJG, Hawker GAGA, Hudak PL, Glazier RH, Mahomed NN, Kreder HJ, et al. Variability in physician opinions about the indications for knee arthroplasty. J Arthroplasty. 2011;26(4):569–75.
3. Van Manen MD, Nace J, Mont MA. Management of primary knee osteoarthritis and indications for total knee arthroplasty for general practitioners. J Am Osteopath Assoc. 2012;112(11):709–15.
4. Conner-Spady BL, Marshall DA, Hawker GA, Bohm E, Dunbar MJ, Frank C, et al. You'll know when you're ready: A qualitative study exploring how patients decide when the time is right for joint replacement surgery. BMC Health Serv Res. 2014;14(1):1–10.
5. Bourne RB, Chesworth BM, Davis AM, Mahomed NN, Charron KDJ, Met D. Patient satisfaction after total knee arthroplasty who is satisfied and who is not? Clin Orthop Relat Res. 2010;468(1):57–63.
6. Santaguida P, Hawker G, Hudak P, Glazier R, Mahomed N, Kreder H, et al. Patient characteristics affecting the prognosis of THR and TKR-a systematic review.pdf. Can J Surg. 2008;51(6):428.
7. Overgaard A, Lidgren L, Sundberg M, Robertsson O, W-Dahl A. Patient-reported 1-year outcome not affected by body mass index in 3,327 total knee arthroplasty patients. Acta Orthop. 2019;90(4):360–5.
8. Lingard EA, Katz JN, Wright EA, Sledge CB. Predicting the outcome of total knee arthroplasty. J Bone Jt Surg - Ser A. 2004;86(10):2179–86.
9. Hawker GA, Badley EM, Borkhoff CM, Croxford R, Davis AM, Dunn S, et al. Which patients are most likely to benefit from total joint arthroplasty? Arthritis Rheum. 2013;65(5):1243–52.
10. Baker PN, Rushton S, Jameson SS, Reed M, Gregg P, Deehan DJ. Patient satisfaction with total knee replacement cannot be predicted from pre-operative variables alone: a cohort study from the National Joint Registry for England and Wales. Bone Joint J. 2013;95(10):1359–65.
11. Youlden DJ, Dannaway J, Enke O. Radiographic severity of knee osteoarthritis and its relationship to outcome post total knee arthroplasty: a systematic review. ANZ J Surg. 2020;90(3):237–42.
12. Jones CA, Voaklander DC, Suarez-Almazor ME. Determinants of function after total knee arthroplasty. Phys Ther. 2003;83(8):696–706.

13. Kuijer PPFM, Kievit AJ, Pahlplatz TMJ, Hooiveld T, Hoozemans MJM, Blankevoort L, et al. Which patients do not return to work after total knee arthroplasty? Rheumatol Int. 2016;36(9):1249–54.
14. De Beer J, Petruccelli D, Gandhi R, Winemaker M. Primary total knee arthroplasty in patients receiving workers' compensation benefits. Can J Surg. 2005;48(2):100–5.
15. Lizaur-utrilla A, Martinez-mendez D, Miralles-muñoz FA, Marco-gomez L, Lopez-prats FA. Negative impact of waiting time for primary total knee arthroplasty on satisfaction and patient-reported outcome. Int Orthop. 2016;40(11):2303–7.
16. Garriga C, Sanchez-Santos MT, Judge A, Perneger T, Hannouche D, Lübbeke A, et al. Development of a model predicting non-satisfaction 1 year after primary total knee replacement in the UK and transportation to Switzerland. Sci Rep. 2018;8(1):1–8.
17. Kunze KN, Akram F, Fuller BC, Zabawa L, Sporer SM, Levine BR. Internal validation of a predictive model for satisfaction after primary total knee arthroplasty. J Arthroplasty. 2019;34(4):663–70.
18. Katz JN, Losina E, Smith SR, Bido J, Collins JE, Yang H, et al. Impact of preoperative opioid use on total knee arthroplasty outcomes. J Bone Jt Surg - Am Vol. 2017;99(10):803–8.
19. Hawker GA, Conner-Spady BL, Bohm E, Dunbar MJ, Jones CA, Ravi B, et al. The relationship between patient-reported readiness for total knee arthroplasty and likelihood of a good outcome at one year. Arthritis Care Res. 2021;
20. Kennedy DM, Stratford PW, Riddle DL, Hanna SE, Gollish JD. Assessing recovery and establishing prognosis following total knee arthroplasty. Phys Ther. 2008;88(1):22–32.
21. Mizner RL, Petterson SC, Snyder-Mackler L. Quadriceps strength and the time course of functional recovery after total knee arthroplasty. J Orthop Sport Phys Ther. 2005;35(7):424–36.
22. Ramkumar PN, Navarro SM, Haeberle HS, Ng M, Piuzzi NS, Spindler KP. No difference in outcomes 12 and 24 months after lower extremity total joint arthroplasty: a systematic review and meta-analysis. J Arthroplasty. 2018;33(7):2322–9.
23. Healy WL, Della Valle CJ, Iorio R, Berend KR, Cushner FD, Dalury DF, et al. Complications of total knee arthroplasty: Standardized list and definitions of the Knee Society. Clin Orthop Relat Res. 2013;471(1):215–20.
24. Le DH, Goodman SB, Maloney WJ, Huddleston JI. Current modes of failure in TKA: Infection, instability, and stiffness predominate. Clin Orthop Relat Res. 2014;472(7):2197–200.
25. Insall JN, Dorr D, Scott R, Scott W. Rationale of The Knee Society Clinical Rating System. Clin Orthop Relat Res. 1989;26(4):13–4.

26. Kahlenberg CA, Nwachukwu BU, Mclawhorn AS, Cross MB, Cornell CN, Padgett DE. Patient satisfaction after total knee replacement: a systematic review. HSS J. 2018;14(2):192–201.
27. Nakano N, Shoman H, Olavarria F, Matsumoto T, Kuroda R, Khanduja V. Why are patients dissatisfied following a total knee replacement? A systematic review. Int Orthop. 2020;44(10):1971–2007.
28. Graham B, Green A, James M, Katz J, Swiontkowski M. Measuring patient satisfaction in orthopaedic surgery. JBJS. 2015;97(1):80–4.
29. Hafkamp FJ, Gosens T, de Vries J, den Oudsten BL. Do dissatisfied patients have unrealistic expectations? A systematic review and best-evidence synthesis in knee and hip arthroplasty patients. EFORT Open Rev. 2020;5(4):226–40.
30. Deakin AH, Smith MA, Wallace DT, Smith EJ, Sarungi M. Fulfilment of preoperative expectations and postoperative patient satisfaction after total knee replacement. A prospective analysis of 200 patients. Knee. 2019;26(6):1403–12.
31. Conner-Spady BL, Bohm E, Loucks L, Dunbar MJ, Marshall DA, Noseworthy TW. Patient expectations and satisfaction 6 and 12 months following total hip and knee replacement. Qual Life Res. 2020;29(3):705–19.
32. Hawker GA, Conner-Spady BL, Bohm E, Dunbar MJ, Jones CA, Ravi B, et al. Patients' preoperative expectations of total knee arthroplasty and satisfaction with outcomes at one year: A prospective cohort study. Arthritis Rheumatol. 2021;73(2):223–31.
33. Hamilton DF, Loth FL, MacDonald DJ, Giesinger K, Patton JT, Simpson AH, et al. Treatment success following joint arthroplasty: Defining thresholds for the Oxford Hip and Knee Scores. J Arthroplasty. 2018;33(8):2392–7.
34. Giesinger JM, Hamilton DF, Jost B, Behrend H, Giesinger K. WOMAC, EQ-5D and Knee Society Score thresholds for treatment success after total knee arthroplasty. J Arthroplasty. 2015;30(12):2154–8.
35. Clement ND, Bardgett M, Weir D, Holland J, Gerrand C, Deehan DJ. What is the minimum clinically important difference for the WOMAC index after TKA? Clin Orthop Relat Res. 2018;476(10):2005–14.
36. Walker LC, Clement ND, Bardgett M, Weir D, Holland J, Gerrand C, et al. The WOMAC score can be reliably used to classify patient satisfaction after total knee arthroplasty. Knee Surgery, Sport Traumatol Arthrosc. 2018;26(11):3333–41.

37. Pham T, van der Heijde D, Altman RD, Anderson JJ, Bellamy N, Hochberg M, et al. OMERACT-OARSI initiative: Osteoarthritis research society international set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthr Cartil. 2004;12(5):389–99.
38. Clement ND, Bardgett M, Weir D, Holland J, Gerrand C, Deehan DJ. The rate and predictors of patient satisfaction after total knee arthroplasty are influenced by the focus of the question: a standard satisfaction question is required. Bone Joint J. 2018;100(6):740–8.
39. Kiran A, Bottomley N, Biant LC, Tr F, Javaid MK, Carr AJ, et al. Variations in good patient reported outcomes after total knee arthroplasty. J Arthroplasty. 2015;30(8):1364–71.
40. Huijbregts HJTAM, Khan RJK, Fick DP, Jarrett OM, Haebich S. The knee prosthetic alignment after total knee replacement is not associated with dissatisfaction or change in oxford knee score a multivariable regression analysis. Knee. 2016;23(3):535–9.
41. Scott CEH, Howie CR, Macdonald D, Biant LC. Predicting dissatisfaction following total knee replacement: a prospective study of 1217 patients. J Bone Joint Surg Br. 2010;92(9):1253–8.
42. Mannion AF, Kämpfen S, Munzinger U, Quervain IK. The role of patient expectations in predicting outcome after total knee arthroplasty. Arthritis Res Ther. 2009;11(5):1–13.
43. Baker PN, van der Meulen JH, Lewsey J, Gregg PJ. The role of pain and function in determining patient satisfaction after total knee replacement: data from the National Joint Registry for England and Wales. J Bone Jt Surg. 2007;89(7):893–900.
44. Franklin PD, Karbassi JA, Li W, Yang W, Ayers DC. Reduction in narcotic use after primary total knee arthroplasty and association with patient pain relief and satisfaction. J Arthroplasty. 2010;25(6):12–6.
45. Du H, Tang H, Gu J, Zhou Y. Patient satisfaction after posterior-stabilized total knee arthroplasty: a functional specific analysis. Knee. 2014;21(4):866–70.
46. Schnurr C, Jarrous M, Güdden I, Eysel P, König DP. Pre-operative arthritis severity as a predictor for total knee arthroplasty patients’ satisfaction. Int Orthop. 2013;37(7):1257–61.
47. Keurentjes JC, Blane D, Bartley M, Keurentjes JJB, Fiocco M, Nelissen RG. Socio-economic position has no effect on improvement in health-related quality of life and patient satisfaction in total hip and knee replacement : a cohort study. PLoS One. 2013;8(3):1–12.
48. Jacobs CA, Christensen CP, Karthikeyan T. Patient and intraoperative factors influencing satisfaction two to five years after primary total knee arthroplasty. J Arthroplasty. 2014;29(8):1576–9.

49. Thambiah MD, Nathan S, Seow BZX, Liang S, Lingaraj K. Patient satisfaction after total knee arthroplasty: an Asian perspective. Singapore Med J. 2015;56(5):259–63.
50. Lingard EA, Sledge CB, Learmonth ID, Kinemax Outcomes Group. Patient expectations regarding total knee arthroplasty: differences among the United States, United Kingdom, and Australia. JBJS. 2006;88(6):1201–7.
51. Pronk Y, Peters MCWM, Brinkman JM. Is patient satisfaction after total knee arthroplasty predictable using patient characteristics and preoperative patient-reported outcomes? J Arthroplasty. 2021;36(7):2458–65.
52. Noble PC, Conditt MA, Cook KF, Mathis KB. The John Insall Award: patient expectations affect satisfaction with total knee arthroplasty. Clin Orthop Relat Res. 2006;452:35–43.
53. Bergkvist LI, Rossiter JR. The predictive validity of multiple-item versus single-item measures of the same constructs. J Mark Res. 2007;44(2):175–84.
54. Jacobs CA, Christensen CP. Factors influencing patient satisfaction two to five years after primary total knee arthroplasty. J Arthroplasty. 2014;29(6):1189–91.
55. Bullens PHJ, Loon CJM Van, Malefijt MCDW, Laan RFJM, Veth RPH. Patient satisfaction after total knee arthroplasty a comparison between subjective and objective outcome assessments. J Arthroplasty. 2001;16(6):740–7.
56. Kim TK, Chang CB, Kang YG, Kim SJ, Seong SC. Causes and predictors of patient's dissatisfaction after uncomplicated total knee arthroplasty. J Arthroplasty. 2009;24(2):263–71.
57. Kunze KN, Polce EM, Sadauskas AJ, Levine BR. Development of machine learning algorithms to predict patient dissatisfaction after primary total knee arthroplasty. J Arthroplasty. 2020;35(11):3117–22.
58. Farooq H, Deckard ER, Ziemba-Davis M, Madsen A, Meneghini RM. Predictors of patient satisfaction following primary total knee arthroplasty: results from a traditional statistical model and a machine learning algorithm. J Arthroplasty. 2020;35(11):3123–30.
59. Harrell F. Regression modeling strategies. Second. New York: Springer Science+Business Media; 2015.
60. Chatterjee S, Hadi A. Regression by example. Fourth Edi. Hoboken, NJ: Wiley & Sons Inc.; 2006. 53–348 p.
61. Kuhn M, Johnson K. Applied predictive modeling. New York: Springer Science+Business Media; 2013.

62. Clement ND, MacDonald D, Howie CR, Biant LC. The outcome of primary total hip and knee arthroplasty in patients aged 80 years or more. J Bone Jt Surg. 2011;93(9):1265–70.
63. Goh GS, Liow MHL, Chen JY, Tay DKJ, Lo NN, Yeo SJ. Can octogenarians undergoing total knee arthroplasty experience similar functional outcomes, quality of life, and satisfaction rates as their younger counterparts? A propensity score matched analysis of 1188 patients. J Arthroplasty. 2020;35(7):1833–9.
64. Williams DP, Price AJ, Beard DJ, Hadfield SG, Arden NK, Murray DW, et al. The effects of age on patient-reported outcome measures in total knee replacements. Bone Joint J. 2008;95(1):38–44.
65. Scott CEH, Oliver WM, MacDonald D, Wade FA, Moran M, Breusch SJ. Predicting dissatisfaction following total knee arthroplasty in patients under 55 years of age. Bone Jt J. 2016;98(12):1625–34.
66. Parvizi J, Nunley RM, Berend KR, Lombardi A V, Ms ELR, Clohisy JC, et al. High level of residual symptoms in young patients after total knee arthroplasty. Clin Orthop Relat Res. 2014;472(1):133–7.
67. Ritter MA, Wing JT, Berend ME, Davis KE, Meding JB. The clinical effect of gender on outcome of total knee arthroplasty. J Arthroplasty. 2008;23(3):331–6.
68. Rissolio L, Sabatini L, Risitano S, Bistolfi A, Galluzzo U, Massè A, et al. Is it the surgeon, the patient, or the device? A comprehensive clinical and radiological evaluation of factors influencing patient satisfaction in 648 total knee arthroplasties. J Clin Med. 2021;10(12):2599.
69. Merle-Vincent F, Couris CM, Schott AM, Conrozier T, Piperno M, Mathieu P, et al. Factors predicting patient satisfaction 2 years after total knee arthroplasty for osteoarthritis. Jt Bone Spine. 2011;78(4):383–6.
70. Fisher DA, Dierckman B, Watts MR, Davis K. Looks good but feels bad: factors that contribute to poor results after total knee arthroplasty. J Arthroplasty. 2007;22(6):6–9.
71. Clement ND, Macdonald D, Simpson AHRW, Burnett R. Total knee replacement in patients with concomitant back pain results in a worse functional outcome and a lower rate of satisfaction. Bone Joint J. 2013;95(12):1632–9.
72. Collados-Maestre I, Lizaur-Utrilla A, Martinez-Mendez D, Marco-Gomez L, Lopez-Prats FA. Concomitant low back pain impairs outcomes after primary total knee arthroplasty in patients over 65 years: a prospective, matched cohort study. Arch Orthop Trauma Surg. 2016;136(12):1767–71.

73. Robertsson O, Dunbar M, Pehrsson T, Knutson K, Lidgren L. Patient satisfaction after knee arthroplasty: a report on 27,372 knees operated on between 1981 and 1995 in Sweden. Acta Orthop Scand. 2000;71(3):262–7.

74. Kennedy LG, Newman JH, Ackroyd CE, Dieppe PA. When should we do knee replacements? Knee. 2003;10(2):161–6.

75. Alosh H, Behery OA, Levine BR. Radiographic predictors of patient satisfaction following primary total knee arthroplasty. Bull Hosp Joint Dis. 2018;76(2):105–12.

76. Liebensteiner M, Wurm A, Gamper D, Oberaigner W, Dammerer D, Krismer M. Patient satisfaction after total knee arthroplasty is better in patients with pre-operative complete joint space collapse. Int Orthop. 2019;43(8):1841–7.

77. Jacobs CA, Christensen CP, Karthikeyan T. An intact anterior cruciate ligament at the time of posterior cruciate ligament–retaining total knee arthroplasty was associated with reduced patient satisfaction and inferior pain and stair function. J Arthroplasty. 2016;31(8):1732–5.

78. Duncan GJ, Daly MC, Mcdonough P, Williams DR. Optimal indicators of socioeconomic status for health research. Am J Public Health. 2002;92(7):1151–7.

79. Dowsey MM, Nikpour M, Choong PFM. Outcomes following large joint arthroplasty: does socio-economic status matter? BMC Musculoskelet Disord. 2014;15(1):1–12.

80. Barrack RL, Ruh EL, Chen J, Lombardi A V, Berend KR, Parvizi J, et al. Impact of socioeconomic factors on outcome of total knee arthroplasty. Clin Orthop Relat Res. 2014;472(1):86–97.

81. Jorn LP, Johnsson R, Toksvig-Larsen S. Patient satisfaction, function and return to work after knee arthroplasty. Acta Orthop Scand. 1999;70(4):343–7.

82. Kamp T, Brouwer S, Hylkema TH, van Beveren J, Rijk PC, Brouwer RW, et al. Psychosocial Working Conditions Play an Important Role in the Return-to-Work Process After Total Knee and Hip Arthroplasty. J Occup Rehabil. 2021;

83. Styron JF, Barsoum WK, Smyth KA, Singer ME. Preoperative predictors of returning to work following primary total knee arthroplasty. J Bone Jt Surg - Ser A. 2011;93(1):2–10.

84. Harris IA, Dao ATT, Young JM, Solomon MJ, Jalaludin BB. Predictors of patient and surgeon satisfaction after orthopaedic trauma. Injury. 2009;40(4):377–84.

85. Bowling A, Rowe G, Lambert N, Waddington M, Mahtani KR, Kenten C, et al. The measurement of patients' expectations for health care: a review and psychometric testing of a measure of patients' expectations. Health Technol Assess (Rockv). 2012;16(30).

86. Mondloch M V, Cole DC, Frank JW. Does how you do depend on how you think you'll do? A systematic review of the evidence for a relation between patients' recovery expectations and health outcomes. CMAJ. 2001;165(2):174–9.

87. Vissers MM, Groot IB De, Reijman M, Bussmann JB, Stam HJ, Verhaar JAN. Functional capacity and actual daily activity do not contribute to patient satisfaction after total knee arthroplasty. BMC Musculoskelet Disord. 2010;11(1):1–8.

88. Becker R, Döring C, Denecke A, Brosz M. Expectation, satisfaction and clinical outcome of patients after total knee arthroplasty. Knee surgery, Sport Traumatol Arthrosc. 2011;19(9):1433–41.

89. Barlow T, Clark T, Dunbar M, Metcalfe A, Griffin D. The effect of expectation on satisfaction in total knee replacements: a systematic review. Springerplus. 2016;5(1):1–7.

90. Nilsdotter AK, Toksvig-Larsen S, Roos EM. Knee arthroplasty: are patients' expectations fulfilled? A prospective study of pain and function in 102 patients with 5-year follow-up. Acta Orthop. 2009;80(1):55–61.

91. Gandhi R, Davey JR, Mahomed N. Patient expectations predict greater pain relief with joint arthroplasty. J Arthroplasty. 2009;24(5):716–21.

92. Noble PC, Scuderi GR, Brekke AC, Sikorskii A, Benjamin JB, Lonner JH, et al. Development of a new Knee Society scoring system. Clin Orthop Relat Res. 2012;470(1):20–32.

93. Oliver RL. Satisfaction: A behavioural perspective on the consumer. Second Edi. New York, NY: M.E. Sharpe; 2010. 61–89 p.

94. Gong L, Dong J-Y. Patient's personality predicts recovery after total knee arthroplasty: a retrospective study. J Orthop Sci. 2014;19(2):263–9.

95. Giurea A, Fraberger G, Kolbitsch P, Lass R, Schneider E, Kubista B, et al. The impact of personality traits on the outcome of total knee arthroplasty. Biomed Res Int. 2016;2016.

96. Ramaesh R, Jenkins P, Lane J V, Knight S, Macdonald D, Howie C. Personality, function and satisfaction in patients undergoing total hip or knee replacement. J Orthop Sci. 2014;19(2):275–81.

97. Singh JA, Lewallen DG. Depression in primary TKA and higher medical comorbidities in revision TKA are associated with suboptimal subjective improvement in knee function. BMC Musculoskelet Disord. 2014;15(1):1–11.

98. Ali A, Lindstrand A, Sundberg M, Flivik G. Preoperative anxiety and depression correlate with

dissatisfaction after total knee arthroplasty: a prospective longitudinal cohort study of 186 patients , with 4-year follow-up. J Arthroplasty. 2017;32(3):767–70.
99. Gandhi R, Davey JR, Mahomed NN. Predicting patient dissatisfaction following joint replacement surgery. J Rheumatol. 2008;35(12):2415–8.
100. Ali A, Sundberg M, Robertsson O, Dahlberg LE, Thorstensson CA, Redlund-Johnell I, et al. Dissatisfied patients after total knee arthroplasty: a registry study involving 114 patients with 8–13 years of follow-up. Acta Orthop. 2014;85(3):229–33.
101. Duivenvoorden T, Vissers MM, Verhaar JAN, Busschbach JJ V, Gosens T, Bloem RM, et al. Anxiety and depressive symptoms before and after total hip and knee arthroplasty: a prospective multicentre study. Osteoarthr Cartil. 2013;21(12):1834–40.
102. Dhurve K, Scholes C, El-tawil S, Shaikh A, Weng LK, Levin K, et al. Multifactorial analysis of dissatisfaction after primary total knee replacement. Knee. 2017;24(4):856–62.
103. Bierke S, Petersen W. Influence of anxiety and pain catastrophizing on the course of pain within the first year after uncomplicated total knee replacement: a prospective study. Arch Orthop Trauma Surg. 2017;137(12):1735–42.
104. Bierke S, Häner M, Petersen W. Influence of somatization and depressive symptoms on the course of pain within the first year after uncomplicated total knee replacement: a prospective study. Int Orthop. 2016;40(7):1353–60.
105. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: development and validation. Psychol Assess. 1995;7(4):524–32.
106. Ross TP, Ross LT, Short SD. The multidimensional health locus of control scale: psychometric properties and form equivalence. Psychol Rep. 2015;116(3):889–913.
107. Gandhi R, de Beer J, Petruccelli D, Winemaker M. Does patient perception of alignment affect total knee arthroplasty outcome? Can J Surg. 2006;50(3):181–6.
108. Gunaratne R, Pratt DN, Banda J, Fick DP, Khan RJK, Robertson BW. Patient dissatisfaction following total knee arthroplasty: a systematic review of the literaturere. J Arthroplasty. 2017;32(12):3854–60.
109. Chua JL, Goh GS-H, Liow MHL, Tay DK-JT, Lo N-N Lo, Yeo S-J. Modern TKA implants are equivalent to traditional TKA implants in functional and patellofemoral joint-related outcomes. Knee Surgery, Sport Traumatol Arthrosc. 2019;27(4):1116–23.
110. Baker PN, Deehan DJ, Lees D, Jameson S, Avery PJ, Gregg PJ, et al. The effect of surgical factors

on early patient-reported outcome measures (PROMS) following total knee replacement. J Bone Joint Surg Br. 2012;94(8):1058–66.

111. Hinarejos P, Puig-Verdie L, Leal J, Pelfort X, Torres-Claramunt R, Sánchez-Soler J, et al. No differences in functional results and quality of life after single-radius or multiradius TKA. Knee Surgery, Sport Traumatol Arthrosc. 2016;24(8):2634–40.

112. Dixon S, Blom AW, Whitehouse MR, Wylde V. Comparison of patient reported outcomes after Triathlon® and Kinemax Plus prostheses. Ann R Coll Surg Engl. 2014;96(1):61–6.

113. Jain NP, Lee SY, Morey VM, Chong S, Kang YG, Kim TK. Early clinical outcomes of a new posteriorly stabilized total knee arthroplasty prosthesis: comparisons with two established prostheses. Knee Surg Relat Res. 2017;29(3):180–8.

114. Choi WC, Lee S, Seong SC, Jung JH, Lee MC. Comparison between standard and high-flexion posterior-stabilized rotating-platform mobile-bearing total knee arthroplasties: A randomized controlled study. J Bone Jt Surg - Ser A. 2010;92(16):2634–42.

115. Lizaur-utrilla A, Sanz-Reig J, Trigueros-Rentero MA. Greater satisfaction in older patients with a mobile-bearing compared with fixed-bearing total knee arthroplasty. J Arthroplasty. 2012;27(2):207–12.

116. Wylde V, Learmonth I, Potter A, Bettinson K, Lingard E. Patient-reported outcomes after fixed-versus mobile-bearing total knee replacement: a multi-centre randomised controlled trial using the Kinemax total knee replacement. J Bone Joint Surg Br. 2008;90(9):1172–9.

117. Collados-Maestre I, Lizaur-Utrilla A, Gonzalez-Navarro B, Miralles-Muñoz FA, Marco-Gomez L, Lopez-Prats FA, et al. Better functional outcome after single-radius TKA compared with multi-radius TKA. Knee Surgery, Sport Traumatol Arthrosc. 2017;25(11):3508–14.

118. Nishio Y, Onodera T, Kasahara Y, Takahashi D, Iwasaki N, Majima T. Intraoperative Medial Pivot Affects Deep Knee Flexion Angle and Patient-Reported Outcomes After Total Knee Arthroplasty. J Arthroplasty. 2014;29(4):702–6.

119. Boese C, Gallo T, Plantikow C. Range of motion and patient satisfaction with traditional and high-flexion rotating-platform knees. Iowa Orthop J. 2011;31:73–7.

120. Kim YH, Choi Y, Kim J-S. Range of motion of standard and high-flexion posterior cruciate-retaining total knee prostheses: A prospective randomized study. J Bone Jt Surg - Ser A. 2009;91(8):1874–81.

121. Fricka KB, Sritulanondha S, McAsey CJ. To cement or not? Two-year results of a prospective,

randomized study comparing cemented vs. cementless total knee arthroplasty (TKA). J Arthroplasty. 2015;30(9):55–8.

122. Reimann P, Brucker M, Arbab D, Lüring C. Patient satisfaction - A comparison between patient-specific implants and conventional total knee arthroplasty. J Orthop. 2019;16(3):273–7.

123. Hernandez-Vaquero D, Noriega-Fernandez A, Suarez-Vazquez A. Total knee arthroplasties performed with a mini-incision or a standard incision. Similar results at six months follow-up. BMC Musculoskelet Disord. 2010;11(1):1–8.

124. Nam D, Nunley RM, Berend KR, Lombardi A V, Barrack RL. The impact of custom cutting guides on patient satisfaction and residual symptoms following total knee arthroplasty. Knee. 2016;23(1):144–8.

125. Jawhar A, Skeirek D, Stetzelberger V, Kollowa K, Obertacke U. No effect of tourniquet in primary total knee arthroplasty on muscle strength, functional outcome, patient satisfaction and health status: a randomized clinical trial. Knee Surgery, Sport Traumatol Arthrosc. 2020;28(4):1045–54.

126. Ali A, Lindstrand A, Nilsdotter A, Sundberg M, Ali A, Lindstrand A, et al. Similar patient-reported outcomes and performance after total knee arthroplasty with or without patellar resurfacing: A randomized study of 74 patients with 6 years of follow-up. Acta Orthop. 2016;87(3):274–9.

127. Burnett RSJ, Boone JL, Rosenzweig S, Steger-may K, Barrack RL. Patellar resurfacing compared with nonresurfacing in total knee arthroplasty. J Bone Jt Surg. 2009;91(11):2562–7.

128. Aunan E, Næss G, Clarke-Jenssen J, Sandvik L, Kibsgard TJ. Patellar resurfacing in total knee arthroplasty: Functional outcome differs with different outcome scores. Acta Orthop. 2016;87(2):158–64.

129. Kaseb MH, Mortazavi J, Firoozabadi MA, Toofan H. Comparison between patellar resurfacing and retention in total knee arthroplasty regarding the postoperative satisfaction of patients and patellar crepitus. Arch Bone Jt Surg. 2019;7(5):441.

130. Roberts DW, Hayes TD, Tate CT, Lesko JP. Selective patellar resurfacing in total knee arthroplasty: a prospective, randomized, double-blind study. J Arthroplasty. 2015;30(2):216–22.

131. Waters TS, Bentley G. Patellar resurfacing in total knee arthroplasty: a prospective, randomized study. J Bone Jt Surg. 2003;85(2):212–7.

132. Kuster MS, Bitschnau B, Votruba T. Influence of collateral ligament laxity on patient satisfaction after total knee arthroplasty: A comparative bilateral study. Arch Orthop Trauma Surg. 2004;124(6):415–7.

133. Kawahara S, Okazaki K, Matsuda S, Nakahara H. Internal rotation of femoral component affects functional activities after TKA - Survey with the 2011 knee society score. J Arthroplasty. 2014;29(12):2319–23.

134. Abhari S, Hsing TM, Malkani MM, Smith AF, Smith LS, Mont MA, et al. Patient satisfaction following total knee arthroplasty using restricted kinematic alignment. Bone Jt J. 2021;103(7):59–66.

135. de Grave PW, Luyckx T, Claeys K, Tampere T, Kellens J, Müller J, et al. Higher satisfaction after total knee arthroplasty using restricted inverse kinematic alignment compared to adjusted mechanical alignment. Knee Surgery, Sport Traumatol Arthrosc. 2020;1–12.

136. Blum CL, Lepkowsky E, Hussein A, Wakelin EA, Plaskos C, Koenig JA. Patient expectations and satisfaction in robotic-assisted total knee arthroplasty: a prospective two-year outcome study. Arch Orthop Trauma Surg. 2021;141(12):2155–64.

137. Scuderi GR, Bourne RB, Noble PC, Benjamin JB, Lonner JH, Scott WN. The new Knee Society Knee Scoring System. Clin Orthop Relat Res. 2011;470(1):3–19.

138. Nazzal MI, Bashaireh KH, Alomari MA, Nazzal MS, Maayah MF, Mesmar M. Relationship between improvements in physical measures and patient satisfaction in rehabilitation after total knee arthroplasty. Int J Rehabil Res. 2012;35(2):94–101.

139. Van Onsem S, Verstraete M, Dhont S, Zwaenepoel B, Van Der Straeten C. Improved walking distance and range of motion predict patient satisfaction after TKA. Knee Surgery, Sport Traumatol Arthrosc. 2018;26(11):3272–9.

140. Razmjou H, Boljanovic D, Wright S, Murnaghan J, Holtby R. Association between neuropathic pain and reported disability after total knee arthroplasty. Physiother Canada. 2015;67(4):311–8.

141. Maratt JD, Lee Y, Lyman S, Westrich GH. Predictors of satisfaction following total knee arthroplasty. J Arthroplasty. 2015;30(7):1142–5.

142. Clement ND, Weir DJ, Holland J, Deehan DJ. Contralateral knee pain reduces the rate of patient satisfaction but does not clinically impair the change in WOMAC score after total knee arthroplasty. Bone Jt J. 2020;102(1):125–31.

143. van Loon C, Baas N, Huey V, Lesko J, Meermans G, Vergroesen D. Early outcomes and predictors of patient satisfaction after TKA: a prospective study of 200 cases with a contemporary cemented rotating platform implant design. J Exp Orthop. 2021;8(30):1–7.

144. Lützner J, Beyer F, Günther KP, Huber J. Higher treatment effect after total knee arthroplasty is

associated with higher patient satisfaction. Knee Surgery, Sport Traumatol Arthrosc. 2021;29(10):3426–32.

145. Mahomed N, Gandhi R, Daltroy L, Katz JN. The self-administered patient satisfaction scale for primary hip and knee arthroplasty. Arthritis. 2011;2011:1–6.

146. Nakahara H, Okazaki K, Mizu-uchi H, Hamai S, Tashiro Y, Matsuda S, et al. Correlations between patient satisfaction and ability to perform daily activities after total knee arthroplasty: why aren't patients satisfied? J Orthop Sci. 2015;20(1):87–92.

147. Chang MJ, Kim SH, Kang YG, Chang CB, Kim TK. Activity levels and participation in physical activities by Korean patients following total knee arthroplasty. BMC Musculoskelet Disord. 2014;15(1):1–6.

148. Bonnin M, Laurent JR, Parratte S, Zadegan F, Badet R, Bissery A. Can patients really do sport after TKA? Knee Surgery, Sport Traumatol Arthrosc. 2010;18(7):853–62.

149. Witjes S, Geenen RCI Van, Koenraadt KLM, van der Hart CP, Blankevoort L, Kerkhoffs GM, et al. Expectations of younger patients concerning activities after knee arthroplasty: are we asking the right questions? Qual Life Res. 2017;26(2):403–17.

150. Matsuda S, Kawahara S. Postoperative alignment and ROM affect patient satisfaction after TKA. Clin Orthop Relat Res. 2013;471(1):127–33.

151. Goudie ST, Deakin AH, Ahmad A, Maheshwari R, Picard F. Flexion contracture following primary total knee arthroplasty: Risk factors and outcomes. Orthopedics. 2011;34(12):855–60.

152. Devers BN, Conditt MA, Jamieson ML, Driscoll MD, Noble PC, Parsley BS. Does greater knee flexion increase patient function and satisfaction after total knee arthroplasty? J Arthroplasty. 2011;26(2):178–86.

153. Kubo M, Maeda T, Kumagai K, Amano Y, Kawasaki T, Imai S. Good postoperative flexion angle improves knee function and improvement of flexion angle increases patient satisfaction after total knee arthroplasty. J Arthroplasty. 2021;36(9):3137–40.

154. Bonnefoy-Mazure A, Favre T, Praplan G, Armand S, Sagawa Junior Y, Hannouche D, et al. Associations between gait analysis parameters and patient satisfaction one year following primary total knee arthroplasty. Gait Posture. 2020;80(December 2019):44–8.

155. Clement ND, Burnett R. Patient satisfaction after total knee arthroplasty is affected by their general physical well-being. Knee Surgery, Sport Traumatol Arthrosc. 2013;21(11):2638–46.

156. Baumann C, Rat AC, Osnowycz G, Mainard D, Cuny C, Guillemin F. Satisfaction with care after

total hip or knee replacement predicts self-perceived health status after surgery. BMC Musculoskelet Disord. 2009;10(1):150.

157. Felix J, Becker C, Vogl M, Buschner P, Plötz W, Leidl R. Patient characteristics and valuation changes impact quality of life and satisfaction in total knee arthroplasty - Results from a German prospective cohort study. Health Qual Life Outcomes. 2019;17(1):1–10.

158. Clement ND, Macdonald D, Burnett R, Simpson AHRW, Howie CR. A patient's perception of their hospital stay influences the functional outcome and satisfaction of total knee arthroplasty. Arch Orthop Trauma Surg. 2017;137(5):693–700.

159. Losina E, Plerhoples T, Fossel AH, Mahomed NN, Barrett J, Creel AH, et al. Offering patients the opportunity to choose their hospital for total knee replacement: Impact on satisfaction with the surgery. Arthritis Care Res. 2005;53(5):646–52.

160. Huang S, Li X, Tang Y, Stiphan S, Yan B, He P, et al. Different patient satisfaction levels between the first and second knee in the early stage after simultaneous bilateral total knee arthroplasty (TKA): A comparison between subjective and objective outcome assessments. J Orthop Surg Res. 2017;12(1):1–6.

161. McLawhorn AS, Bjerke-Kroll BT, Blevins JL, Sculco PK, Lee Y yu, Jerabek SA. Patient-reported allergies are associated with poorer patient satisfaction and outcomes after lower extremity arthroplasty: a retrospective cohort study. J Arthroplasty. 2015;30(7):1132–6.

162. Noble PC, Fuller-lafreniere S, Meftah M, Dwyer MK. Challenges in outcome measurement discrepancies between patient and provider definitions of success. Clin Orthop Relat Res. 2013;471(11):3437–45.

163. Gautreau S, Gould ON, Aquino-Russell C, Forsythe ME. Developing a surgeon-patient communication checklist for total knee arthroplasty. Musculoskeletal Care. 2019;17(1):91–6.

164. Culliton SE, Bryant DM, MacDonald SJ, Hibbert KM, Chesworth BM. Effect of an e-learning tool on expectations and satisfaction following total knee arthroplasty: a randomized controlled trial. J Arthroplasty. 2018;33(7):2153–8.

165. Tolk JJ, Janssen RPA, Haanstra TM, van der Steen MC, Bierma-Zeinstra SMA, Reijman M. The influence of expectation modification in knee arthroplasty on satisfaction of patients: a randomized controlled trial. Bone Joint J. 2021;103(4):619–26.

166. Gautreau SJ, Gould ON, Allanach WW, Clark AE, Massoeurs SJ, Forsythe ME. Total knee arthroplasty communication checklist increases patient satisfaction. J Arthroplasty.

2019;34(3):456–61.

167. Edwards JR. Ten difference score myths. Organ Res methods. 2001;4(3):265–87.

168. Van Onsem S, Van Der Straeten C, Arnout N, Deprez P, Van Damme G, Victor J. A new prediction model for patient satisfaction after total knee arthroplasty. J Arthroplasty. 2016;31(12):2660–7.

169. Goh GS, Bin Abd Razak HR, Tay DKJ, Lo NN, Yeo SJ. Early post-operative oxford knee score and knee society score predict patient satisfaction 2 years after total knee arthroplasty. Arch Orthop Trauma Surg. 2021;141(1):129–37.

170. Zabawa L, Li K, Chmell S. Patient dissatisfaction following total knee arthroplasty: external validation of a new prediction model. Eur J Orthop Surg Traumatol. 2019;29(4):861–7.

171. Calkins TE, Culvern C, Nahhas CR, Della Valle CJ, Gerlinger TL, Levine BR, et al. External validity of a new prediction model for patient satisfaction after total knee arthroplasty. J Arthroplasty. 2019;34(8):1677–81.

172. Steyerberg EW. Clinical prediction models. Second. Cham: Springer International Publishing; 2019.

173. Williams DP, Brien SO, Doran E, Price AJ, Beard DJ, Murray DW, et al. Early postoperative predictors of satisfaction following total knee arthroplasty. Knee. 2013;20(6):442–6.

174. Bryan S, Goldsmith LJ, Davis JC, Hejazi S, MacDonald V, McAllister P, et al. Revisiting patient satisfaction following total knee arthroplasty: A longitudinal observational study. BMC Musculoskelet Disord. 2018;19(1):1–8.

175. Young-Shand KL, Dunbar MJ, Laende EK, Mills Flemming JE, Astephen Wilson JL. Early identification of patient satisfaction two years after total knee arthroplasty. J Arthroplasty. 2021;36(7):2473–9.

3 Objectives

The objectives of this dissertation corresponded to the four studies that have been completed:

Study 1: The objective of this study was to evaluate the ability of a standard logistic regression model and models created using machine learning methods to predict total knee arthroplasty (TKA) dissatisfaction 1-year post-surgery using pre-surgery and surgical variables.

Study 2: This study had two objectives. The first was to evaluate if 1-year post-surgery TKA dissatisfaction could be predicted using pre-surgery variables alone using logistic regression and classification trees. The second was to determine if 1-year TKA dissatisfaction could be predicted with pre-surgery and 3-month post-surgery variables using logistic regression and classification trees.

Study 3: The objective of this study was to create a prediction model that is simple to implement to predict 1-year post-surgery dissatisfaction using 3-month post-surgery individual patient reported outcome measure questionnaire items.

Study 4: The objective of this study was to investigate the met expectations variable and to determine if met expectations significantly moderate the effect of pain and function variables on satisfaction.

4 Logistic Regression and Machine Learning Models Cannot Discriminate Between Satisfied and Dissatisfied Total Knee Arthroplasty Patients

A version of this paper has been published: Munn JS, Lanting BA, MacDonald SJ, Somerville LE, Marsh JD, Bryant DM, Chesworth BM. Logistic Regression and Machine Learning Models Cannot Discriminate Between Satisfied and Dissatisfied Total Knee Arthroplasty Patients. The Journal of Arthroplasty.

4.1 Introduction

Approximately 20% of patients are dissatisfied or unsure of their satisfaction with their total knee arthroplasty (TKA)[1,2]. This is a very popular surgery with approximately 60,000 TKAs performed in 2019-2020 in Canada[3]. This means that in 2019-2020 alone there were approximately 12,000 TKA patients dissatisfied or unsure. This is a substantial problem. If dissatisfied and unsure patients can reliably be identified pre-surgery, specific interventions or the optimization of non-operative modalities could result in a reduction in dissatisfied or unsure patients.

There have been a number of previous attempts at predicting 1-year and 2-year TKA dissatisfaction using pre-surgery and surgical variables[4–8]. Several studies have used logistic regression[7] and ordinal logistic regression[9]. These attempts have either not been successful or failed to generalize to other samples[7,9].

Recently machine learning prediction models have been used to determine if they can outperform logistic regression, with Kunze[4] and Farooq[5] both finding that machine learning methods were superior to logistic regression. Although there has been a recent proliferation in the use of machine learning methods to predict outcomes in arthroplasty and orthopaedic medicine in general[10,11], these studies represent the first attempts at employing machine learning models to predict TKA dissatisfaction. Despite the reported success of these studies, they have not exhaustively explored all potential predictor variables of TKA dissatisfaction. Specifically, neither of these studies included patient reported outcome measures (PROMs).

When evaluating prediction models, it is also important to consider their intended use. If the models are meant to be used as standalone discriminatory tools, then it must be established what kind of performance is necessary for the model to be considered clinically useful. However, if models are intended to be used to provide supplemental information and to help inform surgeon and patient decision making, then it may be more important to evaluate a model's ability to generate an accurate probability that a patient will be dissatisfied than its discriminatory capacity.

The purpose of this study was to evaluate the ability of a standard logistic regression model and models created using machine learning methods to predict TKA dissatisfaction 1-year post-surgery using pre-surgery and surgical variables that can be considered as typically collected as part of a pre-surgery assessment. This study will also aim to evaluate if the variables included in this analysis contain the necessary information to predict dissatisfaction or if additional variables or analysis techniques may be required. Both types of models were evaluated by their ability to discriminate between satisfied and dissatisfied or unsure patients and their ability to generate accurate probabilities that patients will be dissatisfied or unsure. Because of the findings of Kunze[4] and Farooq[5], it was expected that the machine learning models would outperform standard logistic regression.

4.2 Methods

4.2.1 Participants

This was a prognostic observational cohort study that retrospectively analyzed prospectively collected data. An institutional database was used to identify a potential sample of patients who underwent primary TKA at a single large teaching institution (University Hospital London, Ontario, Canada) between August 2012 and December 2016. Demographic and patient reported outcome measures were retrieved using an institutional database. These variables were prospectively collected in a clinical setting as part of a standard pre-surgery and 1-year post-surgery assessment for TKA. A retrospective chart review was completed to collect additional surgical variables.

In order to be included in the analysis, patients were required to have been diagnosed with osteoarthritis, have no missing demographic information, and have completed both a pre-surgery and 1-year post-surgery Knee Society Knee Scoring System (KSS) questionnaire. This study received ethics approval from the institutional ethics review board.

4.2.2 Outcome Measures

The primary outcome measure was a dichotomized version of the KSS satisfaction subscale at 1-year post-surgery. This subscale includes five questions that evaluate a patient's satisfaction with their pain while sitting, pain while lying down, ability to get out of bed, ability to perform light household duties, and ability to perform recreational activities. Possible response options for each question (score) are very satisfied (8), satisfied (6), neutral (4), dissatisfied (2), and very dissatisfied (0). These five questions are summed together to create a score from 0-40 (worst to best). Satisfaction scores were dichotomized into satisfied and dissatisfied groups. Patients with a satisfaction score ≤24 were classified as dissatisfied or neutral and patients with a satisfaction score >24 were considered as satisfied. A cut point of 24 was chosen as this would mean that a patient must have responded to a minimum of one of the satisfaction questions as dissatisfied or three questions as neutral.

4.2.3 Predictor Variables

Candidate predictor variables included: age, gender, body mass index (BMI), American Society of Anesthesiologists (ASA) class , alcohol consumption, smoking status, living status, previous knee surgeries, pre-surgery knee alignment, patella re-surfaced, posterior collateral ligament sacrificed, as well as, subscales from the following questionnaires: the KSS; pre-operative and post-operative versions[12], the Knee Society Clinical Rating System (KSCRS)[13], the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[14], and the 12-Item Short Form Health Survey version 2 (SF-12)[15]. ASA class, alcohol consumption, smoking status, living status, and previous knee surgeries were dichotomized. ASA class was split into two categories, patients scored as class 1 and 2, and patients scored as class 3 and 4. Alcohol consumption was split into medium and light (<5 drinks per week) and heavy drinkers (>5 drinks per week). Living status was split into patients who lived alone and patients who lived with others. Previous surgery was split into patients who had previously undergone knee surgery and those who had not.

4.2.4 Missing Data

Any variables that had greater than 15% of the data missing were not included in the analysis. The KSCRS anteroposterior stability, mediolateral stability, flexion contracture, extension lag, and valgus alignment variables were not included in the analysis, as too much data were missing. Missing data were imputed using k nearest neighbours' imputation using the DMwR2 R package[16].

4.2.5 Statistical Analysis

All analyses were completed in R version 4.0.3[17].

The data was split 70:30 between a training (n=1022) and a test set (n=437). The split was stratified on the dissatisfaction variable to ensure that there was roughly the same ratio of dissatisfied patients in the training and the test sets using the caret R package[18]. Seven different statistical and machine learning models (R package used) were attempted including: standard logistic regression, elastic-net logistic regression (GLMnet)[19], random forest (randomForest)[20], gradient boosted trees (gbm)[21], extreme gradient (XG) boosted trees (xgboost)[22], support vector machines (e1071)[23] (SVM), and neural networks (keras)[24]. For all models except neural networks the caret R package was used to perform 10-fold cross-validation to select the optimal model. For neural networks, a single validation set was used[18].

No explicit sample size calculation was completed as it is generally recommended that as large a sample as possible be collected to create prediction models using machine learning methods.

The standard logistic regression model was selected using backwards elimination. A p-value of 0.20 was used as the cut off so that no potentially important variables were excluded. Variance inflation factor (VIF) was calculated for the model using the car R package, with a VIF>5 being considered problematic[25].

Elastic-net logistic regression adds regularization to a standard logistic regression model to avoid overfitting and improve model performance[26,27].

A random forest is an ensemble tree-based method[26,27]. This method fits a large number of classification trees (usually several thousand) using bootstrapped samples and a random subset of predictor variables, preventing classification trees from being too uniform. The outcome of each of the individual trees is aggregated and a final prediction is generated[26,27].

Gradient boosting and XG boosting are both gradient boosting methods; however, they use different algorithms to select and optimize models[22]. Gradient and XG boosted trees, like random forests, are ensemble tree-based methods. However, these methods fit a series of classification trees sequentially, with each tree created attempting to fix the prediction errors made by the previous tree. This process is typically repeated several hundred or several thousand times.

SVMs are models that attempt to draw a hyperplane that maximizes the distance between the two classes[26,27]. This model type can be linear or non-linear. To create a non-linear model, a kernel can be used, to allow for greater flexibility. This study used a non-linear version of SVMs. Platt scaling was used to calibrate the SVM model[26,27].

Neural network models are meant to emulate the neuronal circuitry of the human brain[27]. Neural networks consist of a series of inputs, hidden layers of nodes, and an output. Data must pass through all the layers and a prediction is generated at the end. These models are extremely flexible and can map complex patterns[27].

4.2.6 Model Evaluation

After the optimal model for each type of machine learning algorithm was generated using the training set, its performance was evaluated using the test set. Metrics that evaluated model discrimination and calibration were used.

4.2.6.1 Discrimination

Discrimination is the model's ability to accurately predict a patient's satisfaction status (i.e. correctly predicting that a dissatisfied patient would be dissatisfied). Model discrimination was evaluated using area under the curve (AUC). Additionally cut points were established that maximized model accuracy and overall accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and kappa scores were calculated[27].

4.2.6.2 Calibration

Calibration is the model's ability to accurately assign probabilities to a patient's outcome (i.e. patients who are assigned a probability of 0.6 are dissatisfied 60% of the time). The models' calibration was evaluated using the Brier score, Cox intercept, and Cox slope. Brier score is a measure of how closely probabilities predicted by the model match the actual outcome[27], with a lower score indicating better performance. Cox intercept and slope are measures of how well the calibration probability matches with an actual outcome. The optimal Cox intercept is 0 and the optimal Cox slope is 1, with values close to 0 and 1 indicating good overall calibration. Calibration plots were generated for the models that performed best[27]. Calibration metrics and plots were calculated using the rms R package[28].

4.2.7 Model Interpretation

Not all machine learning models are equally interpretable or provide information about variable importance. If the algorithm allowed for straightforward interpretation of the importance or effect of individual variables on model performance, that information was included for the best performing models. For logistic regression an output table including odds ratios was generated. For tree-based models, relative influence plots were created, and variables were grouped into high, medium, and low influence categories.

4.3 Results

Of an initial 3000 patients who were identified as candidates for this study, 1,432 met the inclusion criteria. The study population had a mean age of 66.75 and 969 (64.5%) were female. A total of 313 (21.8%) patients were classified as dissatisfied or unsure at 1-year post-surgery. Descriptive statistics including t-tests/chi-squares are included in table 4.1.

Table 4.1. Descriptive statistics including mean, standard deviation and t-test/chi-square p-values. Values in bold indicate a p-value of less than 0.05.

	Dissatisfied or Neutral n=313		Satisfied n=1119		
Variable	Mean/Frequency	SD	Mean/Frequency	SD	T-Test/Chi Square P-Values
Age	66.62	10.88	66.79	9.42	0.809
BMI	33.77	8.24	33.35	7.24	0.106
Gender (Female)	67.7%		63.5%		0.12
Living Status (Live Alone)	25.6%		21.3%		**0.023**
Alcohol Consumption (>5 Drinks/Week)	15.7%		21.1%		0.206
Smoker	47.6%		51.7%		0.884
ASA Class (Three and Four)	60.7%		54.8%		0.171
PCL Sacrificed	85.3%		87.3%		0.353
Patella Resurfaced	22.0%		27.3%		0.059
Valgus Alignment (>10 degrees valgus)	28.4%		24.2%		0.128
Varus Alignment (<2 degrees valgus)	61.0%		72.1%		**<0.001**
Neutral Alignment (2-10 degrees valgus)	10.5%		3.7%		**<0.001**
KSCRS Preop Pain	19.95	6.67	21.53	6.58	**<0.001**
KSCRS Preop Range of Motion Extension	5.05	9.36	4.42	5.77	**0.033**
KSCRS Preop Range of Motion Flexion	102.33	13.85	103.85	13.74	0.205
KSCRS Preop Total Function	45.59	12.36	48.45	12.36	0.883
SF-12 Preop MCS	49.77	10.98	54.01	10.18	0.062
SF-12 Preop PCS	30.27	7.55	32.19	8.33	**<0.001**
WOMAC Preop Pain	42.82	16.04	49.81	17.44	**<0.001**
WOMAC Preop Stiffness	38.07	19.80	42.10	19.94	**0.002**
WOMAC Preop Function	42.69	15.44	49.93	16.72	**<0.001**
WOMAC Preop Total Score	41.74	14.70	48.19	15.58	**<0.001**
KSS Symptoms	7.02	4.81	8.71	5.35	**<0.001**
KSS Preop Satisfaction	12.23	6.60	14.73	7.28	**<0.001**
KSS Preop Expectations	13.20	2.22	13.73	1.67	**<0.001**
KSS Walking Standing	10.50	7.66	12.06	7.58	**0.001**
KSS Standard Activity	11.00	4.84	12.63	4.98	**<0.001**
KSS Advanced Activity	4.40	4.03	5.37	4.30	**<0.001**
KSS Discretionary Activity	4.82	3.14	5.56	3.16	**<0.001**

SD, standard deviation; BMI, body mass index; ASA, American Society of Anesthesiologists; PCL, posterior collateral ligament; KSCRS, Knee Society Clinical Rating System; SF-12, 12-Item Short Form Health Survey version 2; MCS, mental component score; PCS, physical component score; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; KSS, Knee Society Knee Scoring System.

4.3.1 Model Performance

4.3.1.1 Discrimination

Measures of discrimination are included in table 4.2. The models that performed best when considering discrimination metrics were the standard logistic regression and XG boosted trees models. The logistic regression model outperformed the elastic-net logistic regression model and was chosen as the final model. When considering the AUC, accuracy, sensitivity, specificity, and kappa score, the logistic regression model performed best.

4.3.1.2 Calibration

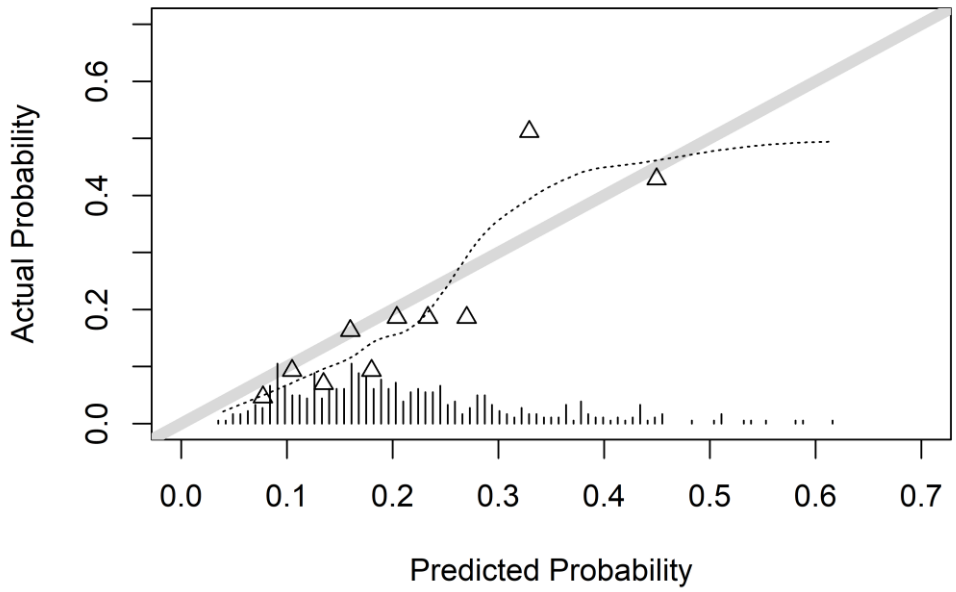
Measures of model calibration are included in table 4.2. When considering Brier score, Cox intercept, and Cox slope, the best performing models were the logistic regression and gradient boosted trees models (figure 4.1A and 4.1C). The XG boosted trees model was very poorly calibrated (figure 4.1B). This is illustrated by the proximity of the dotted line and the triangles to the thick grey line in figure 4.1.

Table 4.2. Discrimination and calibration measures of the test set including AUC, accuracy, sensitivity, specificity, positive predictive value, negative predictive value, kappa score, Brier score, Cox intercept, and Cox slope. Cut points used to generate accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and kappa score were chosen to maximize accuracy.

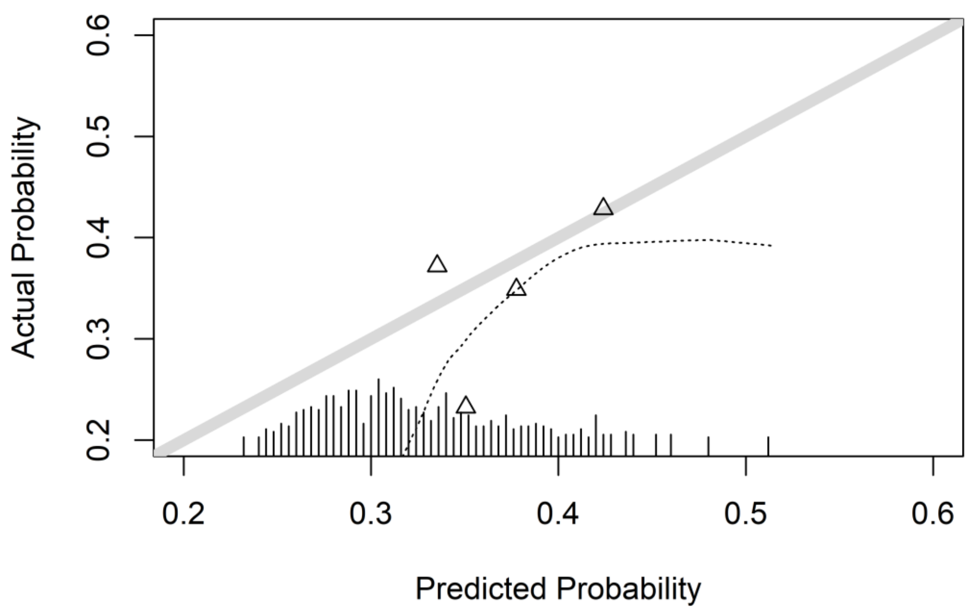
Measure	Standard Logistic Regression	Random Forest	Gradient Boosted Trees	XG Boosted Trees	Support Vector Machines	Neural Networks
Discrimination						
AUC	0.736	0.709	0.689	0.713	0.6	0.702
Accuracy	0.795	0.786	0.79	0.781	0.664	0.804
Sensitivity	0.155	0.202	0.083	0.214	0.44	0.131
Specificity	0.951	0.928	0.962	0.919	0.719	0.968
PPV	0.433	0.405	0.35	0.391	0.276	0.5
NPV	0.822	0.827	0.812	0.828	0.841	0.821
Kappa Score	0.139	0.16	0.064	0.161	0.13	0.137
Calibration						
Brier Score	0.141	0.147	0.149	0.165	0.243	0.153
Cox Intercept	0.241	0.473	0.054	0.798	-1.246	0.112
Cox Slope	1.310	1.092	1.158	3.14	0.098	1.663

XG, extreme gradient; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.

A **Logistic Regression**



B **XG Boosted Trees**



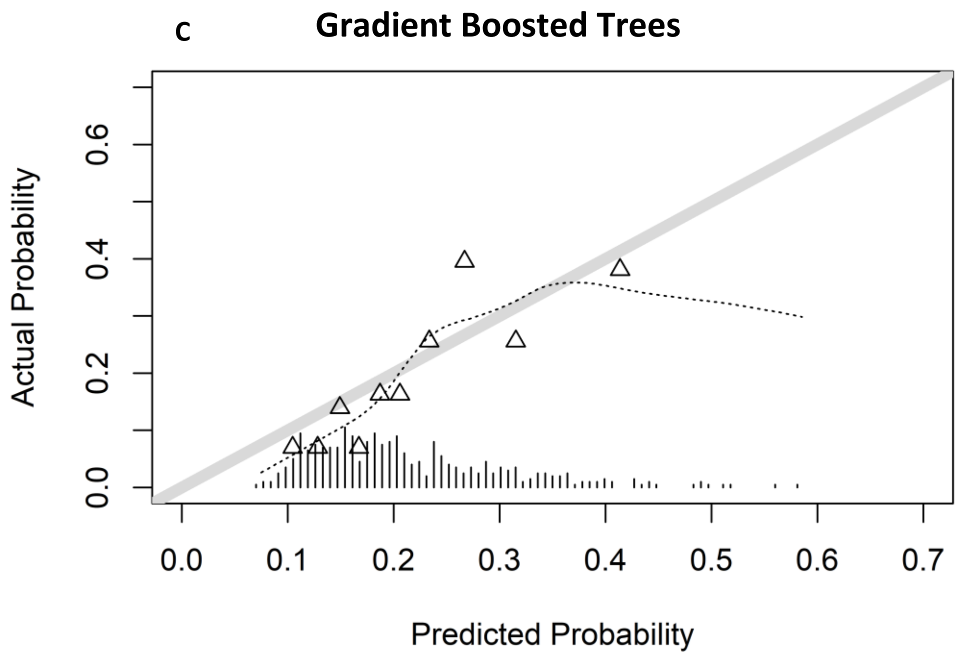


Figure 4.1. Calibration plots for the logistic regression (A), extreme gradient boosted trees (B) and gradient boosted trees models (C). The triangles indicate a group of 40 patients. The average probability for the group is taken and then that is compared to percentage of patients who are actually dissatisfied. The dotted line is a smoothed calibration line. The thick grey line indicates perfect calibration.

4.3.2 Model Interpretation

The output of the logistic regression model is included in table 4.3, outlining important predictors. The relative influence of the individual variables for the XG boosted and gradient boosted models has been included in figure 4.2, with variables being grouped into high, medium, and low influence categories.

Table 4.3. Logistic regression model output table. Age, BMI, and gender were forced into the model regardless of significance.

	Dissatisfaction				
Predictors	Estimates	Standard Error	P-Value	Odds Ratios	Confidence Interval
(Intercept)	3.462	1.153			
Age	0.017	0.009	0.049	1.02	1.00 – 1.03
BMI	0.002	0.011	0.818	1.00	0.98 – 1.02
Gender (Female)	-0.021	0.174	0.901	0.98	0.69 – 1.37
Valgus Alignment	-0.609	0.332	0.066	0.54	0.29 – 1.05
Varus Alignment	-0.856	0.311	0.006	0.42	0.23 – 0.79
Patella Resurfaced	-0.37	0.185	0.045	0.69	0.48 – 0.99
KSCRS Pain	-0.018	0.013	0.147	0.98	0.96 – 1.01
SF-12 MCS	-0.028	0.008	0.001	0.97	0.96 – 0.99
SF-12 PCS	-0.018	0.012	0.149	0.98	0.96 – 1.01
WOMAC Function	-0.014	0.006	0.032	0.99	0.97 – 1.00
KSS Pre-Surgery Satisfaction	-0.016	0.015	0.294	0.98	0.96 – 1.01
KSS Pre-Surgery Expectations	-0.155	0.04	<0.001	0.86	0.79 – 0.93
KSS Walking Standing	0.022	0.012	0.072	1.02	1.00 – 1.05
Observations			1003		

BMI, body mass index; KSCRS, Knee Society Clinical Rating System; SF-12, 12-Item Short Form Health Survey version 2; MCS, mental component score; PCS, physical component score; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; KSS, Knee Society Knee Scoring System.

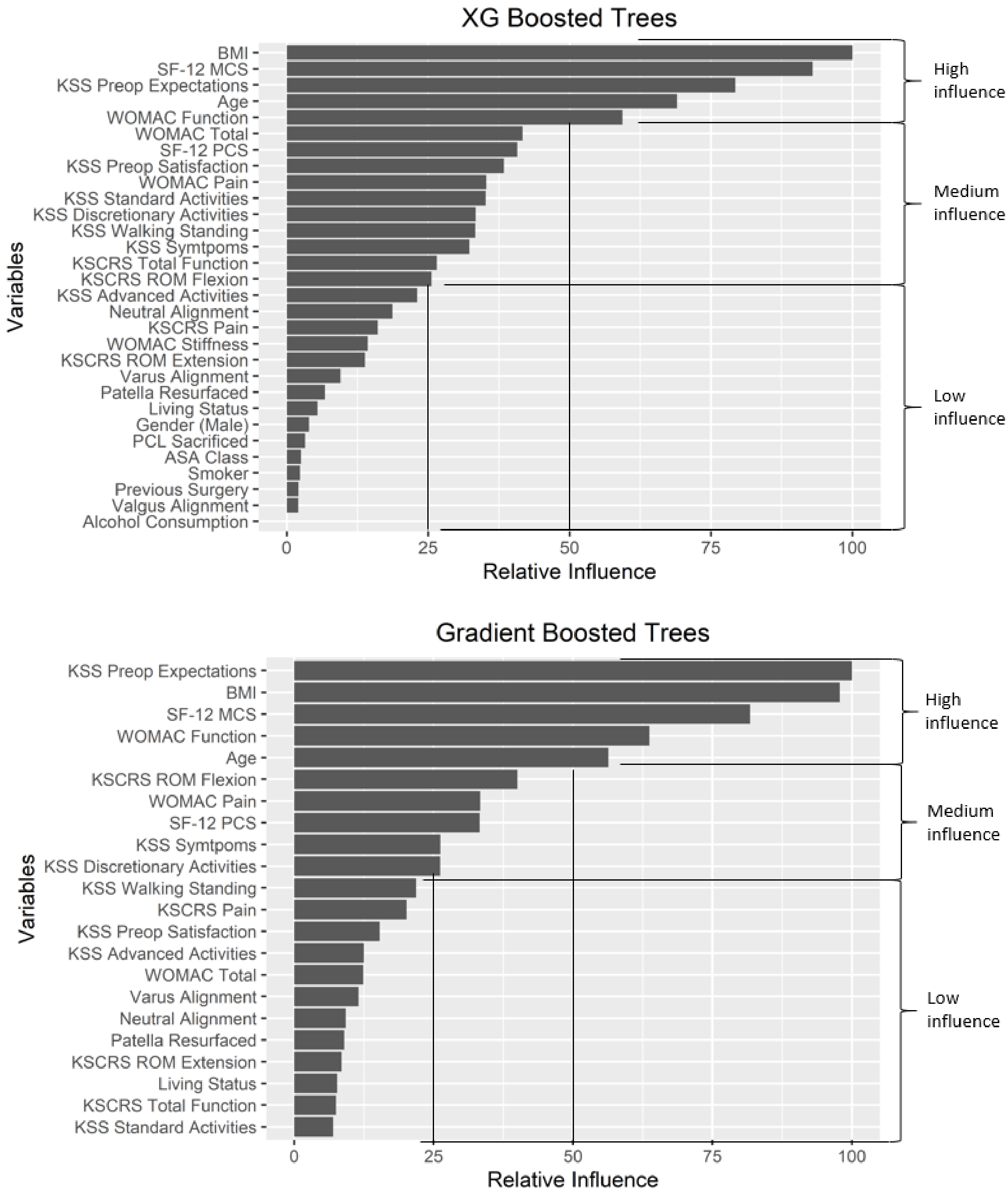


Figure 4.2. This is a graphical representation of relative importance of variables for the extreme gradient boosted trees and gradient boosted trees scaled from 0-100, with the most important variable being scored as 100 and the least important scored as 0.

BMI, body mass index; SF-12, 12-Item Short Form Health Survey version 2; MCS, mental component score; KSS, Knee Society Knee Scoring System; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; PCS, physical component score; KSCRS, Knee Society Clinical Rating System; PCL, posterior collateral ligament; ASA, American Society of Anesthesiologists.

4.4 Discussion

When considering AUC, the best and second-best performing models respectively were standard logistic regression and XG boosted trees. This finding was unexpected and suggests that when predicting dissatisfaction using variables that can be considered as typically collected as part of a pre-surgery assessment, more complex machine learning models may not offer a predictive advantage over standard logistic regression. This was contrary to the findings of Kunze[4] and Farooq[5] who found that machine learning models outperformed standard logistic regression when predicting dissatisfaction. However, this finding was consistent with a systematic review by Christodolou[29], that found in a medical context machine learning methods did not outperform logistic regression. A possible reason for the disparity, is that this study made use of different variables than Kunze[4] and Farooq[5]. Notably, we included the SF-12 and WOMAC as predictor variables. These questionnaires are validated and commonly used with arthroplasty patients. Additionally, a different satisfaction measure was used as the dependent variable. Although the variables used in this analysis have historically been considered as important, it appears that they do not contain the necessary information to accurately predict 1-year post-surgery TKA dissatisfaction and are not capable of fully realizing the benefits of machine learning methods. A wider variety of explanatory variables and analytical methods must be investigated.

Despite acceptable AUC values, when establishing discriminatory cut points and examining accuracy, sensitivity, specificity, and kappa scores none of the models performed particularly well. How a discriminatory cut point is chosen depends on the priority of the decision maker. However, in this study cut points were simply chosen to maximize accuracy. Generally, models performed at or below the no information rate. This means, that when maximizing accuracy, the models' performance was approximately the same or worse as predicting that all patients would be satisfied. Additionally, the models have low sensitivity scores, indicating that they struggle to accurately identify dissatisfied patients. It is possible to improve sensitivity of a model by adjusting the cut point that is chosen. However, this would come at the expense of specificity and overall accuracy.

When evaluating these models, it is also important to consider how the model is going to be used and how well the model needs to perform in order to be clinically useful as a discriminatory tool. This would require a discussion of acceptable sensitivity and specificity values. It would also require that a judgement be made about which is more costly, not performing surgery on a patient who would be satisfied or performing surgery on a patient who will be dissatisfied. The machine learning models presented in this paper likely do not discriminate well enough to be useful as clinical decision tools.

Despite the poor discriminatory capacity of the models, the calibration scores appear to be more promising. This indicates that the models may be better at generating an accurate probability that a patient will be dissatisfied than discriminating between satisfied and dissatisfied patients. Although not optimal, the best calibrated models were the logistic regression and gradient boosted trees models. It would be possible to use the probabilities generated by these models to help inform surgeon and patient decision making, without the models making an explicit prediction about whether a patient is dissatisfied or not. Harrell[30] has argued there can be great value in using medical prediction models in this probabilistic fashion. This approach has the additional benefit of allowing the surgeon's judgement and the patient's viewpoint to be used in treatment decisions[30,31]. In this way information that has not been incorporated into a prediction model can be included in the decision-making process.

When evaluating significant and influential predictors of satisfaction, a consistent theme appeared in the logistic regression, gradient boosted trees, and XG boosted trees models. Specifically, age, pre-surgery WOMAC function, pre-surgery SF-12 MCS, and pre-surgery patient expectations were all identified as significant/influential predictors of satisfaction. It is important to note that unlike logistic regression, the relative influence of a variable does not provide any information about the direction or the context in which the variable is influencing the prediction. For example, in the logistic regression model the odds ratio for age indicates that the older a patient is the more likely they are to be dissatisfied. This relationship cannot be assumed when examining the relative influence plot. The information from logistic regression and tree-based models in some sense can be considered complimentary as they are providing different types of information about predictors of dissatisfaction and if used together could potentially help further our understanding of how these variables influence satisfaction. In the gradient boosted trees and XG boosted trees models, BMI was a highly influential variable, but was not a significant predictor in the logistic regression model. This finding does not necessarily indicate that higher BMI is associated with dissatisfaction. It is possible that BMI has a complex relationship with dissatisfaction and is conditionally important. This could help explain the inconsistent relationship between BMI and satisfaction in the literature[32].

The variables included in this study do not appear to have the necessary information to accurately predict 1-year post-surgery dissatisfaction, and it may be possible that there is no reasonable way to predict dissatisfaction using pre-surgery and surgical variables. However, before making this determination there are other analytical methods and predictors to explore. Part of the value of machine learning methods is their ability to map complex patterns and include a wider variety of data than standard statistical models. Natural language processing (NLP) is a form of machine learning that can be used to process language and predict outcomes. Training NLP models to make use of the information in clinical notes to predict dissatisfaction is a possible avenue for future research. Using machine learning to analyze radiographic images to predict dissatisfaction is also an area to investigate. Lastly, the collection of different PROMs that provide more granular information about the patient may allow for the models to be improved. Specifically, determining what a patient would define as a successful outcome for their surgery. This could include determining, not just pre-surgery expectations, but specific activities that are particularly important or that the patient is motivated to perform. Additionally, psychological factors (i.e. personality, self-efficacy, and locus of control), socio economic status, financial well-being, and social support remain mostly unexplored in the TKA literature. Including these or other variables that may interact with standard PROMs and can take advantage of the flexibility of machine learning methods could potentially improve the performance of prediction models.

This study had several limitations. First the study is a retrospective analysis, although the PROMs were collected prospectively. It could be a potential limitation that this study only included a 1-year follow up, although a systematic review and meta-analysis completed by Ramkumar[33] has found no difference in outcomes 1 and 2-years post-surgery for lower body total joint arthroplasty, so a 1-year follow up time should be acceptable[33]. This study was missing a comorbidity measure, and additionally, several of the KSCRS sub scores had to be removed because of missing data. The choice to dichotomize ASA class, alcohol consumption, smoking status, living status, and previous knee surgeries variables may have adversely affected model performance. Lastly, despite the relatively large sample size of this study, it is likely still too small to generate an optimal prediction model.

In conclusion, all the models in this study do not perform well as discriminatory tools and are early attempts at using machine learning models to predict dissatisfaction. A wider array of variables and analytical techniques need to be investigated to determine if machine learning models can more accurately predict TKA dissatisfaction. In order to make this determination a decision must also be made with regards to how these tools will be used. If the models will not be used in a stand-alone discriminatory capacity, there may be greater value in creating well-calibrated prediction models that generate accurate probabilities than models that are meant to be purely discriminatory.

4.5 References

1. Kahlenberg CA, Nwachukwu BU, Mclawhorn AS, Cross MB, Cornell CN, Padgett DE. Patient satisfaction after total knee replacement: a systematic review. HSS J. 2018;14(2):192–201.
2. Gunaratne R, Pratt DN, Banda J, Fick DP, Khan RJK, Robertson BW. Patient dissatisfaction following total knee arthroplasty: a systematic review of the literaturere. J Arthroplasty. 2017;32(12):3854–60.
3. Hip and Knee Replacements in Canada: CJRR Annual Statistics Summary, 2018–2019. Ottawa, ON; 2020.
4. Kunze KN, Polce EM, Sadauskas AJ, Levine BR. Development of machine learning algorithms to predict patient dissatisfaction after primary total knee arthroplasty. J Arthroplasty. 2020;35(11):3117–22.
5. Farooq H, Deckard ER, Ziemba-Davis M, Madsen A, Meneghini RM. Predictors of patient satisfaction following primary total knee arthroplasty: results from a traditional statistical model and a machine learning algorithm. J Arthroplasty. 2020;35(11):3123–30.
6. Baker P, Muthumayandi K, Gerrand C, Kleim B, Bettinson K, Deehan D. Influence of Body Mass Index (BMI) on Functional Improvements at 3 Years Following Total Knee Replacement : A Retrospective Cohort Study. PLoS One. 2013;8(3):1–9.
7. Garriga C, Sanchez-Santos MT, Judge A, Perneger T, Hannouche D, Lübbeke A, et al. Development of a model predicting non-satisfaction 1 year after primary total knee replacement in the UK and transportation to Switzerland. Sci Rep. 2018;8(1):1–8.
8. Kunze KN, Akram F, Fuller BC, Zabawa L, Sporer SM, Levine BR. Internal validation of a predictive model for satisfaction after primary total knee arthroplasty. J Arthroplasty. 2019;34(4):663–70.
9. Baker PN, Rushton S, Jameson SS, Reed M, Gregg P, Deehan DJ. Patient satisfaction with total knee replacement cannot be predicted from pre-operative variables alone: a cohort study from the National Joint Registry for England and Wales. Bone Joint J. 2013;95(10):1359–65.
10. Ramkumar PN, Haeberle HS, Bloomfield MR, Schaffer JL, Kamath AF, Patterson BM, et al. Artificial intelligence and arthroplasty at a single institution: real-world applications of machine learning to big data, value-based care, mobile health, and remote patient monitoring. J Arthroplasty. 2019;34(10):2204–9.
11. Haeberle HS, Helm JM, Navarro SM, Karnuta JM, Schaffer JL, Callaghan JJ, et al. Artificial intelligence and machine learning in lower extremity arthroplasty: a review. J Arthroplasty.

2019;34(10):2201–3.
12. Noble PC, Scuderi GR, Brekke AC, Sikorskii A, Benjamin JB, Lonner JH, et al. Development of a new Knee Society scoring system. Clin Orthop Relat Res. 2012;470(1):20–32.
13. Insall JN, Dorr D, Scott R, Scott W. Rationale of The Knee Society Clinical Rating System. Clin Orthop Relat Res. 1989;26(4):13–4.
14. Bellamy N, Buchanan W, Goldsmith C, Campbell J, Stitt L. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol. 1988;15(12):1833–40.
15. Ware JE, Krosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996;34(3):220–33.
16. Torgo L. Data mining with R: learning with case studies. 1st ed. Boca Raton, FL: Taylor & Francis; 2010.
17. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for statistical Computing; 2020.
18. Kuhn M. caret: Classification and Regression Training. R package version 6.0-88; 2021.
19. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1–22.
20. Liaw A, Wiener M. Classification and regression by randomForest. R News. 2002;2(3):18–22.
21. Greenwell B, Boehmke B, Cunningham J, Developers G. gbm: Generalized Boosted Regression Models. 2020.
22. He T, Chen K, Mitchell R, Cano I, Lin M. Package ‘ xgboost ’ R topics documented. 2021.
23. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. Package ‘e1071.’ 2021.
24. Allaire J, Chollet F. keras: R Interface to “keras.” 2021.
25. Fox J, Weisberg S. An R Companion to Applied Regression. Third. Thousand Oaks CA: Sage; 2019.
26. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. New York: Springer Science+Business Media; 2013.
27. Kuhn M, Johnson K. Applied predictive modeling. New York: Springer Science+Business Media; 2013.
28. Harrell F. rms: Regression Modeling Strategies. R package version 6.2-0; 2021.
29. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review

shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019;110:12–22.

30. Harrell F. Regression modeling strategies. Second. New York: Springer Science+Business Media; 2015.

31. Spiegelhalter DJ. Probabilistic prediction in patient management and clinical trials. Stat Med. 1986;5(5):421–33.

32. Nakano N, Shoman H, Olavarria F, Matsumoto T, Kuroda R, Khanduja V. Why are patients dissatisfied following a total knee replacement? A systematic review. Int Orthop. 2020;44(10):1971–2007.

33. Ramkumar PN, Navarro SM, Haeberle HS, Ng M, Piuzzi NS, Spindler KP. No difference in outcomes 12 and 24 months after lower extremity total joint arthroplasty: a systematic review and meta-analysis. J Arthroplasty. 2018;33(7):2322–9.

5 Can 1-Year Dissatisfaction with TKA be Predicted Pre-Surgery or Early in the Recovery Process?

5.1 Introduction

Total knee arthroplasty (TKA) is a common and effective surgery for relieving pain and improving function for patients with end-stage osteoarthritis (OA). However, at 1-year post-surgery approximately 20% of patients are dissatisfied or unsure of their satisfaction with the results of their surgery[1,2]. The most consistently identified factors associated with dissatisfaction are pain in the joint[1–8], a lack of functional ability[4,6,8–11], and unmet patient expectations following surgery[4,7,8,12–14]. Two different strategies that aim to improve patient satisfaction have appeared in the literature. The first is identifying patients likely to be dissatisfied pre-surgery[9,15]. The second is the implementation of interventions that improve patient satisfaction[16–18].

Attempts at predicting dissatisfaction using solely pre-operative variables have traditionally been done by dichotomizing the satisfaction variable, at which point logistic regression and machine learning models have been created to predict if a patient is satisfied or not[9,15,19–22]. These attempts have generally not been successful, with Baker[9] concluding that patient satisfaction with TKA cannot be predicted using pre-operative variables alone. When these models do show some promise they have not generalized well to the population outside of the study sample[15,19].

Efforts to improve satisfaction via intervention during the post-surgery recovery process have focused on patient education and generating realistic patient expectations to improve met expectations scores and by association patient satisfaction[16–18,23]. Post-surgery interventions may be an effective way of improving TKA satisfaction. Identifying patients likely to be dissatisfied with their surgery early in the recovery process could allow for the implementation of interventions to improve satisfaction before 1-year post-surgery where outcomes are more likely to be fixed. If early warning signs of dissatisfaction can be identified or prediction models can be created to identify these patients, it may be possible to reduce the number of individuals who are dissatisfied with their TKA.

The purpose of this study was to determine if prediction models can identify patients who are dissatisfied at 1-year post-surgery with pre-surgery variables alone or a combination of pre-surgery and 3-month post-surgery variables using logistic regression and classification trees.

5.2 Methods

5.2.1 Patients

This was a prognostic observational cohort study that retrospectively analyzed prospectively collected data. Patients included in this analysis were originally part of a randomized control trial (RCT) that attempted to improve TKA satisfaction using an e-learning tool that aimed to adjust patient expectations. The e-learning tool was observed to have no effect on patients[16]. As no differences between the treatment and control groups were observed, it was considered acceptable to pool patients into a single sample.

For the initial RCT, patients scheduled to undergo TKA were recruited from a pre-admission clinic at a large teaching institution (University Hospital London, Ontario, Canada) between April 2013 and April 2014. In order to be included in the original RCT, patients were required to be 20 years of age or older, have been diagnosed with knee osteoarthritis, and have the cognitive capacity to consent to participate. Patients undergoing patellar resurfacing, hemi or unicondylar knee arthroplasty, high tibial osteotomy, or knee surgery to address a tumor were excluded. Patients were asked to complete questionnaires at pre-surgery and 6-weeks, 3-months, and 1-year post-surgery. Questionnaires were completed by patients online and stored in a secure database.

In order to be included in the present study, patients were required to have completed questionnaires at pre-surgery, 3-months post-surgery, and 1-year post-surgery.

This study received ethics approval from the university ethics review board.

5.2.2 Outcome Measures

Questionnaires included the Knee Injury and Osteoarthritis Outcome Score (KOOS)[24], the Knee Society Knee Scoring System (KSS; pre-operative and post-operative versions)[25], the 12-Item Short Form Health Survey version 2 (SF-12)[26], the Pain Catastrophizing Scale (PCS)[27], the University of California at Los

Angeles (UCLA) Activity Score[28], the Hospital Anxiety and Depression Scale (HADS)[29], and the Self-Administered Comorbidity Questionnaire (SCQ).

Satisfaction was measured using the KSS satisfaction subscale which yields a patient satisfaction score ranging from 0-40 (least to most satisfied). This scale was dichotomized into a satisfied and dissatisfied group with patients scoring >24 considered to be satisfied and those scoring ≤24 considered to be dissatisfied. A score of 24 or less indicates that the patients must have responded to at least one of the satisfaction questions as dissatisfied or at least three questions as neutral.

5.2.3 Statistical Analysis

All statistical analyses were completed in R version 4.0.3[30].

Descriptive statistics were calculated for all variables in the data set. Changes in satisfaction rates and histograms illustrating the distribution of the continuous satisfaction variable at pre-surgery, 3-months post-surgery, and 1-year post-surgery were created using the ggplot2 package[31].

Patients who did not complete questionnaires at 3-months post-surgery but had completed questionnaires at pre-surgery or pre-surgery and 1-year post-surgery were compared to determine if there were differences between patients who had completed all the questionnaires and those who had not. T-tests were used for continuous variables and chi square tests were used for categorical variables.

Two sets of logistic regression and classification and regression tree (CART) models were created. The first set of models included only pre-surgery variables as candidate independent variables; the second set of models included both pre-surgery and 3-month post-surgery variables as candidate independent variables. The dichotomized KSS satisfaction subscale at 1-year was used as the dependent variable. Correlation with satisfaction and between potential independent variables was used to limit the number of variables considered for inclusion in the models. Highly correlated variables were removed from consideration to avoid multicollinearity. Candidate independent variables included age, body mass index (BMI), and gender as well as the pre-surgery and 3-month post-surgery KOOS symptoms, KOOS activities of daily living (ADL), KOOS quality of life, KSS symptoms, KSS patient satisfaction, KSS patient expectations, SF-12 mental health, SF-12 bodily pain, and SF-12 vitality questionnaire subscales. Values assumed to be missing at random were imputed using k nearest neighbours' imputation using the DMwR2 R package[32].

Two logistic regression models were initially created using forward stepwise variable selection. No alpha cut off value was set for inclusion in the model. Forty-nine patients in the sample were dissatisfied. Therefore, a maximum of four variables were included in each of the models. This was done to ensure there were at least 10 dissatisfied observations per variable[33]. Multicollinearity was assessed using variance inflation factor (VIF) using the car R package[34], with a VIF>5 being considered problematic. For each logistic regression model elastic-net penalization was used to shrink model estimates to attempt to improve model generalizability using the GLMnet R package[35]. Bootstrapping was used to evaluate the consistency of the predictors selected for inclusion in the models. A total of 500 bootstrapped samples were generated and models were selected using an automated forward stepwise selection algorithm that was limited to four variables using the bootStepAIC R package[36]. Alpha and Lambda values for the elastic-net penalization were selected using the caret R package to maximize area under the curve (AUC) using 10-fold cross-validation repeated 10 times[37]. If elastic-net penalization did not improve model performance the standard logistic regression model was used.

To create the classification tree models, the rpart R package was used to employ a recursive partitioning algorithm[38]. An initial tree was selected with the stopping criteria defined as a minimum of 20 observations in each node for a split to be attempted. The initial tree was then pruned to prevent overfitting using a cost-complexity parameter. The cost-complexity parameter was selected with the caret R package using 10-fold cross-validation repeated 10 times. The tree with the highest cross-validated accuracy was chosen[37].

Models were initially trained using all of the observations available in the data set. For the logistic regression models AUC and cross-validated AUC were calculated using the pROC and caret R packages[39]. For both logistic regression and classification tree models an overall and cross-validated prediction accuracy and kappa score was calculated using 10-fold cross-validation repeated 10 times with the caret R package[37]. Model performance was assessed using overall accuracy, sensitivity, specificity, positive likelihood ratio (LR+), negative likelihood ratio (LR-), positive predictive value (PPV), negative predictive value (NPV), cross-validated accuracy, and cross-validated kappa scores. AUC and cross-validated AUC were also used to evaluate the discriminatory capacity of the logistic regression models. Brier score, Cox intercept, and Cox slope were calculated with the rms R package and used to evaluate calibration of the logistic regression models[40]. Contingency tables were also created for each model to identify the number of patients who were correctly predicted as dissatisfied, correctly predicted as satisfied, incorrectly predicted as dissatisfied and incorrectly predicted as satisfied.

No sample size calculation was completed for this analysis as information from a previous or similar prediction model would be required to make the calculation and to the best of the author's knowledge this information was not available.

5.3 Results

A total of 416 patients were recruited to participate in the initial RCT. Of this original sample, 50 did not undergo TKA or were missing measures from the pre-admission clinic. Of the 366 remaining eligible patients, 275 met the inclusion criteria for this study.

Patient outcome scores and demographic information are included in table 5.1. Change in satisfaction rates are included in table 5.2. Histograms including distribution of the continuous satisfaction variable are included in figure 5.1.

Table 5.1. Descriptive statistics for all participants including mean and standard deviation.

	Pre-Surgery n=275		3-Months n=275		1-Year n=275	
Variable	Mean	SD	Mean	SD	Mean	SD
Age	64.65	8.45				
Comorbidities	2.77	1.27				
Sex (Female)*	176 (64%)					
BMI	32.97	7.45				
KSS Patient Satisfaction Dichotomized at 1-Year (Satisfied)*					226 (82%)	
KSS Patient Satisfaction Score	13.27	6.73	28.73	7.92	31.95	8.54
KSS Symptoms	10.07	3.70	17.67	5.04	19.54	4.97
KSS Patient Expectations	13.41	1.90	8.58	2.39	9.16	2.78
KSS Walking Standing	25.11	4.48	23.95	4.74	25.50	4.41
KSS Standard Activity	11.74	4.85	21.53	4.81	23.24	5.23
KSS Advanced Activity	4.57	3.63	9.08	5.59	11.51	5.80
KSS Discretionary Activity	5.43	2.94	9.83	3.43	11.37	3.01
KOOS Pain	41.58	16.79	72.59	16.77	82.06	17.34
KOOS Symptoms	42.00	17.20	66.63	16.00	75.98	15.38
KOOS Activities of Daily Living	46.38	17.64	76.47	15.39	83.43	16.46
KOOS Functional Sport and Recreation	17.68	24.08	42.05	27.39	53.76	28.36
KOOS Quality of Life	20.34	15.61	55.63	20.20	65.75	21.82
PCS Rumination	9.31	3.85	5.99	2.84	5.68	3.00
PCS Magnification	5.34	2.49	3.77	1.56	3.82	1.64
PCS Helplessness	11.28	4.86	7.90	3.01	7.94	3.45
SF-12 Physical Function	28.38	8.13	38.88	10.40	41.51	10.70
SF-12 Role Physical	32.96	9.23	42.15	9.34	45.52	9.69
SF-12 Bodily Pain	29.35	9.19	43.97	10.02	47.47	10.41
SF-12 General Health	49.58	9.04	50.93	8.60	51.02	9.01
SF-12 Vitality	47.13	9.87	50.82	9.13	52.79	9.19
SF-12 Social Function	44.30	11.48	50.53	8.49	51.55	8.19
SF-12 Role Emotional	72.35	27.44	83.85	21.26	88.01	18.65
SF-12 Mental Health	84.89	20.82	92.16	17.73	93.57	17.44
UCLA Activity Score	4.33	1.77	5.12	1.65	5.69	1.75

*Values are represented as a frequency; SD, standard deviation; KSS, Knee Society Knee Scoring System; KOOS, Knee Injury and Osteoarthritis Outcome; PCS, Pain Catastrophizing Scale; Score; SF-12, Short Form Health Survey 12; UCLA, University of California Los Angeles

Table 5.2. Descriptive statistics for satisfaction scores and dichotomized satisfaction frequency including mean/frequency, standard deviation, median and change scores.

Variable	Pre-Surgery Mean/ Frequency	Pre-Surgery SD	Pre-Surgery Median	3-Months Post-Surgery Mean/ Frequency	3-Months Post-Surgery SD	3-Months Post-Surgery Median	1-Year Post-Surgery Mean/ Frequency	1-Year Post-Surgery SD	1-Year Post-Surgery Median	Change Scores: Pre and 3-Months Post-Surgery	Change Scores: 3-Months and 1-Year Post-Surgery
KSS Patient Satisfaction	13.27	6.73	12	28.73	7.92	30	31.95	8.54	34	15.46	3.22
Dichotomized Dissatisfaction Rate*	255 (93%)			72 (26%)			49 (18%)			-0.67	-0.08

*Patients satisfaction was dichotomized with scores ≤24 being considered dissatisfied and scores >24 being considered satisfied; SD, standard deviation; KSS, New Knee Society Knee Scoring System

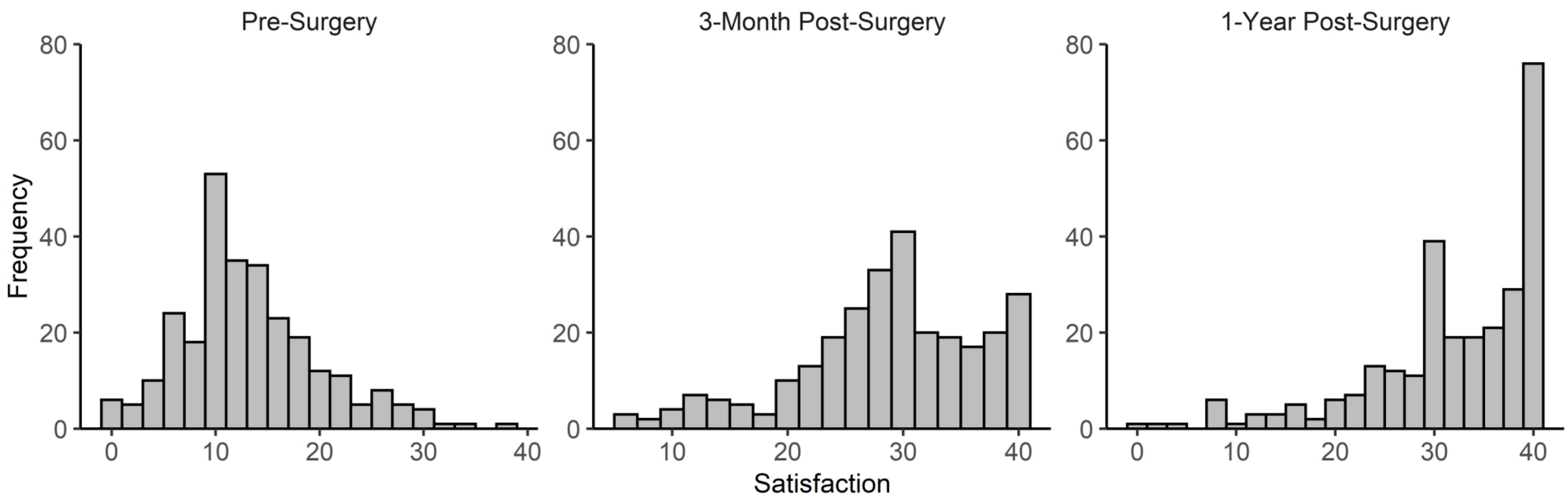


Figure 5.1. Includes histograms of the new knee society scoring system satisfaction scores (0-40 [worst to best]) for patients at pre-surgery, 3-months post-surgery, and 1-year post-surgery.

Approximately 1.2% of all data were found to be missing and was therefore imputed. No individual variable had more than 10% of its data missing.

No significant differences at pre-surgery or 1-year post-surgery were observed between patients who had and had not completed a 3-month post-surgery questionnaire.

For both the pre-surgery only and pre-surgery and 3-months post-surgery logistic ridge regression models, elastic-net penalization was not found to improve model performance. Therefore, standard logistic regression models were reported. The logistic regression model using only pre-surgery independent variables (table 5.3) was found to have an AUC and cross-validated AUC of 0.737 and 0.714 respectively, a Brier score of 0.131, a Cox intercept of 0, and a Cox slope of 1. When establishing a cut point to maximize accuracy, an overall accuracy of 0.825 and a cross-validated accuracy of 0.818 were found. The model's sensitivity was 0.061 and specificity was 0.991. A cut-off value of 0.55 maximized model accuracy in the discrimination analysis, with values above 0.55 predicting that a patient was dissatisfied and below 0.55 predicting that a patient was satisfied (table 5.4). This model included four variables: pre-surgery KSS patient satisfaction, pre-surgery KSS symptoms, pre-surgery KOOS symptoms, and pre-surgery KOOS quality of life. The frequency with which these variables were found in 500 bootstrapped logistic regression models is included in table 5.3. All VIF scores were <5.

Table 5.3. A logistic regression model predicting 1-year TKA patient dissatisfaction using pre-surgery patient reported outcome measures.

	Patient Satisfaction					
Predictors	Estimates	Standard Error	P-Value	Odds Ratios	Confidence Interval	Bootstrap Frequency
Intercept	-0.748	0.622				
KSS Pre-Surgery Patient Satisfaction	-0.129	0.036	**<0.001**	0.88	0.82 – 0.94	0.388
KSS Pre-Surgery Symptoms	-0.148	0.062	**0.018**	0.86	0.76 – 0.97	0.172
KOOS Pre-Surgery Symptoms	0.038	0.012	**0.002**	1.04	1.01 – 1.06	0.646
KOOS Pre-Surgery Quality of Life	0.028	0.014	**0.044**	1.03	1.00 – 1.06	0.536
Observations	275					
AUC/Cross-Validated AUC	0.737/0.714					
Brier Score	0.131					
Cox Intercept	0					
Cox Slope	1					

KSS, Knee Society Score; KOOS, Knee Injury and Osteoarthritis Outcome Score; AUC, area under the curve

Table 5.4. Contingency table for the pre-surgery only logistic regression model illustrating the number of patients correctly predicted as dissatisfied and satisfied, dissatisfied patients incorrectly predicted as satisfied, and satisfied patients incorrectly predicted as dissatisfied. A cut-off value of 0.55 maximized overall model accuracy.

		Actual Outcome		
		Dissatisfied	Satisfied	
Predicted Outcome	Dissatisfied	3	2	5
	Satisfied	46	224	270
		49	226	275

The classification tree model using only pre-surgery variables produced a model that made no splits in the data. This means that the model did not include any variables and simply predicted that all patients would be satisfied. As this model did not perform better than the no information rate, it will not be discussed further in the results section.

The logistic regression model using pre-surgery and 3-month post-surgery variables (table 5.5) was found to have an overall and cross-validated AUC of 0.909 and 0.903 respectively, a Brier score of 0.079, a Cox intercept of 0, and a Cox slope of 1. When establishing a cut point to maximize accuracy, an overall accuracy of 0.898 and a cross-validated accuracy of 0.881 were found. The model's sensitivity was 0.673 and specificity was 0.947. A cut-off of 0.41 maximized model accuracy in the discrimination analysis, with values above 0.41 predicting that a patient was dissatisfied and below 0.41 predicting that a patient was satisfied (table 5.6). There were four variables included in this model: 3-month KOOS ADL, 3-month KSS patient satisfaction, pre-surgery KSS patient satisfaction, and pre-surgery KOOS symptom. The frequency with which the variables appeared in the 500 bootstrapped logistic regression models is included in table 5.5. All VIF scores were <5.

Table 5.5. A logistic regression model predicting 1-year TKA patient dissatisfaction using pre-surgery and 3-month patient reported outcome measures.

	Patient Satisfaction					
Predictors	Estimates	Standard Error	P-Value	Odds Ratios	Confidence Interval	Bootstrap Frequency
Intercept	5.649	1.249				
KOOS 3-Months Activities of Daily Living	-0.089	0.022	**<0.001**	0.92	0.87 – 0.95	0.916
KSS 3-Months Patient Satisfaction	-0.097	0.037	**0.008**	0.91	0.84 – 0.97	0.276
KSS Pre-Surgery Patient Satisfaction	-0.111	0.044	**0.012**	0.89	0.82 – 0.97	0.408
KOOS Pre-Surgery Symptoms	0.066	0.016	**<0.001**	1.07	1.04 – 1.10	0.646
Observations	275					
AUC/Cross-Validated AUC	0.909/0.903					
Brier Score	0.079					
Cox Intercept	0					
Cox Slope	1					

KSS, Knee Society Knee Scoring System; KOOS, Knee Injury and Osteoarthritis Outcome Score; AUC, area under the curve

Table 5.6. Contingency table for the pre-surgery and 3-month post-surgery logistic regression model illustrating the number of patients correctly predicted as dissatisfied and satisfied, dissatisfied patients incorrectly predicted as satisfied, and satisfied patients incorrectly predicted as dissatisfied. A cut-off value of 0.41 was used for the model to maximize overall accuracy.

		Actual Outcome		
		Dissatisfied	Satisfied	
Predicted Outcome	Dissatisfied	33	12	45
	Satisfied	16	214	230
		49	226	275

The classification tree model using pre-surgery and 3-month post-surgery variables had an overall accuracy of 0.884, a cross-validated accuracy of 0.856, a sensitivity of 0.694, and a specificity of 0.925 (figure 5.2 & table 5.7). This model made use of two variables and included two splits and three terminal nodes. The variables included in this model were the 3-month KOOS ADL and 3-month KSS patient satisfaction. This model predicted that patients with a KOOS ADL score <67 and a KSS patient satisfaction score <27 would be dissatisfied, whereas patients with a 3-month activities of daily living score ≥ 67 would be satisfied, and patients with a 3-month KOOS ADL score < 67 but a 3-month KSS patient satisfaction score ≥ 27 would be satisfied.

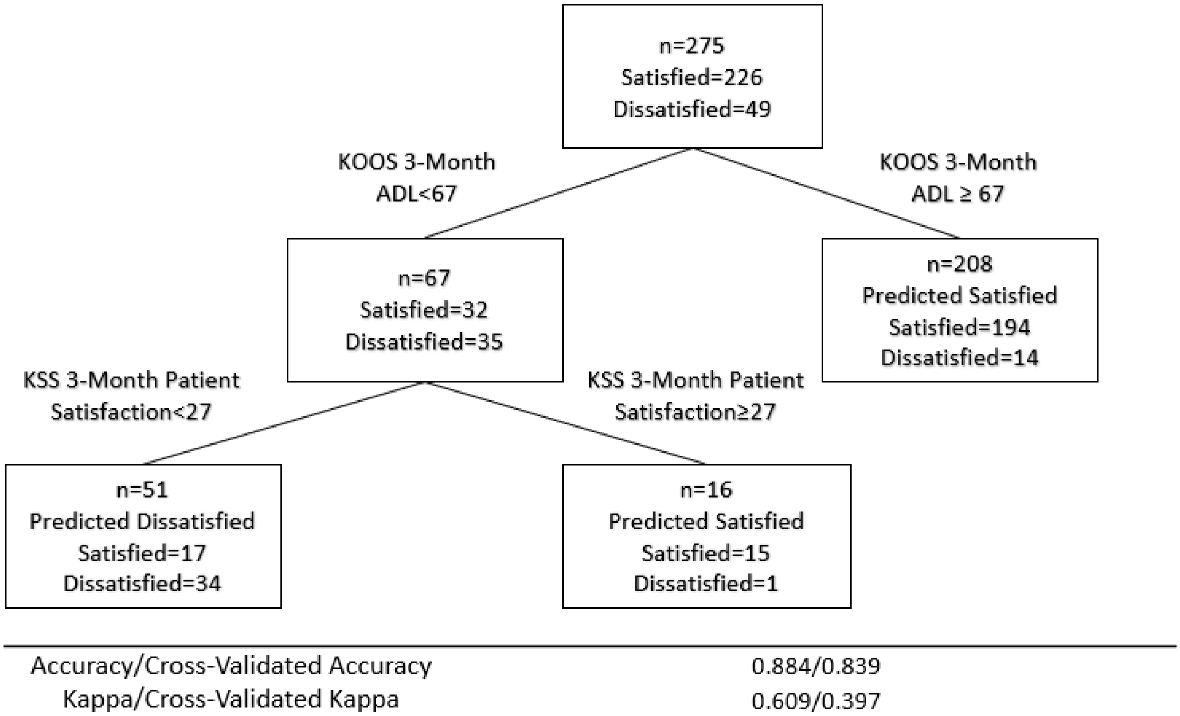


Figure 5.2. A classification tree predicting 1-year TKA patient satisfaction using pre-surgery and 3-month post-surgery patient reported outcome measures. (3-month activities of daily living on the KOOS measurement scale 0-100 [worst to best] and 3-month patient satisfaction KSS measurement scale 0-40 [worst to best]).

KOOS, Knee Injury and Osteoarthritis Outcome Score; ADL, activities of daily living; KSS, Knee Society Knee Scoring System; TKA, total knee arthroplasty.

Table 5.7. Cross tabulated table for the classification tree model using pre-surgery and 3-month post-surgery patient reported outcome measures illustrating the number of patients correctly predicted as dissatisfied and satisfied, dissatisfied patients incorrectly predicted as satisfied and satisfied patients incorrectly predicted as dissatisfied.

		Actual Outcome		
		Dissatisfied	Satisfied	
Predicted Outcome	Dissatisfied	34	17	51
	Satisfied	15	209	224
		49	226	275

The accuracy, cross-validated accuracy, kappa, cross-validated kappa, sensitivity, specificity, LR+, LR-, PPV, and NPV for the logistic regression model using only pre-surgery independent variables, the logistic regression model using pre-surgery and 3-month post-surgery independent variables and the classification tree using pre-surgery and 3-month post-surgery independent variables have been included in table 5.8.

Table 5.8. The accuracy/cross-validated accuracy, kappa/cross-validated kappa, sensitivity, specificity, likelihood of a positive test result, likelihood of a negative test result, positive predictive value and negative predictive value calculations for the logistic regression and classification tree models.

Measure	Logistic Regression Pre-Surgery	Logistic Regression Pre-Surgery and 3-Months Post-Surgery	Classification Tree
Accuracy/Cross-Validated Accuracy	0.825/0.818	0.898/0.881	0.884/0.856
Kappa/Cross-Validated Kappa	0.081/0.071	0.641/0.534	0.609/0.455
Sensitivity	0.061	0.673	0.694
Specificity	0.991	0.947	0.925
LR+	6.918	12.684	9.224
LR-	0.947	0.345	0.331
PPV	0.6	0.733	0.667
NPV	0.83	0.93	0.933

LR+, positive likelihood ratio; LR-, negative likelihood ratio; PPV, positive predictive value; NPV, negative predictive value

5.4 Discussion

The prediction models using only pre-surgery variables generally did not perform well. Despite having acceptable AUC values, the logistic regression model performed only slightly better than simply predicting that all patients are satisfied when evaluating accuracy. Although accuracy is not always the most important measurement in a prediction model, it is still likely desirable that model accuracy be greater than simply predicting that all patients are satisfied. The pre-surgery CART model did not actually include any variables and simply predicted that all patients would be satisfied with their surgery. There does not appear to be enough information using the pre-surgery variables included in this study to predict dissatisfaction at 1-year post-surgery accurately.

In contrast, it appears that patient dissatisfaction at 1-year post-surgery can be predicted using a combination of pre-surgery and 3-month post-surgery data. Both logistic regression and classification tree models predicted dissatisfaction accurately. The logistic regression model performed better than the classification tree, both in overall and cross-validated accuracy. Additionally, the logistic regression model had good AUC and cross-validated AUC values indicating that it has strong discriminatory capacity across a range of cut-off values. The similar overall and cross-validated accuracy and AUC of the logistic regression model suggests that it could generalize well to new data. However, when evaluating the CART model, there was a drop off between overall and cross-validated accuracy for the classification tree. This finding would suggest that this model may have been slightly overfitting the data.

Both logistic regression models completed in this study were constrained by the sample size of the study and included only four variables. The small number of variables may have also been why the elastic-net penalization did not improve model performance for either logistic regression model. When examining the pre-surgery model, higher pre-surgery KSS patient satisfaction and pre-surgery KSS symptoms were negatively associated with dissatisfaction. Whereas higher pre-surgery KOOS symptoms and pre-surgery KOOS quality of life scores were positively associated with dissatisfaction (table 5.2). For the 3-month post-surgery model 3-month KOOS ADL, 3-month KSS patient satisfaction, and pre-surgery KSS patient satisfaction were negatively associated with dissatisfaction. Whereas pre-surgery KOOS symptoms was positively associated with dissatisfaction (table 5.4).

The final classification tree model was simple and included two variables and three terminal nodes. This means that at most two steps or two cut-off values would be required to generate a prediction (figure

5.1). The two variables included in the model were the 3-months KOOS ADL and the 3-months KSS patient satisfaction sub scores. A score of 67 on the 3-months KOOS ADL sub score would indicate that, on average, patients report between mild and moderate discomfort performing the 17 activities included in the sub score. A KSS patient satisfaction score of 27 would indicate that a patient is satisfied, but not extremely satisfied. Both of these scores could indicate minimum thresholds that a patient would need to achieve by the 3-months post-surgery time point in order to be satisfied at 1-year post-surgery.

Both models that used pre-surgery and 3-months post-surgery data performed well. This could be attributable to the pattern of recovery for TKA patients. There is rapid improvement up to 3-months post-surgery, followed by much slower and gradual improvement between 3-months and 1-year post-surgery [41]. Therefore, a patient's general function and pain scores at 3-months post-surgery is indicative of how satisfied they will be at 1-year post-surgery. It appears that the 3-month KOOS ADL, and 3-month post-surgery dissatisfaction variables are particularly important, as both 3-month post-surgery models found that these variables were significant predictors of dissatisfaction at 1-year. The KOOS ADL contains a series of questions that cover a wide range of activities that a patient would be expected to perform daily, for example, light household chores or walking up and down stairs. The inability to perform these tasks adequately 3-months after surgery may be a warning sign that patients are unlikely to be satisfied at 1-year post-surgery. It is possible that targeting the ability to perform the activities included in this domain could be a potential area for intervention. Dissatisfaction at 3-months post-surgery evaluates a patient's satisfaction with their pain while sitting, pain while lying, ability to rise from sitting, ability to perform ADL, and ability to perform recreational activities. It appears that a patient's satisfaction at 3-months is extremely important for their satisfaction at 1-year post-surgery. As patient satisfaction is a measure that includes a number of different domains, the significance of this variable does not suggest any clear avenues for intervention.

Interestingly, in both logistic regression models, higher pre-surgery KOOS symptoms indicate that a patient is less likely to be satisfied at 1-year post-surgery. The KOOS symptoms variable evaluates the physical characteristics of the knee, including, grinding, clicking, swelling, stiffness, and ability to straighten and bend the knee. It appears that if the pre-surgery symptoms in a patients' knee are not as severe, they are more likely to be dissatisfied at 1-year post-surgery.

Despite the apparent success of both the logistic regression and classification tree models that made use of pre-surgery and 3-months post-surgery variables, there is more than model performance to consider when discussing the possible implementation of these models. While the logistic regression model does perform very well, it would be difficult to use in a clinical setting. In order to be implemented, it would require that patients fill out several questionnaires at multiple time points. The questionnaire scores would then need to be calculated and run through the logistic regression model in order to generate a prediction. While this is feasible, implementation could be difficult and expensive. Although the performance of the classification tree was not optimal, it does have several advantages in that it is simple, using only two variables and it is easy to understand. This model could be transported to a clinical setting without much difficulty.

This study had several limitations, with the primary issue being that the sample may not have been large enough to generate a robust prediction model that will generalize well to other samples. The classification tree model is of particular concern here as the cross-validated accuracy was lower than the overall accuracy, indicating that the model may have been overfitting the sample. The lack of stability in several of the predictors when performing the 500 iterations of the bootstrapped sample in both logistic regression models also highlights this issue. Additionally, this study may not have included all variables that are relevant to satisfaction. There may be missing demographic, clinical, or surgical variables that could improve the performance of the models that are not included in this study. Lastly, although no differences were observed at pre-surgery and 1-year post-surgery between patients who did and did not complete a 3-month post-surgery questionnaire, it is possible that there may have been differences between these patients at 3-months post-surgery. For example, patients may not have completed the questionnaire at this time point because they were dissatisfied with their knee at 3-months. This could potentially skew the results of the prediction models.

In conclusion, dissatisfaction at 1-year can be predicted using pre-surgery and 3-month data using both logistic regression and classification trees. However, there are real advantages in terms of interpretability and clinical use to creating an accurate classification tree model. It also appears that the 3-month post-surgery KOOS ADL and 3-month post-surgery KSS patient satisfaction variables are particularly important in determining if a patient will be dissatisfied at 1-year post-surgery. These findings need to be replicated with a larger sample size in order to attempt to improve the performance of the models created in this study. Using prediction models to identify patients likely to be dissatisfied and targeting them for intervention early in the recovery process could help decrease the number of patients who are dissatisfied with their TKA.

5.5 References

1. Kahlenberg CA, Nwachukwu BU, Mclawhorn AS, Cross MB, Cornell CN, Padgett DE. Patient satisfaction after total knee replacement: a systematic review. HSS J. 2018;14(2):192–201.
2. Gunaratne R, Pratt DN, Banda J, Fick DP, Khan RJK, Robertson BW. Patient dissatisfaction following total knee arthroplasty: a systematic review of the literaturere. J Arthroplasty. 2017;32(12):3854–60.
3. Clement ND, Bardgett M, Weir D, Holland J, Gerrand C, Deehan DJ. The rate and predictors of patient satisfaction after total knee arthroplasty are influenced by the focus of the question: a standard satisfaction question is required. Bone Joint J. 2018;100(6):740–8.
4. Bourne RB, Chesworth BM, Davis AM, Mahomed NN, Charron KDJ, Met D. Patient satisfaction after total knee arthroplasty who is satisfied and who is not? Clin Orthop Relat Res. 2010;468(1):57–63.
5. Kim TK, Chang CB, Kang YG, Kim SJ, Seong SC. Causes and predictors of patient's dissatisfaction after uncomplicated total knee arthroplasty. J Arthroplasty. 2009;24(2):263–71.
6. Van Onsem S, Verstraete M, Dhont S, Zwaenepoel B, Van Der Straeten C. Improved walking distance and range of motion predict patient satisfaction after TKA. Knee Surgery, Sport Traumatol Arthrosc. 2018;26(11):3272–9.
7. Vissers MM, Groot IB De, Reijman M, Bussmann JB, Stam HJ, Verhaar JAN. Functional capacity and actual daily activity do not contribute to patient satisfaction after total knee arthroplasty. BMC Musculoskelet Disord. 2010;11(1):1–8.
8. Barlow T, Clark T, Dunbar M, Metcalfe A, Griffin D. The effect of expectation on satisfaction in total knee replacements: a systematic review. Springerplus. 2016;5(1):1–7.
9. Baker PN, Rushton S, Jameson SS, Reed M, Gregg P, Deehan DJ. Patient satisfaction with total knee replacement cannot be predicted from pre-operative variables alone: a cohort study from the National Joint Registry for England and Wales. Bone Joint J. 2013;95(10):1359–65.
10. Pivec R, Issa K, Given K, Harwin SF, Greene KA, Hitt KD, et al. A prospective, longitudinal study of patient satisfaction following total knee arthroplasty using the short-form 36 (sf-36) survey stratified by various demographic and comorbid factors. J Arthroplast. 2015;30(3):374–8.
11. Scott CEH, Howie CR, Macdonald D, Biant LC. Predicting dissatisfaction following total knee replacement: a prospective study of 1217 patients. J Bone Joint Surg Br. 2010;92(9):1253–8.
12. Becker R, Döring C, Denecke A, Brosz M. Expectation, satisfaction and clinical outcome of patients after total knee arthroplasty. Knee surgery, Sport Traumatol Arthrosc. 2011;19(9):1433–

41.

13. Mannion AF, Kämpfen S, Munzinger U, Quervain IK. The role of patient expectations in predicting outcome after total knee arthroplasty. Arthritis Res Ther. 2009;11(5):1–13.

14. Thambiah MD, Nathan S, Seow BZX, Liang S, Lingaraj K. Patient satisfaction after total knee arthroplasty: an Asian perspective. Singapore Med J. 2015;56(5):259–63.

15. Garriga C, Sanchez-Santos MT, Judge A, Perneger T, Hannouche D, Lübbeke A, et al. Development of a model predicting non-satisfaction 1 year after primary total knee replacement in the UK and transportation to Switzerland. Sci Rep. 2018;8(1):1–8.

16. Culliton SE, Bryant DM, MacDonald SJ, Hibbert KM, Chesworth BM. Effect of an e-learning tool on expectations and satisfaction following total knee arthroplasty: a randomized controlled trial. J Arthroplasty. 2018;33(7):2153–8.

17. Gautreau S, Gould ON, Aquino-Russell C, Forsythe ME. Developing a surgeon-patient communication checklist for total knee arthroplasty. Musculoskeletal Care. 2019;17(1):91–6.

18. Padilla JA, Feng JE, Anoushiravani AA, Hozack WJ, Schwarzkopf R, Macaulay WB. Modifying patient expectations can enhance total hip arthroplasty postoperative satisfaction. J Arthroplasty. 2019;34(7):S209–14.

19. Zabawa L, Li K, Chmell S. Patient dissatisfaction following total knee arthroplasty: external validation of a new prediction model. Eur J Orthop Surg Traumatol. 2019;29(4):861–7.

20. Van Onsem S, Van Der Straeten C, Arnout N, Deprez P, Van Damme G, Victor J. A new prediction model for patient satisfaction after total knee arthroplasty. J Arthroplasty. 2016;31(12):2660–7.

21. Kunze KN, Polce EM, Sadauskas AJ, Levine BR. Development of machine learning algorithms to predict patient dissatisfaction after primary total knee arthroplasty. J Arthroplasty. 2020;35(11):3117–22.

22. Farooq H, Deckard ER, Ziemba-Davis M, Madsen A, Meneghini RM. Predictors of patient satisfaction following primary total knee arthroplasty: results from a traditional statistical model and a machine learning algorithm. J Arthroplasty. 2020;35(11):3123–30.

23. Tolk JJ, Janssen RPA, Haanstra TM, van der Steen MC, Bierma-Zeinstra SMA, Reijman M. The influence of expectation modification in knee arthroplasty on satisfaction of patients: a randomized controlled trial. Bone Joint J. 2021;103(4):619–26.

24. Roos EM, Toksvig-Larsen S. Knee injury and Osteoarthritis Outcome Score (KOOS) - validation and comparison to the WOMAC in total knee replacement. Health Qual Life Outcomes. 2003;1(1):17.

25. Noble PC, Scuderi GR, Brekke AC, Sikorskii A, Benjamin JB, Lonner JH, et al. Development of a

new Knee Society scoring system. Clin Orthop Relat Res. 2012;470(1):20–32.

26. Ware JE, Krosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996;34(3):220–33.

27. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: development and validation. Psychol Assess. 1995;7(4):524–32.

28. Naal FD, Impellizzeri FM, Leunig M. Which is the best activity rating scale for patients undergoing total joint arthroplasty? Clin Orthop Relat Res. 2009;467(4):958–65.

29. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand. 1983;67(6):361–70.

30. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for statistical Computing; 2020.

31. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.

32. Torgo L. Data mining with R: learning with case studies. 1st ed. Boca Raton, FL: Taylor & Francis; 2010.

33. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49(12):1373–9.

34. Fox J, Weisberg S. An R Companion to Applied Regression. Third. Thousand Oaks CA: Sage; 2019.

35. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1–22.

36. Rizopolous D. bootStepAIC: Bootstrap stepAIC. R package 1.2-0; 2009.

37. Kuhn M. caret: Classification and Regression Training. R package version 6.0-88; 2021.

38. Therneau T, Atkinson EJ. rpart: Recursive Partitioning and Regression Trees. R package version 4.1-15; 2019.

39. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.

40. Harrell F. rms: Regression Modeling Strategies. R package version 6.2-0; 2021.

41. Stevens-Lapsley JE, Schenkman ML, Dayton MR. Comparison of self-reported Knee Injury and Osteoarthritis Outcome Score to performance measures in patients after total knee arthroplasty. PM R. 2011;3(6):541–9.

6 Can a Pooled Index Predict Dissatisfaction With 1-Year Total Knee Arthroplasty Early in the Recovery Process?

6.1 Introduction

Total knee arthroplasty (TKA) patient satisfaction is a multidimensional patient reported outcome measure (PROM) that provides the patient's evaluation of the success of their surgery. Satisfaction has most consistently been associated with pain[1–6], function[1,3,5–8], and met expectations[2,4,9–12]. Although TKA has been shown to be an effective treatment for end-stage osteoarthritis, approximately 20% of patients are consistently found to be dissatisfied or unsure of their satisfaction with their surgery[6,13]. This 20% dissatisfaction rate has persisted despite improvements in both the quality of implants and surgical techniques[6,13].

If effective prediction models can be developed to identify dissatisfied patients early in the recovery process, it may be possible to reduce dissatisfaction rates. A simple way in which this could be done, while accounting for satisfaction's multidimensional nature, would be through the use of a pooled index composed of PROMs items[14,15]. A pooled index is a method in which a combination of usually five or six different outcome measures are weighted and then summed together to create a single overall score[14,15]. If a computationally simple pooled index can be created and a cut point that discriminates between satisfied and dissatisfied patients can be established, it may be possible to create a simple and easily administered prediction model.

Attempting to create a TKA satisfaction pooled index using existing subscales would require that patients answer a significant number of individual PROM items. This would place a substantial burden on patients, be time consuming to administer, and difficult to score. In order to reduce the burden on patients and limit the number of questions they were required to answer, Van Onsem[16] used individual pre-surgery PROM questionnaire items in regression models to predict 3-month post-surgery patient satisfaction. Although this model did not successfully translate to new data, using individual questionnaire items in the form of a pooled index could allow for the creation of a prediction tool that is easy to administer and can be used early in the recovery process to predict 1-year post-surgery dissatisfaction.

The purpose of this study was to create a pooled index using individual PROM questionnaire items and to perform a cut point analysis on that index to determine if 1-year patient dissatisfaction can be predicted at 3-months post-surgery using a simple prediction tool.

6.2 Methods

6.2.1 Patients

This was a prognostic observational cohort study that retrospectively analyzed prospectively collected data. Patient data used in this analysis was drawn from a randomized control trial (RCT) that aimed to employ an e-learning tool to generate realistic expectations for TKA patients and by association improve patient satisfaction[17]. No differences between the treatment and control groups were observed. As the e-learning tool was observed to have no effect, it was considered acceptable to pool patients into a single sample.

For the initial RCT, patients scheduled to undergo TKA were recruited from a pre-admission clinic at a large teaching institution (University Hospital London, Ontario, Canada) between April 2013 and April 2014. The inclusion criteria for the RCT required that patients be 20 years of age or older, have been diagnosed with knee osteoarthritis, and have the cognitive capacity to consent to participate. Patients undergoing patellar resurfacing, hemi or unicondylar knee arthroplasty, high tibial osteotomy, or knee surgery to address a tumor were excluded. Questionnaires were completed pre-surgery and at 6-weeks, 3-months, and 1-year post-surgery. Patients completed questionnaires online and responses were stored in a secure database.

The inclusion criteria for the present study required that patients have completed questionnaires at pre-surgery, 3-months post-surgery, and 1-year post-surgery.

This study received ethics approval from the university ethics review board.

6.2.2 Outcome Measures

PROMs included in the study were the Knee Society Knee Scoring System (KSS; pre-operative and post-operative versions)[18], the Knee Injury and Osteoarthritis Outcome Score (KOOS)[19], the 12-Item Short Form Health Survey version 2 (SF-12)[20], the Self-Administered Comorbidity Questionnaire (SCQ), the

Hospital Anxiety and Depression Scale (HADS)[21], the University of California at Los Angeles (UCLA) Activity Score[22], and the Pain Catastrophizing Scale (PCS)[23].

The KSS satisfaction subscale was used as the outcome variable. This subscale consists of five questions that ask patients to rate their satisfaction with pain while sitting, pain while lying down, their ability to get out of bed, ability to perform light household duties, and ability to perform recreational activities. Each question includes the response options (score): very satisfied (8), satisfied (6), neutral (4), dissatisfied (2), and very dissatisfied (0). The five individual items are summed to generate a score between 0 and 40 (least to most satisfied). Satisfaction scores were dichotomized into satisfied and dissatisfied groups. Patients with satisfaction scores >24 were considered satisfied and patients with satisfaction scores ≤24 were considered dissatisfied. A score of ≤24 was chosen as this would mean that a patient must have responded to at least one of the satisfaction questions as dissatisfied or three questions as neutral.

6.2.3 Statistical Analysis

All statistical analyses were completed in R version 4.0.3[24].

Descriptive statistics were calculated, and paired t-tests were used to test for differences between 3-month and 1-year post-surgery outcomes. Missing data was imputed using K nearest neighbours imputation using the DMwR2 R package[25].

An initial correlation matrix was calculated using individual SF-12, KOOS, and KSS 3-month post-surgery items as well as the dichotomized 1-year post-surgery satisfaction variable. Area under the curve (AUC) was calculated using each individual KOOS and KSS questionnaire item as the predictor and dichotomized 1-year post-surgery dissatisfaction as the outcome variable. Individual questionnaire items that were most correlated with dissatisfaction and had the highest individual AUC values were identified as candidate variables to be included in the pooled index. Correlations between candidate variables for the pooled index were also evaluated to ensure that variables were not too closely related. A correlation between items of <0.6 was preferred[14].

A number of different potential pooled indices were created with a maximum of six items included in each index. A limit of six items was chosen as there has been little benefit observed in including more than six variables[14]. AUC was calculated for each pooled index that was created using the pROC R package[26]. Bootstrapping with 1000 repetitions was used to calculate confidence intervals. When choosing the final pooled index, both the number of items included and the AUC score were considered, with fewer items being preferable. Only the optimal pooled index was reported. For each pooled index created, items were transformed so as to be oriented in the same direction and on the same scale if necessary. Items in each index were combined by summing the individual variables together. Weights were not applied to any items in the index. This method of combining items was chosen in order to maximize computational simplicity.

For the final pooled index, a cut point analysis maximizing overall accuracy was completed[27]. A second cut point that aimed to balance sensitivity and specificity scores was also established. Cut points were established using the cutpointr R package[28]. Using the selected cut points, accuracy, sensitivity, specificity, positive likelihood ratio (LR+), negative likelihood ratio (LR-), positive predictive value (PPV), and negative predictive value (NPV) were calculated.

No sample size calculation was completed as there is limited information about necessary sample sizes required for the creation of a pooled index.

6.3 Results

A total of 416 patients were recruited to participate in the initial RCT. Of this original sample 50 did not undergo TKA or were missing measures from the pre-admission clinic. Of the 366 remaining eligible patients, 275 met the inclusion criteria for this study.

Descriptive statistics for demographic information and patient outcome scores are included in table 6.1.

The final pooled index that was chosen included five individual items (table 6.2) that were summed together to generate a score ranging from 0-20 (best to worst state). The pooled index was composed of four items from the KOOS and one item from the KSS questionnaires. Items included from the KOOS were the P1 and P6 questions from the pain subscale, and the A3 and A17 questions from the KOOS activities of daily living (ADL) subscales. The item included from the KSS was the third question from the met expectations' subscale (table 6.2). The met expectations variable had its scoring direction reversed to match the KOOS items and was transformed so as to be scored from 0-4 (best to worst) instead of 1-5 (worst to best).

The pooled index was found to have an AUC of 0.89 with a 95% confidence interval of 0.825-0.941. The cut point analysis maximizing accuracy found that if the pooled index was ≥12, the patient was predicted as dissatisfied and if the pooled index was <12, the patient was predicted as satisfied. The cut point analysis balancing sensitivity and specificity found that if the pooled index score was ≥10, the patient was predicted as dissatisfied and if the pooled index was <10, the patient was predicted as satisfied. The measures of model performance are included in table 6.3.

Table 6.1. Descriptive statistics including mean and standard deviations with repeated measures t-tests evaluating significant differences between 3-months and 1-year post-surgery satisfaction.

	3-Months Post-Surgery n=275		1-Year Post-Surgery n=275		T-Test P-Values
Variable	Mean	SD	Mean	SD	
Age	64.65	8.45			
Comorbidities	2.77	1.27			
Sex (Female)*	176 (64%)				
BMI	32.97	7.45			
KSS Patient Satisfaction Dichotomized at 1-Year (Satisfied)*			226 (82%)		
KSS Patient Satisfaction Score	28.73	7.92	31.95	8.54	<0.001
KSS Symptoms	17.67	5.04	19.54	4.97	<0.001
KSS Patient Expectations	8.58	2.39	9.16	2.78	0.009
KSS Walking Standing	23.95	4.74	25.50	4.41	<0.001
KSS Standard Activity	21.53	4.81	23.24	5.23	<0.001
KSS Advanced Activity	9.08	5.59	11.51	5.80	<0.001
KSS Discretionary Activity	9.83	3.43	11.37	3.01	<0.001
KOOS Pain	72.59	16.77	82.06	17.34	<0.001
KOOS Activities of Daily Living	76.47	15.39	83.43	16.46	<0.001
KOOS Functional Sport and Recreation	42.05	27.39	53.76	28.36	<0.001
KOOS Quality of Life	55.63	20.20	65.75	21.82	<0.001
KOOS Symptoms	66.63	16.00	75.98	15.38	<0.001
PCS Rumination	5.99	2.84	5.68	3.00	0.211
PCS Magnification	3.77	1.56	3.82	1.64	0.737
PCS Helplessness	7.90	3.01	7.94	3.45	0.895
SF-12 Physical Function	38.88	10.40	41.51	10.70	0.004
SF-12 Role Physical	42.15	9.34	45.52	9.69	<0.001
SF-12 Bodily Pain	43.97	10.02	47.47	10.41	<0.001
SF-12 General Health	50.93	8.60	51.02	9.01	0.901
SF-12 Vitality	50.82	9.13	52.79	9.19	0.012
SF-12 Social Function	50.53	8.49	51.55	8.19	0.153
SF-12 Role Emotional	83.85	21.26	88.01	18.65	0.015
SF-12 Mental Health	92.16	17.73	93.57	17.44	0.346
UCLA Activity Score	5.12	1.65	5.69	1.75	<0.001

*Variable presented as a frequency (percentage) not a mean; SD, standard deviation; KSS, Knee Society Knee Scoring System; KOOS, Knee Injury and Osteoarthritis Outcome Score; PCS, pain catastrophizing scale; SF-12, 12-Item Short Form Health Survey version 2; UCLA, University of California Los Angeles

Table 6.2. Variables included in the pooled index created from individual questionnaire item scores from the 3-months post-surgery Knee Injury and Osteoarthritis Outcome Score (KOOS).

Variable-1 – Included individual item scores are summed to create a score ranging from 0-20 (Best to Worst).

KOOS Pain subscale

- P1. How often do you experience knee pain? (0-4)*
- What amount of knee pain have you experienced the last week during the following activities?
    - P6. Going up or down stairs (0-4)*

KOOS Activities of Daily Living subscale

- For each of the following activities please indicate the degree of difficulty you have experienced in the last week due to your knee:
    - A3. Rising from sitting (0-4)*
    - A17. Light domestic duties (cooking, dusting, etc) (0-4)*

KSS Met Expectations Subscale

- Compared to what you expected before your knee replacement:
    - 3 – My expectations for being able to do my leisure, recreational or sports activities were… (0-4)*

*Individual questions were scored from best (0) to worst (4); KOOS, Knee Injury and Osteoarthritis Outcome Score; KSS, Knee Society Knee Scoring System

Table 6.3. Outcome measures using a cut point analysis maximizing accuracy and balancing sensitivity and specificity. AUC for the pooled index was 0.89 (0.825-0.941).

Measure	Cut Point Maximizing Accuracy	Cut Point Balancing Sensitivity and Specificity
Cut Point	12	10
Accuracy	0.898	0.829
Sensitivity	0.571	0.837
Specificity	0.969	0.827
Kappa	0.609	0.532
LR+	18.449	4.849
LR-	0.442	0.197
PPV	0.8	0.512
NPV	0.912	0.959

AUC, area under the curve; LR+, positive likelihood ratio; LR-, negative likelihood ratio; PPV, positive predictive value; NPV, negative predictive value

6.4 Discussion

A pooled index with a cut point can be used to predict 1-year dissatisfaction at 3-months post-surgery. The AUC of the pooled index was 0.892 with a 95% confidence interval of 0.825-0.941 which is considered to be good discriminatory capacity. After testing multiple combinations of items, the optimal pooled index included five individual items from the KOOS and KSS. Due to its simplicity, this tool can be quickly administered, easily interpreted, and a prediction can be generated using only a pen and paper. This represents an easy and inexpensive way to accurately identify patients early in the recovery process who are likely to be dissatisfied at 1-year post-surgery. The identification of these patients at 3-months post-surgery may allow for targeted interventions that improve patient satisfaction to be employed before 1-year post-surgery where outcomes tend to be more fixed.

When examining the cut point that maximized accuracy, there was a problem with low sensitivity, or difficulty identifying truly dissatisfied patients as dissatisfied. However, sensitivity can be improved by adjusting the cut point of the scale. Although, this comes at the expense of overall accuracy and specificity as demonstrated in table 6.3. The choice of cut point may depend on the priorities of the decision maker using the scale. For example, a decision maker may choose to maximize accuracy if an intervention that is being implemented is expensive and it is important that people do not receive the intervention unnecessarily. In contrast, if an intervention is inexpensive and can be more easily administered to a variety of patients, having a higher sensitivity score may be more important. In this situation, dissatisfied patients would be identified more accurately, but a larger number of patients would unnecessarily receive an intervention.

The optimal pooled index in this study included five individual items from the KOOS and KSS questionnaires. The items were drawn from two different KOOS subscales (Pain and ADL) and one KSS subscale (Met Expectations). The items included in the pooled index addressed the pain, function and met expectations dimensions of satisfaction in some capacity. The ability for patients to perform specific activities at some sort of baseline level at 3-months post-surgery appears to be important in predicting dissatisfaction at 1-year post-surgery. Additionally, a patient's met expectations with regards to their ability to perform sport and recreational activities appears to be an important component of predicting dissatisfaction.

Typically, a pooled index is used to evaluate a medical intervention or the outcome of treatment for a disease that has multiple clinically important disease manifestations or multiple different outcomes[14]. This type of tool appears to be well suited to predicting satisfaction, as satisfaction is a multifaceted evaluation by the patient of the success of their surgery. Using a pooled index in this context has several advantages. A pooled index avoids multiplicity or evaluating each outcome individually[15]. The condensed score that a pooled index creates can provide a common measure that allows the net effect of the treatment to be evaluated. Additionally, the creation of an index can reduce measurement noise or variance in the score[15]. However, it is not clear that there will be a reduction in variance in this case as a limited number of individual PROMs questions were used to create the index.

Despite its strengths, there are some disadvantages to using a pooled index. A pooled index is not necessarily easily interpretable as it is a combination of potentially unrelated items[15]. However, as the purpose of this index is to be a prediction tool, interpretability is less of a concern. Additionally, the choice of weights or method of combining the variables is not always straightforward and it can be computationally difficult to calculate a score from a pooled index if variables are standardized or transformed[15]. In the case of this study, the transformation process was not difficult, and items were simply combined into a sum facilitating calculation of the score.

This study had a number of limitations. The sample size was small, including only 275 patients. Because of this small sample size, it is difficult to have a high degree of confidence that the optimal items were selected for inclusion for a tool of this nature. Additionally, the pooled index used in this study has not been validated and its properties need to be investigated further. All items included in the pooled index have face validity as they have been pulled from the KOOS and KSS. However, the item scores were retrieved from a completed KOOS and KSS questionnaire and were not administered as a subscale. Patients may not answer the five questions included in the pooled index in the same way if these questions were asked in isolation.

In conclusion, it appears that a small number of individual PROM questions can be used to generate a pooled index that accurately identifies patients likely to be dissatisfied at 1-year post-surgery early in the recovery process. It is possible that this type of measure could be used in conjunction with post-surgery interventions to maximize their effectiveness and improve patient satisfaction.

6.5 References

1. Kim TK, Chang CB, Kang YG, Kim SJ, Seong SC. Causes and predictors of patient's dissatisfaction after uncomplicated total knee arthroplasty. J Arthroplasty. 2009;24(2):263–71.
2. Bourne RB, Chesworth BM, Davis AM, Mahomed NN, Charron KDJ, Met D. Patient satisfaction after total knee arthroplasty who is satisfied and who is not? Clin Orthop Relat Res. 2010;468(1):57–63.
3. Clement ND, Bardgett M, Weir D, Holland J, Gerrand C, Deehan DJ. The rate and predictors of patient satisfaction after total knee arthroplasty are influenced by the focus of the question: a standard satisfaction question is required. Bone Joint J. 2018;100(6):740–8.
4. Vissers MM, Groot IB De, Reijman M, Bussmann JB, Stam HJ, Verhaar JAN. Functional capacity and actual daily activity do not contribute to patient satisfaction after total knee arthroplasty. BMC Musculoskelet Disord. 2010;11(1):1–8.
5. Van Onsem S, Verstraete M, Dhont S, Zwaenepoel B, Van Der Straeten C. Improved walking distance and range of motion predict patient satisfaction after TKA. Knee Surgery, Sport Traumatol Arthrosc. 2018;26(11):3272–9.
6. Kahlenberg CA, Nwachukwu BU, Mclawhorn AS, Cross MB, Cornell CN, Padgett DE. Patient satisfaction after total knee replacement: a systematic review. HSS J. 2018;14(2):192–201.
7. Scott CEH, Howie CR, Macdonald D, Biant LC. Predicting dissatisfaction following total knee replacement: a prospective study of 1217 patients. J Bone Joint Surg Br. 2010;92(9):1253–8.
8. Baker PN, Rushton S, Jameson SS, Reed M, Gregg P, Deehan DJ. Patient satisfaction with total knee replacement cannot be predicted from pre-operative variables alone: a cohort study from the National Joint Registry for England and Wales. Bone Joint J. 2013;95(10):1359–65.
9. Mannion AF, Kämpfen S, Munzinger U, Quervain IK. The role of patient expectations in predicting outcome after total knee arthroplasty. Arthritis Res Ther. 2009;11(5):1–13.
10. Barlow T, Clark T, Dunbar M, Metcalfe A, Griffin D. The effect of expectation on satisfaction in total knee replacements: a systematic review. Springerplus. 2016;5(1):1–7.
11. Becker R, Döring C, Denecke A, Brosz M. Expectation, satisfaction and clinical outcome of patients after total knee arthroplasty. Knee surgery, Sport Traumatol Arthrosc. 2011;19(9):1433–41.
12. Thambiah MD, Nathan S, Seow BZX, Liang S, Lingaraj K. Patient satisfaction after total knee arthroplasty: an Asian perspective. Singapore Med J. 2015;56(5):259–63.
13. Gunaratne R, Pratt DN, Banda J, Fick DP, Khan RJK, Robertson BW. Patient dissatisfaction

following total knee arthroplasty: a systematic review of the literaturere. J Arthroplasty. 2017;32(12):3854–60.

14. Goldsmith CH, Smyth HA, Helewa A. Interpretation and power of a pooled index. J Rheumatol. 1993;20(3):575–8.

15. Roberts RS. Pooled outcome measures in arthritis: the pros and cons. J Rheumatol. 1993;20(3):566–7.

16. Van Onsem S, Van Der Straeten C, Arnout N, Deprez P, Van Damme G, Victor J. A new prediction model for patient satisfaction after total knee arthroplasty. J Arthroplasty. 2016;31(12):2660–7.

17. Culliton SE, Bryant DM, MacDonald SJ, Hibbert KM, Chesworth BM. Effect of an e-learning tool on expectations and satisfaction following total knee arthroplasty: a randomized controlled trial. J Arthroplasty. 2018;33(7):2153–8.

18. Noble PC, Scuderi GR, Brekke AC, Sikorskii A, Benjamin JB, Lonner JH, et al. Development of a new Knee Society scoring system. Clin Orthop Relat Res. 2012;470(1):20–32.

19. Roos EM, Toksvig-Larsen S. Knee injury and Osteoarthritis Outcome Score (KOOS) - validation and comparison to the WOMAC in total knee replacement. Health Qual Life Outcomes. 2003;1(1):17.

20. Ware JE, Krosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996;34(3):220–33.

21. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand. 1983;67(6):361–70.

22. Naal FD, Impellizzeri FM, Leunig M. Which is the best activity rating scale for patients undergoing total joint arthroplasty? Clin Orthop Relat Res. 2009;467(4):958–65.

23. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: development and validation. Psychol Assess. 1995;7(4):524–32.

24. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for statistical Computing; 2020.

25. Torgo L. Data mining with R: learning with case studies. 1st ed. Boca Raton, FL: Taylor & Francis; 2010.

26. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.

27. Thiele C. Cutpointr: determine and evaluate optimal cutpoints in binary classification tasks. R Packag version. 2019;1(1).

28. Thiele C. cutpointr: Determine and Evaluate Optimal Cutpoints in Binary Classification Tasks. R

package version 1.1.0; 2021.

7 Can Met Expectations Moderate the Relationship Between Pain/Function and Satisfaction in Total Knee Arthroplasty?

A version of this paper has been published: Munn JS, Culliton SE, Bryant DM, MacDonald SJ, Chesworth BM. Can Met Expectations Moderate the Relationship Between Pain/Function and Satisfaction in Total Knee Arthroplasty? The Journal of Arthroplasty. 2021;36(6):1942-6.

7.1 Introduction

Total knee arthroplasty (TKA) is a successful and effective treatment for end-stage osteoarthritis (OA). However, approximately 20% of patients report they are dissatisfied following surgery[1].

There has been substantial effort put into understanding what causes patient dissatisfaction and a wide variety of individual factors have been identified including: body mass index (BMI)[2], age[3], comorbidities[4], mental health[5], personality[6], locus of control[7], extended hospital stay[3], met expectations[3,8–12], pre-surgery expectations[11,13], quad strength[14], walking distance[14], joint pain[2–4,10,14,15], and joint function[2,4,5,14–16]. The most commonly found predictors of dissatisfaction are joint pain, joint function, and met expectations.

Met expectations across many fields, including health care, are frequently discussed in relation to satisfaction through the expectation confirmation theory framework[17]. Expectation confirmation theory states that satisfaction and dissatisfaction, in this case with a surgical outcome, are a result of an individual's expectations of surgery being met or unmet post-operatively. Within a medical context, a patient expectation has been defined as the anticipation that a specific event or outcome is likely to occur as a result of a medical intervention[18]. How closely the actual outcome matches the expected outcome would determine if a patient considers their expectations met or unmet[17]. Consequently, the greater the difference between the individual's expected outcome and the actual outcome, the greater that individual's dissatisfaction. This has been suggested as a model of satisfaction for TKA patients[19]. Interventions designed to modify a patient's expectations have been attempted in order to improve patient satisfaction[19,20]. In these studies it is assumed that met expectations are actually measuring how closely a patient's expectations match up with their actual condition[9]. This interpretation of a single met expectation question has been criticized. It has been argued that this is simply a different type of performance question, that may not actually be measuring how closely expectations match up with actual outcomes[21].

If expectation confirmation theory is truly correct, then met expectations may represent a subjective threshold that indicates if a desired level of pain relief or functional activity has been achieved. This could be evaluated by treating met expectations as a moderating variable and analyzing the outcome. Moderation occurs when the relationship between two variables depends on a third. In the case of TKA satisfaction, this would mean that as a met expectations score increased it would be expected that pain and function variables would less strongly predict satisfaction for patients. The purpose of this study was to determine if met expectations significantly moderated the effect of pain and function variables on satisfaction.

7.2 Methods

7.2.1 Patients

This was an observational cohort study in which patients were identified retrospectively from a randomized control trial (RCT) that evaluated the effects of an e-learning tool on patient expectations. The e-learning tool was found to have no effect on patient expectations or satisfaction[19]. As the RCT was not observed to have any effect on outcomes, it was considered acceptable to pool the treatment and control groups. Patients for the RCT were recruited from the pre-admission clinic of a large teaching hospital (University Hospital London, Ontario, Canada) over a 1-year period (April 2013 to April 2014).

To be included in the initial RCT, patients were required to have been diagnosed with knee OA for a primary TKA, have the cognitive capacity to consent to participate, and be 20 years of age or older at the time of surgery. Patients undergoing a revision, hemi or unicondylar resurfacing, patellar resurfacing, high tibial osteotomy, or surgery to address tumors were excluded[19]. Patients were asked to complete questionnaires at pre-surgery, 6-weeks post-surgery, 3-months post-surgery, and 1-year post-surgery. Pre-surgery and follow up questionnaires were completed online and stored in a secured database.

In order to be included in the analysis patients were required to have completed questionnaires at pre-surgery and 1-year post-surgery.

The questionnaires included in the study were: The Knee Society Knee Scoring System (KSS; pre-operative and post-operative versions)[22], Knee Injury and Osteoarthritis Outcome Score (KOOS)[23], the 12-Item Short Form Health Survey version 2 (SF-12)[24], the Hospital Anxiety and Depression Scale (HADS)[25], the Pain Catastrophizing Scale (PCS)[26], Self-Administered Comorbidity Questionnaire (SCQ)[27] and the University of California at Los Angeles (UCLA) Activity Score[28].

The met expectations score was measured using the KSS met expectations subscale. This subscale is comprised of three questions (pain relief, activities of daily living, and recreational or sports activities) that evaluate if a patient's expectations before TKA were met post-operatively. More specifically, patients are asked to indicate if their pre-operative expectations were too high, just right, or too low. These response options (and scoring) are structured into the following five-point ordinal scale: Too High- "I'm a lot worse than I thought" (1), Too High- "I'm somewhat worse than I thought" (2), Just Right- "My expectations were met" (3), Too Low- "I'm somewhat better than I thought" (4) and Too Low- "I'm a lot better than I thought" (5). The scores for the individual items are then summed to create a total score ranging from 3-15 (worst to best).

Satisfaction was measured using the KSS satisfaction subscale. This is a commonly used outcome measure in the TKA literature[15]. The KSS satisfaction subscale is one of the few TKA satisfaction measures with published validation information[22,29]. Although the new KSS scoring system was designed with the intention of limiting floor and ceiling effects, floor and ceiling effects have yet to be published for this subscale[22]. This subscale includes five questions that evaluate patient satisfaction with: pain while sitting, pain while lying in bed, knee function while getting out of bed, knee function while performing light household duties, and knee function while performing leisure and recreational activities. Patients were asked to rate their satisfaction state for each question using the following response options (and scoring): very dissatisfied (0), dissatisfied (2), neutral (4), satisfied (6) and very satisfied (8). The scores of the individual items were then summed into a total satisfaction score ranging from 0-40 (worst to best state).

7.2.2 Statistical Analysis

All statistical analysis was performed using R version 3.4.1[30].

Descriptive statistics were calculated for the sample and a correlation matrix and scatterplots for pairs of variables were examined.

Data were analyzed using two separate linear regression models. Total satisfaction scores on the KSS satisfaction subscale were used as the dependent variable. In order to detect a moderating effect, an interaction term, (i.e. the product of two independent variables already included in the model) was added to each model. The two interaction terms were: met expectations with pain and met expectations with function. Therefore, separate models for pain and function were created using the KSS symptoms score for pain and KOOS activities of daily living (ADL) score for function. Values that were missing at random were imputed using regression imputation with the mice R package[31]. All values were standardized before any regression modeling took place. A backwards stepwise variable selection method was employed to determine significant predictors of satisfaction. An $alpha<0.05$ was used as the threshold of significance to retain terms in the model. Demographic information including age, sex, and BMI, as well as the pre-surgery expectations variable was included in the model even if these variables did not meet the $alpha<0.05$ significance threshold. These variables were included in order to ensure that any significant interactions could not be explained by these variables being absent from the model. Multicollinearity was assessed using the variance inflation factor (VIF) with the car R package[32]. VIF scores greater than 5 were considered as problematic. Goodness of fit was assessed using adjusted R^2 and root mean square error (RMSE). A 10-fold k-fold cross-validation was performed and a mean RMSE score was generated from the average of the 10 folds with the caret R package[33]. Robust standard errors were used to perform tests of significance in order to avoid bias in the linear regression estimators from heteroscedasticity[34].

A simple slopes analysis[35] was completed using the interactions R package for any significant interaction terms with met expectations treated as the moderating variable for both interactions[36]. A simple slopes analysis calculates how the effect of pain or function on satisfaction changes based on the patient's met expectations score. In this case, the change in relationship between pain or function and satisfaction was calculated for met expectations scores at the mean and one standard deviation above and below the mean.

No sample size calculation was completed for this analysis as information about the strength of the interaction term would be required and to the best of the author's knowledge this information is not available.

7.3 Results

A sample of 416 patients were initially recruited to participate in the RCT. Of the initial sample, 50 either did not undergo TKA or were missing measures from pre-admission. Of the remaining 366 eligible patients, 304 met the inclusion criteria for this study.

Descriptive statistics for all patients' outcome scores are included in table 7.1. Less than 1% of the data in the sample was found to be missing and was therefore imputed.

Table 7.1. Means and standard deviations of pre-surgery and 1-year post-surgery variables included in the data set.

Variable	Pre-Surgery N=304 Mean/Frequency	SD	1-Year Post-Surgery N=304 Mean	SD
Age	64.63	8.55		
Sex (Female)*	198 (65%)			
Body Mass Index	32.87	7.46		
Comorbidities	2.87	1.66		
KSS Symptoms	9.92	3.66	19.37	4.93
KSS Patient Satisfaction	13.01	6.70	31.81	8.46
KSS Patient Expectations	13.44	1.90	9.17	2.77
KSS Walking Standing	4.71	4.15	21.30	8.36
KSS Standard Activities	11.52	4.83	23.13	5.09
KSS Advanced Activities	4.46	3.66	11.34	5.84
KSS Discretionary Activities	5.19	3.01	11.25	3.09
KOOS Pain	40.56	16.61	81.43	17.27
KOOS Symptoms	43.18	20.38	75.45	16.97
KOOS Stiffness	37.90	19.00	75.24	19.68
KOOS Activities of Daily Living	45.72	17.70	82.70	16.55
KOOS Functional Sport and Recreation	18.47	25.11	51.90	29.11
KOOS Quality of Life	19.82	15.53	64.88	21.65
SF-12 Physical Function	28.27	8.04	40.9	10.84
SF-12 Role Physical	32.43	9.11	45.04	9.78
SF-12 Bodily Pain	28.75	8.8	46.92	10.69
SF-12 General Health	49.07	9.22	50.28	9.66
SF-12 Vitality	46.61	10.1	52.11	9.59
SF-12 Social Function	44.08	11.7	21.12	8.56
SF-12 Role Emotional	71.57	27.24	87.09	19.52
SF-12 Mental Health	84.43	20.48	92.97	17.88

*The variable is represented as a frequency and not a mean with percentage of the sample included in brackets; SD, Standard Deviation; KSS, Knee Society Knee Scoring System; KOOS, Knee Injury and Osteoarthritis Outcome Score; SF-12, 12-Item Short Form Health Survey version 2

The regression models evaluating the interactions of KSS symptoms (pain) with KSS met expectations and of KOOS ADL with KSS met expectations can be found in table 7.2. Significant non interaction predictors of satisfaction in the models were: KSS symptoms (pain), KSS met expectations, pre-surgery KOOS ADL, KOOS ADL (function), BMI, and SF-12 general health. A significant interaction between KSS symptoms (pain) and KSS met patient expectations was detected with a p-value of 0.043. The interaction between KOOS ADL (function) and KSS met expectations was not found to be significant, but approached significance with a p-value of 0.086. When testing assumptions in the model, heteroscedasticity was observed. Therefore, robust standard errors were used for all individual item tests of significance in the analysis. VIF scores for both models indicated very little collinearity (table 7.3.).

Table 7.2. Output for the linear regression models including each interaction.

Predictors	Pain interaction			Function Interaction		
	Estimates	Confidence Interval	P-Value	Estimates	Confidence Interval	P-Value
Intercept	31.91	31.40 – 32.52		31.92	31.35 – 32.49	
Age	0.34	0.06 – 0.87	0.154	0.34	-0.13 – 0.81	0.161
Sex	-0.09	-0.55 – 0.37	0.707	-0.09	-0.55 – 0.37	0.703
Body Mass Index	0.59	0.09 – 1.10	**0.024**	0.59	0.08 – 1.11	**0.025**
KSS Pre-Surgery Patient Expectations	-0.31	-0.76 – 0.14	0.174	-0.32	-0.78 – 0.14	0.172
KSS 1-Year Pain (Symptoms)	3.48	2.53 – 4.42	**<0.001**	3.73	2.83 – 4.64	**<0.001**
KSS 1-Year Met Patient Expectations	1.51	0.96-2.06	**<0.001**	1.51	0.94 – 2.08	**<0.001**
KOOS Pre-Surgery Activities of Daily Living	-0.74	-1.23 – -0.24	**0.003**	-0.77	-1.27 – -0.28	**0.002**
KOOS 1-Year Activities of Daily Living	3.03	2.17 – 3.90	**<0.001**	2.79	1.79 – 3.79	**<0.001**
SF-12 1-Year General Health	0.74	0.22 – 1.26	**0.005**	0.76	0.24 – 1.29	**0.005**
Interactions						
KSS 1-Year Met Patient Expectations * KSS 1-Year Symptoms	-0.53	-1.04 – -0.02	**0.043**			
KSS 1-Year Met Patient Expectations * KOOS 1-Year Activities of Daily Living				-0.49	-1.04 – 0.07	0.086
Observations		304			304	
Adjusted R^2		0.786			0.784	
RMSE		3.98			4.01	

RMSE, root mean square error; KSS, Knee Society Knee Scoring System; KOOS, Knee Injury and Osteoarthritis Outcome Score; SF-12, 12-Item Short Form Health Survey version 2

Table 7.3. VIF values for the pain and function interaction models.

Predictors	Pain Interaction Model	Function Interaction Model
Age	1.20	1.20
Sex	1.29	1.29
Body Mass Index	1.08	1.09
KSS Pre-Surgery Patient Expectations	1.12	1.13
KSS 1-Year Pain (Symptoms)	3.40	3.08
KSS 1-Year Met Patient Expectations	1.74	1.77
KOOS Pre-Surgery Activities of Daily Living	1.43	1.42
KOOS 1-Year Activities of Daily Living	3.43	4.29
SF-12 1-Year General Health	1.30	1.31
KSS 1-Year Met Patient Expectations * KSS 1-Year Symptoms	1.55	
KSS 1-Year Met Patient Expectations * KOOS 1-Year Activities of Daily Living		1.80

KSS, Knee Society Knee Scoring System; KOOS, Knee Injury and Osteoarthritis Outcome Score; SF-12, 12-Item Short Form Health Survey version 2

The interaction between KSS symptoms (pain) and KSS met expectations was found to be in the negative direction. This means that as KSS met expectations increase, the beta estimate for KSS symptoms (pain) decreases. This is demonstrated in the simple slopes analysis included in table 7.4. For a met expectations score one standard deviation below the mean, the beta estimate for the KSS symptoms (pain) variable is 4.01. For a met expectations score at the mean, the beta estimate is 3.48. Lastly, for a met expectations score one standard deviation above the mean, the beta estimate is 2.95.

Table 7.4. Simple slopes analysis for the symptoms (KSS) and 1-year met expectations (KSS) interaction variable in the regression model in table 7.2.

KSS 1-Year Symptoms Slope Estimate	KSS 1-Year Met Patient Expectations Standard Deviations	Standard Error	T-Value	P-Value
4.01	-1	0.40	9.81	**<0.001**
3.48	0	0.41	8.27	**<0.001**
2.95	1	0.52	5.48	**<0.001**

KSS, Knee Society Knee Scoring System

Although not found to be significant the KOOS ADL and KSS met expectations interaction was also found to be in the negative direction. As this interaction was not found to be significant a simple slopes analysis was not completed.

7.4 Discussion

Our study has demonstrated that met expectations can act as a moderator for the relationship between pain and satisfaction. Met expectations significantly affect how strongly a pain variable predicts patient satisfaction. The KOOS ADL (function) interaction was not found to be significant, but approached significance with a p-value of 0.086. The interaction means that the magnitude of the relationship between pain and satisfaction is conditional upon the met expectations score. For a met expectations score that is lower, say one standard deviation below the mean, the slope for the relationship between pain and satisfaction is 4.01. This means that for a one-unit improvement in pain the satisfaction score will be predicted to increase 4.01 units. For a met expectations score that is higher, say one standard deviation above the mean, the slope for the relationship between pain and satisfaction is 2.95. This means that for a one-unit improvement in pain the satisfaction score will be predicted to increase by 2.95 units. In other words, as met expectations scores increase, further improvements in pain become less important for improving satisfaction.

In terms of clinical application, this model appears to be capturing that patients who have met or exceeded their expected level of pain relief, do not value pain relief as much as patients who have not achieved their expected level of pain relief. If this is correct, then this would indicate that expectations being met represents a subjective threshold for pain scores that the patient would like to achieve. We wonder if this change in importance of pain relief is what surgeons see clinically, which would reinforce the importance to patients and surgeons of realistic expectations.

It is unclear if other variables become more important to patient satisfaction as met expectations increase and pain relief becomes less important. Although only approaching significance, the function interaction is in the same negative direction as the pain interaction. This indicates that it is unlikely that function is valued more by patients as met expectations increase and pain relief becomes less important. However, the function variable that was used was the KOOS ADL subscale. It is possible that a different relationship could be observed with met expectations and recreational activities, although that was not assessed in this study.

The interpretation of the model that has been presented is predicated on the met expectation variable actually representing a met expectation. If the met expectation variable is not actually measuring met expectations, but some other subjective measure of performance, this does not change the interpretation of the interaction term. This remains a subjective evaluation of performance and met expectations appear to be capturing some important performance threshold.

This analysis made use of standardized variables in the regression analyses in order to test for interactions. Standardized variables have not typically been used in the TKA literature. This approach has several advantages. Testing for interactions using standardized variables avoids high levels of multicollinearity. A second benefit of this approach is it facilitates comparison of variables in the model because all variables are on the same scale.

There were a number of limitations in this paper. The comorbidity measure used in this analysis was crude and the study did not examine individual comorbidities. A subset of the patients were also part of an intervention to adjust patient expectations. However, the original analysis showed there was no significant treatment effect[19]. The sample size of this study was only 304 patients and this was potentially not large enough to detect all predictors. Some of the patients in this study were younger, potentially affecting satisfaction rate. The study did not collect data on patients past 1-year post-operative. Future satisfaction studies should examine longer time horizons. Additionally, despite low VIF scores there is still the possibility of collinearity in the interaction models. The patient satisfaction variable for TKA was skewed, with more patients at the higher, or more satisfied, end of the scale. In this instance, using the skewed satisfaction variable was considered preferable to converting patient satisfaction to a categorical variable and thereby losing information.

In conclusion met expectations have been found to moderate the relationship between pain and satisfaction. This finding indicates a more complex relationship between satisfaction, met expectations and pain variables. This relationship needs to be examined in further detail in order to better understand the role of met expectations as a moderating variable. What a met expectation score is actually measuring must also be better understood in order to adequately study the relationship between met expectations and satisfaction.

7.5 References

1. Gunaratne R, Pratt DN, Banda J, Fick DP, Khan RJK, Robertson BW. Patient dissatisfaction following total knee arthroplasty: a systematic review of the literaturere. J Arthroplasty. 2017;32(12):3854–60.
2. Kim TK, Chang CB, Kang YG, Kim SJ, Seong SC. Causes and predictors of patient's dissatisfaction after uncomplicated total knee arthroplasty. J Arthroplasty. 2009;24(2):263–71.
3. Bourne RB, Chesworth BM, Davis AM, Mahomed NN, Charron KDJ, Met D. Patient satisfaction after total knee arthroplasty who is satisfied and who is not? Clin Orthop Relat Res. 2010;468(1):57–63.
4. Clement ND, Bardgett M, Weir D, Holland J, Gerrand C, Deehan DJ. The rate and predictors of patient satisfaction after total knee arthroplasty are influenced by the focus of the question: a standard satisfaction question is required. Bone Joint J. 2018;100(6):740–8.
5. Scott CEH, Howie CR, Macdonald D, Biant LC. Predicting dissatisfaction following total knee replacement: a prospective study of 1217 patients. J Bone Joint Surg Br. 2010;92(9):1253–8.
6. Choi Y, Ra HJ. Patient Satisfaction after Total Knee Arthroplasty. Knee Surg Relat Res. 2016;28(1):1–15.
7. Dhurve K, Scholes C, El-tawil S, Shaikh A, Weng LK, Levin K, et al. Multifactorial analysis of dissatisfaction after primary total knee replacement. Knee. 2017;24(4):856–62.
8. Mannion AF, Kämpfen S, Munzinger U, Quervain IK. The role of patient expectations in predicting outcome after total knee arthroplasty. Arthritis Res Ther. 2009;11(5):1–13.
9. Barlow T, Clark T, Dunbar M, Metcalfe A, Griffin D. The effect of expectation on satisfaction in total knee replacements: a systematic review. Springerplus. 2016;5(1):1–7.
10. Vissers MM, Groot IB De, Reijman M, Bussmann JB, Stam HJ, Verhaar JAN. Functional capacity and actual daily activity do not contribute to patient satisfaction after total knee arthroplasty. BMC Musculoskelet Disord. 2010;11(1):1–8.
11. Becker R, Döring C, Denecke A, Brosz M. Expectation, satisfaction and clinical outcome of patients after total knee arthroplasty. Knee surgery, Sport Traumatol Arthrosc. 2011;19(9):1433–41.
12. Thambiah MD, Nathan S, Seow BZX, Liang S, Lingaraj K. Patient satisfaction after total knee arthroplasty: an Asian perspective. Singapore Med J. 2015;56(5):259–63.
13. Kiran A, Bottomley N, Biant LC, Tr F, Javaid MK, Carr AJ, et al. Variations in good patient reported outcomes after total knee arthroplasty. J Arthroplasty. 2015;30(8):1364–71.

14. Van Onsem S, Verstraete M, Dhont S, Zwaenepoel B, Van Der Straeten C. Improved walking distance and range of motion predict patient satisfaction after TKA. Knee Surgery, Sport Traumatol Arthrosc. 2018;26(11):3272–9.

15. Kahlenberg CA, Nwachukwu BU, Mclawhorn AS, Cross MB, Cornell CN, Padgett DE. Patient satisfaction after total knee replacement: a systematic review. HSS J. 2018;14(2):192–201.

16. Baker PN, Rushton S, Jameson SS, Reed M, Gregg P, Deehan DJ. Patient satisfaction with total knee replacement cannot be predicted from pre-operative variables alone: a cohort study from the National Joint Registry for England and Wales. Bone Joint J. 2013;95(10):1359–65.

17. Oliver RL. Satisfaction: A behavioural perspective on the consumer. Second Edi. New York, NY: M.E. Sharpe; 2010. 61–89 p.

18. Bowling A, Rowe G, Lambert N, Waddington M, Mahtani KR, Kenten C, et al. The measurement of patients' expectations for health care: a review and psychometric testing of a measure of patients' expectations. Health Technol Assess (Rockv). 2012;16(30).

19. Culliton SE, Bryant DM, MacDonald SJ, Hibbert KM, Chesworth BM. Effect of an e-learning tool on expectations and satisfaction following total knee arthroplasty: a randomized controlled trial. J Arthroplasty. 2018;33(7):2153–8.

20. Gautreau S, Gould ON, Aquino-Russell C, Forsythe ME. Developing a surgeon-patient communication checklist for total knee arthroplasty. Musculoskeletal Care. 2019;17(1):91–6.

21. Edwards JR. Ten difference score myths. Organ Res methods. 2001;4(3):265–87.

22. Noble PC, Scuderi GR, Brekke AC, Sikorskii A, Benjamin JB, Lonner JH, et al. Development of a new Knee Society scoring system. Clin Orthop Relat Res. 2012;470(1):20–32.

23. Roos EM, Toksvig-Larsen S. Knee injury and Osteoarthritis Outcome Score (KOOS) - validation and comparison to the WOMAC in total knee replacement. Health Qual Life Outcomes. 2003;1(1):17.

24. Ware JE, Krosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996;34(3):220–33.

25. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand. 1983;67(6):361–70.

26. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: development and validation. Psychol Assess. 1995;7(4):524–32.

27. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self- Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. Arthritis Care Res Off J Am Coll Rheumatol. 2003;49(2):156–63.

28. Naal FD, Impellizzeri FM, Leunig M. Which is the best activity rating scale for patients undergoing total joint arthroplasty? Clin Orthop Relat Res. 2009;467(4):958–65.
29. Culliton SE, Bryant DM, MacDonald SJ, Hibbert KM, Chesworth BM. Validity and internal consistency of the new Knee Society Knee Scoring System. Clin Orth Relat Res. 2018;476(1):77–84.
30. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for statistical Computing; 2020.
31. Buuren S Van, Groothuis-Oudshoorn K. MICE: multivariate imputation by chained equations in r. J Stat Softw. 2010;1–68.
32. Fox J, Weisberg S. An R Companion to Applied Regression. Third. Thousand Oaks CA: Sage; 2019.
33. Kuhn M. caret: Classification and Regression Training. R package version 6.0-88; 2021.
34. Imbens GW, Kolesár M. Robust standard errors in small samples: some practical advice. Rev Econ Stat. 2016;98(4):701–12.
35. Preacher KJ, Curran PJ, Bauer DJ. Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. J Educ Behav Stat. 2006;31(4):437–48.
36. Long JA. interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions. R package version 1.1.0; 2019.

8 Discussion

8.1 Introduction

Dissatisfaction following total knee arthroplasty (TKA) remains a persistent problem. This dissertation aimed to investigate strategies to improve TKA dissatisfaction. Of the four studies completed, three investigated the pre-surgery and early post-surgery identification of dissatisfied patients. The fourth study attempted to better understand the met expectations variable. These studies were informed by the conceptual framework of TKA satisfaction outlined in the introduction (section 1.4).

8.2 Summary of Study Results

8.2.1 Study 1 (Chapter 4)

The first study (chapter 4) showed that logistic regression and machine learning models cannot discriminate between satisfied and dissatisfied patients in a clinically useful manner when using only pre-surgery and surgical explanatory variables. This study (chapter 4) is the first to include patient reported outcome measures (PROMs) in machine learning prediction models. Additionally, no satisfaction prediction models that make use of machine learning have been created using Canadian data. Despite acceptable area under the curve (AUC) values, when maximizing accuracy, the models only marginally outperformed the no information rate. However, the models' calibration performance was more promising and could be an avenue for exploration. In order to improve model performance and fully leverage the value of machine learning models, a wider array of explanatory variables and analytical techniques need to be investigated. It should also be acknowledged that there may be no reasonable way to accurately predict 1-year post-surgery dissatisfaction using pre-surgery variables alone.

8.2.2 Study 2 (Chapter 5)

The second study (chapter 5) demonstrated that TKA satisfaction can be predicted using pre-surgery and 3-month post-surgery explanatory variables with both logistic regression and classification trees. This means that it is feasible to identify patients early in the recovery process who are likely to be dissatisfied at 1-year post-surgery. This could allow for the implementation of targeted interventions. The classification tree has advantages in interpretability and simplicity over the logistic regression model and would be easier to transport to a clinical setting. In both models, the activities of daily living (ADL) Knee

Injury and Osteoarthritis Outcome Score (KOOS) variable was found to be an extremely important predictor of dissatisfaction.

8.2.3 Study 3 (Chapter 6)

The third study (chapter 6) showed that a pooled index using five individual questionnaire items from the 3-month post-surgery KOOS and Knee Society Knee Scoring System (KSS) questionnaires can predict 1-year post-surgery dissatisfaction. The five items included in this index were summed together to create an overall score that showed a strong ability to discriminate between satisfied and dissatisfied patients across a variety of cut points. The discriminatory capacity of the pooled index, paired with the simplicity of administration of this prediction model, means that it would be extremely easy to transport to a clinical setting.

8.2.4 Study 4 (Chapter 7)

The fourth study (chapter 7) demonstrated that met expectations moderated the relationship between pain and satisfaction. As met expectations scores increased, pain was found to become less important for improving satisfaction. The met expectations measure appears to be an important subjective threshold of performance for patients. Interventions geared towards unmet expectations may be effective in improving satisfaction.

8.3 Satisfaction Improvement Methods

Outside of surgical improvements, there are two possible ways to reduce the number of dissatisfied TKA patients. The first is by identifying dissatisfied patients pre-surgery. If patients can be identified pre-surgery, then surgeons can inform patients about their risk of dissatisfaction and a more informed decision can be made about proceeding with surgery. The second strategy is to implement interventions to improve satisfaction. Identifying dissatisfied patients post-surgery as targets for intervention during recovery, could help improve the efficacy of those interventions and by association reduce TKA dissatisfaction at 1-year postoperative.

8.4 Predicting Dissatisfaction

8.4.1 Assessing the Performance of a TKA Clinical Prediction Model

When determining if a model is clinically useful, there are several factors that must be considered: the time point of the prediction (pre-surgery/post-surgery), the intended use of the model, and the decision makers' priorities[1,2]. These factors are not necessarily statistical questions and fall outside of traditional quantitative prediction model evaluation metrics. However, these factors will influence the assessment of the prediction model, as they will inform how well the model needs to perform quantitatively to be considered clinically useful.

8.4.1.1 Time Point of the Prediction (Pre-Surgery/Post-Surgery)

The time point at which the prediction is made will influence the assessment of model clinical utility. A model that is used pre-surgery will be employed to help decide if a patient should proceed with surgery or not. A model that is identifying patients early in the recovery process will be used to help determine if a patient should receive a post-surgery intervention or not. The consequences of the decision that the model is intended to make or to inform will influence the way in which the model is used and how model performance is evaluated.

8.4.1.2 Intended Use

A model can be employed as a decision tool that is meant to discriminate between satisfied and dissatisfied patients, or it can be used to provide probabilistic information that a surgeon and patient will then factor into their decision about surgery. Intended model use will determine what metrics are prioritized when evaluating model performance. In the TKA literature, models have typically focused on discriminatory performance; however, it has been argued by Harrell[1] that there is significant value in using clinical prediction models in a probabilistic fashion. For example, a model produces a probability of dissatisfaction for a given patient. Then that probability is factored into the surgeon and patient's decision about whether or not to proceed with surgery.

8.4.1.2.1 Discriminatory Approach

Medical prediction models are usually used in a discriminatory capacity[2]. This is advantageous as the model provides a clear decision about what action should be taken. However, establishing a decision threshold to determine if a patient is satisfied or dissatisfied is not necessarily straightforward. The costs of the decision and priorities of the decision maker must be factored in when determining how a prediction model will be used to make a decision[2].

8.4.1.2.1.1 Decision Maker Priorities

When evaluating discriminatory performance of a prediction model, it is not as straightforward as evaluating accuracy, sensitivity, or specificity. The outcome that is prioritized by the decision maker must also be considered. For example, a surgeon could consider a false positive (incorrectly predicting that a satisfied patient will be dissatisfied) as more consequential than a false negative (incorrectly predicting that a dissatisfied patient will be satisfied). This weighting of prediction outcomes can be quantified by associating a cost with true positive, false positive, true negative, and false negative outcomes. Costs can include economic costs, the harm caused by providing a treatment that does not benefit the patient, or the harm caused by not providing treatment to a patient who will benefit.

Within the context of dissatisfaction, calculating or quantifying these costs can be difficult. It is not always clear that a patient who is dissatisfied has been harmed or has not received a benefit from undergoing TKA. This is supported by Bourne[3] who found that 62.4% of dissatisfied TKA patients would undergo surgery again, with 18% unsure, and 19.6% stating they would not undergo surgery again. It would seem unlikely that dissatisfied patients who have not received some benefit from surgery would want to undergo surgery again. This ambiguity can make assessing model usefulness particularly complex for dissatisfaction prediction models.

8.4.1.2.2 Probabilistic Approach

Using the probabilities generated by prediction models to inform patients and surgeons about the probability that a patient will be dissatisfied is an alternative approach to using models in a discriminatory capacity. Although it does not provide a clear decision about surgery, there are several advantages to using prediction models probabilistically[1]. This allows decisions to be tailored to patients instead of establishing thresholds that are applied universally. For example, more active and less active patients may have different costs associated with their surgery. A final advantage of a probabilistic approach is that this information could function as its own intervention by providing patients with a realistic understanding of their postoperative chances of dissatisfaction before undergoing surgery.

8.4.2 Thesis Prediction Models' Clinical Utility

8.4.2.1 Pre-surgery Prediction Model Performance (Study 1 Chapter 4)

The first study (chapter 4) in this thesis aimed to create prediction models that could identify dissatisfied patients before surgery has taken place. Despite the current ambiguity surrounding the necessary performance of a prediction model, clinically useful discrimination performance in study 1 (Chapter 4) was defined as outperforming the no information rate, as well as conveying a high degree of certainty that a patient predicted as dissatisfied will be dissatisfied. No clear performance thresholds were established for calibration metrics.

Study 1 (chapter 4) found pre-surgery prediction models only marginally outperformed the no information rate and failed to convey a high degree of certainty that a dissatisfied patient will be dissatisfied. Additionally, the AUC scores calculated for the models created by this study perform slightly worse than what has been reported by Kunze[4] and Farooq[5].

The models in study 1 (chapter 4) did show some ability to generate accurate probabilities that a patient will be dissatisfied. Despite the more promising calibration scores, they were still not optimal and would most likely not be useful in helping patients and surgeons make decisions about whether or not to undergo surgery.

8.4.2.2 Early Post-Surgery Prediction Model Performance (Study 2 and 3 Chapters 5 and 6)

If patients cannot be identified pre-surgery, then satisfaction must be improved through post-surgery interventions. At present, interventions aimed to improve TKA dissatisfaction have had mixed results (section 2.4). An important part of improving the performance of these interventions could be the identification of dissatisfied patients early in the recovery process.

The second (chapter 5) and third (chapter 6) studies in this thesis aimed to create prediction models that could identify dissatisfied patients early in the recovery process (e.g. at 3-months postoperative). To evaluate clinical utility of these models and to assess the costs of a decision, the details of a specific intervention would be required. However, regardless of the intervention, it is likely that these models could be considered clinically useful if they show strong discriminatory capacity. As these models will be deciding if a patient should receive an intervention and are occurring post-surgery, it was considered less harmful if predictions were incorrect. This was because there is no scenario where a model can influence a decision about surgery and potentially persuade a patient who would benefit not to undergo surgery.

Simplicity and ease of administration are an additional consideration for clinical usefulness of early post-surgery prediction models. Because the decision is being made post-surgery, it could be considered less consequential than the decision made by a pre-surgery prediction model. It was thought that more complex and expensive-to-implement models would be far less likely to be used.

The second study (chapter 5) in this thesis demonstrated strong discriminatory capacity particularly for the logistic regression model. Patients dissatisfied at 1-year post-surgery can accurately be identified at 3-months post-surgery. The third study (chapter 6) in this thesis expanded upon that finding by creating a simple to implement pooled index that included five items from commonly used patient reported outcome measures. This model showed strong discriminatory capacity across a wide variety of cut points and appeared as if it could be clinically useful. The pooled index model is likely preferrable to the logistic regression and classification tree models as it performed similarly to these models while being substantially simpler than the logistic regression model and much more flexible than the classification tree. To implement the logistic regression model in a clinical setting, patients would be required to fill out multiple questionnaires and a computer would be needed generate a prediction. The classification tree can be scored by hand, but the thresholds for this model are fixed. This means that the priority of the decision maker cannot be factored in while using this model. The pooled index addresses these issues by being simple enough to score with a pen and paper, while also allowing priorities of the decision maker to be factored in, as the cut point of the index can be easily adjusted. The models created in studies 2 and 3 (chapters 5 and 6) had substantially higher AUC scores than the univariate logistic regression models reported by Goh[6] which used 6-month post-surgery data to predict dissatisfaction at 2-years post-surgery.

Not only could the early identification of dissatisfied patients allow for these individuals to be targeted for specific interventions, but these models could also provide information on functional milestones that are likely necessary for long term satisfaction. For example, if patients cannot adequately perform ADL at 3-months post-surgery, they are unlikely to be satisfied at 1-year post-surgery. Identifying and targeting patients who will be dissatisfied early in the recovery process is a feasible undertaking.

8.5 Met Expectations and Post-Surgery Interventions

Studies that have implemented met expectations interventions to improve satisfaction have had mixed results[7–9]. These differences can partly be attributed to mode of delivery and content of the intervention. However, these inconsistencies also highlight that the met expectations variable may not be that well understood[10,11]. A definition of met expectations is rarely provided in the literature. The fourth study (chapter 7) in this thesis aimed to examine met expectations in greater detail in order to better understand how met expectations relate to pain, function, and satisfaction. This study found that met expectations represent an important subjective threshold for patients. Although further investigation of the met expectations variable is required, this may indicate that dissatisfaction interventions should be directed towards areas where patient expectations are unmet and may be modifiable.

8.6 Implications for Clinicians and Researchers

Implications for both researchers and surgeons can be drawn from the studies included in this dissertation. The findings from study 1 suggest that pre-surgery TKA dissatisfaction cannot be predicted using explanatory variables that can be considered as a standard part of a pre-surgery assessment. Study 2 and study 3 indicate that not only are post-surgery detection models able to discriminate well, but they can be easily transported to a clinical setting and implemented using only a pen and paper. Additionally, both of these studies identified functional milestones that a patient will likely need to meet at 3-months post-surgery in order to be satisfied at 1-year post-surgery. Lastly, study 4 found that unmet expectations represent important subjective thresholds of performance. Targeting unmet expectations may be a fruitful avenue for intervention, either through educational programs or physiotherapy.

8.7 Limitations

The studies in this dissertation had several limitations. First, all studies in this thesis were secondary data analyses. However, all data were prospectively collected and dissatisfaction rates were similar to those reported in the literature. Second, all studies in this analysis examined satisfaction at 1-year post-surgery and it has been suggested that 2-years should be the minimum follow up time. However, a 1-year post-surgery follow up time is frequently used in the literature and Ramkumar[12] found that there were no significant differences between 1 and 2-year post-surgery outcomes for arthroplasty patients.

Third, sample size may also have been problematic for all of the studies. Although study 1 used a fairly large sample size to predict 1-year post-surgery dissatisfaction using pre-surgery variables, this sample may still be too small to maximize the advantages of machine learning models[13]. For studies 2, 3, and 4, small sample sizes were used, and this work likely needs to be replicated with a larger sample to have confidence in the predictors identified and models created. In study 2, sample size was particularly problematic as the number of dissatisfied patients artificially constrained the number of predictors that could be included in the model, affecting the conclusions that can be drawn[14]. Lastly, for study 4, the KSS continuous satisfaction scale was used. This scale suffers from ceiling effects which can have an affect on the interpretation of the models. However, as this study was geared to better understanding the relationship between met expectations and satisfaction, it was considered more desirable not to lose information by dichotomizing the satisfaction variable.

8.8 Recommendations for Future Directions

Further investigation of untested explanatory variables and new analytical techniques is required to improve the ability to detect dissatisfied patients either pre-surgery or early in the recovery process. Additionally, the met expectations variable and its relationship with satisfaction is still not well understood and must be further investigated.

There are a number of analytical techniques and pre-surgery variables that remain uninvestigated in the TKA dissatisfaction literature. This is particularly true when employing machine learning methods. Natural language processing (NLP) is a type of machine learning that can process text and be used to predict an outcome. NLP has previously been used in conjunction with clinical notes in a number of different scenarios to predict health outcomes, including the automated detection of periprosthetic infections following total joint arthroplasty[15]. Image recognition models using radiographic data could also be employed to determine if dissatisfaction can be predicted using radiographic information. This has been used to automate the detection of loosening implants for TKA patients[16].

There are a number of PROMs that have not been sufficiently investigated that could potentially improve the performance of prediction models. This could include determining what a patient would define as a successful outcome for their surgery or the identification of activities that the patient would consider important or that they are particularly motivated to perform. For example, a modified KSS discretionary activities score could also include a question that evaluates the importance of the activity.

Additionally, psychological predictors have not been investigated in great detail in the TKA literature including personality, locus of control, social support, and financial well-being[17–19].

A discussion to determine how well a model must perform to be considered clinically useful is also required to determine how successful these models are. In association with this evaluation, establishing the costs associated with accurately identifying and misidentifying dissatisfied patients is required to assess model performance more concretely.

8.9 Conclusions

This thesis determined that neither logistic regression nor machine learning methods can discriminate between satisfied and dissatisfied patients using commonly measured pre-surgery and surgical variables. However, patients likely to be dissatisfied at 1-year post-surgery can be identified at 3-months post-surgery and this can be done using a prediction tool that is simple enough to administer with a pen and paper. Lastly, this dissertation found that met expectations represent an important subjective performance threshold for patients and it may be reasonable to target unmet expectations to improve satisfaction.

8.10 References

1. Harrell F. Regression modeling strategies. Second. New York: Springer Science+Business Media; 2015.
2. Steyerberg EW. Clinical prediction models. Second. Cham: Springer International Publishing; 2019.
3. Bourne RB, Chesworth BM, Davis AM, Mahomed NN, Charron KDJ, Met D. Patient satisfaction after total knee arthroplasty who is satisfied and who is not? Clin Orthop Relat Res. 2010;468(1):57–63.
4. Kunze KN, Polce EM, Sadauskas AJ, Levine BR. Development of machine learning algorithms to predict patient dissatisfaction after primary total knee arthroplasty. J Arthroplasty. 2020;35(11):3117–22.
5. Farooq H, Deckard ER, Ziemba-Davis M, Madsen A, Meneghini RM. Predictors of patient satisfaction following primary total knee arthroplasty: results from a traditional statistical model and a machine learning algorithm. J Arthroplasty. 2020;35(11):3123–30.
6. Goh GS, Bin Abd Razak HR, Tay DKJ, Lo NN, Yeo SJ. Early post-operative oxford knee score and knee society score predict patient satisfaction 2 years after total knee arthroplasty. Arch Orthop Trauma Surg. 2021;141(1):129–37.
7. Culliton SE, Bryant DM, MacDonald SJ, Hibbert KM, Chesworth BM. Effect of an e-learning tool on expectations and satisfaction following total knee arthroplasty: a randomized controlled trial. J Arthroplasty. 2018;33(7):2153–8.
8. Tolk JJ, Janssen RPA, Haanstra TM, van der Steen MC, Bierma-Zeinstra SMA, Reijman M. The influence of expectation modification in knee arthroplasty on satisfaction of patients: a randomized controlled trial. Bone Joint J. 2021;103(4):619–26.
9. Gautreau SJ, Gould ON, Allanach WW, Clark AE, Massoeurs SJ, Forsythe ME. Total knee arthroplasty communication checklist increases patient satisfaction. J Arthroplasty. 2019;34(3):456–61.
10. Hafkamp FJ, Gosens T, de Vries J, den Oudsten BL. Do dissatisfied patients have unrealistic expectations? A systematic review and best-evidence synthesis in knee and hip arthroplasty patients. EFORT Open Rev. 2020;5(4):226–40.
11. Edwards JR. Ten difference score myths. Organ Res methods. 2001;4(3):265–87.
12. Ramkumar PN, Navarro SM, Haeberle HS, Ng M, Piuzzi NS, Spindler KP. No difference in outcomes 12 and 24 months after lower extremity total joint arthroplasty: a systematic review

and meta-analysis. J Arthroplasty. 2018;33(7):2322–9.

13. Kuhn M, Johnson K. Applied predictive modeling. New York: Springer Science+Business Media; 2013.

14. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49(12):1373–9.

15. Fu S, Wyles CC, Osmon DR, Carvour ML, Sagheb E, Ramazanian T, et al. Automated detection of periprosthetic joint infections and data elements using natural language processing. J Arthroplasty. 2021;36(2):688–92.

16. Shah RF, Bini SA, Martinez AM, Pedoia V, Vail TP. Incremental inputs improve the automated detection of implant loosening using machine-learning algorithms. Bone Jt J. 2020;102(6):101–6.

17. Gong L, Dong J-Y. Patient's personality predicts recovery after total knee arthroplasty: a retrospective study. J Orthop Sci. 2014;19(2):263–9.

18. Dhurve K, Scholes C, El-tawil S, Shaikh A, Weng LK, Levin K, et al. Multifactorial analysis of dissatisfaction after primary total knee replacement. Knee. 2017;24(4):856–62.

19. Bierke S, Petersen W. Influence of anxiety and pain catastrophizing on the course of pain within the first year after uncomplicated total knee replacement: a prospective study. Arch Orthop Trauma Surg. 2017;137(12):1735–42.

Appendices

Appendix A – Post Hoc Sample Size Considerations

The sample size calculations used in this appendix require information about the sample that was not available before completing the analysis. These calculations were performed to determine if samples used in this thesis were appropriate.

Study 1

Necessary sample size for the logistic regression model was calculated as 1194 using Riley[1] method. For the machine learning models, no explicit sample size calculation was completed. There generally are not clear discrete methods for calculating sample size for machine learning methods. Necessary sample size is typically considered to be problem specific. Ogink[2] found that a median sample of 625.5 was used to predict patient reported outcome measures. The sample size of 1450 was considered as likely to be sufficient.

Study 2

Necessary sample size for the pre-surgery logistic regression model was calculated as 478. Necessary sample size for the pre-surgery and 3-month post-surgery logistic regression model was calculated as 311. Logistic regression sample sizes were calculated using Riley[1] method. Necessary sample size was not calculated for the classification tree model. A heuristic requiring at least 20 observations per node was used to prevent overfitting. The sample size of 275 was considered as slightly smaller than what would be considered ideal.

Study 3

Sample size requirement information was not calculated for the pooled index as information about necessary sample size for evaluating the predictive quality of a pooled index is not available.

Study 4

Sample size required to determine if the interaction can be detected with a power of 0.8 was calculated as 455 using Shieh's[3] method. Using a sample size of 304, this study had a power of 0.64 to detect the reported interaction.

References

1. Riley RD, Ensor J, Snell KI, Harrell FE, Martin GP, Reitsma JB, Moons KG, Collins G, Van Smeden M. Calculating the sample size required for developing a clinical prediction model. Bmj. 2020;368.
2. Ogink PT, Groot OQ, Karhade AV, Bongers ME, Oner FC, Verlaan JJ, Schwab JH. Wide range of applications for machine-learning prediction models in orthopedic surgical outcome: a systematic review. Acta Orthopaedica. 2021:1-6.
3. Shieh G. Detecting interaction effects in moderated multiple regression with continuous variables power and sample size considerations. Organizational Research Methods. 2009;12(3):510-28.

Appendix B – Data Quality Considerations

Two data sets were used in this thesis. Data sets made use of well validated questionnaires. Data checks were included to prevent incorrect response options from being recorded. Study 1 was paper based collection. During transcription, electronic data checks were included to prevent incorrect response options from being recorded. Study 2, 3, and 4 data collection was electronic. Patients could not input a response option outside of the correct range of responses. All questionnaire scores were calculated using algorithms (i.e. not by hand) to avoid errors.

Sex, age, and body mass index (BMI) were compared between data sets. Additionally, respondents and non respondents within each data set were compared to determine if there were differences between patients. Lastly, the sample was compared to data taken from the Canadian Joint Replacement Registry[1].

Table B.1. Means and standard deviations of age, body mass index, and gender for all available respondents and respondents included in data sets used in this thesis.

	Canadian Joint Replacement Registry (2017-2018)	Study 1 All Respondents		Respondents Included in Study 1 Analysis		Study 2, 3, and 4 All Respondents		Respondents Included in Study 2 and 3 Analysis		Respondents Included in Study 4 Analysis	
		n=3000		n=1432		n=342		n=275		n=304	
Variable	Mean	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	68.3	67.28	9.69	66.75	9.75	64.31	8.53	64.65	8.45	64.63	8.55
BMI	NA	33.32	7.41	33.43	7.47	33.00	7.52	32.97	7.45	32.87	7.47
Gender (female)	0.61	0.64		0.65		0.65		0.64		0.65	

BMI, body mass index

References

1. Canadian Institute for Health Information. Hip and Knee Replacements in Canada, 2017–2018: Canadian Joint Replacement Registry Annual Report. Ottawa, ON: CIHI; 2019.

Appendix C – Ethics Approvals

Western Research

Date: 18 December 2017

To: Dr. Bert Chesworth

Project ID: 110693

Study Title: Examining the Relationship Between Patient Satisfaction and Expectations in Total Knee Arthroplasty

Application Type: HSREB Initial Application

Review Type: Delegated

Meeting Date / Full Board Reporting Date: 09/Jan/2018

Date Approval Issued: 18/Dec/2017

REB Approval Expiry Date: 18/Dec/2018

Dear Dr. Bert Chesworth

The Western University Health Science Research Ethics Board (HSREB) has reviewed and approved the above mentioned study, as of the HSREB Initial Approval Date noted above. This research study is to be conducted by the investigator noted above. All other required institutional approvals must also be obtained prior to the conduct of the study.

Documents Approved:

Document Name	Document Type
Data Collection Ethics	Other Data Collection Instruments

Documents Acknowledged:

Document Name	Document Type
References	References
Study Protocol	Protocol

No deviations from, or changes to, the protocol or WREM application should be initiated without prior written approval of an appropriate amendment from Western HSREB , except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University HSREB operates in compliance with, and is constituted in accordance with, the requirements of the TriCouncil Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2); the International Conference on Harmonisation Good Clinical Practice Consolidated Guideline (ICH GCP); Part C, Division 5 of the Food and Drug Regulations; Part 4 of the Natural Health Products Regulations; Part 3 of the Medical Devices Regulations and the provisions of the Ontario Personal Health Information Protection Act (PHIPA 2004) and its applicable regulations. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Patricia Sargeant, Ethics Officer on behalf of Dr. Marcelo Kremenchutzky, HSREB Vice-Chair

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Date: 14 April 2021

To: Dr. Jacquelyn Marsh

Project ID: 109321

Study Title: Patient Satisfaction and Resource Use after Total Knee Arthroplasty

Application Type: Continuing Ethics Review (CER) Form

Review Type: Delegated

REB Meeting Date: 04/May/2021

Date Approval Issued: 14/Apr/2021

REB Approval Expiry Date: 10/May/2022

Dear Dr. Jacquelyn Marsh,

The Western University Research Ethics Board has reviewed the application. This study, including all currently approved documents, has been re-approved until the expiry date noted above.

REB members involved in the research project do not participate in the review, discussion or decision.

Western University REB operates in compliance with, and is constituted in accordance with, the requirements of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2); the International Conference on Harmonisation Good Clinical Practice Consolidated Guideline (ICH GCP); Part C, Division 5 of the Food and Drug Regulations; Part 4 of the Natural Health Products Regulations; Part 3 of the Medical Devices Regulations and the provisions of the Ontario Personal Health Information Protection Act (PHIPA 2004) and its applicable regulations. The REB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Please do not hesitate to contact us if you have any questions.

Sincerely,

The Office of Human Research Ethics

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Appendix D – Copyright Permission Information for Chapters 4 and 7

Published in *The Journal of Arthroplasty*

Elsevier copyright information.

Can I use material from my Elsevier journal article within my thesis/dissertation? –

As an Elsevier journal author, you have the right to Include the article in a thesis or dissertation (provided that this is not to be published commercially) whether in full or in part, subject to proper acknowledgment; see the Copyright page for more information. No written permission from Elsevier is necessary.

This right extends to the posting of your thesis to your university's repository provided that if you include the published journal article, it is embedded in your thesis and not separately downloadable.

Author rights in Elsevier's proprietary journals	**Published open access**	**Published subscription**
Retain patent and trademark rights	√	√
Retain the rights to use their research data freely without any restriction	√	√
Receive proper attribution and credit for their published work	√	√
Re-use their own material in new works without permission or payment (with full acknowledgement of the original article): 1. Extend an article to book length 2. Include an article in a subsequent compilation of their own work 3. Re-use portions, excerpts, and their own figures or tables in other works.	√	√
Use and share their works for scholarly purposes (with full acknowledgement of the original article): 1. In their own classroom teaching. Electronic and physical distribution of copies is permitted 2. If an author is speaking at a conference, they can present the article and distribute copies to the attendees 3. Distribute the article, including by email, to their students and to research colleagues who they know for their personal use 4. Share and publicize the article via Share Links, which offers 50 days' free access for anyone, without signup or registration 5. Include in a thesis or dissertation (provided this is not published commercially) 6. Share copies of their article privately as part of an invitation-only work group on commercial sites with which the publisher has a hosting agreement	√	√

Appendix E – R Packages Used for Each Study

Study 1

car[1], caret[2], DMwR2[3], e1701[4], gbm[5], GLMnet[6], keras[7], pROC[8], randomForest[9], rms[10], and xgboost[11]

Study 2

bootStepAIC[12], car[1], caret[2], DMwR2[3], ggplot2[13], pROC[8], rms[10], and rpart[14]

Study 3

cutpointr, DMwR2[3], and pROC[8]

Study 4

car[1], caret[2], interactions[15], and mice[16]

References

1. Fox J, Weisberg S. An R Companion to Applied Regression. Third. Thousand Oaks CA: Sage; 2019.
2. Kuhn M. caret: Classification and Regression Training. R package version 6.0-88; 2021.
3. Torgo L. Data mining with R: learning with case studies. 1st ed. Boca Raton, FL: Taylor & Francis; 2010.
4. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. Package 'e1071.' 2021.
5. Greenwell B, Boehmke B, Cunningham J, Developers G. gbm: Generalized Boosted Regression Models. 2020.
6. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1–22.
7. Allaire J, Chollet F. keras: R Interface to "keras." 2021.
8. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.
9. Liaw A, Wiener M. Classification and Regression by randomForest. R News. 2002;2(3):18–22.
10. Harrell F. rms: Regression Modeling Strategies. R package version 6.2-0; 2021.
11. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, et al. xgboost: Extreme Gradient Boosting. R package version 1.4.1.1; 2021.
12. Rizopolous D. bootStepAIC: Bootstrap stepAIC. R package 1.2-0; 2009.
13. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.

14. Therneau T, Atkinson EJ. rpart: Recursive Partitioning and Regression Trees. R package version 4.1-15; 2019.
15. Long JA. interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions. R package version 1.1.0; 2019.
16. Buuren S Van, Groothuis-Oudshoorn K. MICE: multivariate imputation by chained equations in r. J Stat Softw. 2010;1–68.

Curriculum Vitae

Name:	Joseph Munn
Post-Secondary Education	University of Toronto Toronto, Ontario, Canada 2007-2011 Honours BPHE University of Western Ontario London, Ontario, Canada 2014-2016 MSc (Kinesiology) University of Western Ontario London, Ontario, Canada 2016-present PhD Candidate
Honours and Awards	University of Western Ontario Graduate Research Scholarship 2016-2020
Related Work Experience	Teaching Assistant The University of Western Ontario – Kinesiology 2015 Lab Coordinator The University of Western Ontario – Kinesiology 2016 Teaching Assistant The University of Western Ontario – Health and Rehabilitation Science 2017-2020 Teaching Assistant The University of Western Ontario – Biology 2020-2021

Publications:

Refereed Published Papers

Munn JS, Culliton SE, Bryant DM, MacDonald SJ, Chesworth BM. Can Met Expectations Moderate the Relationship Between Pain/Function and Satisfaction in Total Knee Arthroplasty?. The Journal of Arthroplasty. 2021;36(6):1942-6.

Refereed Accepted Papers

Munn JS, Lanting BA, MacDonald SJ, Somerville LE, Marsh JD, Bryant DM, Chesworth BM. Logistic Regression and Machine Learning Models Cannot Discriminate Between Satisfied and Dissatisfied Total Knee Arthroplasty Patients. The Journal of Arthroplasty.

Furlano JA, Morava A, Wong MYS, Bray NW, Sui W, Munn J, Prapavessis H. Exercise behaviours, perceived barriers and motivators to exercise, and use of on- and off-campus exercise resources among graduate students at a Canadian university: A cross-sectional study. Journal of American College Health

Presentations

Munn J, Volker N. Using an aerial drone to investigate lateral movement in sweep rowers (2016) Poster presentation at The 19th Biennial meeting of the Canadian Society for Biomechanics, Hamilton, Ontario, Canada.

Munn J, Volker N. Using an aerial drone to investigate lateral movement in sweep rowers (2016). Oral presentation at the Kinesiology Graduate Student Association Symposium. London, Ontario, Canada.

Munn J, Volker N. Using an aerial drone to investigate lateral movement in sweep rowers (2016). Oral presentation at Bodies of Knowledge 2016. Toronto, Ontario, Canada.

Munn J, Volker N. Using an aerial drone to investigate lateral movement in sweep rowers (2016). Oral presentation at the Ontario Biomechanics Conference, Alliston, Ontario, Canada.

Munn J, Culliton SE, Bryant DM, Cipriano LE, MacDonald SJ, Chesworth BM. Comparing Analytic Approaches for Predicting Patient Satisfaction in Total Knee Arthroplasty (2020). Oral presentation at the Health and Rehabilitation Science Graduate Research Conference.