Data Science Strategies for Modelling Total Knee Arthroplasty Patient Variability

by

Kathryn L. Young-Shand

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
February 2021

Dedicated to Rory Anna,

& to her younger sister who we look forward to welcoming.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Symbols Used

In alphabetical order:

AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
BMI	Body Mass Index (mass[kg]/height[m]^2)
CART	Classification and Regression Tree
CONSORT	Consolidated Standards of Reporting Trials
EMG	Electromyography
EQ-VAS	EuroQual Visual Analog Scale
EQ-5D	EuroQual Five-Dimension Questionnaire
HSD	Honestly Significant Difference
KOOS	Knee Injury and Osteoarthritis Outcome Score
KRES	Knee Replacement Expectation Survey
MCD	Minimum Detectable Change
MCID	Minimum Clinically Important Difference
OA	Osteoarthritis
OR	Odds Ratio
OKS	Oxford Knee Score
PCA	Principal Component Analysis
PC	Principal Component Vector
PC scores	Principal Component Scores
PCS	Pain Catastrophizing Score
PROM	Patient Reported Outcome Measure
RR	Risk Ratio
SF-36	Short Form 36-item health survey
SOM	Self Organizing Map
TKA	Total Knee Arthroplasty
VAS	Visual Analog Scale
WOMAC	Likert Western Ontario and McMaster Universities Arthritis Index

Abstract

Knee osteoarthritis (OA) can present through self-reported patient symptoms and joint-level manifestations, impairing mobility and function. Symptom relief and functional restoration of end-stage OA is typically treated with total knee arthroplasty (TKA) surgery. TKA success it not universal; over 20% of patients do not report satisfaction or clinically meaningful pain and function improvements post-TKA. This thesis aimed to improve our understanding of multidimensional variability among OA patients, and investigate how variability manifests into different TKA outcomes. Four studies investigated relationships between patient-reported measures and knee kinematics and kinetics during walking gait, characterized demographic and knee biomechanical variability (clusters) among OA and TKA populations, and addressed relationships between clusters and functional outcomes after TKA.

The first study objective examined demographic and patient-reported factors pre-TKA to two-years post-TKA associated with patient-reported TKA satisfaction using longitudinal analysis. Study 2 examine demographics, pre-TKA knee kinematics and kinetics during gait, and post-TKA gait changes associated with self-reported pain and function improvements one-year post-TKA using regression models. Studies 3 and 4 characterize demographic and knee biomechanical variability among (3) pre-TKA and (4) asymptomatic to post-TKA observations using machine-learning cluster analysis.

Studies revealed that (1) pre-operative self-reported symptoms were not predictive of longitudinal satisfaction; however, findings support the ability to identify less satisfied patients as early as six-weeks post-TKA. (2) Frontal and sagittal knee kinematic patterns during gait pre-TKA, and less adduction angle reductions post-TKA were associated with greater self-reported pain and function improvements post-TKA. (3) TKA candidates and (4) patients along the OA continuum can be characterized by demographic and knee biomechanic clusters, separated by features corresponding to clinical OA severity.

Collectively, these studies characterized temporal and multidimensional patient variability encompassing self-reported symptoms and knee biomechanics using data science and machine-learning strategies. Knee biomechanics provided important insights into mechanical factors impacting the patient experience, and biomechanical cluster profiling supported the ability to classify patients who may benefit most from TKA. Findings support the utility to reveal novel insights into the patient experience using advanced data science strategies, providing direction for innovations in OA management and TKA care. Presented methodologies are directly applicable to other clinical applications.

Acknowledgements

I appreciate the opportunity to extend my thanks to the community of individuals whose help and support made this work possible.

I particularly would like to thank my supervisors and mentors. Janie Astephen Wilson, for unconditional support from start to finish. I am grateful you provided me the freedom to shape this scope, and encouragement to explore analysis tools and classrooms of various departments. Perhaps most importantly, thank you for the invaluable mentoring academically, in motherhood, and life. Michael Dunbar, for always advocating for me, providing expertise, bold ideas, and a motivational perspective. I've always found your ideas the best anchors for my work. Raza Abidi and Joanna Mills Flemming, I am so grateful to have had access to your exceptional knowledge and mentorship. Thank you for contributing so substantially to my technical and professional growth - I learned so much from you both.

I would also like to express sincere thanks to Patrice Roy and JoAnne Douglas, the support of the Dynamics of Human Motion group and the orthopaedic research community at the Halifax Infirmary. In addition, Elysia Davies and Elise Laende for always raising the bar.

Finally, I would like to thank my family. Andrew for endless auxiliary support, truly making this possible. Rory, for demonstrating true determination and persistent happiness.

Chapter 1. Background and Literature Review

1.1. Knee Osteoarthritis: A Disease of Phenotypes

Knee osteoarthritis (OA) is a degenerative joint disease that involves the entire joint during the disease process, including disruptions to joint cartilage, underlying bone, and the soft-tissue structures of the synovium, ligaments, and musculature surrounding the joint (Felson et al., 2000). Disease prevalence increases with age, and obesity, disproportionally affecting females (Dillon et al., 2006; Felson et al., 1987). OA development has also been attributed to joint level biomechanical factors, such as abnormal loading in response to deformity, obesity, or posttraumatic joint injury, in addition to systemic factors, such as genetics and metabolic influences (Bijlsma et al., 2011; Felson et al., 2000). OA can present in terms of symptoms characterized by joint pain, swelling, and stiffness, or radiographically as shown by cartilage degradation and osteophyte presence within the joint structure (Bijlsma et al., 2011; Kellgren and Lawrence, 1957; Schaible, 2012). These burdens largely contribute to patient disability (Lang et al., 2018), difficulty walking (King et al., 2018), and all-cause mortality (Liu et al., 2017), making knee OA the fifth most common reason for inpatient hospitalization in Canada (CIHI, 2015a).

In recognition of the complex onset and manifestation pathways of OA, recent efforts have been made to characterize the heterogeneity of OA in terms of phenotypes. Dell'Isola and Steultjens discussed phenotypes driven by pain, joint disease severity, mechanical alignment, inflammatory factors, metabolic disorders, and bone and cartilage metabolism (Dell'Isola and Steultjens, 2018). Kittleson et al. characterized phenotypes associated with pain, psychological distress, and radiographic severity (Kittelson et al., 2016). Knoop et al. identified phenotypes by radiographic severity, muscle strength, body mass index (BMI), and depression scores (Knoop et al., 2011). Elbaz et al. reported phenotypes using spatiotemporal gait parameters (Elbaz et al., 2014). Further, Waarsing et al. identified four clusters by demographics, cartilage features, mechanical alignment, and self-reported symptoms, with clusters strongly characterized by joint-level structural

degradation (Waarsing et al., 2015), indicative of variable functional loadings (Andriacchi et al., 2009). Despite variability among studies, understanding possible phenotypes is relevant to developing treatments targeting individual OA manifestations or deficiencies.

However, current phenotyping research remains in initial stages of study, and there are a number of factors that inhibit clinical utility of current findings. Reported phenotypes differ by study design and distinguishing features, whether they are descriptive of disease manifestation processes, or in terms of their clinical presentation. They tend to lack analysis on clinical relevance, such as treatment outcomes between phenotypes, or information on repeatability or stability on external datasets. Phenotypes have also been proposed to evolve over the course of the disease process (Castaneda et al., 2013), where end stage joint degradation converges along a common pathway for all severe knee OA patients (Felson, 2010). Yet, current phenotypes tend to be derived from cross-sectional datasets, captured at various times within the OA progression pathway (Deveza et al., 2017; Knoop et al., 2011; Waarsing et al., 2015). A clear approach to the context in which phenotypes are characterized is important if we want to distill multidisciplinary phenotype information into relevant and targeted OA management and intervention approaches specific to phenotype manifestations (Van Spil et al. 2020).

1.2. Total Knee Arthroplasty

Total knee arthroplasty (TKA) is the most prevalent end state treatment for knee OA, and the second most common inpatient surgery performed in Canada (CIHI, 2015a). It is an elective surgery which involves the removal of native disease-compromised bone and replacing it with an articulating joint prosthesis composed of metal and plastic. The longstanding goal is the restoration of normal, natural function of the knee joint, and to relieve patient pain (Andriacchi, 1993; Noble et al., 2005). Treatment efficacy for TKA has been well established for end stage knee OA (Giesinger et al., 2014; Skou et al., 2015), contributing to the increasing volumes of joint replacements being performed each year (CIHI, 2015b). However, surgical success is not guaranteed. Patient reported

satisfaction rates remain around 80%, low relative to other orthopaedic procedures; over 2% of patients are at risk of revisions; and readmissions impact over 6% of patients within three months (Bourne et al., 2010; Husted et al., 2008; Noble et al., 2006; Scott et al., 2012). Surgical prevalence is also increasing among a changing demographic of younger and more physically demanding patients with higher functional expectations following surgery (Ravi et al., 2012; Scott et al., 2012), influencing surgical outcomes. For example, the Canadian National Joint Registry has shown that TKA recipients under age 65 have a greater three-year revision rate relative to the national average, with more revisions occurring among females (CIHI, 2015b). Registry studies in England have found patients under 55 to have greater incidence of revisions, and lower rates of satisfaction (Williams et al., 2013b). Adverse outcome frequency among specific demographics of patients has raised concerns of the potential overuse of joint replacement, and performing costly procedures without an indication of potential risks or benefits (Clavel et al., 2016). Poor outcomes in a subset of patients after TKA also signals the continued need to improve our understanding of outcome variability after TKA, and determine which patients are at risk for adverse outcomes, to drive evidence-based surgical selection (Clavel et al., 2016; Kurtz et al., 2009) and appropriate models of care (Allen et al., 2016).

A Standard Surgical Approach

Despite evidence of variability among patients presenting for TKA, the surgical standard of care in traditional TKA is to induce a neutral (within 0 ± 3˚) "mechanical axis" alignment in all individuals through femoral positioning using intramedullary guides and two-dimensional frontal plane radiographs. This was historically associated with higher survival rates and reduced polyethylene wear (Bargren et al., 1983; Fang et al., 2009; Lotke and Ecker, 1977), however, a number of authors have proposed that this standard approach to joint alignment may not be optimal for all patient populations (Mooney et al., 2016; Parratte et al., 2010; Vanlommel et al., 2013). For instance, healthy populations have demonstrated large ranges in frontal plane alignment, statistically deviating from the defined 0 ± 3˚of a straight mechanical axis (Eckhoff et al., 2005;

Moreland et al., 1987). Despite intraoperative efforts to prevent high-risk loads, alignment achieved intraoperatively does not consistently relate to dynamic loading as captured during gait (Orishimo et al. 2012; Miyazaki et al. 2002; Rodriguez et al. 2016) and more immediate post-TKA loading does not reflect longitudinal patterns (Orishimo et al. 2012). Further, Vanlommel et al. and Salzmann et al. reported significantly better self-reported knee scores in pre-operatively varus individuals, whose post-operative alignment remained mildly varus (Salzmann et al., 2017; Vanlommel et al., 2013). Parratte et al. found no longitudinal survival improvement in 389 modern prostheses that deviated from a conventional axis after arthroplasty, attributing possible causes of failure to patient factors that included dynamic impact and gait (Parratte et al., 2010). These findings motivate the notion that TKA standard of care may not be optimal for all individuals, and in an effort to extend longevity, dynamic function may be a more effective target over frontal plane alignment alone. Innovations in TKA, such as robotic surgery will aim to address patient-specific mechanical targets. However, these innovative approaches are in their early stages, where current robotic surgeries only improve the precision of conventional TKA alignment practices (Mancino et al., 2020).

Selecting Patients for TKA

There is no standardized criteria for appropriate patient selection in TKA, where current triaging is based on clinical opinion. Radiographic scores, high patient pain and low quality of life have been ranked as the greatest contributors to TKA patient selection by Canadian orthopaedic surgeons (Frankel et al., 2016). However, a statistically driven patient selection protocol would require an understanding of the symptom state most associated with a defined optimal outcome after arthroplasty (Hawker et al., 2013; Losina and Katz, 2013). Since current knee arthroplasty methodology can be regarded as a standardized procedure catering to the norm, the clinical utility in characterizing TKA candidate variability may be in the ability to statistically identify patients who are at high risk pre-operatively, and to make clinical decisions that are appropriate to the patient's health situation, as some candidates may fair well from an altered conservative treatment strategy (such as exercise and physiotherapy) over TKA (Skou et al., 2015). In the

Canadian context, efforts have been made to inform the development of a surgeon-patient decision-support tool (Hawker et al., 2015). Its development incorporated a mixed panel to address the needs of OA patients (Frankel et al., 2012), clinicians (Frankel et al., 2016), and policy-makers (Clavel et al., 2016), from which six baseline criteria were defined. Although the effective development a decision-support tool could improve patient selection and mitigate the burdens associated with poor outcomes, the utility has been met with skepticism (Frankel et al., 2016; Wright et al., 2002), and at current, tools may not meet all desired needs from a policy standpoint (Clavel et al., 2016). More work is required to statistically identify high risk patients through an appropriate selection criterion, and mitigate the burdens associated with poor outcomes (Clavel et al., 2016; Kurtz et al., 2009).

1.3. Assessing TKA Outcomes

1.3.1. Patient-Reported Outcome Measures

TKA is commonly evaluated through patient-reported outcome measures (PROMs), which reflect patient symptoms and symptomatic relief associated with arthroplasty. PROMs are paper or electronic-based form tools that allow patients to self-report their perceptions of pain, quality of life, and function by rating perceived symptoms or difficulty performing daily tasks (e.g., walking, or climbing stairs). Treatment efficacy for TKA has been well established for end stage knee OA using PROMs criteria (Giesinger et al., 2015; Skou et al., 2015). PROMs in orthopaedics have been encouraged (Hurwitz et al., 2000), and tied to reimbursement in the United States (Centers for Medicare and Medicaid Services 2015). In England, national PROMs collection has been mandated, with the goal of using PROMs to understand risk factors for poor outcomes and prioritize patients for surgery (Health and Social Care Information Centre 2011).

Part of the utility of PROMs reflects the overall success of TKA. Typical clinical endpoint indicators such as mortality or revision are relatively rare after TKA. For example, revisions might impact approximately 2.3% of TKA patients in Canada (CIHI, 2015b), while up to half of TKA patients have been shown to report no clinically

meaningful improvement in subjective outcome scores (Hawker et al., 2013). Therefore, in addition to providing important information on the patient experience, PROMs also provide a more sensitive metric due to the higher incidence of poor responses relative to clinical indicators. This balanced incidence has advantages during outcomes modeling, improving the power of prediction models when limited sample sizes are available. As we aim to continue to improve TKA outcomes, PROMs provide an appropriate measure for surgical success.

Types of PROMS

As defined by the Patient-Reported Outcomes Measures Working Group of the International Society of Arthroplasty Registries, PROMs can be categorized based on two application criteria: generic (or general health) scores and site specific scores (Rolfson et al., 2016). The former explains scores capturing overall patient health status, applicable in both the presence or absence of a disease or symptom. Examples include the Short Form 36-item (SF-36) general health score, using eight domains to address overall physical functioning, bodily pain, general health, vitality, social functioning, and emotional and mental health (McHorney et al., 1994, 1993). Similarly, the EuroQual (EQ) Visual Analog Scale (VAS), and the EuroQual five dimension (EQ-5D) general health score capture dimensions representing difficulty with walking, self-care, performing usual activities, pain/discomfort and anxiety/depression (Bansback et al., 2012; Brooks, 1996).

Specific scores are designed to capture perceptions about a distinct disease or symptom, focusing on an area of interest. In knee arthroplasty, common joint-specific surveys include the Oxford-12 Knee Score (OKS) having domains for both knee pain and function (Dawson et al., 1998), and the Knee Injury and Osteoarthritis Outcome Score (KOOS) with five domains for pain, other symptoms, function in daily living, function in sport and recreation, and knee related quality of life (Roos et al., 1998). The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) addresses three dimensions for pain, stiffness, and physical function (Griffiths et al., 1993), and in addition, the VAS Pain Score can be applied to a specific pain site of interest (Dolan and

Sutton, 1997; Ferraz et al., 1990; Huskisson, 1974; Singer and Jr., 1998). Despite common use, satisfaction scores (e.g., VAS 0[unsatisfied] to 100[completely satisfied]) (Bullens et al., 2001; Dolan and Sutton, 1997; Singer and Jr., 1998), and expectation fulfillment (e.g., Knee Replacement Expectations Survey (KRES, 19[highest]-95[lowest])) (Mancuso et al., 2001), do not clearly fall under the clinical PROM definition (Rolfson et al., 2016). However, such scores can capture the patient experience not addressed otherwise and have utility when used in combination with other measures.

Both general health scores and disease specific tools are designed to measure independent constructs. In the assessment of OA and TKA, the score selected should be optimized to "accurately detect the type of change being quantified" (Beaton et al., 2001). Joint-specific scores have been demonstrated to be more sensitive to post-operative changes from TKA (Giesinger et al., 2014; Impellizzeri et al., 2011), and less susceptible to floor and ceiling effects (Escobar et al., 2007; Giesinger et al., 2014), with common tools such as WOMAC and OKS generally performing well among arthroplasty populations (Dunbar et al., 2001; Giesinger et al., 2014). However, the joint is not in isolation, and joint-specific scores alone may fail to capture overall health or constructs such as general satisfaction. As such, combinations of scores are often used to provide a more comprehensive picture of the patient experience.

Using PROMs in Risk Factor Analysis

Extensive research has been conducted to address risk factors associated with undesirable outcomes from TKA. Many pre-operative variables have been identified related to improved likelihood of optimal surgical outcomes. For example, patients who present with less pain while sitting or lying down (Bourne et al., 2010), more self-reported pain and function severity (Scott et al., 2010), better general health scores (Baker et al., 2013), better mental health (Scott et al., 2010; Vissers et al., 2010), or not having a previous diagnosis of depression (Baker et al., 2013) have been associated with greater likelihood of TKA satisfaction. Post-operatively, improved satisfaction has been attributed to pain relief (Bourne et al., 2010; Mannion et al., 2009; Scott et al., 2010; Vissers et al., 2010), greater health-related quality of life (Scott et al., 2010; Vissers et al.,

2010), better self-reported function (Bourne et al., 2010; Scott et al., 2010; Vissers et al., 2010), and expectations being met (Bourne et al., 2010; Vissers et al., 2010). Older age (but not too advanced in age (Bourne et al., 2010)) and male sex have also been associated with better satisfaction relative to female (Baker et al., 2013; Williams et al., 2013a). Patients presenting with worse self-reported pain, function, less co-morbidities, less pain in other joints, have been associated with greater improvements in self-reported WOMAC and OKS scores (Hawker et al., 2013; Judge et al., 2012). Finally, contrary to satisfaction findings, younger age has been associated with more improvement in OKS and EQ-5D outcomes (Williams et al., 2013b).

Despite exhaustive literature and routine PROMs collection, we continue to have an insufficient understanding of key drivers of patient outcomes, and what features (or combinations of) are most meaningful in outcome prediction. Research methods and outcome definitions also vary. Outcomes can be described in terms of "the destination" such as satisfaction (Bourne et al., 2010; Scott et al., 2010; Williams et al., 2013a) and expectation fulfillment (Scott et al., 2012). Alternatively, we can address outcomes in terms of patient responsiveness, "the journey", captured by improvements in metrics such as pain or function (Escobar et al., 2013; Hawker et al., 2013; Williams et al., 2013a). This variability has resulted in conflicting responses (such as the influence of patient age), and a poor ability to translate findings into clinical use. As exceptionally stated by Losina and Katz "laying out a clear definition of what success means would allow us to more effectively target and address the challenges that prevent some patients from achieving it" (Losina and Katz, 2012).

Using PROMs as Responsiveness Scores

Recent emphasis on PROMs collection has led to the use of "responsive" scores to analyze changes in patient reported measures, i.e., the journey. Two examples of response measures include the minimal clinically important difference (MCID), and the minimally detectable change (MDC) criteria, which define statistical thresholds for pre to post-TKA PROMs changes. The MCID anchors PROM scores next to patients' perceptions of "improvement" to detect clinically meaningful score changes, while MDC

is a distribution-based method, representing the smallest within-individual change that is larger than that of measurement error. Assessing group average change, or the percent of a group that achieved an MCID in an aggregated context is applicable (Quintana et al., 2006), but more complexities need to be considered when making decisions for an individual patient (King, 2011). This is because MCID scores defined by anchoring methods can be heavily influenced by symptom severity at baseline (Escobar et al., 2013), such that global responsiveness scores are not generalizable across all patients. Response scores also need to be specific to the procedure they are evaluating, and MCID developed for another treatment may not be sufficient for arthroplasty. For example, in a study conducted by Hawker et al. (Hawker et al., 2013) an MCID cut-off of 13.4 was applied in a pain domain of the WOMAC score for TKA patients based on a rehabilitation intervention among OA populations (Angst et al., 2002), differing from a 23 point cut-off for post-TKA populations using the same questionnaire (Escobar et al., 2007). Finally, PROM score ceiling effects may restrict rates of patients meeting MCID thresholds, despite having improved. Still, responsiveness scores have been recommended as part of PROMs assessment strategies by the International Society of Arthroplasty Registries Working Group (Rolfson et al., 2016), due to their ability to improve within and between-patient score interpretations from interventions. The limitations of these variables needs to be carefully understood prior to application to statistical techniques with the intention of informing patient care strategies.

Limitations of PROMs

The pre-operative condition of a patient has been shown to have a substantial impact on status post-operatively (destination), and the level of gain achieved (journey) (Dunbar et al., 2013; Fortin et al., 1999; Hawker et al., 2013; Jiang et al., 2017; Judge et al., 2012). In an early multicenter study conducted by Fortin et al., patients were stratified into high and low functioning groups pre-TKA based on median WOMAC function scores, and 6-month patient-reported outcomes were compared between groups. Although both patient groups improved pre to post-TKA, the largest relative improvement was reported in the low-functioning group, even though post-TKA scores remained worse than the scores of

the high functioning cohort (Fortin et al., 1999). This concept was confirmed by other authors (Hawker et al., 2013; Jiang et al., 2017; Judge et al., 2012), where patients with worse pre-operative scores pre-TKA generally experienced a larger degree of improvement from surgery, however mean post-operative status may not reach that of the higher-functioning cohort. If a single patient were to be selected for surgery by their potential for functional improvement (journey) relative to their post-operative performance alone (destination), each measure would favor a different candidate. This concept is part of the ongoing determination of whether surgical success should be based on the degree of patient-perceived improvement, or the discrete status achieved post-operatively (Losina and Katz, 2013).

As a qualitative measure, PROMs experience a large degree of variability independent of arthroplasty and OA disease, and clinicians and researchers are well aware of the confounding limitations of PROMs. PROMs can be influenced by co-morbidities (Dunbar et al., 2004; Hawker et al., 2013), baseline score status (Dunbar et al., 2004; Hawker et al., 2013; Jiang et al., 2017), patient demographics (Dunbar et al., 2004; Jiang et al., 2017; Williams et al., 2013a), mental health (Jiang et al., 2017), expectations (Sullivan et al., 2011), perceptions of self-efficacy (Maly et al., 2006), injustice (Yakobov et al., 2014), and pain (Dave et al., 2017; Yakobov et al., 2014). This poor sensitivity means that a proportion of a poor functional response may be a reflection of auxiliary dimensions of patient health or perceptions. Further, some confounding factors such as mental health, depression, expectations, and anxiety pose challenges because they are seldom captured, and difficult to quantify. It is also often assumed that the independent measures included as patient factors adequately covered all modes of variability that would be required in determining the outcome. However, stochastic (time varying) patient responses may reflect biological or behavioral variability associated with day-to-day changes and perceptions, where the cause may be unknown. These inferences are therefore liable to induce response bias when data is only assessed at one or two time points, limitations that can be reduced through analysis using repeated measures.

OA symptom severity described by a patient does not always align with the disease progression governed by radiographic scores (Bedson and Croft, 2008), therefore PROMs

do not provide an accurate indication of severity. Missing data is also commonplace in healthcare, and there is no established "acceptable" response frequency in PROMs registry-based research. To fill this gap, the Internarial Society of Arthroplasty Registries PROMs Working Group recently proposed a response threshold of 60%, recognizing patient reported outcome scores can be difficult to capture for reasons that are independent of survey logistics, with some jurisdictions performing better than others (Rolfson et al., 2016). In sight of these limitations, all incomplete or missing values should be reported at a per-item level to improve understanding of response bias (Rolfson et al., 2016). Data from questionnaire respondents who do respond may also not be representative of the true clinical population, reflecting a non-random subset of the population. Those with missing elements in arthroplasty studies have been characterized by worse pain and function (e.g., 15.2 vs 18.2, OKS scored 0-48) (Jiang et al., 2017), poorer mental health status, a greater prevalence of co-morbidities (Dave et al., 2017), and lower rates of satisfaction (Dunbar et al., 2001), resulting in analyses being conducted on an optimistic subset of the population. Finally, it is critical to note that most PROMs scores were designed to capture perceived symptom status (such as pain or function). They were not designed for predictive applications, and there is little evidence to support use of them for purposes such as prioritizing access to care (Baker et al., 2013; Judge et al., 2011).

Unfortunately, from assessing PROM responses alone, it is impossible to tease out potential underlying factors associated with poor responses. For example, from PROMs we know that upwards of 11% of patients will show no clinical improvement in terms of pain and function after TKA (Alzahrani et al., 2011; Hawker et al., 2013). Patients who report poor responses immediately following TKA will continue to experience pain and dissatisfaction at 10-15 years, without a quantifiable radiographic or clinical premise for their response (Ali et al., 2014; Robertsson et al., 2000). When compared to age-matched control groups with no knee disorders, TKA recipients also report larger functional limitations on self-reported scores (Noble et al., 2005). Perceptions in domains of function poorly translation into objective functional status, as they can be bias multidimensional confounding, narrowing the utility of PROMs when assessed in

isolation. In short, subjective measures scored between 0-100 provide clinicians with little insights into understanding the mechanisms associated with perceived limitations. Objective measures, including objective gait function might provide a complementary measure for understanding the cause of poor self-reported outcomes, and provide a vehicle for characterizing which patients experience the most functional improvement from arthroplasty and why.

1.3.2. Biomechanical Gait Assessment

Although not routinely incorporated into TKA evaluation, functional assessments provide an alternative means to measure patient functional status or outcome. Performance-based functional tests include the Timed Up and Go (TUG), Six-Minute Walk, Stair Climbing Tests and Sit to Stand Tasks, which demonstrate a patient's ability to complete tasks unobstructed by confounders associated with patient perceptions. However, not all functional assessments are able to objectively quantify how a joint performs mechanically. Three-dimensional high accuracy dynamic gait analysis during walking provides a means to capture dynamic function at the joint level. Gait mechanics have been shown to worsen with osteoarthritis progression (Astephen et al., 2008a), severity seen on radiographs (Wilson et al., 2011) and pain (Henriksen et al., 2012; Maly et al., 2008). TKA aims to improve this worsened knee function, in part by improving patient gait (Andriacchi, 1993; Noble et al., 2005), where walking has been rated the most important activity of daily living by TKA patients (Noble et al. 2011; Scott et al. 2012). Therefore joint-level gait assessment provides a relevant means to objectively assess function and performance of the joint both pre and post-operatively.

Modern gait analysis uses rigid limb segment kinematic data and ground reaction forces to a mechanical model to determine joint motion and loading. It can provide an approximation of three-dimensional knee (and other lower extremity joint) angles and external moments defined anatomically in terms of flexion-extension, abduction-adduction, and internal-external rotation. Raw data is captured using motion capture technology and a force platform system to acquire external ground reaction forces. Knee kinematics (knee motion) is captured by high accuracy optoelectronic motion capture systems, describing the three-dimensional motion of the joint, such as flexion range of

motion, in a limb segment rigid body coordinate system (Grood and Suntay, 1983). Kinetics (knee moments), are calculated using ground reaction forces and joint segment models, working up the kinetic chain from the most distal segment to the joint of interest (Costigan et al., 1992; Li et al., 1993; Vaughan et al., 1999). These models account for acceleration and mass moment of inertia at the segment level, providing approximations of the net resultant forces and moments acting on the joint. With appropriate biomechanical modeling, gait assessment provides an objective approach to examining joint-level performance with particular insights into the mechanical joint motion and loading environments.

Gait Mechanics During OA Progression

OA progression is marked by increasing pain and symptoms, and structural joint-level changes. During OA progression, severity can typically be characterized by the "stiffening" of the knee-joint during gait. In the sagittal plane, flexion and extension range of motion and range of moment magnitudes reduce with OA severity, resulting in less bi-phasic gait patterns (Astephen et al., 2008b). In the frontal plane, magnitudes of the knee adduction moment increase (Astephen et al., 2008a), with less mid-stance adduction moment un-loading (Hatfield et al., 2011), describing greater sustained medial compressive loads during weight bearing (Schipplein and Andriacchi, 1991). Greater overall knee adduction moment loading has been associated with cartilage volume loss (Bennell et al., 2011), and progression to TKA (Hatfield et al., 2015a). Both pre and post-TKA, larger adduction moment magnitudes have also been linked to tibial component migration (Hilding et al., 1996; Wilson et al., 2010), an early indicator for implant failure (Ryd et al., 1995), making loading an important risk factor for joint longevity.

There are generally two perspectives on "stiffening" of knee joint patterns with OA progression. One hypothesis suggests that some irregular gait patterns pre-date symptoms, and observed features such as a lack of extension are inducing greater moments on articular surfaces not adapted to handle maximal loads, thereby resulting in degeneration and pain (Andriacchi et al., 2004). The alternative proposes that "stiffer" features in the sagittal and frontal planes, such as less dynamic range of motion and

prolonged moment magnitudes, are a protective guarding mechanism against pain during loadbearing and avoidance of large flexion moments. Altered mechanics could also reflect a combination of these proposed manifestations, where in both interpretations, decreased function coincides with pain. Sagittal plane range of knee motion has been reported to explain up to 18% of pain intensity experienced in OA patients, where a 1° decrease in flexion and extension range corresponded to a 3% increase in patient-reported pain (Maly et al., 2008), with similar associations post-TKA after the removal of native articular surfaces (Bonnefoy-Mazure et al., 2017; Turcot et al., 2013). Inducing knee pain in healthy subjects using hypertonic saline injections, has also been found to result in a significant decrease in peak knee flexion moments, extension moments, and adduction moments (Henriksen et al., 2010). Similarly, pain relief through prescription drug use has been shown to result in increased flexion angles, adduction moment peaks, and gait speed in sample of 18 patients (Schnitzer et al., 1993). Lastly, in a recent analysis between asymptomatic and symptomatic individuals, both with evidence of structural knee OA (Kellgren-Lawrence Grade of 2), those with worse self-reported WOMAC pain, stiffness, and function scores, walked slower, with lower overall magnitudes of knee flexion moments, and less dynamic adduction moment magnitudes, suggesting symptoms, not the disease, are the key drivers to gait changes (Kellgren and Lawrence, 1957; Wilson et al., 2017).

Gait Mechanics after TKA

Post-TKA, biomechanical gait kinematic and kinetic patterns alter by shifting towards normative asymptomatic gait ranges, interpreted as a functional improvement post-TKA (Hatfield et al., 2011; Sosdian et al., 2014). For example, gait after surgical intervention has decreased overall knee adduction angles (Naili et al., 2017) and knee adduction moment magnitudes (Hatfield et al., 2011; Wilson et al., 2015). The latter contributing to more bi-phasic unloading of the joint during stance, an improvement relative to severe OA mechanics. In the sagittal plane, corrections include increases in knee flexion angle magnitudes (Hatfield et al., 2011; Naili et al., 2017; Wilson et al., 2015), flexion and extension motion (Bonnefoy-Mazure et al., 2017; Naili et al., 2017), and moment ranges

(Hatfield et al., 2011; Wilson et al., 2015). However, post-operative patterns do not return to healthy asymptomatic patterns (Paul Robert Biggs et al., 2019; McClelland et al., 2010; Naili et al., 2017; Outerleys et al., 2021), with the most discriminating characteristics between asymptomatic and post-TKA mechanics occurring in the sagittal plane (Outerleys et al., 2021). The concept of a biomechanical ceiling effect, such that post-TKA gait patterns statistically map to disease-state gait over asymptomatic gait has been proposed in two recent studies (Paul Robert Biggs et al., 2019; Outerleys et al., 2021). Gait analysis provides a means to quantify improvements, capture deficiencies in TKA gait relative to normative function, and inform operative strategies targeting specific mechanical corrections. Further, a lack of flexion angle and moment improvement post-TKA has been associated with patient-reported pain after TKA (Naili et al., 2017; Smith et al., 2004). These findings suggested perceived post-operative improvements experienced by patients may incorporate dynamic improvements in the sagittal plane.

Only more recently has gait analysis been used to investigate how standard of care surgery may impact patients differently, such as male and female cohorts (Wilson et al., 2015), varus and valgus alignment (Rodriguez et al., 2016), and obesity (Paterson et al., 2020, 2017). For example, women have been shown to present for TKA less range of flexion and extension motion, less biphasic sagittal stance moments, and with greater flexion moment magnitudes, descriptive of a stiffer gait with a more constant sagittal loading pattern (Paterson et al., 2017; Wilson et al., 2015). Alternately, male surgical candidates have been characterized by greater magnitudes of frontal plane adduction moments. Although males and females both improved from surgery, females lagged in sagittal loading corrections, and male adduction moments magnitudes remained relatively heightened post operatively. Therefore, both sexes presented with, and resulted in, different gait features that would be characterized as unfavorable (Hilding et al., 1996; Wilson et al., 2010), with surgery addressing different function deficits between them. The reason for this is likely related to pre-operative functional status (Naili et al., 2017), sex-specific joint morphologies (Mahfouz et al. 2011) and neuromuscular control (Wilson et al., 2015). Similar indications have been captured in the examination of

alignment cohorts. Valgus aligned individuals pre-TKA have been shown to experience a 14 times increase in knee adduction moment impulses from surgery (Rodriguez et al., 2016). This degree of mechanical alteration was much larger than that experienced by their varus aligned counterparts, who showed a more typical adduction moment reduction from surgery. Patients whose surgical alignment more closely resembled their pre-operative alignment have also been demonstrated to report better outcomes (Parratte et al., 2010). Similarly, Naili et al. proposed poor patient-reported outcomes to be partially explained by a lack of dynamic kinematic and kinetic corrections, despite alignment corrections in the frontal plane, a feature that surgery may be most able to address biomechanically (Naili et al., 2017). Therefore, despite the prescribed mechanics from standard TKA, patient-specific characteristics, such as sex or alignment, may be inducing variable mechanical changes, potentially contributing to the causes of functional and self-reported limitations among some cohorts. It has been expressed that changes in post-operative gait parameters should be investigated against patient-reported factors of pain, function, and quality of life, to determine the clinical relevance of gait corrections (Sosdian et al., 2014). If this variability can be properly characterized, there is an opportunity to treat patients on an individual basis incorporating functional parameters.

Orthopaedic surgeons have expressed a need to measure function in a way that is meaningful (Frankel et al., 2016). When applied to clinical practice, gait may offer an objective assessment of the disease and functional recovery in a form that is less obstructed by the noise of co-founding dimensions, while also playing an important role in extending prosthesis longevity. In research, joint-level analysis can provide orthopedic surgeons and biomechanists with an objective approach to restoring normal function and directing them to the functional mechanisms that may be inhibiting task performance. Further investigations of mechanical cohorts and linking gait mechanics to other forms of outcomes may benefit our understanding of what corrections are clinically valuable among patient groups, and enable us to predict who may benefit most from standard of care procedures, and who might benefit from different functional targets.

1.4. Data Science Applications in Total Knee Arthroplasty

Directives to measure the value delivered from current care practices, and to select and prioritize patients accordingly (Clavel et al., 2016; Hawker et al., 2013), have been accompanied by a surge in registry programs and new demands on observational databases not designed for predictive modeling. Early valuations of these programs have reported no meaningful changes in patient selection, patient outcomes, or change in provider behaviors (Varagunam et al., 2014). A poor ability to realize and translate clinical and policy benefits from these programs leads to wasted expenditure, a lack of confidence in providers, and negative preconceptions and skepticism toward subsequent innovation (Frankel et al., 2016; Wright et al., 2002). Traditional analysis techniques on independently collected datasets may be limiting the value derived from current outcome programs, and overlook important signals in our data. There is a need to utilize the data we have, conduct linkages between diverse multidisciplinary datasets, and apply them to novel model frameworks. An open-minded approach towards new statistical and computer science data analysis strategies on linked datasets may provide novel information on the nature of our existing data. This may point us towards data features that are meaningful, and provide evidence for where time and resources should be dedicated for future data collection, guiding targeted research and deriving valuable clinical utility for OA and TKA care.

1.4.1. Principal Component Analysis

Principal Component Analysis (PCA) is a multivariate statistical technique and data reduction tool. Applications of PCA on gait waveforms has demonstrated the ability to characterize pattern changes in waveforms to improve analysis (Deluzio and Astephen, 2007). Therefore, instead of conducting analysis on discrete points in a waveform, PCA offers a means to capture and "score" temporal changes, and apply overall waveform scores to statistical tests. Starting with a data matrix, X, data is centered about zero by subtracting the mean from each variable ($X_{centered} = [X - \bar{x}]$), and the covariance matrix, S, is calculated. The sum of the diagonal of S, returns the total variance in the dataset. The eigenvectors, U, of the covariance matrix describe the principal component vectors, which we call PCs; these are the dominant features in the original dataset. Next,

a change of basis is applied to represent the uncorrelated data set as a linear combination of the original matrix X using the equation $Z = U'[X - \bar{x}]$. This describes a projection of the original subjects' waveforms onto the PCs, resulting in a new uncorrelated dataset, defined by matrix Z (PC scores). PC scores can be applied to waveform interpretation and hypothesis testing to quantify group differences. The variability explained by each PC can be determined by dividing the eigenvalue by the total variability. The variability explained is highest for the first PC and descends thereafter, and therefore the majority of the variation can be explained by the first few PCs.

Our research group has applied PCA to capture unique kinematic, kinetic and electromyography (EMG) pattern changes in gait at different states of OA progression and detected significant waveform pattern changes pre to post-TKA (Astephen et al., 2008a; Hatfield et al., 2011, 2015a). PCA has been used to link gait patterns with patient symptoms after TKA (Smith et al., 2004), and to discriminate joint morphology features among 1000 knees stratified by sex and ethnic groups, used to described six morphological joint shapes (Mahfouz et al., 2011). To understand biomechanical variability, PCA was also advantageous in the characterization of intraoperative navigation kinematics in a sample of 340 curves, mathematically defining six dominant interpretable patterns, and characterizing how patterns change between the pre to post-implant states (Young et al., 2015).

1.4.2. Regression and Longitudinal Analysis

Multivariate regression models are the most commonly used statistical tool in the determination of predictive factors of outcomes after surgery. Regression techniques can estimate factors that contribute to TKA progression (Hatfield et al., 2015a), and improved patient satisfaction (Baker et al., 2007; Bourne et al., 2010; Noble et al., 2006; Scott et al., 2010), mobility, function and pain (Alzahrani et al., 2011; Hawker et al., 2013; Turcot et al., 2013) and to understand the relationship between gait variables and PROMs (Smith et al., 2004; Turcot et al., 2013).

Longitudinal data analysis models are an expanded case of general multivariate regression modeling, however longitudinal data offers the opportunity to account for

dependencies caused by repeated measures (i.e., within-individual changes in a response variable over time). Longitudinal analysis also provides a means to provide more accurate estimates of covariates and predictive ability, which could be subject to bias when assessed by cross-sectional inference alone. Further, generalized mixed effects models are robust in the handling of missing data (Gibbons et al., 2010), and making inferences about an individual (not a population) through the inclusion of random effects. Despite this, of the number of studies that have assessed predictors for poor patient outcomes in the literature, very few have assessed outcomes with more than two time points in the short or long term (Dave et al., 2017; Edwards et al., 2009; Høvik et al., 2016; Jiang et al., 2017; Williams et al., 2013a). Fewer still employed longitudinal data analysis techniques (Edwards et al., 2009; Høvik et al., 2016; Jiang et al., 2017; Williams et al., 2013a).

1.4.3. Data Mining and Machine Learning

Applications for data mining and machine learning are emerging as new tools in the study of health outcome assessment. They are capable of extracting interesting and novel patterns from large datasets, and offer an improved ability to build novel relationships and predictions that can aid in our characterization of patient variability, and relate variability to outcomes after TKA.

Data-mining and machine learning methodologies can generally be divided into i) unsupervised learning, and ii) supervised learning strategies. During unsupervised learning, the data is not labeled into a class or category, therefore inherent structures within the input data is deduced or learned without *a priori* ("from the former") knowledge. During supervised learning, all data is labeled into specific class or categories. Relationships between the input data and the class label are learned, typically to determine a prediction of the class label using input data available. The following section will introduce supervised and unsupervised learning strategies applied in this thesis.

Cluster Analysis

Multivariate unsupervised data partitioning (i.e., data clustering) can stratify data objects into several groups (clusters, or phenotypes) by modes of statistical similarity and separation (dissimilarity) without *a priori* knowledge of class separation (i.e., unsupervised learning). Cluster analysis has been recommended for OA phenotype research, to provide novel insights, extending beyond current knowledge (Spil et al., 2020).

The most common types of clustering methods include *k* -means, *k* -medoids, and hierarchical agglomerative (bottom-up) clustering. When clustering, typically the number of clusters, *k*, is defined, and the algorithm works to optimally separate the data into *k* distinct groups. Clustering by *k* -means methodology separates data objects into clusters based on Euclidean distance in space (Han et al., 2011). Starting with the initial dataset in multivariate feature space, *k* centroids are randomly established, representing the centroid of each cluster. Following a stepwise process, the Euclidean distance between each observation and each centroid is calculated. The observation of interest is then assigned to the cluster of the closest centroid. A new centroid of the winning cluster is calculated, the mean of the observations assigned to that cluster. This process is repeated for each data observation until convergence. *K*-medoids is very similar, except the medoid, not the mean, is calculated to define cluster centroid, making it a more robust method against outlier observations (Han et al., 2011). The most commonly used unsupervised clustering technique in the OA literature is hierarchical clustering (Deveza et al., 2017). Initially, each observation is considered a single cluster (a leaf) in a theoretical tree. At each step, the two clusters (leaves) whose merging is associated with the minimal increase in variance via Ward's criteria are combined (two leaves are merged), forming a new cluster (Clatworthy et al., 2005; Ward, 1963). This process is repeated until all observations are members of a single root. Number of clusters from k=1:$n_{observations}$ can be examined from the single tree, based on the number of splits at each node.

Few studies have applied unsupervised methodologies to OA populations. Pinedo-Villanueva et al. used hierarchical cluster analysis to identify and characterize a high pain

cluster after TKA (Pinedo-Villanueva et al., 2018). Behrend et al. used hierarchical cluster analysis to characterize three clusters based on self-reported Forgotten Joint Score outcomes (Behrend et al., 2016). Knoop et al. identified five clusters by *k*-means using features of radiographic severity, muscle strength, BMI, and depression scores (Knoop et al., 2011). Incorporating objective functional measures, Elbaz et al. used *k*-means to report four clusters using spatiotemporal gait parameters (Elbaz et al., 2014). Finally, Waarsing et al. used a model-based clustering approach to identify four clusters by demographics, cartilage features, mechanical alignment, and self-reported symptoms, with clusters strongly characterized by joint-level structural degradation (Waarsing et al., 2015), indicative of variable functional loading (Andriacchi et al., 2009). Overall, clustering literature in OA and TKA remains in its initial stages, with few clustering variables typically addressing one specialty of OA or TKA research. Although valuable information on patient variability is has been found, common and relevant clustering themes between studies and on multidimensional datasets needs to be synthesized before we can translate findings into information that is meaningful clinically.

Self Organizing Maps

Self Organizing Maps (SOMs) are artificial neural networks that project high-dimensional data onto a connected network of nodes. Each node is represented by a weighted vector equation, enabling input data to be mapped to the lower-dimensional SOM space (Kohonen, 1990, 1982). SOMs are similarity graphs and cluster diagrams, where similar features in the input feature space remain spatially proximal in the lower-dimensional mapped space.

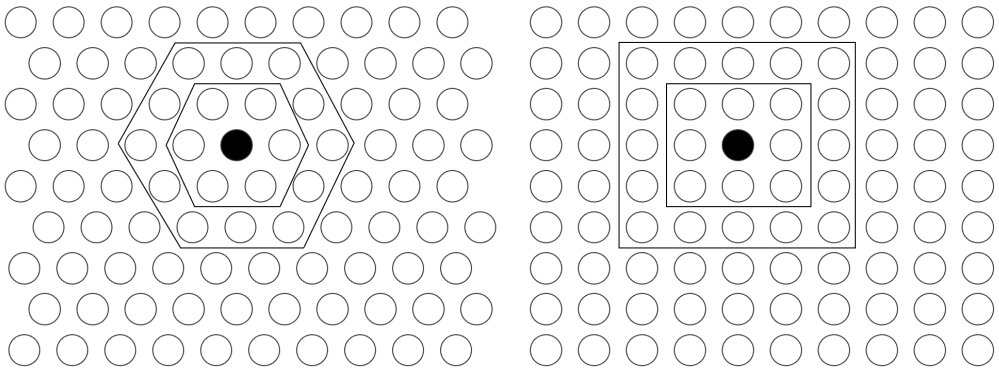

Figure 1.1. Hexagonal (left) and rectangular node (right) SOM neighborhoods.

SOMs typically represent a two-dimensional network of nodes. These nodes connect to adjacent nodes via a neighborhood relationship, generally represented by proximal hexagonal or rectangular configurations (Figure 1.1). Each node can be defined by the equation, $m_i=[v_1,v_2,...,v_n]$, where i denotes the index on the SOM grid, and v denotes vector weights for n clustering features. Nodes are initialized by randomly assigning an observation from the training dataset to each node. The SOM methodology then follows a recursive, stepwise learning process (Kohonen, 2013, 1990, 1982). First, a training observation is projected onto the SOM. A winning node is defined, satisfying the minimum Euclidean distance between the training observation and each node's vector equation (i.e., "competitive learning"). The vector weights, v, of the winning node and its neighboring nodes are then adjusted towards the input training observation. The magnitude of this adjustment is a function of the distance to the winning node, and a specified learning rate, l. The learning rate parameter is decreased linearly to 0 over the learning process. The learning process is repeated, such that each observation is presented to the SOM iteratively until node weights converge, or based on a specified training time, t, the number of times the training set is presented to the SOM.

SOMs have not been applied to OA or TKA populations. However, they have been applied to optical nerve imaging of glaucoma patients, demonstrating an ability to map glaucoma disease progression among across five different disease subtypes (Abidi et al., 2018). SOMs were also paired with PCA in a Parkinson disease dataset to remotely track Parkinson disease progression, including motor feature data items (Nilashi et al., 2020), and identify limping and normal gait patterns using kinematics during treadmill walking (Caldas et al., 2018).

Classification and Regression Trees

Classification and Regression Trees (CART) is an application of top-down recursive partitioning, which has the ability to handle multidimensional features (Breiman et al., 1984). CART learning leads to the creation of a flow-chart decision tree structure

comprised of a series of decision nodes defined by data factors, terminating at a final class definition, or the "leaf". Resultant trees can be translated into interpretable rulesets for each route to a class label, where rules are both mutually exclusive and exhaustive (Han et al., 2011). The CART algorithm determines the hierarchy of the attributes based on how adequately they discriminate the data into a given class, using a measure of the partition's "purity". An ideally "pure" partition would be characterized as having exactly different class labels for each branch of the partition. Mathematically, the impurity of a full dataset, D, is calculated by the Gini index, which represents the probability, p_i, of patients belonged to class i, out of m possible classes (Breiman et al., 1984; Han et al., 2011) . During attribute selection, the impurity of a dataset given a possible partition node is calculated using a weighted sum of the impurity of the entire partition. Therefore, given patient attribute, A, a binary split, the partition would separate the dataset into D_1 and D_2, respectively. In Equation 1, $|D|$ denotes the absolute size (number of elements) of the respective dataset. This is repeated for each possible splitting attribute or splitting combination (if not binary). The attribute that maximizes the reduction in impurity is selected as the splitting criteria for the decision node, Equation 2.

$$Gini_A(D) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \quad [1]$$

$$\Delta Gini(A) = Gini(D) - Gini_A(D) \quad [2]$$

Two recent studies by Lungu et al. demonstrated the application CART approaches in outcome prediction models for both TKA and total hip arthroplasty candidates, classifying patients based on self-reported WOMAC scores and their degree of artificial joint perception after surgery (Lungu et al., 2015, 2014). In the TKA study, prediction rules were developed based on a sample of 141 patients, to identify poor outcomes at six months post-TKA classified using WOMAC scores (Lungu et al., 2014). Using Pre-TKA WOMAC alone, Lungu et al.'s prediction rule reported an acceptable performance, with an area under Receiver Operating Characteristic (ROC) curves (AUC) score of 0.77, and a recall (true positive rate) of 82.1%. A study by Bade et al. applied CART classification to develop decision rules for predictions of functional performance in 119 patients at six months post-TKA using Timed Up and Go, Six-Minute Walk and Stair Climbing Tests

(Bade et al., 2012). Model independent variables (or attributes) included pre-TKA functional score tests and patient demographics. Bade et al. and Lungu et al. both reported pre-TKA functional tests to be the best predictors for post-TKA functional scores.

1.5. Collaboration and Utilizing Available Data

Diverse multidisciplinary research and clinical groups have contributed extensive knowledge of the key drivers of patient variability, and how variability relates to outcomes after interventions such as TKA. Despite the amount of information collected on patients, currently knowledge remains insufficient to drive clinical decision-making. Some of the limitations of analyses to date may be reflect the nature in which phenotypes have been studied; by independent clinical and research groups, or using secondary data specific to their field of research. Current findings may represent research-area-specific phenotypes, without fully capturing the full multifactorial complexities of OA disease (Andriacchi et al., 2014). Using linkages to consolidate the diverse OA data assets immediately available, exploring new analysis methods capable of modelling complex multidimensional data, and incorporating multidisciplinary expertise to analysis and interpretation may uncover greater potential from our existing data assets.

Chapter 2. Thesis Objectives and Format

The theme of this thesis was to investigate if symptomatic and functional variability among patients may be related to self-reported and objective functional outcomes after TKA. Four objectives investigated relationships between self-reported symptoms, knee biomechanical variability and self-reported outcomes after TKA, characterized demographic and knee biomechanical variability (clusters) among OA and TKA populations, and investigated the relationships between clusters and functional outcomes after TKA.

2.1. Objectives and Hypotheses

Objective 1

To characterize changes in patient-reported general satisfaction from six weeks to two years following TKA, and to identify self-reported factors and demographics associated poor general satisfaction longitudinally.

Hypothesis

A longitudinal approach would be able to identify patient factors that contribute to the odds of poor satisfaction after TKA, and identify when factors become meaningful in high-risk of poor satisfaction patient profiling.

Rationale

Despite broad knowledge of factors associated with satisfaction, we still have a poor ability to predict satisfied or dissatisfied patients. If we cannot identify which patients will be dissatisfied prior to arthroplasty (Baker et al., 2013; Judge et al., 2011), there is impetus to identify potential dissatisfied patients as early as possible in the care process post-surgery, as appropriate supports or interventions could be essential in achieving desirable outcomes longitudinally.

Objective 2

i) To compare pre-TKA demographic and gait biomechanic (knee-joint kinematic and kinetic) differences between patients who self-report clinically meaningful improvements in pain and function post-TKA (responders) and those who do not (non-responders).

ii) To compare pre to post-TKA gait biomechanic (knee-joint kinematic and kinetic) changes between responders and non-responders, and assess correlations between post-operative gait improvements and self-reported pain and function improvements after TKA.

Hypothesis

Knee-joint kinematic and kinetic factors pre-operatively and biomechanical gait changes experienced from surgery would be associated with PROMs improvements after TKA.

Rationale

PROMs improvements after arthroplasty have been associated with baseline gait mechanics (Naili et al., 2017; Smith et al., 2004; Turcot et al., 2013). TKA is inherently a mechanical surgery, and gait mechanics worsen with OA progression (Astephen et al., 2008a), radiographic severity (Wilson et al., 2011), and with symptoms of pain (Henriksen et al., 2012; Maly et al., 2008). Objective assessment of gait severity at baseline may aid in identifying functional features most associated with PROM improvements post-TKA, providing important information for pre-operative candidate selection and expectation management. TKA also aims to improve knee function, in part by improving patient gait (Andriacchi, 1993; Hatfield et al., 2011; Noble et al., 2005). It remains unknown if patients who self-report poor outcomes do improve objectively in terms of gait function, and further, what gait function improvements are associated with PROMs improvements (Sosdian et al., 2014). Exploring this could motivate the importance of surgically targeting specific gait function changes during arthroplasty.

Objective 3

i) To identify knee joint gait biomechanics phenotypes (clusters) among TKA candidates based on similarities in patient demographics, frontal and sagittal plane knee kinematics and kinetics during gait.

ii) To compare objective functional gait improvements between phenotypes after standard TKA.

Hypothesis

Distinct biomechanical phenotypes exist within TKA populations, varying by levels of knee function severity. Patient phenotype categorization will be associated with the degree of objective functional improvement experience after TKA.

Rationale

Phenotyping discussions consistently propose the existence of biomechanically-driven OA subtypes (Andriacchi et al., 2014; Bannuru et al., 2019; Castaneda et al., 2013). To date, joint-level biomechanical variability among TKA candidates has primarily been assessed through *a priori* group definition, stratifying OA patients by sex (Paterson et al., 2017; Wilson et al., 2015), frontal plane alignment (Turcot et al., 2012), obesity (Paterson et al., 2017), or patient-reported symptoms before (Thorp et al., 2007; Wilson et al., 2017), and after TKA (Naili et al., 2017; Young-Shand et al., 2020). We still have little evidence into how patient demographic and joint-level features naturally separate. Unsupervised multivariate data partitioning strategies (i.e., data clustering) can characterize phenotypes by modes of statistical separation and similarity, and have been recommended for OA phenotype research (Spil et al., 2020). Few studies have applied unsupervised methodologies to OA populations, and none have investigated joint-level biomechanics, or addressed outcomes of interventions such as TKA at a cluster/group level. With the goal of individualizing treatment plans to patient groups, there is a need to better understand TKA candidate biomechanical variability and understand how biomechanical variability relates to outcomes after TKA.

Objective 4

To quantify OA profiles through the continuum of OA progression using a novel unsupervised machine learning framework, and visually map variability in demographic and knee joint kinematic and kinetics during gait using self-organized maps (SOM).

Hypothesis

Mapped regions were hypothesized to demonstrate knee biomechanics variability associated with clinical OA disease severity, and provide evidence of OA phenotypes traveling a multitude of patient progression pathways which would have relevance for individual treatment strategies.

Rationale

To date, OA and TKA phenotypes have only been derived from cross-sectional datasets, captured at various static time points within the OA progression pathway, defined using variable severity criteria (Deveza et al., 2017; Knoop et al., 2011; Waarsing et al., 2015). Temporal snapshots may lack insight into phenotypes that span the longitudinal OA disease progression process, limiting our ability to understand if phenotype-specific progression pathways exist. As we aim to propose phenotype-specific prevention and intervention strategies, a longitudinal view of the OA progression pathway is required to identify OA phenotypes that impact clinical utility and personalized treatments (Felson, 2010).

2.2. Thesis Format and Data Compilation

This thesis is presented in four stand-alone chapters composed as original and independent manuscripts with a natural progression; each answering questions posed from the prior chapter. The first study, Chapter 3, focused on the patient experience over time. It explored longitudinal changes in patient satisfaction after TKA to identify risk factors for poor satisfaction outcomes. It was written for the Journal of Arthroplasty, and this manuscript is currently undergoing revisions to address reviewer comments.

Building off the patient experience, Chapter 4 brought in a second perspective, joint mechanics during gait. It aimed to investigate relationships between self-reported outcomes and changes in gait biomechanics after TKA, demonstrating that biomechanical variability relates to the patient experience. This work was written for and has been published in the Journal of Bone and Joint Surgery (Young-Shand et al., 2020).

Motivated by knowledge that patient biomechanical variability relates to the patient experience, Chapter 5 aimed to use unsupervised partitioning strategies to investigate and characterize knee joint-level biomechanical phenotypes among patients prior to TKA. This work was the recipient of two awards at the Orthopaedic Research Society Conference in Austin Texas in 2019, the Orthopaedic Research Society Implant Section Poster Award, and the American Academy of Orthopaedic Surgeons Women's Health Advisory Board Poster Award. As such, it was written for and submitted to the Journal of Orthopaedic Research in December of 2020.

Finally, Chapter 6, takes a step back temporally. It aimed to provide indications of multivariable phenotypes incorporating patient demographics and knee joint-level biomechanics occurring longitudinally over the course of the OA disease process (from asymptomatic to post-TKA). It was written for and submitted to the Journal of Biomechanics in January of 2020.

In order to answer our questions spanning different patient features and stage of the OA disease process, this work was conducted on compilated data sets available between research groups at Dalhousie University, and clinical orthopaedic registry and research datasets collected with the Nova Scotia Health Authority. Successful collaboration between our academic biomedical engineering group and orthopaedic surgeons has resulted in numerous academic publications and engineering innovations. The next phase of collaboration was to link our high-resolution biomechanics datasets to comprehensive clinical registries, with the aim of identifying novel linkages between patient factors and health outcomes. This work provides evidence and frameworks for strategic data collection and analysis going forward.

Chapter 3. Early Identification of Satisfaction After TKA

3.1. Introduction

Total Knee Arthroplasty (TKA) is a high volume joint replacement surgery with increasing prevalence (CIHI, 2015b; Kurtz et al., 2011). Procedure rates are rising among younger, and more physically demanding individuals with high functional expectations (Ravi et al., 2012; Scott et al., 2012). Although TKA is widely recognized as an effective procedure, patient reported satisfaction rates remain around 80%, low relative to other orthopaedic procedures; over 2% of patients are at risk of revisions; and readmissions impact over 6% of patients within three months (Bourne et al., 2010; Husted et al., 2008; Noble et al., 2006; Scott et al., 2012). Common reports of poor PROMs in a subset of patients after TKA signals the continued need to improve our understanding of which patients are at risk for adverse outcomes to inform evidence-based models of care and maximize outcomes (Allen et al., 2016).

General satisfaction with TKA can be influenced by a variety of factors, and has been attributed pre-operatively to severity of pain (Bourne et al., 2010; Scott et al., 2010), self-reported function (Scott et al., 2010), and mental health scores (Scott et al., 2010; Vissers et al., 2010). Post-operatively, satisfaction has been attributed to pain relief (Bourne et al., 2010; Mannion et al., 2009; Scott et al., 2010; Vissers et al., 2010), health-related quality of life (Scott et al., 2010; Vissers et al., 2010), and self-reported function scores (Bourne et al., 2010; Scott et al., 2010; Vissers et al., 2010), in addition to considerations of expectations (Bourne et al., 2010; Vissers et al., 2010), procedure complications (Bourne et al., 2010), pain in other joints (Mannion et al., 2009; Scott et al., 2010), and personality traits (Giurea et al., 2016). Despite broad knowledge of features associated with satisfaction, we still have a poor ability to predict satisfied or dissatisfied patients. Specifically, pre-operative patient reported measures lack the predictive ability to identify satisfied patients post-operatively (Baker et al., 2013; Judge et al., 2011), and a recent review has suggested that many associations with satisfaction are controversial or equivocal due to methodological differences in study timelines, cohorts, and outcomes (Gibon et al., 2020). Although the myriad of dimensions that influence satisfaction are

complex, patients who do report poor responses in the years following TKA will continue to experience more pain or dissatisfaction at 10-15 years, without a quantifiable radiographic or clinical premise for their response (Ali et al., 2014; Robertsson et al., 2000). This suggests that some patients are being missed who may have benefitted from early post-operative care strategies, conservative treatment strategies (Skou et al., 2015), or approaches that deviate from the standard of care (Vanlommel et al., 2013; Young-Shand et al., 2020). Most post-operative assessments of satisfaction lack this longitudinal approach by measuring satisfaction at a single time point, typically between six months to two years post-TKA. Only one investigation has addressing earlier post-operative outcomes at three months (Williams et al., 2013a). If we cannot identify dissatisfied patients prior to arthroplasty (Baker et al., 2013; Judge et al., 2011), there is impetus to identify dissatisfied patients as early as possible in the care processes, as appropriate supports or interventions could be essential in achieving desirable outcomes longitudinally.

The objective of this study was to characterize changes in patient-reported general satisfaction from six weeks to two years following TKA, and to identify patient-reported outcomes and demographics that are associated with dissatisfaction. Longitudinal data analysis was used to identify when particular factors become meaningful in high-risk patient profiling, while confirming the relevance of factors up to two years post-TKA.

3.2. Methods

This was a secondary study on a subset of patients recruited for a radiostereometric analysis implant migration study (Laende et al., 2019). Surgeries were performed by five high-volume surgeons at a single site for primary TKA between 2011 and 2014. Patients were asked to voluntarily participate in this study; inclusion criteria was ethical consent from patients. Pre-TKA patient factors and PROMs were collected during preadmission surgical visits, which included demographics (age, sex), body mass index (BMI), and the following questionnaires: 1) the Hospital for Special Surgery Knee Replacement Expectations Survey (KRES, 19[highest]-95[lowest]) (Mancuso et al., 2001), 2) the Pain

Catastrophizing Score (PCS, 0[least]-52[most]) (Sullivan et al., 1995) reflecting anxious preoccupation and a sense of helplessness regarding pain, shown to be an independent predictor of post-TKA chronic pain (Burns et al., 2015), 3) the joint-specific functional Oxford-12 Knee Score (OKS) (0[worst]-48[best]) (Dawson et al., 1998), 4) the visual analog scale (VAS) Pain Score (0[worst pain imaginable]-100[no pain]) (Whitehouse et al., 2003), 5) the UCLA Physical Activity Score (0[lowest]-10[highest]) (Naal et al., 2009), 6) the EuroQoL EQ-VAS general health score (0[worst]-100[best]), and 7) the EQ-5D questionnaire based on five questions regarding difficulties with i) walking, ii) self-care, ii) performing usual activities, iv) experiencing pain/discomfort and v) anxiety/depression (5[best]-15[worst]) (Bansback et al., 2012). Post-TKA outcomes were collected longitudinally through follow-up mail-outs at six weeks, twelve weeks, six months, one and two years post-TKA. Follow-up questionnaires included numbers 3-7 from those listed above. A Satisfaction VAS questionnaire was also asked pre-TKA and at each follow-up. Patients were asked "How satisfied are you with your knee today, in your opinion", and indicated on a scale from 0 (unsatisfied) to 100 (completely satisfied). The satisfaction score was used to define a binary outcome; "satisfied" for scores ≥ 90, and "not fully satisfied" for scores ≤ 89. A binary cut-off score of 90 was selected as it approximated mean satisfaction scores at one year. Further, a study by Noble et al. (Noble et al., 2011) reported satisfaction on a similar population using both a VAS scale and the five-factor satisfaction scale of the New Knee Society Scoring System more commonly used (Bourne et al., 2010; Mannion et al., 2009; Robertsson and Dunbar, 2001; Scott et al., 2010; Turcot et al., 2013; Vissers et al., 2010). Those categorized as "satisfied" using the five factor scale had mean satisfaction scores post-operatively that approximated 90 on the VAS scale, supporting the satisfaction threshold assumption applied for the purposes of examining clinically less satisfied TKA patients. The score of 90 also agreed with an expected ~80% satisfaction rate for TKA candidates at one year.

Patients missing pre-TKA satisfaction scores, and three or more satisfaction responses after TKA (i.e., missing ≥4/6 responses total) were removed from analysis. Statistical analysis was performed using R (2015, R Foundation for Statistical

Computing, Vienna, Austria). Data analysis was performed on a de-identified database, under Research Ethics Board approval.

3.2.1. Satisfaction at One Year

Satisfaction scores at one year post-TKA were used to create stratified groups of "satisfied" and "not fully satisfied" at a common and stable time-point (Bourne et al., 2010; Scott et al., 2010; Williams et al., 2013a). Mann-Whitney U tests, un-paired t-tests and chi-squared tests were used to identify differences between one-year satisfaction stratified groups in PROMs and demographics at each collection time.

3.2.2. Longitudinal Satisfaction

Wilcoxon signed-rank tests were used to identify changes in mean satisfaction scores between each follow-up time point. A generalized linear mixed effects model with a binomial logit link function was applied to examine pre and post-TKA PROMs and demographics associated with longitudinal satisfaction. This model accounted for dependencies caused by repeated measures and within subject variability. Patient-specific random effects were included, as were fixed effects to test for the effect of demographic and questionnaire factors. Influence of factors were presented as odds ratios (OR). For clinical interpretability, Pain VAS, EQ-VAS and OKS were standardized such that ORs represented 10% changes. Models were assessed using randomized quantile residual and Q-Q plots, Shapiro-Wilk normality tests, and the Akaike Information Criterion (AIC). Significant effects were those with a p-value ≤ 0.05.

3.3. Results

Demographic and pre-TKA satisfaction scores were available for 110 primary TKA patients. After correcting for missing data points (dropouts), 86 patients with pre-TKA satisfaction responses were included. Questionnaire responses ($n_{individuals}$=86, $j_{observations}$=483) and missing data elements at each time point are summarized in Table 3.1. Comparing demographic and PROM responses between drop-out and the study

group captured higher VAS pain scores in the drop-out group relative to the study group (p=0.03).

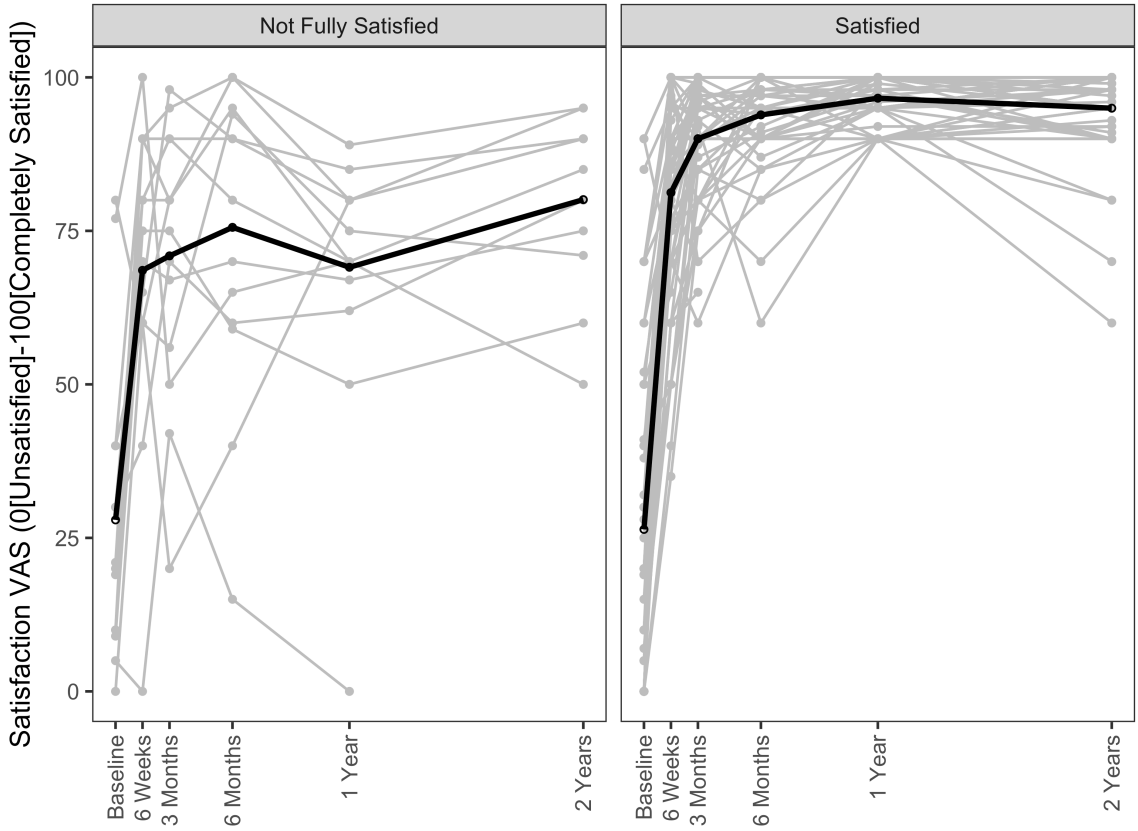

Figure 3.1. Mean patient-reported satisfaction at each follow-up. Satisfied (right) is defined by VAS Satisfaction scores 90 ≥ at one-year post-TKA (n=80 unique individuals).

Satisfaction at One Year

Eighty patients completed satisfaction follow-up scores at one-year post-TKA, and were included in the one-year stratified analysis (n=80, j=455). Mean satisfaction scores at one-year were 91.8±14.4 (Table 3.1). Using a satisfaction threshold of ≥ 90, 82.5% (n=66/80) were categorized as satisfied (Table 3.1), with mean satisfaction scores of 96.6±3.9 among the satisfied group, relative to 69.1±22.6 among those not fully satisfied (Table 3.2). Time series satisfaction scores between these groups is shown in Figure 3.1. Differences between those categorized as satisfied and not fully satisfied at one year were found in the pre-TKA data, with the not fully satisfied group having higher PCS (24.3 vs. 13.0, p=0.02) and lower EQ-5D scores (0.46 vs. 0.63, p<0.001) (Table 3.2). There were

no differences in sex, or BMI between the one-year satisfied and not fully satisfied groups pre-operatively. Additional differences were found at six weeks post-TKA, with the one year not fully satisfied group, reporting lower OKS (25.2 vs. 30.3, p=0.01), worse Pain VAS (57.6 vs. 71.7, p<0.03), and lower EQ-VAS scores (62.3 vs. 77.3, p<0.01).

Table 3.1. Patient factors, mean PROM responses, standard deviations, and missing data pre-TKA and at each follow-up time (n=86 unique individuals; j=483 observations). ‖Denotes time independent features captured pre-operatively.

	Baseline (Pre-TKA)	**6 Weeks**	**3 Months**	**6 Months**	**1 Year**	**2 Years**
Responses	86	85	83	81	80	68
Age (years)‖	63.1 (8.7)					
Sex (F:M)‖	55: 31					
BMI (kg/m^2) ‖	34.6 (7.7)					
Knee Replacement Expectations‖	41.2 (13.6)					
Missing	39					
Pain Catastrophizing Scale‖	15.4 (13.0)					
Missing	18					
Mean Satisfaction VAS	26.3 (25.7)	78.3 (18.3)	86.9 (14.6)	90.9 (14.2)	91.8 (14.4)	92.6 (10.5)
% Satisfied	3.5%	37.6%	62.7%	81.5%	82.5%	85.3%
Missing	0	1	3	5	6	18
Oxford Knee Score	21.3 (6.5)	29.3 (7.3)	36.4 (7.3)	38.8 (6.5)	39.6 (6.9)	39.9 (6.6)
Missing	0	0	0	0	0	0
Pain VAS	47.6 (22.5)	67.6 (22.9)	84.1 (17.4)	86.6 (21.3)	88.2 (19.6)	90.2 (16.0)
Missing	0	1	2	0	0	0
EQ-5D	0.60 (0.16)	0.73 (0.12)	0.79 (0.14)	0.82 (0.15)	0.84 (0.15)	0.83 (0.15)
Missing	0	0	0	0	1	1
EQ-VAS	67.2 (17.5)	74.4 (16.5)	79.9 (15.9)	83.4 (14.6)	83.7 (15.2)	82.4 (15.3)
Missing	0	1	0	0	0	1
UCLA Activity Score	4.7 (2.0)	4.6 (1.3)	5.2 (1.4)	5.7 (1.5)	5.5 (1.8)	5.5 (1.5)
Missing	0	3	0	0	0	1

Longitudinal Satisfaction

Mean satisfaction scores increased over time, with significant differences between pre-TKA and six weeks (p=0.001), six weeks and three months (p=0.001), and three

months and six months (p=0.01). No statistical differences in satisfaction scores were captured between six-months up to the two-year follow-up point (p>0.4).

Longitudinal analysis conducted on the entire dataset (n=86, j=483) found the strongest contributors for improved odds of satisfaction in an individual to be higher OKS (OR = 2.08, p<0.001), less pain (OR=1.69, p<0.001), and higher EQ-VAS scores (OR =1.34, p=0.03), where a one-point change in the OR of the β coefficient represented a 10% increase in each score, Table 3.3. Although total EQ-5D Total scores were not significant in the model (p=0.07), breaking down EQ-5D responses by each of the five questions, question one was significant, which described greater odds of a satisfaction response in patients with less difficulty walking (one-unit decrease in EQ-5D question one: OR=2.33, p<0.05). Question two neared significance (p=0.07), which captured dimensions of self-care, included to optimize model AIC. All features remained significant after EQ-5D question two's removal. When testing this model using a continuous VAS satisfaction outcome variable, only OKS and VAS Pain factors contributed to an improved satisfaction response (p<0.001). Patient age, sex, and BMI were not significant in the longitudinal analysis.

Table 3.2. Mean patient scores and standard deviations, based on one year Fully Satisfied and Not Fully Satisfied group stratification. Note: Six participants used in the longitudinal analysis did not have a satisfaction result at 1-year, and were therefore not included in the stratified group analysis (n=80 unique individuals; j=455 observations), ‖Denotes time independent features captured pre-operatively; significant ($p<0.05$) features in bold.

	Baseline (Pre-TKA)			6 Weeks			3 Months			6 Months			1 Year			2 Years		
	1-Year Satisfied	1-Year Not Fully Satisfied	*p*	1-Year Satisfied	1-Year Not Fully Satisfied	*p*	1-Year Satisfied	1-Year Not Fully Satisfied	*p*	1-Year Satisfied	1-Year Not Fully Satisfied	*p*	1-Year Satisfied	1-Year Not Fully Satisfied	*p*	1-Year Satisfied	1-Year Not Fully Satisfied	*p*
Age (years) ‖	62.6 (8.6)	65.4 (9.6)	0.3															
Sex (F:M)‖	41 : 25	10 : 4	0.5															
BMI (kg/m^2)‖	34.0 (7.3)	36.0 (8.9)	0.4															
Knee Replacement Expectations‖	41.0 (13.3)	43.1 (16.7)	0.9															
Pain Catastrophizing Scale‖	13.0 (11.3)	24.3 (15.8)	**0.02**															
Mean Satisfaction VAS	26.4 (26.4)	27.9 (24.7)	0.6	81.2 (16.4)	68.6 (25.5)	0.05	90.0 (9.9)	70.9 (22.2)	**0.001**	93.9 (7.6)	75.6 (25.6)	**0.01**	96.6 (3.9)	69.1 (22.6)	**<0.001**	95.0 (7.8)	80.1 (14.8)	**<0.001**
Oxford Knee Score	21.8 (6.3)	18.6 (7.5)	0.2	30.3 (7.4)	25.2 (6.3)	**0.01**	37.8 (6.0)	30.1 (10.2)	**<0.01**	40.6 (4.5)	30.9 (8.2)	**<0.001**	41.3 (5.3)	31.1 (7.6)	**<0.001**	41.8 (4.8)	31.4 (8.4)	**<0.001**
Pain VAS	49.2 (23.6)	38.1 (19.1)	0.1	71.7 (22.0)	57.6 (22.3)	**0.03**	88.4 (11.8)	63.1 (23.5)	**<0.001**	90.1 (18.7)	67.6 (25.4)	**<0.001**	94.9 (9.2)	56.9 (24.8)	**<0.001**	92.2 (14.9)	78.6 (19.3)	**0.001**
EQ-5D	0.63 (0.1)	0.46 (0.2)	**<0.001**	0.74 (0.1)	0.67 (0.1)	**<0.01**	0.82 (0.1)	0.67 (0.2)	**<0.01**	0.85 (0.1)	0.71 (0.2)	**<0.01**	0.88 (0.1)	0.69 (0.1)	**<0.001**	0.87 (0.1)	0.70 (0.1)	**<0.001**
EQ-VAS	68.9 (16.5)	64.1 (19.8)	0.4	77.3 (15.3)	62.3 (16.8)	**<0.01**	83.7 (12.2)	64.7 (18.2)	**<0.001**	85.4 (13.4)	74.6 (17.1)	**0.01**	87.7 (11.7)	65.0 (15.9)	**<0.001**	84.0 (15.2)	74.6 (13.9)	**0.02**
UCLA Activity Score	4.9 (1.9)	4.5 (2.2)	0.4	4.7 (1.2)	4.3 (1.4)	0.4	5.2 (1.3)	5.1 (1.8)	0.5	5.9 (1.4)	4.9 (1.5)	**0.03**	5.6 (1.7)	4.6 (2.1)	**0.04**	5.7 (1.4)	4.8 (1.4)	0.07

Table 3.3. Factors associated with patient satisfaction (score ≥ 90) as determined using a generalized linear mixed effects model. Pain VAS, EQ-VAS and OKS were standardized such that ORs represent 10% changes (n=86 unique individuals; j=483 observations); significant ($p<0.05$) features in bold.

	β	95% CI β	Odds ratio (β)	Std. Error (β)	Z value	*p*
(Intercept)	-8.68	(-12.44, -4.92)	0.00	1.92	-4.52	**<0.001**
Follow-up Time	0.02	(0.00, 0.03)	1.02	0.01	2.45	**0.01**
Pain VAS	0.52	(0.32, 0.72)	1.69	0.10	5.08	**<0.001**
EQ-5D Q1 (Mobility)	-0.84	(-1.68, 0.00)	0.43	0.43	-1.97	**<0.05**
EQ-5D Q2 (Self-Care)	-1.48	(-3.10, 0.15)	0.23	0.83	-1.78	0.07
EQ-VAS	0.29	(0.03, 0.55)	1.34	0.13	2.21	**0.03**
Oxford Knee Score	0.73	(0.38, 1.08)	2.08	0.18	4.10	**<0.001**

3.4. Discussion

Results revealed distinct temporal responses between satisfied and not fully satisfied groups that have not been previously captured (Figure 3.1). Patient satisfaction scores significantly improved within the first six months post-TKA ($p \leq 0.01$), and stabilized thereafter ($p>0.4$). Although there may be clinically meaningful information in satisfaction metrics closely following surgery, final satisfaction perceptions do not stabilize until 6 months post-operatively, likely coinciding with the period of post-operative healing and improved physical function, with muscle recovery plateauing at approximately six months post-TKA (Stevens-Lapsley et al., 2010).

At one year post-TKA, the incidence of a satisfied response (82.5%) aligned with satisfaction rates commonly reported in the literature (Bourne et al., 2010; Mannion et al., 2009; Robertsson and Dunbar, 2001; Scott et al., 2010; Turcot et al., 2013; Vissers et al., 2010; Williams et al., 2013a), yet the three month response was lower than satisfaction scores reported by Turcot et al. (62.7% in Table 3.1 vs. 79.5%) (Turcot et al., 2012). The Turcot patient cohort was slightly older with a lower mean BMI and less pre and post-TKA pain severity than our study, which may account for some of this difference. Pre-

operatively, satisfied patients in our study also self-reported less pain catastrophizing (PCS, $p=0.02$) relative to the not fully satisfied group by one-year post-TKA stratification. Although pre-operative pain catastrophizing has previously been association with poor TKA outcomes in terms of pain (Baert et al., 2016; Burns et al., 2015) and quality of life (Yakobov et al., 2018), during longitudinal modeling, pain catastrophizing was not significant in influencing satisfaction outcomes (Table 3.3). Pain catastrophizing metrics were only available pre-operatively (a time independent-variable), which included missing responses (18/86). Still, this feature remained insignificant after conducting a sensitivity analysis using a data subset with complete pain catastrophizing cases. Similarly, pre-operative general health scores were better among satisfied patients pre-operatively (EQ-5D, $p<0.001$), yet this score was also not relevant in longitudinal satisfaction modeling. It has been suggested that pre-operative data alone may not have the predictive capacity to identify satisfied patients post-operatively (Baker et al., 2013; Judge et al., 2011). Indeed, neither PCS nor EQ-5D total scores, which differed between satisfaction groups pre-operatively, were stable in separating satisfied patients using longitudinal models. Although general health and satisfaction are not clearly correlated (Fenton, 2012), other authors have found post-operative mental health scores to be associated with satisfaction post-TKA (Giurea et al., 2016; Scott et al., 2010; Vissers et al., 2010), and further work is needed to understand the relative importance of pain catastrophizing captured post-operatively on longitudinal satisfaction. Our findings support the notion that pre-operative scores should not be used to prioritize access to care, and if we wish to attempt to do so, examination of the next available time point would be more effective.

As early as six weeks post-TKA, differences between satisfaction groups were identified in pain severity (VAS), joint-specific health (OKS), and general health (EQ-VAS), agreeing with prior findings at three (Williams et al., 2013a), and six months (Scott et al., 2010), but also providing novel insight into divergence in these outcomes earlier than previously reported. Early patient perceptions of pain intensity, general health, and joint specific health at six weeks post-operatively has a lasting influence on satisfaction outcomes. Therefore, the potential influence of the underlying factors

captured in these PROMs begins very early in the post-operative period, but has a longitudinal influence on outcomes up to two years post-TKA. Early post-operative education and intervention may warrant further study to aid in changing this long-term outcome trajectory.

Pain relief is a main expectation of patients undergoing TKA (Scott et al., 2012). In a comprehensive analysis of patient satisfaction and PROMs by Robertsson and Dunbar, pain-related domains were found to be the largest contributor to satisfaction followed by physical domains (Dunbar et al., 2013; Robertsson and Dunbar, 2001), supported in later studies (Baker et al., 2007; Bourne et al., 2010; Scott et al., 2010). We also found changes in pain to have a large effect on satisfaction (Pain VAS OR 1.69±0.10), however, the longitudinal effect of the function-based OKS was greater (OKS OR 2.08±0.18). Similarly, examination of patient responses longitudinally suggested that physical domains and general health contributed most to early satisfaction stratification, demonstrated by clinically significant score changes between baseline and six weeks. For example, clinically significant improvements in the OKS (Beard et al., 2015) were apparent only in the satisfied group at six weeks, yet clinically significant improvements in VAS pain (Katz et al., 2015; Tubach et al., 2005) occurred in both the satisfied and not fully satisfied groups at six weeks (Table 3.2). These findings suggest that less satisfied patients may be more perceptive to a lack of improvement in overall function early in the recovery process, having utility for early patient profiling. Further, the longitudinal satisfaction model specifically included the EQ-5D question 1 (perceived difficulty walking) over other dimensions of general health, which is a unique finding. The TKA patients in this study had a mean age of 63, and were therefore younger than the typically reported age of 70 in previous studies (Bourne et al., 2010; Robertsson and Dunbar, 2001; Scott et al., 2010), potentially resulting in heightened functional awareness relative to populations previously studied. Still, the importance of functional measures on satisfaction agrees with prior findings using diverse tools including the OKS (Scott et al., 2010), Western Ontario and McMaster Universities Arthritis Index (Bourne et al., 2010; Vissers et al., 2010), other functional scores (Noble et al., 2006; Scott et al., 2010), and from gait measures (Turcot et al., 2013). Recent work has also demonstrated a link

between objective joint-level gait mechanics and self-reported outcomes after TKA (Naili et al., 2017; Turcot et al., 2013). As TKA demographics become younger, with greater functional expectations (Ravi et al., 2012; Scott et al., 2012), these findings collectively demonstrate the importance of objective functional considerations to the patient experience longitudinally if we want to achieve optimal outcomes for all individuals (Wilson et al., 2019).

Although visually distinctive temporal differences were found between the satisfied and not fully satisfied group response profiles (Figure 3.1), significant differences between satisfaction scores were not apparent until 12 weeks, lagging other PROMs (Table 3.2). This result may be reflective of an acclimatization after the expiry of support sessions, such as physiotherapy, and return to everyday activities (e.g., work), as previous studies have identified associations between less social supports and low self-reported quality of life, and between living alone and dissatisfaction post-TKA (Bourne et al., 2010; Desmeules et al., 2009). This delayed response might also reflect patient realization that the procedure may not meet expectations, or a surgical team's ability to manage expectations closely following surgery, but be less effective with long-term management. There is variability in regards to the role of expectations on satisfaction (Culliton et al., 2012), where met expectations have been characterized as a leading factor when assessed post-operatively (Bourne et al., 2010; Noble et al., 2006; Scott et al., 2010, 2012), but a poor predictor pre-operatively (Culliton et al., 2012; Mannion et al., 2009; Vissers et al., 2010). This study also found expectations to be a poor pre-operative predictor for satisfaction. Although it should be noted that our pre-operative expectation data included missing responses (39/86), this feature also remained insignificant after conducting a sensitivity analysis using a data subset with complete expectation scores. Future studies could include considerations of post-TKA expectations, surgical team interaction, post-operative programs and program duration to characterize possible contributors to response patterns. This would be needed to inform the development of targeted extended treatment, expectation management or shared decision-making programs (Allen et al., 2016; Bozic et al., 2013).

Achieving patient satisfaction post-TKA is complex and should not be a definitive goal for improving patient care, as satisfaction may be influenced by contradictory or non-modifiable factors (Gibon et al., 2020). Therefore, the anchoring of this study on satisfaction and self-reported tools, which can be heavily biased by external factors, was a limitation. As with most satisfaction studies, our "satisfaction" threshold of 90 overlooks any distribution characteristics of the outcome variable, yet as discussed, this score agreed with an expected ~80% satisfaction rate for TKA candidates at one year, and our stratified groups were representative of prior literature. Other limitations include missing data. We restricted our study sample to include patients with three or more satisfaction responses after TKA, yet also had high missing data specifically in the mentioned domains of pain catastrophizing (missing 18/86) and expectations (missing 39/86) which were only measured pre-operatively. We also found greater baseline pain in the excluded populations relative to our study population ($p=0.03$). Drop-out populations have been shown to be less satisfied (Robertsson and Dunbar, 2001), and results may be biased in terms of overoptimistic outcome scores. We also observed large response variance in self-reported scores, particularly among not fully satisfied individuals (Table 3.2) and visible in Figure 3.1. For example, an individual with a satisfaction score of 0 could be observed at 1-year post-TKA. This patient went on to undergo revision surgery. Sensitivity analysis removing this individual from our model did not alter our results. Longitudinal analysis is robust in the handling of missing data (Gibbons et al., 2010), and a sample of 86 individuals with 483 observations lends to the stability our findings with valuable temporal insights into individual patient outcome trajectories.

This study provided novel insights into longitudinal changes in satisfaction follow TKA. Although it may not be feasible to identify patients at high-risk of long-term dissatisfaction pre-operatively (Baker et al., 2013; Judge et al., 2011), our findings support the potential to identify them as early as six weeks after surgery, suggesting the trajectory for recovery is set in the early post-operative period. Records of satisfaction scores over two-years of follow-up presents an opportunity to identify high-risk patients early in the post-operative care process, and inform early targeted supports or interventions. Our findings also highlight the importance of function in the patient

experience. We recognize that global satisfaction cannot be the only outcome target for surgical and management refinement, and can be influenced by many and varied factors (Gibon et al., 2020). However, understanding the time course of satisfaction along with associated functional and quality of life metrics captured provides novel insight into the time course of the post-operative patient experience, making these findings relevant to providing appropriate supports for patients on poor trajectories, and achieving desirable outcomes longitudinally.

Chapter 4. Individual Gait Features are Associated with Clinical Improvements After TKA

4.1. Introduction

Upwards of 20% of patients with knee osteoarthritis (OA) do not report clinically meaningful improvements in pain and function or satisfaction after TKA (Escobar et al., 2007; Hawker et al., 2013; Quintana et al., 2006; Robertsson et al., 2000), raising concerns over the potential overuse of TKA (Clavel et al., 2016). Appropriate patient selection would thus require an understanding of the symptom state most associated with meaningful improvements after arthroplasty, previously termed the "sweet spot" (Hawker et al., 2013; Losina and Katz, 2013). While patients with worse self-reported pain and function preoperatively experience greater PROM improvements from TKA (Hawker et al., 2013; Jiang et al., 2017; Judge et al., 2012) common PROM tools lack the predictive ability to identify optimal candidates pre-operatively (Hawker et al., 2013; Judge et al., 2012, 2011). Used in isolation, PROMs also provide limited insights into potential underlying biomechanical mechanisms associated with why patients fare well or poorly.

PROMs improvements after arthroplasty have been associated with baseline gait mechanics (Naili et al., 2017; Smith et al., 2004; Turcot et al., 2013). TKA is inherently a mechanical surgery, and gait mechanics worsen with OA progression (Astephen et al., 2008a), radiographic severity (Wilson et al., 2011), and with symptoms of pain (Henriksen et al., 2012; Maly et al., 2008). Objective assessment of gait severity at baseline may aid in identifying functional features most associated with PROM improvements post-TKA, providing important information for pre-operative candidate selection and expectation management. TKA also aims to improve knee function, in part by improving patient gait (Andriacchi, 1993; Hatfield et al., 2011; Noble et al., 2005). It remains unknown if patients who self-report poor outcomes do improve objectively in terms of gait function, and further, what gait function improvements are associated with PROMs improvements (Sosdian et al., 2014). Exploring this could motivate investigations that assess the efficacy of surgically targeting specific functional deficits.

This was an explorative study that aimed to compare pre-TKA demographic and knee-joint gait mechanic (kinematic and kinetic) differences between patients who self-report clinically meaningful improvements in pain and function post-TKA (responders) and those who do not (non-responders), and to model preoperative demographics and gait features descriptive of responders. The secondary aim was to examine pre to post-TKA knee-joint gait mechanic changes among pain and function responders and non-responders, and to examine correlations between gait changes and self-reported improvements.

4.2. Methods

4.2.1. Patients and Surgery

Patients with end-stage knee OA scheduled to receive a primary TKA at a high-volume academic orthopedic clinic (2003-2016) underwent gait assessment one-week prior to (n=135) and one-year post-TKA (n=109) (Figure 4.1). Patients were included if they were able to walk six meters unassisted, and were excluded if they screened positive for neurological disease, or other conditions affecting their gait or ability to safely participate. Surgeries followed a standard medial parapatellar approach, with distal femur cuts set to 5° of valgus, and tibial cuts targeting neutral mechanical alignment. The measured resection technique was used to obtain a balanced flexion-extension gap. Standard post-operative inpatient physiotherapy was used, with immediate weightbearing. Median hospital stay was three days. Outpatient physiotherapy was not standardized and was optional. Informed participant consent was obtained according to the institution ethics board.

4.2.2. Gait Biomechanics

Data on age, sex, mass, height, and OA severity graded by an orthopaedic surgeon using Kellgren-Lawrence (KL) global radiographic scores (Kellgren and Lawrence, 1957) were collected as part of the pre-operative assessment. Infrared light-emitting markers were placed on participants according to a standardized protocol, which included four triads of markers attached to the pelvis, thigh, shank, and foot to establish limb

segment rigid body coordinate systems (Landry et al., 2007). To define local anatomical joint axes, four individual light-emitting markers were placed on boney landmarks, and eight digitized points (Landry et al., 2007). Participants walked along a five-meter walkway wearing comfortable shoes at a self-selected speed. Three-dimensional external ground reaction forces were recorded at 2000Hz with an AMTI Biomechanics Platform System (Advanced Medical Technology Inc., Watertown, MA) embedded in the walkway, synchronized to an Optotrak™ (Northern Digital Inc., Waterloo, ON) optoelectronic motion capture system sampling marker positions at 100Hz. Knee joint angles were calculated according to the joint coordinate system (Grood and Suntay, 1983), and net resultant knee joint moments by inverse dynamics (Deluzio et al., 1993; Deluzio and Astephen, 2007; Li et al., 1993), amplitude normalized to body mass (Nm/kg). Following this protocol (Landry et al., 2007), a minimum of four walking trials were averaged and normalized for each participant to one gait cycle (0-100%) for flexion/extension angles, and to stance-phase (0-100%) for moments and adduction angles.

Principal Component Analysis (PCA) was used to capture major features of participant variability in knee angle and moments waveforms, having demonstrated improved sensitivity over discrete gait parameters (Hatfield et al., 2015b; Smith et al., 2004; Wilson et al., 2011). A large sample of pre-TKA (n=135) and one year (n=109) post-TKA patient waveforms were used to create robust PC models using a standardized protocol (Deluzio and Astephen, 2007). Three knee adduction moment, adduction angle, and flexion moment Principal Components (PCs), and four knee flexion/extension angle PCs were retained (Appendix A). These features have been previously shown to describe the major modes of variability in TKA (Hatfield et al., 2011) and osteoarthritis gait (Astephen et al., 2008b), or were features typically applied to functional assessment post-TKA (Bonnefoy-Mazure et al., 2017; McClelland et al., 2011; Smith et al., 2004). Individual patient data was projected onto each PC, providing individual subject PC scores used in hypothesis testing.

4.2.3. PROMs

WOMAC (Bellamy et al., 1988) PROM questionnaires were collected on a portion of gait participants one-week pre (n=59) and one-year post-TKA (n=46), meeting international PROMs collection standards (ICHOM, 2017). Patients with matched pre and post-TKA WOMAC scores (n=46) were included in analysis (Figure 4.1). Pre to post-TKA WOMAC (0[worst]-100[best]) pain and function score changes were used to categorize patients as responders in pain (≥23) and function (≥19) independently, using established WOMAC minimal clinically important difference (MCID) criteria (Escobar et al., 2007; Roos, 2018). Non-responder follow-up scores were assessed for ceiling effects (post-operative score of 100), ensuring WOMAC boundaries did not contribute to non-responder classification.

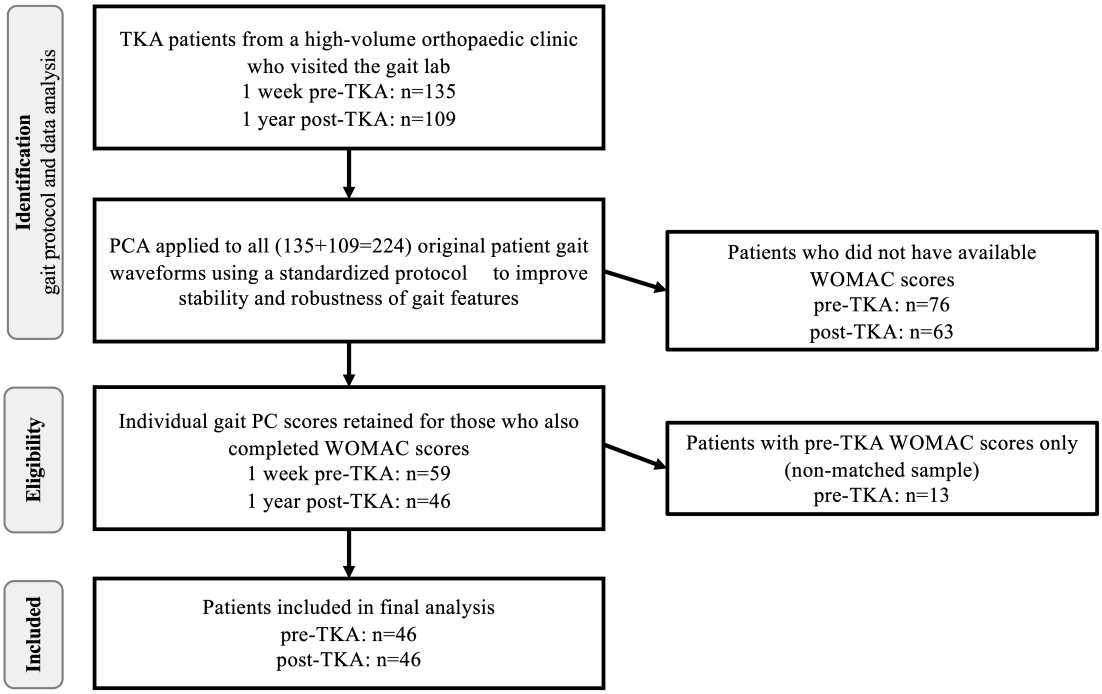

Figure 4.1. CONSORT (Consolidated Standards of Reporting Trials) diagram of patient eligibility and selection processes. All participants were screened for neurological, previous lower extremity surgeries (e.g., arthroplasty in another joint), or other existing pathologies (e.g., rheumatoid arthritis) prior to gait recruitment. No baseline WOMAC differences were detected between subjects who did (n=46) and did not (n=13) have completed 304 post-TKA scores in any WOMAC domain ($p>0.5$).

4.2.4. Statistical Analysis

Baseline Analysis (primary aim): Baseline variables (age, BMI, KL grade, WOMAC scores, gait speed) and PC score differences between pain and function responders and non-responders were compared using un-paired t-tests and Mann-Whitney U tests. Correlations between baseline demographics and gait PC scores, against WOMAC pain and function changes (post−pre) were examined using Pearson's correlation coefficients. Variables showing significant correlations with WOMAC pain and function were retained for multiple regression analyses. Binomial generalized linear models examined baseline demographics and baseline gait features associated pain and function responder classification independently, assessed using Akaike's information criterion. Final models were presented using modified Poisson regression (Zou, 2004) for improved clinical interpretation (Knol et al., 2012), representing coefficients as relative risk ratios (RRs) and 95% confidence intervals (CI) derived from standard errors using the robust sandwich estimator. Features were scaled (0-10), where a one-point increase in RR was associated with a 10% change in PC score. All analyses were conducted in an exploratory fashion with p-values <0.05 considered significant in interpretation.

Pre to Post-TKA Changes (secondary aim): Differences between pre and post-TKA gait features within pain and function responder and non-responder groups were compared using paired t-tests. Correlations between changes in PC scores (post−pre), against changes in WOMAC pain and function were examined using Pearson's correlation coefficients.

4.3. Results

4.3.1. Baseline Analysis

Pain: Seventy-four percent (34/46) of patients met pain domain MCID improvement criteria and were classified as responders in WOMAC pain; 26% (12/46) were classified as pain non-responders. Pre-operatively, pain responders were more radiographically severe (p=0.03), more symptomatic (WOMAC Total, 54.7 95%CI 38.2-75.0 vs. 45.5 95%CI 19.1-71.0, p=0.04), and walked at faster gait speeds (0.93±0.19m/s vs.

0.80±0.18m/s, p=0.04) than non-responders (Table 4.1). Pain responders also walked with lower adduction angle PC1 scores relative to pain non-responders pre-operatively (p=0.03), indicating lower overall knee adduction angle magnitudes (less consistently varus) throughout the stance phase of gait (Table 4.2, Figure 4.2).

Patients who had less stance-phase flexion-extension angle range (PC4, r=-0.32, p=0.03), and lower stance-phase varus magnitudes (PC1, r=-0.37, p=0.01) pre-operatively, were associated with more WOMAC pain improvement (Figure 4.3a-b). In multivariate modeling, lower stance-phase varus (adduction angle) magnitude was the only pre-operative feature predictive of being a pain responder (PC1, RR=0.92, p<0.05, Table 4.3).

Function: Seventy-eight percent (36/46) of patients met function domain MCID improvement criteria and were classified as responders in WOMAC function; 22% (10/46) were classified as function non-responders. Pre-operatively, function responders were more radiographically severe (p=0.03) than function non-responders (Table 4.1). Function responders also had lower stance-phase varus (adduction angle) magnitudes (PC1, p<0.05) and less stance-phase flexion-extension angle range than non-responders pre-operatively (PC4, p=0.01) (Table 4.4, Figure 4.4).

Patients who were younger (r=-0.41, p=0.005), had less stance-phase flexion-extension angle range (PC4, r=-0.38, p=0.009) and lower varus magnitudes (PC1, r=-0.34, p=0.01) pre-operatively, were associated with more WOMAC function improvement (Figure 4.3d-f). In multivariable modeling, function responder likelihood increased only if patients walked with less stance-phase flexion-extension angle range pre-operatively (PC4, RR=0.90, p=0.01, Table 4.3).

4.3.2. Pre to Post-TKA Changes

Pain and function responders demonstrated typically reported pre to post-TKA gait improvements (toward asymptomatic) in magnitude and pattern of adduction moment, flexion moment, and flexion angle features (Tables 4.2,4.4) (Hatfield et al., 2011; Wilson et al., 2015). The only gait changes captured among pain and function non-responders

included reductions in stance-phase varus magnitudes post-TKA relative to pre-operatively (PC1 p≤0.005; Tables 4.2,4.4; Figures 4.2,4.4). Pain non-responders alone also showed more stance-phase flexion moment range post-TKA relative to pre-operatively (PC2, p=0.03, Table 4.2).

Patients who experienced less pre to post-TKA reduction in varus magnitudes from surgery (ΔPC1, r=0.47, p=0.001), were associated with more WOMAC pain improvement (Figure 4.3c).

Patients who experienced less reduction in stance-phase varus magnitude from surgery (ΔPC1, r=0.38, p=0.009) and showed larger increases in early to mid-stance adduction moment range (ΔPC2, r=0.32, p=0.03), were associated with more WOMAC function improvement (Figure 4.3g-h).

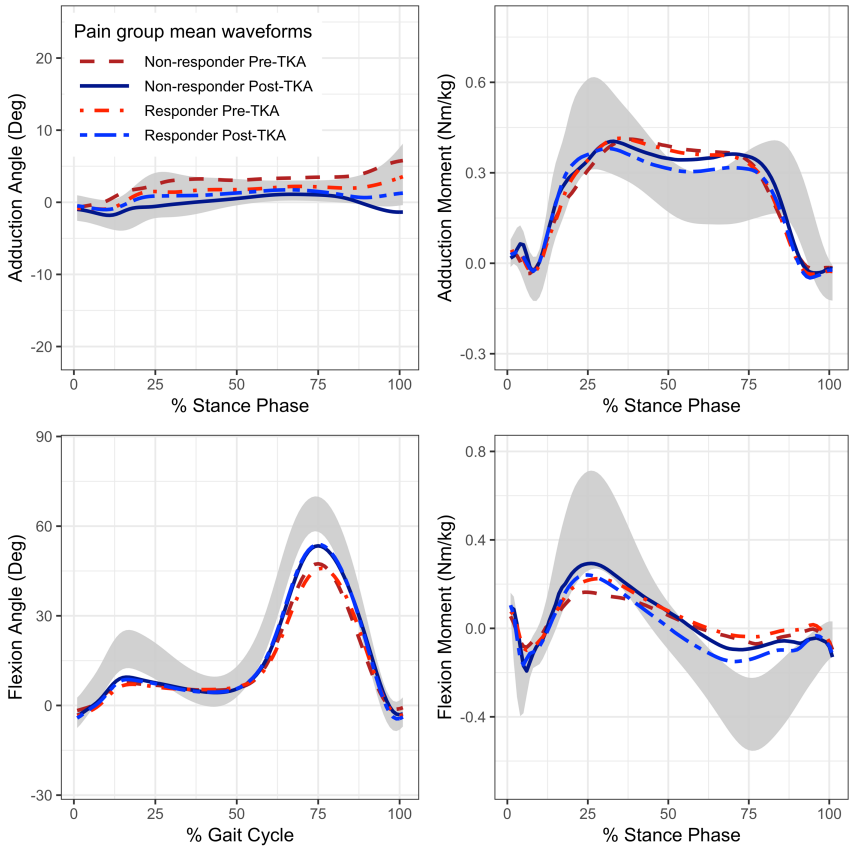

Figure 4.2. Mean waveforms of pain responder (n=34) and non-responder (n=12) groups collected one-week pre-TKA and one-year post-TKA. Grey denotes 209 previously captured asymptomatic (mean age 51.6 ±9.0, BMI 26.7±4.5) gait patterns (mean ± 1SD).

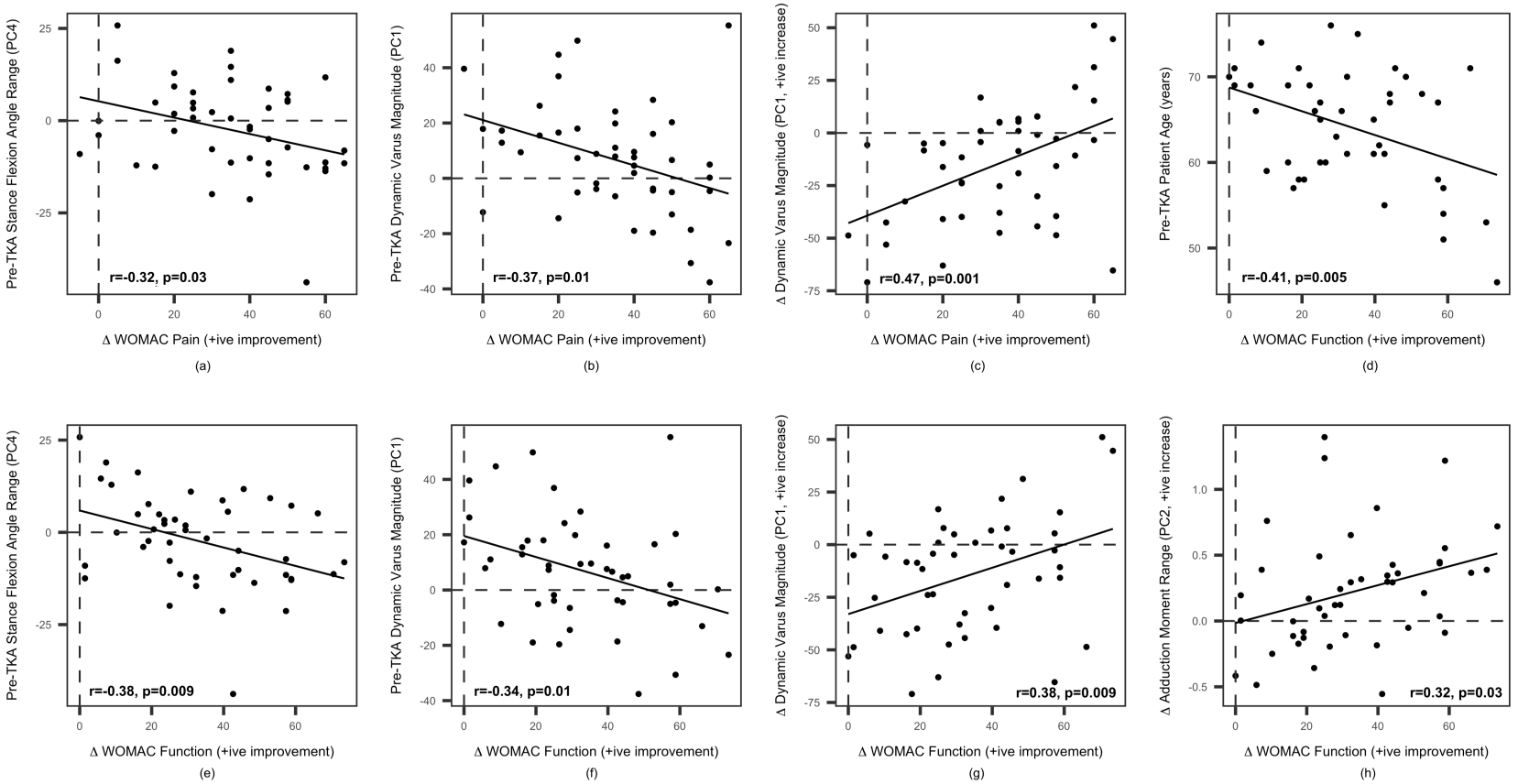

Figure 4.3. Associations between demographic and gait features, and pre to post-TKA changes in WOMAC pain and function. (c & g) Positive changes in stance-phase varus magnitude (adduction angle PC1) represented an increase in varus alignment during stance, while negative changes represented more varus magnitude reduction (varus to valgus change). Less stance-phase varus magnitudes at baseline, and less pre to post-TKA reductions in stance-phase varus magnitudes post-TKA were each independently associated with more self-reported improvements in pain and function. (h) Positive changes in adduction moment range (adduction moment PC2) represented more medial compartment loading/unloading range during stance. Larger increases in dynamic loading range were associated with more improvement in self-reported function

Table 4.1. Baseline demographic and self-reported WOMAC scores of pain and function responder and non-responder groups. Normally distributed variables represented as means and standard deviations, non-normally distributed variables represented as medians and 95% confidence intervals. Significant ($p<0.05$) features in bold.

	Total	WOMAC Pain Domain			WOMAC Physical Function Domain		
		Responder	Non-Responder	*p*	Responder	Non-responder	*p*
Subjects	46	34	12		36	10	
Males	17	12	5		13	4	
Females	29	22	7		23	6	
Age (years)	64.1(6.6)	63.6 (7.0)	65.7(5.4)	0.6	63.5(6.8)	66.4(5.7)	0.2
BMI (kg/m^2)	32.6(5.7)	32.7 (6.2)	32.5(4.0)	0.9	32.4(6.1)	33.3(4.1)	0.7
Kellgren-Lawrence (KL) Grade‖	4.0(3, 4)	4.0 (3, 4)	3.0(3, 3)	**0.03**	4.0(3, 4)	3.0(3, 3)	**0.03**
KL Score 0	0	0	0		0	0	
KL Score 1	0	0	0		0	0	
KL Score 2	0	0	0		0	0	
KL Score 3	13	9	4		9	4	
KL Score 4	14	14	0		14	0	
Pre-TKA Gait Speed (m/s)	0.9(0.2)	0.9 (0.2)	0.8(0.2)	**0.04**	0.9(0.2)	0.9(0.2)	0.5
Post-TKA Gait Speed (m/s)	1.0(0.2)	1.1 (0.2)	1.0(0.2)	0.2	1.1(0.2)	1.0(0.2)	0.2
WOMAC Total (/100)	47.9(21.3, 75.6)	45.5 (19.1, 71.0)	54.7(38.2, 75.0)	**0.04**	46.1(19.3, 70.7)	56.9(38.3, 75.1)	0.1
Pain (/100)	50.0(26.3, 75 .0)	45.0 (24.1, 70.9)	62.5(37.8, 78.6)	**0.007**	47.5(24.3, 75.0)	60.0(37.2, 77.8)	0.07
Joint Stiffness (/100)	50.0(12.5, 75.0)	50.0 (10.3, 75.0)	50.0(19.4, 75.0)	0.08	50.0(10.9, 75.0)	43.8(15.3, 75.0)	0.6
Physical Function (/100)	47.1(25.6, 80.1)	44.1 (22.9, 82.4)	47.8(35.2, 73.1)	0.2	46.9(23.6, 74.5)	58.1(34.7, 86.0)	0.07

‖Grades available for 27/46 participants, reasonably distributed between groups: Available for 23/34 pain responders, 4/12 pain non-responders, 23/36 function responders, and 4/10 function non-responders.

Table 4.2. Principal component results of pain responder and non-responder groups pre and post-TKA represented as PC score means and standard deviations. Significant ($p<0.05$) features in bold.

WOMAC Pain Domain

Feature	Interpretation	Var. Exp.	Pre-TKA			Post-TKA			Within group difference	
			Responder (n=34)	Non-responder (n=12)	*p*	Responder (n=34)	Non-responder (n=12)	*p*	Responder *p*	Non-responder *p*
Flexion Angle										
PC1	Gait cycle flexion angle magnitude	65.09%	-13.70 (57.97)	-10.55 (58.16)	0.9	21.44 (60.69)	21.35 (41.31)	0.1	**0.02**	0.1
PC2	Stance to swing angle range	15.79%	-2.23 (40.75)	-7.12 (31.53)	0.7	7.88 (31.40)	4.67 (29.45)	0.2	0.3	0.4
PC3	Phase shift: timing of stance and peaks	11.91%	-8.20 (31.07)	2.22 (32.64)	0.3	4.38 (22.68)	-0.99 (28.72)	0.8	0.06	0.8
PC4	Stance-phase range of motion	2.60%	-4.18 (12.70)	2.55 (11.83)	0.1	0.16 (11.16)	2.45 (17.43)	0.5	0.4	0.9
Adduction Angle										
PC1	Stance-phase adduction angle magnitude	57.40%	3.10 (19.75)	17.52 (18.29)	**0.03**	-5.21 (20.04)	-15.14 (20.77)	**0.001**	0.09	**<0.001**
PC2	Mid-stance to terminal stance range	24.04%	3.10 (19.75)	-1.21 (9.16)	0.8	-2.02 (13.00)	-2.03 (7.53)	0.8	0.6	0.8
PC3	Heal strike to mid-stance range	8.60%	0.10 (17.11)	0.45 (12.01)	0.9	0.20 (5.66)	3.26 (7.79)	0.9	0.9	0.5
Flexion Moment										
PC1	Gait cycle flexion moment magnitude	72.59%	0.08 (2.03)	-0.17 (1.62)	0.7	-0.48 (1.21)	0.01 (1.69)	0.5	0.2	0.8
PC2	Stance-phase flexion moment range	16.53%	-0.21 (0.63)	-0.33 (0.43)	0.6	0.19 (0.61)	0.21 (0.65)	**0.01**	**0.01**	**0.03**
PC3	Phase shift: timing of flexion peaks	3.90%	0.00 (0.47)	-0.01 (0.27)	0.9	0.00 (0.29)	-0.05 (0.26)	0.9	0.9	0.7
Adduction Moment										
PC1	Stance-phase adduction moment magnitude	83.17%	0.06 (1.67)	0.05 (1.63)	0.9	-0.15 (0.88)	0.07 (0.86)	0.6	0.5	0.9
PC2	First peak and mid-stance range	8.40%	-0.14 (0.33)	-0.28 (0.40)	0.2	0.09 (0.34)	-0.09 (0.36)	**0.003**	**0.006**	0.2
PC3	Mid-stance and second peak range	3.20%	-0.07 (0.27)	-0.08 (0.31)	0.9	0.13 (0.23)	0.17 (0.32)	**0.02**	**0.002**	0.06

Table 4.3. Baseline gait features and change in gait features contributing to clinically meaningful improvements in self-reported pain and function from multivariable modified Poisson regression. Items were scaled (0-10), where a one-point increase in RR was associated with a 10% change in PC score. Significant ($p<0.05$) features in bold.

	Interpretation	RR	95% CI	Estimate	Std. Error	*p*
Pain domain ($r^2 = 0.14$[‖]):						
Adduction Angle PC1	Pre-TKA magnitude of stance-phase varus alignment	0.915	(0.838, 0.998)	-0.089	0.045	**<0.05**
Function domain ($r^2 = 0.15$[‖]):						
Flexion Angle PC4	Pre-TKA flexion angle range of motion during stance	0.898	(0.827, 0.976)	-0.107	0.042	**0.01**

[‖]Linear models were applied using the magnitude of WOMAC domain improvement as the independent variable to provide estimates of r^2

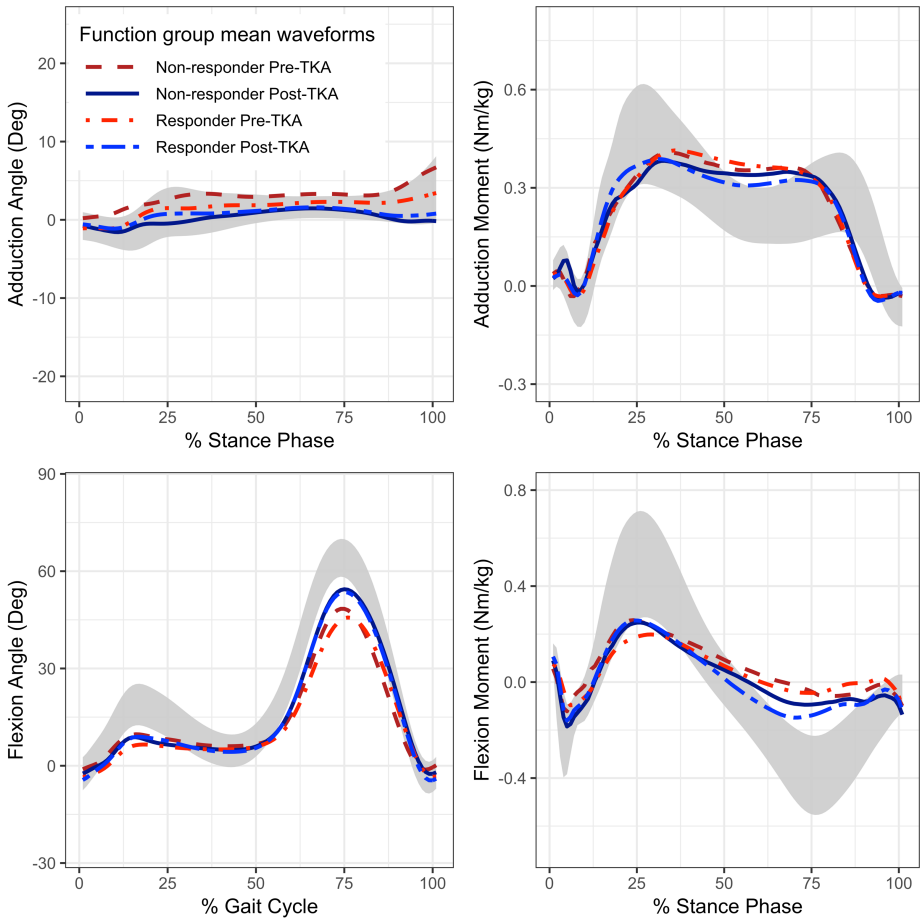

Figure 4.4. Mean waveforms of function responder (n=36) and non-responder (n=10) groups collected one-week pre-TKA and one-year post-TKA. Grey denotes 209 previously captured asymptomatic (mean age 51.6 ±9.0, BMI 26.7±4.5) gait patterns (mean ± 1SD).

Table 4.4. Principal component results of function responder and non-responder groups pre and post-TKA represented as PC score means and standard deviations. Significant ($p<0.05$) features in bold.

WOMAC Physical Function Domain

Feature	Interpretation	Var. Exp.	Pre-TKA			Post-TKA			Within group difference	
			Responder (n=36)	Non-responder (n=10)	*p*	Responder (n=36)	Non-responder (n=10)	*p*	Responder *p*	Non-responder *p*
Flexion Angle										
PC1	Gait cycle flexion angle magnitude	65.09%	-14.90 (56.42)	-5.59 (63.33)	0.7	20.43 (58.94)	24.96 (45.40)	0.2	**0.01**	0.2
PC2	Stance to swing angle range	15.79%	-0.33 (40.32)	-14.92 (28.65)	0.3	7.06 (28.97)	6.98 (37.71)	**0.03**	0.4	0.2
PC3	Phase shift: timing of stance and peaks	11.91%	-8.91 (30.64)	6.86 (32.91)	0.2	3.89 (24.13)	-0.29 (25.38)	0.8	0.05	0.6
PC4	Stance-phase range of motion	2.60%	-4.99 (11.58)	6.78 (12.84)	**0.008**	0.95 (10.85)	0.06 (19.36)	0.2	**0.03**	0.4
Adduction Angle										
PC1	Stance-phase adduction angle magnitude	57.40%	3.75 (20.31)	18.07 (16.15)	**<0.05**	-6.79 (19.81)	-11.47 (23.49)	**0.001**	**0.03**	**0.005**
PC2	Mid-stance to terminal stance range	24.04%	0.02 (16.64)	-1.18 (10.03)	0.8	-2.24 (12.82)	-1.26 (7.02)	0.8	0.5	0.9
PC3	Heal strike to mid-stance range	8.60%	1.21 (9.91)	-2.79 (8.69)	0.3	0.49 (5.95)	2.83 (7.62)	0.2	0.7	0.1
Flexion Moment										
PC1	Gait cycle flexion moment magnitude	72.59%	-0.03 (1.99)	0.18 (1.68)	0.8	-0.39 (1.23)	-0.20 (1.80)	0.2	0.5	0.6
PC2	Stance-phase flexion moment range	16.53%	-0.29 (0.56)	-0.07 (0.66)	0.3	0.24 (0.57)	0.03 (0.77)	0.2	**<0.001**	0.8
PC3	Phase shift: timing of flexion peaks	3.90%	0.00 (0.43)	0.01 (0.45)	0.9	0.01 (0.29)	-0.08 (0.27)	0.9	0.9	0.6
Adduction Moment										
PC1	Stance-phase adduction moment magnitude	83.17%	0.06 (1.64)	0.04 (1.73)	0.9	-0.10 (0.85)	-0.05 (0.99)	0.7	0.6	0.9
PC2	First peak and mid-stance range	8.40%	-0.19 (0.38)	-0.12 (0.27)	0.6	0.10 (0.35)	-0.13 (0.31)	0.07	**0.001**	0.9
PC3	Mid-stance and second peak range	3.20%	-0.08 (0.26)	-0.04 (0.36)	0.6	0.15 (0.24)	0.13 (0.31)	0.07	**<0.001**	0.3

4.4. Discussion

Functional responders were characterized biomechanically by less stance-phase flexion-extension angle range, and lower adduction angle magnitudes pre-operatively (Table 4.4, Figure 4.4). In multivariate modeling, less stance-phase flexion-extension range was the only feature predictive of being a functional responder (Table 4.3). This is in agreement with a similar study by Naili et al. (n=28) who reported less sagittal plane knee angle range pre-TKA (stance to swing, 45±6° vs. 52±5°) in patients meeting minimal detectable change criteria in knee-related quality-of-life scores (Naili et al., 2017). Less sagittal range is typically associated with "more severe" or stiffer sagittal plane kinematics, resembling more severe OA pattern norms (Astephen et al., 2008a, 2008b) (Figure 4.4). Younger age was also associated with more WOMAC function improvements univariately (r=-0.41, p=0.005, Figure 4.3). Although younger patients typically report less satisfaction post-TKA (Williams et al., 2013a), they have been associated with more self-reported improvements (Alzahrani et al., 2011; Williams et al., 2013a), attributed to improved functional abilities captured within activities of daily living scores. Of the function responders ≤ 55 in this study, most (4/5) demonstrated stance-phase flexion-extension angle ranges (PC4) below the norm pre-operatively, potentially representing a young stiff sagittal kinematics subset. Stiff kinematics, coupled with more radiographic severity (p=0.03), and trends towards more symptomatic severity (Table 4.1) aligns with previous inferences that more severe patients (typically measured by PROMs) tend to have better arthroplasty outcomes (Fortin et al., 1999; Hawker et al., 2013; Jiang et al., 2017; Naili et al., 2017). This study suggests severity could be captured objectively during gait with knee kinematics. Further, these kinematic features could be detectable in clinical settings through simple wearable or marker-less motion capture.

The only biomechanical gait feature descriptive of pain responders pre-TKA were lower stance-phase varus magnitudes, suggested by comparative tests and in multivariate modeling (Table 4.2,4.3). Conversely, pain non-responders appeared more varus during stance pre-operatively. While static and dynamic varus alignment have both been associated with further medial compartment OA severity (Henriksen et al., 2012;

Miyazaki et al., 2002; Orishimo et al., 2012), less radiographic and symptom severity in pain non-responders (Table 4.1) suggests a potential kinematic subgroup of constitutionally varus aligned (Bellemans et al., 2012) or kinematically varus individuals (Chang et al., 2010). Although interesting, these results should be interpreted with caution. Our exploratory approach did not account for multiple comparisons. This, coupled with small non-responder group sizes increased the possibility of Type I errors, and resulted in large confidence intervals around our estimates. However, visualizations of kinematic data did suggest that 10/12 pain (and 9/10 function) non-responders had pre-operative varus angle magnitudes above the norm. If the soft tissue and musculature surrounding the joint has adapted to native varus kinematics (Young et al., 2015), mechanics after standard arthroplasty might be perceived as un-natural, potentially contributing to less self-reported pain and function improvements. It has been suggested that standardized alignment may not be optimal for all patients (Blakeney et al., 2018; Parratte et al., 2010; Vanlommel et al., 2013). Vanlommel et al. reported significantly better function and knee scores in pre-operatively varus individuals, whose post-operative surgical alignment remained mildly varus (Vanlommel et al., 2013). Under these assumptions, native varus magnitudes might be a false signal during arthroplasty patient selection, or this presentation with an absence of other severe OA features might characterize clinical candidate subgroups for whom neutral corrections are not "clinically relevant" (Sosdian et al., 2014). Investigating patient biomechanical variability with respect to outcomes in larger studies is an important area for further research. These groups might benefit from altered clinical or surgical approach (such as individualized alignment, or a high tibial osteotomy), relative to standard of care arthroplasty.

Patients who self-report less pain and function improvement post-operatively appeared to demonstrate less objective functional improvements during gait. Non-responders showed significantly reduced stance-phase varus angles post-TKA, yet lagged in terms of sagittal kinematic and kinetic loading pattern corrections typically reported in population average studies (Table 4.2,4.4) (Bonnefoy-Mazure et al., 2017; Hatfield et al., 2011; Smith et al., 2004; Wilson et al., 2015). Naili et al. proposed poor patient-reported outcomes might be partially explained by a lack of dynamic kinematic and kinetic

corrections, despite alignment corrections in the frontal plane, a feature surgery may be most able to biomechanically address (Naili et al., 2017). Although our results suggested less three-dimensional corrections among non-responders overall (Tables 4.2,4.4), we did find frontal plane changes associated with self-reported pain and function gains. Specifically, less varus magnitude reductions (in both pain and function), and larger increases in dynamic frontal plane loading (PC2) (in function alone), were independently associated with more PROMs improvements (Figure 4.3). This was a unique finding, supporting our interpretation that standard arthroplasty might not be optimal for a subset of patients. Post-hoc tests also found no difference in the magnitude of varus reduction imposed among the five surgeons involved ($p \geq 0.8$). Further work should investigate if individualized frontal plane mechanics approaches during surgery and rehabilitation have subsequent benefits to the three-dimensional gait features not consistently addressed among non-responders by standard arthroplasty.

Despite being a relatively large three-dimensional gait study sample, fewer non-responders relative to responders makes generalizing results to the TKA population difficult. We instead aimed to provide insights between comprehensive biomechanical and clinical datasets, and share valuable information to guide targeted research. Our exploratory approach did not correct for multiple comparisons, and results should be interpreted as preliminary evidence of patient subgroups that may benefit from altered treatment approaches. Further, the power in our ability to detect pre to post-TKA gait changes among non-responders was low (9-32%). However, small permutations between pain and function non-responder groups (non-responder overlap of 8/12 and 8/10 respectively), operated as a natural sensitivity analysis, improving confidence in findings reported in both domains. Radiographs for KL grade were not available for all individuals, nor were full-leg standing radiographs, limiting our ability to translate stance-phase findings to static alignment traditionally considered surgically. Using MCID thresholds to dichotomize outcomes was also not without limitations. MCID thresholds are not applicable for measuring individual change for all patients, nor do they translate well to global metrics, such as satisfaction (Katz et al., 2015; Williams et al., 2013b). MCID can further be influenced by symptom severity pre-operatively (Escobar et al.,

2013), and questionnaire ceiling effects may restrict rates of patients meeting MCID thresholds, despite having improved. Still, PROM responsiveness scores have been recommended by the International Society of Arthroplasty Registries Working Group (Rolfson et al., 2016) and others (Roos, 2018), due to their ability to improve within and between-patient score interpretations from interventions. Pain and function domains were selected as they tend to be key outcomes assessed post-TKA, and the domains most associated with satisfaction (Scott et al., 2010). Seventy-four and 78% of patients met MCID pain and function thresholds, greater than a previous Canadian study (Hawker et al., 2013), but aligning closely with others (Escobar et al., 2007; Quintana et al., 2006) and the 20% dissatisfaction rate typically reported post-TKA (Robertsson and Dunbar, 2001). WOMAC pain and function domains also tend to be less susceptible to floor and ceiling effects over joint stiffness (Dunbar et al., 2001); no study non-responders experienced ceilings post-operatively.

This study contributes to the growing body of evidence which suggests variability in patient-reported outcomes may be partially explained by a combination of clinical and objectively measured knee joint biomechanical factors. Specifically, more "severe" objective gait features pre-operatively tended to be associated with a larger potential for both objective and self-reported functional improvements (Naili et al., 2017). A unique finding of this study was the indication of varus kinematic subgroups who may be susceptible to less pain and function improvements from standard arthroplasty, and that larger reductions in stance-phase varus alignment may be un-favourable to some patients. These trends warrant further investigation. Objective functional assessment pre-operatively may aid in identifying the optimal functional state (the "sweet spot") associated with patient-reported improvements, and help identify those who may benefit from an individualized approach; informing triaging, surgical planning and expectation management strategies. Our findings support the notion that TKA innovation needs to better understand three-dimensional knee mechanics at an individual level, to provide expected improvements for all patients.

Chapter 5. Biomechanics Phenotypes Among TKA Candidates

5.1. Introduction

Osteoarthritis (OA) populations are increasingly being characterized as heterogeneous, composed of multiple distinct subtypes or phenotypes, presenting clinically through various forms of knee joint-level and symptom manifestations (Andriacchi et al., 2014; Bannuru et al., 2019; Castaneda et al., 2013; Lane et al., 2011). Regardless of OA phenotype presentation, the most common end-stage treatment for knee OA is total knee arthroplasty (TKA), an inherently mechanical surgery which follows a standardized approach for most patients regardless of the heterogeneity. However, patient-specific joint mechanics during gait pre-operatively, and the magnitude of change in joint mechanics as a result of surgical intervention may contribute to both patient-reported and objective functional outcomes after arthroplasty (P R Biggs et al., 2019; Naili et al., 2017; Young-Shand et al., 2020). Specifically, OA patients with lower functioning gait patterns pre-operatively may experience greater self-reported and objective benefits from arthroplasty (Naili et al., 2017; Young-Shand et al., 2020). This work also proposed the existence of gait kinematic phenotypes, for whom standard TKA might yield better (and worse) patient-reported outcomes (Naili et al., 2017; Young-Shand et al., 2020). Understanding these patient phenotypes is relevant to intervention strategies (Felson, 2010), as treatment targeting individual manifestations or deficiencies might yield greater self-reported improvements (Naili et al., 2017; Young-Shand et al., 2020), and optimize the number of patients whose gait function is restored to healthy norms (Paul Robert Biggs et al., 2019; Outerleys et al., 2021).

To date, joint-level biomechanical variability during gait among TKA candidates has primarily been assessed through *a priori* group definition, stratifying OA patients by sex (Paterson et al., 2017; Wilson et al., 2015), frontal plane alignment (Turcot et al., 2012), obesity (Paterson et al., 2017), or patient-reported symptoms before (Thorp et al., 2007; Wilson et al., 2017), and after arthroplasty (Naili et al., 2017; Young-Shand et al., 2020). While helpful in understanding how specific patient groups respond to standard of care

surgery, these studies provide little evidence into how patient demographic and joint-level features naturally separate. Data-driven strategies such as multivariate unsupervised data partitioning (i.e., data clustering) can characterize phenotypes by modes of statistical separation and similarity, and have been recommended for OA phenotype research (Spil et al., 2020). Few studies have applied unsupervised methodologies to OA populations. Knoop et al. identified five clusters by radiographic severity, muscle strength, body mass index (BMI), and depression scores (Knoop et al., 2011). Elbaz et al. reported four clusters using spatiotemporal gait parameters (Elbaz et al., 2014). Most recently, Waarsing et al. identified four clusters by demographics, cartilage features, mechanical alignment, and self-reported symptoms, with clusters strongly characterized by joint-level structural degradation (Waarsing et al., 2015), indicative of variable functional loading (Andriacchi et al., 2009). Phenotyping discussions consistently propose the existence of biomechanical-driven OA subtypes (Andriacchi et al., 2014; Bannuru et al., 2019; Castaneda et al., 2013), however prior studies lack information on joint-level biomechanics, and have not addressed outcomes of interventions such as TKA at a cluster/group level. With the goal of individualizing treatment plans to patient groups, there is a need to better understand TKA candidate biomechanical variability to inform targeted non-invasive rehabilitation approaches, surgical correction and decision-making strategies, or prosthesis innovations addressing phenotype-specific biomechanical deficits.

The primary aim of this study was to identify knee joint biomechanical phenotypes (clusters) among TKA candidates based on similarities in patient demographics, frontal and sagittal plane knee kinematics and kinetics during gait. The secondary aim was to compare objective functional gait improvements between phenotypes after standard TKA. It was hypothesized that distinct biomechanical phenotypes exist within TKA populations, varying by levels of knee function severity (Young-Shand et al., 2020), and that patient phenotype categorization would be associated with the degree of objective functional improvement experience after TKA.

5.2. Methods

This was a secondary study of severe knee OA patients scheduled for TKA between 2003 and 2016 recruited from a high-volume academic orthopaedic clinic to undergo three-dimensional (3D) instrumented gait analysis one week prior to (n=135) and one-year post-TKA (n=109) (Young-Shand et al., 2020). Included patients were able to walk six meters unassisted without a walking aid. Patients were excluded if they screened positive for neurological disease, or other conditions that might affect their gait or ability to safely participate in walking trails. TKA followed a standard medial parapatellar approach, with femur and tibial cuts targeting neutral mechanical alignment. The measured resection technique was used to achieve a balanced flexion-extension gap. Informed consent was obtained from all participants according to the institution ethics board.

5.2.1. Gait Analysis Data Collection

Patient factors and demographics, including age, sex, mass, height, and OA structural severity, as graded by an orthopaedic surgeon using the Kellgren-Lawrence (KL) global radiographic score (Kellgren and Lawrence, 1957) were collected as part of a gait assessment protocol. Four triads of infrared light-emitting markers were attached to the surgical lower limb segments of the pelvis, thigh, shank, and foot to establish local rigid body coordinate systems. Individual light-emitting markers were placed on boney landmarks of the shoulder, greater trochanter, lateral epicondyle and lateral malleolus, while eight digitized points measured during quiet standing were used to define local anatomical joint axes following a standardized protocol (Landry et al., 2007). Participants walked along a five-meter walkway wearing comfortable walking shoes at a self-selected speed. An Optotrak™ (Northern Digital Inc., Waterloo, ON, Canada) optoelectronic motion capture system sampling at 100 Hz captured three-dimensional kinematics of the lower limb, and was synchronized to a walkway embedded AMTI Biomechanics Platform System (Advanced Medical Technology Inc., Watertown, MA) sampling at 2000 Hz to capture three-dimensional external ground reaction forces. Knee joint angles during gait were calculated using the joint coordinate system (Grood and Suntay, 1983). Net resultant knee joint moments were calculated using inverse dynamics (Costigan et al.,

1992; Deluzio et al., 1993; Li et al., 1993), amplitude normalized to body mass. Up to seven (minimum four) walking trials were averaged and normalized for each participant to one gait cycle (0-100%) for flexion/extension angles, and to stance phase (0-100%) for 3D knee moments and adduction angles.

5.2.2. Data Preparation for Clustering

Principal Component Analysis (PCA), a multivariate statistical analysis and data reduction technique (Deluzio and Astephen, 2007), was used to extract temporal gait angle and moment features, having demonstrated improved sensitivity over discrete feature selection alone (Hatfield et al., 2011). Previously reported (Young-Shand et al., 2020) frontal and sagittal knee angle and moment gait PCs that described major features of gait variability were extracted. Pre-TKA demographics (age, sex), body mass index (BMI), walking speed, and pre-TKA frontal and sagittal gait angle and moment PC score features previously found to differentiate sex (Wilson et al., 2015), OA severity (Astephen et al., 2008a, 2008b), symptoms (Thorp et al., 2007; Wilson et al., 2017), and self-reported patient outcomes (Naili et al., 2017; Young-Shand et al., 2020), were selected for cluster analysis (15 features, Table 5.1).

Clustering can be sensitive to outliers, so all features were standardized to z-scores (mean=0, standard deviation=1) and assessed for outliers (Milligan and Cooper, 1988). One pre-TKA session with a feature exceeding Tukey's outer (3*IQR) fence was removed, resulting in a cluster dataset of $n_{pre\text{-}TKA}$=134 patient waveforms. Of these patients, one-year matched pre to post-TKA instances were available from $n_{post\text{-}TKA}$=105 instances. One missing observation was due to data collection errors, the remaining were lost to follow up at the time of data extraction.

5.2.3. Clustering and Statistical Analysis

To gain insights into variations of joint-level biomechanical phenotypes among TKA candidates, machine learning based cluster analysis was applied to the pre-TKA dataset ([134x15]; 15 features). Most clustering algorithms use proximity distance metrics (e.g., Euclidian or Manhattan distances) to summarize similarity (or dissimilarity) between data points. However, when using high-dimensional data, the distance between the furthest

and closest point (relative contrast) diminishes, reducing the ability of distance metrics to discriminate near and far neighbours (Lee and Verleysen, 2007). Dimension reduction methods can transform the high-dimensional data into a lower-dimensional space, while preserving the cluster structure and therefore restoring the relevance of distance metrics (Lee and Verleysen, 2007). As such, two-dimensional multidimensional scaling (MDS) (Gower, 1966), a dimension reduction technique using Manhattan distances was applied. This represented demographic and pre-TKA PC score data in lower-dimensional coordinates ([134x2]; 2 features) while preserving inherent multi-dimensional distances between observations from the original higher-dimensional space (Lee and Verleysen, 2007). Hierarchical agglomerative (bottom-up) cluster analysis using Ward's minimum variance criteria (Clatworthy et al., 2005; Ward, 1963) was then applied to the MDS coordinates, the most commonly used unsupervised technique in the OA literature (Deveza et al., 2017). Each observation is considered a single cluster (a leaf), and at each step, the two clusters associated with the minimal increase in variance via Ward's criteria are combined (two leaves are merged), forming a new cluster. This process is repeated until all observations are members of a single root. Several clustering solutions (number of clusters k=2:10) were assessed using mean silhouette width criterion (Rousseeuw, 1987). A silhouette width coefficient s of observation x measures how similar observation x is to its own cluster (cohesion) compared to other clusters (separation). A silhouette width close to -1 indicates observation x is "misclassified", much closer to a neighbouring cluster over its own. A silhouette width close to 1 indicates observation x is "well clustered", much closer to its own cluster over a neighbouring one. Thus, the best clustering solutions are the ones with the highest mean silhouette widths. We used an additional criterion, the Adjusted Rand Index (ARI), to quantify cluster stability, a metric bound by ± 1 (0[completely random]-1[perfect agreement]) (Hubert, 1985). The cluster model of the full data set was compared with 100 subsets of the full dataset (fraction of 0.8) for k=2:10, and mean ARI was reported. The selected clustering solution (k clusters) with the greatest mean s and ARI was chosen, reported in the Results.

To validate clusters and to understand how clusters differ in TKA response, between-cluster differences in person factors, demographics and gait features were examined pre-

TKA (n=134) and pre to post-TKA in terms of gait changes ($\Delta PCscore = Post_{PCscore} - Pre_{PCscore}$, n=105 matched pre to post-TKA instances). Chi-squared (sex), Kruskal-Wallace (KL score), and k-way ANOVA (age, BMI, gait speed, pre-TKA PC score, ΔPC score) tests with Tukey's honestly significant difference (HSD) were used. Intra-cluster pairwise pre vs. post-TKA (n=105) gait differences within clusters were examined using paired Student's t-tests. Using Bonferroni corrections to adjust for multiple comparisons, observations at the 0.001 significance level were considered significant. All analyses were conducted using R (R Foundation for Statistical Computing, Vienna, Austria).

5.3. Results

By applying silhouette and ARI criteria (s=0.37, ARI=0.57, Appendix B, Table B.1, Figure B.1), the best clustering solution had four (k=4) TKA candidate clusters with inherent similarities.

5.3.1. Inter-Cluster Characterization Pre-TKA

Cluster 1: 'Higher functioning males' was a small compact cluster (n=7, Appendix B Figure B.1) consisting of only males. Cluster 1 patients walked with faster self-selected gait speeds than clusters 2 and 3 (mean 1.2 SD 0.2 m/s; 3<2<,4,1; p<0.001), and had the largest flexion angle (PC4; 3,4,2<1; p<0.001) and flexion moment range (PC2; 3<2<4<1; p<0.001) among the clusters during stance pre-TKA. Cluster 1 also had the highest stance-phase adduction moment magnitudes (PC1; 3,4<2,1; p<0.001), and adduction moment range from the first peak to mid-stance (PC2; 3,2<4<1; p<0.001), and from mid-stance to second peak (PC3; 3,2<1; p=0.001). Cluster 1 therefore showed the greatest dynamic loading/unloading kinetics and kinematic range patterns, representing a higher functioning male subset, most resembling asymptomatic patterns (Astephen et al., 2008a) (Figure 5.1, Table 5.1).

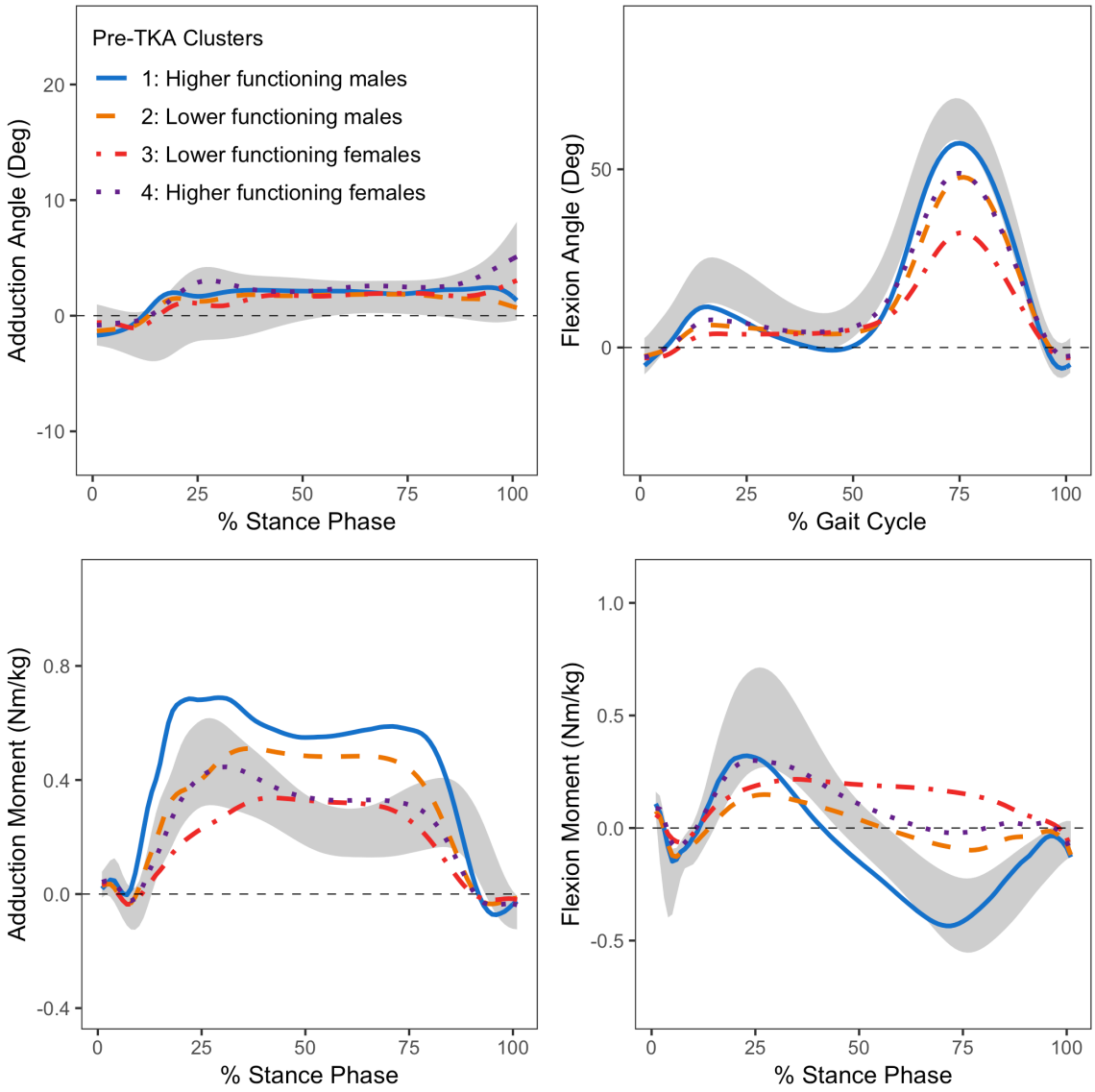

Figure 5.1. Mean pre-TKA gait waveforms by cluster; cluster 1 (solid), cluster 2 (dashed), cluster 3 (dot-dashed), and cluster 4 (dotted). Grey denotes 209 previously captured asymptomatic (mean age 51.6 ±9.0, BMI 26.7±4.5) gait patterns (mean ± 1SD).

Cluster 2: 'Lower functioning males' also consisted of mostly males (44/47) who walked with slower gait speeds than clusters 1 and 4 (mean 0.8 SD 0.2 m/s; 3<2<4,1; p<0.001), and trended towards being older (mean 67.0 SD 7.4 years; 1,4<2; p=0.006). Cluster 2 also had higher overall knee adduction moment magnitudes (PC1; 3,4<2,1; p<0.001) than clusters 3 and 4, yet walked with less biphasic kinetic and kinetic range

(more “stiff-kneed”) than cluster 1, demonstrated by less adduction moment range from first peak to mid-stance (PC2; 3,2<4<1; p<0.001), from mid-stance to the second peak (PC3; 3,2<1; p=0.001), and less stance-phase flexion moment range during stance (PC2; 3<2<4<1; p<0.001). Cluster 2 therefore represented a slower, lower functioning stiff-kneed male subset (Figure 5.1, Table 5.1).

Cluster 3: ‘Lower functional females’ was mostly (32/34) female with the slowest gait speeds (mean 1.2 SD 0.2 m/s; 3<2<4,1; p<0.001) among the clusters. Cluster 3 also had the lowest overall flexion angle magnitudes (PC1; 3<2,4,1; p<0.001). Kinetically, cluster 3 had large overall flexion moment magnitudes (PC1; 1,4<3 & 2<3; p<0.001), the least stance-phase flexion moment range (PC2; 3<2<4<1; p<0.001), and less adduction moment change from first peak to mid-stance (PC2; 3,2<4<1; p<0.001), and mid-stance to the second peak (PC3; 3,2<1; p=0.001). Cluster 3 therefore captured a slow, lower functioning female subset, with the stiffest-kneed gait among the clusters (Figure 5.1, Table 5.1).

Cluster 4: ‘Higher functioning females’ also consisted of mostly females (43/46) who walked with faster gait speeds (mean 1.0 SD 0.1 m/s; 3<2<4,1; p<0.001). Similar to the other female cluster, cluster 4 had lower overall knee adduction moment magnitudes (PC1; 3,4<2,1; p<0.001). However, cluster 4 also had more dynamic kinetic frontal (adduction moment PC2; 3,2<4<1; p<0.001) and sagittal (flexion moment PC2; 3<2<4<1; p<0.001) loading range relative to clusters 2 and 3, yet was more stiff-kneed relative to cluster 1. Cluster 4 therefore represented a higher functioning female subset (Figure 5.1, Table 5.1).

OA radiographic severity did not differ between clusters (Kellgren-Lawrence global rating, p=0.4 from 102/134 cases, Table 5.1). Pre-TKA, predominantly male clusters (1 and 2) had greater overall adduction moment magnitudes relative to mostly female clusters (PC1; 3,4<2,1; p<0.001). Higher functioning clusters (1 and 4), walked at faster gait speeds (3<2<4,1; p<0.001) and had more dynamic loading/un-loading kinetics, shown by more stance-phase flexion (PC2; 3<2<4<1; p<0.001) and adduction (PC2; 3,2<4<1; p<0.001) moment range relative to clusters 2 and 3.

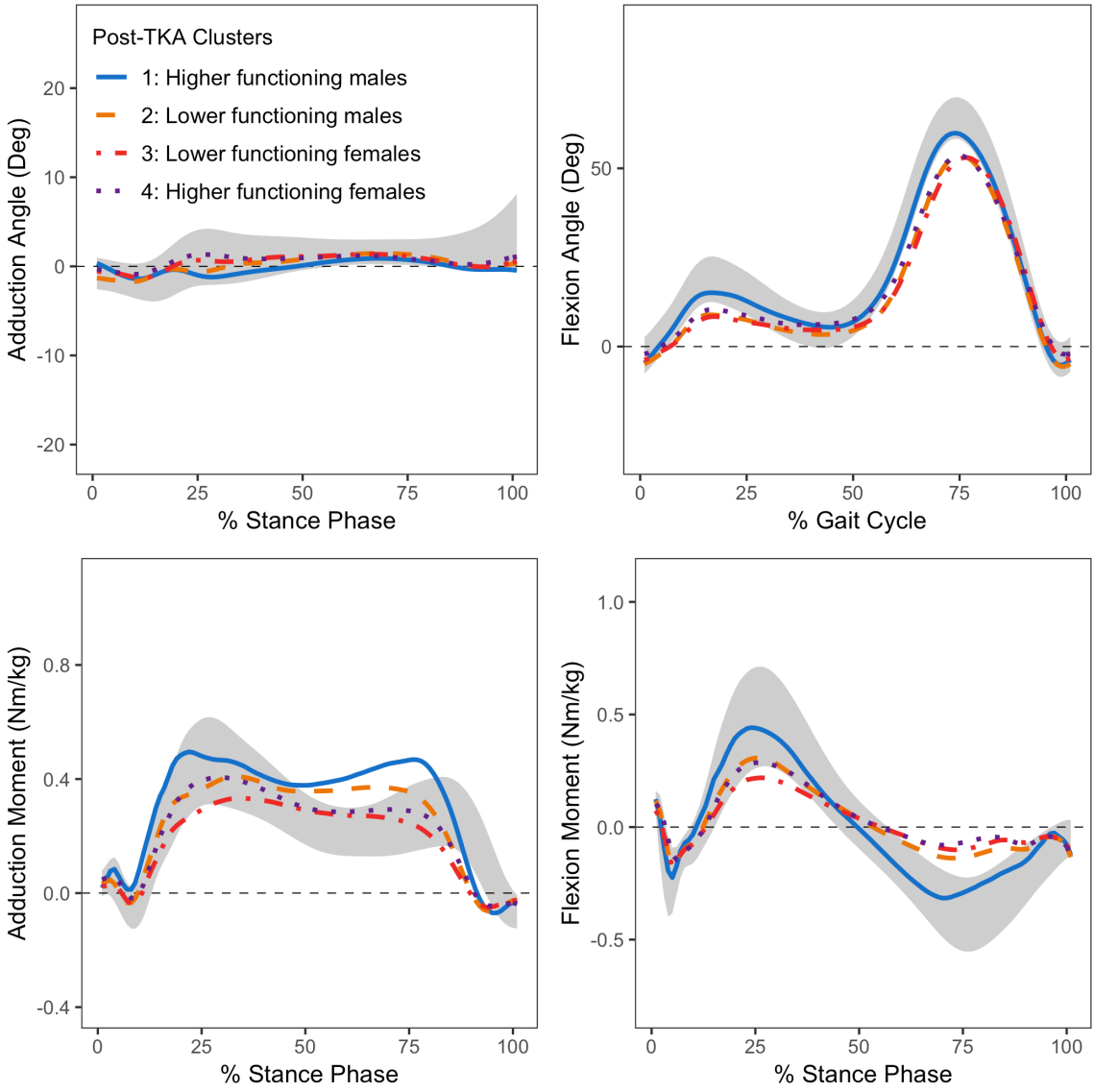

Figure 5.2. Mean post-TKA gait waveforms by cluster; cluster 1 (solid), cluster 2 (dashed), cluster 3 (dot-dashed), and cluster 4 (dotted). Grey denotes 209 previously captured asymptomatic (mean age 51.6 ±9.0, BMI 26.7±4.5) gait patterns (mean ± 1SD).

5.3.2. Inter-Cluster and Intra-Cluster Changes Pre to Post-TKA

Pre to post-TKA (n=105, Figure 5.2, Table 5.2), both lower functioning male and female clusters (2 and 3) demonstrated greater increases in gait speed (4<2,3, p<0.001), and tended to demonstrate more improvements in knee flexion moment, and knee adduction moment loading patterns pre to post-TKA (Table 5.2), changes typically

expected to increase (improve) from surgery (Hatfield et al., 2011). For example, cluster 2, lower functioning males, experienced improvements in frontal plane adduction moment features, shown by pre to post-TKA changes in adduction moment PCs 1-3. However, the magnitude of these changes did not differ by those experienced by other clusters. Cluster 3, lower functioning females, did demonstrate more gait feature improvements than the other clusters, generally in the sagittal plane. This was shown by greater increases in overall flexion angle magnitude (ΔPC1; 2,4<3; $p=0.001$), stance to swing range (ΔPC2; 1,4,2<3; $p<0.001$), and a larger decrease in overall flexion moment loading magnitude (ΔPC1; 3<4,2,1; $p<0.001$).

Conversely, the highest functioning phenotype, cluster 1, higher functioning males, experienced no significant improvement in any gait feature pre to post-TKA ($p>0.02$). Instead, this group showed trends towards reductions in flexion angle range (PC2; $p=0.05$) and increases in flexion loading magnitudes (PC1; $p=0.04$), changes that would typically move patterns further away from asymptomatic norms. Higher functioning females, cluster 4, only demonstrated a decrease in adduction angle magnitude post-TKA (PC1; $p<0.001$), with no other kinematic or kinetic gait features significantly improving.

Table 5.1. Baseline demographic and gait feature (PC score, % variance explained) differences between clustered groups, with significant features bolded. Normally distributed variables represented as means and standard deviations, non-normally distributed variables represented as medians and 95% confidence intervals. PCs were defined such that larger values indicate greater magnitudes in the direction of the provided interpretation. E.g., larger Knee Flexion Angle PC1 indicates greater flexion angle magnitudes.

Feature	1: Higher functioning males	2: Lower functioning males	3: Lower functioning females	4: Higher functioning females	*p*	Tukey HSD
Male, Female (n)	0, 7	3, 44	32, 2	43, 3	**<0.001**	
Age (years)	59.1 (8.5)	67.0 (7.4)	62.9 (8.1)	62.6 (6.7)	0.006	
BMI (Nm/kg)	31.3 (5.0)	33.1 (6.2)	35.1 (6.6)	32.2 (5.7)	0.2	
KL Global Score	4.0 (3,4)	4.0 (3,4)	4.0 (3,4)	4.0 (3,4)		
KL Global of 0:1:2:3:4	0:0:0:2:4	0:0:1:9:25	0:0:1:12:15	0:0:0:16:17	0.4	
Gait speed (m/s)	1.2 (0.2)	0.8 (0.2)	0.7 (0.1)	1.0 (0.1)	**<0.001**	3<2<4,1
Knee Flexion Angles						
Gait cycle flexion angle magnitude (PC1; 65.1%)	30.3 (32.6)	-12.2 (54.8)	-77.5 (60.8)	-1.4 (55.6)	**<0.001**	3<2,4,1
Stance to swing angle range (PC2, 15.8%)	24.2 (6.7)	6.0 (26.6)	-15.9 (37.1)	-1.5 (35.0)	0.004	
Phase shift: earlier terminal stance toe off (PC3; 11.9%)	12.6 (32.0)	-7.1 (28.1)	-8.4 (32.2)	3.4 (22.8)	0.08	
Stance-phase range of motion (PC4; 2.6%)	21.0 (12.8)	3.3 (10.5)	-5.1 (8.6)	0.0 (13.2)	**<0.001**	3,4,2<1; 3<2
Knee Adduction Angles						
Stance-phase adduction angle magnitude (PC1; 57.4%)	3.7 (12.0)	-1.1 (20.4)	0.5 (24.1)	10.2 (20.7)	0.06	
Knee Flexion Moments						
Gait cycle flexion moment magnitude (PC1; 72.6%)	-1.5 (1.0)	-0.4 (1.7)	1.0 (1.8)	0.4 (1.7)	**<0.001**	1<4,3; 2<3
Stance-phase flexion moment range (PC2; 16.5%)	1.2 (0.5)	-0.3 (0.5)	-0.8 (0.4)	0.0 (0.6)	**<0.001**	3<2<4<1
Phase shift: earlier flexion/extension peaks (PC3; 3.9%)	0.3 (0.5)	-0.1 (0.4)	0.0 (0.4)	0.1 (0.4)	0.05	
Knee Adduction Moments						
Stance-phase adduction moment magnitude (PC1; 83.2%)	2.2 (1.2)	0.9 (1.1)	-0.5 (2.0)	0.0 (1.6)	**<0.001**	3,4<2,1
First peak and mid-stance range (PC2; 8.4%)	0.5 (0.6)	-0.2 (0.3)	-0.4 (0.3)	0.1 (0.3)	**<0.001**	3,2<4<1
Mid-stance and second peak range (PC3; 3.2%)	0.3 (0.3)	-0.1 (0.3)	-0.1 (0.3)	0.0 (0.3)	**0.001**	3,2<1

Table 5.2. Change (Δ: post-TKA – pre-TKA) in gait speed and gait features (PC score) between and within clustered groups, with significant features bolded. Variables represented as means and standard deviations.

	1: Higher functioning males				2: Lower functioning males				3: Lower functioning females				4: Higher functioning females				*p*
	Pre-TKA	**Post-TKA**	**Δ**	***p*** $_{Pre/Post}$	**Pre-TKA**	**Post-TKA**	**Δ**	***p*** $_{Pre/Post}$	**Pre-TKA**	**Post-TKA**	**Δ**	***p*** $_{Pre/Post}$	**Pre-TKA**	**Post-TKA**	**Δ**	***p*** $_{Pre/Post}$	Δ ANOVA
Female, Male			0, 6				3, 34				26, 1				33, 2		
Gait speed (m/s)	1.1 (0.1)	1.2 (0.1)	0.1 (0.1)	0.02	0.8 (0.2)	1.0 (0.2)	0.2 (0.2)	**<0.0001**	0.7 (0.1)	1.0 (0.2)	0.3 (0.2)	**<0.001**	1.1 (0.1)	1.1 (0.2)	0.1 (0.1)	0.005	**<0.001**
Knee Flexion Angles																	
PC1	28.7(35.4)	57.1(43.7)	28.3(47.2)	0.2	-6.5(56.4)	15.9(45.4)	22.4(43.9)	0.004	-75.9(63.3)	15.8(50.2)	91.7(71.8)	**<0.001**	-2.5(58.5)	22.2(64.7)	24.7(89.2)	0.1	**0.001**
PC2	22.8(6.1)	-7.6(30.9)	-30.4(28.8)	0.05	3.2(27.7)	11.3(25.3)	8.1(34.2)	0.2	-20.2(35.8)	13.1(30.1)	33.4(36.3)	**<0.001**	0.1(34.5)	-2.8(28.1)	-2.9(40.6)	0.7	**<0.001**
PC3	10.1(34.3)	20.0(22.7)	9.9(15.4)	0.2	-10.7(28.1)	4.0(27.2)	14.7(27.6)	0.003	-11.1(30.2)	-10.8(22.0)	0.3(28.9)	0.9	3.9(18.0)	3.3(22.7)	-0.5(23.4)	0.9	0.06
PC4	21.9(13.7)	14.1(9.8)	-7.8(8.6)	0.08	2.9(11.3)	5.0(15.0)	2.2(16.2)	0.4	-4.8(8.6)	-2.1(11.9)	2.7(11.6)	0.2	-0.2(13.0)	-0.4(10.8)	-0.2(16.1)	0.9	0.4
Knee Adduction Angles																	
PC1	1.0(10.5)	-13.1(20.3)	-14.1(16.6)	0.09	1.1(21.1)	-10.7(17.9)	-11.9(24.6)	0.006	1.1(26.3)	-7.7(25.5)	-8.8(34.8)	0.2	9.1(20.0)	-6.2(17.0)	-15.3(23.2)	**<0.001**	0.8
Knee Flexion Moments																	
PC1	-1.8(0.9)	-0.7(1.3)	1.1(0.9)	0.04	-0.6(1.8)	-0.2(1.6)	0.4(1.9)	0.3	1.3(1.8)	-0.3(1.3)	-1.6(1.6)	**<0.001**	0.3(1.7)	0.0(1.1)	-0.3(1.7)	0.3	**<0.001**
PC2	1.1(0.5)	1.4(0.4)	0.3(0.5)	0.1	-0.3(0.5)	0.4(0.8)	0.6(0.8)	**<0.001**	-0.8(0.4)	-0.1(0.6)	0.7(0.7)	**<0.001**	0.0(0.6)	0.2(0.5)	0.1(0.7)	0.2	0.009
PC3	0.4(0.5)	0.1(0.5)	-0.2(0.3)	0.2	0.0(0.4)	0.0(0.3)	0.0(0.4)	0.9	0.0(0.4)	-0.1(0.3)	0.0(0.4)	0.5	0.1(0.5)	0.0(0.3)	-0.1(0.4)	0.3	0.7
Knee Adduction Moments																	
PC1	2.2(1.3)	0.9(0.6)	-1.3(0.9)	0.02	0.9(1.0)	0.1(0.9)	-0.8(1.1)	**<0.001**	-0.5(2.1)	-0.6(0.9)	-0.1(1.8)	0.9	-0.2(1.7)	-0.1(0.9)	0.2(1.6)	0.5	0.008
PC2	0.4(0.5)	0.2(0.5)	-0.2(0.3)	0.3	-0.3(0.3)	0.0(0.4)	0.3(0.4)	**0.001**	-0.4(0.3)	0.0(0.3)	0.4(0.3)	**<0.001**	0.1(0.3)	0.2(0.4)	0.1(0.4)	0.07	0.01
PC3	0.2(0.3)	0.5(0.3)	0.2(0.3)	0.09	-0.1(0.3)	0.1(0.3)	0.2(0.2)	**<0.001**	-0.1(0.3)	0.0(0.2)	0.1(0.2)	0.01	0.0(0.3)	0.1(0.3)	0.1(0.3)	0.005	0.3

Tukey HSD results:
Gait Speed: 4<2,3
Knee Flexion Angle PC1: 2,4<3
Knee Flexion Angle PC2: 1,4,2<3
Knee Flexion Moment PC1: 3<4,2,1

5.4. Discussion

Severe knee OA TKA candidates are heterogenous with respect to demographics and knee joint kinematics and kinetics during gait, and can be characterized by four clusters. Although previous studies have provided valuable evidence of *a priori* variability among TKA candidates (typically using bi-variate stratification), this is the first study to provide indications of multivariable patient profiles, and insights into feature combinations that represent inherent similarities and natural heterogenous subsets among typical TKA candidates, incorporating both patient demographics and knee joint-level biomechanics.

Patient sex, in combination with knee kinematic and kinetic gait features were dominant in group separation, resulting in phenotypes of mostly male or female clusters. Few studies have addressed sex difference among TKA patients. Both Astephen Wilson et al. (Wilson et al., 2015) and Paterson et al. (Paterson et al., 2017) reported greater magnitudes of frontal plane adduction moments among men pre-TKA. Female TKA candidates have also been shown to present with less biphasic sagittal stance moments, greater flexion moment magnitudes, and less range of flexion and extension motion, descriptive of a stiffer gait with a more constant sagittal loading pattern (Wilson et al., 2015). Our findings support this, such that both male clusters (1 and 2) had higher adduction moment magnitudes (PC1, 3,4<2,1), and females tended to have higher flexion moment loading magnitudes (PC1, 1<4,3; 2<3). However, our results also identified a sizable, lower functioning male phenotype (cluster 2), who had less dynamic biphasic flexion moments (PC2, 3<2<4<1) than the higher functioning females (cluster 4). As our lower functional females (cluster 3) exhibited distinctly more “severe” (Astephen et al., 2008a) mechanics (Figure 5.1), mean sagittal plane kinetics of both our female groups combined would be consistent with less asymptomatic patterns than males. Therefore, less biphasic flexion-extension moments may not characterize pre-TKA females in general. It was interesting that a lower functioning female cluster did exist, exhibiting notably “more severe” knee joint kinematics and kinetics mechanics than any other cluster. Conversely, the lower functioning male cluster was most similar to the higher functioning female cluster. Females in an older adult population are predisposed to the

development of OA (Buckwalter and Lappin, 2000), have a greater incidence of radiographic and symptomatic OA severity (Felson et al., 1987; Srikanth et al., 2005), have been associated with worse self-reported disability (King et al., 2018), and tend to have less encouragement to undergo TKA (Borkhoff et al., 2009). These factors coupled with delayed interventions may contribute to the slower gait speed (Astephen et al., 2008a), and more severe biomechanical presentation amongst female OA patients in this study, despite trending towards being younger (Table 5.1). Differences could further be associated with sex-specific joint morphologies (Mahfouz et al., 2011), perhaps influencing joint-level structural damage manifestations, and sex-specific neuromuscular control (Wilson et al., 2015). Unsupervised separation strategies enabled us to identify more complex combinations of patient sex and gait interactions inherent within the TKA candidate population, better complimenting individual variability, and aiding our ability to provide targeted treatment and management for all OA individuals (Bannuru et al., 2019; Bierma-Zeinstra and Verhagen, 2011). Further, biomechanical assessment and cluster assignment may aid in the development of patient prioritization practices that are independent of possible gender bias, enabling timely interventions for all patients.

Gait improvements post-TKA were cluster specific. Both male clusters had greater adduction moment magnitudes pre-TKA (PC1, 3,4<2,1), and post-TKA reductions in adduction moment magnitude were significant in the lower functioning male cluster ($p<0.001$), with trends towards reductions among higher functioning males ($p=0.02$), Table 5.2. No adduction moment reductions occurred among either female clusters ($p>0.5$), corroborating prior sex-specific findings (Wilson et al., 2015). Similarly, of the female clusters presenting with greater flexion moment loading (PC1, 1<4,3), reductions were only significant for the lower functioning female cluster ($p<0.001$), and not for the higher functioning female, or male cluster (Wilson et al., 2015). In general, the lower functioning female cluster, followed by the lower function male cluster, demonstrated the most improvement in knee joint kinematics and kinetics from arthroplasty relative to their higher functioning sex-matched clusters (Table 5.2). These improvements are typically expected after TKA, moving waveforms toward asymptomatic norms (Hatfield et al., 2011) (Figures 5.1-5.2). Conversely, higher functioning females (cluster 4) only

demonstrated decreased adduction angle magnitudes during stance (PC1). Among TKA candidates with higher static and dynamic adduction angles pre-TKA, reductions in this feature have been associated with worse post-operative self-reported pain, function (Vanlommel et al., 2013; Young-Shand et al., 2020) and quality of life scores (Naili et al., 2017). High functioning males (cluster 1) showed no improvements among any gait feature pre to post-TKA ($p \geq 0.02$). Although this group was small, limiting the statistical power to detect changes, it is notable that some features trended away from asymptomatic patterns (i.e., worsened) pre to post-TKA, such as decreases in stance to swing flexion angle range ($p=0.05$) (Table 5.2). Overall, these higher functioning clusters lagged in terms of knee kinematic and kinetic improvements seen among the lower functioning clusters from arthroplasty (Table 5.2).

In the patient-reported outcome literature, patients with worse self-reported symptoms preoperatively have been well characterized to experience the largest relative improvements in self-reported scores post-operatively (Hawker et al., 2013; Jiang et al., 2017). In gait studies, both Naili et al. (Naili et al., 2017) and Young-Shand et al. (Young-Shand et al., 2020) both stratified TKA patients based on self-reported improvements after TKA. Cohorts with more "severe" gait patterns pre-TKA tended to experience greater degrees of biomechanical improvement and self-reported improvement from arthroplasty. This is perhaps because patients with worse gait mechanics pre-operatively have more capacity for objective improvements from surgery, and are less likely to experience an improvement ceiling effect (Paul Robert Biggs et al., 2019; Outerleys et al., 2021). The findings of this study support this theory, but further suggest the presence of a small (6/105, 6%) higher functioning male phenotype who may not experience any functional benefit from standard TKA, and a sizable (35/105, 33%) higher functioning female cohort whose biomechanical gain from TKA was negligible to small (Figures 5.1-5.2). A study by Knoop et al. (Knoop et al., 2011) used cluster analysis to identify subsets of OA patients. They characterized a "minimal joint disease phenotype", a cohort of patients that were less symptomatic and may not have needed specific interventions to manage OA disease and symptoms. This "minimal joint disease phenotype" might coincide with our identified higher functioning male and female

clusters, despite not identifying OA radiographic severity levels differences between groups (Table 5.1). Knee mechanic characterization of phenotypes may enable us to prioritize patients who may most benefit from TKA. They may also identify those who might benefit from altered non-invasive clinical (Allen et al., 2016; Skou et al., 2015) or more conservative surgical approaches (such as unicompartment or biocompartment) as these techniques improve with robotic surgeries. Understanding of clusters might also provide valuable information around expectation management at a cohort level. Further, understanding the three-dimensional biomechanical corrections not addressed by standard arthroplasty among certain patient cohorts can inform surgical methodologies, prosthesis innovations, or rehabilitation strategies to correct cluster-specific deficits.

A limitation of our study was the lack of information on self-reported patient scores (such as pain), which would have improved our ability to further characterize our subgroups in terms of illness or symptom presentation (Andriacchi et al., 2014; Bannuru et al., 2019; Castaneda et al., 2013; Lane et al., 2011). Prior work has found less severe symptoms to be weakly associated with less severe gait features (Thorp et al., 2007; Wilson et al., 2017), therefore we might expect the higher functioning cohorts to be less symptomatic. Designing future biomechanical studies specifically for phenotyping analysis can help to ensure comprehensive feature collection (Spil et al., 2020). Specifically, future work on larger samples should examine biomechanical variability in combination with attributes of self-reported scores pre and post-operatively, coupled with radiographic severity, static alignment, biomarkers, cartilage wear patterns, and mechanics at other joints (Paul Robert Biggs et al., 2019; Dell'Isola and Steultjens, 2018; Knoop et al., 2011; Waarsing et al., 2015), to enable a broader look at disease interactions of OA, and avoid research-area-specific phenotypes (Andriacchi et al., 2014). It might also be valuable to include a longitudinal sample of OA observations, to better characterize phenotypes associated with function along the disease severity continuum. Moderate to low cluster separation metrics by silhouette coefficient (s=0.37) was another limitation of our study, such that the presented clusters cannot be interpreted as truly distinct groups. This is unlikely a reflection of the data quality, but the continuum of gait patterns that patients demonstrate during the OA disease severity process (Astephen et

al., 2008a), and previously characterized interaction effects of symptoms such as pain (Thorp et al., 2007; Wilson et al., 2017), and sex (Paterson et al., 2017; Wilson et al., 2015). Further work including the additional proposed attributes may include attributes lending to better separation. Another potential limitation was the inclusion of sex, a categorical feature and dominant in cluster characterization. Clustering algorithms are dependent on distance metrics, and there is no true "distance" between binary representation of male and female sex. Transforming all clustering features to z-scores, and conducting MDS prior to clustering ensured even feature weighting during unsupervised learning. Still, clustering methodologies were repeated without sex in an exploratory analysis. These waveforms are provided in Appendix B (Figure B.2-B.3), and were quite similar to those presented here, particularly in the sagittal planes. Without including sex (Figure B.2), the highest functioning group (cluster 1) again had the largest proportion of males (8:5), and the lowest functioning group (cluster 4) was mostly female (16:9). When removing sex, clusters were best characterized demographically by age and BMI, but not sex ($p=0.4$). The importance of sex on our study results, with only artifacts of sex being significant after this features' removal limits the generalization of our findings, however, results are no less valid than a sex-stratified study design. It needs to be determined which classification method (including or excluding sex), is most relevant clinically and to patient outcomes as we look into applications to translate findings to general clinical practice.

Phenotyping investigations consistently propose the existence of biomechanical-driven OA subtypes (Andriacchi et al., 2014; Bannuru et al., 2019; Castaneda et al., 2013; Waarsing et al., 2015). This was the first study to characterize, in an unsupervised manner, the inherent variability among TKA candidates using biomechanical and demographic features. Four biomechanical and sex-specific clusters were characterized. Post-TKA, functional gains were cluster-specific; lower functioning clusters experienced more improvement, and results suggest the presence of cohorts who may not benefit in terms of joint-level biomechanics from TKA. These findings provide valuable information for patient triaging, expectation management, or aiding in the identification of gait deficits that standard arthroplasty is less likely to address at a cohort level.

Building on this structure, cluster profiling applied to larger longitudinal datasets may aid in developing patient-specific surgical and non-surgical OA management approaches to meet individual function needs.

Chapter 6. Assessing OA Severity and Biomechanical Changes after TKA using Self-Organizing Maps

6.1. Introduction

Knee osteoarthritis (OA) is a complex degenerative disease with marked heterogeneity, increasingly being characterized in terms of phenotypes (Andriacchi et al., 2014; Castaneda et al., 2013; Deveza et al., 2017). Proposed phenotypes often include biomechanical or structural components deemed to emerge in both OA disease initiation and manifestation processes (Andriacchi et al., 2014; Bannuru et al., 2019; Bruyère et al., 2015; Castaneda et al., 2013; Deveza et al., 2017). A growing number of studies have applied epidemiological or machine learning based unsupervised partitioning (i.e., clustering) to OA populations, to discover phenotypes that characterize biomechanically-related variability associated with symptoms, muscle strength, obesity, radiographic severity, cartilage wear, knee alignment, or spatiotemporal features (Dell'Isola and Steultjens, 2018; Elbaz et al., 2014; Esch et al., 2015; Waarsing et al., 2015). This characterization of biomechanical OA variability is relevant to determining appropriate intervention strategies, as care guidelines propose treatments personalized to biomechanics profiles (Allen et al., 2016; Bruyère et al., 2015).

OA variability in terms of phenotypes are thought to evolve over the course of the disease process (Castaneda et al., 2013), with end stage joint degradation converging along a common pathway for all patients (Felson, 2010). To date, phenotypes have only been derived from cross-sectional datasets, captured at various static time points within the OA progression pathway, defined using variable severity criteria (Deveza et al., 2017; Knoop et al., 2011; Waarsing et al., 2015). These temporal snapshots lack insight into phenotypes that span the longitudinal OA disease progression process, thus limiting our ability to understand if phenotype-specific progression pathways exist. We argue that as we aim to propose phenotype-specific prevention and intervention strategies, a longitudinal view of the OA progression pathway is required to identify OA phenotypes that impact clinical utility and personalized treatments (Felson, 2010).

This study aimed to quantify OA profiles using a novel unsupervised machine learning framework, and visually map variability in demographic and knee joint kinematic and kinetics during gait using self-organized maps (SOM). The map encompassed changing gait patterns along the spectrum of knee OA clinical pathways: asymptomatic, moderate OA, severe OA (pre-TKA), and post-TKA. A decision tree characterizing patient cluster classification was used in SOM region and progression pathway interpretation. Mapped regions and clusters were hypothesized to demonstrate knee biomechanics variability associated with disease severity, and provide evidence of OA phenotypes traveling a multitude of patient progression pathways relevant to individual treatment strategies.

6.2. Methods

6.2.1. Dataset and feature selection for clustering

This is a secondary study using knee joint kinematic and kinetic gait laboratory data collected between 2001-2018 (n=945 gait session/observations; j=502 knees; Figure 6.1; Appendix C Table C.1). Participants belonged to six OA clinical groups: i) asymptomatic adults with no musculoskeletal injury, disease, or recent surgeries; ii) moderate knee OA patients diagnosed clinically according to the American College of Rheumatology criteria and not deemed TKA candidates (Altman et al., 1986); iii) severe knee OA patients, seen one-month pre-TKA; iv) TKA recipients, seen one and/or two years post-TKA; v) TKA revision patients seen one month pre-revision; vi) and one and/or two years post-revision.

As part of the gait protocol, each participant walked along a five-meter walkway wearing comfortable walking shoes at a self-selected speed. Lower-limb external ground reaction forces and kinematics were captured using a force platform sampling at 2000 Hz, synchronized to an optoelectronic motion capture system sampling at 100 Hz. Knee joint angles during gait were calculated using the joint coordinate system (Grood and Suntay, 1983). Net resultant knee joint moments also represented in the joint coordinate system were calculated using inverse dynamics (Costigan et al., 1992; Deluzio et al., 1993; Li et al., 1993), and amplitude normalized to body mass. Principal Component Analysis (PCA)

was applied to frontal and sagittal plane angle and moment waveforms, resulting in a new uncorrelated dataset of PC scores and PC loading vectors describing the major modes of variability throughout the gait cycle (Deluzio and Astephen, 2007). Five flexion angle, adduction angle, flexion moment, and adduction moment PCs were retained to capture ≥ 95% of the variance explained for each waveform (Jackson, 1993) (Appendix C Figures C.1-C.5).

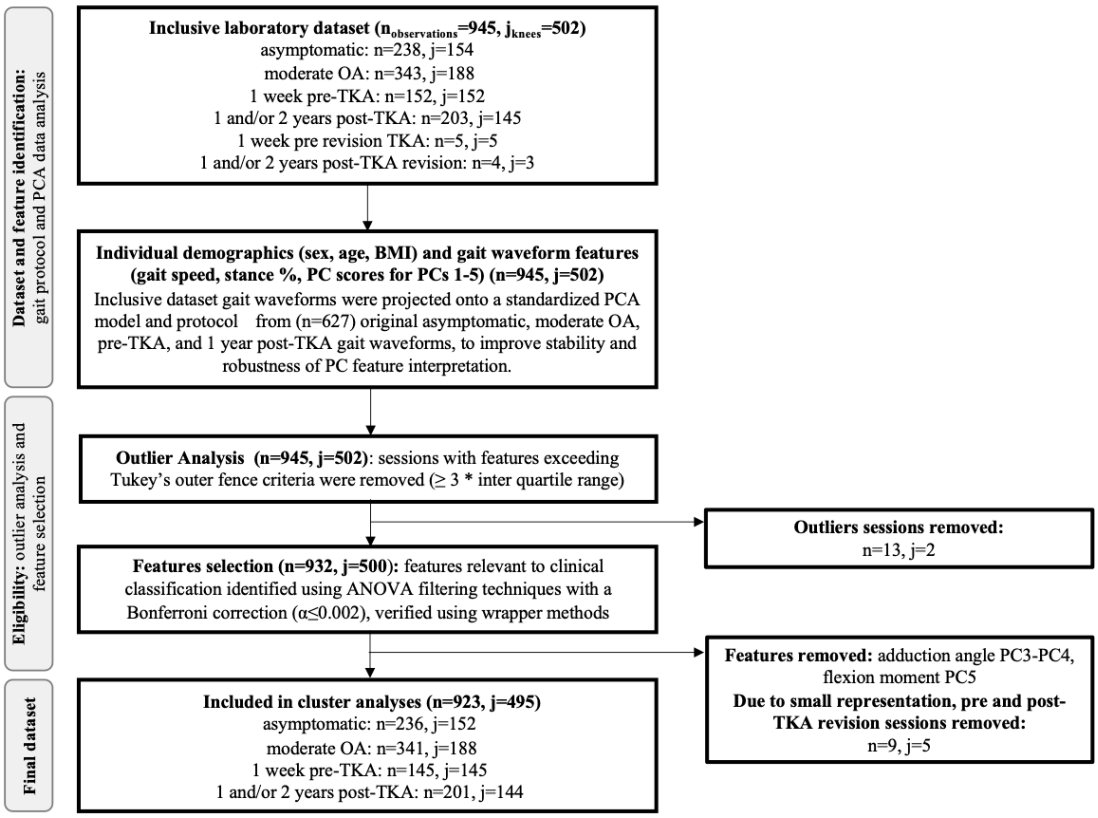

Figure 6.1. CONSORT diagram of patient eligibility and selection processes. All participants were screened for previous lower-extremity surgery (e.g., arthroplasty in another joint) as well as neurological and other existing pathological conditions (e.g., rheumatoid arthritis) prior to recruitment for the gait study.

Kinematic and kinetic gait PC scores, participant age, sex, body mass index (BMI), gait speed, and percent of gait cycle in stance from each session comprised our initial study dataset ([945x25]; 25 features). Distance measures applied during clustering can be sensitive to magnitude and scale differences of input features, and outliers (Milligan and Cooper, 1988). The dataset was therefore standardized to z-scores (mean=0, standard

deviation=1), and 13 outlier observations exceeding Tukey's outer fence (3*IQR) were removed ([932x25]) (Milligan and Cooper, 1988). To achieve a non-redundant and parsimonious feature set, redundancy was assessed using Pearson's Correlation Coefficients (no features removed; maximum r=0.71). Feature relevance was assessed in terms of clinical group separation by ANOVA f statistic feature ranking (Boyd and Little, 2005; Guo and Nixon, 2009) using a Bonferroni correction, accepting observations at the 0.002 significance level ($\alpha/25$). Three weakly relevant features (p>0.002) were removed (adduction angles PC3-4, flexion moment PC5). Consistent with prior studies (Elbaz et al., 2014; Knoop et al., 2011; Waarsing et al., 2015), patient sex was also not included in cluster analysis, yet features of sex were expected to present in gait waveforms (Paterson et al., 2017; Wilson et al., 2015). Finally, due to small sample representation relative to the other clinical groups (n=9), pre and post-TKA revision sessions were removed. The final dataset included 923 gait observations and 21 features ([923x21]; j=495; Table 6.1), divided into training (f=0.95, [878x21]; j=484), and test sets (f=0.05, [45x21], j=45). This maximized the size of our training data to achieve good clustering output, whilst maintaining the ratio of clinically classified patients in the test data.

6.2.2. Clustering

Self-Organizing Map (SOM)

An unsupervised machine learning Self-Organizing Map (SOM) (Kohonen, 1990, 1982) framework was applied to gain insights into the natural organization of OA gait biomechanics. SOMs are artificial neural networks that project high-dimensional data onto a connected (typically two-dimensional) network of nodes, thus providing an interactive visualization of the emergent clusters. Each node is represented by a weighted vector equation, enabling input data to be mapped to the lower-dimensional SOM space (Kohonen, 1990, 1982). SOMs are similarity graphs and cluster diagrams, where similar features in the input feature space remain spatially proximal in the lower-dimensional mapped space.

The SOM was initialized with an 11x11 (i =121 node) feature space with hexagonal nodes to achieve 5-10 observations per node (Westerlund, 2005). Each node was defined by the equation, $m_i=[v_1,v_2,...,v_{21}]$, where i denotes the index on the 121-node SOM, and v denotes vector weights for the 21 clustering features. Nodes are initialized by randomly assigning an observation from the training dataset to each node. The SOM methodology then followed a recursive, stepwise learning process (Kohonen, 2013, 1990, 1982). First, a training observation is projected onto the SOM. A winning node is defined, satisfying the minimum Euclidean distance between the training observation and each node's vector equation (i.e., "competitive learning"). The vector weights, v, of the winning node and its neighboring nodes are then adjusted towards the input training observation. The magnitude of this adjustment is a function of the distance to the winning node, and a specified learning rate, l. The learning rate parameter decreased linearly to 0 over the learning process, with initial learning rates l=0.06:1 tested in this analysis. This learning process is then repeated, such that each observation is presented to the SOM iteratively until node weights converge or based on a specified training time, t, the number of times the training set is presented to the SOM. Here, training times t=1000:3000 were tested, defined to approximate 10 times the number of mapped nodes (121*10=1210) (Westerlund, 2005). SOMs were generated for each learning rate and training time condition (Wehrens and Buydens, 2007). Under these conditions, ten SOM models were retained that minimized the Euclidean norm difference between each observation input vector and their assigned node vector, also termed quantization error, qe (Kohonen, 1990; Kohonen et al., 2009) (Appendix C Figure C.6).

Hierarchical Clustering of the SOM and Statistical Analysis

Each of the 10 retained SOMs were applied to hierarchical (bottom-up) clustering using Ward's minimum variance criteria (Clatworthy et al., 2005; Ward, 1963), minimizing total within-cluster variance; the most commonly used approach in the OA literature (Deveza et al., 2017). When applied to a SOM, clustering is conducted on the weighted vector equation values (m_i) of each node (Kohonen, 1990). Clustering solutions with number of clusters k=2:10 were assessed. The quality of each cluster was determined using silhouette width criterion (Rousseeuw, 1987). Silhouette width, s,

measured how similar node i was to its own cluster (cohesion) compared to other clusters (separation) bound by ± 1 (-1[misclassified]-1[well-clustered]). The cluster model with the greatest count of positive s coefficients among nodes was selected. This selection approach maximized for the number of nodes better represented by their own cluster over a neighbouring one, and avoided representing some nodes very well, and others poorly within each cluster (Appendix C Table C.3).

Final clusters were validated by examining inter-cluster differences among features highly relevant during feature ranking, and found to characterize OA clinical severity (Astephen et al., 2008a, 2008b), symptoms (Wilson et al., 2017), self-reported outcomes (Young-Shand et al., 2020), and sex (Wilson et al., 2015). Chi-squared (sex), k-way ANOVA (PC scores, gait speed, percent stance, age, BMI), Kruskal Wallace tests (OA clinical classification) were used. Post-hoc tests were conducted using Tukey's HSD criterion for parametric features, and pair-wise Pearson's chi-squared tests for categorical (nominal) features. Bonferroni corrections adjusted for multiple comparisons, with p-values ≤ 0.002 accepted.

6.2.3. Decision Rule

Classification and regression trees (CART) (Breiman et al., 1984) were used to aid in SOM region interpretation. CART is an supervised modeling framework that predicts a target class by building a decision tree ruleset from input observations. Rulesets predicting which cluster each observation was most likely to belong to were derived from a second training set, composed of 80% of our original training data (n=703, j=438). Tree pruning (removal of decision nodes and branches) simplifies rulesets using only the most influential features, while preventing over-fitting. Ten-fold cross validation with 10 repeats using a complexity parameter (*cp*), a penalty based on the number of nodes was used to prune the tree (Trost et al., 2016). Accuracy was optimized when the *cp*=0.01. Cluster classification performance was assessed on a secondary withheld test set from our original training data (f=0.20, n=175, j=159).

6.2.4. Outcomes by Cluster

One hundred and two patients had matched gait data available pre and one-year post-TKA. Student's t-tests and chi-squared tests were used to compare baseline demographics and gait changes ($PCscore_{postTKA} - PCscore_{preTKA}$) between defined clusters. All analyses were conducted using R (R Foundation for Statistical Computing, Vienna, Austria).

6.3. Results

6.3.1. SOM and Hierarchical Clustering

The final SOM model is presented in Figure 6.2. Each node represents a different combination of knee biomechanics and demographics features, with similar feature combinations mapping proximal to each other. After applying hierarchical clustering, a three cluster (k=3) model was selected, having the least number of negative nodes (5/121) by silhouette coefficient (model s=0.16; Appendix C Figures C.8). The following provides an observation-based interpretation of each clusters.

Cluster 1: 'High Knee Function' contained the largest proportion of the training data's asymptomatic (86%, 193/225) and moderate OA (45%, 147/324) observations. Observations in this cluster were from younger participants (53.8±9.9 years; 1<2,3; p<0.001), with the lowest BMIs (27.9±4.7 kg/m2; 1<3<2; p<0.001), and fastest self-selected walking speeds (1.4±0.2 m/s; 2<3<1; p<0.001). This cluster also spent the least amount of their gait cycle in stance (63.4±0.2 %; 1<3<2; p<0.001). In the sagittal plane, this cluster walked with greater overall knee flexion angle magnitudes throughout the gait cycle (PC1, 2<3<1, p<0.001), with more flexion angle range of motion (PC4, 3<2<1, p<0.001) and more biphasic flexion/extension loading/un-loading moments during stance (PC2, 2,3<1, p<0.001). In the frontal plane, they had higher overall stance-phase adduction angle magnitudes (more varus, PC1, 2,3<1, p<0.001) and more dynamic frontal plane un-loading/loading range patterns (first peak to mid-stance PC2, 2<3<1, p<0.001; mid-stance to second peak PC3, 2<3<1, p<0.001) (Figure 6.3, Tables 6.1-6.2).

Cluster 2: 'Low Knee Function' differed the most from Cluster 1. It contained the greatest proportion of severe knee OA (78%, 108/138) and post-TKA (58%, 111/191) participants. This cluster had a greater ratio of female observations relative to the first cluster (56% vs. 39%, $p<0.001$), with the greatest mean BMI (34.4±6.2kg/m2), and slowest walking speeds (1.0±0.2 m/s) among the clusters. Gait observations in this cluster spent a greater amount of the gait cycle in stance (66.9±2.1%), walked with the greatest knee flexion moment loading magnitudes overall (PC1; 2<1,3; $p<0.001$), the least knee flexion angle magnitudes (PC1; 2<3<1; $p<0.001$), least knee extension at heal strike and late stance (PC3; 2<1,3; $p<0.001$), and lowest flexion (PC2; 2,3<1; $p<0.001$) and adduction moment (PC2-PC3; 2<3<1; $p<0.001$) loading/un-loading ranges during stance (Figure 6.3, Tables 6.1-6.2).

Table 6.1. Demographic and spaciotemporal features of clusters. Three-way ANOVA and pair-wise Chi-squared (sex) post-hoc tests examined differences between clusters.

	1: High Knee Function		2: Low Knee Function		3: Moderate Knee Function		*p*	Post-hoc
Training Dataset (n=878)								
Clinical Group								
Asymptomatic (n/225, %)	193	86%	7	3%	25	11%		
Moderate (n/324, %)	147	45%	66	20%	111	34%		
Severe/Pre-TKA (n/138, %)	8	6%	108	78%	22	16%		
Post-TKA (n/191, %)	31	16%	111	58%	49	26%	**<0.001**	
Sex								
Female (n, %)	149	39%	162	56%	119	58%		1 vs 2 p<0.001
Male (n, %)	230	61%	130	45%	88	43%	**<0.001**	1 vs 3 p<0.001
Age (years, SD)	53.8	(9.9)	63.9	(8.7)	63.0	(8.0)	**<0.001**	1<2,3
BMI (kg/m^2, SD)	27.9	(4.7)	34.4	(6.2)	30.7	(4.6)	**<0.001**	1<3<2
Stance Percent (%, SD)	63.4	(1.5)	66.9	(2.1)	64.7	(1.5)	**<0.001**	1<3<2
Speed (m/s, SD)	1.4	(0.2)	1.0	(0.2)	1.2	(0.1)	**<0.001**	2<3<1
Test Dataset (n=45)								
Clinical Group								
Asymptomatic (n, %)	9	43%	0	0%	2	17%		
Moderate (n, %)	11	52%	2	17%	4	33%		
Severe/Pre-TKA (n, %)	0	0%	6	50%	1	8%		
Post-TKA (n, %)	1	5%	4	33%	5	42%	**<0.001**	

Cluster 3: 'Moderate Knee Function' was spatially between the first two clusters visually in the SOM. It contained the largest proportion of moderate OA (34%, 111/324), and the second greatest representation of post-TKA (26%, 49/191) observations. This cluster also had a higher ratio of female observations (58% vs. 39%, $p<0.001$) than the first cluster, and mean BMIs (30.7±4.6 kg/m2), gait durations in stance (64.7±1.5%), and walking speeds (1.2±0.1 m/s) between the first two clusters. Features of knee kinematics and kinetics during gait in frontal and sagittal planes were between Clusters 1 and 2. The only exception to this being stance-phase flexion angle range, which was lower in this intermediate group than Clusters 1 and 2 (PC4, 3<2<1, $p<0.001$) (Figure 6.3, Tables 6.1-6.2).

Mapping the test set (n=45) onto the final model, most asymptomatic individuals (9/11) mapped to 'High Knee Function', and most severe OA/pre-TKA individuals (6/7) mapped to 'Low Knee Function', aiding in interpretation validation (Table 6.1).

From cluster and node-level interpretation (Abidi et al., 2018), progression of knee biomechanics severity (i.e., worsening) and OA clinical severity could generally be represented from bottom left (mostly asymptomatic, Cluster 1) to top right (mostly pre or post-TKA, Cluster 2), across the x-axis of the SOM, with age and BMI increasing, and gait speed decreasing stepwise across clusters (Table 6.1). Knee biomechanics during gait also worsened from bottom left to top right along the SOM, captured by decreases in knee flexion angle magnitudes (PC1; 2<3<1; $p<0.001$), flexion angle range (PC4, 3<2<1, $p<0.001$) adduction moment loading/un-loading range (PC2-PC3; 2<3<1; $p<0.001$) across clusters, and node-level interpretation of flexion and adduction moment PC2 (Appendix C Figure C.7).

To demonstrate the utility of the SOM in characterizing OA phenotypes and progression pathways, the mapped SOM locations of six participants' gait observations longitudinally have been illustrated in Figure 6.2. These cases support a left-to-right progression pattern over time, but provide evidence of variable two-dimensional pathways during moderate OA progression (Figure 6.2b,d) and during post-TKA recovery (Figure 6.2e,f).

6.3.2. Decision Rule

Cluster classification applied to CART using 10-fold cross-validation resulted in a tree-based decision rule (Figure 6.4). Overall accuracy applied to a withheld test set was 76.6% 95% CI [70.0, 82.6], with the best classification performance achieved in Clusters 1 and 2 (Table 6.3)

6.3.3. Outcomes by Cluster

Due to low representation of individuals classified in the 'High Knee Function' cluster pre-TKA (n=4), this high function cluster, and the intermediate 'Moderate Knee Function' cluster (n=15) were merged. Pre to post-TKA gait feature changes between this merged subset, and the 'Low Knee Function' cluster pre-TKA (n=83) were compared.

TKA patients pre-operatively classified to the more severe 'Low Knee Function' cluster experienced greater improvements in objective function relative to the other clusters. For example, they experienced larger increases in gait speed ($p=0.01$), knee extension angle magnitudes (PC3, $p<0.001$), and frontal plane loading and un-loading moment range (PC2, $p=0.02$), with more decreases in adduction moment loading magnitudes (PC1, $p=0.01$). These changes move gait patterns towards more asymptomatic gait patterns (Astephen et al., 2008a). Alternatively, the merged 'High Knee Function' and 'Moderate Knee Function' clusters tended to show a decline in mean knee extension angle magnitudes (PC3), and an overall increase in adduction moment magnitude (PC1) during stance; changes that move gait patterns further from asymptomatic post operatively (Hatfield et al., 2011), Table 6.4.

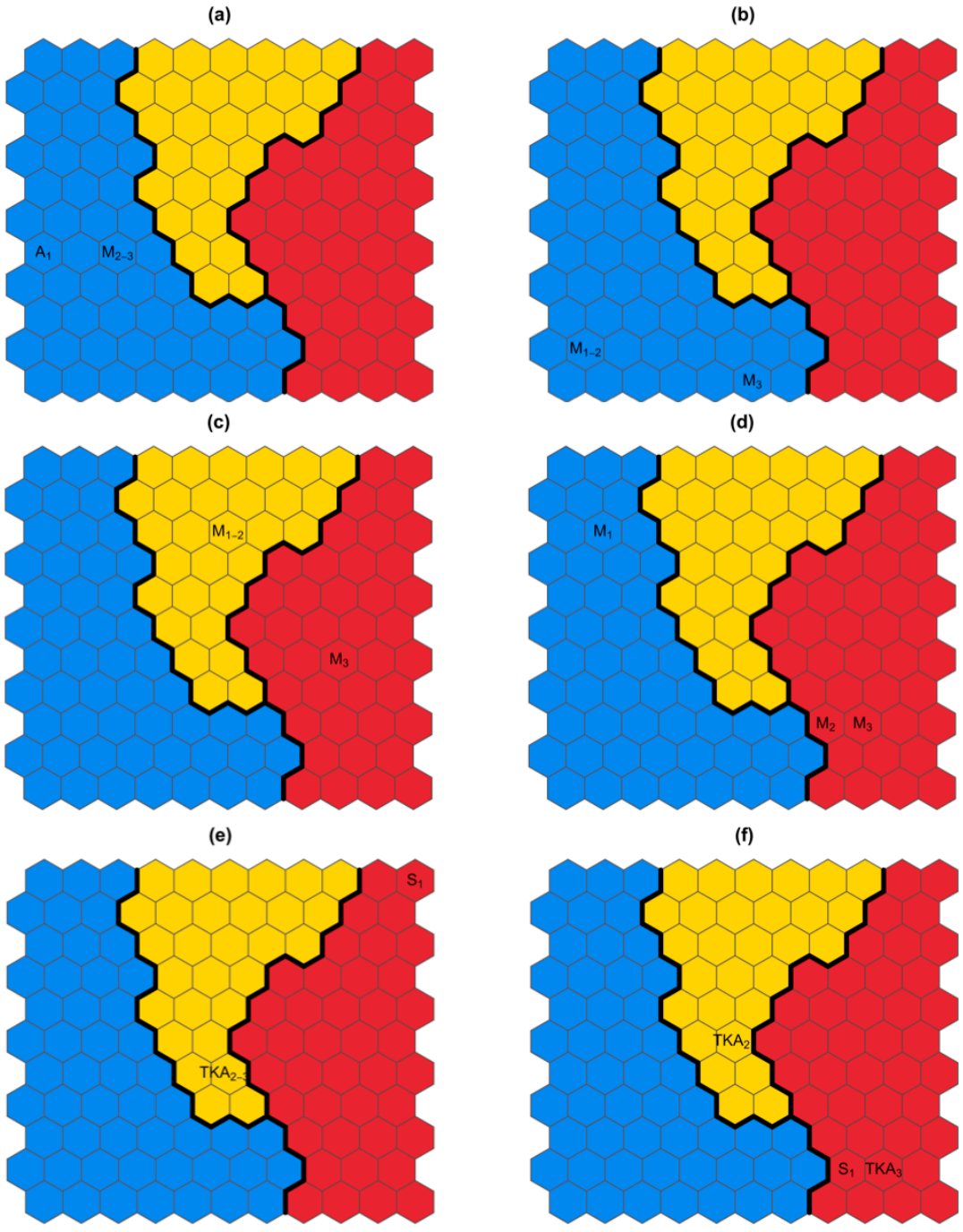

Figure 6.2. Final SOM Model (11x11 grid), depicting cluster 1 (High Knee Function, blue), cluster 2 (Low Knee Function, red), and cluster 3 (Moderate Knee Function, yellow), generally describing OA progression from left (highest functioning) to right (lowest functioning) across the SOM. Example pathways of six individual patient pathways are shown in (a)-(f), described by A (asymptomatic), M (moderate), S (severe OA/pre-TKA) and TKA (post-TKA) clinical classifications. Subscripts represent the order of individual observations to provide longitudinal context.

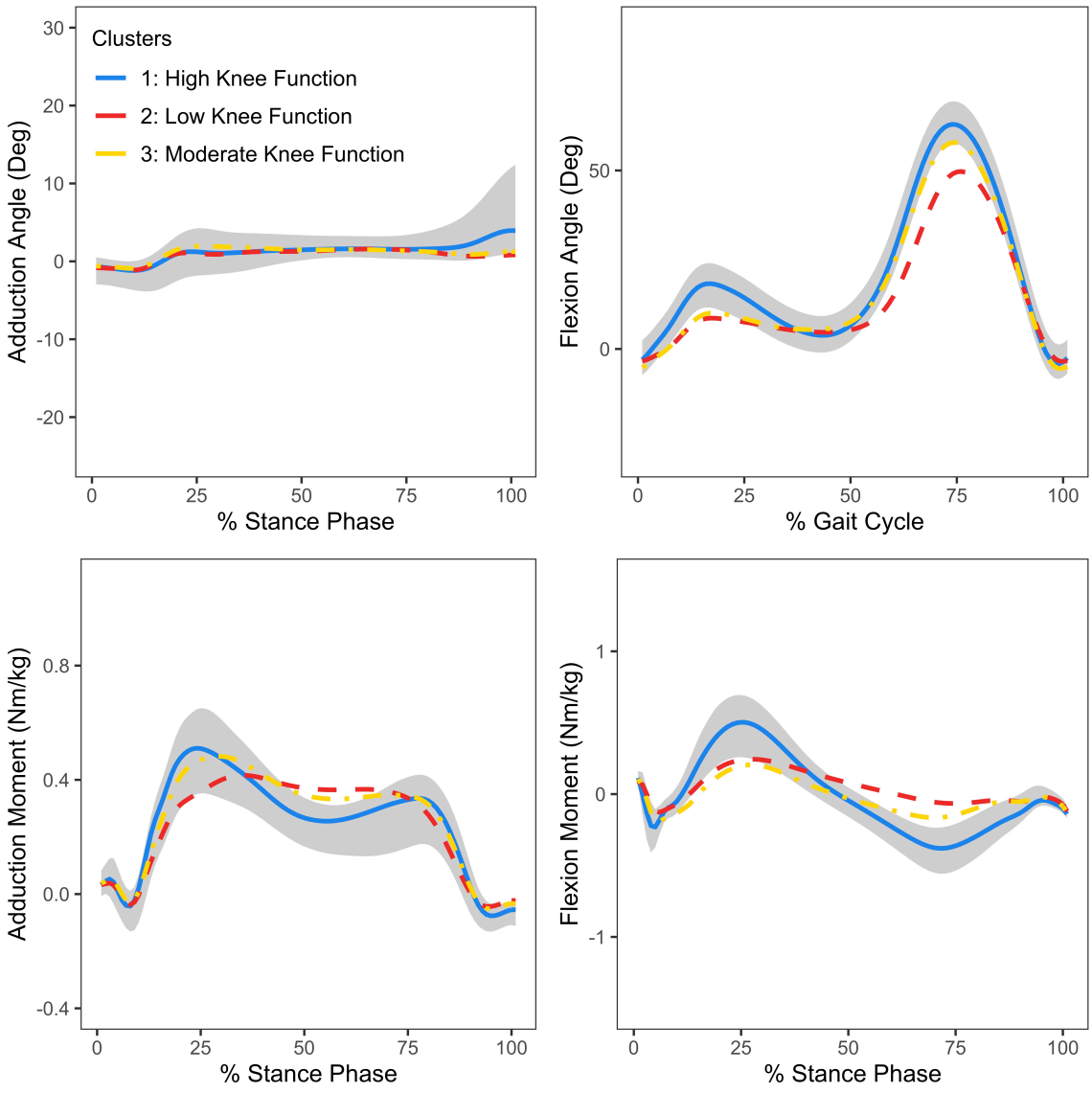

Figure 6.3. Mean pre-TKA gait waveforms of Cluster 1 (blue solid), Cluster 2 (red dashed), and Cluster 3 (yellow dot-dashed). Grey denotes mean±sd of the clinically classified asymptomatic group (n= 236).

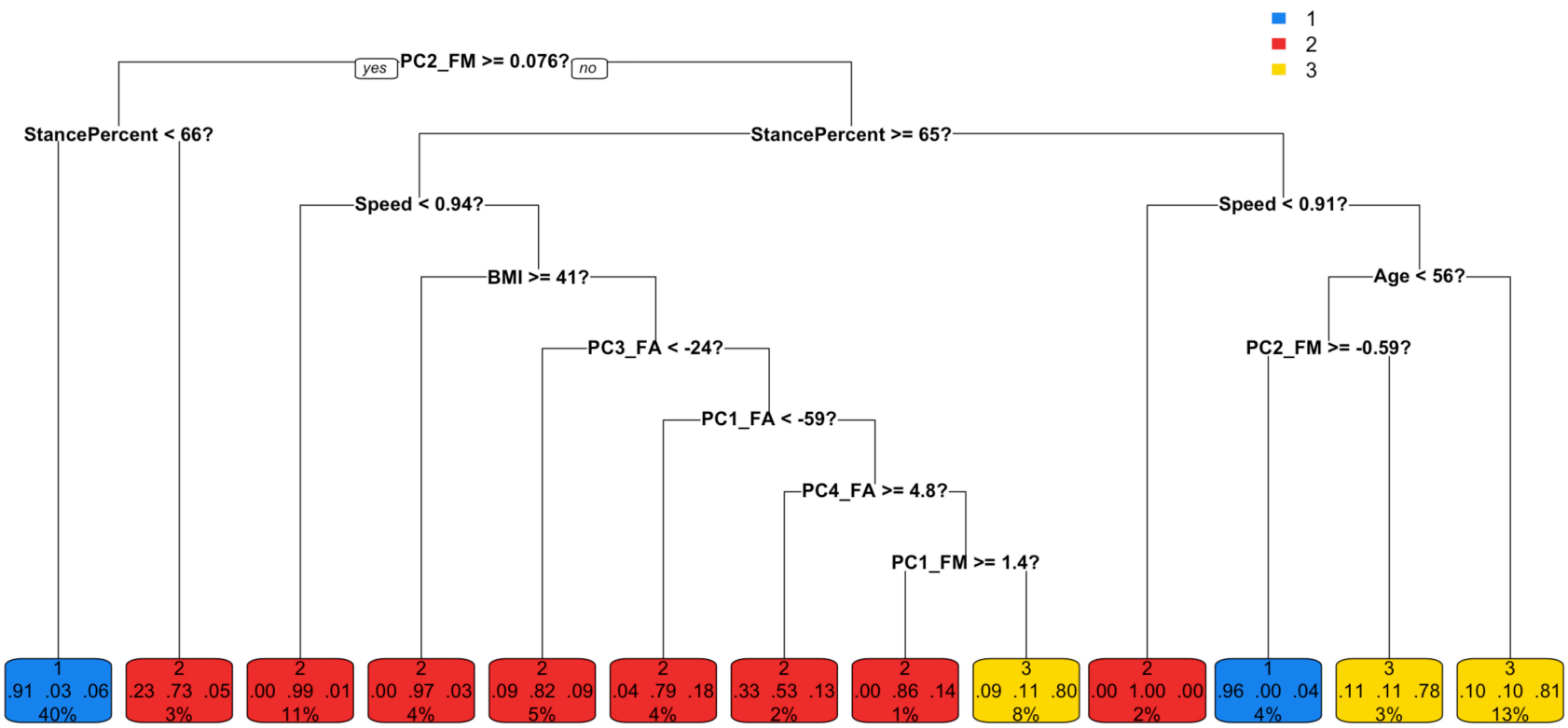

Figure 6.4. Classification and regression tree. Terminal node rows denote i) cluster classification of observations at node; ii) probability of observations belonging to clusters 1 (High Knee Function), 2 (Low Knee Function), and 3 (Moderate Knee Function) at node; iii) percent of total observations at node.

Table 6.2. PC score knee kinematic and kinetic gait features between clusters that were relevant during feature ranking and previously found to be clinically interpretable (Astephen et al., 2008a; Hatfield et al., 2011). Three-way ANOVA and Tukey HSD criteria examined mean differences between clusters.

Interpretation (PC; Variance Explained)	1: High Knee Function	2: Low Knee Function	3: Moderate Knee Function	p-value	Tukey HDS
Flexion Angle (FA)					
Higher score: greater overall flexion angle magnitude throughout the gait cycle (PC1; 65%)	32.46(44.6)	-45.47(60.7)	-4.79(46.8)	**<0.001**	**2<3<1**
Higher scores: more knee extension in heal strike and late stance with earlier occurring peak flexion in swing (PC3; 11%)	5.65(20.5)	-17.50(28.3)	8.43(21.9)	**<0.001**	**2<1,3**
Higher score: more flexion and extension angle range in stance (PC4; 5%)	8.46(16.7)	-2.09(14.1)	-12.75(15.5)	**<0.001**	**3<2<1**
Adduction Angle (AA)					
Higher score: greater overall adduction angle magnitude in stance (PC1; 56%)	0.76(19.9)	-6.91(21.3)	-4.00(19.0)	**<0.001**	**2,3<1**
Flexion Moment (FM)					
Higher score: greater overall flexion moment magnitude in stance (PC1; 50%)	-0.25(1.0)	0.60(1.6)	-0.11(1.3)	**<0.001**	**1,3<2**
Higher scores: more flexion to extension moment range in stance (PC2; 38%)	0.83(0.9)	-0.80(0.8)	-0.81(0.9)	**<0.001**	**2,3<1**
Adduction Moment (AM)					
Higher scores: greater overall adduction moment magnitude in stance (PC1; 71%)	0.04(1.0)	0.02(1.2)	0.22(1.2)	0.09	
Higher score: more first peak to mid-stance adduction moment range in stance (PC2; 15%)	0.32(0.5)	-0.43(0.3)	-0.05(0.4)	**<0.001**	**2<3<1**
Higher score: more mid-stance to second peak adduction moment range in stance (PC3; 6%)	0.15(0.4)	-0.13(0.3)	-0.06(0.3)	**<0.001**	**2<3<1**

Table 6.3. Classification and regression tree classification performance by cluster.

	1: High Knee Function	2: Low Knee Function	3: Moderate Knee Function
Accuracy	0.869	0.851	0.811
Sensitivity (True Positive Rate)	0.836	0.806	0.585
Specificity (True Negative Rate)	0.889	0.880	0.881

Table 6.4. Baseline demographic and pre to post-TKA PC score changes between defined groups examined by Student's t-tests and chi-squared tests.

Feature	Merged 1: High Knee Function 3: Moderate Knee Function	2: Low Knee Function	p-value
Sex			
Female (n, %)	12, 63%	45, 54%	
Male	7, 37%	38, 46%	0.5
Age Pre-TKA (years, SD)	64.1(5.1)	64.4(7.5)	0.9
BMI Pre-TKA (kg/m^2, SD)	32.0(4.9)	33.6(6.0)	0.3
Δ Stance Percent (%, SD)	0.0(1.1)	-0.4(2.1)	0.5
Δ Speed (m/s, SD)	0.1(0.2)	0.2(0.2)	**0.01**
Δ Flexion Angle (FA) PC Scores (PC; Variance Explained)			
Higher score: increase in overall flexion angle magnitude throughout the gait cycle (PC1; 65%)	21.1(57.9)	42.0(74.4)	0.3
Higher scores: increase in knee extension in heal strike and late stance (PC3; 11%)	-9.5(19.2)	15.1(30.3)	**0.001**
Higher score: increase in flexion and extension angle range in stance (PC4; 5%)	-4.9(17.1)	-0.4(15.3)	0.3
Δ Adduction Angle (AA) PC Scores (PC; Variance Explained)			
Higher score: increase in overall adduction angle magnitude in stance (PC1; 56%)	-12.5(22.5)	-11.7(27.0)	0.9
Δ Flexion Moment (FM) PC Scores (PC; Variance Explained)			
Higher score: increase in overall flexion moment magnitude in stance (PC1; 50%)	-0.1(1.3)	-0.5(2.0)	0.5
Higher scores: increase in flexion to extension moment range in stance (PC2; 38%)	0.2(0.8)	0.5(0.8)	0.2
Δ Adduction Moment (AM) PC Scores (PC; Variance Explained)			
Higher scores: increase in overall adduction moment magnitude in stance (PC1; 71%)	0.3(1.8)	-0.6(1.2)	**0.01**
Higher score: increase in first peak to mid-stance adduction moment range in stance (PC2; 15%)	0.1(0.4)	0.3(0.4)	**0.02**
Higher score: increase in mid-stance to second peak adduction moment range in stance (PC3; 6%)	0.1(0.3)	0.2(0.3)	0.2

6.4. Discussion

SOMs quantified and visually represented the quality of a person's knee biomechanics during gait along a multidimensional continuum. Cluster analysis aided in the directional interpretation of gait mechanicals within the SOM, where gait biomechanics worsening across the SOM, correlating to changes in clinical OA severity in terms of asymptomatic, moderate OA, and severe knee OA disease state descriptors (Astephen et al., 2008a). This provided construct validity, where OA progression status could generally be interpreted from bottom left to top right along the SOM, mirroring changes typically observed during OA progression (Astephen et al., 2008a), and post-TKA (Hatfield et al., 2011; Young-Shand et al., 2020). The unsupervised nature of this analysis, objectively mapping demographic and gait observations without *a priori* knowledge, provides novel insights into heterogenous feature combinations (phenotypes), and how they progress longitudinally during the OA disease process and post-TKA.

The three clusters described within the SOM model can be further characterized through subgroup interpretation at the node-level, where gait variability within clusters is not lost. For instance, exploratory analysis clustering the SOM into k=4 clusters defined a fourth cluster within the 'High Knee Function' region in the bottom left corner of the SOM. This new cluster had lower BMIs, shorter stance percent, and more flexion/extension moment range (PC2), corresponding to greater majority of clinically asymptomatic observations (Appendix Figure C.7). The inclusion of a fourth "asymptomatic" cluster captured the variability between asymptomatic and knee OA gait observations (Astephen et al., 2008a; Paul Robert Biggs et al., 2019; Outerleys et al., 2021) within the larger 'High Knee Function' cluster. However, it also separated the remaining 'High Knee Function' nodes into two sections, suggestive of two diverging pathways to Moderate Knee Function regions. Higher order cluster and node-level interpretation allows us to characterize a high volume (up to 121) of functional regions or phenotypes within the SOM. This type of analysis may be advantageous for a disease such as OA. Although we speak of distinct phenotypes (Deveza et al., 2017), recent studies have demonstrated large degrees of overlap between clinically classified groups when attempting to separate them statistically (Paul Robert Biggs et al., 2019; Outerleys

et al., 2021). Therefore, distinct statistical boundaries between biomechanical phenotypes through OA progression and after interventions such as TKA may not exist. The acceptance of this continuum of variability is important as OA treatment and surgical interventions propose approaches specific to patient biomechanics-related profiles (Allen et al., 2016; Bruyère et al., 2015; Salzmann et al., 2017; Vanlommel et al., 2013).

OA variability is generally discussed cross-sectionally at static time points within the OA disease process (Elbaz et al., 2014; Knoop et al., 2011; Paterson et al., 2017; Waarsing et al., 2015; Wilson et al., 2017, 2015; Young-Shand et al., 2020), with end-stage progression considered to converge into common journey (Castaneda et al., 2013). Our SOM framework lends support that not all pathways are common. Figure 6.2b-d illustrates this with three moderate OA individuals who map to different SOM regions during their first gait observations, and progressed along different pathways. Further, individuals b and c both demonstrated stable mapping during their first two gait observations, with a change in functional mapping at observation three. In this case, a SOM map could provide novel utility to objectively illustrate and monitor individual OA journeys longitudinally. Monitoring these pathways would provide new longitudinal understanding of variability in disease manifestation, with the potential to predict patient trajectories or identify sudden declines in functional status. A valuable next step would be the investigation into frequent pathways from asymptomatic to end-stage OA. This could inform optimal timing of non-invasive therapies targeting current and anticipated symptoms or functional deficiencies (Allen et al., 2016; Bannuru et al., 2019), lending to more preventative and personalized care strategies.

SOMs have unique utility for assessing treatment effects of non-invasive and surgical intervention strategies. For instance, post-TKA, the majority (58%) of gait observations mapped to the most severe ‘Low Knee Function’ cluster. The remaining 16% and 26% of observations mapped to the higher functioning ‘High Knee Function’ and ‘Moderate Knee Function’ clusters. Although some post-TKA mapping to more severe OA groups could be a factor of older age or greater BMI, decision trees classification was predominantly governed by gait features, with age and BMI ranked at lower importance in the CART (Figure 6.4). The concept of a biomechanical ceiling effect, such that post-

TKA gait patterns statistically map to disease-state gait over asymptomatic gait is consistent with two recent studies (Paul Robert Biggs et al., 2019; Outerleys et al., 2021), highlighting deficiencies in TKA to restore normative function. It has also been demonstrated that gait improvements from arthroplasty are cluster specific (Young-Shand et al., 2019), and that individuals with higher gait functioning pre-operatively experience the least functional and self-reported improvements post-TKA (Naili et al., 2017; Young-Shand et al., 2020). We also found TKA candidates mapping to 'Low Knee Function' regions experienced more functional benefit from arthroplasty (Table 6.4). Decision tree classification rules suggest an ability to identify this cluster with accuracies of 85.1% (Table 6.3). Features required for CART cluster prediction were also predominantly spatiotemporal and demographic with sagittal knee kinematics (Figure 6.4). With the exception of flexion moment range, these features could be easily captured in clinical settings through consumer-grade motion capture technologies. Future models should explore training SOMs on clinically captured gait features, in combination self-reported OA outcomes. Anchoring regions where conservative therapeutics or surgical intervention occurred against self-reported and biomechanics improvements presents an opportunity to assess multi-modal treatment efficacy, and systemize intervention timing and selection by SOM regions. This may be particularly relevant for TKA wait list management, to objectively prioritize patients by region with the greatest need and improvement potential (Clavel et al., 2016; Frankel et al., 2016). Post-operatively, SOMs can monitor health status and identify sudden declines (Figure 6.2f). Visually interpretable patient mapping may offer an unbiased vehicle for individualized intervention decision making, prioritization, and monitoring, or developing interventions targeting deficiencies tailored to SOM regions, potentially yielding greater response success.

Females made up the majority of the asymptomatic group in the training dataset (138; 61%), yet the minority in the 'High Knee Function' (149; 39%). Pair-wise post-hoc tests also found smaller ratios of females in the 'High Knee Function' cluster than the other more severe clusters (Table 6.1). Gait differences exist between sexes, such as less stance-phase knee flexion moment range and stance phase flexion moment magnitudes

among healthy females (McKean et al., 2007) and lower knee adduction moments (Paterson et al., 2017; Wilson et al., 2015), knee extension at terminal stance among pre-TKA females (Wilson et al., 2015), with some differences favouring female classification among more severe clusters in the decision tree (Figure 6.4). However, females in our clinically classified asymptomatic, moderate OA, and pre-TKA groups did not differ from males in BMI, age, or gait speed, with the exception of moderate OA females walking slightly slower than males (1.2m/s vs. 1.3m/s, $p=0.004$), yet well above the speed cut-off in the decision tree ruleset which would govern classification into low (<0.9m/s) vs. moderate or high knee function (>0.9m/s) clusters (Figure 6.4). Thus, severity-based gait features (not demographic or spaciotemporal features) were interpreted as the greatest influencer in classifying clinically asymptomatic females among the two lower functioning clusters. Our model supports evidence that female OA prevalence is underrepresented relative to males (Hawker et al., 2000), despite evidence of greater radiographic incidence and symptomatic severity of OA among females (Felson et al., 1987; Srikanth et al., 2005). For example, females demonstrate more self-reported and objective function disability from OA (King et al., 2018; Wilson et al., 2015), wait longer to seek medical attention (Karlson et al., 1997), and are then less likely to be referred for TKA relative to males, despite standardized clinical presentation (Borkhoff et al., 2009; Hawker et al., 2000). After surgery, females also lag functionally relative to males (Wilson et al., 2015). This unfavourable patient journey is not specific to sex, but exists among racial boundaries as well (Ibrahim, 2010; Srikanth et al., 2005). These findings again demonstrate utility for patient mapping to support unbiased decision making around OA monitoring care, intervention timing, and surgical prioritization practices.

Our exploratory approach interpreted and considered ten SOM models, selecting the best performing model by silhouette criteria. Although each model demonstrated clusters that mirrored typical OA progression patterns, the ten models were variable in terms of cluster size, demonstrative of weak model repeatability. The selected model also had a low mean silhouette width (0.16, Appendix C Figure C.8), meaning clusters should not be interpreted as distinct groups. This is unlikely a reflection of the sample size, data type

or quality, but the continuum of gait patterns observed through knee OA progression (Astephen et al., 2008a), post-TKA (Outerleys et al., 2021) and variability associated with symptoms of pain (Thorp et al., 2007; Wilson et al., 2017), sex (Paterson et al., 2018; Wilson et al., 2015), and obesity (Harding et al., 2012), which may not separate well with distinct boundaries. Methodologies such as "fuzzy" clusters could be considered which do not assume distinct cluster separation. Similar to most phenotyping studies, this study used secondary data study from a large, longitudinal gait database. The lack access to complete self-reported OA outcomes, physical and biochemical OA attributes (pain patterns, alignment, radiographic severity, biomarkers, cartilage wear patterns) (Paul Robert Biggs et al., 2019; Carlesso et al., 2020; Dell'Isola and Steultjens, 2018; Knoop et al., 2011; Waarsing et al., 2015) for SOM training was a limitation, restricting SOM region and pathway interpretation to being knee joint-specific (Andriacchi et al., 2014). However, our dataset does contain some information on which OA patients progress radiographically or symptomatically, and self-reported improvements post-TKA. Future work will address if radiographic and symptomatic OA progression follows different pathways (Costello et al., 2020), identify SOM regions associated with symptom presentation, and improved outcomes post-TKA. This model has the unique ability to characterize longitudinal observational data during the natural OA disease processes, and explore treatment responses, demonstrating clinical utility (Spil et al., 2020).

This study was the first to propose a machine-learning framework to characterize and cluster multi-dimensional knee kinetic, kinematic and demographic data for diverse patients along the OA disease continuum; up to 121 location-based phenotypes can be characterized. Three large clusters were identified, aiding in SOM directional interpretation and coinciding with clinical OA severity from asymptomatic to end stage OA. A unique aspect of this framework is the ability to objectively track and characterize multivariable OA progression pathways longitudinally, and measure the effect of interventions on knee joint kinematic and kinetic function. Next steps require anchoring mapped locations relative to patient-reported outcomes after intervention, to identify high/low risk intervention regions. Validated progression maps could provide individual

trajectory models, aiding in intervention planning and outcome prediction, developing patient prioritization practices, or tailoring treatment to SOM regions targeting individual manifestations, potentially driving improved treatment success.

Chapter 7. Discussion

7.1. Summary of Findings

This thesis investigated variability among patients with knee OA incorporating both the patient-experience, captured using self-reported outcome measures, and objective knee function, determined from joint-level gait biomechanics. It investigated relationships between individual variability and self-reported and gait function outcomes after TKA intervention. The following summarizes the key findings of this thesis.

7.1.1. Function is an Important Part of the TKA Patient Experience

Function, measured using both self-reported scores and joint-level biomechanics, contributes to the patient experience after arthroplasty. This was first demonstrated in our longitudinal satisfaction study (Chapter 3), where function-based OKS and patient perceptions with difficulty walking both surpassed the influence of pain domains in predicting satisfaction longitudinally (Table 3.3). Although pain is typically considered the most dominant domain in satisfaction determination (Scott et al., 2012), the relative importance of functional measures agrees with prior findings using diverse tools including the OKS (Scott et al., 2010), WOMAC (Bourne et al., 2010; Vissers et al., 2010), self-reported activities of daily living (Noble et al., 2006), and from gait measures (Turcot et al., 2013). TKA is inherently a mechanical surgery altering patient biomechanics. Patients may therefore perceive a lack of improvement in overall function early in the post-TKA recovery process. Function captured through PROMs (or potentially during gait) may contribute to stratification of good and poor outcome groups as early as six weeks after TKA, having utility for early patient profiling and initiating appropriate support interventions.

The second study of this thesis (Chapter 4) linked gait biomechanics to self-reported outcomes after TKA. Patients who report less pain and function improvement after TKA (non-responders) were found to demonstrate less objective functional improvements during gait. Non-responders showed significantly reduced stance-phase varus angles after

TKA, yet lagged in terms of sagittal kinematic and kinetic loading pattern corrections typically observed after arthroplasty (Hatfield et al., 2011). Surgical corrections have historically focused on frontal plane alignment and ligament balancing, treating patients uniformly. Naili et al. proposed that poor patient-reported outcomes might be partially explained by a lack of dynamic kinematic and kinetic corrections, despite alignment corrections in the frontal plane, a feature that surgery may be most able to address biomechanically address (Naili et al., 2017). Chapter 4 did identify novel links between frontal plane features and self-reported pain and function improvement. Specifically, *less* reduction in stance-phase varus (adduction angle) magnitude was independently associated with more improvement in PROM scores (in both the pain and function domains), as were *larger* increases in dynamic frontal plane loading (PC2) (in the function domain alone) (Figure 4.3). It has also been suggested that investigating patient factors such as pain, function, and quality of life relative to changes in specific gait parameters after TKA can aid in understanding the clinical relevance of surgical corrections (Sosdian et al. 2014). Chapter 4 provided links between biomechanical variability, biomechanical changes after TKA, and the patient experience. Objective gait function can provide new and valuable insights into patient variability and the mechanisms associated with why some patients may fair well or poorly. This evidence can be used to inform further investigations that aim to incorporate three-dimensional knee mechanics into personalized surgical innovation and care strategies.

7.1.2. Pre-TKA Biomechanical Variability Impacts Outcomes

To date, joint-level biomechanical variability during gait among TKA candidates has primarily been assessed through *a priori* group definition, such as the MCID stratification applied in Chapter 4. The findings of Chapter 4 demonstrated that knee biomechanical variability may influence the patient experience. The motivation behind Chapters 5 and 6 was the need to better understand biomechanical variability in an unsupervised nature, by modes of statistical separation and similarity. This work was the first to apply multivariate partitioning strategies to characterize patient phenotypes incorporating both knee joint-level biomechanics and patient demographics.

Chapter 5 focused on the TKA candidate population by modeling variability among severe OA individuals. Four clusters were characterized, differing by sex and joint mechanic changes mirroring those observed during OA progression. This demonstrated the functional variability of patients undergoing TKA, likely representing patients undergoing surgery at different points in the severity pathway, and the variable biomechanical manifestations of OA. In general, clusters with more "severe" gait patterns pre-TKA tended to experience greater degrees of biomechanical improvement and self-reported improvement from arthroplasty. This finding agrees with *a priori* results in Chapter 4, and the patient-reported outcome literature, where patients with worse self-reported symptoms preoperatively typically experience the largest relative improvements in self-reported scores post-operatively (Fortin et al., 1999; Hawker et al., 2013; Jiang et al., 2017; Judge et al., 2012; Robertsson and Dunbar, 2001). We propose that patients with worse gait mechanics pre-operatively may have more capacity for objective self-reported (Chapter 4) and objective (Chapter 5) improvements from surgery, and are less likely to experience an functional improvement ceiling effect (Paul Robert Biggs et al., 2019; Outerleys et al., 2021). Knee biomechanic characterization of phenotypes may enable us to prioritize patients who may benefit most for TKA in terms of the patient "journey", and those whose biomechanical gain may be negligible to small. For example, Chapter 5 identified a small (6/105, 6%) higher functioning male phenotype who may not experience any functional benefit from standard TKA, and a sizable (35/105, 33%) higher functioning female cohort whose biomechanical gain from TKA only included reductions to front-plane varus angles during stance. Objective phenotyping classification may provide a means to prioritize patients for surgery, or better manage patient expectations through discussions of phenotype-specific functional predictions. Further work addressing these clusters relative to self-outcomes might also aid in identifying those who could benefit from an altered or non-invasive clinical approach (Allen et al., 2016; Skou et al., 2015).

7.1.3. Sex-Based Biomechanical Profiles May Reflect Bias in Arthroplasty

An interesting finding of our TKA clustering study (Chapter 5), was the dominant separation of clusters by male/female sex. The lower functioning females exhibited

distinctly more clinically "severe" (Astephen et al., 2008a) gait mechanics than any other cluster. Conversely, the lower functioning male cluster was most similar to the higher functioning female cluster. We saw similar trends in our longitudinal SOM clustering paper (Chapter 6). Females made up the majority of the asymptomatic group in the training dataset (138; 61%), yet the minority in the 'high-functioning gait' cluster (149; 39%). Pair-wise post-hoc tests also found smaller ratios of females in the high-functioning cluster than the more severe clusters (Table 6.1). Our model classifying more "asymptomatic" females among lower functioning clusters using objective criteria supports evidence that clinical OA prevalence in females is underrepresented relative to males (Hawker et al., 2000). Females in an older adult population are predisposed to the development of OA (Buckwalter and Lappin, 2000), have a greater incidence of radiographic and symptomatic OA severity (Felson et al., 1987; Srikanth et al., 2005), and have been associated with worse self-reported disability (King et al., 2018) and objective function disability (Wilson et al., 2015). We also know females wait longer to seek medical attention (Karlson et al., 1997), and are then less likely to be referred for TKA than males, despite standardized clinical presentation (Borkhoff et al., 2009; Hawker et al., 2000). After surgery, they lag functionally relative to males (Wilson et al., 2015). This unfavourable patient journey is not specific to sex, but exists among racial boundaries as well (Ibrahim, 2010; Srikanth et al., 2005). Medical research and clinical practices cannot continue to fail to address systemic biases among females and racial ethnicities. Objective patient phenotyping by clusters classification or SOM patient mapping may provide an unbiased vehicle for decision making around OA care. Integrating objective and evidence measures into regular practice may support fair intervention timing, surgical prioritization practices, and patient monitoring.

7.1.4. Data Science Strategies Can Support a Multidimensional and Longitudinal Perspective of OA Variability and TKA Patient Outcomes

Utilizing SOM methodologies, Chapter 6 quantified and visually represented the quality of a person's knee biomechanics during gait along a multidimensional continuum in a large longitudinal population. Gait biomechanics worsened across the SOM, correlating to changes associated with clinically defined OA severity (Astephen et al., 2008a). Chapter 6 demonstrated that higher order cluster and node-level interpretation

allows us to characterize a high volume of functional regions or phenotypes within a SOM, with the potential to characterize biomechanical variability associated with natural patient variability, and differences associated with sex and symptoms of pain. SOM also provided a novel ability to objectively illustrate and monitor an individual's OA journey over time. Regional mapped location interpretation, and the ability to observe an individual's change relative to their prior state may be advantageous for OA monitoring, such that it enables a patient's baseline status to operate as a personalized "asymptomatic control". Further, as distinct statistical boundaries between biomechanical phenotypes through OA progression and after TKA interventions may not exist (Paul Robert Biggs et al., 2019; Outerleys et al., 2021), there may not be a region or phenotype for TKA intervention that results in a globally optimal outcome for all individuals. Instead, we can identify a number of regions based on expected individual or group-level disease state status. This group-level approach may provide more sensitivity relative to a global model, potentially improving confidence in predictions and an ability to better tailor OA management and treatment to individuals.

In addition to being an interpretable feedback and monitoring tool, SOMs have the potential to incorporate all the key considerations and findings from this thesis i) a longitudinal trajectory or view (Chapter 3), ii) the ability to incorporate both self-reported and objective functional measures in patient mapping (Chapter 4), or any comprehensive collection of multidimensional patient variables, iii) the inherent incorporation of baseline functional and self-reported status for outcome prediction (Chapters 3, 4 and 5), iv) the ability characterize multi-dimensional variability or phenotypes (Chapter 5), and finally v) to support un-bias patient decision making around conservative or surgical interventions. SOMs have also uncovered novel phenotypes associated with patient self-reported domains when modeled using PROMs data alone (Appendix D). Future work is required for model refinement and validation. First steps on our existing datasets will address if radiographic and symptomatic OA progression follows different pathways (Costello et al., 2020), and identify SOM regions associated with symptom presentation and improved outcomes post-TKA. Next steps could include incorporating new physical and biochemical OA attributes to SOM training (e.g., self-reported symptoms, pain

patterns, alignment, radiographic severity, biomarkers, cartilage wear patterns, frailty indicators) (Paul Robert Biggs et al., 2019; Carlesso et al., 2020; Dell'Isola and Steultjens, 2018; Knoop et al., 2011; Schmucker et al., 2019; Waarsing et al., 2015), or developing a minimal yet comprehensive dataset encompassing relevant features that could be captured clinically. Final utility might include anchoring all SOM regions against self-reported outcomes for OA various intervention strategies (including exercise or other conservative care). Tailoring treatment to SOM regions targeting individual manifestations has the potential to drive evidence-based decision making, preventative, and personalized care strategies in OA. My hope is that the collective findings of this work provide new insights and inspires new investigations employing unique data driven approaches to OA and TKA investigations incorporating multi-dimensional patient dimensions including self-reported and objective functional scores. I believe we have an opportunity to continue to utilize the data and modeling tools we have available to us, to expose new relationships, ultimately enabling us to "think what nobody has yet thought, about that which everybody sees" (Erwin Schrodinger).

7.2. Research Limitations

OA is a complex disease encompassing joint-level degradation, symptomatic presentation, mobility limitations, an inflammatory response, and metabolic influences which changes over the course of the disease's progress. This thesis focused on mobility limitations as captured using knee joint level gait analysis, and symptomatic features as measured using self-reported scores. Although some components of this thesis were able to capture a longitudinal view of the OA disease process and TKA recovery, incorporating other dimensions of OA disease presentation (such as biomarkers or radiographic information) is required to create a more comprehensive picture of OA patient variability, perhaps uncovering feature combinations particularly meaningful in variability stratification and TKA outcome modeling.

This work was conducted on a series of pooled retrospective datasets, designed for studies independent of this thesis project. Gait and navigation data collected during my

Master's thesis, and time spent in industry examining intervention outcomes among frail elderly individuals helped motivate this work. I believe this thesis demonstrates an ability to extract novel findings from our existing datasets by creating linkages and looking at our data using different tools; providing evidence and frameworks to guide future OA and TKA patient modeling. Although this work was conducted on a unique high-volume dataset, containing high-accuracy biomechanics with self-reported outcomes, working with retrospective datasets alone has limitations. Datasets were originally composed with objectives that differ from those of this thesis, therefore recruitment, screening, inclusion criteria, and data collected may influence relationships and phenotypes derived (Spil et al., 2020). Physical study inclusion criteria, such as the ability to walk un-assisted, also suggests findings may not generalize well to more global TKA populations. Not all studies collected self-reported metrics, radiographic severity or alignment information, resulting in incomplete variables in some instances. The variables collected also limit this work to a research-area-specific area of variability characterization and phenotyping (Andriacchi et al., 2014), best described by knee joint level biomechanical and self-reported outcomes. Our retrospective datasets were all collected in the Halifax Regional Municipality, from the Dynamics of Human Motion Laboratory, or the Halifax Infirmary Orthopedic site. This means our patients population may be regionally different from other Canadian sites (more obese, with more co-morbid conditions (Canada and santé, 2011)). Patients also saw a select number of TKA surgeons, who may not be representative of surgical centers at other sites within the province of Nova Scotia, or nationally. A multicentered approach and measuring stability by re-producing modeling frameworks on external datasets is required to achieve clinical utility.

Recent studies (Paul Robert Biggs et al., 2019; Outerleys et al., 2021), and the clustering chapters presented here found a failure to reliably discriminate clinically classified groups when attempting to separate them statistically. This was interpreted to reflect the continuum of natural variability of OA patients and over the disease spectrum, where distinct boundaries between individual phenotypes and across clinically classified groups may not exist. Therefore, presented phenotypes cannot be interpreted as mutually exclusive groups. This has implications for surgical decision making, and we cannot

assume that an average response among a phenotype will apply for every individual. It should be noted that most machine learning clustering quality metrics, such as the silhouette coefficient presented here, assume that an optimal and complete separation within the data does exit. This may not be the case in many real-life data applications. However, future studies with larger datasets, more variables, and the sensitivity to characterize more phenotypes may be valuable to create clusters of similar phenotypes, and better understand expected variability within larger phenotype groups.

Specific to modeling TKA outcomes, there remains a lack of consensus on what a good or poor outcome entails. Part of that is a reflection of the success of TKA, where strong clinical indicators such as mortality and infection (with very low incidence rates) generally do not apply. It has been proposed that "expected health benefit [should] exceed the expected negative consequences by a sufficiently wide margin to make TKA worth performing" (Escobar et al., 2003), in the context of the "average patient seen by an average physician" (Brooks, 1996). A number of endpoints have been proposed for TKA outcomes assessment, with this thesis selecting satisfaction and MCID criteria in our studies, generally based on data availability and our desire to incorporate endpoint metrics that reflect the patient experience. Still, arthroplasty is a generally successful procedure, resulting in 74-83% of patients meeting "responder" or good outcome criteria in our studies (Chapters 3-4). The unbalanced nature of our success groups results in a lower power to detect changes within non-responder groups. Further, self-reported outcome scores are complex, and should not be a definitive goal for improving patient care, where scores may be influenced by contradictory or non-modifiable external factors (Gibon et al., 2020). We did not have access to auxiliary dimensions of depression or anxiety in most cases. Therefore, our results should be interpreted acknowledging the limitations of self-reported scores, discussed in detail in section 1.4.1.

7.3. Future Directions

The findings of this work have demonstrated that patient function captured using gait mechanics can provide important insights into patient biomechanical variability and the mechanical basis behind functional limitations of TKA intervention. Joint-level mechanics have not routinely been collected clinically, historically burdened by time and financial investment of data collection practices, coupled with a lack of convincing clinical utility. Modern gait collection protocols have the potential to efficiently capture detailed mechanical information on all arthroplasty candidates in clinically settings using consumer-grade marker-less motion capture. When gait mechanics are paired with other dimensions of patient health such as patient-reported outcomes, novel understanding of the mechanisms that may be impacting the patient experience can be gained.

Determining the basis for poor arthroplasty outcomes has been a debate over the 30 year history of the modern prosthesis. This work has demonstrated the utility of establishing unique linkages between diverse research datasets, and how data science and machine-learning approaches can enable us to analyze temporal and multidimensional datasets collectively, capturing new perspectives and providing novel insights into our datasets, such as how patient variability might be meaningful for patient care and outcomes. For example, this work has suggested that patients who present for TKA with greater varus kinematic magnitudes during stance-phase of gait and an absence of other severe OA features may characterize clinical candidate subgroups for whom neutral corrections may not be clinically relevant to self-reported improvements in pain and function. It has also characterized variability in TKA and OA continuum cohorts, and demonstrated a link between variability and gait improvement after arthroplasty. Investigating patient biomechanical variability with respect to outcomes in larger studies is an important area for further research, which may enable us to anchor post-intervention gait changes next to self-reported criteria such as “somewhat better” or “a great deal better”, and identify gait changes that are clinically meaningful to individuals (Escobar et al., 2007). Diverse longitudinal clinical datasets, paired with self-reported metrics, and other clinical indicators (such as comorbidities, radiographic severity, alignment, cartilage wear patterns, biomarkers, or frailty indicators) could populate a rich and

valuable dataset, and improve our understanding of the holistic patient picture. New feature inclusion may identify attributes that lend to better statistical separation and enable us to train and validate stronger variability and outcome models. This work could also inform strategic collection of features specifically optimized for phenotyping analysis (a multi-disciplinary minimal dataset) while supporting a comprehensive patient view (Spil et al., 2020). Therefore, a focus on working collaboratively, linking datasets between independent disciplines of OA and TKA research and utilizing data science and machine-learning strategies should be priority for future clinical and research investment, with the methodologies presented here directly applicable to other clinical applications. When paired with outcome information, findings can support better informed clinical decision making, and provide valuable direction for innovations in individualized OA management and TKA care, with the ultimate goal of improving outcomes for patients.

References

Abidi, S., Roy, P.C., Shah, M.S., Yu, J., Yan, S., 2018. A Data Mining Framework for Glaucoma Decision Support Based on Optic Nerve Image Analysis Using Machine Learning Methods. J Healthc Inform Res. https://doi.org/10.1007/s41666-018-0028-7

Ali, A., Sundberg, M., Robertsson, O., Dahlberg, L.E., Thorstensson, C.A., Redlund-Johnell, I., Kristiansson, I., Lindstrand, A., 2014. Dissatisfied patients after total knee arthroplasty: a registry study involving 114 patients with 8-13 years of followup. Acta orthopaedica 85, 229–233. https://doi.org/10.3109/17453674.2014.916487

Allen, K.D., Choong, P.F., Davis, A.M., Dowsey, M.M., Dziedzic, K.S., Emery, C., Hunter, D.J., Losina, E., Page, A.E., Roos, E.M., Skou, S.T., Thorstensson, C.A., Esch, M. van der, Whittaker, J.L., 2016. Osteoarthritis: Models for appropriate care across the disease continuum. Best Practice & Research Clinical Rheumatology 30, 503–535. https://doi.org/10.1016/j.berh.2016.09.003

Altman, R., Asch, E., Bloch, D., Bole, G., Borenstein, D., Brandt, K., Christy, W., Cooke, T.D., Greenwald, R., Hochberg, M., Howell, D., Kaplan, D., Koopman, W., Longley, S., Mankin, H., McShane, D.J., Medsger, T., Meenan, R., Mikkelsen, W., Moskowitz, R., Murphy, W., Rothschild, B., Segal, M., Sokoloff, L., Wolfe, F., 1986. Development of criteria for the classification and reporting of osteoarthritis: Classification of osteoarthritis of the knee. Arthritis & Rheumatology 29, 1039–1049. https://doi.org/10.1002/art.1780290816

Alzahrani, K., Gandhi, R., Debeer, J., Petruccelli, D., Mahomed, N., 2011. Prevalence of clinically significant improvement following total knee replacement. The Journal of rheumatology 38, 753–759. https://doi.org/10.3899/jrheum.100233

Andriacchi, T.P., 1993. Functional analysis of pre and post-knee surgery: total knee arthroplasty and ACL reconstruction. Journal of biomechanical engineering 115, 575–581.

Andriacchi, T.P., Favre, J., Erhart-Hledik, J.C., Chu, C.R., 2014. A Systems View of Risk Factors for Knee Osteoarthritis Reveals Insights into the Pathogenesis of the Disease. Annals of Biomedical Engineering 43, 376–387. https://doi.org/10.1007/s10439-014-1117-2

Andriacchi, T.P., Koo, S., Scanlan, S.F., 2009. Gait Mechanics Influence Healthy Cartilage Morphology and Osteoarthritis of the Knee. The Journal of Bone and Joint Surgery (American) 91, 95–101. https://doi.org/10.2106/jbjs.h.01408

Andriacchi, T.P., Mündermann, A., Smith, R.L., Alexander, E.J., Dyrby, C.O., Koo, S., 2004. A Framework for the in Vivo Pathomechanics of Osteoarthritis at the Knee.

Annals of Biomedical Engineering 32, 447–457. https://doi.org/10.1023/b:abme.0000017541.82498.37

Angst, F., Aeschlimann, A., Michel, B.A., Stucki, G., 2002. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. The Journal of rheumatology 29, 131–138.

Astephen, J.L., Deluzio, K.J., Caldwell, G.E., Dunbar, M.J., 2008a. Biomechanical changes at the hip, knee, and ankle joints during gait are associated with knee osteoarthritis severity. Journal of Orthopaedic Research 26, 332–341. https://doi.org/10.1002/jor.20496

Astephen, J.L., Deluzio, K.J., Caldwell, G.E., Dunbar, M.J., Hubley-Kozey, C.L., 2008b. Gait and neuromuscular pattern changes are associated with differences in knee osteoarthritis severity levels. Journal of Biomechanics 41, 868–876. https://doi.org/10.1016/j.jbiomech.2007.10.016

Bade, M.J., Wolfe, P., Zeni, J.A., Stevens-Lapsley, J.E., Snyder-Mackler, L., 2012. Predicting poor physical performance after total knee arthroplasty. null 30, 1805–1810. https://doi.org/10.1002/jor.22140

Baert, I.A.C., Lluch, E., Mulder, T., Nijs, J., Noten, S., Meeus, M., 2016. Does pre-surgical central modulation of pain influence outcome after total knee replacement? A systematic review. Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society 24, 213–223. https://doi.org/10.1016/j.joca.2015.09.002

Baker, P.N., Meulen, J.H. van der, Lewsey, J., Gregg, P.J., Wales, N.J.R. for E. and, 2007. The role of pain and function in determining patient satisfaction after total knee replacement. Data from the National Joint Registry for England and Wales. The Journal of bone and joint surgery. British volume 89, 893–900. https://doi.org/10.1302/0301-620x.89b7.19091

Baker, P.N., Rushton, S., Jameson, S.S., Reed, M., 2013. Patient satisfaction with total knee replacement cannot be predicted from pre-operative variables alone. The Bone and Joint Journal 95-B, 1359–1365. https://doi.org/10.1302/0301-620x.95b10

Bannuru, R.R., Osani, M.C., Vaysbrot, E.E., Arden, N.K., Bennell, K., Bierma-Zeinstra, S.M.A., Kraus, V.B., Lohmander, L.S., Abbott, J.H., Bhandari, M., Blanco, F.J., Espinosa, R., Haugen, I.K., Lin, J., Mandl, L.A., Moilanen, E., Nakamura, N., Snyder-Mackler, L., Trojian, T., Underwood, M., McAlindon, T.E., 2019. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthr Cartilage 27, 1578–1589. https://doi.org/10.1016/j.joca.2019.06.011

Bansback, N., Tsuchiya, A., Brazier, J., Anis, A., 2012. Canadian valuation of EQ-5D health states: preliminary value set and considerations for future valuation studies. PloS one 7, e31115. https://doi.org/10.1371/journal.pone.0031115

Bargren, J.H., Blaha, J.D., Freeman, M.A.R., 1983. Alignment in Total Knee Arthroplasty Correlated Biomechanical and Clinical Observations. Clin Orthop Relat R 173, 178–183. https://doi.org/10.1097/00003086-198303000-00024

Beard, D.J., Harris, K., Dawson, J., Doll, H., Murray, D.W., Carr, A.J., Price, A.J., 2015. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. Journal of clinical epidemiology 68, 73–79. https://doi.org/10.1016/j.jclinepi.2014.08.009

Beaton, D.E., Bombardier, C., Katz, J.N., Wright, J.G., 2001. A taxonomy for responsiveness. J Clin Epidemiol 54, 1204–1217. https://doi.org/10.1016/s0895-4356(01)00407-3

Bedson, J., Croft, P.R., 2008. The discordance between clinical and radiographic knee osteoarthritis: a systematic search and summary of the literature. BMC Musculoskeletal Disorders 9, 116. https://doi.org/10.1186/1471-2474-9-116

Behrend, H., Zdravkovic, V., Giesinger, J., Giesinger, K., 2016. Factors Predicting the Forgotten Joint Score After Total Knee Arthroplasty. The Journal of Arthroplasty 31, 1927–1932. https://doi.org/10.1016/j.arth.2016.02.035

Bellamy, N., Buchanan, W.W., Goldsmith, C.H., Campbell, J., Stitt, L.W., 1988. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. The Journal of rheumatology 15, 1833–1840.

Bellemans, J., Colyn, W., Vandenneucker, H., Victor, J., 2012. The Chitranjan Ranawat award: is neutral mechanical alignment normal for all patients? The concept of constitutional varus. Clinical Orthopaedics and Related Research 470, 45–53. https://doi.org/10.1007/s11999-011-1936-5

Bennell, K.L., Bowles, K.A., Wang, Y., Cicuttini, F., Davies-Tuck, M., Hinman, R.S., 2011. Higher dynamic medial knee load predicts greater cartilage loss over 12 months in medial knee osteoarthritis. Annals of the rheumatic diseases 70, 1770–1774. https://doi.org/10.1136/ard.2010.147082

Bierma-Zeinstra, S.M.A., Verhagen, A.P., 2011. Osteoarthritis subpopulations and implications for clinical trial design. Arthritis research & therapy 13, 213–9. https://doi.org/10.1186/ar3299

Biggs, P R, Whatling, G.M., Wilson, C., Holt, C.A., 2019. Correlations between patient-perceived outcome and objectively-measured biomechanical change following Total Knee Replacement. Gait & Posture 70, 65–70. https://doi.org/10.1016/j.gaitpost.2019.02.028

Biggs, Paul Robert, Whatling, G.M., Wilson, C., Metcalfe, A.J., Holt, C.A., 2019. Which osteoarthritic gait features recover following total knee replacement surgery? PloS one 14, e0203417. https://doi.org/10.1371/journal.pone.0203417

Bijlsma, J.W.J., Berenbaum, F., Lafeber, F.P.J.G., 2011. Series Arthritis 1 Osteoarthritis: an update with relevance for clinical practice. The Lancet 377, 2115–26.

Blakeney, W., Clément, J., Desmeules, F., Hagemeister, N., Rivière, C., Vendittoli, P.-A., 2018. Kinematic alignment in total knee arthroplasty better reproduces normal gait than mechanical alignment. Knee Surgery, Sports Traumatology, Arthroscopy 32, 2133–8. https://doi.org/10.1007/s00167-018-5174-1

Bonnefoy-Mazure, A., Armand, S., Sagawa, Y., Suvà, D., Miozzari, H., Turcot, K., 2017. Knee Kinematic and Clinical Outcomes Evolution Before, 3 Months, and 1 Year After Total Knee Arthroplasty. Journal of Arthroplasty 32, 793–800. https://doi.org/10.1016/j.arth.2016.03.050

Borkhoff, C.M., Hawker, G.A., Kreder, H.J., Glazier, R.H., Mahomed, N.N., Wright, J.G., 2009. Patients' gender affected physicians' clinical decisions when presented with standardized patients but not for matching paper patients. Journal of clinical epidemiology 62, 527–541. https://doi.org/10.1016/j.jclinepi.2008.03.009

Bourne, R.B., Chesworth, B.M., Davis, A.M., Mahomed, N.N., Charron, K.D.J., 2010. Patient satisfaction after total knee arthroplasty: who is satisfied and who is not? Clinical orthopaedics and related research 468, 57–63. https://doi.org/10.1007/s11999-009-1119-9

Boyd, J.E., Little, J.J., 2005. Biometric Gait Recognition, in: Undefined, Advanced Studies in Biometrics. pp. 19–42. https://doi.org/10.1007/11493648_2

Bozic, K.J., Belkora, J., Chan, V., Youm, J., Zhou, T., Dupaix, J., Bye, A.N., III, C.H.B., Chenok, K.E., III, J.I.H., 2013. Shared Decision Making in Patients with Osteoarthritis of the Hip and Knee. The Journal of Bone and Joint Surgery (American) 95, 1633–1639. https://doi.org/10.2106/jbjs.m.00004

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees, Chapman and Hall/CRC. Chapman and Hall/CRC.

Brooks, R., 1996. EuroQol: the current state of play. Health policy (Amsterdam, Netherlands) 37, 53–72.

Bruyère, O., Cooper, C., Arden, N., Branco, J., Brandi, M.L., Herrero-Beaumont, G., Berenbaum, F., Dennison, E., Devogelaer, J.-P., Hochberg, M., Kanis, J., Laslop, A., McAlindon, T., Reiter, S., Richette, P., Rizzoli, R., Reginster, J.-Y., 2015. Can we identify patients with high risk of osteoarthritis progression who will respond to

treatment? A focus on epidemiology and phenotype of osteoarthritis. Drugs & aging 32, 179–187. https://doi.org/10.1007/s40266-015-0243-3

Buckwalter, J.A., Lappin, D.R., 2000. The Disproportionate Impact of Chronic Arthralgia and Arthritis Among Women. Clin Orthop Relat R 372, 159–168. https://doi.org/10.1097/00003086-200003000-00018

Bullens, P.H., Loon, C.J. van, Malefijt, M.C. de W., Laan, R.F., Veth, R.P., 2001. Patient satisfaction after total knee arthroplasty: a comparison between subjective and objective outcome assessments. Journal of Arthroplasty 16, 740–747. https://doi.org/10.1054/arth.2001.23922

Burns, L.C., Ritvo, S.E., Ferguson, M.K., Clarke, H., Seltzer, Z., Katz, J., 2015. Pain catastrophizing as a risk factor for chronic pain after total knee arthroplasty: a systematic review. Journal of pain research 8, 21–32. https://doi.org/10.2147/jpr.s64730

Caldas, R., Rátiva, D., Neto, F.B. de L., 2018. Clustering of Self-Organizing Maps as a means to support gait kinematics analysis and symmetry evaluation. Med Eng Phys 62, 46–52. https://doi.org/10.1016/j.medengphy.2018.09.007

Canada, P.H.A. of, santé, C.I. for H.I./ I. canadien d'information sur la, 2011. Obesity in Canada.

Carlesso, L.C., Hawker, G.A., Torner, J., Lewis, C.E., Nevitt, M., Neogi, T., 2020. Association of intermittent and constant knee pain patterns with knee pain severity, radiographic knee osteoarthritis duration and severity. Arthritis Care & Research.

Castaneda, S., Roman-Blas, J.A., Largo, R., Herrero-Beaumont, G., 2013. Osteoarthritis: a progressive disease with changing phenotypes. Rheumatology 53, 1–3. https://doi.org/10.1093/rheumatology/ket247

Centre, H. and S.C.I., 2011. Hospital Episode Statistics: Finalised Patient Reported Outcome Measures (PROMs) in England. Health and Social Care Information Centre.

Chang, A., Hochberg, M., Song, J., Dunlop, D., Chmiel, J.S., Nevitt, M., Hayes, K., Eaton, C., Bathon, J., Jackson, R., Kwoh, C.K., Sharma, L., 2010. Frequency of varus and valgus thrust and factors associated with thrust presence in persons with or at higher risk of developing knee osteoarthritis. Arthritis & Rheumatology 62, 1403–1411. https://doi.org/10.1002/art.27377

CIHI, 2015a. Inpatient Hospitalizations, Length of Stay, Surgeries and Newborn Indicators in 2013–2014. Canadian Institute for Health Information / Institut canadien d'information sur la santé.

CIHI, 2015b. Hip and Knee Replacements in Canada: Canadian Joint Replacement Registry 2015 Annual Report. Canadian Institute for Health Information / Institut canadien d'information sur la santé.

Clatworthy, J., Buick, D., Hankins, M., Weinman, J., Horne, R., 2005. The use and reporting of cluster analysis in health psychology: a review. British journal of health psychology 10, 329–358. https://doi.org/10.1348/135910705x25697

Clavel, N., Coster, C.D., Pomey, M.-P., Sanmartin, C., Bohm, E., Dunbar, M.J., Frank, C.Y., Hawker, G., Noseworthy, T., 2016. Appropriateness for Total Joint Replacement: Perspectives of Decision-Makers. Healthcare policy = Politiques de sante 11, 80–92.

Costello, K.E., Wilson, J.L.A., Stanish, W.D., Urquhart, N., Hubley-Kozey, C.L., 2020. Differences in Baseline Joint Moments and Muscle Activation Patterns Associated With Knee Osteoarthritis Progression When Defined Using a Clinical Versus a Structural Outcome. J Appl Biomech 36, 39–51. https://doi.org/10.1123/jab.2019-0127

Costigan, P.A., Wyss, U.P., Deluzio, K.J., Li, J., 1992. Semiautomatic three-dimensional knee motion assessment system. Medical Biological Eng Comput 30, 343–350. https://doi.org/10.1007/bf02446973

Culliton, S.E., Bryant, D.M., Overend, T.J., MacDonald, S.J., Chesworth, B.M., 2012. The relationship between expectations and satisfaction in patients undergoing primary total knee arthroplasty. The Journal of Arthroplasty 27, 490–492. https://doi.org/10.1016/j.arth.2011.10.005

Dave, A.J., Selzer, F., Losina, E., Usiskin, I., Collins, J.E., Lee, Y.C., Band, P., Dalury, D.F., Iorio, R., Kindsfater, K., Katz, J.N., 2017. The association of pre-operative body pain diagram scores with pain outcomes following total knee arthroplasty. Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society 25, 667–675. https://doi.org/10.1016/j.joca.2016.12.013

Dawson, J., Fitzpatrick, R., Murray, D., 1998. Questionnaire on the perceptions of patients about total knee replacement. … of Bone & Joint Surgery ….

Dell'Isola, A., Steultjens, M., 2018. Classification of patients with knee osteoarthritis in clinical phenotypes: Data from the osteoarthritis initiative. PloS one 13, e0191045. https://doi.org/10.1371/journal.pone.0191045

Deluzio, K.J., Astephen, J.L., 2007. Biomechanical features of gait waveform data associated with knee osteoarthritis. Gait & Posture 25, 86–93. https://doi.org/10.1016/j.gaitpost.2006.01.007

Deluzio, K.J., Wyss, U.P., Li, J., Costigan, P.A., 1993. A procedure to validate three-dimensional motion assessment systems. Journal of Biomechanics 26, 753–759. https://doi.org/10.1016/0021-9290(93)90037-f

Desmeules, F., Dionne, C.E., Belzile, É., Bourbonnais, R., Frémont, P., 2009. Waiting for total knee replacement surgery: factors associated with pain, stiffness, function and quality of life. BMC Musculoskeletal Disorders 10, 52. https://doi.org/10.1186/1471-2474-10-52

Deveza, L.A., Melo, L., Yamato, T.P., Mills, K., Ravi, V., Hunter, D.J., 2017. Knee osteoarthritis phenotypes and their relevance for outcomes: a systematic review. Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society 25, 1926–1941. https://doi.org/10.1016/j.joca.2017.08.009

Dillon, C.F., Rasch, E.K., Gu, Q., Hirsch, R., 2006. Prevalence of knee osteoarthritis in the United States: arthritis data from the Third National Health and Nutrition Examination Survey 1991-94. The Journal of rheumatology 33, 2271–2279.

Dolan, P., Sutton, M., 1997. Mapping visual analogue scale health state valuations onto standard gamble and time trade-off values. Social science & medicine 44, 1519–1530. https://doi.org/10.1016/s0277-9536(96)00271-7

Dunbar, M.J., Richardson, G., Robertsson, O., 2013. I can't get no satisfaction after my total knee replacement: rhymes and reasons. The bone & joint journal 95-B, 148–152. https://doi.org/10.1302/0301-620x.95b11.32767

Dunbar, M.J., Robertsson, O., Ryd, L., 2004. What's all that noise? The effect of co-morbidity on health outcome questionnaire results after knee arthroplasty. Acta orthopaedica Scandinavica 75, 119–126. https://doi.org/10.1080/00016470412331294355

Dunbar, M.J., Robertsson, O., Ryd, L., Lidgren, L., 2001. Appropriate questionnaires for knee arthroplasty. Bone & Joint Journal 83-B, 339–344x. https://doi.org/10.1302/0301-620x.83b3.11134

Eckhoff, D.G., Bach, J.M., Spitzer, V.M., Reinig, K.D., Bagur, M.M., Baldini, T.H., Flannery, N.M.P., 2005. Three-dimensional mechanics, kinematics, and morphology of the knee viewed in virtual reality. The Journal of Bone and Joint Surgery (American) 87 Suppl 2, 71–80. https://doi.org/10.2106/jbjs.e.00440

Edwards, R.R., Haythornthwaite, J.A., Smith, M.T., Klick, B., Katz, J.N., 2009. Catastrophizing and depressive symptoms as prospective predictors of outcomes following total knee replacement. Pain research & management 14, 307–311.

Elbaz, A., Mor, A., Segal, G., Debi, R., Shazar, N., Herman, A., 2014. Novel classification of knee osteoarthritis severity based on spatiotemporal gait analysis.

Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society 22, 457–463. https://doi.org/10.1016/j.joca.2013.12.015

Esch, M. van der, Knoop, J., Leeden, M. van der, Roorda, L.D., Lems, W.F., Knol, D.L., Dekker, J., 2015. Clinical phenotypes in patients with knee osteoarthritis: a study in the Amsterdam osteoarthritis cohort. Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society 23, 544–549. https://doi.org/10.1016/j.joca.2015.01.006

Escobar, A., Pérez, L.G., Herrera-Espiñeira, C., Aizpuru, F., Sarasqueta, C., Tejada, M.G.S. de, Quintana, J.M., Bilbao, A., 2013. Total knee replacement; minimal clinically important differences and responders. Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society 21, 2006–2012. https://doi.org/10.1016/j.joca.2013.09.009

Escobar, A., Quintana, J.M., Aróstegui, I., Azkárate, J., Güenaga, J.I., Arenaza, J.C., Garai, I., 2003. Development Of Explicit Criteria For Total Knee Replacement. International Journal of Technology Assessment in Health Care 19, 57–70. https://doi.org/10.1017/s0266462303000060

Escobar, A., Quintana, J.M., Bilbao, A., Aróstegui, I., Lafuente, I., Vidaurreta, I., 2007. Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. Osteoarthritis and Cartilage 15, 273–280. https://doi.org/10.1016/j.joca.2006.09.001

Fang, D.M., Ritter, M.A., Davis, K.E., 2009. Coronal alignment in total knee arthroplasty: just how important is it? The Journal of Arthroplasty 24, 39–43. https://doi.org/10.1016/j.arth.2009.04.034

Felson, D.T., 2010. Identifying different osteoarthritis phenotypes through epidemiology. Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society 18, 601–604. https://doi.org/10.1016/j.joca.2010.01.007

Felson, D.T., Lawrence, R.C., Dieppe, P.A., Hirsch, R., Helmick, C.G., Jordan, J.M., Kington, R.S., Lane, N.E., Nevitt, M.C., Zhang, Y., Sowers, M., McAlindon, T., Spector, T.D., Poole, A.R., Yanovski, S.Z., Ateshian, G., Sharma, L., Buckwalter, J.A., Brandt, K.D., Fries, J.F., 2000. Osteoarthritis: new insights. Part 1: the disease and its risk factors., in: Undefined, Annals of Internal Medicine. pp. 635–646.

Felson, D.T., Naimark, A., Anderson, J., Kazis, L., Castelli, W., Meenan, R.F., 1987. The prevalence of knee osteoarthritis in the elderly. The Framingham Osteoarthritis Study. Arthritis & Rheumatism 30, 914–918. https://doi.org/10.1002/art.1780300811

Fenton, J.J., 2012. The Cost of Satisfaction. Archives of internal medicine 172, 405–7. https://doi.org/10.1001/archinternmed.2011.1662

Ferraz, M.B., Quaresma, M.R., Aquino, L.R., Atra, E., Tugwell, P., Goldsmith, C.H., 1990. Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. The Journal of rheumatology 17, 1022–1024.

Fortin, P.R., Clarke, A.E., Joseph, L., Liang, M.H., Tanzer, M., Ferland, D., Phillips, C., Partridge, A.J., Bélisle, P., Fossel, A.H., Mahomed, N., Sledge, C.B., Katz, J.N., 1999. Outcomes of total hip and knee replacement: preoperative functional status predicts outcomes at six months after surgery. Arthritis & Rheumatism 42, 1722–1728. https://doi.org/10.1002/1529-0131(199908)42:8<1722::aid-anr22>3.0.co;2-r

Frankel, L., Sanmartin, C., Conner-Spady, B., Marshall, D.A., Freeman-Collins, L., Wall, A., Hawker, G.A., 2012. Osteoarthritis patients' perceptions of "appropriateness" for total joint replacement surgery. Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society 20, 967–973. https://doi.org/10.1016/j.joca.2012.05.008

Frankel, L., Sanmartin, C., Hawker, G., Coster, C.D., Dunbar, M., Bohm, E., Noseworthy, T., 2016. Perspectives of orthopaedic surgeons on patients' appropriateness for total joint arthroplasty: a qualitative study. Journal of evaluation in clinical practice 22, 164–170. https://doi.org/10.1111/jep.12449

Gibbons, R.D., Hedeker, D., DuToit, S., 2010. Advances in analysis of longitudinal data. Annual review of clinical psychology 6, 79–107. https://doi.org/10.1146/annurev.clinpsy.032408.153550

Gibon, E., Goodman, S.B., Goodman, M.J., 2020. Patient Satisfaction After Total Knee Arthroplasty: A Realistic or Imaginary Goal? Orthopedic Clinics of North America 48, 421–431.

Giesinger, J.M., Hamilton, D.F., Jost, B., Behrend, H., Giesinger, K., 2015. WOMAC, EQ-5D and Knee Society Score Thresholds for Treatment Success After Total Knee Arthroplasty. The Journal of Arthroplasty 30, 2154–2158. https://doi.org/10.1016/j.arth.2015.06.012

Giesinger, K., Hamilton, D.F., Jost, B., Holzner, B., Giesinger, J.M., 2014. Comparative responsiveness of outcome measures for total knee arthroplasty. Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society 22, 184–189. https://doi.org/10.1016/j.joca.2013.11.001

Giurea, A., Fraberger, G., Kolbitsch, P., Lass, R., Schneider, E., Kubista, B., Windhager, R., 2016. The Impact of Personality Traits on the Outcome of Total Knee Arthroplasty. BioMed research international 2016, 5282160. https://doi.org/10.1155/2016/5282160

Gower, J.C., 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53, 325–338.

Griffiths, G., Bellamy, N., Kean, W.F., Campbell, J., Gerecz-Simon, E., 1993. A study of the time frame dependency of responses to the WOMAC osteoarthritis index. InflammoPharmacology 2, 85–87. https://doi.org/10.1007/bf02663745

Grood, E.S., Suntay, W.J., 1983. A joint coordinate system for the clinical description of three-dimensional motions: application to the knee. Journal of biomechanical engineering 105, 136–144.

Guo, B., Nixon, M.S., 2009. Gait Feature Subset Selection by Mutual Information. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 39, 36–46. https://doi.org/10.1109/tsmca.2008.2007977

Han, J., Pei, J., Kamber, M., 2011. Data Mining: Concepts and Techniques, Elsevier. Elsevier.

Harding, G.T., Hubley-Kozey, C.L., Dunbar, M.J., Stanish, W.D., Wilson, J.L.A., 2012. Body mass index affects knee joint mechanics during gait differently with and without moderate knee osteoarthritis. Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society 20, 1234–1242. https://doi.org/10.1016/j.joca.2012.08.004

Hatfield, G.L., Hubley-Kozey, C.L., Wilson, J.L.A., Dunbar, M.J., 2011. The effect of total knee arthroplasty on knee joint kinematics and kinetics during gait. The Journal of Arthroplasty 26, 309–318. https://doi.org/10.1016/j.arth.2010.03.021

Hatfield, G.L., Stanish, W.D., Hubley-Kozey, C.L., 2015a. Three-dimensional biomechanical gait characteristics at baseline are associated with progression to total knee arthroplasty. Arthritis care & research 67, 1004–1014. https://doi.org/10.1002/acr.22564

Hatfield, G.L., Stanish, W.D., Hubley-Kozey, C.L., 2015b. Relationship between knee adduction moment patterns extracted using principal component analysis and discrete measures with different amplitude normalizations: Implications for knee osteoarthritis progression studies. Clinical biomechanics (Bristol, Avon) 30, 1146–1152. https://doi.org/10.1016/j.clinbiomech.2015.08.011

Hawker, G., Bohm, E.R., Conner-Spady, B., Coster, C.D., Dunbar, M., Hennigar, A., Loucks, L., Marshall, D.A., Pomey, M.-P., Sanmartin, C., Noseworthy, T., 2015. Perspectives of Canadian Stakeholders on Criteria for Appropriateness for Total Joint Arthroplasty in Patients With Hip and Knee Osteoarthritis. Arthritis & rheumatology (Hoboken, N.J.) 67, 1806–1815. https://doi.org/10.1002/art.39124

Hawker, G.A., Badley, E.M., Borkhoff, C.M., Croxford, R., Davis, A.M., Dunn, S., Gignac, M.A., Jaglal, S.B., Kreder, H.J., Sale, J.E.M., 2013. Which patients are most likely to benefit from total joint arthroplasty? Arthritis & Rheumatism 65, 1243–1252. https://doi.org/10.1002/art.37901

Hawker, G.A., Wright, J.G., Coyte, P.C., Williams, J.I., Harvey, B., Glazier, R., Badley, E.M., 2000. Differences between men and women in the rate of use of hip and knee arthroplasty. New England Journal of Medicine 342, 1016–1022. https://doi.org/10.1056/nejm200004063421405

Henriksen, M., Aaboe, J., Bliddal, H., 2012. The relationship between pain and dynamic knee joint loading in knee osteoarthritis varies with radiographic disease severity. A cross sectional study. The Knee 19, 392–398. https://doi.org/10.1016/j.knee.2011.07.003

Henriksen, M., Graven-Nielsen, T., Aaboe, J., Andriacchi, T.P., Bliddal, H., 2010. Gait changes in patients with knee osteoarthritis are replicated by experimental knee pain. Arthritis care & research 62, 501–509. https://doi.org/10.1002/acr.20033

Hilding, M.B., Lanshammar, H., Ryd, L., 1996. Knee joint loading and tibial component loosening. RSA and gait analysis in 45 osteoarthritic patients before and after TKA. The Journal of bone and joint surgery. British volume 78, 66–73.

Høvik, L.H., Winther, S.B., Foss, O.A., Gjeilo, K.H., 2016. Preoperative pain catastrophizing and postoperative pain after total knee arthroplasty: a prospective cohort study with one year follow-up. BMC Musculoskeletal Disorders 17, 214. https://doi.org/10.1186/s12891-016-1073-0

Hubert, L., 1985. Comparing Partitions. Journal of Classification 2, 193–218. https://doi.org/10.1016/s0021-9673(00)91529-2

Hurwitz, S.R., Slawson, D., Shaughnessy, A., 2000. Orthopaedic information mastery: applying evidence-based information tools to improve patient outcomes while saving orthopaedists' time. The Journal of Bone and Joint Surgery (American) 82, 888–894.

Huskisson, E.C., 1974. Measurement of pain. The Lancet.

Husted, H., Holm, G., Jacobsen, S., 2008. Predictors of length of stay and patient satisfaction after hip and knee replacement surgery: fast-track experience in 712 patients. Acta orthopaedica 79, 168–173. https://doi.org/10.1080/17453670710014941

Ibrahim, S.A., 2010. Racial variations in the utilization of knee and hip joint replacement: an introduction and review of the most recent literature. Current orthopaedic practice 21, 126–131. https://doi.org/10.1097/bco.0b013e3181d08223

ICHOM, I.C. for H.O.M., 2017. Hip & Knee Osteoarthritis Data Collection Reference Guide.

Impellizzeri, F.M., Mannion, A.F., Leunig, M., Bizzini, M., Naal, F.D., 2011. Comparison of the reliability, responsiveness, and construct validity of 4 different

questionnaires for evaluating outcomes after total knee arthroplasty. The Journal of Arthroplasty 26, 861–869. https://doi.org/10.1016/j.arth.2010.07.027

Jackson, D.A., 1993. Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. Ecology Society of America 74, 2204–2214. https://doi.org/10.2307/1939574

Jiang, Y., Sanchez-Santos, M.T., Judge, A.D., Murray, D.W., Arden, N.K., 2017. Predictors of Patient-Reported Pain and Functional Outcomes Over 10 Years After Primary Total Knee Arthroplasty: A Prospective Cohort Study. The Journal of Arthroplasty 32, 92-100.e2. https://doi.org/10.1016/j.arth.2016.06.009

Judge, A., Arden, N.K., Kiran, A., Price, A., 2012. Interpretation of patient-reported outcomes for hip and knee replacement surgery. The Journal of Bone and Joint Surgery (American). https://doi.org/10.1302/0301-620x.94b3

Judge, A., Arden, N.K., Price, A., Glyn-Jones, S., Beard, D., Carr, A.J., Dawson, J., Fitzpatrick, R., Field, R.E., 2011. Assessing patients for joint replacement: can pre-operative Oxford hip and knee scores be used to predict patient satisfaction following joint replacement surgery and to guide patient selection? The Journal of bone and joint surgery. British volume 93, 1660–1664. https://doi.org/10.1302/0301-620x.93b12.27046

Karlson, E.W., Daltroy, L.H., Liang, M.H., Eaton, H.E., Katz, J.N., 1997. Gender differences in patient preferences may underlie differential utilization of elective surgery. The American journal of medicine 102, 524–530. https://doi.org/10.1016/s0002-9343(97)00050-8

Katz, N.P., Paillard, F.C., Ekman, E., 2015. Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions. Journal of orthopaedic surgery and research 10, 24. https://doi.org/10.1186/s13018-014-0144-x

Kellgren, J.H., Lawrence, J.S., 1957. Radiological assessment of osteo-arthrosis. Annals of the rheumatic diseases.

King, L.K., Kendzerska, T., Waugh, E.J., Hawker, G.A., 2018. Impact of Osteoarthritis on Difficulty Walking: A Population-Based Study. Arthritis care & research 70, 71–79. https://doi.org/10.1002/acr.23250

King, M.T., 2011. A point of minimal important difference (MID): a critique of terminology and methods. Expert review of pharmacoeconomics & outcomes research 11, 171–184. https://doi.org/10.1586/erp.11.9

Kittelson, A.J., Stevens-Lapsley, J.E., Schmiege, S.J., 2016. Determination of Pain Phenotypes in Knee Osteoarthritis: A Latent Class Analysis Using Data From the

Osteoarthritis Initiative. Arthritis care & research 68, 612–620. https://doi.org/10.1002/acr.22734

Knol, M.J., Cessie, S.L., Algra, A., Vandenbroucke, J.P., Groenwold, R.H.H., 2012. Overestimation of risk ratios by odds ratios in trials and cohort studies: alternatives to logistic regression. Canadian Medical Association Journal = journal de l'Association medicale canadienne 184, 895–899. https://doi.org/10.1503/cmaj.101715

Knoop, J., Leeden, M. van der, Thorstensson, C.A., Roorda, L.D., Lems, W.F., Knol, D.L., Steultjens, M.P.M., Dekker, J., 2011. Identification of phenotypes with different clinical outcomes in knee osteoarthritis: Data from the osteoarthritis initiative. Arthritis care & research 63, 1535–1542. https://doi.org/10.1002/acr.20571

Kohonen, T., 2013. Essentials of the self-organizing map. Neural networks : the official journal of the International Neural Network Society 37, 52–65. https://doi.org/10.1016/j.neunet.2012.09.018

Kohonen, T., 1990. The self-organizing map. P Ieee 78, 1464–1480. https://doi.org/10.1109/5.58325

Kohonen, T., 1982. Self-Organizing Formation of Topologically Correct Feature Maps. Biological Cybermetrics 43, 59–69.

Kohonen, T., Nieminen, I.T., Honkela, T., 2009. On the Quantization Error in SOM vs. VQ: A Critical and Systematic Study. Proceedings of the 7th International Workshop on Advances in Self-Organizing. https://doi.org/10.1007/978-3-642-02397-2_16

Kurtz, S.M., Lau, E., Ong, K., Zhao, K., Kelly, M., Bozic, K.J., 2009. Future young patient demand for primary and revision joint replacement: national projections from 2010 to 2030. null 467, 2606–2612. https://doi.org/10.1007/s11999-009-0834-6

Kurtz, S.M., Ong, K.L., Lau, E., Widmer, M., Maravic, M., Gómez-Barrena, E., Pina, M. de F. de, Manno, V., Torre, M., Walter, W.L., Steiger, R. de, Geesink, R.G.T., Peltola, M., Röder, C., 2011. International survey of primary and revision total knee replacement. International Orthopaedics 35, 1783–1789. https://doi.org/10.1007/s00264-011-1235-5

Laende, E.K., Wilson, J.L.A., Flemming, J.M., Valstar, E.R., Richardson, C.G., Dunbar, M.J., 2019. Equivalent 2-year stabilization of uncemented tibial component migration despite higher early migration compared with cemented fixation: an RSA study on 360 total knee arthroplasties. Acta Orthop 90, 172–178. https://doi.org/10.1080/17453674.2018.1562633

Landry, S.C., McKean, K.A., Hubley-Kozey, C.L., Stanish, W.D., Deluzio, K.J., 2007. Knee biomechanics of moderate OA patients measured during gait at a self-selected

and fast walking speed. Journal of Biomechanics 40, 1754–1761. https://doi.org/10.1016/j.jbiomech.2006.08.010

Lane, N.E., Brandt, K., Hawker, G., Peeva, E., Schreyer, E., Tsuji, W., Hochberg, M.C., 2011. OARSI-FDA initiative: defining the disease state of osteoarthritis. Osteoarthritis and Cartilage 19, 478–482. https://doi.org/10.1016/j.joca.2010.09.013

Lang, J.J., Alam, S., Cahill, L.E., Drucker, A.M., Gotay, C., Kayibanda, J.F., Kozloff, N., Mate, K.K.V., Patten, S.B., Orpana, H.M., 2018. Global Burden of Disease Study trends for Canada from 1990 to 2016. Cmaj 190, E1296–E1304. https://doi.org/10.1503/cmaj.180698

Lee, J.A., Verleysen, M., 2007. Nonlinear Dimensionality Reduction. Springer New York. https://doi.org/10.1007/978-0-387-39351-3

Li, J., Wyss, U.P., Costigan, P.A., Deluzio, K.J., 1993. An integrated procedure to assess knee-joint kinematics and kinetics during gait using an optoelectric system and standardized X-rays. Journal of biomedical engineering 15, 392–400.

Liu, Q., Niu, J., Li, H., Ke, Y., Li, R., Zhang, Y., Lin, J., 2017. Knee Symptomatic Osteoarthritis, Walking Disability, NSAIDs Use and All-cause Mortality: Population-based Wuchuan Osteoarthritis Study. Sci Rep-uk 7, 3309. https://doi.org/10.1038/s41598-017-03110-3

Losina, E., Katz, J.N., 2013. Total joint replacement outcomes in patients with concomitant comorbidities: a glass half empty or half full? Arthritis & Rheumatology 65, 1157–1159. https://doi.org/10.1002/art.37903

Losina, E., Katz, J.N., 2012. Total knee replacement: pursuit of the paramount result. Rheumatology (Oxford, England) 51, 1735–1736. https://doi.org/10.1093/rheumatology/kes199

Lotke, P.A., Ecker, M.L., 1977. Influence of positioning of prosthesis in total knee replacement. The Journal of Bone and Joint Surgery (American) 59, 77–79.

Lungu, E., Desmeules, F., Dionne, C.E., Belzile, É.L., Vendittoli, P.-A., 2014. Prediction of poor outcomes six months following total knee arthroplasty in patients awaiting surgery. BMC Musculoskeletal Disorders 15, 299. https://doi.org/10.1186/1471-2474-15-299

Lungu, E., Vendittoli, P.-A., Desmeules, F., 2015. Identification of patients with suboptimal results after hip arthroplasty: development of a preliminary prediction algorithm. BMC Musculoskeletal Disorders 1–10. https://doi.org/10.1186/s12891-015-0720-1

Mahfouz, M., Fatah, E.E.A., Bowers, L.S., Scuderi, G., 2011. Three-dimensional Morphology of the Knee Reveals Ethnic Differences. null 470, 172–185. https://doi.org/10.1007/s11999-011-2089-2

Maly, M.R., Costigan, P.A., Olney, S.J., 2008. Mechanical factors relate to pain in knee osteoarthritis. Clinical Biomechanics 23, 796–805. https://doi.org/10.1016/j.clinbiomech.2008.01.014

Maly, M.R., Costigan, P.A., Olney, S.J., 2006. Determinants of self efficacy for physical tasks in people with knee osteoarthritis. Arthritis & Rheumatism 55, 94–101. https://doi.org/10.1002/art.21701

Mancino, F., Cacciola, G., Malahias, M.-A., Filippis, R.D., Marco, D.D., Matteo, V.D., Gu, A., Sculco, P.K., Maccauro, G., Martino, I.D., 2020. What are the benefits of robotic-assisted total knee arthroplasty over conventional manual total knee arthroplasty? A systematic review of comparative studies. Orthop Rev 12, 8657. https://doi.org/10.4081/or.2020.8657

Mancuso, C.A., Sculco, T.P., Wickiewicz, T.L., Jones, E.C., Robbins, L., Warren, R.F., Williams-Russo, P., 2001. Patients' expectations of knee surgery. The Journal of Bone and Joint Surgery (American) 83-A, 1005–1012.

Mannion, A.F., Kämpfen, S., Munzinger, U., Quervain, I.K., 2009. The role of patient expectations in predicting outcome after total knee arthroplasty. Arthritis research & therapy 11, R139. https://doi.org/10.1186/ar2811

McClelland, J.A., Webster, K.E., Feller, J.A., Menz, H.B., 2011. Knee kinematics during walking at different speeds in people who have undergone total knee replacement. The Knee 18, 151–155. https://doi.org/10.1016/j.knee.2010.04.005

McClelland, J.A., Webster, K.E., Feller, J.A., Menz, H.B., 2010. Knee kinetics during walking at different speeds in people who have undergone total knee replacement. Gait & Posture 32, 205–210. https://doi.org/10.1016/j.gaitpost.2010.04.009

McHorney, C.A., Jr, J.E.W., Lu, J., Sherbourne, C.D., 1994. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Medical care. https://doi.org/10.2307/3766189

McHorney, C.A., Jr, J.E.W., Raczek, A.E., 1993. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Medical care. https://doi.org/10.2307/3765819

McKean, K.A., Landry, S.C., Hubley-Kozey, C.L., Dunbar, M.J., Stanish, W.D., Deluzio, K.J., 2007. Gender differences exist in osteoarthritic gait. JCLB 22, 400–409. https://doi.org/10.1016/j.clinbiomech.2006.11.006

Milligan, G.W., Cooper, M.C., 1988. A study of standardization of variables in cluster analysis. Journal of Classification 5, 181–204. https://doi.org/10.1007/bf01897163

Miyazaki, T., Wada, M., Kawahara, H., Sato, M., Baba, H., Shimada, S., 2002. Dynamic load at baseline can predict radiographic disease progression in medial compartment knee osteoarthritis. Annals of the rheumatic diseases 61, 617–622. https://doi.org/10.1136/ard.61.7.617

Mooney, L.T., Smith, A., Sloan, K., Clark, G.W., 2016. The effect of the native kinematics of the knee on the outcome following total knee arthroplasty. The bone & joint journal. https://doi.org/10.1302/0301-620x.98b11

Moreland, J.R., Bassett, L.W., Hanker, G.J., 1987. Radiographic analysis of the axial alignment of the lower extremity. The Journal of Bone and Joint Surgery (American) 69, 745–749.

Naal, F.D., Impellizzeri, F.M., Leunig, M., 2009. Which is the best activity rating scale for patients undergoing total joint arthroplasty? null 467, 958–965. https://doi.org/10.1007/s11999-008-0358-5

Naili, J.E., Wretenberg, P., Lindgren, V., Iversen, M.D., Hedström, M., Broström, E.W., 2017. Improved knee biomechanics among patients reporting a good outcome in knee-related quality of life one year after total knee arthroplasty. BMC Musculoskeletal Disorders 18, 122. https://doi.org/10.1186/s12891-017-1479-3

Nilashi, M., Ahmadi, H., Sheikhtaheri, A., Naemi, R., Alotaibi, R., Alarood, A.A., Munshi, A., Rashid, T.A., Zhao, J., 2020. Remote tracking of Parkinson's Disease progression using ensembles of Deep Belief Network and Self-Organizing Map. Expert Syst Appl 159, 113562. https://doi.org/10.1016/j.eswa.2020.113562

Noble, P.C., Conditt, M.A., Cook, K.F., Mathis, K.B., 2006. The John Insall Award: Patient Expectations Affect Satisfaction with Total Knee Arthroplasty. Clinical orthopaedics and related research 452, 35–43. https://doi.org/10.1097/01.blo.0000238825.63648.1e

Noble, P.C., Gordon, M.J., Weiss, J.M., Reddix, R.N., Conditt, M.A., Mathis, K.B., 2005. Does Total Knee Replacement Restore Normal Knee Function? Clinical Orthopaedics and Related Research NA;, 157–165. https://doi.org/10.1097/01.blo.0000150130.03519.fb

Noble, P.C., Scuderi, G.R., Brekke, A.C., Sikorskii, A., Benjamin, J.B., Lonner, J.H., Chadha, P., Daylamani, D.A., Scott, W.N., Bourne, R.B., 2011. Development of a New Knee Society Scoring System. null 470, 20–32. https://doi.org/10.1007/s11999-011-2152-z

Orishimo, K.F., Kremenic, I.J., Deshmukh, A.J., Nicholas, S.J., Rodriguez, J.A., 2012. Does total knee arthroplasty change frontal plane knee biomechanics during gait? Clinical Orthopaedics and Related Research 470, 1171–1176. https://doi.org/10.1007/s11999-011-2196-0

Outerleys, J., Dunbar, M.J., Richardson, G., Hubley-Kozey, C.L., Wilson, J.L.A., 2021. Quantifying achievable levels of improvement in knee joint biomechanics during gait after total knee arthroplasty relative to osteoarthritis severity. Journal of applied biomechanics 37, 1–31.

Parratte, S., Pagnano, M.W., Trousdale, R.T., Berry, D.J., 2010. Effect of postoperative mechanical axis alignment on the fifteen-year survival of modern, cemented total knee replacements. The Journal of Bone and Joint Surgery (American) 92, 2143–2149. https://doi.org/10.2106/jbjs.i.01398

Paterson, K.L., Sosdian, L., Bennell, K.L., Metcalf, B.R., Wrigley, T.V., Kasza, J., Dowsey, M.M., Choong, P.F., Hinman, R.S., 2020. The influence of sex and pre-operative obesity on biomechanics two years after total knee arthroplasty_ A longitudinal cohort study. Gait & Posture 76, 74–84. https://doi.org/10.1016/j.gaitpost.2019.10.031

Paterson, K.L., Sosdian, L., Hinman, R.S., Wrigley, T.V., Kasza, J., Dowsey, M., Choong, P., Bennell, K.L., 2018. Effects of sex and obesity on gait biomechanics before and six months after total knee arthroplasty_ A longitudinal cohort study. Gait & Posture 61, 263–268. https://doi.org/10.1016/j.gaitpost.2018.01.014

Paterson, K.L., Sosdian, L., Hinman, R.S., Wrigley, T.V., Kasza, J., Dowsey, M., Choong, P., Bennell, K.L., 2017. The influence of sex and obesity on gait biomechanics in people with severe knee osteoarthritis scheduled for arthroplasty. Clinical biomechanics (Bristol, Avon) 49, 72–77. https://doi.org/10.1016/j.clinbiomech.2017.08.013

Pinedo-Villanueva, R., Khalid, S., Wylde, V., Gooberman-Hill, R., Soni, A., Judge, A., 2018. Identifying individuals with chronic pain after knee replacement: a population-cohort, cluster-analysis of Oxford knee scores in 128,145 patients from the English National Health Service. BMC Musculoskeletal Disorders 19, 354. https://doi.org/10.1186/s12891-018-2270-9

Quintana, J.M., Escobar, A., Arostegui, I., Bilbao, A., Azkárate, J., Goenaga, J.I., Arenaza, J.C., 2006. Health-related quality of life and appropriateness of knee or hip joint replacement. Archives of internal medicine 166, 220–226. https://doi.org/10.1001/archinte.166.2.220

Ravi, B., Croxford, R., Reichmann, W.M., Losina, E., Katz, J.N., Hawker, G.A., 2012. The changing demographics of total joint arthroplasty recipients in the United States and Ontario from 2001 to 2007. Best Practice & Research Clinical Rheumatology 26, 637–647. https://doi.org/10.1016/j.berh.2012.07.014

Robertsson, O., Dunbar, M., Pehrsson, T., Knutson, K., Lidgren, L., 2000. Patient satisfaction after knee arthroplasty: a report on 27,372 knees operated on between 1981 and 1995 in Sweden. Acta orthopaedica Scandinavica 71, 262–267. https://doi.org/10.1080/000164700317411852

Robertsson, O., Dunbar, M.J., 2001. Patient satisfaction compared with general health and disease-specific questionnaires in knee arthroplasty patients. Journal of Arthroplasty 16, 476–482. https://doi.org/10.1054/arth.2001.22395a

Rodriguez, J.A., Bas, M.A., Orishimo, K.F., Robinson, J., Nicholas, S.J., 2016. Differential Effect of Total Knee Arthroplasty on Valgus and Varus Knee Biomechanics During Gait. The Journal of Arthroplasty 31, 248–253. https://doi.org/10.1016/j.arth.2016.06.061

Rolfson, O., Chenok, K.E., Bohm, E., Lübbeke, A., Denissen, G., Dunn, J., Lyman, S., Franklin, P., Dunbar, M., Overgaard, S., Garellick, G., Dawson, J., Registries, P.-R.O.M.W.G. of the I.S. of A., 2016. Patient-reported outcome measures in arthroplasty registries. Acta orthopaedica 87 Suppl 1, 3–8. https://doi.org/10.1080/17453674.2016.1181815

Roos, E.M., 2018. 3 steps to improve reporting and interpretation of patient-reported outcome scores in orthopedic studies. Acta orthopaedica 89, 1–2. https://doi.org/10.1080/17453674.2017.1407058

Roos, E.M., Roos, H.P., Lohmander, L.S., Ekdahl, C., 1998. Knee Injury and Osteoarthritis Outcome Score (KOOS)—development of a self-administered outcome measure. Journal of Orthopaedic & … 28, 88–96. https://doi.org/10.2519/jospt.1998.28.2.88

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Ryd, L., Albrektsson, B.E., Carlsson, L., Dansgård, F., Herberts, P., Lindstrand, A., Regnér, L., Toksvig-Larsen, S., 1995. Roentgen stereophotogrammetric analysis as a predictor of mechanical loosening of knee prostheses. The Journal of bone and joint surgery. British volume 77, 377–383. https://doi.org/10.1007/978-1-4939-2645-9

Salzmann, M., Fennema, P., Becker, R., Hommel, H., 2017. Does Postoperative Mechanical Axis Alignment Have an Effect on Clinical Outcome of Primary Total

Knee Arthroplasty? A Retrospective Cohort Study. The open orthopaedics journal 11, 1330–1336. https://doi.org/10.2174/1874325001711011330

Schaible, H.-G., 2012. Mechanisms of chronic pain in osteoarthritis. Current rheumatology reports 14, 549–556. https://doi.org/10.1007/s11926-012-0279-x

Schipplein, O.D., Andriacchi, T.P., 1991. Interaction between active and passive knee stabilizers during level walking. null 9, 113–119. https://doi.org/10.1002/jor.1100090114

Schmucker, A.M., Hupert, N., Mandl, L.A., 2019. The Impact of Frailty on Short-Term Outcomes After Elective Hip and Knee Arthroplasty in Older Adults: A Systematic Review. Geriatric Orthop Surg Rehabilitation 10, 2151459319835109. https://doi.org/10.1177/2151459319835109

Schnitzer, T.J., Popovich, J.M., Andersson, G.B.J., Andriacchi, T.P., 1993. Effect of piroxicam on gait in patients with osteoarthritis of the knee. Arthritis Rheumatism 36, 1207–1213. https://doi.org/10.1002/art.1780360905

Scott, C., Howie, C.R., MacDonald, D., 2010. Predicting dissatisfaction following total knee replacement. Bone & Joint … 92-B, 1253–1258. https://doi.org/10.1302/0301-620x.92b9

Scott, C.E.H., Bugler, K.E., Clement, N.D., MacDonald, D., Howie, C.R., Biant, L.C., 2012. Patient expectations of arthroplasty of the hip and knee. The Journal of bone and joint surgery. British volume 94, 974–981. https://doi.org/10.1302/0301-620x.94b7.28219

Services, C. for M. and M., 2015. Federal Register. Department of Health and Human Services, Centers for Medicare & Medicaid Services.

Singer, A.J., Jr., H.C.T., 1998. Determination of the Minimal Clinically Significant Difference on a Patient Visual Analog Satisfaction Scale. Acad Emerg Med 5, 1007–1011. https://doi.org/10.1111/j.1553-2712.1998.tb02781.x

Skou, S.T., Roos, E.M., Laursen, M.B., Rathleff, M.S., Arendt-Nielsen, L., Simonsen, O., Rasmussen, S., 2015. A Randomized, Controlled Trial of Total Knee Replacement. The New England journal of medicine 373, 1597–1606. https://doi.org/10.1056/nejmoa1505467

Smith, A.J., Lloyd, D.G., Wood, D.J., 2004. Pre-surgery knee joint loading patterns during walking predict the presence and severity of anterior knee pain after total knee arthroplasty. Journal of orthopaedic research : official publication of the Orthopaedic Research Society 22, 260–266. https://doi.org/10.1016/s0736-0266(03)00184-0

Sosdian, L., Dobson, F., Wrigley, T.V., Paterson, K., Bennell, K., Dowsey, M., Choong, P., Allison, K., Hinman, R.S., 2014. Longitudinal changes in knee kinematics and moments following knee arthroplasty: a systematic review. The Knee 21, 994–1008. https://doi.org/10.1016/j.knee.2014.09.009

Spil, W.E. van, Bierma-Zeinstra, S.M.A., Deveza, L.A., Arden, N.K., Bay-Jensen, A.C., Kraus, V.B., Carlesso, L., Christensen, R., Esch, M. van der, Kent, P., Knoop, J., Ladel, C., Little, C.B., Loeser, R.F., Losina, E., Mills, K., Mobasheri, A., Nelson, A.E., Neogi, T., Peat, G.M., Rat, A.C., Steultjens, M., Thomas, M.J., Valdes, A.M., Hunter, D.J., 2020. A consensus-based framework for conducting and reporting osteoarthritis phenotype research. Arthritis research & therapy 22, 1926–7. https://doi.org/10.1186/s13075-020-2143-0

Srikanth, V.K., Fryer, J.L., Zhai, G., Winzenberg, T.M., Hosmer, D., Jones, G., 2005. A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. Osteoarthritis and Cartilage 13, 769–781. https://doi.org/10.1016/j.joca.2005.04.014

Stevens-Lapsley, J.E., Balter, J.E., Kohrt, W.M., Eckhoff, D.G., 2010. Quadriceps and Hamstrings Muscle Dysfunction after Total Knee Arthroplasty. null 468, 2460–2468. https://doi.org/10.1007/s11999-009-1219-6

Sullivan, M., Bishop, S.R., Pivik, J., 1995. The pain catastrophizing scale: development and validation. Psychological assessment. https://doi.org/10.1037/1040-3590.7.4.524

Sullivan, M., Tanzer, M., Reardon, G., Amirault, D., Dunbar, M., Stanish, W., 2011. The role of presurgical expectancies in predicting pain and function one year following total knee arthroplasty. Pain 152, 2287–2293. https://doi.org/10.1016/j.pain.2011.06.014

Thorp, L.E., Sumner, D.R., Wimmer, M.A., Block, J.A., 2007. Relationship between pain and medial knee joint loading in mild radiographic knee osteoarthritis. Arthritis & Rheumatism 57, 1254–1260. https://doi.org/10.1002/art.22991

Trost, S.G., Fragala-Pinkham, M., Lennon, N., O'Neil, M.E., 2016. Decision Trees for Detection of Activity Intensity in Youth with Cerebral Palsy. Medicine Sci Sports Exerc 48, 958–966. https://doi.org/10.1249/mss.0000000000000842

Tubach, F., Ravaud, P., Baron, G., Falissard, B., Logeart, I., Bellamy, N., Bombardier, C., Felson, D., Hochberg, M., Heijde, D. van der, Dougados, M., 2005. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Annals of the rheumatic diseases 64, 29–33. https://doi.org/10.1136/ard.2004.022905

Turcot, K., Armand, S., Lübbeke, A., Fritschy, D., Hoffmeyer, P., Suvà, D., 2012. Does knee alignment influence gait in patients with severe knee osteoarthritis? Clinical Biomechanics 28, 34–39. https://doi.org/10.1016/j.clinbiomech.2012.09.004

Turcot, K., Sagawa, Y., Fritschy, D., Hoffmeyer, P., Suvà, D., Armand, S., 2013. How gait and clinical outcomes contribute to patients' satisfaction three months following a total knee arthroplasty. The Journal of Arthroplasty 28, 1297–1300. https://doi.org/10.1016/j.arth.2013.01.031

Vanlommel, L., Vanlommel, J., Claes, S., Bellemans, J., 2013. Slight undercorrection following total knee arthroplasty results in superior clinical outcomes in varus knees. Knee Surgery, Sports Traumatology, Arthroscopy 21, 2325–2330. https://doi.org/10.1007/s00167-013-2481-4

Varagunam, M., Hutchings, A., Neuburger, J., Black, N., 2014. Impact on hospital performance of introducing routine patient reported outcome measures in surgery. Journal of health services research & policy 19, 77–84. https://doi.org/10.1177/1355819613506187

Vaughan, C.L., Davis, B.L., Jeremy, C.O., 1999. Dynamics of Human Gait 2nd Edition, Gait & Posture. Kiboho Publishers. https://doi.org/10.1016/j.gaitpost.2009.06.017

Vissers, M.M., Groot, I.B. de, Reijman, M., Bussmann, J.B., Stam, H.J., Verhaar, J.A., 2010. Functional capacity and actual daily activity do not contribute to patient satisfaction after total knee arthroplasty. BMC Musculoskeletal Disorders 11, 121. https://doi.org/10.1186/1471-2474-11-121

Waarsing, J.H., Bierma-Zeinstra, S.M.A., Weinans, H., 2015. Distinct subtypes of knee osteoarthritis: data from the Osteoarthritis Initiative. Rheumatology 54, 1650–1658. https://doi.org/10.1093/rheumatology/kev100

Ward, J.H., 1963. Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association 58, 236–244.

Wehrens, R., Buydens, L.M.C., 2007. Self- and Super-organizing Maps in R: The kohonen Package. Journal of Statistical Software 21, 1–19. https://doi.org/10.18637/jss.v021.i05

Westerlund, M.L., 2005. Classification of Kohonen Self-Organizing Maps. Soft Computing, Haskoli Islands 1–16.

Whitehouse, S.L., Lingard, E.A., Katz, J.N., 2003. Development and testing of a reduced WOMAC function scale. Bone & Joint …. https://doi.org/10.1302/0301-620x.85b5.13681

Williams, D.P., O'Brien, S., Doran, E., Price, A.J., Beard, D.J., Murray, D.W., Beverland, D.E., 2013a. Early postoperative predictors of satisfaction following total knee arthroplasty. The Knee 20, 442–446. https://doi.org/10.1016/j.knee.2013.05.011

Williams, D.P., Price, A.J., Beard, D.J., Hadfield, S.G., Arden, N.K., Murray, D.W., Field, R.E., 2013b. The effects of age on patient-reported outcome measures in total knee replacements. Bone Jt J 95-B, 38–44. https://doi.org/10.1302/0301-620x.95b1.28061

Wilson, J.L.A., Deluzio, K.J., Dunbar, M.J., Caldwell, G.E., Hubley-Kozey, C.L., 2011. The association between knee joint biomechanics and neuromuscular control and moderate knee osteoarthritis radiographic and pain severity. Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society 19, 186–193. https://doi.org/10.1016/j.joca.2010.10.020

Wilson, J.L.A., Dunbar, M.J., Hubley-Kozey, C.L., 2015. Knee joint biomechanics and neuromuscular control during gait before and after total knee arthroplasty are sex-specific. The Journal of Arthroplasty 30, 118–125. https://doi.org/10.1016/j.arth.2014.07.028

Wilson, J.L.A., Lamontagne, M., Wilson, D.R., Beaulé, P.E., Mwale, F., Yee, A., 2019. Specific Functional Analysis: The Key to the Next Revolution Towards the Treatment of Hip and Knee Osteoarthritis. Journal of Orthopaedic Research 37, 1–6. https://doi.org/10.1002/jor.24317

Wilson, J.L.A., Stanish, W.D., Hubley-Kozey, C.L., 2017. Asymptomatic and symptomatic individuals with the same radiographic evidence of knee osteoarthritis walk with different knee moments and muscle activity. null 35, 1661–1670. https://doi.org/10.1002/jor.23465

Wilson, J.L.A., Wilson, D.A., Dunbar, M.J., Deluzio, K.J., 2010. Preoperative gait patterns and BMI are associated with tibial component migration. Acta orthopaedica 81, 478–486. https://doi.org/10.3109/17453674.2010.501741

Wright, C.J., Chambers, G.K., Robens-Paradise, Y., 2002. Evaluation of indications for and outcomes of elective surgery. Canadian Medical Association Journal = journal de l'Association medicale canadienne 167, 461–466.

Yakobov, E., Scott, W., Stanish, W., Dunbar, M., Richardson, G., Sullivan, M., 2014. The role of perceived injustice in the prediction of pain and function after total knee arthroplasty. Pain 155, 2040–2046. https://doi.org/10.1016/j.pain.2014.07.007

Yakobov, E., Stanish, W., Tanzer, M., Dunbar, M., Richardson, G., Sullivan, M.J.L., 2018. The prognostic value of pain catastrophizing in health-related quality of life judgments after Total knee arthroplasty. Health and quality of life outcomes 16, 126. https://doi.org/10.1186/s12955-018-0955-2

Young, K.L., Dunbar, M.J., Richardson, G., Wilson, J.L.A., 2015. Intraoperative passive knee kinematics during total knee arthroplasty surgery. Journal of orthopaedic research : official publication of the Orthopaedic Research Society 33, 1611–1619. https://doi.org/10.1002/jor.22945

Young-Shand, K., Roy, P., Abidi, S.R., Dunbar, M., Wilson, J.A., 2019. Clinical and Biomechanical Cluster Classification Before TKA Impacts Functional Outcome - Women's Health Advisory Board of the American Academy of Orthopaedic Surgeons Award; Implant Section Award, in: Knee - Arthroplasty and Other Surgical Interventions. Presented at the Orthopaedic Research Society Conference, Austin Texas.

Young-Shand, K.L., Dunbar, M.J., Wilson, J.L.A., 2020. Individual Gait Features Are Associated with Clinical Improvement After Total Knee Arthroplasty. JBJS Open Access 5, e0038-11. https://doi.org/10.2106/jbjs.oa.19.00038

Zou, G., 2004. A modified poisson regression approach to prospective studies with binary data. American journal of epidemiology 159, 702–706. https://doi.org/10.1093/aje/kwh090

Appendix A Supplementary Materials for Chapter 4

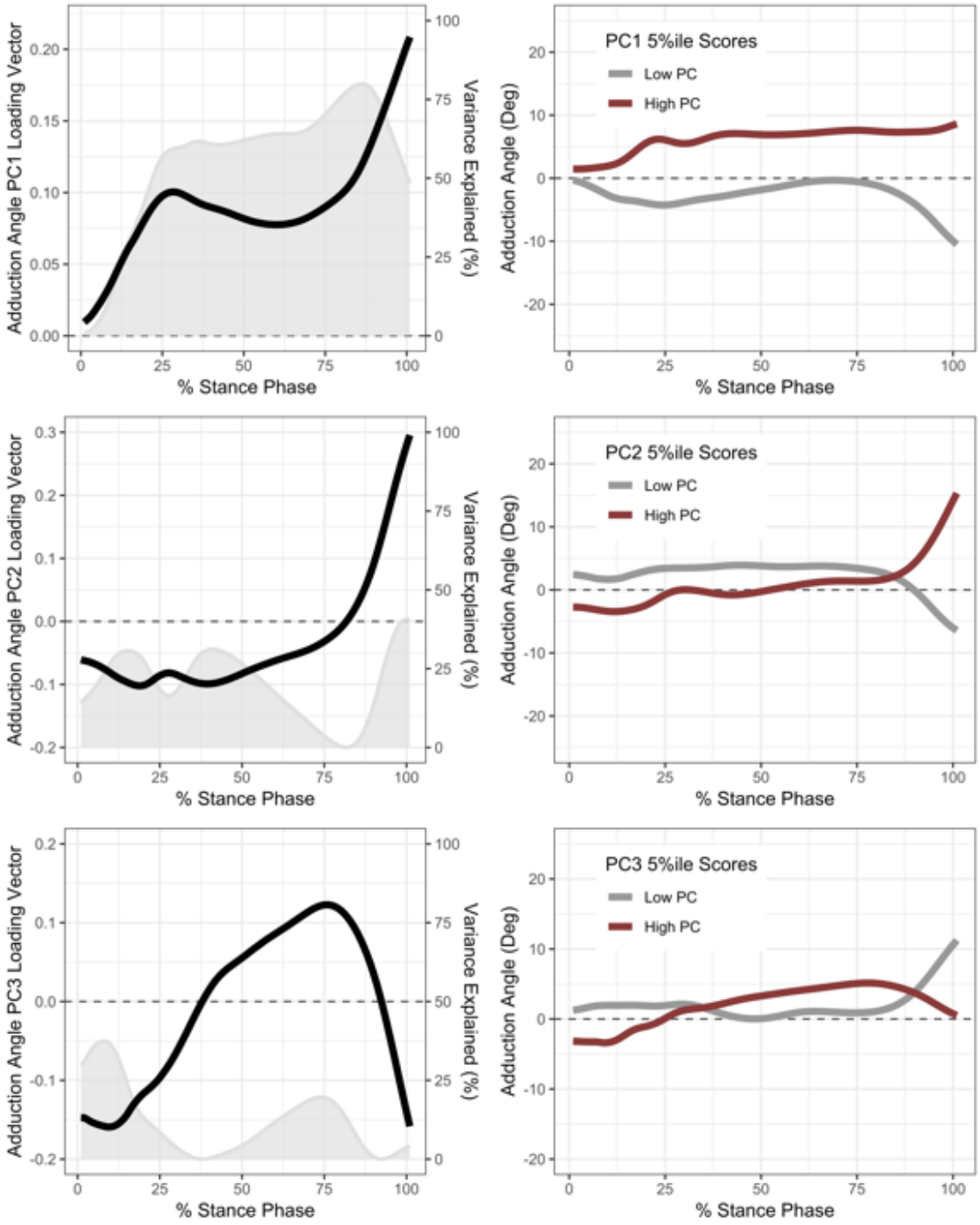

Figure A.1. Knee adduction angle principal components. Left) loading vectors (solid) and corresponding variance explained (grey shaded region) across the stance-phase of one gait cycle for PCs 1-3. Right) Example waveforms with high (95th percentile) and low (5th percentile) PC scores across the stance-phase of one gait cycle for PCs 1-3.

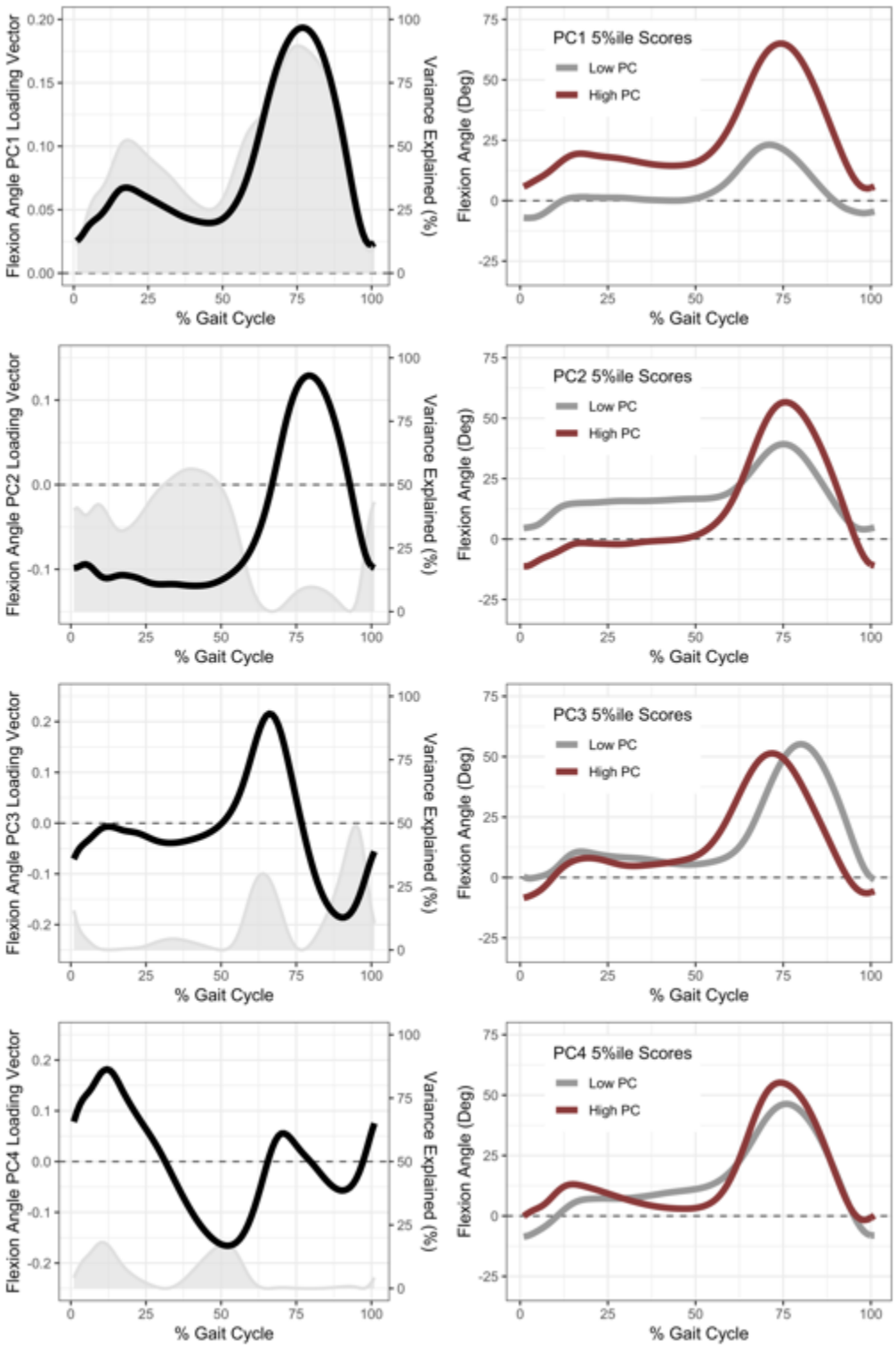

Figure A.2. Knee flexion angle principal components. Left) loading vectors (solid) and corresponding variance explained (grey shaded region) across one complete gait cycle for PCs 1-4. Right) Example waveforms with high (95th percentile) and low (5th percentile) PC scores across one complete gait cycle for PCs 1-4.

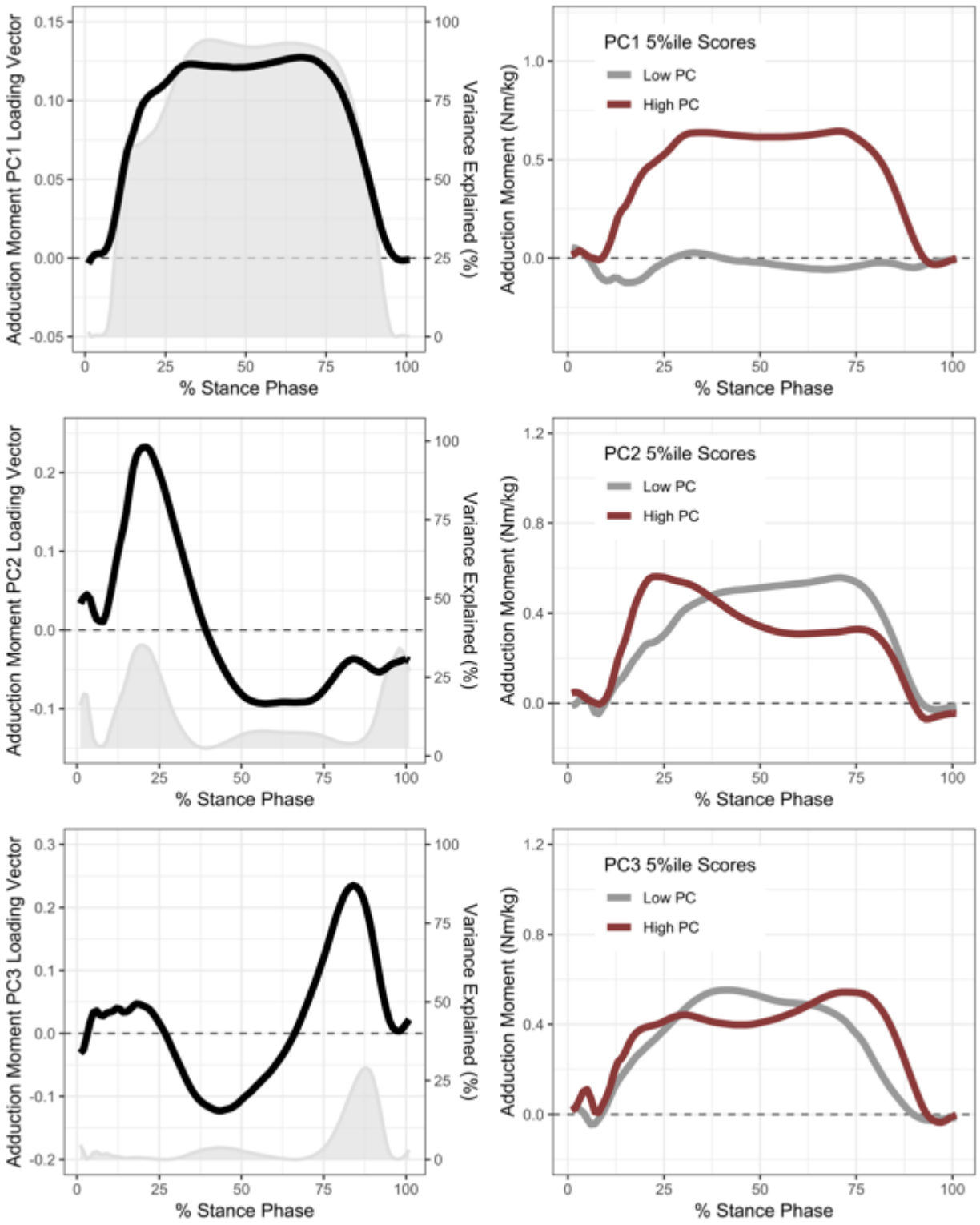

Figure A.3. Knee adduction moment principal components. Left) loading vectors (solid) and corresponding variance explained (grey shaded region) across the stance-phase of one gait cycle for PCs 1-3. Right) Example waveforms with high (95th percentile) and low (5th percentile) PC scores across the stance-phase of one gait cycle for PCs 1-3.

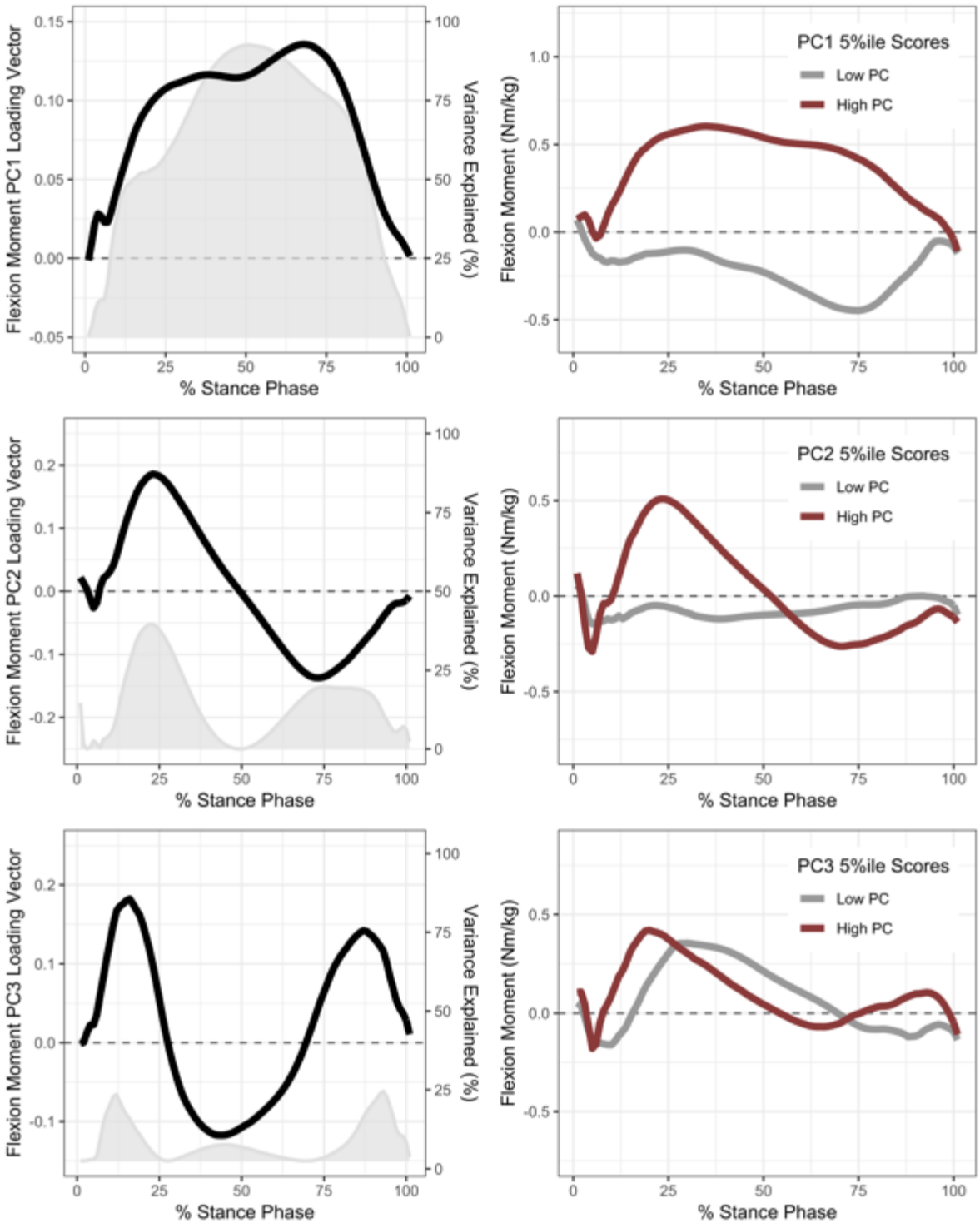

Figure A.4. Knee flexion moment principal components. Left) loading vectors (solid) and corresponding variance explained (grey shaded region) across the stance-phase of one gait cycle for PCs 1-3. Right) Example waveforms with high (95th percentile) and low (5th percentile) PC scores across the stance-phase of one gait cycle for PCs 1-3.

Appendix B Supplementary Materials for Chapter 5

Table B.1. Mean silhouette coefficient, *s*, and Adjusted Rand Index, ARI, for clustering solutions 2:10. Bold indicates best clustering solution.

Feature	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
s	0.31	0.35	**0.37**	0.35	0.36	0.33	0.34	0.35	0.35
ARI	0.15	0.54	**0.57**	0.55	0.53	0.49	0.48	0.49	0.49

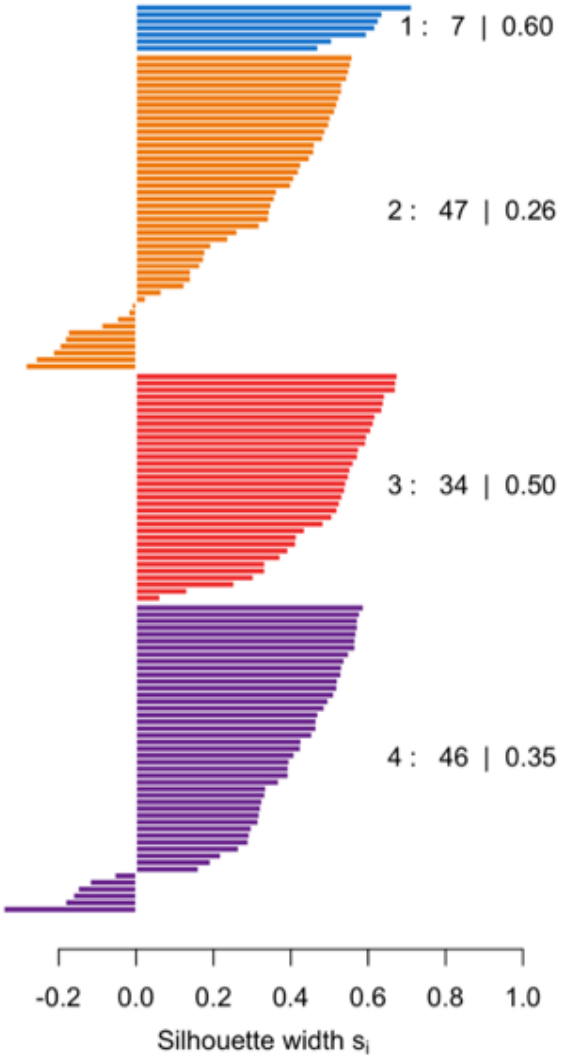

Figure B.1. Silhouette plot of accepted clustering solution. Figure illustrates $n_{\text{patients in cluster}} | s_{\text{mean cluster silhouette width}}$ for 134 patients among cluster 1 (blue, higher functioning males), cluster 2 (orange, lower functioning males), cluster 3 (red, lower functioning females), and cluster 4 (purple, higher functioning females).

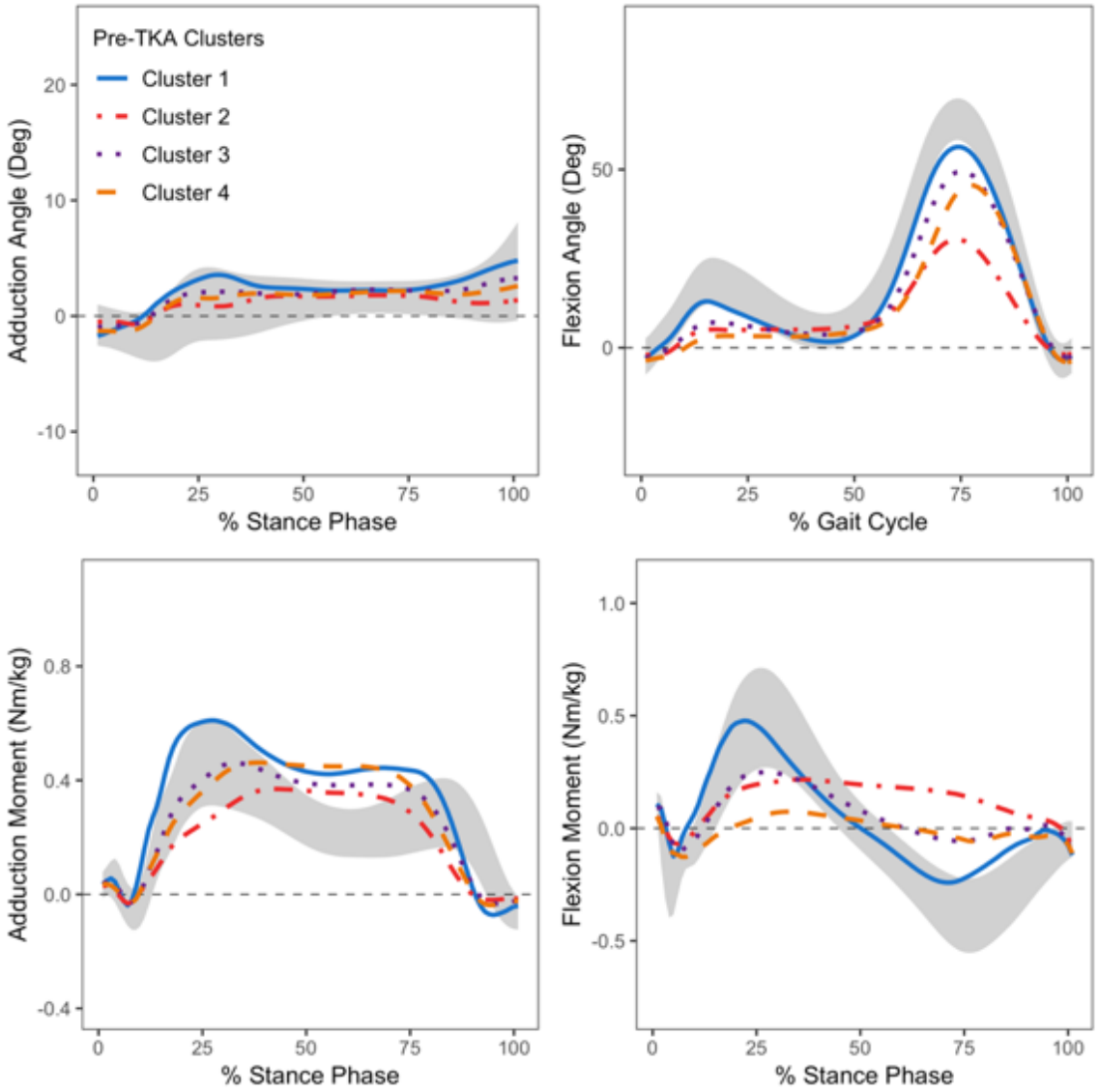

Figure B.2. Mean pre-TKA gait waveforms by cluster; Cluster 1 (solid), Cluster 2 (dot-dashed), Cluster 3 (dotted), and Cluster 4 (dashed) when sex is removed from clustering analysis prior to MDS coordinate definition. Grey denotes 209 previously captured asymptomatic (mean age 51.6 ±9.0, BMI 26.7±4.5) gait patterns (mean ± 1SD).

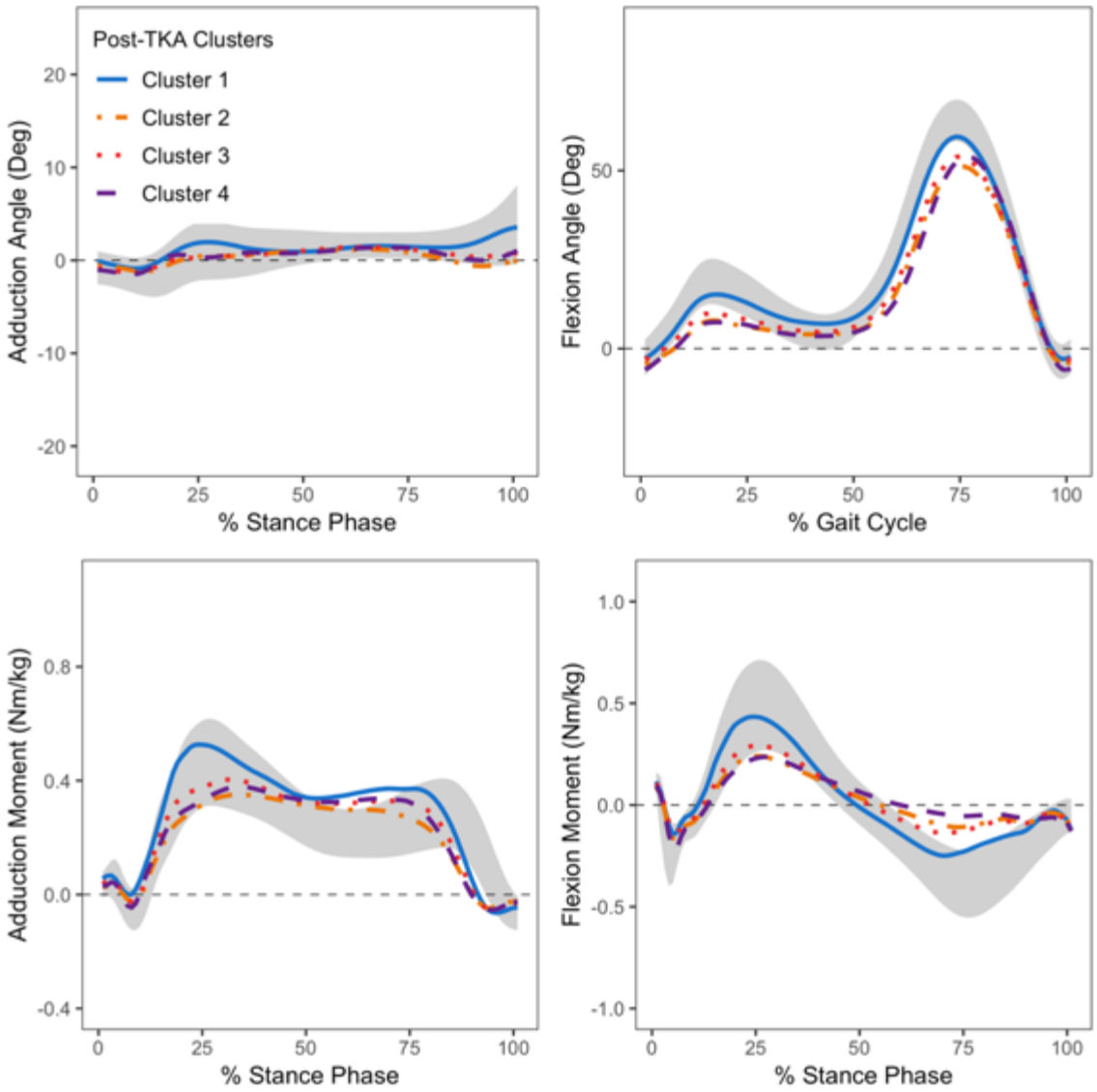

Figure B.3. Mean post-TKA gait waveforms by cluster; Cluster 1 (solid), Cluster 2 (dot-dashed), Cluster 3 (dotted), and Cluster 4 (dashed) when sex is removed from clustering analysis prior to MDS coordinate definition. Grey denotes 209 previously captured asymptomatic (mean age 51.6 ±9.0, BMI 26.7±4.5) gait patterns (mean ± 1SD).

Appendix C Supplementary Materials for Chapter 6

Table C.1. Mean (± 1 standard deviation) demographic, spatiotemporal, and PC score gait waveform features between the clinical groups applied to cluster analysis. Four-way ANOVA tests examined mean differences between groups. Full dataset (n=923).

	PC Var Exp.	All Clinical Groups	Asymptomatic	Moderate OA	Pre-TKA	Post-TKA	*p*
Sex							
Female, n (%)		454 (49)	146 (62)	114 (33)	79 (54)	115 (57)	
Male, n (%)		469 (51)	90 (38)	227 (67)	66 (46)	86 (43)	
Age (years)		59.36 ± 10.22	51.55 ± 10.74	59.19 ± 8.45	63.75 ± 7.96	65.64 ± 7.28	<0.001
BMI (kg/m^2)		30.63 ± 5.93	26.80 ± 4.59	30.75 ± 5.22	33.02 ± 5.96	33.20 ± 6.11	<0.001
Stance Percent (%)		64.85 ± 2.32	63.43 ± 1.67	64.53 ± 1.96	66.25 ± 2.52	66.08 ± 2.21	<0.001
Speed (m/s)		1.18 ± 0.24	1.34 ± 0.17	1.24 ± 0.19	0.91 ± 0.21	1.09 ± 0.18	<0.001
Flexion Angle							
PC1	65%	-1.87 ± 61.04	41.61 ± 43.92	5.91 ± 49.22	-63.54 ± 63.18	-21.61 ± 49.37	<0.001
PC2	14%	1.39 ± 28.30	-4.14 ± 24.50	3.01 ± 27.16	-3.18 ± 36.78	8.41 ± 25.57	<0.001
PC3	11%	-1.54 ± 26.03	2.79 ± 21.29	5.41 ± 24.49	-14.48 ± 27.31	-9.07 ± 27.41	<0.001
PC4	5%	-0.04 ± 17.72	4.46 ± 17.37	0.07 ± 20.67	-2.65 ± 12.80	-3.63 ± 14.40	<0.001
PC5	3%	1.06 ± 13.00	0.57 ± 12.04	3.31 ± 14.72	-2.64 ± 11.74	0.49 ± 11.04	<0.001
Adduction Angle							<0.001
PC1	56%	-3.09 ± 20.35	1.03 ± 19.44	-3.27 ± 20.54	1.80 ± 18.90	-11.16 ± 19.67	<0.001
PC2	26%	-1.72 ± 13.38	4.24 ± 12.50	-5.45 ± 13.26	-4.11 ± 14.03	-0.69 ± 11.39	<0.001
PC5	2%	-0.36 ± 3.45	0.68 ± 3.07	-0.85 ± 3.61	-1.37 ± 3.69	-0.00 ± 3.07	<0.001
Flexion Moment							<0.001
PC1	50%	0.05 ± 1.34	-0.18 ± 0.94	-0.16 ± 1.32	0.66 ± 1.75	0.25 ± 1.24	<0.001
PC2	38%	-0.09 ± 1.17	0.80 ± 0.87	-0.11 ± 1.29	-0.97 ± 0.72	-0.46 ± 0.77	<0.001
PC3	4%	-0.01 ± 0.37	0.09 ± 0.35	-0.07 ± 0.37	0.05 ± 0.39	-0.06 ± 0.32	<0.001
PC4	2%	0.00 ± 0.29	0.10 ± 0.25	0.06 ± 0.31	-0.16 ± 0.23	-0.09 ± 0.27	<0.001
Adduction Moment							<0.001
PC1	71%	0.07 ± 1.13	-0.30 ± 0.78	0.42 ± 1.07	0.26 ± 1.55	-0.23 ± 0.97	<0.001
PC2	15%	-0.02 ± 0.52	0.26 ± 0.44	0.10 ± 0.53	-0.47 ± 0.41	-0.24 ± 0.38	<0.001
PC3	6%	0.01 ± 0.34	0.11 ± 0.35	0.03 ± 0.36	-0.16 ± 0.30	-0.01 ± 0.27	<0.001
PC4	3%	-0.02 ± 0.24	0.06 ± 0.24	-0.02 ± 0.25	-0.07 ± 0.22	-0.09 ± 0.19	<0.001
PC5	2%	0.01 ± 0.17	0.01 ± 0.17	0.03 ± 0.20	-0.02 ± 0.15	-0.02 ± 0.14	0.0069

Table C.2. Mean (± 1 standard deviation) for all PC score gait waveform features between clusters. Three-way ANOVA and Tukey HSD criteria examined mean differences between clusters. Training dataset (n=878).

Interpretation (PC; Variance Explained)	Cluster 1	Cluster 2	Cluster 3	*p*	Tukey HDS
Flexion Angle					
Higher score: greater overall flexion angle magnitude throughout the gait cycle (PC1; 65%)	32.46(44.6)	-45.47(60.7)	-4.79(46.8)	**<0.001**	**2<3<1**
Higher scores: Less difference between mean stance magnitude and flexion peak (PC2; 14%)	-3.74(23.1)	6.24(33.3)	3.62(29.0)	**<0.001**	**1<2,3**
Higher scores: more knee extension in heal strike and late stance with earlier occurring peak flexion in swing (PC3; 11%)	5.65(20.5)	-17.50(28.3)	8.43(21.9)	**<0.001**	**2<1,3**
Higher score: more flexion and extension angle range in stance (PC4; 5%)	8.46(16.7)	-2.09(14.1)	-12.75(15.5)	**<0.001**	**3<2<1**
Not interpreted (PC5; 3%)	0.72(12.3)	2.18(12.5)	0.76(15.2)	0.3	
Adduction Angle					
Higher score: greater overall adduction angle magnitude in stance (PC1; 56%)	0.76(19.9)	-6.91(21.3)	-4.00(19.0)	**<0.001**	**2,3<1**
Higher score: more abduction to adduction angle range in stance (PC2; 26%)	1.46(12.6)	-3.69(12.9)	-5.76(14.2)	**<0.001**	**2,3<1**
Not interpreted (PC5; 2%)	-0.26(3.3)	-0.59(3.7)	-0.37(3.5)	0.5	
Flexion Moment					
Higher score: Greater overall flexion moment magnitude in stance (PC1; 50%)	-0.25(1.0)	0.60(1.6)	-0.11(1.3)	**<0.001**	**1,3<2**
Higher scores: more flexion to extension moment range in stance (PC2; 38%)	0.83(0.9)	-0.80(0.8)	-0.81(0.9)	**<0.001**	**2,3<1**
Higher scores: earlier occurring flexion moment loading acceptance (PC3, 4%)	0.03(0.4)	-0.08(0.4)	0.02(0.3)	**<0.001**	**2<1,3**
Higher score: more extension to flexion moment range from heal strike to early stance (PC4; 2%)	0.060.3	-0.13(0.2)	0.09(0.3)	**<0.001**	**2<1,3**

Adduction Moment					
Higher scores: greater overall adduction moment magnitude in stance (PC1; 71%)	0.04(1.0)	0.02(1.2)	0.22(1.2)	0.09	
Higher score: more first peak to mid-stance adduction moment range in stance (PC2; 15%)	0.32(0.5)	-0.43(0.3)	-0.05(0.4)	**<0.001**	**2<3<1**
Higher score: more mid-stance to second peak adduction moment range in stance (PC3; 6%)	0.15(0.4)	-0.13(0.3)	-0.06(0.3)	**<0.001**	**2<3<1**
Higher score: heal strike to weight acceptance adduction moment range (PC4; 3%)	-0.01(0.3)	-0.08(0.2)	0.03(0.2)	**<0.001**	**2<1,3**
Higher score: more heal strike and terminal stance abduction moment magnitude (PC5; 2%)	0.02(0.2)	0.01(0.1)	-0.02(0.2)	**0.030**	**3<1**

Table C.3. Hierarchical clustering on the 10 best performing SOM models defined in Figure C.6. Here, cluster performance was defined by counts of positive and negative node silhouette values (left) and mean silhouette width (right) for k clusters 2:5.

	k=3		**k=4**		**k=5**	
	Negative Nodes	**Positive Nodes**	**Negative Nodes**	**Positive Nodes**	**Negative Nodes**	**Positive Nodes**
Model 1	-16	105	-25	96	-24	97
Model 2	-9	112	-11	110	-20	101
Model 3	-13	108	-19	102	-20	101
Model 4	-18	103	-20	101	-16	105
Model 5	-12	109	-7	114	-10	111
Model 6	-10	111	-16	105	-23	98
Model 7	-12	109	-14	107	-10	111
Model 8	-9	112	-15	106	-15	106
Model 9	-13	108	-13	108	-13	108
Model 10	**-5**	**116**	-9	112	-19	102

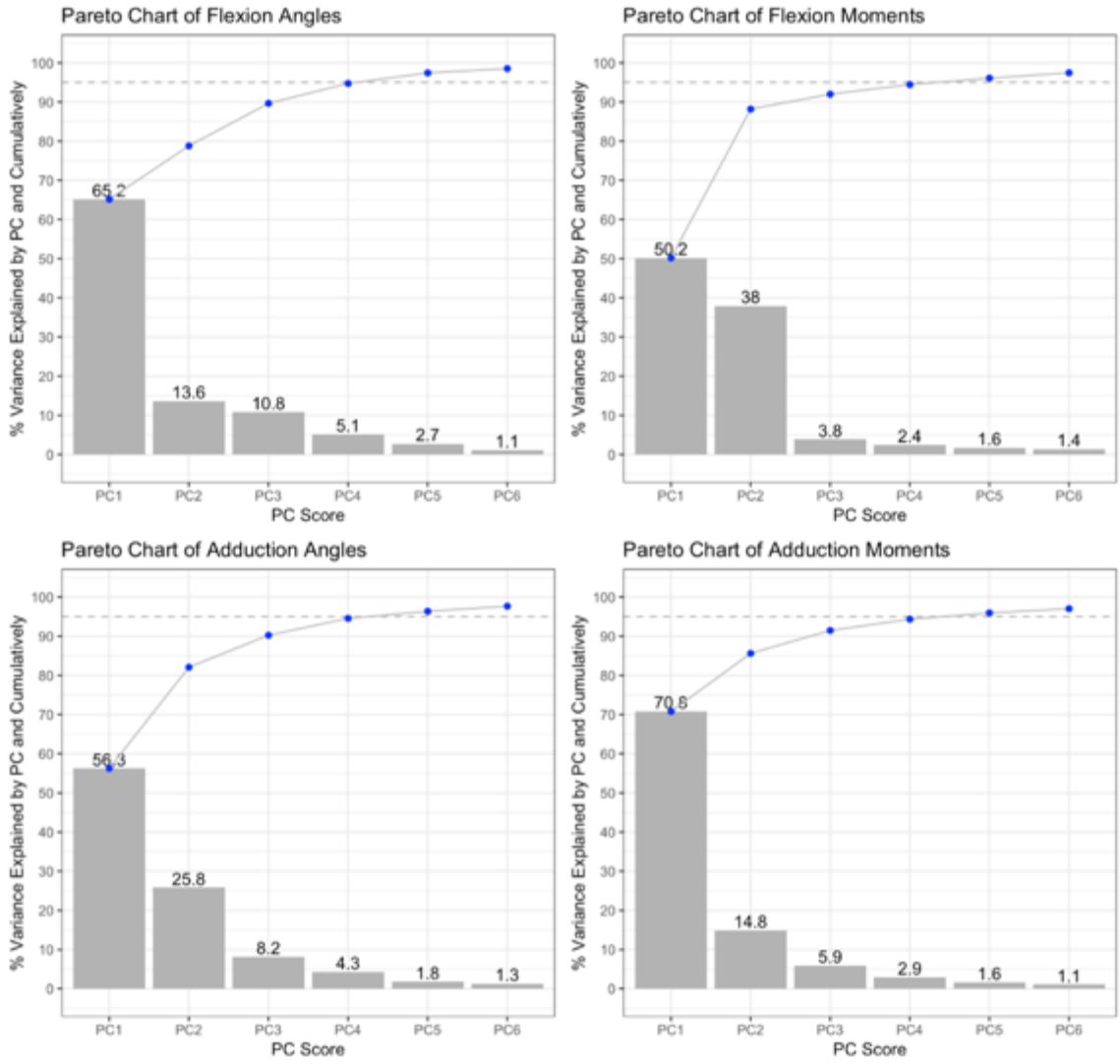

Figure C.1. Pareto / Scree Plots illustrating PC score variance explained. Five PCs were retained for flexion angle (cumulatively 97.4% variance explained), adduction angle (96.4%), flexion moment (96.1%) and adduction moment (95.6%) features.

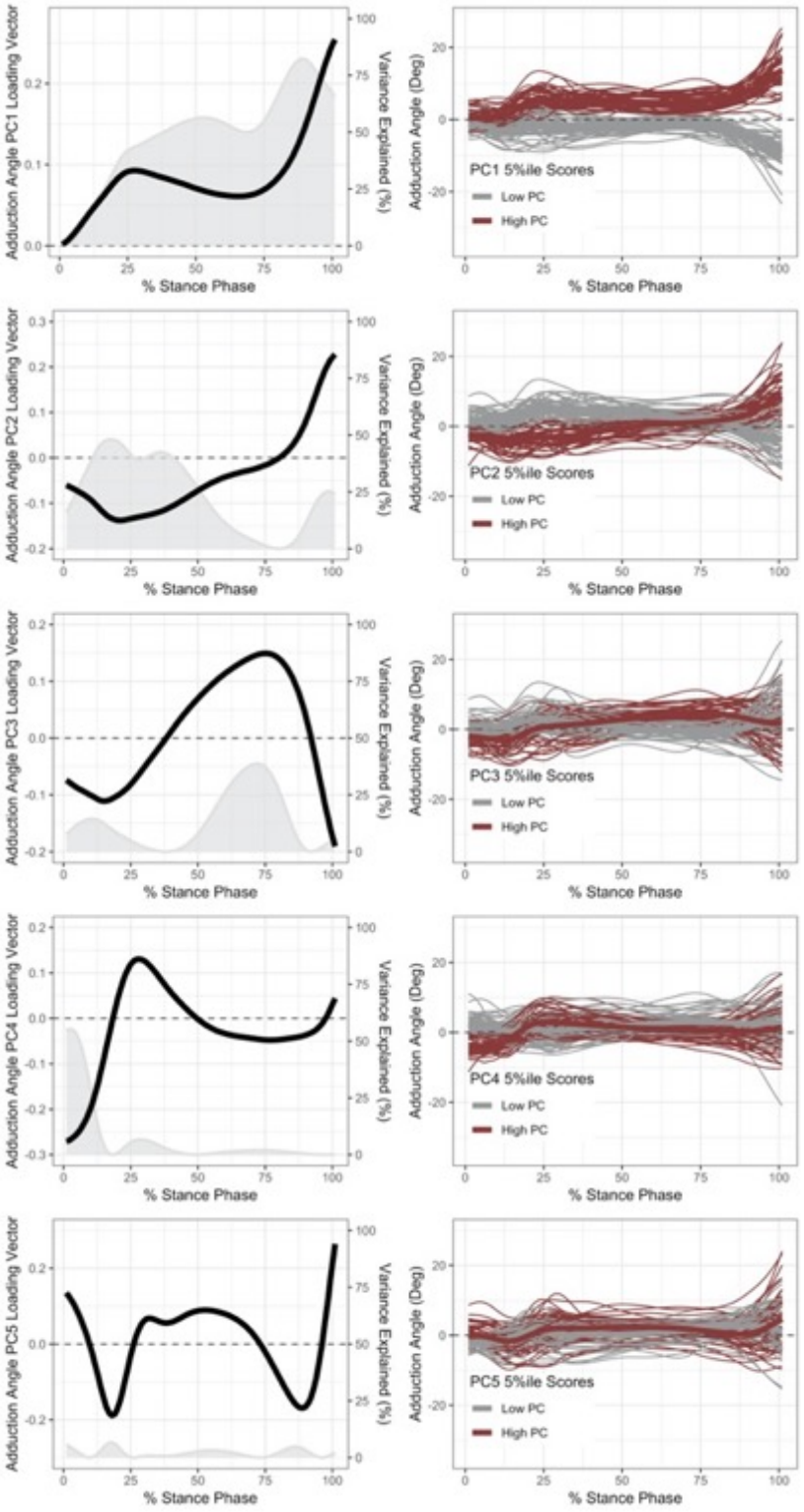

Figure C.2. Knee adduction angle principal components. Left) loading vectors (solid) and corresponding variance explained (grey shaded region) across the stance-phase of one gait cycle for PCs 1-5. Right) Example waveforms with high (95th percentile) and low (5th percentile) PC scores across the stance-phase of one gait cycle for PCs 1-5.

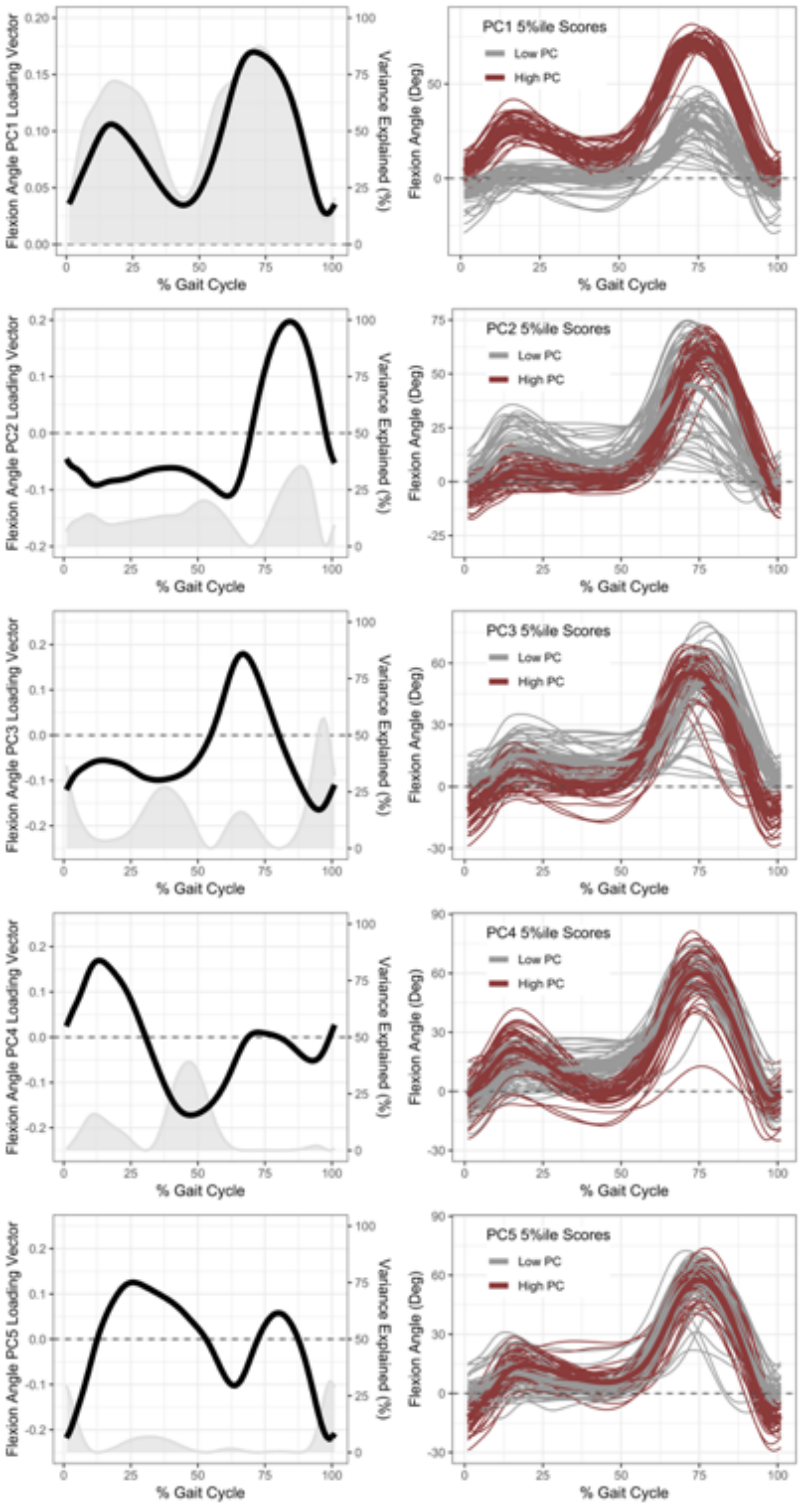

Figure C.3. Knee flexion angle principal components. Left) loading vectors (solid) and corresponding variance explained (grey shaded region) across one complete gait cycle for PCs 1-5. Right) Example waveforms with high (95th percentile) and low (5th percentile) PC scores across one complete gait cycle for PCs 1-5.

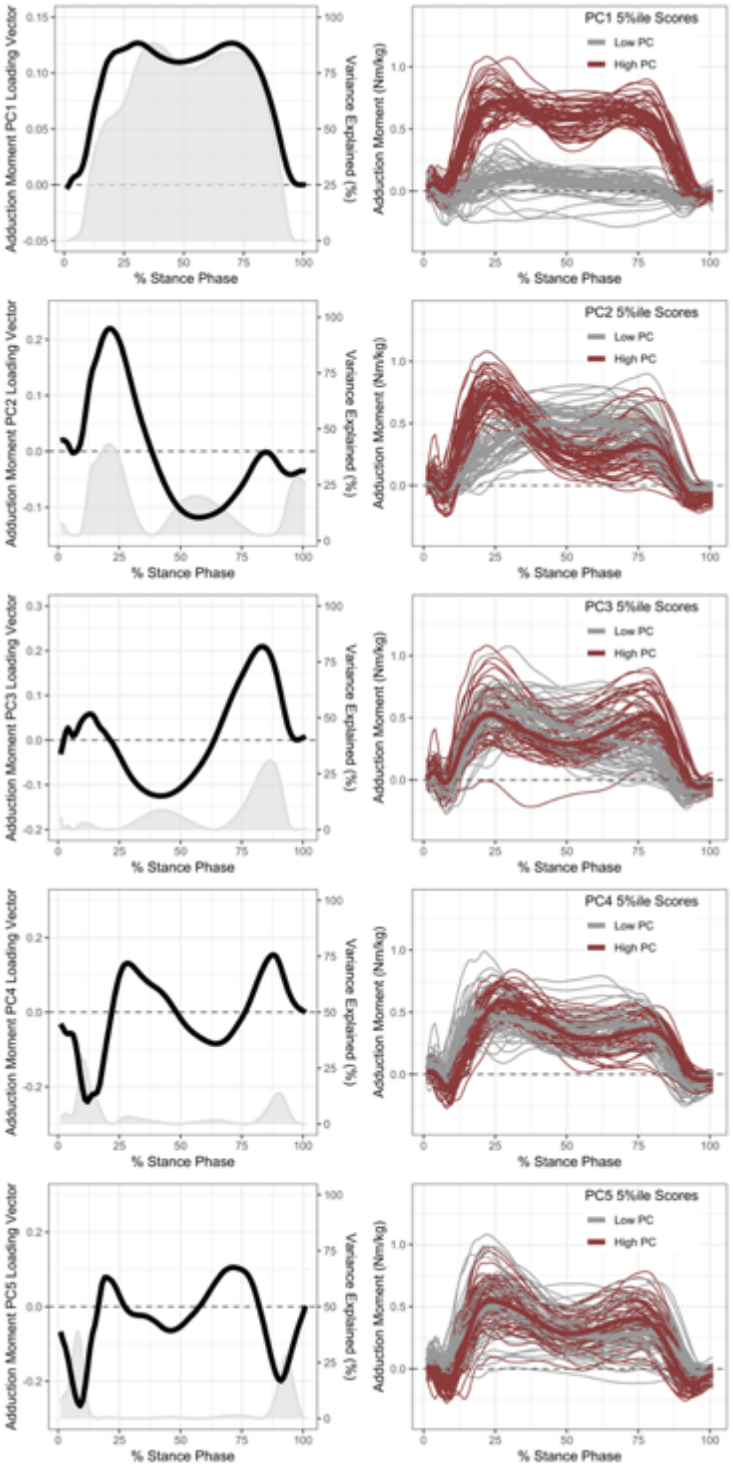

Figure C.4. Knee adduction moment principal components. Left) loading vectors (solid) and corresponding variance explained (grey shaded region) across the stance-phase of one gait cycle for PCs 1-5. Right) Example waveforms with high (95th percentile) and low (5th percentile) PC scores across the stance-phase of one gait cycle for PCs 1-5.

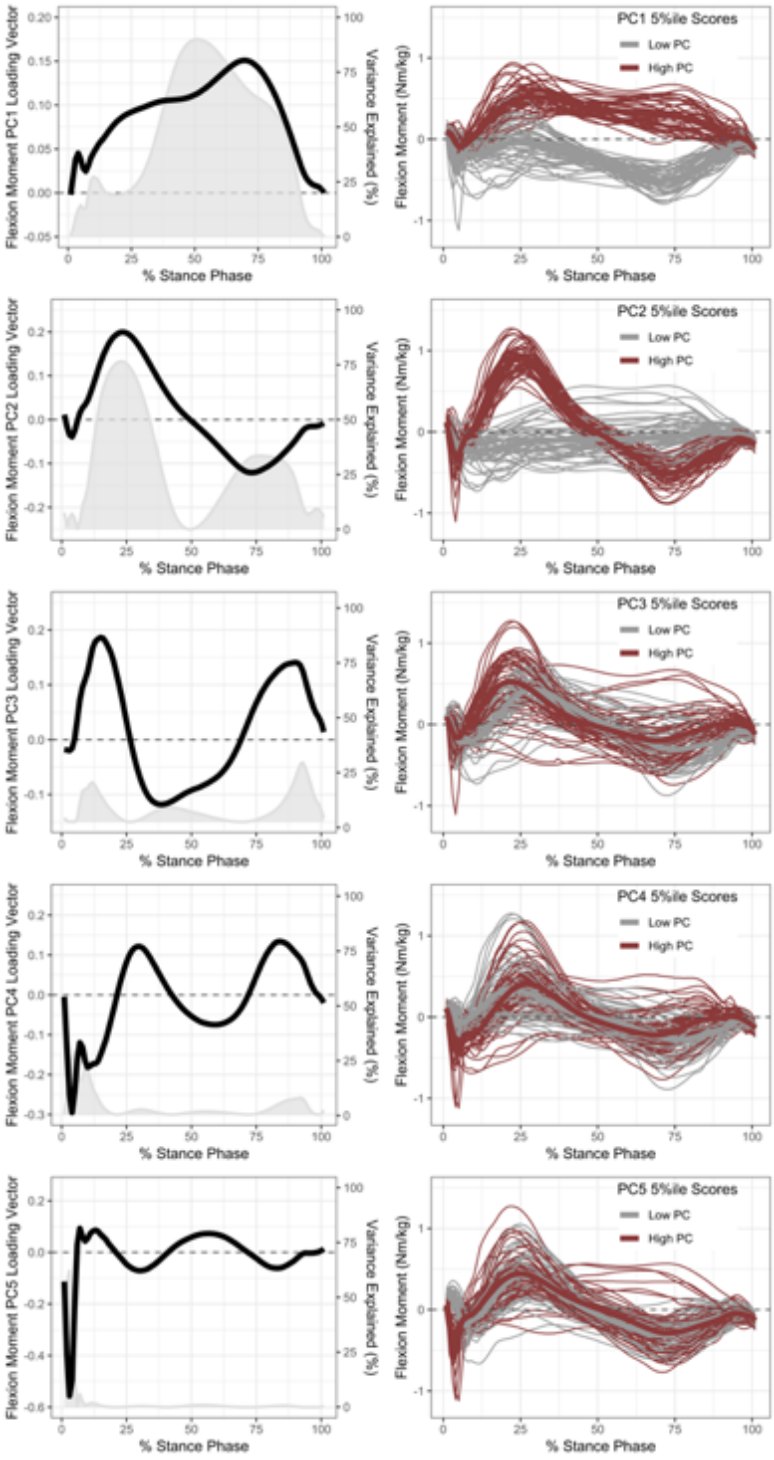

Figure C.5. Knee flexion moment principal components. Left) loading vectors (solid) and corresponding variance explained (grey shaded region) across the stance-phase of one gait cycle for PCs 1-5. Right) Example waveforms with high (95th percentile) and low (5th percentile) PC scores across the stance-phase of one gait cycle for PCs 1-5.

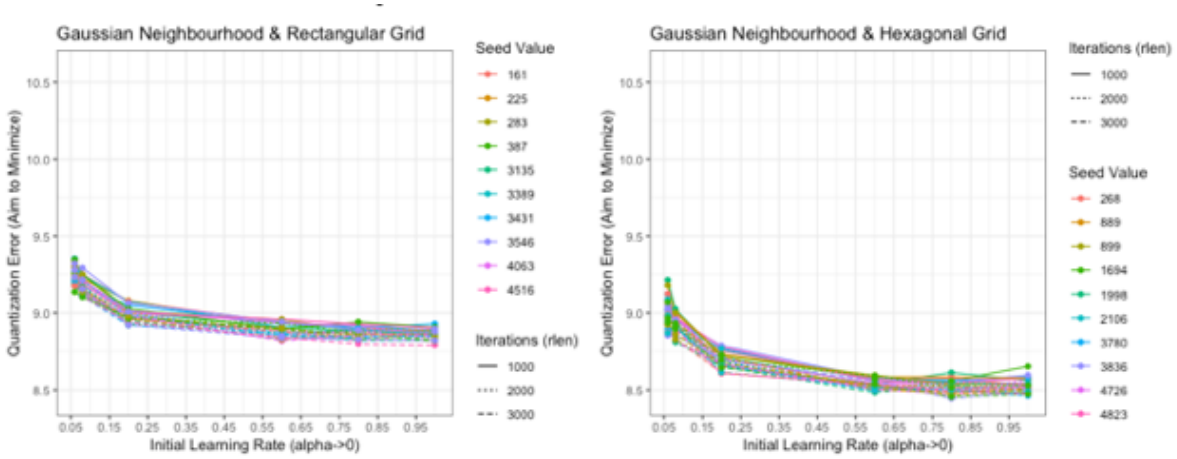

Figure C.6. SOM training under a number of initializing and tuning conditions, including training time (t, defined as "rlen" above), initial learning rate, and (l, defined as "alpha" above), under differing grid criteria for randomly selected seed values. The 10 best performing SOM models by quantization error (qe) performance criterion where retained.

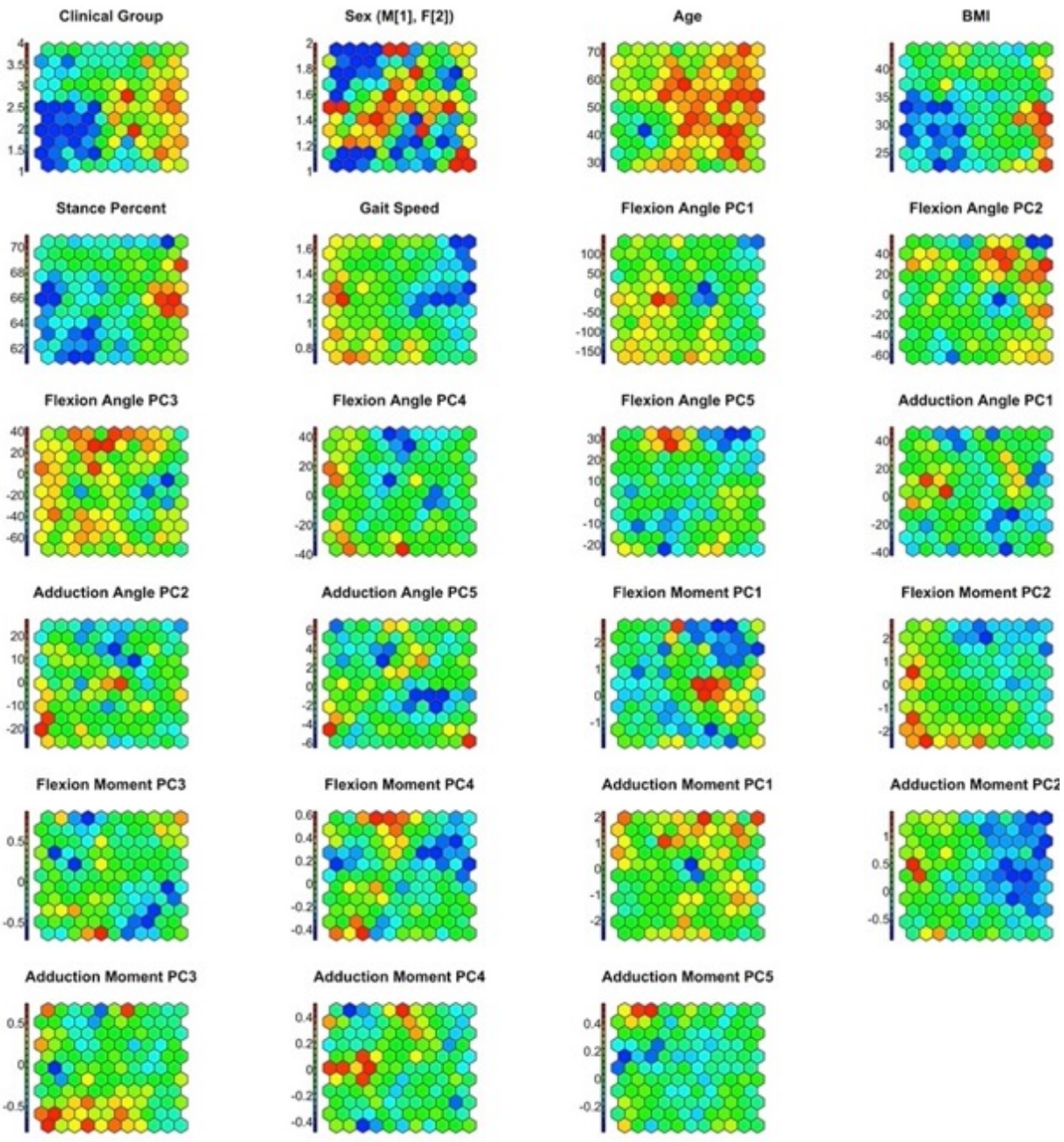

Figure C.7. Heat map of SOM at the feature level: Clinical Group, descriptive of 1[asymptomatic], 2[moderate OA], 3[pre-TKA], 4[post-TKA] clinical classification (not applied to clustering); Sex, denoted as 1[male] and 2[female] (not applied to clustering); Age, participant age at the time of the gait observation in years; BMI, body mass index at the time of the gait observation in kg/m^2; Stance Percent, percent of gait cycle in stance; Gait Speed, in m/s; Flexion Angle, Adduction Angle, Flexion Moment and Adduction Moment PC scores applied to clustering. Colours correspond to feature bar legends, low[blue]-high[red].

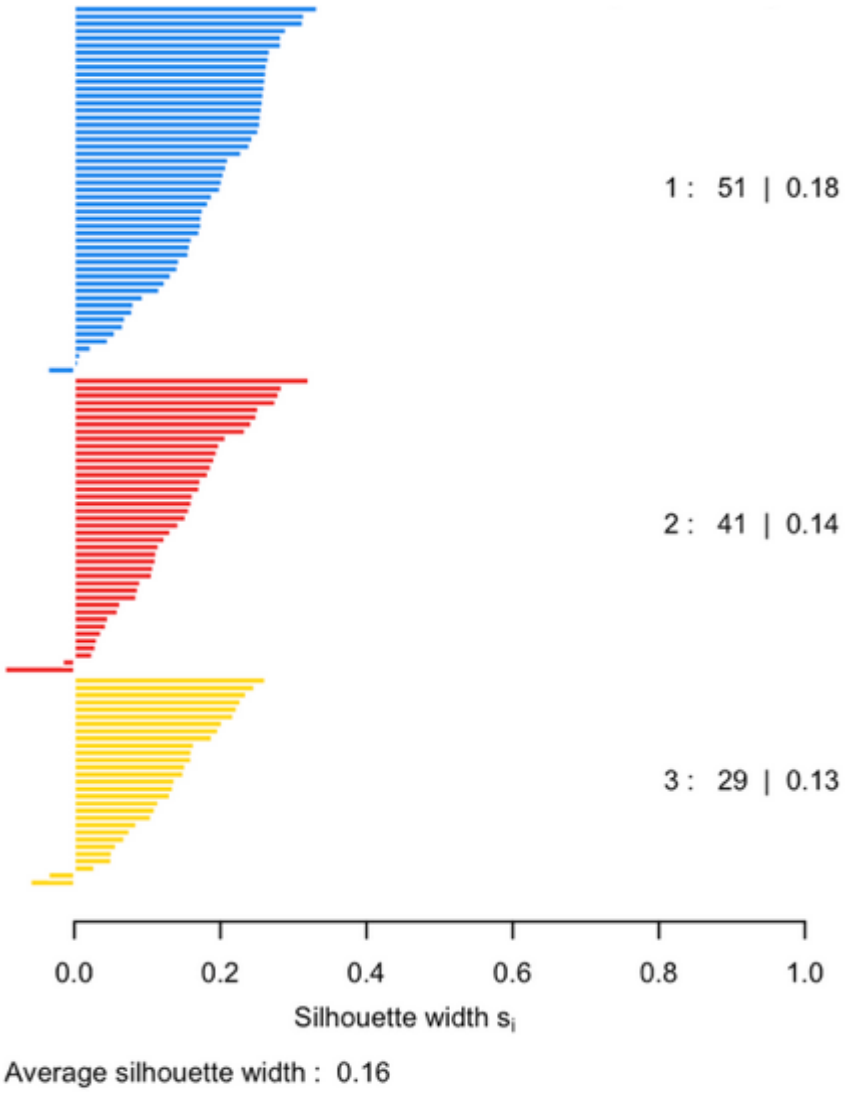

Figure C.8. Silhouette plot of accepted SOM clustering solution, where s_i is the silhouette width of each node for clusters 1 to 3. The number of nodes belonging to each cluster, and the mean silhouette width of each cluster are also provided (nodes in cluster k | mean silhouette width of cluster).

Appendix D Abstract: SOM Applied to Self-Reported Data

Submitted to the Canadian Orthopaedic Association/Canadian Orthopaedic Research Society Conference 2021:

Phenotypes Associated with Self-Reported Pain, Function and Pain Catastrophizing Among TKA Populations Using Machine Learning Based Self-Organizing Maps

Purpose: Patient-reported outcome measures (PROMs) of osteoarthritis (OA) patients before and after Total Knee Arthroplasty (TKA) allows patients to share their self-reported perceptions of pain, quality of life, and function; meaningful insights into patient health. However, responses can be influenced by external dimensions. Understanding variability and "phenotypes" within patient-reported responses is relevant to interpretation and tailoring interventions to the individual. The objective of this study was to use a data mining framework to map demographic and PROM response variability among patients pre and post-TKA, and to identify patient phenotypes within the map.

Methods: Primary and revision TKA patients (n=876) completed joint-specific Oxford Knee Score (OKS, 0[worst]-48[best]) and Pain Catastrophizing (PCS, 0[least]-52[most]) questionnaires pre-TKA (n=608), and post-TKA at 6-months (n=84), one year (n=145), and two years (n=39). Of the full dataset (j=876), training (j=834), and test sets (j=42) were defined. Using the training dataset, patient age at the time of surgery, OKS (questions 1-12), and PCS (questions 1-13) were applied to an unsupervised Self-Organizing Map (SOM) framework followed by hierarchical clustering. The ten best performing SOM models by quantization error were applied to hierarchical clustering. Cluster models (k=2:5) were assessed by counts of negative nodes using silhouette coefficients, where the cluster model with the least negative nodes was selected. Clusters were validated by examining inter-cluster differences by chi-squared and k-way ANOVA, Kruskal-Wallis, and Fisher Tests.

Results: A SOM model with four (k=4) clusters yielded optimum performance metrics. All clusters differed significantly in OKS dimensions of Pain ($p<0.001$), Function ($p<0.001$) and Total scores ($p<0.001$), and PCS dimension of Helplessness ($p<0.001$), Rumination ($p<0.001$), Magnification ($p<0.001$), and Total scores ($p<0.001$). In general, Cluster 1 was interpreted as a "high catastrophizing, poor pain and function phenotype", with the highest PCS and lowest (worst) OKS among the clusters. Of the pre-TKA patients in the training set, n=141/578 were classified to Cluster 1. Cluster 2 was a "low catastrophizing, low pain and high function phenotype" with the lowest PCS and highest (best) OKS among the clusters (n=29/578 pre-TKA patients). Cluster 3 was a "moderate-to-high catastrophizing, poor pain and function phenotype", with moderate PCS and poor OKS (n=178/578 pre-TKA patients). Cluster 4 was a "low catastrophizing, poor pain and function phenotype", with low PCS and poor OKS (n=230/578 pre-TKA patients). The test dataset projected on the SOM yielded very similar cluster distributions, providing validity to phenotype interpretation.

Conclusions: Two-dimensional mapping quantifies and visually represents dimensions of self-reported pain catastrophizing and joint-level pain and function scores. This analysis identified "high catastrophizing, poor pain and function" (Cluster 1), and "low catastrophizing, poor pain and function" (Cluster 4) phenotypes, demonstrating that PCS and self-reported pain and function are not linearly associated for all patients. PCS has been demonstrated to be an independent predictor for chronic pain after TKA, and the ability to identify high-risk phenotypes prior pre-TKA is relevant to tailoring interventions to individuals. Future work incorporating detailed demographic and self-reported dimensions to SOM clustering could aid in developing detailed patient profiles, and personalized care strategies.

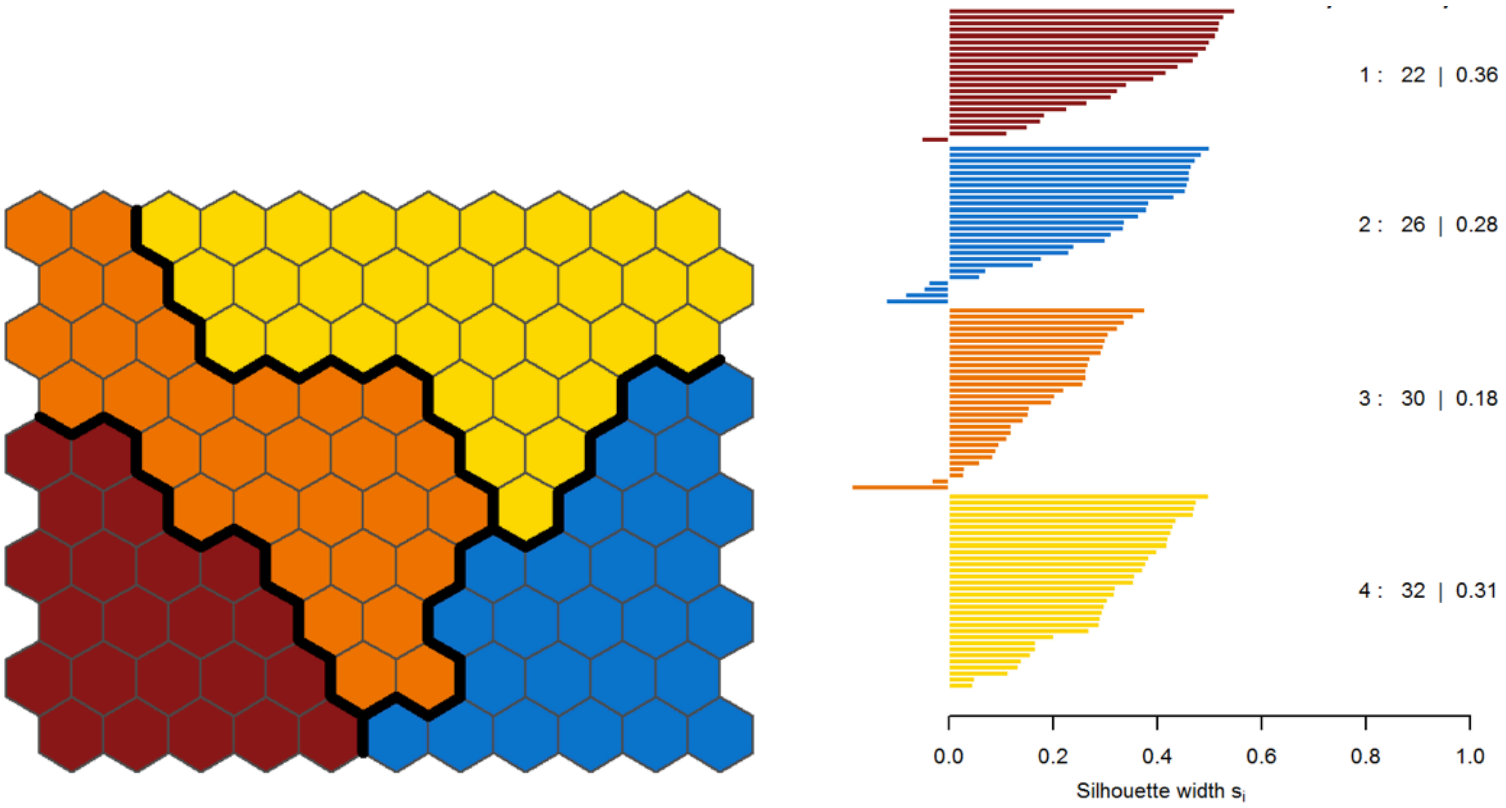

Figure D.1. Left) SOM model; right) silhouette plot of accepted SOM clustering solution (mean silhouette width of 0.28), where s_i is the silhouette width of each node for clusters 1 to 4. The number of nodes belonging to each cluster, and the mean silhouette width of each cluster are also provided (nodes in cluster k | mean silhouette width of cluster).

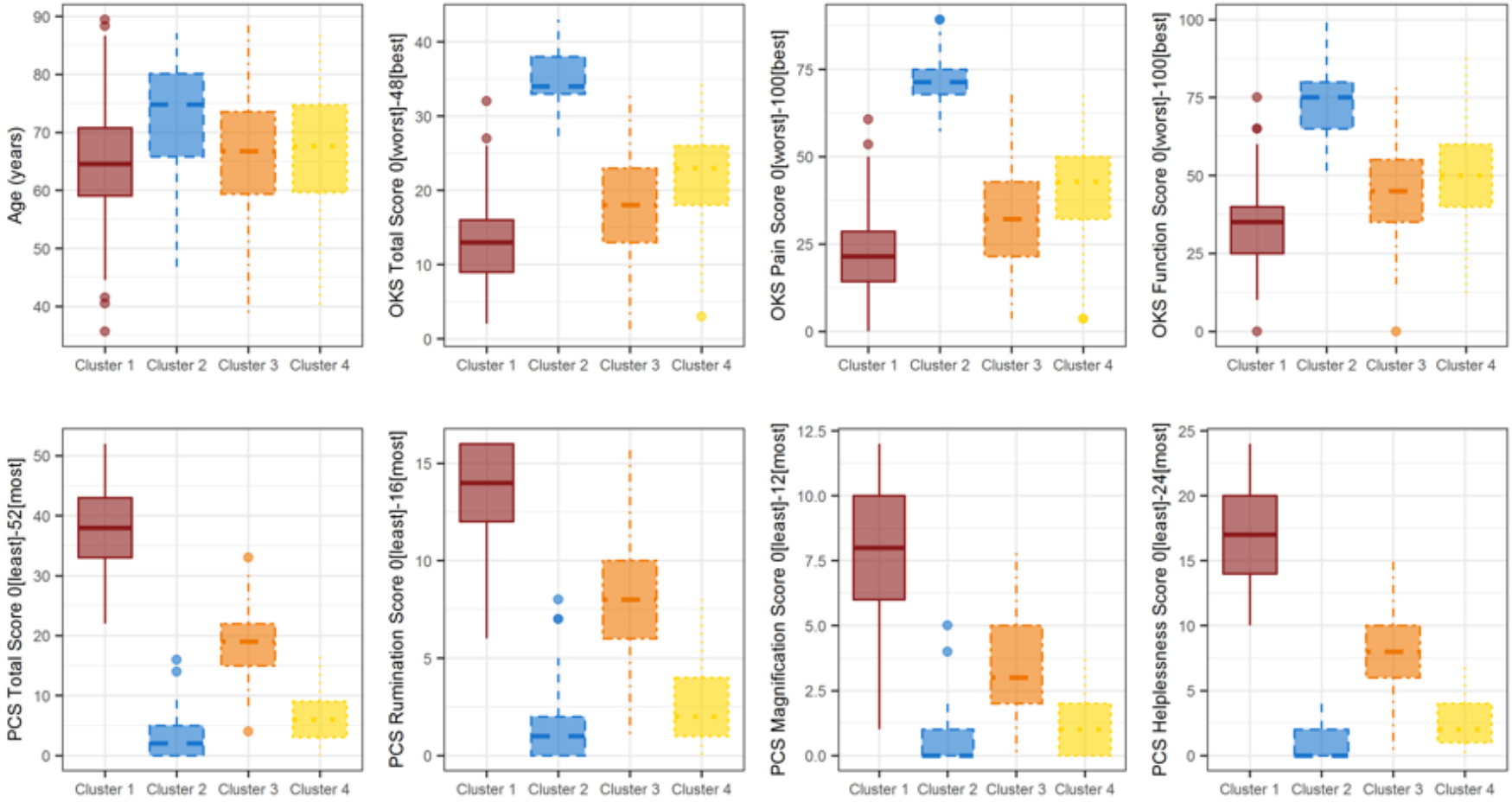

Figure D.2. Box plot of cluster distribution by clustered features.

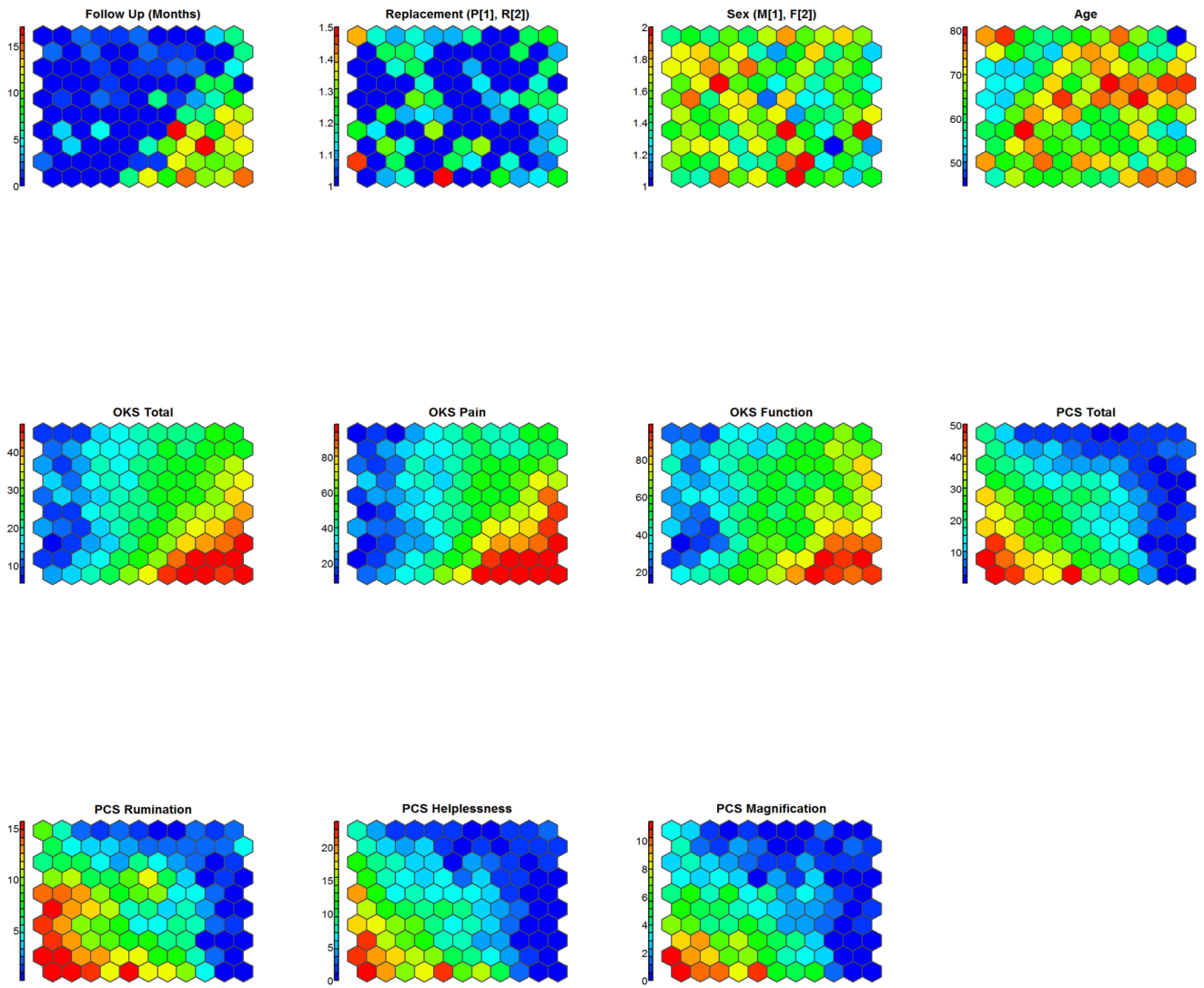

Figure D.3. Heat map of SOM at the feature level. Colours correspond to feature bar legends, low[blue]-high[red].a