MEASURE WHAT MATTERS: CHANGES IN HEALTH-RELATED QUALITY OF LIFE OUTCOMES AMONG PATIENTS UNDERGOING SURGERY FOR END-STAGE ANKLE ARTHRITIS

by

Shanika Rajapakshe

B.Sc., The University of British Columbia, 2017

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Population and Public Health)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2019

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis entitled:

MEASURE WHAT MATTERS: CHANGES IN HEALTH-RELATED QUALITY OF LIFE OUTCOMES AMONG PATIENTS UNDERGOING SURGERY FOR END-STAGE ANKLE ARTHRITIS

submitted by Shanika Rajapakshe in partial fulfillment of the requirements for

the degree of Master of Science

in Population and Public Health

Examining Committee:

Dr. Jason Sutherland
Supervisor

Dr. Wei Zhang
Supervisory Committee Member

Supervisory Committee Member

Dr. Aslam Anis
Additional Examiner

Additional Supervisory Committee Members:

Dr. Anne Gadermann
Supervisory Committee Member

Supervisory Committee Member

Abstract

End-stage ankle arthritis (ESAA) is a debilitating condition that negatively affects patients' health-related quality of life (HRQoL). If patients fail conservative treatment, they are treated with ankle arthrodesis (AA) or total ankle replacement (TAR). These surgical procedures have been associated with changes in general and ankle-specific HRQoL. However, there is little understanding of how these treatments affect other aspects of HRQoL, such as depression symptoms and pain, or whether these changes are meaningful to patients.

This thesis is based on a cohort of prospectively recruited patients treated with AA or TAR for their ESAA. This thesis aims to 1) measure changes in aspects of HRQoL that have not been previously investigated, such as depression symptoms and pain, 2) determine if changes in HRQoL over the peri-operative period are meaningful to patients, and 3) determine the effects of potentially modifiable health-system factors, such as wait times and utilization of post-operative physiotherapy, on patients' changes in HRQoL.

Out of 190 eligible patients, 89 were included in the analysis. Participants with the worst pre-operative HRQoL experienced the largest changes. Changes in ankle-specific HRQoL were correlated with changes in pain and depression symptoms. The effects of wait times and post-operative physiotherapy on changes in HRQoL were found to be minimal. However, while most participants saw statistically significant changes in all aspects of HRQoL that were measured, not all participants found these changes to be meaningful. More research is required to understand why certain ESAA patients treated with AA and TAR do not see meaningful changes to aspects of their HRQoL.

Lay Summary

End-stage ankle arthritis is a condition associated with very poor health. Failing medical management, patients with this condition are surgically treated with either ankle arthrodesis or total ankle replacement. These procedures are known to improve aspects of patients' health, including ankle mobility and function. However, it is unclear if other aspects of patients' health, such as pain and depression, improve after surgery, and if these improvements translate to changes that patients really notice. Additionally, there is uncertainty regarding the effects of waiting for surgery and physiotherapy on these improvements.

This thesis investigates how different aspects of end-stage ankle arthritis patients' health-related quality of life improve following treatment by ankle arthrodesis and total ankle replacement.

It was found that almost all patients saw improvement in some aspects of HRQoL, and that waiting for surgery and physiotherapy had little effect on these improvements. However, many of the improvements may not have been meaningful to patients.

Preface

This thesis is based on analysis of patient-reported outcomes data from the Value and Limitations in Hospital Utilization and Expenditures (VALHUE) project, which is a partnership between Vancouver Coastal Health (VCH), Providence Health Care (PHC), and the University of British Columbia (UBC). The research was approved by the University of British Columbia's Behavioral Research Ethics Board (BREB approval number: H12-02062).

This thesis is based on work conducted at UBC's Centre for Health Services and Policy Research and is "in press" as "Health and quality of life outcomes among patients undergoing surgery for end-stage ankle arthritis" published by SAGE Publications in the journal Foot and Ankle International. The author, Shanika Rajapakshe, was responsible for data preparation, statistical analysis, interpretation of results, and writing most of the aforementioned manuscript. Dr. Guiping Liu helped with data-cleaning. Kate Redfern and Drs. Kevin Wing, Trafford Crump, Murray Penner, Alastair Younger and Andrew Veljkovic were involved in manuscript edits. Dr. Jason Sutherland supervised the project and provided guidance regarding methods, interpretation of results, and additional manuscript edits.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AA – Ankle arthrodesis

AAOS-FAM – American academy of orthopedic surgeons foot and ankle module

AAS – Ankle arthritis scale

AOFAS – American foot and ankle society ankle and hindfoot scale

AOS – Ankle osteoarthritis scale

DAD – Discharge abstract database

EQ-5D – EuroQol Five-Dimension Survey

ESAA – End-stage ankle arthritis

FAAM – Foot and ankle ability measure

HRQoL – Health-related quality of life

MAR – Missing at random

MCAR – Missing completely at random

MNAR – Missing not at random

MID – Minimally important difference

PEG – Pain intensity (P), interference with enjoyment (E) of life and general (G) activity instrument

PHC – Providence Health Care

PHQ-9 – Patient health questionnaire

PRO – Patient reported outcome

QoL – Quality of Life

SES – Socio-economic status

SF-36 – Short-form health survey 36 item

SMFA – Short musculoskeletal function assessment

TAR – Total ankle replacement

UBC – University of British Columbia

VALHUE - Value and Limitation in Hospital Utilization and Expenditures study

VAS – Visual analogue scale

VCH – Vancouver Coastal Health

WHO – World Health Organization

WOMAC – Western Ontario and McMaster Universities Arthritis Index

Acknowledgements

I would like to thank my supervisor Dr. Jason Sutherland for accepting me into his research group to conduct this research. His expertise, patience, and understanding have been critical in molding me into the researcher I am today. Without his guidance and encouragement, I would still have trouble writing my way out of a paper bag.

I would also like to thank my committee members Dr. Anne Gadermann and Dr. Wei Zhang for their constructive and insightful feedback while drafting this thesis.

This project would not have been possible without funding from the Canadian Institutes of Health Research, and I am grateful for the opportunities this financial assistance has provided me throughout my Master of Science Degree.

I had the great fortune of working with some incredible colleagues, Kate Redfern, Guiping Liu, and Ernest Lai, who have been indispensable in helping me with analysis and supporting my work at UBC, VCH and PHC.

I need to acknowledge my close friends, the “Kerrisdale crew”, Camilla Piatkowski, Chad Fibke, and Jessi Steinke, for making my program more fun and helping me get through it with a smile. I raise a beaker in your names!

Finally, to all those who have impacted my life along the way, you have my gratitude.

Dedication

To my family, for your unwavering support. You are my light.

Chapter 1: Introduction

1.1 End Stage Ankle Arthritis

End stage ankle arthritis (ESAA) is a degenerative condition of the ankle's tibiotalar joint resulting from cartilage damage (1). Although the ankle joint is subject to more force per square centimeter and is injured more often than other joints in the body (2), ankle arthritis is less prevalent than arthritis of other major joints. Compared to arthritis of the knee, with a prevalence of between 27% and 44% (3), ankle arthritis is estimated to have a prevalence 1/8 to 1/10 of knee arthritis (4–6). Unlike the hip or knee, ankle joints are rarely affected by primary osteoarthritis (7). Instead, previous ankle trauma is the most common origin of ankle arthritis (5,8–10), with some studies finding nearly 80% of all ESAA are attributable to previous ankle injuries or trauma (7,9,11). These injuries include fractures of the ankle, pilon, tibia, fibula, and talus or soft tissue injuries such as sprains, dislocations, or persistent ankle instability (4). The latency, or time, between injury and development of ankle arthritis is thought to range from 12 and 22 years (12,13). Primary osteoarthritis and rheumatoid arthritis can also give rise to ESAA (14), but at a much lower rate than ESAA following ankle injury.

Risk factors for ESAA, such as obesity, age, low muscle strength or low neuromuscular control, alter the biomechanics of the ankle joint or directly damage ankle cartilage (7,13). ESAA symptoms include general or localized pain, stiffness, or decrease in range of motion of the tibiotalar joint in the affected ankle (7). A series of weight-bearing (i.e., while the ankle is under weight) radiographs are used to determine the severity of ESAA. Computed tomography or magnetic resonance imaging can also be utilized for improved specificity and sensitivity of ESAA detection (7,15).

1.1.1 Non-Surgical Treatment

For patients with ESAA, conservative treatments are exhausted prior to proceeding with invasive surgery. One of the most common conservative treatments is the use of a functional orthotic (16), a device meant to be inserted into the shoe to alter the motion or position of the foot. For patients with ESAA, orthotics are meant to change alignment within the ankle or hindfoot in order to shift the loading pressure on the joint (16). Braces, devices worn around the ankle, are used in situations where orthotics have not been successful or are inappropriate and can either be purchased off-the-shelf or custom-made for the patient.

A variety of pharmacologic interventions are also available, including acetaminophen and other over-the-counter medications. The injection of corticosteroids into the affected area can be used to decrease inflammation and pain. Biological or 'regenerative' interventions, such as injections of platelet-rich plasma, amnion, or hyaluronic acid (16,17) may also be recommended.

Patients are also encouraged to modify or stop vigorous activities entirely (2), although this may not be desirable for some patients.

1.1.2 Ankle Arthrodesis

Ankle arthrodesis (AA), also known as ankle fusion, has long been considered the gold-standard for ESAA treatment (14,18,19). In this procedure, the bones of the ankle are fused into one piece, limiting the pain-inducing ankle movement attributable to ESAA (20). This is done by removing cartilage present in the space between the bones. The bones are then compressed together, allowing new bone to grow across the joint similar to a fracture healing process (21). Two or more screws are used to maintain compression of the ankle to ensure that bone grows across the joint (20,22).

Arthrodesis of the ankle was first performed in the 1900s as a method to stabilize the ankle and foot (14). A variety of arthrodesis techniques have emerged since the procedure's introduction, including external fixation (attachment of hardware outside of the body) (14), internal fixation (23), and arthroscopic approaches (14). The choice of technique is dictated by the surgeon's comfort, previous hardware installed in the patient, and the condition of soft tissue surrounding the operating site (19,24).

There are several indications for AA. Patients should have severe ankle pain that interferes with their ability to stand or walk and should have failed conservative treatment for at least three months. Additionally, AA can be recommended after a failure of a total ankle replacement following removal of previous hardware (25,26). AA is avoided in children and adolescents with open growth plates (14). Other contra-indications include active infection, insufficient vascular supply, and inadequate soft tissue envelope, conditions that negatively impact the wound's ability to heal (14,24).

The direct costs attributable to AA are significant, but lower than other major joint surgeries. The estimated total direct costs of AA in 2013 in the US were $16,754 USD, while the combination of direct and indirect costs (due to missed work) amounted to $37,368 (27). In Canada, when considering the costs attributable to operating room time, hospital stay, surgeon billing, and implant cost, AA was found to have a much lower average cost than total ankle replacement (TAR), total hip replacements, or total knee replacements with a total cost of approximately $5,500 in Canadian dollars (2006). Approximately $442 of this was due to surgeon billing and $1200 due to the cost of hardware (i.e., screws and plates) (28).

Union rate, or the percentage of patients who achieve a successful union of the bones of the tibiotalar joint, is a common clinical outcome measure that is used to assess success of AA

(20,22,29). Union rate is approximately 90% (ranging from 82% to 100%) (29–31). Patients with non-unions often undergo revision AA surgery to achieve union, or below-the knee amputations if the ankle site is chronically infected (14). Other complications following AA include infections, damage to nerves, bleeding, blood clots, misalignment, new arthritis in nearby joints (14).

1.1.3 Total Ankle Replacement

The goal of total ankle replacement (TAR) is to restore ankle function while retaining motion of the ankle and reducing pain (32,33). This is done by replacing some or all surfaces of the ankle joint with prosthetic components. TAR was seen as an alternative to AA that, when successful, would provide additional range of motion for patients (18,33). First generation replacements were developed in the 1970s (32,34). These first-generation prostheses had a variety of problems; many were irrationally designed, limited in size, and did not account for the mechanics of the ankle joint (32). At first, studies that investigated the use of these prostheses were promising. However, longer-term studies found that these first-generation prostheses had a high risk of failure due to loosening of the device (32). Rates of loosening ranged from 22 to 75% (10). Since then, a variety of second and third-generation prostheses have been developed (32,34,35).

Indications for TAR are similar to those of AA – severe arthritis whose symptoms have not been alleviated by conservative treatments / medical management (32). Contraindications include diabetes mellites, vascular diseases, previous infections of the foot, or severe neurological disorders (32).

The estimated direct hospital costs of TAR in 2013 in the US were $21,423 USD. In Canada, the total direct cost, noted as cost due to operating room time, hospital stay, surgeon billing, and implant/hardware cost was found to be $13,500 in 2006 dollars in 2015. Of this, $419 was attributable to surgeon billing and $6,420 was attributable to the cost of the implant (28). These costs were found to be similar to those of total knee and hip replacements but were higher than costs for AA.

Complications following TAR include those that occur intraoperatively, such as nerve and tendon lacerations and fractures or component misalignment, or post-operative complications such as fracture of the replacement component, and joint stiffness (32). The primary reason for revision of TAR is due to the position of the components or failure of components to be fixed securely (32,33,36). Revision rates for TAR range from 10 to 22% (31,37–39) at five years and 20 to 24% at 10 years (38,40). Patients presenting with unsalvageable TAR procedures can have a conversion to ankle fusion or amputation (27,32). Younger and more active patients are at higher risk of revision than older and more sedentary patients (32,39).

1.2 Health-Related Quality of Life and Patient Reported Outcomes

1.2.1 Health-Related Quality of Life

Although many definitions exist, health-related quality of life (HRQoL) is often defined as self-perceived health status, or how one views one's own health (41,42). HRQoL generally focuses on functioning and well-being in the physical, psychological, and social domains (42–46) and is influenced by patient and societal preferences (43). While there have been calls to use the term 'self-perceived health-status' rather than 'HRQoL' to define how patients feel about the status of their health (41,42), the term HRQoL is used in this thesis as the latter term persists and is recognized by clinicians in the literature (42,47–49).

HRQoL is closely related to both health and quality of life (QoL) (42). Health is defined by the WHO as “a state of complete physical, mental and social well-being, and not merely the absence of disease and infirmity” (50). The WHO defines QoL in a broader sense, stating that QoL is “individuals’ perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns.” This definition encompasses health, as well as social relationships and aspects of the environment, including physical safety, transport, and work satisfaction (51). Consequently, HRQoL is as an aspect of health, which is in turn an aspect of QoL (41,42).

Wilson and Cleary’s conceptual model of patient outcomes notes that aspects of health, including HRQoL, fall along a continuum of increasing biological, social, and psychological complexity (43,52). Biological measures, such as red blood cell count and heart rate, are found at one end of the continuum, while more complex measures, such as physical functioning and mental health status, are at the other. Traversing across the continuum described in this model can be seen as moving from the cell to the individual to the interaction of the individual and society (43). This model combines two paradigms of health, the clinical paradigm (ie. the biomedical model focusing on biological, physiological, and clinical outcomes) and the social-science paradigm (i.e., HRQoL, focusing on overall well-being and functioning). Combining the two paradigms allows causal relationships to be made between the health concepts found in both paradigms (ie. the ability to relate biological measures of health to HRQoL).

HRQoL touches on a variety of domains of health and can therefore be divided into a number of separate aspects including functional status, symptom severity, and psychological well-being. Certain aspects of HRQoL have been found to be highly correlated, such as pain status and depression symptoms (53).

1.2.2 Patient Reported Outcomes

As HRQoL is subjective (44), it is often measured directly from patients through patient reported outcomes (PROs), measures of patients' self-reported health (54,55). PROs are obtained through validated and psychometrically sound self-administered surveys, asking patients to rate aspects of their HRQoL including functional status and symptom severity (54,56). PROs are increasingly being collected in a variety of domains, including orthopedic surgery. They have many applications including evaluation of clinical outcomes (57), effectiveness of surgical techniques and care pathways (58,59), quality improvement (60), and cost-effectiveness analyses (58,61). PROs can be broken down into two groups; generic instruments and disease-specific instruments.

1.2.2.1 Generic Instruments

PROs that are classified as generic instruments are used to measure aspects of general HRQoL, such as mobility, the ability to perform usual-activities, and self-care (62). Thus, they can be used to make comparisons between populations with different health conditions. However, they are known to be less sensitive to changes after surgery and more sensitive to comorbidities compared to disease-specific instruments (44,63–65). Examples of generic instruments include the Short Form-36 (SF-36) (66), and the EuroQol-5D (67).

1.2.2.2 Disease-Specific Instruments

PROs that are classified as disease-specific instruments are used to elucidate very specific aspects of HRQoL that are relevant to particular diseases. Examples include the Patient Health Questionnaire (PHQ-9) for depression, and the Ankle Osteoarthritis Score (AOS) for ankle arthritis (68). These instruments are designed to be sensitive to changes in functional limitations and/or disease-specific symptoms that may not be captured by generic instruments.

1.2.3 Minimally Important Difference

Although PROs have seen rapid use over the past few decades in clinical practice and research, their interpretability remains a challenge (45,69). For example, if a particular PRO score in a treatment group is three points higher than a control group, what does that mean, and to whom? Would this be considered a large enough treatment effect to justify the choice of one intervention over another?

The minimally important difference (MID) is a concept defined as the smallest change in a given measure that a patient would perceive as beneficial (70). The MID can help clinicians understand the size of treatment effects and compare the possible beneficial and harmful outcomes attributable to a procedure (69). MIDs are frequently derived from changes in PRO scores (71,72). In these cases, the MID is noted as the smallest change in a PRO score that a patient perceive as beneficial (69). For the purpose of this thesis, the MID is defined as the smallest change in a PRO score that a patient perceives as beneficial (70).

The MID is related to the minimal clinically important difference (MCID), which refers to the smallest change in a measure that would be noted as being *clinically* meaningful (70). However, both the MID and MCID have been used in literature to define changes important to patients (73). This synonymy (same meaning for varying terms) can make it difficult to differentiate the two concepts (74). However, it is important to keep in mind that both the MID and MCID are thresholds that indicate whether a change is considered meaningful to patients or clinicians respectively.

There are two primary approaches to calculating MIDs: distribution- and anchor-based methods (69,75). Distribution based methods compare changes in a PRO score to a measure of variability, such as the standard deviation or standard error of the distribution of scores. Anchor based methods compare a change in a PRO score to an independent measure that indicates a meaningful change, such as a self-reported global assessment question or change in value of another PRO. While distribution-based methods may not directly ascertain that a change is meaningful to patients, it is assumed that MIDs derived from this approach are meaningful to patients (75). Both methods are common, and there is uncertainty over which should be preferred (75). Therefore, many studies present MIDs calculated using several variations of both distribution and anchor-based methods (71,76).

1.3 Changes in HRQoL of ESAA Patients Treated with AA and TAR

Ankle surgery patients overwhelmingly seek relief from pain and improvements in function related to their ankle, such as walking or performing daily activities (77,78). Therefore, for patients treated with AA or TAR, previous research into changes in HRQoL in the peri-operative period has focused on ankle-specific and general HRQoL (30,37,72,79,80).

Hendrickx (30) explored the post-operative HRQoL of ESAA patients treated with AA through a retrospective follow-up study of 66 AA procedures. The authors measured general HRQoL with the SF-36 and ankle-specific HRQoL with the American Orthopaedic Foot and

Ankle Society Ankle and Hindfoot scale (AOFAS) and the Foot and Ankle Ability Measure (FAAM). Patients completed PROs post-operatively during annual physical examinations, and the mean follow-up time was 9 years. However, this study lacked pre-operative scores. To compensate for this, the authors compared general health scores to those of the general Dutch population, finding patients' post-operative general HRQoL were similar. However, physical function and bodily pain were significantly lower in the patient cohort compared to age-matched controls from the general Dutch population. The authors found that AA patients' AOFAS post-operative scores were similar to post-operative scores of AA patients in other studies. The authors did not state whether patients experienced meaningful changes in their ankle-specific HRQoL. The authors noted that 91% of the cohort was satisfied with their clinical result, but did not detail this evaluation, aside from stating that "patients were asked if they were satisfied with their situation" (30).

Koivu et al. (81) retrospectively reviewed a cohort of 34 TARs performed on 33 patients suffering from ESAA operated or supervised by a senior orthopedic surgeon over a 15-year period. Patients completed the Kofoed ankle score, a measure of ankle-specific HRQoL that measures pain, function, and range of motion associated with the ankle. The instrument was administered pre-operatively, one year after surgery, then every second year thereafter. The median follow-up time was 13.3 years. The authors found that patient's ankle-related pain and function, as measured by the Kofoed ankle score, improved considerably, from a median pre-operative score of 42 to a median post-operative score of 88 points at 1 year, 82 points at 5 years, 81.5 points at 10 years, and 79.5 points at 15 years. This translated to 'excellent' ankle-specific HRQoL at 1 year, and 'good' ankle-specific HRQoL at every point thereafter. The authors did not indicate if these changes surpassed the instrument's MID.

Lachman et al. (82) retrospectively reviewed a prospectively collected cohort of 29 ankle arthritis patients with failed TAR and had undergone revision surgery. Patients had completed the AOFAS, SF-36, and Short Musculoskeletal function assessment (SMFA) pre-operatively, then 6 months and annually post-operatively. Patients had an average time to revision of 3.9 years and an average follow-up of 3.2 years after revision. Both primary and revision patients experienced statistically significant changes in ankle-specific and general HRQoL following surgery. Similar to previous studies, the authors did not indicate whether these changes were meaningful to patients. Average changes in all instrument scores were greater in primary than in revision TAR.

Desai et al. (83) reported on a prospectively-recruited cohort study, identifying changes in HRQoL for bilateral and unilateral ESAA patients treated with TAR. Patients completed the SF-36 and AAOS-FAM to elucidate general and ankle-specific HRQoL. Surveys were administered once pre-operatively, then annually post-operatively with a mean follow-up time of 5.9 years. Patients in both the bilateral and unilateral groups experienced statistically significant changes in general and ankle-specific HRQoL post-operatively. While pre-operative scores differed between groups, with unilateral patients exhibiting better general and ankle-specific HRQoL, post-operative scores were similar. The authors did not determine if these changes in score were meaningful to patients.

Studies have also compared changes between AA and TAR, finding that changes in HRQoL attributable to both procedures are nearly equal. Daniels et al. (37) investigated the post-operative HRQoL of 107 AA and 287 TAR surgeries for ESAA patients through a prospective cohort study. The authors compared post-operative ankle related health, as measured by AOS, a measure of ankle-specific health, and general HRQoL as measured by the SF-36, among AA and

TAR patients. Patients completed instruments pre-operatively, one year following surgery, then annually thereafter. The average follow-up was 5.5 years. Among both AA and TAR patients, the authors found that ankle-specific and general HRQoL improved post-operatively. AOS scores of AA patients improved by roughly 20 points, and TAR patients by 26 points. SF-36 scores improved by points 6.6 for AA patients and 5.4 for TAR patients. While these changes were considered statistically significant, the authors did not determine if they were larger than each instruments' MID and therefore did not determine if these changes were meaningful to patients. The authors also performed a linear mixed effects regression model, which compared changes in scores of AA and TAR patients while adjusting for demographic variables (age, sex, presence of comorbidities such smoking status, body mass index, and inflammatory arthritis diagnosis), clinical variables (surgeon and side of surgery), and pre-operative PRO scores. After these adjustments, the authors found that changes in AOS and SF-36 scores were similar for AA and TAR patients.

Slobogean et al. (84) compared changes in general HRQoL among 46 AA and 61 TAR procedures in a prospectively collected cohort of ESAA patients. General HRQoL was measured by the SF-36 instrument. SF-36 scores were then converted to a scale ranging from zero to one (zero referring to worst health and one referring to best possible health). Patients completed the SF-36 pre-operatively, then one-year post-operatively. AA patients had a mean pre-operative score of 0.66, compared to 0.67 for TAR patients. Mean post-operative scores for both AA and TAR patients were the same, at 0.73. Changes in the SF-36 were found to be statistically significant for both groups. However, the authors did not determine if these changes were meaningful to patients.

1.3.1 Gaps in the Literature

Although there has been a significant volume of scholarship surrounding changes in HRQoL of ESAA patients treated with AA and TAR, gaps in the literature remain. These gaps are 1) aspects of HRQoL that have not been explored, 2) the meaning/significance of changes in HRQoL to patients, and 3) the effects of health-system delivery factors such as utilization of physiotherapy and surgical wait times on these changes.

1.3.1.1 Missing Aspects of HRQoL

The available literature is primarily focused on ankle-specific and general HRQoL measures. While these aspects of HRQoL are important to patients (77,78), other aspects of HRQoL may still be relevant. For example, previous research has indicated that anxiety (85) and pain-catastrophizing (86) are present in osteoarthritis patients. Another understudied aspect of HRQoL in ESAA patients is depression symptoms. Depression, also known as major depressive disorder (MDD), is recognized as a period of at least 2 weeks during which a patient exhibits either depressed mood or the loss of interest or pleasure in nearly all activities (87). The impact of depression is immense; the 12-month prevalence of depression globally is estimated at 4.4% and is the largest contributor to global disability according to the World Health Organization (WHO), attributable to 7.5% of all years lived with disability in 2015 (88). The average lifetime prevalence for depression is estimated to be 11.1% in low- to middle-income countries and 14.6% in high-income countries (89). Depression effects people from all walks of life, including members of the military (90), children (91), and elective surgery patients (92).

There is evidence that the prevalence of depression among patients suffering from osteoarthritis is nearly double that of patients without arthritis (93,94), and joint surgery patients suffering from depression have been found to have more complications and worse post-operative outcomes than patients without depression (92,95,96). While there are some studies that have looked into changes in mental health status following AA or TAR (37,97), they do not explicitly measure depression symptoms.

Depression is also closely related to pain. The prevalence of pain in depressed cohorts is higher than that of non-depressed cohorts (98). Similarly, the prevalence of depression in pain cohorts is higher than that of non-pain cohorts (53). There is a large body of evidence detailing the interaction between depression and pain symptoms, which has been labeled depression-pain syndrome or the pain-depression dyad (99,100).

Pain is an important indication for surgery and a major opportunity for post-operative improvement among orthopedic patients (53,99–101). While ankle-specific pain is often a domain of PROs that elucidate ankle-specific HRQoL (68), none of the studies identified in section 1.3 used instruments specifically designed to measure general pain. The connection between ankle-specific pain and general pain has not been investigated, meaning there is uncertainty around how general pain changes following surgery. Furthermore, given the relationship between pain and depression, if a relationship between general pain and ankle-specific pain is firmly established, then there could be a relationship between depression and ankle-specific HRQoL (since ankle-related pain is often considered an aspect of ankle-specific HRQoL). Additionally, the focus on pain overall rather than pain specific to the ankle would allow for comparisons of changes in HRQoL to other major joint procedures such as hip or knee replacements.

1.3.1.2 Relevance of Changes in HRQoL to Patients

While studies investigating changes in HRQoL following AA or TAR for treatment of ESAA have found changes to be statistically significant, the importance of these changes to patients is largely unstudied. Comparing changes in HRQoL to a PRO's MID is one method to evaluate the significance of these changes to patients (45). However, many PRO instruments do not have established MIDs, leading to their absence from many HRQoL studies. Therefore, there is an opportunity to use PROs with established MIDs in order to measure if changes in HRQoL after AA or TAR surgery are meaningful to patients.

1.3.1.3 The Effects of Post-Operative Physiotherapy and Wait times on HRQoL

Of the previous work on changes in HRQoL of ESAA patients, only the study by Daniels et al. (37) investigated changes while adjusting for demographic variables, such as age, sex, and presence of comorbidities. However, other variables could also influence changes in HRQoL following surgery. For example, health-system delivery factors have been known to influence patients' health outcomes (102). Two of these potentially modifiable factors, time waiting for surgery and post-operative delivery of physiotherapy, are especially relevant to ESAA patients treated with AA and TAR.

Physiotherapy is often routinely prescribed as a form of treatment in the post-operative period, with the goal to minimize pain and optimize recovery after a hospital stay (103). There have been a variety of systematic reviews investigating the effectiveness of post-operative physiotherapy in lower extremity orthopedic patients. Post-operative physiotherapy is associated with improved short and medium-term (under one year) levels of physical function, pain, and range of motion following replacement of the hip or knee (104,105). Although there has been significant research investigating physiotherapy following surgery of the hip and knee, there is very little information available for ankle surgery. Recommendations that physiotherapy be administered following surgery exist for ankle patients (106,107), but no scientific studies have published on the use of physiotherapy among ESAA patients treated with AA or TAR in the peri-operative period.

Many patients, including those in Canada, report that their wait for surgery, known as their ‘wait time,’ is unacceptable (108). Wait time for surgery, or the amount of time between being placed on a surgical waitlist and receiving that surgery, is used as a benchmark for effectiveness or equity of access between health systems (109,110). There is uncertainty surrounding the effects of surgical wait times on changes to orthopedic patients HRQoL in the peri-operative period.

There are two randomized controlled trials (111,112) and three prospective cohort studies (113–115) measuring the association between the wait time and changes in HRQoL during the wait for hip and knee replacements. The two randomized controlled trials, one for hip and one for knee replacement, found no evidence of an association between longer wait times with worse health at the end of the waiting period, as measured by the Harris hip score or Knee Society Clinical Rating System respectively (111,112). The cohort studies by Ostendorf et al. and Kelly et al. also found no significant changes in HRQoL as measured by the Western Ontario and McMaster Universities Arthritis Index (WOMAC) and SF-36 for hip revision patients and total knee/hip replacement respectively while on the surgical waitlist (114,115). However, the cohort study by Desmeules et al. found that patients who waited more than nine months for knee replacement experienced significant worsening of pain and function while on the waitlist (113).

There has been little investigation into the effects of wait times on changes in HRQoL of AA and TAR patients.

Chapter 2: Study Purpose and Rationale

This study will measure changes in different aspects of HRQoL (general HRQoL, ankle-specific HRQoL, pain, and depression symptoms) during the peri-operative period among patients who receive AA or TAR as treatment for ESAA, and whether these changes are considered meaningful/significant to patients. Additionally, the effects of demographic variables and modifiable health system factors, such as time on the wait list and utilization of post-operative physiotherapy, on these changes will be investigated.

2.1 Investigate Changes in Multiple Aspects of HRQoL as Measured by PROs

Most studies investigating changes to HRQoL of ESAA patients treated with AA and TAR have been focused on ankle-specific and general HRQoL. Therefore, changes in other aspects of HRQoL, such as depression symptoms and pain, are not well understood. Although understudied, these aspects of HRQoL may still be important to ESAA patients treated with AA or TAR as many orthopedic and arthritic patients have high levels of depression symptoms and pain (93,96,101). For example, as pain and depression are closely related (53), it is possible that following surgery, patient's depression symptoms, in addition to pain, will change.

This thesis will explore changes in general HRQoL, ankle-specific HRQoL, pain, and depression symptoms as measured by PROs of ESAA patients treated with AA or TAR. Possible correlations between changes in different aspects of HRQoL will also be evaluated. Since it is unknown if changes in HRQoL of ESAA patients treated with AA or TAR are meaningful, changes in PRO scores will be compared to each instruments' MID. By examining aspects of HRQoL under-investigated elsewhere, investigating their interrelatedness, and determining their significance to patients, this thesis aims to provide additional insight into changes in HRQoL for ESAA patients treated with AA or TAR.

2.2 Investigate Effects of Post-Operative Physiotherapy and Wait Time on HRQoL

The effects of utilization of physiotherapy in the post-operative period or duration on the surgical waitlist on changes in HRQoL have not been evaluated among ESAA patients treated with AA and TAR. Previous research into the effects of physiotherapy and wait times on HRQoL have focused on other joints such as the hip or knee. This thesis aims to target this gap in knowledge through exploratory analysis of the effects of these factors on changes in HRQoL. The results can help discharge planning following AA or TAR or encourage effective utilization of care while ESAA patients are on the surgical waitlist.

Chapter 3: Data

3.1 Data Source Overview

This study uses data collected from AA and TAR patients participating in the Value and Limitation in Hospital Utilization and Expenditures (VALHUE) project. This population-based study collects PROs during the peri-operative period. The VALHUE Project is a partnership between Vancouver Coastal Health (VCH), Providence Health Care (PHC) and the University of British Columbia (UBC). Data collection among orthopedic patients has been ongoing since October 2015. Information collected by the VALHUE project is linked with sources of administrative data and clinical information regarding patient's treatment in the peri-operative period.

3.2 Participant Recruitment, Inclusion and Exclusion Criteria

Patients are identified from VCH's population-based registry of BC residents prospectively scheduled for primary AA or TAR surgery by one of four foot and ankle surgeons practicing in the provincial referral centre for foot and ankle surgery. Therefore, it is likely that almost all patients who receive these procedures in the province of BC during the recruitment period are identified for inclusion in this cohort irrespective of where they reside in the province.

To be eligible for this study, patients have to be community dwelling, 19 years of age or older, able to answer survey questions in English with or without assistance and scheduled for TAR or AA. Patients are excluded if their surgery is emergent, scheduled within two weeks of being enrolled on the waitlist, or scheduled for a revision surgery. Eligible patients are contacted by phone by research staff at VCH to participate during normal business hours (9am to 5pm on weekdays). Participants were recruited from October 2015 to August 2018. Post-operative PROs were collected until February 2019.

Participants complete a battery of four PROs, the EurQol Five-Dimension three-level instrument (EQ-5D(3L)), Ankle Osteoarthritis Scale (AOS), Pain intensity, interference with Enjoyment of life, and General activity instrument (PEG), and the Patient Health Questionnaire (PHQ-9), which are discussed in more detail below. The PROs are administered to participants before and six months after their surgery. Almost all participants complete their pre-operative PROs in the three months before their surgery. Participants who do not complete their survey packages within two weeks received reminder phone calls or emails.

VCH Legal and Privacy Office completed a Privacy Impact Assessment and the University of British Columbia's Research Ethics Board approved data collection and linkage (approval number: H12-02062).

3.3 PRO Instruments

3.3.1 EuroQol Five Dimension Scale

Many studies investigating changes in HRQoL following AA or TAR use a generic PRO instrument to measure general HRQoL. The EQ-5D(3L)'s Visual Analogue Scale (VAS), referred to as the EQ-5D VAS, was selected to measure general HRQoL, based on its brevity, comparability, and psychometric properties (67). The EQ-5D VAS ranges from 0 (i.e., "Worst imaginable health state") to 100 (i.e., "Best imaginable health state"). The EQ-5D VAS has demonstrated strong correlations to other measures of HRQoL important to orthopedics patients such as pain and disability (116). The EQ-5D VAS has also been found to be equally responsive as the SF-36 and WOMAC, other instruments that are frequently administered to orthopedic patients (117). The EQ-5DVAS has seen extensive use in patients with ESAA (118–121).

The EQ-5D(3L) also includes five items corresponding to the domains of mobility, self-care, usual activities, pain/discomfort, and anxiety/depression ranked on a three-level scale. Scores of each domain are combined to denote one of 243 (3^5) unique health states. These health states can be transformed into utility scores, subjective values of health states, used by health economists in cost-utility analyses (27,122). While these utility scores have been considered measures of HRQoL (42), they are cohort or country specific and make generalizations to other jurisdictions challenging (122). Furthermore, the EQ-5D(3L) items have exhibited strong ceiling effects in lower extremity orthopedic patients (123), and have been criticized for not adequately capturing milder health problems (124,125). This in turn has led to bimodality in the distribution of EQ-5D(3L) utility scores among orthopedics patients, making them harder to incorporate into in statistical models (125). In this study, patterns of responses to the five items of the EQ-5D are not explored.

3.3.2 Ankle Osteoarthritis Scale

The ankle osteoarthritis scale (AOS) measures self-reported disability and symptom severity attributable to ESAA (1,68). Patients complete each of the 18 items by placing a mark along a 100mm horizontal line, bounded by “No pain/difficulty” on the left and “Worst pain imaginable / So difficult unable” on the right.

The 18 items are split into two nine-item subscales for pain and disability. Averaging the two subscale scores generates the overall score, which ranges from 0 (best ankle function) to 100 (worst ankle function). The AOS has exhibited excellent test-retest reliability and strong criterion validity, finding an intraclass correlation coefficient of 0.97 between surveys completed a week apart, and strong correlations to the WOMAC and SF-36 (126). Construct validity, or how well a measure reflects a given latent variable or underlying concept (127), has also been assessed through correlation with a clinical measure of functional disability, finding that the instrument captures both pain and disability in the ankle (1,68,126). Differential item functioning, or differences in measurement qualities of an instrument between two or more categories (i.e., sex, type of surgery, etc.), has been found to be negligible. The instrument has shown minimal ceiling and floor effects when administered to patients with advanced ankle arthritis (128). The AOS has seen extensive use among ESAA patient populations (37,121,129).

3.3.3 Pain intensity, interference with enjoyment of life and general activity (PEG)

The Pain intensity (P), interference with enjoyment (E) of life and general (G) activity instrument is used to measure general pain. Known as PEG and based on the Brief Pain Inventory (BPI), this three-item instrument consists of one pain intensity item and two pain interference items. Items are scored on a scale of 0 (i.e., No pain/interference) to 10 (i.e., pain "as bad as you can imagine/complete interference").

The PEG instrument's total score is calculated as the average of the three items. Scores greater than three are indicative of high levels of pain (130). The instrument has shown strong psychometric properties such as good reliability and construct validity and is as responsive as other validated measures of pain such as the Brief Pain Index (131,132) among primary care patients. Though ceiling and floor effects for this instrument have not been reported, the PEG has seen use among patients presenting with musculoskeletal pain in primary care (131,132), and orthopedic patients (121).

3.3.4 Patient Health Questionnaire (PHQ-9)

The patient health questionnaire (PHQ-9) is an instrument used to measure depressive symptoms in patients (133). This instrument measures functional impairment and symptoms due to depression. Participants respond to nine items, each ranked on a four-point Likert scale ranging from 0 (“Not at all bothered”) to three (“Bothered nearly every day”).

The instrument’s total score is calculated by summing the nine item scores. Scores greater than 10 are considered clinically significant depression, while scores above 15 and 20 correspond to moderate and severe depression respectively (134). The PHQ-9 has demonstrated strong floor effects in primary care and obstetrics-gynecology patients (135). The instrument has been noted as having good construct validity and test-retest reliability (136) and has seen use among orthopedic populations (93,137).

3.4 Demographic and Clinical Variables

VCH links PROs data to the Discharge Abstract Database (DAD) and the BC Surgical Registry. The DAD is a database that holds information regarding the population of discharges from acute care hospitals in BC. The BC Surgical Registry is the provincial database that holds information for patients on the surgical waitlist for elective procedures. Research staff also link data from participant’s electronic medical records located in the province’s referral centre for foot and ankle surgery.

The remainder of this sub-section details the rational for choosing the study’s analytic variables. Age and sex are commonly used as covariates and summarizing cohorts in HRQoL research (37,71,121), and are obtained from surgical registry data supplied by VCH.

The Charlson index, a measure of the severity of comorbid conditions (138), is included as patients with more comorbid conditions often have worse post-operative outcomes (139). The Charlson index is calculated using all comorbidities identified from the Discharge Abstract Database (DAD) linked with participant's PROs.

Socio-economic status (SES) is included in the analyses since lower SES is associated with poorer post-operative outcomes, including increased pain and reduced physical function among joint replacement patients (140,141). SES is obtained from Canadian census data independent of this study (142). SES is calculated for BC's census dissemination areas as a composite index of six separate indices. Each of the six indices is a combination of three or four variables and acts as a measure of regional hardship. Participants are assigned the SES of their census dissemination block, which is determined from their postal code. While participants' family SES would be the preferred, previous work by Mustard et al has shown that SES status derived from postal code income information works as a reasonable proxy for family-level SES (143). Since two types of surgery are investigated (AA and TAR), a dichotomous variable is included that indicates which procedure each participant receives. This information is obtained from linking with the DAD.

Wait times and utilization of post-operative physiotherapy are included in order to explicitly investigate their effects on changes to HRQoL. Wait time is measured as the number of weeks between participants being placed on the BC surgical registry and the surgery date and is also sourced from the surgical registry. As ESAA is a degenerative condition, wait time may be confounded with severity of ESAA-related symptoms (i.e., the longer a patient is on the waitlist, the worse their ESAA).

To measure post-operative physiotherapy utilization, counts of participants' visits to publicly funded post-operative physiotherapy in the 180 days following discharge are observed from participants' electronic medical records. The 180 day threshold is chosen since most post-operative physiotherapy following orthopedic surgery occurs in this period (104,144). In this study, privately paid physiotherapy, which is thought to be common, is not observable and is a limitation of the findings.

Chapter 4: Methods

The main investigation is to measure changes in HRQoL (general HRQoL, ankle-specific HRQoL, pain, and depression symptoms) as measured by PROs in the peri-operative period for ESAA patients treated with AA or TAR.

This study also measures whether participants' changes in HRQoL are greater than each PRO instrument's MID, indicating whether changes are meaningful to patients. The role of surgical wait time, utilization of post-operative physiotherapy, type of surgery, and demographic variables on changes in HRQoL are also investigated.

4.1 Missing Data

Missing data is common in health research utilizing PROs values (145). There are three types of missing data: Missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). If the probability of a variable miss is not dependent on any other variable found in the dataset, or on the value of the variable itself, then that variable is MCAR. Dropping data that is MCAR does not bias results (146). If the probability of missing a variable is dependent on one or more observed variables, that variable is MAR. If one conditions on observed variables, the probability that the variable of interest is missing becomes MCAR. In this situation, other variables can account for the missingness, and can adjust for bias or loss of power. If the probability of a variable missing is dependent on unobserved variables, that variable is MNAR. This is the most difficult kind of missing data to deal with, as it cannot be easily adjusted for effects on statistical inference. This in turn may produce biased results (145,147).

A common method of dealing with missing categorical data that is MCAR, MAR, or MNAR is the inclusion of a separate category for missing values (148). For missing continuous data, multiple imputation, replacing missing data using a sophisticated algorithm that incorporates information from the rest of the dataset is often employed. The most common algorithms for multiple imputation are Markov Chain Monte Carlo and iterative chained equations (146). If data is MCAR, using multiple imputations allows one to compensate for a lack of power (149). Multiple imputations can also remove bias if data is MAR and can minimize bias to less than 5% if data is MNAR (149).

Missing data in the present cohort is limited to SES information and PRO scores. SES could be missing due to several reasons, including due to participants living in new housing developments not yet assigned to an SES indicator, living on reserve, being homeless, or living outside of BC. Participants with missing SES information are grouped into a separate SES category. This 'Missing SES' category accounts for roughly 16% of participants. PROs scores are found to be missing in roughly 1% of participants. Many studies that use PROs make the assumption that missing PROs scores are MAR, and are suitable to augmentation by multiple imputation (150,151). Thus, multiple imputation by chained equations (146) is employed to account for missing PROs scores, assuming that most data is MAR. Simulations have shown that many imputations are necessary to ensure sufficient power (152). For the purpose of this thesis, 100 imputations are used.

4.2 Presentation of Sample Demographics and Clinical Variables

Available sample demographic information includes age, sex, SES, and Charlson index. Age is summarized by the sample mean and standard deviation. To make presentation of demographic information easier to interpret, all demographic variables are split into categories. Age is split into 50 years and under, between 51 and 60, between 61 and 70, and greater than 70 years. Sex is split into male and female. SES is categorized into 5 quintiles. As mentioned in section 4.1, a sixth quintile is included for participants with missing SES data. Participants are split into two categories based on Charlson index; an index of zero indicating no comorbid conditions or an index greater than or equal to one indicating the presence of one or more comorbid conditions.

Additional categories are created from clinical information, including type of surgery, time on surgical wait list, and utilization of post-operative physiotherapy. Surgery is split into either AA or TAR. Wait time is measured as the number of weeks between being placed on the BC surgical registry and the surgery date. Wait time is summarized with three categories: 12 weeks or less, between 13 and 26 weeks and greater than 26 weeks, thresholds considered ‘good’, ‘acceptable’, and ‘unacceptable’ by clinicians. Wait times for related procedures in Canada (hip and knee replacement) have a benchmark wait time of less than 26 weeks (110). Participant’s physiotherapy visits are categorized; zero/no utilization and at least one physiotherapy visit.

Demographic and clinical characteristics of participants, as well as utilization of publicly funded outpatient physiotherapy and wait time categories, are summarized by counts and percentages to give an overview of the cohort and provide support that this cohort is similar to other cohorts of ESAA patients treated with AA or TAR.

4.3 Comparing Changes in HRQoL to MIDs

To investigate the significance of changes in HRQoL, this study uses the MIDs of each PRO instrument. These MIDs are obtained in two ways – from previous research and the current participant cohort.

Previous research using an anchor-based approach to calculating MIDs found that the MID for the EQ-5D VAS to be 7.0 (153). This value was based on a cohort of 534 cancer patients, and may not be the suitable to the current patient cohort. However, no other estimates of the MID for the EQ-5D VAS were available from previous research and supports this research's calculation of an MID value from this cohort of ESAA participants.

The MID of the AOS has been reported as 28, and was derived from a cohort of 238 ESAA patients treated with AA and using an anchor-based approach (72). A summary of previous research for pain instruments found that the MID of the PEG was 1.0, calculated via a distribution-based approach (76). For the PHQ-9, a change of 5.0 points was found to be the MID, calculated using a distribution-based approach using scores of 434 patients in a primary care setting (154).

MID values for each of the PROs used in this study are also generated from the current cohort for comparative analysis of MIDs. To create these MIDs, a common distribution-based calculation method, 0.5 times the standard deviation of change in each PRO measure (75), is utilized. In this approach, the MID of a given instrument is defined as 0.5 times the sample standard deviation of the change in that instrument. Although distribution-based methods have been criticized for not explicitly asking patients what they believe is a meaningful change and for being sample specific (75), this method of MID calculation has seen extensive use in PROs research (71,155).

For each PRO instrument and participant, indicator variables are calculated representing whether the participant's change in PRO score exceeds the instrument's MID. This is performed for both the MIDs sourced from previous research and calculated from this cohort.

Participants PRO values are summarized by categories noted in section 4.2. Mean pre- and post-operative PRO scores are reported for variables' categories. Paired t-tests are performed, for the entire cohort and each demographic and clinical category, to determine if mean pre- and post-operative scores differed statistically. P-values of these t-tests are presented. As this analysis is exploratory, adjustments for multiple comparisons are not applied (156).

4.4 Changes in Multiple Aspects of HRQoL

When investigating changes of multiple PROs scores, one may be interested in the interrelatedness of these changes. Related changes in multiple PRO scores could provide evidence that instruments measure overlapping concepts. For example, the AOS includes an ankle-specific pain domain, which could overlap with pain as measured by the PEG, and result in an association between changes in the AOS and PEG. To explore potential relationships between changes in PRO scores, correlations between changes in each instrument are reported.

To visualize multiple aspects of HRQoL simultaneously, several figures showing changes among instruments' values are presented.

4.5 Variables Associated with Changes in HRQoL

In order to determine the effects of utilization of post-operative physiotherapy, wait times for surgery, demographic variables, and the type of surgery, on changes in HRQoL, two analyses are performed. The first is a series of four multivariable linear regression models, one for each

PRO instrument. The dependent variable is the change in PRO value (defined as the post-operative score minus the pre-operative score).

$$\Delta\ (change)\ in\ a\ PRO\ score\ = postoperative\ score - preoperative\ score$$

This is then adjusted for participant demographic information (age, sex, SES, Charlson index) as well as the type of surgery performed. Additionally, the models include a term for pre-operative PRO scores as an explanatory variable to adjust for pre-operative HRQoL. The models can be summarized by the general equation,

$$change\ in\ a\ PRO\ score = \ \beta_1 * A + \ \beta_2 * C + \ \beta_3 * D + \ \beta_4 * E + \varepsilon$$

where *A* represents each pre-operative PRO score, *C* represents demographic and clinical variables, *D* represents the number of post-operative physiotherapy visits a participant had, and *E* represents the wait for surgery in weeks. ε refers to the error term. The β_i terms represent the coefficients associated with each explanatory variable. Coefficients and p-values are reported for these four regression models.

In order for linear regression models to be valid, they need to satisfy four key assumptions: that the dependent variable can be expressed as a linear function of the independent variables, that error terms are independent, that error terms are normally distributed, and that variation of error terms is constant. To check for these assumptions, plots of dependent variables vs independent variables and plots of residuals are produced and evaluated.

In order not to enforce arbitrary structure on the data, the continuous variables 'wait time in weeks' and 'number of physiotherapy visits' are selected over their categorical counterparts introduced in section 4.2. However, there is an argument for using the categorical versions of these variables. The utility of wait time categories, which are defined by clinicians, can examined in greater detail upon their inclusion in the aforementioned regression models. Most participants did not receive any physiotherapy in the post-operative period and treating physiotherapy visits as continuous may violate the linearity assumption present in regression models. Dichotomizing physiotherapy visits into zero visits and one or more visits, can overcome this possible violation (157). With this in mind, the four models above are re-run using the categorical versions of wait time and physiotherapy visits as a sensitivity analysis. Coefficients and p-values for these four regression models are reported in the appendix.

The second analysis, performed to measure whether demographic variables, type of surgery, wait times, and utilization of physiotherapy affect changes in all instruments simultaneously, is a multivariate analysis of covariance (MANCOVA). This is an extension of the analysis of covariance (ANCOVA) and the multiple analysis of variance (MANOVA). In the case of MANCOVA, variation in the means of several continuous dependent variables are assessed across multiple independent categorical variables simultaneously while considering the effects of independent continuous variables.

The design of the MANCOVA model is as follows:

$$change\ in\ PRO\ scores = \beta_1 * A + \beta_2 * C + \beta_3 * D + \beta_4 * E + \varepsilon.$$

where *A* represents all pre-operative PRO scores, *C* represents demographic and clinical variables, *D* represents the number of post-operative physiotherapy visits and *E* represents surgical wait time in weeks. The β_i terms represent the coefficients associated with each explanatory variable. ε refers to the errors for the vector of dependent variables. The main difference between this model and the previous analyses is that the MANCOVA can determine the effects of all independent variables on all of the dependent variables simultaneously. Thus, the 'change in PRO scores' on the left of the above equation refers to all changes in *all* PRO scores measured. Additionally, the variable corresponding to *A* in the above equation refers to *all* pre-operative PRO scores.

For a MANCOVA to be sound, several assumptions must hold: that the observations are independent, that dependent variables are normally distributed, that covariances of dependent variables are equal, that independent continuous variables are uncorrelated, that regression slopes for each category do not differ significantly, and that independent continuous variables are not related to the independent categorical variables. Observations are assumed to be independent, as it is unlikely that the changes in PRO scores of one participant affects that of another. Histograms of dependent variables are used to check for normality. Covariance matrices of dependent variables are used to check for equal covariances across dependent variables. If covariances are not found to be equal, the dependent variables are z-score standardized (defined as subtracting the mean and dividing by the standard deviation), and covariances are re-evaluated. Correlations between continuous independent variables are inspected. If two variables are highly correlated, one is dropped from the model. To determine that regression slopes do not differ between categories, the effects of interaction terms are assessed. Relationships between categorical and continuous variables are checked using boxplots. For the MANCOVA model, adjusted coefficients and p-values are reported.

Like the separate linear regression models above, a sensitivity analysis is performed using categorical variables. Adjusted coefficients and p-values for this model are reported in the appendix.

4.6 Variables Associated with Meaningful Changes in HRQoL

A series of four logistic regression models (one for each instrument) are created to determine if there is a relationship between experiencing a meaningful change in HRQoL and utilization of post-operative physiotherapy, wait times category, demographic variables, and the type of surgery participants undergo. Logistic regression assumes that the outcome is binary and that observations are independent.

In situations where there are few participants in combinations of categorical independent and dependent variable, estimates of regression coefficients can be biased away from the null hypothesis (158). Also known as sparse data bias, this phenomenon can be worsened by the exponentiation of coefficients in logistic regression. To assess sparsity, 2×2 tables are tabulated for categorical independent variables and the dependent variable (i.e., exceeding a given MID) are generated. If sparsity is associated with a particular categorical variable, a new categorical variable is generated which avoids sparsity while attempting to maintain the underlying pattern of the data. The 2×2 tables of the independent variable and categorical dependent variables that show sparsity will be presented, along with 2×2 tables of the independent and new categorical variables. Regression coefficients, p-values, odds ratios, and 95% confidence intervals of odds ratios of the four logistic regression models are reported.

Participants are also categorized based on the number of MIDs they exceeded. That is, if a participant experiences changes in the AOS and EQ-5D VAS that are above each instrument's MID, but does not experience the same for the PEG or PHQ-9, then that participant falls into the 'Two MIDs exceeded' category. This classification corresponds to five categories, exceeding zero MIDs, exceeding one MID, exceeding two MIDs, exceeding three MIDs and exceeding four MIDs. This classification is performed for MIDs identified from previous research only. The percentage of participants in each of the five categories are visually presented in a stacked bar chart.

Additionally, an ordinal logistic regression is performed to determine the association between the collected variables with the probability of participants experiencing zero meaningful changes in HRQoL. Ordinal logistic regression is a method to determine the relationship of independent variables to a dependent ordered categorical variable. The major assumption underlying this approach is the proportional odds assumption or the parallel regression assumption. This assumes that the coefficients describing the relationship between the lowest versus all higher categories of the response variable are the same as those that describe the relationship between the next lowest category and all higher categories, and so on, for the remaining categories. The proportional odds assumption is assessed by comparing the coefficients from a series of binary logistic regressions with different cut-points on the dependent variable. Ordinal logistic regression can also be affected by sparse data. To assess sparsity, 2 × 2 tables of categorical independent variables and the dependent variable are tabulated. If sparsity is present, the "number of MIDs exceeded" (dependent variable) is split into fewer categories and, if necessary, new categorical variables are generated in a similar manner to the instrument-specific logistic regression models detailed earlier. The 2 × 2 tables of the independent variable and categorical dependent variables that show sparsity will be presented, along with 2 × 2 tables of the independent variable and new categorical variables.

Regression coefficients, p-values, odds ratios, and 95% confidence intervals of odds ratios of the ordinal logistic regression are reported.

4.7 Statistical Power and Sample Size

This study is based on recruiting and retaining participants scheduled for AA or TAR surgery from the population of these surgeries scheduled in VCH's hospitals. Power calculations could not be performed a priori since all patients were contacted to participate during the study period during the period of funding available – and lower extremity orthopedics' patients started being recruited later in the larger study's funded period. Therefore, as change in PROs was intended to be explanatory in nature, the study's sample size was not powered for specific changes in PROs.

However, ex poste power is calculated to provide insight into whether the results should detect changes in PROs values following surgery. The ex poste calculation assumes the difference between pre- and post-operative PRO scores – the effect size – is the MID, for each instrument respectively. The analysis assumes a two-sided alpha of 5%. Using the observed means and standard deviations, the minimum sample size to ensure power of 80% is 79 participants for the EQ-5D VAS. Other instruments required smaller sample sizes to achieve 80% power.

This study's sample size of 89 is highly likely to detect changes as least as large as the MIDs for each of the PROs used in this study. Power calculations were computed using R's 'power.t.test' command.

To give further strength to the findings of this work, previous publications report significant differences in PROs with far smaller datasets with similar demographic compositions (30,81,82). This lends additional support to the reliability of our sample size's ability to detect meaningful changes in HRQoL as measured by changes in PROs.

There are some limitations to this analysis of power; the analyses are based on each PRO individually and do not account for correlation between PROs. However, the multivariate analyses adjust for sources of variation thought to be associated with participant's outcomes, so this factor may increase power to some extent.

4.8 Statistical Software

Analysis, including demographic information and regressions was carried out in SAS 9.4 (SAS Institute Inc., Cary, NC, USA) and R 3.5.3 "Great Truth" (159). Figures were produced using R 3.5.3 and the ggplot2 package (160).

Chapter 5: Results

This chapter is split into five subsections. Subsection 5.1 summarizes the sample's demographic and clinical variables, wait times and post-operative physiotherapy utilization. Subsection 5.2 presents pre- and post-operative scores of each instrument and includes an investigation on the percentage of participants that improve above the MID. Relationships between changes in different PROs are shown in subsection 5.3. The effects of the collected variables on changes in HRQoL are examined in subsection 5.4. This chapter will conclude with subsection 5.5, which explores the effects of the collected variables on exceeding MIDs of the various instruments.

5.1 Demographics

The unadjusted demographic and clinical variables of the cohort are summarized in Table 1. There were 190 patients eligible to participate. Among the 190 eligible patients, participation was 60.0% leaving 114 participants that completed the preoperative PROs. Study dropout was 22% and 89 participants completed the postoperative PROs.

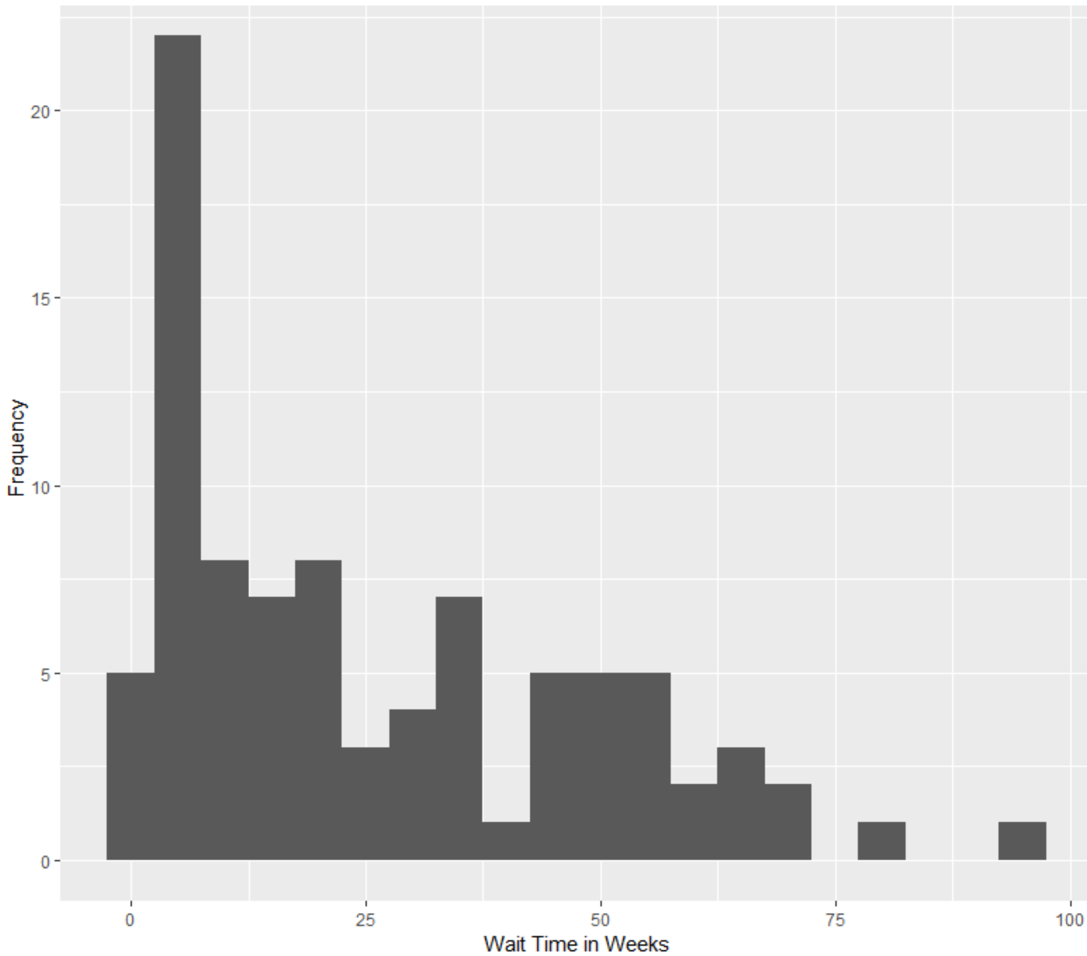
There was no difference between participants and non-participants based on sex. Participants were, on average, three years older than non-participants ($p < 0.01$.) No other demographic characteristics were observable on non-participants, a topic developed further in the study's limitations.

Descriptive statistics of the characteristics of the sample are provided in Table 1. Participants had a mean age of 60 years. Most participants were female (60.7%) and did not have other chronic health conditions (88.8%) reported in their hospital discharge. AA was more common (68.5%) than TAR (31.5%). Participants were distributed evenly across age and SES categories. Most participants received surgery within 26 weeks. A further breakdown of wait times is shown in Figure 1, showing that a plurality of participants received surgery in the first 12 weeks on the waitlist. However, a non-trivial number of participants waited over 26 weeks for their surgery. Only 9% of participants received post-operative physiotherapy in the 180 days following their surgery. This measure of physiotherapy only includes participants who received publicly-funded physiotherapy, as privately-funded physiotherapy could not be included.

Table 1 Cohort Demographics

Characteristic	N	%
Overall	*89*	*100.0*
Age (mean, *sd*)	60	(10.2)
Age Category		
≤ 50	16	18.0
51 – 60	32	36.0
61 – 70	26	29.2
> 70	15	16.9
Sex		
Female	54	60.7
Male	35	39.3
Charlson Index		
0	79	88.8
≥ 1	10	11.2
SES		
1 (Highest)	13	14.6
2	20	22.5
3	17	19.1
4	11	12.4
5 (Lowest)	14	15.7
Missing	14	15.7
Surgery Category		
AA	61	68.5
TAR	28	31.5
Wait Time in Weeks		
0 - 12	34	38.2
13 - 26	19	21.3
> 26	36	40.4
Post-Operative Physio Visits		
0	81	91.0
≥ 1	8	9.0

Figure 1 Histogram of wait time in weeks



5.2 Results of Comparisons between Changes in HRQoL and MIDs

5.2.1 EQ-5D VAS

Table 2 displays information for the EQ-5D VAS. The overall unadjusted mean pre-operative EQ-5D VAS score was 65.4, while the post-operative mean score was 72.0. These means were significantly different (p-value < 0.05).

As the MIDs identified from previous research and generated from this cohort were almost the same (7.0 vs. 7.06), only percentages of participants that improved above the MID identified from previous research were shown.

Nearly 43% of participants experienced a change in EQ-5D VAS score that was above the MID. This finding suggests that although most participants see *statistically* significant changes in general HRQoL, less than half of participants found these changes to be meaningful. Considering that most participants expect benefits following surgery that are noticeable, such as improved ability to walk and decreased pain (77), this finding was counterintuitive.

For participants in the 50 years and younger age category, paired t-tests revealed a statistically significant difference between mean pre- and post-operative EQ-5D VAS scores (p < 0.05). Additionally, 56.3% of participants in this age category experienced changes in score that exceeded the instrument's MID. Participants in other age categories did not see a statistically significant difference in mean pre- and post-operative scores, although it was close to significant for participants in the 61-70 category (p-value of 0.077). Participants in the 50 years and under age category experienced both worse (lower) mean pre-operative and better (higher) mean post-operative EQ-5D VAS scores than most other age categories. This suggests that younger participants had worse general HRQoL pre-operatively and saw a larger change in general HRQoL.

Both sexes had similar mean pre- and post-operative scores (66.2 and 71.6 for females, 64.2 and 72.5 for males, respectively). The difference between mean pre- and post-operative scores was statistically significant for males, and almost statistically significant in females (p-value = 0.058). A similar percentage of participants in both sex categories exceed the EQ-5D VAS MID (40.7% for females, 45.7% for males). These finding suggest that changes in general HRQoL after AA or TAR do not differ between the sexes.

Among participants with a Charlson index greater than or equal to one, 50% experienced changes in EQ-5D VAS scores that were meaningful. Participants with a Charlson index greater than or equal to one had lower mean pre-operative general HRQoL and saw more meaningful changes in general HRQoL compared to participants with Charlson indexes of zero. These findings are at odds with previous research, which has found patients with higher Charlson indexes often have worse post-operative outcomes following orthopedic surgery.

Changes in general HRQoL varied widely across different SES categories. Mean pre- and post-operative EQ-5D VAS scores were only found to be statistically different for participants in the 3rd SES quintile, and just over 70% of these participants experienced changes that were above the instrument's MID. Participants in the highest and second highest SES categories had a higher mean pre-operative general HRQoL than those in the third, fourth, fifth, or missing categories. However, mean post-operative scores were lower for participants in the highest and second highest SES categories compared to those in the third, fifth, and missing SES categories. These findings suggest that participants with higher SES have better pre-operative general HRQoL, but do not improve as much as those with lower SES. Literature on SES supports this trend, suggesting that participants in higher SES categories tend to have better health than participants in lower SES categories (161). It is possible that the lower SES participants, having worse

HRQoL than their higher SES peers, achieve greater changes in their general HRQoL in the peri-operative period. However, due to the small sample sizes across SES categories, these findings should be interpreted with caution.

The 53.6% of participants that received TAR experienced a change in EQ-5D VAS score that they found to be meaningful. This is compared to 37.7% of participants that received AA who experienced a change in EQ-5D VAS score that was meaningful. However for both AA and TAR participants, the differences between pre- and post-operative EQ-5D VAS scores were found to be statistically significant. These findings provide additional context to the work by Daniels et al. (37), which reported that differences between pre- and post-operative general HRQoL were similar for AA and TAR patients. The varying percentage of participants experiencing changes above the instrument's MID across AA and TAR categories suggests that the changes in general HRQoL attributable to AA may not be as important to participants as changes attributable to TAR. This appears reasonable, as AA reduces the ankle's range of motion compared to TAR, which could lead to less AA participants finding their changes to be meaningful.

Participants with wait times of 12 weeks or under had better mean pre-operative general HRQoL (score of 68.5) compared to participants who waited between 13 and 26 weeks and participants who waited over 26 weeks (60.1 and 65.4 respectively). However, 52.6% of participants who waited between 13 and 26 weeks for surgery experienced meaningful changes in EQ-5D VAS scores, compared to 32.4% and 47.2% of participants who waited 12 weeks or under and over 26 weeks respectively. It is unclear why this is the case.

Participants who received post-operative physiotherapy in the 180 days following surgery had similar mean pre-operative EQ-5D VAS scores to participants who did not receive physiotherapy. However, participants who received post-operative physiotherapy had lower mean EQ-5D VAS scores relative to participants who did not receive post-operative physiotherapy. Among participants who did not receive outpatient physiotherapy, 44% experienced meaningful changes compared to only 25% of participants who utilized post-operative physiotherapy.

It is possible that participants that utilized post-operative physiotherapy experienced selection bias, wherein participants with poorer prognosis are referred to post-operative physiotherapy at higher rates. For example, these participants could have experienced complications following surgery and utilized physiotherapy as a means of mitigating these complications, though this study design did not provide an opportunity for participants' qualitative assessment of their access to physiotherapy. As this study only captured physiotherapy utilization that was publicly funded, it is possible that these participants were of lower SES categories. However, upon inspection of 2 x 2 tables of physiotherapy categories and SES, this did not appear to be the case.

Table 2 Summarized pre- and post-operative EQ-5D VAS scores and percentage of participants exceeding MID

Characteristic	EQ-5D VAS			
	Pre-Operative Mean (SD)	Post-Operative Mean (SD)	Difference P-Value	Percent Exceeding MID (Literature)
Overall	65.4 (18.8)	72.0 (16.2)	0.003	42.7
Age Category				
≤ 50	57.1 (17.1)	74.0 (16.0)	0.002	56.3
51 – 60	65.9 (16.3)	67.7 (16.1)	0.645	37.5
61 - 70	69.0 (18.7)	76.3 (14.6)	0.077	38.5
> 70	66.9 (24.5)	71.5 (18.5)	0.309	46.7
Sex				
Female	66.2 (18.3)	71.6 (16.8)	0.058	40.7
Male	64.2 (19.6)	72.5 (15.4)	0.019	45.7
Charlson Index				
0	66.9 (17.9)	72.6 (16.3)	0.016	41.8
≥ 1	53.7 (22.4)	67.6 (15.2)	0.047	50.0
SES				
1 (Highest)	71.5 (15.8)	70.0 (15.3)	0.759	23.1
2	68.9 (19.2)	70.8 (16.4)	0.887	40.0
3	59.1 (21.3)	72.8 (15.1)	0.003	70.6
4	62.1 (22.9)	68.7 (24.3)	0.212	36.4
5 (Lowest)	67.9 (18.2)	77.8 (11.8)	0.070	35.7
Missing	63.0 (14.3)	71.1 (15.6)	0.156	42.9
Surgery Flag				
AA	65.7 (18.3)	71.7 (16.6)	0.030	37.7
TAR	64.9 (20.2)	72.7 (15.5)	0.031	53.6
Wait Time in Weeks				
0 - 12	68.5 (14.3)	73.6 (14.5)	0.076	32.4
13 - 26	60.1 (24.1)	68.9 (20.4)	0.081	52.6
> 26	65.4 (19.0)	72.2 (15.3)	0.080	47.2
Post-Operative Physio Visits				
0	65.6 (17.9)	73.0 (15.8)	0.001	44.4
≥ 1	63.4 (27.1)	61.9 (17.3)	0.845	25.0

5.2.2 AOS

Table 3 summarizes findings regarding the AOS instrument. The mean pre-operative score for the AOS instrument was 60.5 while the mean post-operative scores was 31.7. The difference between mean pre- and post-operative AOS scores was statistically significant (p-value < 0.05). Many participants experienced a change in AOS that was greater than the instrument's MID, although this differed depending on the MID used (45.5% when the MID identified from previous research was considered, 65.9% for the cohort-calculated MID).

Participants in all age categories experienced statistically significant differences between mean pre- and post-operative AOS scores, only the 51-60 and greater than 70 age categories saw more than 50% of participants experience changes greater than the MID identified from previous research (71.4% and 62.5% respectively). On the other hand, all age categories saw more than 50% of participants experience changes greater than the MID calculated from the cohort.

Both females and males saw statistically significant differences between mean pre- and post-operative AOS scores. However, males had lower mean pre- and post-operative scores than females. Differences also persisted when comparing the percentage of participants that exceeded MIDs. For males, 53.9% of participants experienced a change in AOS score that was greater than the MID identified from previous research compared to 33.3% of females. This discrepancy between sexes was also reflected when comparing changes greater than the MID - 76.9% of males saw changes greater than that MID compared to 50% of females. This suggests that less females found their changes in ankle-specific HRQoL to be meaningful compared to males. There is evidence that suggests females expect to achieve complete improvement in ankle function following ankle surgeries at higher rates than men (78). This difference in expectation could play a role in the differences in exceeding the MID between sexes.

Mean pre- and post-operative AOS scores were similar for participants in both Charlson index categories. Differences between mean pre- and post-operative scores were statistically significant for participants in both Charlson index categories. A similar percentage of participants in both Charlson index categories experienced changes greater than the MID of the AOS identified from previous research (44.4 and 50.9%, respectively). This was also the case for the MID calculated from the cohort. These findings are contrary to previous research, which has found that patients with more comorbidities often have worse outcomes following orthopedic surgeries than patients with less comorbidities (139). However, as few participants had comorbidities, these findings need to be interpreted with caution.

Changes in AOS scores were similar across most SES categories. However, participants in the second-lowest SES category (quintile 4) saw worse mean post-operative scores than other SES categories. This translated to zero participants in this SES category improving above the MID identified from previous research. It was unclear why this was the case, though the MID calculated from the cohort suggests that 50% of these participants did experience meaningful changes in the AOS. Again, these results could be due to small sample sizes across SES groups.

Both AA and TAR participants saw statistically significant differences between mean pre- and post-operative AOS scores. Participants in both categories had similar mean pre- and post-operative AOS scores. However, the relevance of these changes to participants differed; 56.7% of AA participants experienced changes in the AOS greater than the instrument's MID identified from previous research compared to 21.4% of TAR participants. This pattern also appeared when applying the cohort-calculated MID, finding that 76.7% of AA participants improved above this threshold compared to 42.9% of TAR participants. These findings suggest that participants who received AA were more likely to experience meaningful changes in ankle-specific HRQoL compared to TAR participants. As TAR is associated with more range of motion than AA (32), and one might expect the opposite result.

Differences between mean pre- and post-operative AOS scores were statistically significant for participants in all three wait time categories. Participants in the 12 weeks or under wait time category had a lower mean pre-operative AOS score (53.2) compared to the other two wait time categories (67.4 and 63.1 respectively). Mean post-operative scores were similar between wait-time categories, with participants in the greater than 26 weeks category seeing lower mean AOS scores (27.7) compared to participants in the other two wait time categories (32.0 and 37.9). Participants in the greater than 26 weeks wait time category also experienced more changes in AOS score above the MID identified from previous research (50%) than participants in the other two wait time categories. When applying the cohort-calculated MID, all three wait times categories had a similar percentage of participants improved above the MID (68.8%, 60.0%, and 66.7% respectively).

Participants in both post-operative physiotherapy categories saw similar mean pre-operative AOS scores. For participants who did not utilize physiotherapy in the 180 days following surgery, the difference between mean pre- and post-operative scores was statistically significant. This was not the case for participants who utilized physiotherapy. Furthermore, 46.3% of participants who did not receive physiotherapy experienced changes in score greater than the MID identified from previous research, compared to 33.3% of participants who did receive physiotherapy. However, when applying the cohort-calculated MID, the percentage of participants that saw meaningful changes in the AOS was similar across both physiotherapy categories. This again suggests that the choice of MID was critical in detecting if a change in PRO score was meaningful.

Table 3 Summarized pre- and post-operative AOS scores and percentage of participants exceeding MID

Characteristic	AOS				
	Pre-Operative Mean (SD)	Post-Operative Mean (SD)	Difference P-Value	Percent Exceeding MID (Literature)	Percent Exceeding MID (Calculated from Cohort)
Overall	60.5 (19.7)	31.7 (20.2)	< 0.0001	45.5	65.9
Age Category					
≤ 50	66.3 (17.8)	43.6 (20.2)	0.005	25.0	62.5
51 – 60	64.5 (16.2)	28.9 (19.0)	< 0.0001	71.4	78.6
61 - 70	48.9 (24.6)	30.0 (21.6)	0.049	21.4	50.0
> 70	67.1 (10.8)	27.3 (19.6)	0.001	62.5	75.0
Sex					
Female	65.8 (13.3)	39.6 (20.8)	< 0.0001	33.3	50.0
Male	56.9 (22.7)	27.1 (18.8)	< 0.0001	53.9	76.9
Charlson Index					
0	60.6 (18.7)	32.7 (20.6)	< 0.0001	44.4	66.7
≥ 1	60.0 (26.1)	27.2 (19.1)	0.038	50.0	62.5
SES					
1 (Highest)	69.0 (14.6)	35.7 (14.9)	0.001	57.1	57.1
2	55.9 (16.0)	30.5 (23.2)	0.018	33.3	66.7
3	64.1 (21.6)	31.6 (20.6)	< 0.0001	66.7	88.9
4	66.9 (20.0)	49.1 (27.3)	0.120	0.0	50.0
5 (Lowest)	61.5 (16.3)	36.2 (16.1)	0.229	66.7	66.7
Missing	55.0 (24.5)	30.6 (17.8)	0.013	41.7	58.3
Surgery Flag					
AA	61.6 (18.3)	31.2 (19.7)	< 0.0001	56.7	76.7
TAR	58.1 (23.3)	33.2 (22.7)	0.027	21.4	42.9
Wait Time in Weeks					
0 - 12	53.2 (21.4)	32.0 (16.5)	0.001	43.8	68.8
13 - 26	67.4 (20.1)	37.9 (15.8)	0.009	40.0	60.0
> 26	63.1 (17.1)	27.7 (25.9)	< 0.0001	50.0	66.7
Post-Operative Physio Visits					
0	60.1 (19.9)	30.5 (19.8)	< 0.0001	46.3	65.9
≥ 1	66.3 (20.4)	45.0 (24.4)	0.343	33.3	66.7

5.2.3 PEG

Table 4 summarizes the pain instrument's findings. The pain MID described by Dworkin was nearly the same as the MID obtained from this cohort (1.0 compared to 1.09, respectively).

Overall, participants saw statistically significant differences between mean pre- and post-operative PEG scores and 64% of participants experienced meaningful changes in PEG scores, the highest among any of the instruments surveyed. Differences in mean pre- and post-operative PEG scores were statistically significant in all but one participant category.

A gradient in mean pre-operative pain was observed across age categories. Participants in the 50 years and under category had the highest mean pre-operative PEG scores, while participants in the older than 70 years category had the lowest mean pre-operative scores. This pattern was not maintained post-operatively. Mean post-operative scores and the percentage of participants that exceeded the MID varied across age categories with no clear pattern.

Mean pre- and post-operative PEG scores were similar between males and females. Likewise, participants in both sex categories saw statistically significant differences in pre- and post-operative PEG scores ($p < 0.05$). Although mean post-operative scores for males and females differed by 0.2 (Post-operative PEG scores of 3.6 in males, 3.4 in females), 74.3% of males experienced changes above the MID compared to only 57.4% of females. These results provide further evidence that a change in score that is statistically significant may not be meaningful to participants. This also suggests that fewer female participants experienced meaningful changes in pain compared to males.

Mean pre-operative PEG scores differed greatly between Charlson index categories - participants with a Charlson index greater than or equal to one had a mean pre-operative PEG score of 7.2, compared to a mean score of 5.7 for participants with a Charlson index of zero.

While participants in both Charlson index categories saw statistically significant differences in mean pre- and post-operative scores, the percentage of participants who found these changes meaningful varied; 60.8% of participants with a Charlson index of zero saw their mean PEG scores change in a meaningful way, compared to 90% of participants with a Charlson index greater than or equal to one. This finding was counterintuitive, as orthopedic patients with higher Charlson indexes have seen worse outcomes following surgery compared to those with lower Charlson indexes (139). However, the results here suggest that more participants with higher Charlson indexes experience meaningful changes in their pain status.

Mean pre- and post-operative PEG scores were similar across SES categories. The difference in mean pre- and post-operative scores was statistically significant for participants in the third SES category (SES quintile 3) only. The percentage of participants that saw changes in PEG scores that exceeded the MID varied across SES categories, with no clear pattern.

Both AA and TAR participants saw similar mean pre- and post-operative levels of pain as measured by the PEG. The difference between pre- and post-operative PEG scores was statistically significant for participants in both surgical categories. The percentage of participants that saw changes in scores that exceeded the PEG's MID was also similar across surgical category. This suggests that the ability of both surgeries to alleviate pain is similar.

Participants who waited between 13 and 26 weeks for surgery had higher mean pre-and post-operative PEG scores than participants in the other two wait time categories. The differences in mean pre- and post-operative scores was statistically significant for participants in all wait time categories. Additionally, the percentage of participants that saw meaningful changes in pain was similar across wait time categories. These findings suggest that time on the wait list does not affect the percentage of participants that experienced meaningful changes or whether these changes were statistically significant.

Participants who had at least one physiotherapy visit had lower mean pre-operative PEG scores and higher mean post-operative PEG scores than participants with no physiotherapy. While participants without physiotherapy saw differences in mean pre- and post-operative scores that were statistically significant, this was not the case for participants who had at least one physiotherapy visit. Furthermore, only 37.5% of participants that received physiotherapy saw changes in pain that exceeded the PEG's MID, compared to 66.7% of participants that did not receive physiotherapy. This implies that participants who received physiotherapy saw less statistically significant and meaningful changes in pain status over the peri-operative period.

Table 4 Summarized pre- and post-operative PEG scores and percentage of participants exceeding MID

	PEG			
Characteristic	Pre-Operative Mean (SD)	Post-Operative Mean (SD)	Difference P-Value	Percent Exceeding MID (Literature)
Overall	5.8 (2.5)	3.5 (2.5)	< 0.001	64.0
Age Category				
≤ 50	6.1 (2.7)	3.4 (2.6)	0.001	62.5
51 – 60	6.0 (2.4)	3.9 (2.3)	< 0.001	59.4
61 - 70	5.6 (2.4)	3.1 (2.7)	0.001	73.1
> 70	5.5 (3.0)	3.6 (2.5)	0.012	60.0
Sex				
Female	5.8 (2.4)	3.6 (2.4)	< 0.001	57.4
Male	5.8 (2.8)	3.4 (2.6)	< 0.001	74.3
Charlson Index				
0	5.7 (2.5)	3.5 (2.5)	< 0.001	60.8
≥ 1	7.2 (2.6)	4.1 (2.5)	0.014	90.0
SES				
1 (Highest)	5.5 (3.0)	3.7 (2.4)	0.009	61.5
2	5.3 (2.4)	3.6 (2.3)	0.023	55.0
3	6.5 (2.2)	3.8 (2.4)	< 0.001	76.5
4	6.4 (2.3)	3.5 (2.8)	0.003	63.6
5 (Lowest)	5.4 (3.0)	3.1 (3.3)	0.033	50.0
Missing	6.1 (2.5)	3.4 (2.1)	0.004	78.6
Surgery Flag				
AA	5.8 (2.6)	3.5 (2.5)	< 0.001	63.9
TAR	6.0 (2.5)	3.6 (2.6)	< 0.001	64.3
Wait Time in Weeks				
0 - 12	5.5 (2.6)	3.3 (2.2)	< 0.001	67.7
13 - 26	6.4 (2.6)	4.1 (2.5)	0.002	63.2
> 26	5.8 (2.5)	3.5 (2.8)	< 0.001	61.1
Post-Operative Physio Visits				
0	5.9 (2.6)	3.4 (2.5)	< 0.001	66.7
≥ 1	5.5 (2.6)	4.8 (2.3)	0.470	37.5

5.2.4 PHQ-9

Table 5 summarizes the statistics of participants' PHQ-9. The MID identified from previous research as described by Löwe et al. was much larger than the MID calculated using data from this cohort (5.0 compared to 1.74).

The difference between unadjusted mean pre- and post-operative PHQ-9 scores was statistically significant. Additionally, according to the MID identified from previous research, 22.5% of participants experienced a change in depression symptoms that was considered meaningful. There was no previously established evidence that would suggest participants treated with AA or TAR would see changes in depression symptoms, or that some participants would find these changes meaningful. However, when the cohort-calculated MID is applied, 52.8% of participants reported changes to be meaningful. This again shows the variability in clinical significance due to MID selection.

Mean pre-operative PHQ-9 scores were higher in participants 50 years and under compared to participants in other age categories. This could be related to the inability to perform daily functions or physical activities as a result of ESAA, which could leave younger participants with more depression symptoms than older participants. Participants in the youngest and oldest age categories also experienced mean pre- and post-operative PHQ-9 scores that differed in a statistically significant fashion. Mean post-operative PHQ-9 scores were similar to those of participants in other age categories. The percentage of participants in most age categories that experienced meaningful changes in the PHQ-9 were similar, except for participants in the 61 – 70 age category. Only 11% of participants in this age category saw meaningful changes in PHQ-9 scores as denoted by the MID identified from previous research, roughly half that of other age categories. This also persisted when considering the percentage of participants that exceeded the cohort-calculated MID. It was unclear why this was the case.

Mean pre-operative scores for males and females were similar (6.2 and 6.0). However, females saw higher mean post-operative PHQ-9 scores compared to males (4.9 compared to 3.8). The differences in mean pre- and post-operative PHQ-9 scores were statistically significant in males, and nearly significant in females (p-value of 0.058). Additionally, slightly more males experienced a change in PHQ-9 score greater than the instrument's MID identified from previous research than females (25.7% compared to 20.4%). When the cohort-calculated MID was considered, then slightly more females were noted to have had a meaningful change in PHQ-9 scores than males.

Participants with a Charlson index of zero had lower mean-preoperative PHQ-9 scores than participants with a Charlson index greater than or equal to one. However, participants in both Charlson index categories saw similar mean post-operative scores, and differences in pre- and post-operative scores were statistically significant across Charlson index categories. The percentage of participants with a Charlson index greater than or equal to one that saw changes in PHQ-9 scores which exceeded the MID identified from previous research was double that that of participants with a Charlson index of zero (40% compared to 20.3%). When the cohort-calculated MID was considered, the difference in the percentage of participants that saw a meaningful change in score was less pronounced.

Mean pre- and post-operative PHQ-9 scores differed across SES categories with no clear pattern. Additionally, the percentage of participants that saw a change in PHQ-9 greater than the MID identified from previous research was similar across most SES categories, with the notable exception of participants in the SES category four (quintile four). Zero of these participants improved above the MID sourced from previous research. However, when the cohort-calculated MID was applied, the percentage of participants that saw a meaningful change in PHQ-9 score in quintile four was similar to that of other SES quintiles. Again, these results should be interpreted with caution due to the small samples across SES categories.

AA and TAR participants had similar mean pre- operative PHQ-9 scores. However, mean post-operative scores differed between the two categories, with AA participants seeing a mean PHQ-9 score greater than the TAR participants (4.8 compared to 3.7, respectively). Participants in both surgical categories saw differences in mean pre- and post-operative PHQ-9 scores that were statistically significant. Furthermore, according to the MID identified from previous research, the percentage of participants that saw changes in the PHQ-9 that were meaningful was similar between surgical categories. However, more TAR participants saw changes in PHQ-9 scores above the cohort-calculated MID than AA participants (64.3% compared to 47.5%).

Participants that waited 12 weeks or less for surgery had a lower mean pre-operative PHQ-9 score than participants in the other two wait time categories. Post-operative PHQ-9 scores were similar across wait time categories. Additionally, the percentage of participants that saw changes in PHQ-9 scores above the MID identified from previous research was similar across wait time categories. This was also the case when the cohort-calculated MID was considered.

Mean pre-operative PHQ-9 scores were similar for participants that both received and did not receive post-operative physiotherapy in the 180 days following surgery. Mean post-operative scores were different between physiotherapy categories. Participants with no physiotherapy experienced differences in pre- and post-operative scores that were statistically significant. This was not the case for participants that received physiotherapy. Only 12.5% of participants that received physiotherapy experienced a change in score that was greater than the MID identified from previous research compared to 23.5% of participants that did not receive physiotherapy. When the cohort-calculated MID was considered, the percentage of participants that saw meaningful changes in depression symptoms was similar between physiotherapy groups.

Table 5 Summarized pre- and post-operative PHQ-9 scores and percentage of participants exceeding MID

	PHQ-9				
Characteristic	Pre-Operative Mean (SD)	Post-Operative Mean (SD)	Difference P-Value	Percent Exceeding MID (Literature)	Percent Exceeding MID (Calculated from Cohort
Overall	6.1 (5.7)	4.4 (4.4)	0.002	22.5	52.8
Age Category					
≤ 50	8.6 (6.9)	5.1 (5.3)	0.034	25.0	68.8
51 – 60	6.3 (5.3)	5.1 (4.8)	0.247	28.1	56.3
61 - 70	4.5 (5.0)	3.8 (3.1)	0.390	11.5	30.8
> 70	6.1 (5.6)	3.4 (4.3)	0.008	26.7	66.7
Sex					
Female	6.2 (5.6)	4.9 (4.8)	0.056	20.4	53.7
Male	6.0 (5.9)	3.8 (3.6)	0.008	25.7	51.4
Charlson Index					
0	6.0 (5.6)	4.5 (4.4)	0.010	20.3	51.9
≥ 1	7.3 (6.1)	4.3 (4.1)	0.033	40.0	60.0
SES					
1 (Highest)	5.4 (6.5)	3.7 (3.8)	0.265	38.5	46.2
2	7.4 (6.0)	4.2 (4.3)	0.014	30.0	70.0
3	7.6 (4.9)	4.8 (3.7)	0.005	23.5	64.7
4	5.3 (3.7)	5.3 (4.9)	1.000	0.0	54.6
5 (Lowest)	4.7 (5.8)	3.8 (5.3)	0.423	21.4	42.9
Missing	5.4 (6.5)	5.1 (4.8)	0.871	14.3	28.6
Surgery Flag					
AA	6.2 (5.9)	4.8 (4.7)	0.039	21.3	47.5
TAR	6.1 (5.1)	3.7 (3.7)	0.014	25.0	64.3
Wait Time in Weeks					
0 - 12	5.7 (5.5)	4.4 (5.0)	0.108	23.5	52.9
13 - 26	6.9 (7.2)	4.8 (4.2)	0.138	26.3	47.4
> 26	6.1 (5.0)	4.3 (3.9)	0.027	19.4	55.6
Post-Operative Physio Visits					
0	6.1 (5.7)	4.2 (4.2)	0.001	23.5	53.1
≥ 1	6.6 (6.0)	6.8 (6.0)	0.951	12.5	50.0

5.3 Changes in Different Aspects of HRQoL

Table 6 displays the Pearson correlation matrix of changes in PRO scores. Correlations are compared to Cohen's rules of thumb; correlations of 0.1 are small, 0.3 are medium, and 0.5 are large (162). We see that changes in many instruments were highly correlated. Changes in the EQ-5D VAS exhibited small negative correlations with the PHQ-9, and medium negative correlations with the AOS and PEG. The negative value of the correlations was due to the directionality of the different scales – the EQ-5D VAS attributes higher scores to better general HRQoL while the AOS, PEG, and PHQ-9 deem higher scores to be worse ankle-specific HRQoL, pain, and depression symptoms respectively. Thus, a positive change in EQ-5D VAS was an improvement. Conversely, negative changes in the AOS, PEG, and PHQ-9 would be considered improvements. When the change in EQ-5D VAS was considered, correlation between the PEG was the strongest (-0.464), followed by the AOS (-0.337) and PHQ-9 (-0.211). This suggests that the EQ-5D may have captured information that was also captured by these other instruments. Conceptually, general HRQoL should take all other aspects of HRQoL into account to some extent. The stronger association with changes in the PEG suggests that changes in pain were more associated to changes in general HRQoL than either ankle-specific HRQoL or depression symptoms.

Changes in the AOS exhibited strong positive correlations with changes in the PHQ-9 and PEG. This suggests that changes in pain and ankle-specific HRQoL were highly related, and that the instruments overlapped. The ankle-specific pain component of the AOS was most likely driving this result. The strength of the correlation between the AOS and the PHQ-9 could be explained by the pain-depression dyad. Since pain and depression are so strongly linked (100), and that ankle-specific HRQoL was strongly correlated to pain, then ankle-specific HRQoL could be related to depression through its relationship to pain.

Changes in the PEG exhibited medium positive correlations with the PHQ-9. This confirms previous research exploring the pain-depression dyad (99,100). However, the correlation between the PEG and PHQ-9 was weaker than the correlation between the AOS and the PHQ-9. This suggests that constructs measured by the AOS show additional association with the PHQ-9 over pain alone.

Additional insight into the differences in these correlations can be drawn from the scatterplots shown in Figure 2. Each point refers to a participant. These scatterplots illustrate how participants AOS, PEG, and PHQ-9 scores change over the peri-operative period. Participants closer to the upper-right corner were those with worse pain/ankle-specific HRQoL and more depression symptoms. We see in both sub-plots that more participants moved towards the lower-left corner, illustrating improvements in all three instruments. However, the movement of participants appeared to be more consistent in the AOS-PHQ-9 subplot compared to the PEG-PHQ-9 subplot, further illustrating the strong association in changes of the AOS and PHQ-9.

Table 6 Correlation matrix of changes in PRO scores

	VAS	AOS	PEG	PHQ-9
VAS	1	-0.33743	-0.46393	-0.21179
AOS	-0.33743	1	**0.732259**	**0.536154**
PEG	-0.46393	**0.732259**	1	0.36174
PHQ-9	-0.21179	**0.536154**	0.36174	1

Figure 2 Pre and post-operative AOS, PEG and PHQ-9 values

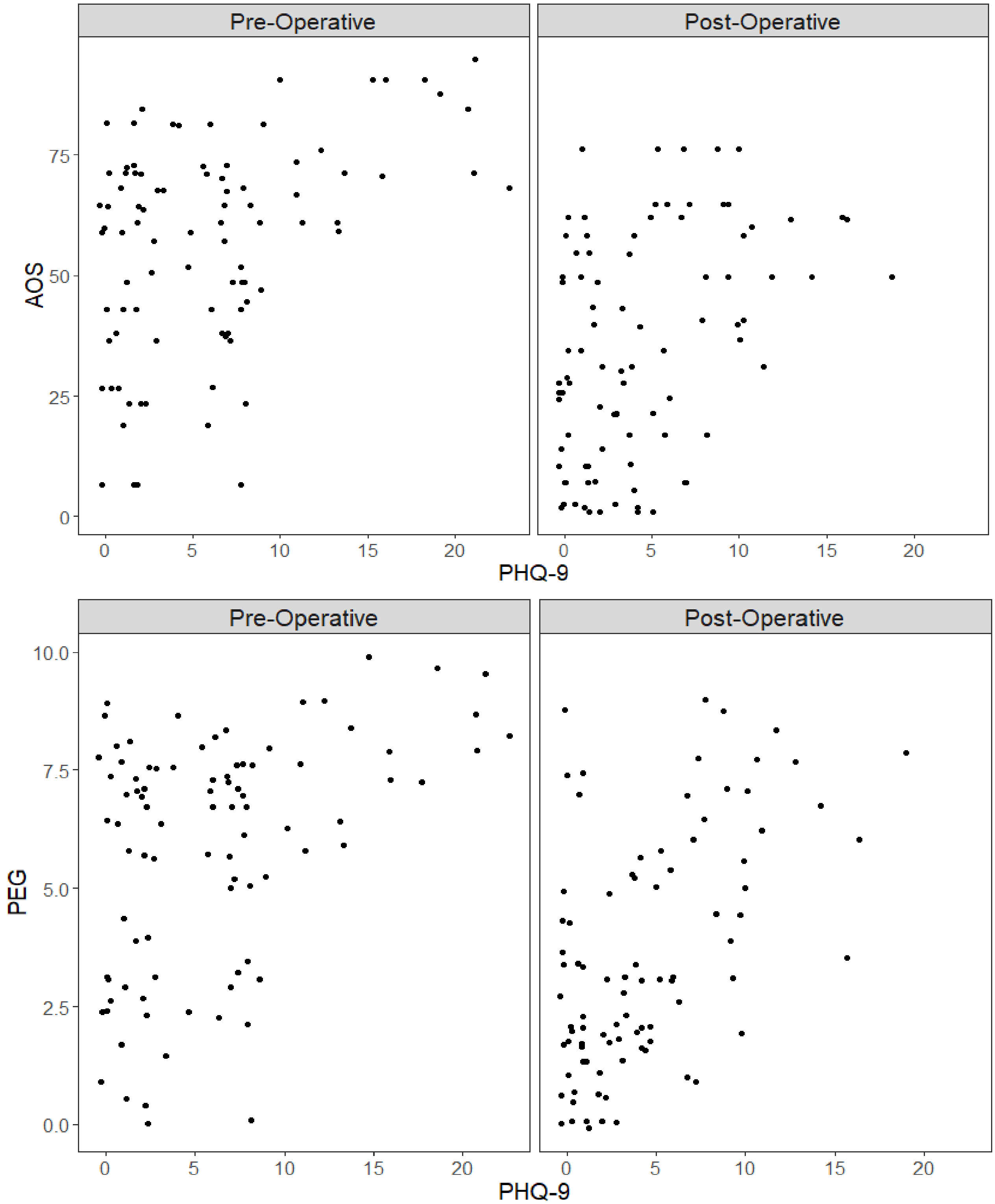


Figure 3 illustrates boxplots of pre- and post-operative scores for the four instruments used in this study. Dashed lines indicate a change in mean pre-operative score that exceeded the instrument’s MID identified from previous research. For the EQ-5D VAS subplot, values of the right-hand boxplot that were above the dashed line indicated improvements in general health that exceeded the instrument’s MID of 7.0. For the AOS, PEG, and PHQ-9 subplots, values in each right-hand boxplot that were below the dashed line indicated improvements in ankle-specific health, pain, and depression symptoms that were greater than each instruments MIDs of 28, 1, and 5, respectively.

The figure illustrates that the mean change in the PEG exceeded the MID by a significant margin. This supports the results of the PEG MID analysis from section 5.2.3, which found that most participants experienced a change in pain scores that was meaningful. The mean change of the AOS and EQ-5D VAS marginally exceeded their MIDs. These results also support the results from MID analysis for each instrument, as just over 40% of participants experienced changes in each PRO score above their respective MIDs. The mean change of the PHQ-9 did not exceed its MID, which was understandable given that under 25% of participants saw changes above this MID.

Figure 3 Side-by-side boxplots of pre- and post-operative PROs scores – dashed lines represent the threshold for a meaningful improvement

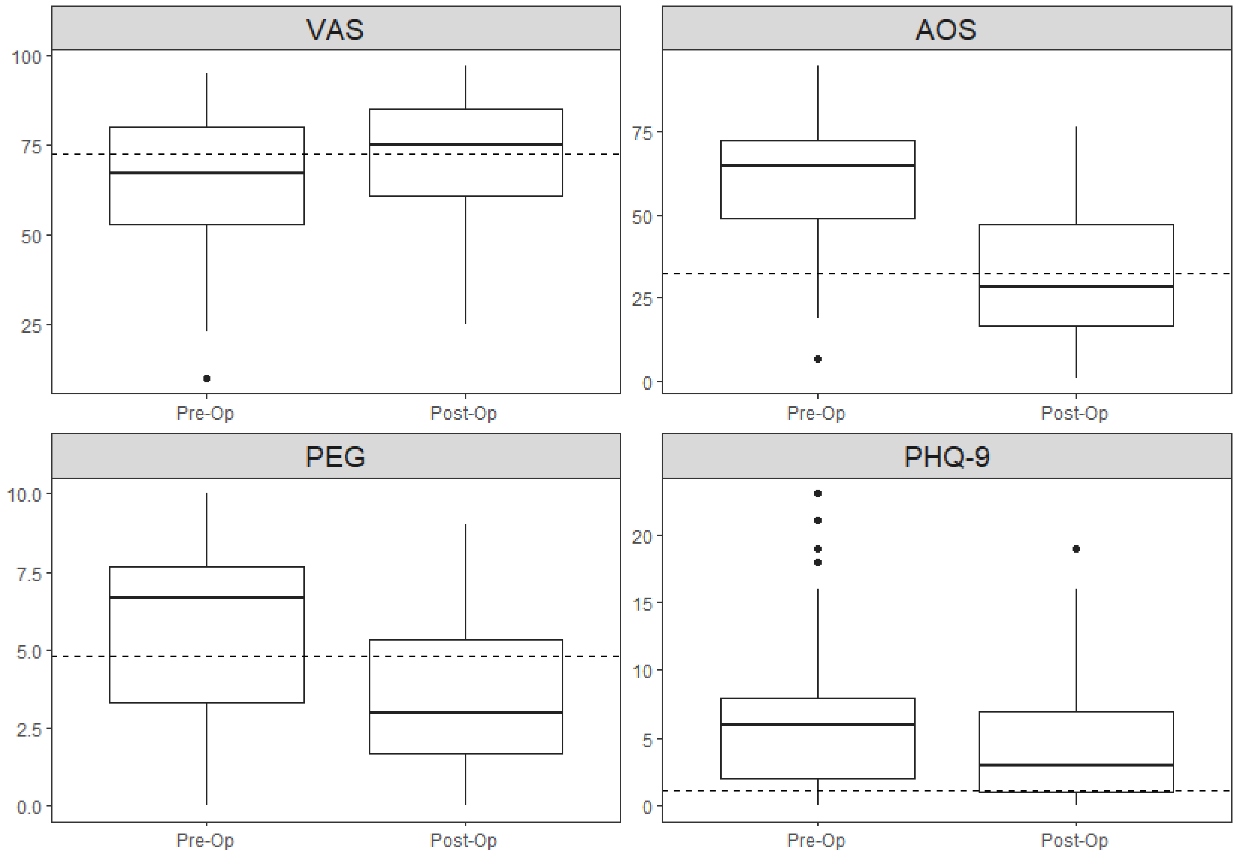
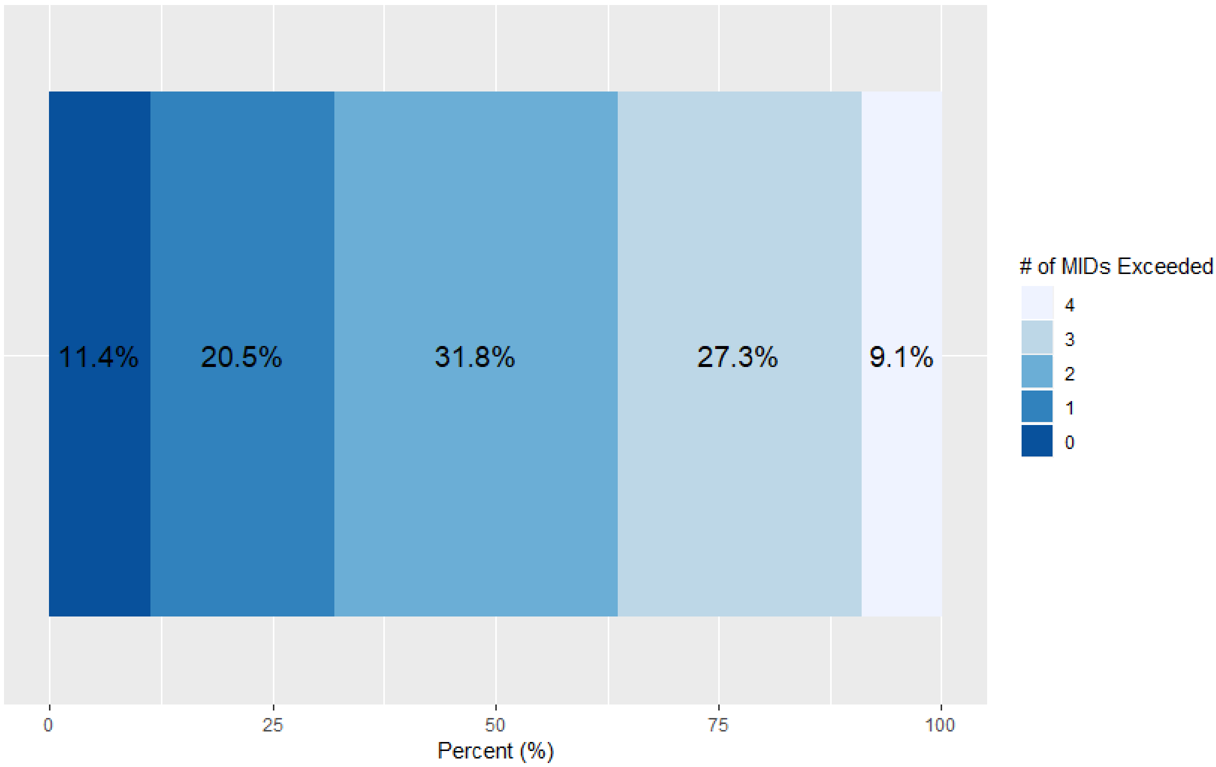


Figure 4 illustrates the percentage of the cohort that saw changes that exceeded multiple MIDs identified from previous research. The majority of participants exceeded at least one MID, and a plurality exceeded two. This could be due to the inter-relatedness of several of the instruments, as demonstrated previously. It is possible that if a participant saw a change in one PRO that exceeds its MID, they would also see a meaningful change in a closely related PRO.

It seems that for 11.4% of participants, no aspects of HRQoL improve in a meaningful way. It is possible that these participants saw decrements in their HRQoL during the peri-operative period, but this cannot be determined from the current study.

Figure 4 Percent of patients exceeding MIDs



5.4 Models Investigating Variables Associated with Changes in HRQoL

5.4.1 Changes in EQ-5D VAS Scores

Table 7 displays results from the regression analysis detailing the effects of clinical and demographic variables, wait times, and utilization of post-operative physiotherapy on changes in the EQ-5D VAS. Pre-operative PRO scores and the number of outpatient physiotherapy visits were significantly associated with changes in general HRQoL.

The regression coefficient associated with the number of post-operative physiotherapy visits was negative (-1.459). This implies that participants with more post-operative physiotherapy visits experienced smaller changes in general HRQoL as measured by the VAS. There were several reasons why this could be the case. One possibility is that participants with post-operative complications could be referred to physiotherapy more often than those without complications. However, as participants were not asked about their complications following surgery, this cannot be determined from the present study.

Age, sex, Charlson index, and SES were not found to be significantly associated with changes in score. This suggests that participant's demographic profile had little effect on changes in general HRQoL when the effects of pre-operative general HRQoL are considered.

The type of surgery and the wait time in weeks were not found to be associated with changes in EQ-5D VAS scores. This suggests that the surgery participants receive for treatment of ESAA and the length of time they spend on the wait list do not affect changes in general HRQoL as measured by the EQ-5D VAS when pre-operative health is accounted for. The findings also support the claim by Daniels et al. (37) that changes in general HRQoL do not differ between AA and TAR patients.

The sensitivity analysis using categorical variables for wait times and physiotherapy visits confirmed the above results, finding same variables to be significant.

Table 7 Regression analysis for EQ-5D VAS

Regression Variable	EQ-5D VAS	
	Coefficient (SE)	P-Value
Intercept	41.531 (11.813)	0.001
Number of Post-Operative Physio Visits	-1.459 (0.605)	0.019
Wait time	0.016 (0.075)	0.826
Pre-Operative PRO Score	-0.646 (0.091)	< 0.001
Age	0.023 (0.169)	0.896
Sex		
Female	Reference Category	
Male	1.483 (3.345)	0.662
Charlson Index		
0	Reference Category	
≥ 1	-0.141 (5.582)	0.972
SES		
1 (Highest)	Reference Category	
2	3.687 (5.498)	0.517
3	9.095 (5.762)	0.117
4	3.511 (6.477)	0.584
5 (Lowest)	9.752 (5.919)	0.101
Missing	5.199 (5.952)	0.379
Type of Surgery		
AA	Reference Category	
TAR	2.741 (3.635)	0.444

5.4.2 Changes in AOS Scores

Table 8 displays results from the regression analysis for the AOS. Pre-operative AOS scores was found to be the only variable associated with larger changes in AOS scores. Demographic variables, clinical variables, the number of physiotherapy visits, and time on the waitlist did not appear to be associated with changes in AOS. This differed to the EQ-5D VAS model, which found that post-operative physiotherapy was associated with changes in general HRQoL.

The lack of association with the type of surgery implies that changes in AOS scores are similar for both procedures. When this finding is considered alongside the findings of Table 3 (detailing the different percentages of participants that experienced meaningful changes between the two surgical groups), it provides evidence that although changes are statistically similar between AA and TAR participants, the significance of these changes to participants differs depending on the surgery they received.

The sensitivity analysis using categorical variables for wait times and physiotherapy visits confirmed the results of the presented regression model by finding the same variables to be significant.

Table 8 Regression analysis for AOS

	AOS	
Regression Variable	Coefficient (SE)	P-Value
Intercept	13.686 (21.717)	0.533
Number of Post-Operative Physio Visits	1.012 (0.838)	0.232
Wait time	0.034 (0.117)	0.770
Pre-Operative PRO Score	-0.567 (0.154)	0.001
Age	0.109 (0.279)	0.700
Sex		
Female	Reference Category	
Male	-8.017 (5.401)	0.145
Charlson Index		
0	Reference Category	
≥ 1	-5.684 (5.303)	0.288
SES		
1 (Highest)	Reference Category	
2	-13.182 (8.964)	0.149
3	-9.420 (8.959)	0.299
4	-4.749 (11.819)	0.691
5 (Lowest)	-16.071 (11.846)	0.187
Missing	-7.607 (11.723)	0.408
Type of Surgery		
AA	Reference Category	
TAR	4.436 (5.876)	0.455

5.4.3 Changes in PEG Scores

Table 9 displays the results from the regression analysis for the PEG. Higher pre-operative PEG scores – more pain – were associated with larger changes in PEG scores. Demographic variables, type of surgery, number of post-operative physiotherapy visits, and time on the wait list were not found to be associated with changes in PEG scores.

The sensitivity analysis using categorical variables for wait times and physiotherapy visits also found that pre-operative PEG scores were the only significant variable.

Table 9 Regression analysis for PEG

	PEG	
Regression Variable	Coefficient (SE)	P-Value
Intercept	2.718 (1.876)	0.152
Number of Post-Operative Physio Visits	0.147 (0.098)	0.139
Wait time	-0.005 (0.012)	0.684
Pre-Operative PRO Score	-0.623 (0.105)	< 0.001
Age	-0.015 (0.027)	0.585
Sex		
Female	Reference Category	
Male	-0.186 (0.546)	0.735
Charlson Index		
0	Reference Category	
≥ 1	0.022 (0.643)	0.972
SES		
1 (Highest)	Reference Category	
2	-0.140 (0.894)	0.876
3	-0.511 (0.920)	0.580
4	-0.806 (1.027)	0.435
5 (Lowest)	-0.619 (0.950)	0.516
Missing	-0.662 (0.950)	0.488
Type of Surgery		
AA	Reference Category	
TAR	0.030 (0.583)	0.959

5.4.4 Changes in PHQ-9 Scores

Table 10 shows the results from the regression analysis of the PHQ-9. Like the EQ-5D VAS, AOS, and PEG, pre-operative PHQ-9 scores were associated with changes in the instrument's score. Similar to the AOS and PEG, demographic variables, type of surgery, number of physiotherapy visits, and time on the surgical wait list were not found to be associated with changes in the PHQ-9 score.

The sensitivity analysis using categorical variables for wait times and physiotherapy visits also found pre-operative PHQ-9 scores to be the only significant variable.

Table 10 Regression analysis for PHQ-9

	PHQ-9	
Regression Variable	Coefficient (SE)	P-Value
Intercept	1.605 (2.861)	0.577
Number of Post-Operative Physio Visits	0.234 (0.149)	0.122
Wait time	-0.004 (0.018)	0.830
Pre-Operative PRO Score	-0.560 (0.074)	< 0.001
Age	0.013 (0.043)	0.763
Sex		
Female	Reference Category	
Male	-1.292 (0.831)	0.124
Charlson Index		
0	Reference Category	
≥ 1	-0.164 (0.972)	0.867
SES		
1 (Highest)	Reference Category	
2	-0.867 (1.373)	0.530
3	-0.352 (1.399)	0.802
4	1.293 (1.547)	0.406
5 (Lowest)	0.288 (1.441)	0.842
Missing	1.527 (1.439)	0.292
Type of Surgery		
AA	Reference Category	
TAR	-1.277 (0.885)	0.153

5.4.5 Changes in Multiple PRO Scores

The results from the MANCOVA model are displayed in Table 11. Prior to generating this model, correlations between the independent variables (demographic and clinical variables, wait times, physiotherapy utilization, pre-operative PRO scores) were evaluated. No interaction terms were found to be statistically significant, meaning that there is evidence suggesting that the regression slopes would not differ in the MANCOVA analysis. Visual inspection of plots between factors and continuous variables did not find evidence of relationships between these variables.

Covariance matrices of dependent variables found that covariances were heterogenous, most likely due to differing scales across PROs. To account for this, each dependent variable was z-score standardized before input into the model. Additionally, pre-operative PRO scores were also z-score standardized, as they are used in calculating changes in PRO scores.

It was found that pre-operative AOS scores and pre-operative PEG scores were highly correlated (Pearson correlation of 0.749). The pre-operative AOS score was selected to be dropped, as both pre-operative PEG and AOS scores had similar correlations to the other instruments.

The effects of pre-operative EQ-5D VAS scores were strongly associated with changes in all other instruments. No other pre-operative PRO score was associated with changes in the instruments investigated when the EQ-5D VAS was considered. This finding is intuitive as the EQ-5D VAS is a measure of general HRQoL. It is not a surprise that other aspects of HRQoL would be summarized by this construct.

The MANCOVA appeared to echo results from the previous regression models; that pre-operative HRQoL was significantly associated with changes in HRQoL. Overall, it appeared that, even with the additional power afforded by the MANCOVA, demographic variables and type of surgery did not have an effect on changes in PROs scores after taking the effect of pre-operative scores into account.

The sensitivity analysis using categorical variables for wait times and physiotherapy visits found the same variables to be significant. Details can be found in Table 22 in Appendix A

Table 11 MANCOVA results

Regression Variable	Estimate	P value
Intercept	-0.068	0.955
Pre-Operative PRO Score		
EQ-5D VAS	-0.740	0.001
PEG	-0.059	0.773
PHQ-9	-0.283	0.165
Number of Post-Operative Physio Visits	-0.069	0.289
Wait Time	0.002	0.796
Age	-0.005	0.772
Sex		
Female	Reference Category	
Male	0.105	0.769
Charlson Index		
0	Reference Category	
≥ 1	0.022	0.959
SES		
1 (Highest)	Reference Category	
2	0.218	0.715
3	0.540	0.379
4	0.152	0.824
5 (Lowest)	0.463	0.461
Missing	0.225	0.721
Surgery		
AA	Reference Category	
TAR	0.156	0.683

5.5 Models Investigating Variables Associated with Meaningful Changes in HRQoL

5.5.1 Instrument-Specific Logistic Regression Results

Sparsity was present in the 2 × 2 tables of SES categories and exceeding the AOS and PHQ-9 MIDs (shown in Table 12 and Table 14). To prevent this sparsity from affecting the models, SES categories were combined to form four new categories: SES quintile one, SES quintiles two and three, SES quintiles four and five, and missing SES, for all instrument-specific logistic regression models. The 2 × 2 tables of these new SES categories and exceeding AOS and PHQ-9 MIDs are shown in Table 13 and Table 15.

Table 16 displays the results from the logistic regression models of exceeding the EQ-5D VAS and AOS MIDs while Table 17 does the same for the logistic regression models of exceeding the PEG and PHQ-9 MIDs. For the EQ-5D VAS model, participants in the second or third SES quintile saw an odds of exceeding the MID that was 6.991 times as large as that of participants in the first SES quintile. It is unclear why this is the case, as previous research has indicated that patients in highest SES categories have better pre- and post-operative HRQoL scores (141). It is possible that this was due to sparsity of the data still present in the data. No other variables were associated with improving above the EQ-5D VAS MID.

For the AOS model, males had an odds of exceeding the AOS MID that was 6.068 times greater than that of females. This could be due to differences in expectations between the sexes, which have been documented (78). It is also possible that differences are attributable to sex-based differences in bone density, as females are more likely to develop less dense bones and suffer from osteoarthritis than males (163,164). It is unlikely that this was due to sparsity of the data. Participants who underwent AA had an odds of exceeding the AOS MID that was 4.83 times greater than that of participants who received TAR. This counters previous evidence which suggests that the changes in HRQoL are similar between the two procedures (37,84). No other variables were associated with improving above the AOS MID.

No variables were associated with exceeding the MID of either the PEG or the PHQ-9.

Table 12 Two-by-two table of exceeding AOS MID and old SES categories

	1	2	3	4	5	Missing
Did not Exceed AOS MID	9	17	11	11	12	9
Exceeded AOS MID	4	3	6	0	2	5

Table 13 Two-by-two table of exceeding AOS MID and new SES categories

	1	2-3	4-5	Missing
Did not Exceed AOS MID	9	28	23	9
Exceeded AOS MID	4	9	2	5

Table 14 Two-by-two table of exceeding PHQ-9 MID and old SES categories

	1	2	3	4	5	Missing
Did not Exceed PHQ-9 MID	8	14	13	11	11	12
Exceeded PHQ-9 MID	5	6	4	0	3	2

Table 15 Two-by-two table of exceeding PHQ-9 MID and new SES categories

	1	2-3	4-5	Missing
Did not Exceed PHQ-9 MID	9	28	23	9
Exceeded PHQ-9 MID	4	9	2	5

Table 16 Logistic regression results for exceeding EQ-5D VAS and AOS MIDs

	VAS				AOS			
Regression Variable	Estimate	P-value	Odds Ratio	OR 95% CI	Estimate	P-value	Odds Ratio	OR 95% CI
Intercept	-0.703 (1.648)	0.670	n/a		-2.740 (2.159)	0.204	n/a	
Age	-0.015 (0.024)	0.519	0.985	(0.939, 1.032)	-0.005 (0.032)	0.871	0.995	(0.934, 1.060)
Physio Category								
0		Reference Category				Reference Category		
≥ 1	-1.785 (1.017)	0.079	0.168	(0.017, 1.042)	-1.066 (1.284)	0.406	0.344	(0.015, 3.448)
Wait Time in Weeks								
0 - 12		Reference Category				Reference Category		
13 - 26	1.190 (0.651)	0.068	3.286	(0.940, 12.376)	-0.196 (0.821)	0.811	0.822	(0.151, 3.998)
> 26	0.919 (0.549)	0.094	2.506	(0.873, 7.639)	0.218 (0.666)	0.743	1.244	(0.334, 4.728)
Sex								
Female		Reference Category				Reference Category		
Male	0.295 (0.494)	0.550	1.343	(0.510, 3.585)	1.803 (0.616)	0.003	6.068	(1.899, 21.864)
Charlson Index								
0		Reference Category				Reference Category		
≥ 1	0.314 (0.550)	0.568	1.369	(0.463, 4.311)	0.690 (0.684)	0.313	1.994	(0.531, 8.438)
SES								
1 (Highest)		Reference Category				Reference Category		
2-3	1.945 (0.812)	0.017	6.991	(1.575, 40.537)	-0.099 (0.830)	0.905	0.906	(0.180, 4.967)
4-5	0.961 (0.831)	0.248	2.614	(0.548, 15.350)	-1.748 (1.049)	0.096	0.174	(0.018, 1.270)
Missing	1.078 (0.911)	0.237	2.939	(0.514, 19.697)	0.110 (0.953)	0.908	1.117	(0.170, 7.632)
Surgery								
TAR		Reference Category				Reference Category		
AA	-0.917 (0.532)	0.085	0.400	(0.136, 1.115)	1.575 (0.813)	0.053	4.830	(1.137, 30.253)

Table 17 Logistic regression results for exceeding PEG and PHQ-9 MIDs

	PEG				PHQ-9			
Regression Variable	Estimate	P-value	Odds Ratio	OR 95% CI	Estimate	P-value	Odds Ratio	OR 95% CI
Intercept	-0.479 (1.654)	0.065	n/a		1.796 (1.895)	0.343	n/a	
Age	0.010 (0.024)	0.664	1.010	(0.964, 1.059)	-0.040 (0.029)	0.166	0.961	(0.906, 1.016)
Physio Category								
0		Reference Category				Reference Category		
≥ 1	-1.482 (0.872)	0.089	0.227	(0.368, 6.101)	-0.840 (1.242)	0.499	0.432	(0.018, 3.471)
Wait Time in Weeks								
0 - 12		Reference Category				Reference Category		
13 - 26	-0.268 (0.628)	0.669	0.765	(0.222, 2.675)	-0.015 (0.714)	0.984	0.985	(0.229, 3.932)
> 26	-0.184 (0.536)	0.731	0.832	(0.287, 2.394)	-0.222 (0.625)	0.723	0.801	(0.228, 2.738)
Sex								
Female		Reference Category				Reference Category		
Male	0.666 (0.498)	0.181	1.946	(0.748, 5.349)	0.487 (0.561)	0.386	1.627	(0.536, 4.968)
Charlson Index								
0		Reference Category				Reference Category		
≥ 1	0.902 (0.774)	0.244	2.464	(0.655, 16.102)	0.657 (0.591)	0.266	1.929	(0.565, 6.133)
SES								
1 (Highest)		Reference Category				Reference Category		
2-3	0.423 (0.705)	0.548	1.527	(0.655, 16.102)	-0.297 (0.716)	0.678	0.743	(0.184, 3.169)
4-5	0.011 (0.735)	0.988	1.011	(0.368, 6.101)	-1.460 (0.869)	0.093	0.232	(0.037, 1.228)
Missing	0.942 (0.915)	0.303	2.565	(0.230, 4.282)	-1.574 (1.019)	0.123	0.207	(0.022, 1.361)
Surgery								
TAR		Reference Category				Reference Category		
AA	0.091 (0.534)	0.865	1.095	(0.376, 3.112)	-0.238 (0.608)	0.695	0.788	(0.242, 2.722)

5.5.2 Ordinal Logistic Regression Results

Sparsity was present across most categorical variables when the "number of MIDs exceeded" was categorized as zero, one, two, three, or four. These five categories were collapsed into three (zero, one to two, three to four MIDs exceeded). Even after this re-categorization, sparsity remained among some SES categories. SES was then summarized into a new set of categories (highest quintile, 2^{nd} and 3^{rd} highest quintile, 4^{th} and 5^{th} quintiles, missing). The corresponding 2×2 tables of are shown below in Table 18, Table 19, and Table 20

Results from the ordinal logistic regression are displayed in Table 21. Two variables were associated with participants experiencing zero meaningful changes in HRQoL – physiotherapy utilization and sex. Participants who utilized physiotherapy had an odds of experiencing zero meaningful changes in HRQoL that was 9.025 times greater than that of participants who did not utilize physiotherapy. This relationship aligned with results from the tables in section 5.2. The magnitude of the relationship could be due to the small sample size of participants who utilized physiotherapy (8 out of 89 or 9% of the cohort).

Female participants had an odds of experiencing zero meaningful changes in HRQoL that was 3.513 times greater than that of males. While females have higher expectations of expecting complete recovery following foot and ankle surgeries (78), this result could be driven by differences in bone and joint health between the sexes (163,164).

No other variable was associated with an increased odds of seeing zero meaningful changes in the four aspects of HRQoL investigated in this study.

Table 18 Two-by-two table of exceeding multiple MIDs (five categories) and old SES categories

	SES Categories					
Number off MIDs Exceeded	1	2	3	4	5	Missing
0	4	6	2	4	5	2
1	3	4	2	3	3	4
2	2	6	7	4	5	5
3	3	4	5	0	0	2
4	1	0	1	0	1	1

Table 19 Two-by-two table of exceeding multiple MIDs (three categories) and old SES categories

	SES Categories					
Number off MIDs Exceeded	1	2	3	4	5	Missing
0	4	6	2	4	5	2
1-2	5	10	9	7	8	9
3-4	4	4	6	0	1	3

Table 20 Two-by-two table of exceeding multiple MIDs (three categories) and new SES categories

	SES Categories			
Number off MIDs Exceeded	1	2-3	4-5	Missing
0	4	8	9	2
1-2	5	19	15	9
3-4	4	10	1	3

Table 21 Results from ordinal logistic regression

Regression Variable	Estimate	P-Value	Odds Ratio	OR 95% CI
Intercept				
3-4|1-2	-0.540	0.714	n/a	
1-2|0	2.446	0.104	n/a	
Age	0.012	0.591	1.012	(0.970, 1.056)
Physio Category				
0		Reference Category		
≥ 1	2.200	0.011	9.025	(1.789, 55.683)
Wait time in Weeks				
0 - 12		Reference Category		
13 - 26	0.042	0.944	1.043	(0.320, 3.394)
> 26	-0.298	0.538	0.742	(0.285, 1.918)
Sex				
Male		Reference Category		
Female	1.256	0.007	3.513	(1.428, 9.099)
Charlson Index				
0		Reference Category		
≥ 1	-1.195	0.053	0.303	(0.084, 0.982)
SES				
1 (Highest)		Reference Category		
2-3	-0.718	0.286	0.488	(0.128, 1.828)
4-5	0.610	0.385	1.841	(0.465, 7.467)
Missing	-0.242	0.762	0.785	(0.160, 3.763)
Surgery				
AA		Reference Category		
TAR	-0.310	0.519	0.734	(0.283, 1.882)

Chapter 6: Discussion

The results of this study demonstrate that participants with end-stage ankle arthritis treated by ankle arthrodesis or total ankle replacement saw statistically significant changes in all aspects of HRQoL in the peri-operative period. However, not all these changes were large enough to be considered meaningful to all participants, indicating a discrepancy between statistical significance and meaning/significance to participants.

Most participants experienced improvements in pain (as measured by the PEG) that were large enough to be considered meaningful. Almost one half of participants saw changes in AOS (ankle-specific HRQoL) and EQ-5D VAS (general HRQoL) scores that exceeded the MIDs sourced from previous research. Furthermore, 22.5% of participants also saw changes in the PHQ-9 that exceeded its MID. This could be due to the strong association between changes in PHQ-9 scores and changes in PEG and AOS scores.

The choice of MID impacted the percentage of participants that were noted as having a meaningful change in HRQoL, providing support for the inclusion of multiple MID estimates. Further investigation is required to determine additional MID estimates for ESAA patients treated with AA and TAR.

Changes in different aspects of HRQoL were found to be highly correlated, and pre-operative scores of some instruments were related with changes in others, suggesting that many of these instruments are measuring similar concepts. This could be driven by specific aspects of each questionnaire, like the EQ-5D having a domain for anxiety and depression, and the AOS having a domain for ankle-specific pain. Notably, changes in the AOS were correlated with changes in the PEG and PHQ-9, suggesting that the pain-depression dyad may be especially important to those suffering from ESAA.

The adjusted results showed that participants that had more post-operative physiotherapy visits had smaller changes in EQ-5D VAS values. This finding may be an artifact of selection bias, as participants with poorer prognosis were referred to outpatient physiotherapy at higher rates.

Pre-operative EQ-5D VAS scores were found to be strongly associated with change across all four instruments when adjusted for other pre-operative scores, demographic variables, physiotherapy visits, and wait times. This strengthens the notion that general HRQoL encompasses ankle-specific HRQoL, pain, and depression symptoms.

Participants in the second or third SES quintile had odds of exceeding the EQ-5D VAS MID that was higher than participants in the first SES quintile. Males and AA participants had odds of exceeding the AOS MID that were higher than females and TAR participants respectively. Additionally, female sex was associated with increased odds of not improving in a meaningful way across any instrument. This warrants further investigation into the possibility that higher pre-operative expectations or biological differences between the sexes are driving this phenomenon. For example, females are more likely than males to see degradation in their bone health at older ages due to differences in bioavailable estradiol (a key regulator in bone metabolism form both males and females) between the sexes (164). Additionally, males are at a reduced risk of developing knee osteoarthritis than females, which is also thought to be due to differences in sex hormones (163).

This study found that in each dimension of health surveyed, participants with the poorest pre-operative HRQoL experienced the greatest changes in HRQoL. A finding that supports the notion that when a surgeon and patient arrive at a mutually agreeable decision to proceed with surgery for treatment of their ESAA, the most severely affected participants benefit the most across multiple dimensions of health, and beyond simply improving their ankle function.

Many participants waited more than 26 weeks for their surgery. However, this study found no evidence of a relationship between the wait for surgery and change in HRQoL after adjustment for patient-level and health service utilization characteristics. This finding was unexpected, given the relationship between delayed access to surgery and poorer outcomes for other lower extremity orthopedic conditions – knee arthroplasties and hallux valgus procedures (113,165).

There were a non-trivial percentage of participants (11.4%) that did not see any meaningful changes in HRQoL. It is possible these participants did not receive any benefit from surgery, or suffered complications following their procedure. Identifying these patients and providing them with additional support may be warranted.

Some aspects of this study align with previous research; overall mean changes in AOS values of 28.8 were similar to those reported by others (37,82,83). Interestingly, this study found meaningful changes in self-reported depressive symptoms among 22% of participants, a finding which is not concordant with previous research (166). There are a number of explanations for why findings in the present study differ from previous research, including that this cohort had a low number of participants with comorbidities and a larger sample size compared to other studies.

There are several important limitations to this study's findings. Participants were recruited from a single surgical clinic in a metropolitan region and, although this study found no significant differences between participants and non-participants, there may be unobserved participation bias. SES was calculated at a block level and may not be reflective each participants actual SES. In addition, this study measured the change in PROs six months following surgery; other studies have suggested different follow up durations. Also, this study investigated primary elective AA and TAR surgeries and may not be reflective of revision surgeries or emergent procedures.

It is possible that pre-operative scores used in the linear regression and MANCOVA models were confounded by other factors, such as age, Charlson index, or is SES. However, routine preliminary data exploration investigating the interrelatedness of the study variables did not find strong associations with any pre-operative PRO scores. However, other factors such as pre-operative anxiety (96) and the presence of injuries were not explored in this study and could also have a confounding effect on pre-operative PRO scores.

Finally, the findings regarding a potential relationship between post-operative outcomes and physiotherapy are limited to this study's use of observable publicly-funded occurrences, and likely significantly under represent the totality of participant's physiotherapy utilization. The reported lack of associating between wait for surgery and failure to see expected changes post-operatively may be limited by the small sample size and associated lack of power to detect such an association.

6.1 Future directions

Future studies could include other aspects of HRQoL that may be relevant to ESAA patients treated with AA or TAR. Concepts such as anxiety (85) and pain-catastrophizing (86) have been observed in osteoarthritis patients, though not included in the present study. As these concepts are closely related to pain and depression symptoms in orthopedic surgery patients

(167,168), it is recommended that future research investigates these aspects of HRQoL to determine if they have a synergistic affect.

As some participants did not see meaningful changes any in any aspect of HRQoL, it would beneficial to determine if pre-operative factors not captured by this study could have some effect on not seeing meaningful changes in HRQoL.

If this study were to be repeated, based on these findings, it is recommended that instruments with fewer overlapping concepts be utilized. The Ankle Arthritis Scale could be used instead of the AOS (169). Rather than elucidating ankle-specific pain (which may overlap with the PEG) and function like the AOS, the AAS focuses on basic and advanced activities of the ankle. Although the PEG and the PHQ-9 are highly correlated, it would be beneficial to retain both instruments, as at present there are no instruments which measure the pain-depression dyad. Finally, the EQ-5D VAS should be retained as a measure of general HRQoL, to allow for comparisons to other procedures or patient groups. General HRQoL may also be supplemented by the EQ-5D(5L) items, which, due to an increase in the range of responses (five rather than three) are less effected by celling and floor effects than those of the EQ-5D(3L). This would also allow for utility values to be generated, which can be used in cost-utility analyses (27,122).

A qualitative study may also to provide insight into the findings; for example, asking patients who did not see meaningful changes in their HRQoL why they did not experience meaningful changes could provide additional contextual information regarding factors not explored in this study. Additionally, asking patients whether the changes in HRQoL they experienced matched what they expected from their surgery could help shed light on possible relationships between pre-operative expectations and changes in HRQoL.

Considering the findings of this study, it would be beneficial for health-system decision makers to understand the characteristics of the patients, health system factors and surgeon's practices that lead to patients being less likely to experience meaningful changes in HRQoL. Understanding the pathway which results in better outcomes may lead to resources being allocated to improve the likelihood of patients experiencing meaningful gains in health. However, in order to do so, systematic collection, integration, and evaluation of patients' HRQoL data into routine data collection systems would be necessary – a difficult feat given the complexity of integrating additional streams, and burden of, data collection. Furthermore, there may be hesitance by some clinicians to and other health-system decision makers to make decisions based on HRQoL data, given that these are subjective measures of health.

6.2 Conclusion

This study found that most participants treated with AA and TAR for their ESAA reported changes in general HRQoL, ankle-specific HRQoL, pain, and depression symptoms in the peri-operative period. Participants experienced the largest changes in pain, followed by general and ankle-specific HRQoL. Some participants also experienced changes in depression symptoms. Participants with the lowest pre-operative HRQoL saw the largest changes, and that duration on the wait list and utilization of physiotherapy did not affect these changes. However, while all changes were statistically significant, they did not always translate to changes that exceeded each instrument's MID, suggesting that changes are not always meaningful to participants. Future research into changes in HRQoL of ESAA patients treated with AA and TAR should consider additional concepts into PROs that are meaningful to patients and select instruments that have less conceptual overlap.

References

1. Liu G, Peterson AC, Wing K, Crump T, Younger A, Penner M, et al. Validation of the Ankle Osteoarthritis Scale Instrument for Preoperative Evaluation of End-Stage Ankle Arthritis Patients Using Item Response Theory. Foot ankle Int [Internet]. 2019 Jan 10;1071100718818573. Available from: http://journals.sagepub.com/doi/10.1177/1071100718818573

2. Thomas RH, Daniels TR. Ankle arthritis. J Bone Joint Surg Am. 2003 May;85(5):923–36.

3. Felson DT, Naimark A, Anderson J, Kazis L, Castelli W, Meenan RF. The prevalence of knee osteoarthritis in the elderly. the framingham osteoarthritis study. Arthritis Rheum. 1987 Aug;30(8):914–8.

4. Weatherall JM, Mroczek K, McLaurin T, Ding B, Tejwani N. Post-traumatic ankle arthritis. Bull Hosp Jt Dis. 2013;71(1):104–12.

5. Huch K, Kuettner KE, Dieppe P. Osteoarthritis in ankle and knee joints. Semin Arthritis Rheum. 1997 Feb;26(4):667–74.

6. Cushnaghan J, Dieppe P. Study of 500 patients with limb joint osteoarthritis. I. Analysis by age, sex, and distribution of symptomatic joint sites. Ann Rheum Dis. 1991 Jan;50(1):8–13.

7. Barg A, Pagenstert GI, Hügle T, Gloyer M, Wiewiorski M, Henninger HB, et al. Ankle osteoarthritis: etiology, diagnostics, and classification. Foot Ankle Clin [Internet]. 2013 Sep;18(3):411–26. Available from: http://dx.doi.org/10.1016/j.fcl.2013.06.001

8. Stauffer RN, Chao EY, Brewster RC. Force and motion analysis of the normal, diseased, and prosthetic ankle joint. Clin Orthop Relat Res. 1977;(127):189–96.

9. Saltzman CL, Salamon ML, Blanchard GM, Huff T, Hayes A, Buckwalter JA, et al.

Epidemiology of ankle arthritis: report of a consecutive series of 639 patients from a tertiary orthopaedic center. Iowa Orthop J. 2005 Jul 30;25(3):44–6.

10. Saltzman CL, Mclff TE, Buckwalter JA, Brown TD. Total Ankle Replacement Revisited. J Orthop Sport Phys Ther. 2000;30(2):56–67.

11. Valderrabano V, Horisberger M, Russell I, Dougall H, Hintermann B. Etiology of ankle osteoarthritis. Clin Orthop Relat Res. 2009 Jul;467(7):1800–6.

12. Horisberger M, Valderrabano V, Hintermann B. Posttraumatic ankle osteoarthritis after ankle-related fractures. J Orthop Trauma. 2009 Jan;23(1):60–7.

13. Lübbeke A, Salvo D, Stern R, Hoffmeyer P, Holzer N, Assal M. Risk factors for post-traumatic osteoarthritis of the ankle: an eighteen year follow-up study. Int Orthop. 2012 Jul;36(7):1403–10.

14. Nihal A, Gellman RE, Embil JM, Trepman E. Ankle arthrodesis. Foot Ankle Surg. 2008;14(1):1–10.

15. Lopez-Ben R. Imaging of the Subtalar Joint. Foot Ankle Clin. 2015 Jun;20(2):223–41.

16. Gentile MA. Nonsurgical Treatment of Ankle Arthritis. Clin Podiatr Med Surg. 2017 Oct;34(4):415–23.

17. Witteveen AGH, Hofstad CJ, Kerkhoffs GMMJ. Hyaluronic acid and other conservative treatment options for osteoarthritis of the ankle. Cochrane database Syst Rev. 2015 Oct 17;10(10):CD010643.

18. Lawton CD, Butler BA, Dekker RG, Prescott A, Kadakia AR. Total ankle arthroplasty versus ankle arthrodesis—a comparison of outcomes over the last decade. J Orthop Surg Res. 2017 Dec 18;12(1):76.

19. Morash J, Walton DM, Glazebrook M. Ankle Arthrodesis Versus Total Ankle

Arthroplasty. Foot Ankle Clin. 2017 Jun;22(2):251–66.

20. Yasui Y, Hannon CP, Seow D, Kennedy JG. Ankle arthrodesis: A systematic approach and review of the literature. World J Orthop. 2016;7(11):700.

21. Huang MJ. Ankle Fusion. In: Operative Dictations in Orthopedic Surgery [Internet]. New York, NY: Springer New York; 2013 [cited 2019 May 16]. p. 209–10. Available from: https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/ankle-fusion

22. Mendicino SS, Kreplick AL, Walters JL. Open Ankle Arthrodesis. Clin Podiatr Med Surg [Internet]. 2017 Oct;34(4):489–502. Available from: http://dx.doi.org/10.1016/j.cpm.2017.05.006

23. Charnley J. Compression arthrodesis of the ankle and shoulder. J Bone Joint Surg Br. 1951 May;33B(2):180–91.

24. Elmlund AO, Winson IG. Arthroscopic ankle arthrodesis. Foot Ankle Clin [Internet]. 2015 Mar;20(1):71–80. Available from: http://dx.doi.org/10.1016/j.fcl.2014.10.008

25. Groth HE, Fitch HF. Salvage procedures for complications of total ankle arthroplasty. Clin Orthop Relat Res. 1987 Nov;(224):244–50.

26. Gross C, Erickson BJ, Adams SB, Parekh SG. Ankle Arthrodesis After Failed Total Ankle Replacement. Foot Ankle Spec. 2015 Apr 5;8(2):143–51.

27. Nwachukwu BU, McLawhorn AS, Simon MS, Hamid KS, Demetracopoulos CA, Deland JT, et al. Management of End-Stage Ankle Arthritis: Cost-Utility Analysis Using Direct and Indirect Costs. J Bone Jt Surgery-American Vol. 2015 Jul 15;97(14):1159–72.

28. Younger ASE, MacLean S, Daniels TR, Penner MJ, Wing KJ, Dunbar M, et al. Initial Hospital-Related Cost Comparison of Total Ankle Replacement and Ankle Fusion With Hip and Knee Joint Replacement. Foot Ankle Int. 2015 Mar 3;36(3):253–7.

29. Haddad SL. Intermediate and Long-Term Outcomes of Total Ankle Arthroplasty and Ankle Arthrodesis. J Bone Jt Surg [Internet]. 2007 Sep 1;89(9):1899. Available from: http://jbjs.org/cgi/doi/10.2106/JBJS.F.01149

30. Hendrickx RPM, Stufkens SAS, de Bruijn EE, Sierevelt IN, van Dijk CN, Kerkhoffs GMMJ. Medium- to long-term outcome of ankle arthrodesis. Foot ankle Int. 2011 Oct;32(10):940–7.

31. Saltzman CL, Mann RA, Ahrens JE, Amendola A, Anderson RB, Berlet GC, et al. Prospective Controlled Trial of STAR Total Ankle Replacement versus Ankle Fusion: Initial Results. Foot Ankle Int. 2009 Jul;30(7):579–96.

32. Southerland JT, Boberg JS, Downey M, Nakra A, Rabjon L V. McGlamry's Comprehensive Textbook of Foot and Ankle Surgery. 4th ed. Philadelphia: Wolters Kluwer Health; 2013.

33. Sadoghi P, Roush G, Kastner N, Leithner A, Sommitsch C, Goswami T. Failure modes for total ankle arthroplasty: a statistical analysis of the Norwegian Arthroplasty Register. Arch Orthop Trauma Surg. 2014 Oct 1;134(10):1361–8.

34. Prusinowska A, Krogulec Z, Turski P, Przepiórski E, Małdyk P, Księżopolska-Orłowska K. Total ankle replacement – surgical treatment and rehabilitation. Reumatologia/Rheumatology. 2015;1(1):34–9.

35. Jastifer JR, Coughlin MJ. Long-Term Follow-Up of Mobile Bearing Total Ankle Arthroplasty in the United States. Foot Ankle Int. 2015 Feb 8;36(2):143–50.

36. Sadoghi P, Liebensteiner M, Agreiter M, Leithner A, Böhler N, Labek G. Revision Surgery After Total Joint Arthroplasty: A Complication-Based Analysis Using Worldwide Arthroplasty Registers. J Arthroplasty. 2013 Sep;28(8):1329–32.

37. Daniels TR, Younger AS, Penner M, Wing K, Dryden PJ, Wong H, et al. Intermediate-Term Results of Total Ankle Replacement and Ankle Arthrodesis. J Bone Jt Surgery-American Vol. 2014 Jan 15;96(2):135–42.

38. S Fevang B-T, A Lie S, I Havelin L, G Brun J, Skredderstuen A, Furnes O. 257 ankle arthroplasties performed in Norway between 1994 and 2005. Acta Orthop. 2007 Jan 8;78(5):575–83.

39. Henricson A, Skoog A, Carlsson Å. The Swedish Ankle Arthroplasty Register: An analysis of 531 arthroplasties between 1993 and 2005. Acta Orthop. 2007 Jan 8;78(5):569–74.

40. Wood PLR, Prem H, Sutton C. Total ankle replacement. J Bone Joint Surg Br. 2008 May;90-B(5):605–9.

41. Moons P. Why Call it Health-Related Quality of Life When You Mean Perceived Health Status? Eur J Cardiovasc Nurs. 2004 Dec 24;3(4):275–7.

42. Karimi M, Brazier J. Health, Health-Related Quality of Life, and Quality of Life: What is the Difference? Pharmacoeconomics. 2016 Jul 18;34(7):645–9.

43. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. JAMA. 1995 Jan 4;273(1):59–65.

44. Guyatt GH. Measuring Health-Related Quality of Life. Ann Intern Med. 1993 Apr 15;118(8):622.

45. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Control Clin Trials. 1989 Dec;10(4):407–15.

46. Centers for Disease Control and Prevention. Well-Being Concepts [Internet]. 2018 [cited 2019 Jun 16]. Available from: https://www.cdc.gov/hrqol/wellbeing.htm#eight

47. Sutherland JM, Liu G, Crump RT, Karimuddin AA. Waiting for surgery : is waiting bad for anyone or everyone ? Hernia. 2017;

48. Jull A, Parag V, Walker N, Rodgers A. Responsiveness of generic and disease-specific health-related quality of life instruments to venous ulcer healing. Wound Repair Regen. 2010;18(1):26–30.

49. Newton E, Janjua A, Lai E, Liu G, Crump T, Sutherland JM. The impact of surgical wait time on patient reported outcomes in sinus surgery for chronic rhinosinusitis. Int Forum Allergy Rhinol. 2017;00(0):1–6.

50. World Health Organization. Constitution of the World Health Organization. 48th ed. Geneva: Basic Documents of the World Health Organization; 2014. 7 p.

51. The WHOQOL Group. The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. Soc Sci Med. 1998 Jun;46(12):1569–85.

52. Ferrans CE, Zerwic JJ, Wilbur JE, Larson JL. Conceptual Model of Health-Related Quality of Life. J Nurs Scholarsh [Internet]. 2005 Dec;37(4):336–42. Available from: http://doi.wiley.com/10.1111/j.1547-5069.2005.00058.x

53. Bair MJ, Robinson RL, Katon W, Kroenke K. Depression and Pain Comorbidity. Arch Intern Med. 2003 Nov 10;163(20):2433.

54. Basch E. Patient-Reported Outcomes - Harnessing Patients' Voices to Improve Clinical Care. N Engl J Med. 2017 Jan 12;376(2):105–8.

55. Valderas JM, Kotzeva a, Espallargues M, Guyatt G, Ferrans CE, Halyard MY, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. Qual Life Res. 2008;17(2):179–93.

56. Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S. Patient reported outcome measures in practice. BMJ. 2015 Feb 10;350(feb10 14):g7818–g7818.

57. Conti MS, Jones MT, Savenkov O, Deland JT, Ellis SJ. Outcomes of Reconstruction of the Stage II Adult-Acquired Flatfoot Deformity in Older Patients. Foot Ankle Int. 2018 Sep 18;39(9):1019–27.

58. Cooke M, Marsh J, Clark M, Nakash R, Jarvis R, Hutton J, et al. Treatment of severe ankle sprain: a pragmatic randomised controlled trial comparing the clinical effectiveness and cost-effectiveness of three types of mechanical ankle support with tubular bandage. The CAST trial. Health Technol Assess (Rockv). 2009 Feb;13(13):iii, ix–x, 1–121.

59. Gooch K, Marshall DA, Faris PD, Khong H, Wasylak T, Pearce T, et al. Comparative effectiveness of alternative clinical pathways for primary hip and knee joint replacement patients: a pragmatic randomized, controlled trial. Osteoarthr Cartil [Internet]. 2012 Oct;20(10):1086–94. Available from: http://dx.doi.org/10.1016/j.joca.2012.06.017

60. Marshall DA, Christiansen T, Smith C, Squire Howden J, Werle J, Faris P, et al. Continuous Quality Improvement Program for Hip and Knee Replacement. Am J Med Qual. 2015 Sep 23;30(5):425–31.

61. Loveday DT, Barr L V., Loizou CL, Barton G, Smith G. A comparative prospective cohort health economic analysis comparing ankle fusion, isolated great toe fusion and hallux valgus surgery. Foot Ankle Surg [Internet]. 2018 Feb;24(1):54–9. Available from: https://doi.org/10.1016/j.fas.2016.11.008

62. Rabin R, Charro F de. EQ-SD: a measure of health status from the EuroQol Group. Ann Med. 2001 Jan 8;33(5):337–43.

63. Bombardier C, Melfi CA, Paul J, Green R, Hawker G, Wright J, et al. Comparison of a

generic and a disease-specific measure of pain and physical function after knee replacement surgery. Med Care. 1995 Apr;33(4 Suppl):AS131-44.

64. Calvert MJ, Freemantle N. Use of health-related quality of life in prescribing research. Part 1: why evaluate health-related quality of life? J Clin Pharm Ther [Internet]. 2003 Dec;28(6):513–21. Available from: http://doi.wiley.com/10.1046/j.0269-4727.2003.00521.x

65. Patrick DL, Deyo RA. Generic and Disease-Specific Measures in Assessing Health Status and Quality of Life. Med Care. 1989 Mar;27(Supplement):S217–32.

66. Ware J, Snow K, Kosinski M, Gandek B. SF-36 Health Survey-manual and interpretation guide. 1993.

67. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy. 1990 Dec;16(3):199–208.

68. Domsic RTRTT, Saltzman CLCL. Ankle osteoarthritis scale. Foot ankle Int. 1998 Jul;19(7):466–71.

69. Johnston BC, Ebrahim S, Carrasco-Labra A, Furukawa TA, Patrick DL, Crawford MW, et al. Minimally important difference estimates and methods: a protocol. BMJ Open. 2015 Oct 1;5(10):e007953.

70. Schunemann HJ, Guyatt GH. Commentary - Goodbye M(C)ID! hello MID, where do you come from? Health Serv Res. 2005;40(2):593–7.

71. Desai S, Peterson AC, Wing K, Younger A, Crump T, Liu G, et al. Minimally Important Difference in the Foot and Ankle Outcome Score Among Patients Undergoing Hallux Valgus Surgery. Foot Ankle Int [Internet]. 2019 Jun 15;40(6):694–701. Available from: http://journals.sagepub.com/doi/10.1177/1071100719831392

72. Coe MP, Sutherland JM, Penner MJ, Younger A, Wing KJ. Minimal Clinically Important Difference and the Effect of Clinical Variables on the Ankle Osteoarthritis Scale in Surgically Treated End-Stage Ankle Arthritis. J Bone Jt Surgery-American Vol. 2015;97(10):818–23.

73. Engel L, Beaton DE, Touma Z. Minimal Clinically Important Difference: A Review of Outcome Measure Score Interpretation. Rheum Dis Clin North Am [Internet]. 2018;44(2):177–88. Available from: https://doi.org/10.1016/j.rdc.2018.01.011

74. King MT. A point of minimal important difference (MID): a critique of terminology and methods. Expert Rev Pharmacoecon Outcomes Res. 2011 Apr;11(2):171–84.

75. Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). J Man Manip Ther [Internet]. 2012 Aug 12;20(3):160–6. Available from: http://www.tandfonline.com/doi/full/10.1179/2042618612Y.0000000001

76. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the Clinical Importance of Treatment Outcomes in Chronic Pain Clinical Trials: IMMPACT Recommendations. J Pain. 2008 Feb;9(2):105–21.

77. Cody EA, Mancuso CA, MacMahon A, Marinescu A, Burket JC, Drakos MC, et al. Development of an Expectations Survey for Patients Undergoing Foot and Ankle Surgery. Foot Ankle Int [Internet]. 2016 Dec;37(12):1277–84. Available from: http://journals.sagepub.com/doi/10.1177/1071100716666260

78. Cody EA, Mancuso CA, Burket JC, Marinescu A, MacMahon A, Ellis SJ. Patient Factors Associated With Higher Expectations From Foot and Ankle Surgery. Foot Ankle Int [Internet]. 2017 May 14;38(5):472–8. Available from:

http://journals.sagepub.com/doi/10.1177/1071100717690807

79. Glazebrook M, Daniels T, Younger A, Foote CJ, Penner M, Wing K, et al. Comparison of Health-Related Quality of Life Between Patients with End-Stage Ankle and Hip Arthrosis. J Bone Jt Surgery-American Vol. 2008 Mar;90(3):499–505.

80. Pakzad H, Thevendran G, Penner MJ, Qian H, Younger A. Factors Associated with Longer Length of Hospital Stay After Primary Elective Ankle Surgery for End-Stage Ankle Arthritis. J Bone Jt Surgery-American Vol. 2014 Jan 1;96(1):32–9.

81. Koivu H, Kohonen I, Mattila K, Loyttyniemi E, Tiusanen H. Long-term Results of Scandinavian Total Ankle Replacement. Foot Ankle Int [Internet]. 2017 Jul 23;38(7):723–31. Available from: http://journals.sagepub.com/doi/10.1177/1071100717698695

82. Lachman JR, Ramos JA, Adams SB, Nunley JA, Easley ME, DeOrio JK. Patient-Reported Outcomes Before and After Primary and Revision Total Ankle Arthroplasty. Foot Ankle Int. 2019;40(1):34–41.

83. Desai SJ, Glazebrook M, Penner MJ, Wing KJ, Younger ASE, Pinsker E, et al. Quality of Life in Bilateral Vs. Unilateral End-Stage Ankle Arthritis and Outcomes of Bilateral Vs. Unilateral Total Ankle Replacement. J Bone Jt Surg. 2017 Jan 18;99(2):133–40.

84. Slobogean GP, Younger A, Apostle KL, Marra CA, Wing K, Penner M, et al. Preference-Based Quality of Life of End-Stage Ankle Arthritis Treated with Arthroplasty or Arthrodesis. Foot Ankle Int [Internet]. 2010 Jul;31(7):563–6. Available from: http://journals.sagepub.com/doi/10.3113/FAI.2010.0563

85. Tan V, Jinks C, Chew-Graham C, Healey EL, Mallen C. The triple whammy anxiety depression and osteoarthritis in long-term conditions. BMC Fam Pract [Internet]. 2015;16(1):15–7. Available from: http://dx.doi.org/10.1186/s12875-015-0346-2

86. Somers TJ, Keefe FJ, Pells JJ, Dixon KE, Waters SJ, Riordan PA, et al. Pain Catastrophizing and Pain-Related Fear in Osteoarthritis Patients: Relationships to Pain and Disability. J Pain Symptom Manage. 2009 May;37(5):863–72.

87. Depressive Disorders. In: Diagnostic and Statistical Manual of Mental Disorders [Internet]. American Psychiatric Association; 2013. Available from: https://psychiatryonline.org/doi/10.1176/appi.books.9780890425596.dsm04

88. World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimates [Internet]. Geneva; 2017 [cited 2019 May 20]. Available from: https://www.who.int/mental_health/management/depression/prevalence_global_health_estimates/en/

89. Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, et al. Cross-national epidemiology of DSM-IV major depressive episode. BMC Med. 2011 Dec 26;9(1):90.

90. Gadermann AM, Engel CC, Naifeh JA, Nock MK, Petukhova M, Santiago PN, et al. Prevalence of DSM-IV Major Depression Among U.S. Military Personnel: Meta-Analysis and Simulation. Mil Med. 2012 Aug;177(8S):47–59.

91. Thomson KC, Richardson CG, Gadermann AM, Emerson SD, Shoveller J, Guhn M. Association of Childhood Social-Emotional Functioning Profiles at School Entry With Early-Onset Mental Health Conditions. JAMA Netw Open [Internet]. 2019 Jan 4;2(1):e186694. Available from: http://jamanetworkopen.jamanetwork.com/article.aspx?doi=10.1001/jamanetworkopen.2018.6694

92. Mollon B, Mahure SA, Ding DY, Zuckerman JD, Kwon YW. The influence of a history

of clinical depression on peri-operative outcomes in elective total shoulder arthroplasty. Bone Joint J [Internet]. 2016 Jun;98-B(6):818–24. Available from: http://online.boneandjoint.org.uk/doi/10.1302/0301-620X.98B6.37208

93. Rosemann T, Backenstrass M, Joest K, Rosemann A, Szecsenyi J, Laux G. Predictors of depression in a sample of 1,021 primary care patients with osteoarthritis. Arthritis Rheum [Internet]. 2007 Apr 15;57(3):415–22. Available from: http://doi.wiley.com/10.1002/art.22624

94. Guglielmo D, Hootman JM, Boring MA, Murphy LB, Theis KA, Croft JB, et al. Symptoms of Anxiety and Depression Among Adults with Arthritis — United States, 2015–2017. MMWR Morb Mortal Wkly Rep. 2018 Oct 5;67(39):1081–7.

95. Browne JA, Sandberg BF, D'Apuzzo MR, Novicoff WM. Depression is associated with early postoperative outcomes following total joint arthroplasty: a nationwide database study. J Arthroplasty [Internet]. 2014 Mar;29(3):481–3. Available from: http://dx.doi.org/10.1016/j.arth.2013.08.025

96. Hanusch BC, O'Connor DB, Ions P, Scott A, Gregg PJ. Effects of psychological distress and perceptions of illness on recovery from total knee replacement. Bone Joint J [Internet]. 2014 Feb;96-B(2):210–6. Available from: http://online.boneandjoint.org.uk/doi/10.1302/0301-620X.96B2.31136

97. Kennedy S, Barske H, Wing K, Penner M, Daniels T, Glazebrook M. SF-36 Mental Component Summary (MCS) Score Does Not Predict Functional Outcome After Surgery for End-Stage Ankle Arthritis. J Bone Jt Surg Am. 2015;1702–7.

98. Arnow BA, Hunkeler EM, Blasey CM, Lee J, Constantino MJ, Fireman B, et al. Comorbid depression, chronic pain, and disability in primary care. Psychosom Med.

2006;68(2):262–8.

99. Li J-X. Pain and depression comorbidity: A preclinical perspective. Behav Brain Res. 2015 Jan;276(10):92–8.

100. Bair MJ, Robinson RL, Katon W, Kroenke K. Depression and Pain Comorbidity, A Literature Review. JAMA Intern Med. 2003;163(20):2433–45.

101. Lindberg MF, Miaskowski C, Rustøen T, Rosseland LA, Paul SM, Lerdal A. Preoperative Pain, Symptoms, and Psychological Factors related to Higher Acute Pain Trajectories during Hospitalization for Total Knee Arthroplasty. Mouraux A, editor. PLoS One. 2016 Sep 1;11(9):e0161681.

102. Sobolev B, Guy P, Sheehan KJ, Bohm E, Beaupre L, Morin SN, et al. Hospital mortality after hip fracture surgery in relation to length of stay by care delivery factors. 2017;0(March).

103. Mitchell C, Walker J, Walters S, Morgan AB, Binns T, Mathers N. Costs and effectiveness of pre- and post-operative home physiotherapy for total knee replacement: randomized controlled trial. J Eval Clin Pract [Internet]. 2005 Jun;11(3):283–92. Available from: http://doi.wiley.com/10.1111/j.1365-2753.2005.00535.x

104. Artz N, Elvers KT, Lowe CM, Sackley C, Jepson P, Beswick AD. Effectiveness of physiotherapy exercise following total knee replacement: systematic review and meta-analysis. BMC Musculoskelet Disord. 2015 Dec 7;16(1):15.

105. Minns Lowe CJ, Barker KL, Dewey ME, Sackley CM. Effectiveness of physiotherapy exercise following hip arthroplasty for osteoarthritis: a systematic review of clinical trials. BMC Musculoskelet Disord. 2009 Dec 4;10(1):98.

106. Barg A, Wimmer MD, Wiewiorski M, Wirtz DC, Pagenstert GI, Valderrabano V. Total

Ankle Replacement. Dtsch Aerzteblatt Online. 2015 Mar 13;112(11):177–84.

107. Smith CL. Physical Therapy Management of Patients with Total Ankle Replacement. Phys Ther. 1980 Mar 1;60(3):303–6.

108. Sanmartin C, Berthelot J-M, McIntosh C. Determinants of Unacceptable Waiting Times for Specialized Services in Canada. Healthc Policy | Polit Santé. 2007 Feb 15;2(3):e140-54.

109. Siciliani L, Borowitz M, Moran V. Executive Summary. New Dir Youth Dev [Internet]. 2013 Dec;2013(140):5–8. Available from: http://dx.doi.org/10.1787/9789264179080-en

110. CIHI. Wait Times for Priority Procedures in Canada, 2017 [Internet]. Canadian Institute for Health Information. 2017 [cited 2019 Feb 27]. Available from: https://secure.cihi.ca/estore/productSeries.htm?pc=PCC395%5Cnh

111. Hirvonen J, Blom M, Tuominen U, Seitsalo S, Lehto M, Paavolainen P, et al. Evaluating waiting time effect on health outcomes at admission: a prospective randomized study on patients with osteoarthritis of the knee joint. J Eval Clin Pract. 2007 Oct;13(5):728–33.

112. Hirvonen J, Tuominen U, Seitsalo S, Lehto M, Paavolainen P, Hietaniemi K, et al. The effect of waiting time on health-related quality of life, pain, and physical function in patients awaiting primary total hip replacement: a randomized controlled trial. Value Heal. 2009 Sep;12(6):942–7.

113. Desmeules F, Dionne CE, Belzile E, Bourbonnais R, Fremont P. The burden of wait for knee replacement surgery: effects on pain, function and health-related quality of life at the time of surgery. Rheumatology. 2010 May 1;49(5):945–54.

114. Kelly KD, Voaklander DC, Johnston DW, Newman SC, Suarez-Almazor ME. Change in pain and function while waiting for major joint arthroplasty. J Arthroplasty. 2001

Apr;16(3):351–9.

115. Ostendorf M, Buskens E, van Stel H, Schrijvers A, Marting L, Dhert W, et al. Waiting for total hip arthroplasty: avoidable loss in quality time and preventable deterioration. J Arthroplasty. 2004 Apr;19(3):302–9.

116. Boonstra AM, Schiphorst HR, Reneman MF, Posthumus JB, Stewart RE. Reliability and validity of the visual analogue scale for disability in patients with chronic musculoskeletal pain Boonstra, Anne M.; Schiphorst Preuper, Henrica R.; Reneman, Michiel F.; Posthumus, Jitze B.; Stewart, Roy E. Int J Rehabil Res. 2008;31:165–9.

117. Zampelis V, Ornstein E, Franzén H, Atroshi I. A simple visual analog scale for pain is as responsive as the WOMAC, the SF-36, and the EQ-5D in measuring outcomes of revision hip arthroplasty. Acta Orthop [Internet]. 2014 Apr 30;85(2):128–32. Available from: http://www.tandfonline.com/doi/full/10.3109/17453674.2014.887951

118. Grobet C, Marks M, Tecklenburg L, Audige L. Application and measurement properties of EQ-5D to measure quality of life in patients with upper extremity orthopaedic disorders: a systematic literature review. Arch Orthop Trauma Surg [Internet]. 2018;138(7):953–61. Available from: http://dx.doi.org/10.1007/s00402-018-2933-x

119. 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 [Internet]. 2015 Dec;30(12):2154–8. Available from: http://dx.doi.org/10.1016/j.arth.2015.06.012

120. Henricson A, Kamrad I, Rosengren B, Carlsson Å. Bilateral Arthrodesis of the Ankle Joint: Self-Reported Outcomes in 35 Patients From the Swedish Ankle Registry. J Foot Ankle Surg. 2016 Nov;55(6):1195–8.

121. Sutherland JM, Rajapakshe S, Younger A, Crump T, Wing K, Liu G, et al. Do preoperative patient-reported outcomes predict hospital length of stay for surgically-treated end-stage ankle osteoarthritis patients? Foot Ankle Surg. 2019 Feb 2;

122. Bansback N, Tsuchiya A, Brazier J, Anis A. Canadian Valuation of EQ-5D Health States: Preliminary Value Set and Considerations for Future Valuation Studies. van Baal PHM, editor. PLoS One [Internet]. 2012 Feb 6;7(2):e31115. Available from: https://dx.plos.org/10.1371/journal.pone.0031115

123. Giesinger K, Hamilton DF, Jost B, Holzner B, Giesinger JM. Comparative responsiveness of outcome measures for total knee arthroplasty. Osteoarthr Cartil [Internet]. 2014 Feb;22(2):184–9. Available from: http://dx.doi.org/10.1016/j.joca.2013.11.001

124. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5D-5L value set for England. Health Econ [Internet]. 2018 Jan;27(1):7–22. Available from: http://doi.wiley.com/10.1002/hec.3564

125. Conner-Spady BL, Marshall DA, Bohm E, Dunbar MJ, Loucks L, Khudairy A Al, et al. Reliability and validity of the EQ-5D-5L compared to the EQ-5D-3L in patients with osteoarthritis referred for hip and knee replacement. Qual Life Res [Internet]. 2015;24(7):1775–84. Available from: http://dx.doi.org/10.1007/s11136-014-0910-6

126. Domsic RT, Saltzman CL. Ankle osteoarthritis scale. Foot ankle Int / Am Orthop Foot Ankle Soc [and] Swiss Foot Ankle Soc. 1998;19(7):466–71.

127. Hubley AM;, Zumbo BD. A dialectic on validity: Where we have been and where we are going ABI/INFORM Collection pg. 207. J Gen Psychol. 1996;123.

128. Veltman ES, Hofstad CJ, Witteveen AGH. Are current foot- and ankle outcome measures appropriate for the evaluation of treatment for osteoarthritis of the ankle?: Evaluation of

ceiling effects in foot- and ankle outcome measures. Foot Ankle Surg [Internet]. 2017 Sep;23(3):168–72. Available from: http://dx.doi.org/10.1016/j.fas.2016.02.006

129. Croft S, Wing KJ, Daniels TR, Glazebrook M, Dryden P, Younger A, et al. Association of Ankle Arthritis Score With Need for Revision Surgery. Foot Ankle Int [Internet]. 2017;107110071771289. Available from: http://journals.sagepub.com/doi/10.1177/1071100717712891

130. Miller KK, Combs SA, Van Puymbroeck M, Altenburger PA, Kean J, Dierks TA, et al. Fatigue and pain: relationships with physical performance and patient beliefs after stroke. Top Stroke Rehabil [Internet]. 2013;20(4):347–55. Available from: http://www.tandfonline.com/doi/full/10.1310/tsr2004-347

131. Krebs EE, Lorenz KA, Bair MJ, Damush TM, Wu J, Sutherland JM, et al. Development and Initial Validation of the PEG, a Three-item Scale Assessing Pain Intensity and Interference. J Gen Intern Med. 2009 Jun 6;24(6):733–8.

132. Krebs EE, Bair MJ, Damush TM, Tu W, Wu J, Kroenke K. Comparative Responsiveness of Pain Outcome Measures Among Primary Care Patients With Musculoskeletal Pain. Med Care. 2010 Nov;48(11):1007–14.

133. Spitzer RL, Williams JBW, Kroenke K, Hornyak R, McMurray J. Validity and utility of the PRIME-MD Patient Health Questionnaire in assessment of 3000 obstetric-gynecologic patients: The PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. Am J Obstet Gynecol. 2000 Sep;183(3):759–69.

134. Kroenke K, Spitzer RL, Williams JBW, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. Gen Hosp Psychiatry. 2010 Jul;32(4):345–59.

135. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. J Gen Intern Med. 2001 Sep;16(9):606–13.

136. Bombardier CH, Kalpakjian CZ, Graves DE, Dyer JR, Tate DG, Fann JR. Validity of the patient health questionnaire-9 in assessing major depressive disorder during inpatient spinal cord injury rehabilitation. Arch Phys Med Rehabil [Internet]. 2012;93(10):1838–45. Available from: http://dx.doi.org/10.1016/j.apmr.2012.04.019

137. Becher S, Smith M, Ziran B. Orthopaedic Trauma Patients and Depression. J Orthop Trauma. 2014 Oct;28(10):e242–6.

138. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. J Chronic Dis. 1987 Jan;40(5):373–83.

139. Chikuda H, Yasunaga H, Horiguchi H, Takeshita K, Sugita S, Taketomi S, et al. Impact of age and comorbidity burden on mortality and major complications in older adults undergoing orthopaedic surgery: an analysis using the Japanese diagnosis procedure combination database. BMC Musculoskelet Disord [Internet]. 2013 May 28;14(1):173. Available from: BMC Musculoskeletal Disorders

140. Hollowell J, Grocott MPW, Hardy R, Haddad FS, Mythen MG, Raine R. Major elective joint replacement surgery: socioeconomic variations in surgical risk, postoperative morbidity and length of stay. J Eval Clin Pract [Internet]. 2010 Mar;16(3):529–38. Available from: http://doi.wiley.com/10.1111/j.1365-2753.2009.01154.x

141. D'Apuzzo MR, Villa JM, Alcerro JC, Rossi MD, Lavernia CJ. Total Joint Arthroplasty: A Granular Analysis of Outcomes in the Economically Disadvantaged Patient. J Arthroplasty. 2016 Sep;31(9):41–4.

142. Vincent K, Sutherland JM. A Review of Methods for Deriving an Index for Socioeconomic Status in British Columbia [Internet]. Vancouver, BC; 2013 [cited 2019 Mar 5]. Available from: http://www.healthcarefunding.ca/files/2013/04/Review-of-Methods-for-SES-Index-for-BC.pdf

143. Mustard CA, Derksen S, Berthelot JM, Wolfson M. Assessing ecologic proxies for household income: A comparison of household and neighbourhood level income measures in the study of population health status. Heal Place. 1999;5(2):157–71.

144. Lowe CJM, Barker KL, Dewey M, Sackley CM. Effectiveness of physiotherapy exercise after knee arthroplasty for osteoarthritis: systematic review and meta-analysis of randomised controlled trials. BMJ [Internet]. 2007 Oct 20;335(7624):812. Available from: http://www.bmj.com/lookup/doi/10.1136/bmj.39311.460093.BE

145. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. Stat Methods Med Res. 2014;23(5):440–59.

146. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res [Internet]. 2011 Mar;20(1):40–9. Available from: http://doi.wiley.com/10.1002/mpr.329

147. Mustillo S, Kwon S. Auxiliary Variables in Multiple Imputation When Data Are Missing Not at Random. J Math Sociol [Internet]. 2015;39(2):73–91. Available from: http://dx.doi.org/10.1080/0022250X.2013.877898

148. Vittinghoff E, Glidden D V., Shiboski SC, McCulloch CE. Regression Methods in Biostatistics. 2nd ed. Boston, MA: Springer US; 2012. 130–135 p. (Statistics for Biology and Health).

149. Johnson DR, Young R. Toward best practices in analyzing datasets with missing data:

Comparisons and recommendations. J Marriage Fam. 2011;73(5):926–45.

150. Gomes M, Gutacker N, Bojke C, Street A. Addressing Missing Data in Patient-Reported Outcome Measures (PROMS): Implications for the Use of PROMS for Comparing Provider Performance. Health Econ [Internet]. 2016 May;25(5):515–28. Available from: http://doi.wiley.com/10.1002/hec.3173

151. Crump RT, Liu G, Chase M, Sutherland JM. Patient-reported outcomes and surgical triage: A gap in patient-centered care? Qual Life Res. 2016 Nov 6;25(11):2845–51.

152. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prev Sci. 2007;8(3):206–13.

153. Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health Qual Life Outcomes. 2010;8(1):4.

154. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care. 2004 Dec;42(12):1194–201.

155. Beaton DE, Boers M, Wells G a. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. Curr Opin Rheumatol. 2002;14(2):109–14.

156. Bender R, Lange S. Adjusting for multiple testing--when and how? J Clin Epidemiol [Internet]. 2001 Apr;54(4):343–9. Available from: http://dx.doi.org/10.1016/S0895-4356(00)00314-0

157. Streiner DL. Breaking up is hard to do: The heartbreak of dichotomizing continuous data. Can J Psychiatry. 2002;47(3):262–6.

158. Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain

sight. BMJ [Internet]. 2016 Apr 27;353:i1981. Available from: http://www.bmj.com/lookup/doi/10.1136/bmj.i1981

159. Team R Development Core. R: a language and environment for statistical computing. 2011;0.

160. Wickham H. ggplot2. Cham: Springer International Publishing; 2016. (Use R!).

161. Claussen B. Socioeconomic Status and Health [Internet]. Second Edi. Vol. 21, International Encyclopedia of the Social & Behavioral Sciences: Second Edition. Elsevier; 2015. 931–936 p. Available from: http://dx.doi.org/10.1016/B978-0-08-097086-8.14043-7

162. Cohen J. Statistical Power Analysis for the Behavioural Sciences. 2nd ed. Lawrence Erlbaum Associates; 1988.

163. Srikanth VK, Fryer JL, Zhai G, Winzenberg TM, Hosmer D, Jones G. A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. Osteoarthr Cartil. 2005;13(9):769–81.

164. Cauley JA. Estrogen and bone health in men and women. Steroids [Internet]. 2015 Jul;99(Pt A):11–5. Available from: http://dx.doi.org/10.1016/j.steroids.2014.12.010

165. Sutherland JM, Wing K, Penner M, Younger A, Liu G, Veljkovic A, et al. Quantifying Patient-Reported Disability and Health While Waiting for Bunion Surgery. Foot Ankle Int [Internet]. 2018;39(9):1047–55. Available from: https://doi.org/10.1177/1071100718772037

166. Kennedy S, Barske H, Wing K, Penner M, Daniels T, Glazebrook M. After Surgery for End-Stage Ankle Arthritis. J Bone Jt Surg Am. 2015;1702–7.

167. Dave AJ, Selzer F, Losina E, Klara KM, Collins JE, Usiskin I, et al. Is There an Association Between Whole-body Pain With Osteoarthritis-related Knee Pain, Pain

Catastrophizing, and Mental Health? Clin Orthop Relat Res. 2015;473(12):3894–902.

168. Beleckas CM, Prather H, Guattery J, Wright M, Kelly M, Calfee RP. Anxiety in the orthopedic patient: using PROMIS to assess mental health. Qual Life Res [Internet]. 2018;27(9):2275–82. Available from: http://dx.doi.org/10.1007/s11136-018-1867-7

169. Wing KJ, Chapinal N, Coe MP, Daniels TR, Glazebrook M, Dryden P, et al. Measuring the Operative Treatment Effect in End-Stage Ankle Arthritis: Are We Asking the Right Questions? A COFAS Multicenter Study. Foot ankle Int [Internet]. 2017 Oct;38(10):1064–9. Available from: http://journals.sagepub.com/doi/10.1177/1071100717714953

Appendices

Appendix A Results of Models Using Categorical Variables

Table 22 MANCOVA Sensitivity Analysis

Regression Variable	Estimate	P value
Intercept	-0.059 (1.162)	0.960
Pre-Operative PRO Score		
EQ-5D VAS	-0.743 (0.213)	0.001
PEG	-0.048 (0.210)	0.819
PHQ-9	-0.266 (0.206)	0.202
Age	-0.004 (0.019)	0.825
Physio Group		
0		
≥ 1	-0.601 (0.610)	0.328
Wait time in Weeks		
0 - 12	Reference Group	
13 - 26	-0.018 (0.474)	0.970
> 26	0.090 (0.400)	0.823
Sex		
Female	Reference Group	
Male	0.060 (0.363)	0.869
Charlson Index		
0	Reference Group	
≥ 1	-0.035 (0.426)	0.934
SES		
1 (Highest)	Reference Group	
2	0.248 (0.612)	0.687
3	0.458 (0.611)	0.456
4	0.094 (0.668)	0.889
5 (Lowest)	0.465 (0.621)	0.456
Missing	0.184 (0.622)	0.768
Surgery		
AA	Reference Group	
TAR	0.161 (0.388)	0.679

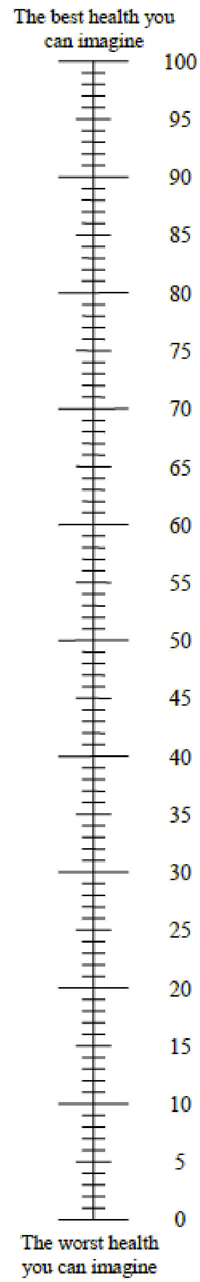
Table 23 Linear regression sensitivity analysis (with categorical versions of wait group and physiotherapy visits)

	EQ-5D VAS		AOS		PEG		PHQ-9	
Regression Variable	Coefficient (SE)	P-Value	Coefficient (SE)	P-Value	Coefficient (SE)	P-Value	Coefficient (SE)	P-Value
Intercept	42.965 (12.070)	0.001	15.809 (21.963)	0.477	2.370 (1.845)	0.203	1.596 (2.794)	0.570
Pre-Operative PRO Score	-0.638 (0.097)	< 0.0001	-0.580 (0.162)	0.001	-0.618 (0.106)	< 0.0001	-0.550 (0.073)	< 0.0001
Age	0.002 (0.174)	0.991	0.027 (0.299)	0.929	-0.014 (0.027)	0.601	0.015 (0.042)	0.723
Physio Category								
0		Reference Group		Reference Group		Reference Group		Reference Group
≥ 1	-11.841 (5.824)	0.046	8.947 (8.997)	0.325	1.573 (0.934)	0.096	2.924 (1.406)	0.051
Wait Time Group (Weeks)								
0 - 12		Reference Group		Reference Group		Reference Group		Reference Group
12 - 26	-1.331 (4.526)	0.770	1.699 (6.382)	0.791	0.453 (0.715)	0.529	-0.200 (1.076)	0.853
> 26	1.258 (3.749)	0.738	-2.510 (6.056)	0.681	-0.089 (0.592)	0.881	-0.648 (0.894)	0.471
Sex								
Female		Reference Group		Reference Group		Reference Group		Reference Group
Male	1.144 (3.419)	0.739	-5.055 (6.088)	0.412	-0.161 (0.545)	0.769	-1.225 (0.826)	0.142
Charlson Index								
0		Reference Group		Reference Group		Reference Group		Reference Group
≥ 1	-0.900 (4.097)	0.827	-3.878 (5.402)	0.476	0.079 (0.637)	0.901	-0.083 (0.957)	0.931
SES								
1 (Highest)		Reference Group		Reference Group		Reference Group		Reference Group
2	3.724 (5.732)	0.518	-9.588 (10.633)	0.375	-0.103 (0.905)	0.910	-1.082 (1.383)	0.437
3	8.431 (5.748)	0.147	-5.472 (9.973)	0.587	-0.449 (0.909)	0.623	-0.362 (1.375)	0.793
4	2.845 (6.395)	0.658	5.686 (11.888)	0.636	-0.740 (1.017)	0.469	1.346 (1.522)	0.379
5 (Lowest)	9.865 (5.967)	0.103	-11.894 (11.750)	0.321	-0.601 (0.953)	0.530	0.115 (1.436)	0.936
Missing	4.403 (5.965)	0.463	-3.443 (10.180)	0.737	-0.537 (0.952)	0.574	1.588 (1.431)	0.271
Type of Surgery								
AA		Reference Group		Reference Group		Reference Group		Reference Group
TAR	2.459 (3.676)	0.506	-0.570 (5.391)	0.916	0.040 (0.586)	0.946	-1.351 (0.883)	0.131

.

Appendix B PRO Instruments

B.1 EQ-5D VAS



- We would like to know how good or bad your health is TODAY.
- This scale is numbered from 0 to 100.
- 100 means the <u>best</u> health you can imagine.
  0 means the <u>worst</u> health you can imagine.
- Mark an X on the scale to indicate how your health is TODAY.
- Now, please write the number you marked on the scale in the box below.

YOUR HEALTH TODAY = ☐

B.2 AOS

Instructions

The line next to each item represents the amount of pain you typically had in each situation. On the far left is "No pain" (0 points) and on the far right is "The worst pain imaginable" (100 points). Place a mark on the line to indicate how bad your **ankle** pain was in each of the following situations in the **past week**. If you were not involved in one or more of these situations, mark that item N/A.

Pain:

How severe was your *right ankle* pain?	No pain 0 — Worst pain imaginable 100	N/A
1. At its worst?	______________________________	☐
2. Before you got up in the morning?	______________________________	☐
3. When you walked barefoot?	______________________________	☐
4. When you stood barefoot?	______________________________	☐
5. When you walked wearing shoes?	______________________________	☐
6. When you stood wearing shoes?	______________________________	☐
7. When you walked wearing shoe inserts or braces?	______________________________	☐
8. When you stood wearing shoe inserts or braces?	______________________________	☐
9. At the end of the day?	______________________________	☐

Vancouver CoastalHealth
Promoting wellness. Ensuring care.

Providence HEALTH CARE
How you want to be treated.

Instructions

The line next to each item represents the amount of difficulty you had performing an activity. On the far left is "No difficulty" (0 points) and on the far right is "So difficult unable" (100 points). Place a mark on the line to indicate how much difficulty you had performing each activity because of your **ankle** in the **past week**. If you did not perform an activity during the past week, mark that item N/A.

Disability:

How much difficulty did you have?	No difficulty 0 — So difficult unable 100	**N/A**
1. Walking around the house?	________________________	☐
2. Walking outside on uneven ground?	________________________	☐
3. Walking four blocks or more?	________________________	☐
4. Climbing stairs?	________________________	☐
5. Descending stairs?	________________________	☐
6. Standing on tip toes?	________________________	☐
7. Getting out of a chair?	________________________	☐
8. Climbing up or down curbs?	________________________	☐
9. Walking fast or running?	________________________	☐

Vancouver CoastalHealth
Promoting wellness. Ensuring care.

Providence HEALTH CARE
How you want to be treated.

B.3 PEG

1. What number best describes your pain on average in the **past week**?

0	1	2	3	4	5	6	7	8	9	10
No pain										Pain as bad as you could imagine

2. What number best describes how, during the **past week**, pain has interfered with you enjoyment of life?

0	1	2	3	4	5	6	7	8	9	10
Does not interfere										Completely interferes

3. What number best describes how, during the **past week**, pain has interfered with you general activity?

0	1	2	3	4	5	6	7	8	9	10
Does not interfere										Completely interferes

PEG pain scale developed aby Drs. Erin Krebs, Karl Lorenz, Matthew Bair, and colleague.

Vancouver CoastalHealth
Promoting wellness. Ensuring care.

Providence HEALTH CARE
How you want to be treated.

B.4 PHQ-9

Instructions

Over the **past 2 weeks**, how often have you been bothered by any of the following problems?

	Not at all	Several days	More than half the days	Nearly every day
1. Little interest or pleasure in doing things.	☐	☐	☐	☐
2. Feeling down, depressed or hopeless.	☐	☐	☐	☐
3. Trouble falling asleep, staying asleep, or sleeping too much.	☐	☐	☐	☐
4. Feeling tired or having little energy.	☐	☐	☐	☐
5. Poor appetite or overeating.	☐	☐	☐	☐
6. Feeling bad about yourself—or that you're a failure or have let yourself or your family down.	☐	☐	☐	☐
7. Trouble concentrating on things, such as reading the newspaper or watching television.	☐	☐	☐	☐
8. Moving or speaking so slowly that other people could have noticed. Or, the opposite—being so fidgety or restless that you have been moving around a lot more than usual.	☐	☐	☐	☐
9. Thoughts that you would be better off dead or of hurting yourself in some way.	☐	☐	☐	☐

Patient Health Questionnaire (PHQ9) developed by Drs. Robert L. Spitzer, Janet B.W. Williams, Kurt Kroenke and colleagues.

Vancouver CoastalHealth
Promoting wellness. Ensuring care.

Providence HEALTH CARE
How you want to be treated.