# A FRAMEWORK FOR FREE-LIVING ASSESSMENT

# INTEGRATING WEARABLE GAIT ANALYSIS FOR INFORMED DECISION MAKING IN LATE-STAGE OSTEOARTHRITIS: A FRAMEWORK FOR FREE-LIVING ASSESSMENT

by MATTHEW COLIN RUDER, M.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

McMaster University © Copyright by Matthew Ruder, June 2025

## Descriptive Note

DOCTOR OF PHILOSOPHY (2025)
(Kinesiology)
McMaster University, Hamilton, Ontario

TITLE: Integrating Wearable Gait Analysis for Informed Decision Making in Late-Stage Osteoarthritis: A Framework for Free-Living Assessment

AUTHOR: Matthew Colin Ruder, MSc

SUPERVISOR: Dylan Kobsar, PhD

NUMBER OF PAGES: xvii, 180

## LAY ABSTRACT

Osteoarthritis of the knee is on the rise and often changes the way people walk, yet these changes are hard to measure outside a laboratory. To better understand how to best measure walking inside and outside of a laboratory, this thesis used a combination of camera-based and sensor-based measurement systems to understand how people move before and after surgery. First, it builds a new computer model that finds walking within sensor signals more accurately. Next, it shows that data from motion capture and sensors agree on how someone moves in the lab and how that relates to what happens during walking in the real world. Finally, both tools track patients before and after knee replacement surgery, revealing the strong points and limits of each method. Together, they could give doctors and patients a clearer, real world picture of recovery and can be adapted to monitor other injuries, sports, or daily activities.

## ABSTRACT

Late-stage knee osteoarthritis (OA) is a growing musculoskeletal disease affecting millions of older adults. Objective clinical assessments may be a means to improve surgical outcomes but often requires dedicated laboratory equipment and space. Wearable sensors may be easier to collect but are limited by more complex analysis and interpretation of results. This thesis introduces a modular, open-source framework that for collecting gait with bilateral shank inertial measurement units (IMUs). The processing pipeline automates data alignment, incorporates deep learning gait segmentation, stride event detection, and metric extraction, enabling seamless analysis across laboratory and free-living settings.

Three studies establish the framework's value. Study 1 retrained an existing ResNet + BiLSTM using healthy and OA datasets. The model reached ~97 % classification accuracy and decreased walking bout fragmentation compared to a heuristic frequency method, especially at slower walking speeds. Study 2 demonstrated strong in-lab agreement between motion capture- and sensor-derived spatiotemporal and kinematic variables. However, week-long free-living recordings revealed systematically slower and more variable gait, confirming that laboratory snapshots may overestimate real-world mobility. Notably, peak mediolateral shank angular velocity, a native IMU metric, remained well-correlated with Oxford Knee Score, highlighting its clinical promise. Study 3 delivered the first longitudinal, head-to-head sensitivity comparison between measurement systems in 42 arthroplasty patients. Metrics from motion capture were able to capture early postoperative gains, whereas data from IMUs tracked day-to-day function.

Collectively, these findings show that pairing laboratory precision with ecological breadth from inertial sensors could yield a richer picture of OA gait than either modality alone, while also demonstrating strengths and weaknesses of both measures. The framework's sensor-agnostic

design, evidence for clinically relevant native IMU variables, and demonstration of complementary sensitivity advance the field toward scalable, data-driven monitoring and personalised rehabilitation.

## ACKNOWLEDGMENTS

This thesis is the culmination of the past four years that would not have been possible without many people helping people along the way. Coming back for my PhD in my thirties, just after the birth of my first child with another planned before I finished, and in another country, wasn't an easy decision. I'm happy to say now at the end, it was the correct one.

To my advisor, Dylan Kobsar, thank you for being an excellent mentor, giving me the space to develop my own ideas but also guidance when I needed it. You always had my back with a positive attitude about everything that really helped, especially when a model wasn't quite working or the stride length calculation (still) wasn't coming out like I wanted it to. Being your first PhD student was not something I took lightly, and it was important to me to help form the foundation for the lab and set the bar high for those that came next. You were the advisor I needed when I had finally found the area that I felt passionate enough about to finish my PhD.

To my committee members: Dr. Peter Keir, Dr. Janie Wilson, and Dr. Rong Zheng. Thank you for all your guidance and feedback throughout. I set out to complete an ambitious project and I could not have completed it without all your help. Your collective experiences provided me with helpful insights and constructive criticism that led to my successful dissertation research.

To my external examiner, Dr. Katherine Boyer. Thank you for your feedback and taking time to review my dissertation. I was honoured when you agreed to be my external examiner, as you are someone I have looked up to in the field of biomechanics for a long time, and greatly enjoyed our conversation during the defence.

To the other members of the Kobsar Lab. Thank you all for being great teammates and friends. Grad school can be a grind at times, and I hope I was helpful when you asked about different problems you were having with your projects or questions about jobs or life in general. It was great to work with all of you, and the collaboratory nature of our lab was one of the things I was most excited about when decided to join. This is truly what a grad school research lab should be like. In particular, thank you to Zaryan Masood and Josh Keogh, for being so welcoming when I arrived in the lab and really made it feel like home.

To clinic staff at St. Joe's. Thank all of you for your willingness to help with placing countless sensors and collecting data. Thanks to Kim Madden for keeping things organized, Monica Malek for running most of the data collections and processing, and Kim Irish for picking everything up during staffing changeovers. This dissertation truly does not happen without all your support from the beginning to the end.

Finally, thank you to my family. To my wife, Sara, this PhD is half yours since without your love and support throughout it would not have happened. We packed up and moved with an infant to a new country so I could do this PhD, and you made sacrifices along the way so I could get that done. The willingness to say yes to that are among the reasons I love you, and I can't wait to see where our life takes us together. To my children, Gideon and Ada, I love you both and love all the extra time I got to spend with you both because I was doing grad school again. I hope I am an inspiring figure in your life. Lastly, to my grandfather, who passed away last year. You always were telling me I was destined to get a PhD, and I wish you could have seen me complete it. I hope you would be proud.

## DECLARATION OF ACADEMIC ACHIEVEMENT

FORMAT AND ORGANIZATION OF THESIS

This thesis is prepared in accordance with the "sandwich" format outlined in the School of Graduate Studies Guide for the Preparation of Master's and Doctoral Theses. It includes a literature review (Chapter 1), a motivational chapter describing the development of a data-informed framework (Chapter 2), three studies prepared in journal article format (Chapters 3-5), and a general discussion chapter (Chapter 6). At the time of this thesis, Chapters 3-5 were prepared for submission. For all papers with multiple authorship, the contributions of the candidate and all coauthors are outlined below using the Contributor Role Taxonomy (CRediT; https://credit.niso.org/), commonly used during journal submission.

**Chapter 3 (Study 1):**

**Ruder M.C.**, Di Bacco V.E., Patel K., Zheng R., Madden K., Adili A., Kobsar D. Augmenting a ResNet+BiLSTM deep learning model with clinical mobility data outperforms heuristic frequency-based model for walking bout segmentation. *Prepared for submission to Sensors.*

*Contributions*

Conceptualization: MCR, KD

Data curation: MCR, VED

Formal analysis: MCR

Funding acquisition: KD

Investigation: MCR, KP

Methodology: MCR, KP, RZ, KM, KD

Project administration: MCR

Resources: KD, AA, KM

Software: MCR, KP

Supervision: MCR, KD

Validation: MCR, KD

Visualization: MCR

Writing – original draft: MCR

Writing – review and editing: VED, KP, RZ, KM, AA, KD

**Chapter 4 (Study 2):**

**Ruder M.C.**, Di Bacco V.E., Madden K., Adili A., Kobsar D. Comparing Gait Metrics from In-Lab Gait Analyses to Free-Living Assessment from Wearable Sensors in End-Stage Osteoarthritis Patients. *Prepared for submission to Journal of Biomechanics.*

*Contributions*

Conceptualization: MCR, KD

Data curation: MCR, VED, KM

Formal analysis: MCR, VED

Funding acquisition: KD, KM

Investigation: MCR,

Methodology: MCR, KM, KD

Project administration: MCR, VED, KM

Resources: KD, AA, KM

Software: MCR

Supervision: MCR, KD

Validation: MCR, KD

Visualization: MCR

Writing – original draft: MCR

Writing – review and editing: VED, KM, AA, KD

**Chapter 5 (Study 3):**

**Ruder M.C.**, Di Bacco V.E., Madden K., Adili A., Kobsar D. Evaluating the Longitudinal Sensitivity of In-Lab and Free-Living Gait Assessments in Knee Osteoarthritis. *Prepared for submission to Osteoarthritis and Cartilage.*

*Contributions*

Conceptualization: MCR, KD

Data curation: MCR, VED, KM

Formal analysis: MCR, VED

Funding acquisition: KD, KM

Investigation: MCR,

Methodology: MCR, KM, KD

Project administration: MCR, VED, KM

Resources: KD, AA, KM

Software: MCR

Supervision: MCR, KD

Validation: MCR, KD

Visualization: MCR

Writing – original draft: MCR

Writing – review and editing: VED, KM, AA, KD

# TABLE OF CONTENTS

**Descriptive Note** ..... ii
**LAY ABSTRACT** ..... iii
**ABSTRACT** ..... iv
**ACKNOWLEDGMENTS** ..... vi
**DECLARATION OF ACADEMIC ACHIEVEMENT** ..... vii
**TABLE OF CONTENTS** ..... x
**List of Figures:** ..... xiv
**List of Tables:** ..... xv
**List of Abbreviations:** ..... xvi
**Glossary of Terms** ..... xvii
**Chapter 1: Literature Review** ..... 1
**1. Introduction** ..... 1
**1.1 Knee Osteoarthritis** ..... 5
**1.2 Gait Analysis Overview** ..... 6
**1.2.1 Traditional Laboratory-based Gait Analyses** ..... 8
**1.2.2 Limitations of Laboratory-based Gait Analysis** ..... 13
**1.2.3 Wearable Sensor-based Gait Analysis** ..... 15
**1.2.4 Limitations of Wearable Sensor-based Gait Analysis** ..... 21
**1.3 Machine Learning Applications for Gait Analysis** ..... 23
**1.3.1 Machine Learning Overview** ..... 23
**1.3.2 Machine Learning for Gait Segmentation** ..... 26
**1.3.3 Limitations of Machine Learning in Gait Segmentation** ..... 31
**1.3.4 Applying a Machine Learning Framework for Osteoarthritis Gait Segmentation** ..... 31
**1.4 Gaps in literature** ..... 34
**1.5 Study Objectives** ..... 36
**Chapter 2: Development of a Data-Informed, Clinically Viable Gait Framework** ..... 38
**Preamble** ..... 38
**2.1 Framework Considerations** ..... 39
**2.1.1 Placement and Location** ..... 39
**2.1.2 Sensor Location** ..... 40

**2.1.3 Number of Sensors** ... 40
**2.1.4 Repeatability and Population Considerations** ... 41
**2.2 Foundational Methodological Studies** ... 41
**2.2.1 Researcher-Placed vs Self-Placed Sensors** ... 41
**2.2.2 Repeatability of Gait Metrics from Out-of-Lab Gait Assessment in OA Patients** 42
**2.2.3 Implications of Findings** ... 43
**2.3 Description of Gait Framework** ... 44
**2.3.1 Preprocessing** ... 46
**2.3.2 Gait Segmentation** ... 46
**2.3.3 Event Detection and Stride Segmentation** ... 47
**2.3.4 Outlier Detection** ... 48
**2.3.5 Metric Extraction** ... 49
**2.4 Conclusions** ... 49
**Chapter 3: Augmenting a ResNet + BiLSTM deep learning model with clinical mobility data outperforms heuristic frequency-based model for walking bout segmentation (Study 1)** ... **51**
**Preamble** ... 51
**Abstract** ... 52
**3.1 Introduction** ... 53
**3.2 Methods** ... 57
**3.2.1 Description of Datasets** ... 57
**3.2.2 Preprocessing** ... 61
**3.2.3 ResNet + BiLSTM Model Framework** ... 63
**3.2.4 Model Training and Performance Analysis** ... 65
**3.3 Results** ... 69
**3.3.1 Dataset composition** ... 69
**3.3.2 Model Training, Validation, and Testing** ... 70
**3.3.3 Model Comparison to Heuristic Gait Detection** ... 74
**3.3.4 Effect of Gait Speed** ... 75
**3.4 Discussion** ... 76
**3.5 Bridge Section – Comparison of Free-Living Bout Identification from Frequency Method and ML Method** ... 81
**3.6 Supplemental Material** ... 84

**3.6.1 Healthy Lab Dataset** ..... 84
**3.6.2 Clinical Lab Dataset** ..... 85
**Chapter 4: Comparing Gait Metrics from In-Lab Gait Analyses to Free-Living Assessment from Wearable Sensors in End-Stage Osteoarthritis Patients (Study 2)** ..... **89**
**Preamble** ..... 89
**Abstract** ..... 90
**4.1 Introduction** ..... 90
**4.2 Materials and Methods** ..... 93
**4.2.1. Participants** ..... 93
**4.2.2 Data Collection Protocol** ..... 93
**4.2.3. Data Analysis** ..... 95
**4.2.4 Statistical Analysis** ..... 98
**4.3 Results** ..... 99
**4.4 Discussion** ..... 106
**4.5 Conclusions** ..... 109
**4.6 Supplemental Figure** ..... 111
**Chapter 5: Evaluating the Longitudinal Sensitivity of In-Lab and Free-Living Gait Assessments in Knee Osteoarthritis (Study 3)** ..... **112**
**Preamble** ..... 112
**Abstract** ..... 112
**5.1 Introduction** ..... 114
**5.2 Materials and Methods** ..... 117
**5.2.1 Participants** ..... 117
**5.2.2 Data Collection Protocol** ..... 117
**5.2.3. Data Analysis** ..... 119
**5.2.4 Statistical Analysis** ..... 121
**5.3 Results** ..... 122
**5.3.1 Patient-Reported Outcomes** ..... 123
**5.3.2 Motion Capture** ..... 125
**5.3.3 Wearable Sensors** ..... 131
**5.4 Discussion** ..... 136
**5.5 Conclusions** ..... 140
**5.6 Supplemental Tables** ..... 141

**Chapter 6: General Discussion ..... 145**
**6.1 Research Implications** ..... 145
**6.2 Contributions to Existing Knowledge** ..... 148
**6.3 Limitations and Considerations** ..... 150
**6.4 Recommendations for Future Work** ..... 153
**6.5 Conclusions** ..... 155
**Chapter 7: References ..... 156**

## List of Figures:


| Figure Number | Description | Page |
|---|---|---|
| Figure 1.1 | Radiographs of osteoarthritis in knee, grades 1 to 4 | 6 |
| Figure 1.2 | Simplified stance diagram depicting varus thrust | 10 |
| Figure 1.3 | Tables depicting rise of publications investigating wearable sensors, along placement of sensors | 17 |
| Figure 1.4 | Longitudinal study design utilized by Bolam, et al. | 20 |
| Figure 1.5 | Deep learning framework proposed by Li and Wang | 29 |
| Figure 2.1 | Simplified workflow for processing pipeline | 45 |
| Figure 3.1 | Sensor placements for PAMAP2, healthy dataset, and clinical dataset | 58 |
| Figure 3.2 | Deep learning model architecture, adapted from Li and Wang | 65 |
| Figure 3.3 | Confusion matrices broken out individually for PAMAP2, healthy, and clinical testing sets | 73 |
| Figure 4.1 | Sensor locations on shank, located medial and inferior to the tibial tuberosity | 94 |
| Figure 4.2 | Bland-Altman plots for each variable between motion capture and in-lab IMU data | 101 |
| Figure 4.3 | Representative plot of stride time distribution for in-lab data and free-living IMU data | 103 |
| Figure 4.4 | Scatter plots for free-living peak angular velocity vs OKS, and gait speed (from motion capture) vs OKS | 106 |
| Figure 5.1 | Self-reported function and pain, plotted for each time point | 124 |
| Figure 5.2 | Gait speed and stride time, derived from motion capture, plotted for each timepoint | 129 |
| Figure 5.3 | Knee joint flexion excursion at each timepoint and normalized knee joint flexion angles from participants who completed all each timepoints | 130 |
| Figure 5.4 | Frontal plane excursion at each timepoint and normalized flexion and abduction/adduction knee joint angles from participants who completed all each timepoints | 131 |
| Figure 5.5 | Peak frontal acceleration asymmetry and peak sagittal angular velocity asymmetry, plotted for each timepoint | 135 |

## List of Tables:


| Table Number | Description | Page |
|---|---|---|
| Table 3.1 | Description of each model iteration in terms of training, validation, and testing data used in model development and evaluation | 67 |
| Table 3.2 | Activity breakdown by percentage for each dataset, both individually as well as total breakdown by dataset combinations | 70 |
| Table 3.3 | Overall performance on training, validation, and test sets for each model. | 71 |
| Table 3.4 | Median performances on individual test sets from each data, based on the top performing model, when used for gait detection and the heuristic frequency-based method | 74 |
| Table 3.5 | Method performance on participants in clinic test set by speed. | 75 |
| Table 3.6 | Performance of the frequency-based and machine-learning bout-detection methods in free-living data. | 83 |
| Table 4.1 | Descriptive statistics for the sample | 99 |
| Table 4.2 | Concurrent validity between motion capture and in-lab sensor data | 100 |
| Table 4.3 | Ecological validity of gait variables from in-lab sensors and free-living sensors | 102 |
| Table 4.4 | Pearson correlations between gait metrics d patient reported outcome measures relating to function, depressive symptoms, and pain | 105 |
| Table 5.1 | Results of linear mixed-effects models for self-reported outcomes at each study timepoint. | 125 |
| Table 5.2 | Results of linear mixed-effects models for variables derived from markerless motion capture | 127 |
| Table 5.3 | Results of linear mixed-effects models for variables derived from wearable sensors | 133 |

**List of Abbreviations**:

| Abbreviation | Definition |
|---|---|
| CMC | Coefficient of Multiple Correlation |
| CNN | Convolutional Neural Network |
| FFT | Fast Fourier Transform |
| GRF | Ground Reaction Force |
| ICC | Intraclass Correlation Coefficient |
| IMU | Inertial Measurement Unit |
| KAM | Knee Adduction Moment |
| KL | Kellgren-Lawrence |
| LSTM | Long Short-Term Memory |
| ML | Medial-Lateral |
| OA | Osteoarthritis |
| OKS | Oxford Knee Score |
| PCA | Principal Component Analysis |
| PHQ | Patient Health Questionnaire |
| RMSE | Root Mean Squared Error |
| RNN | Recurrent Neural Network |
| SVM | Support Vector Machine |
| TKA | Total Knee Arthroplasty |

## Glossary of Terms

| Term | Definition |
|---|---|
| Acceleration | Rate of change of velocity over time, expressed in $m/s^2$. Measured by accelerometer component of IMU to quantify segment speeding up or slowing down. |
| Accuracy | The proportion of total predictions that were correct, reflecting overall correctness of a model. |
| Angular velocity | Rate of rotation around an axis, expressed in degrees per second (deg/s or degree/second). Measured by gyroscope component of IMU to quantify segment rotation. |
| Coefficient of Multiple Correlation | A statistical measure used to compare how similar one waveform is to a reference waveform across multiple points. |
| Deep Learning | An advanced machine learning method using neural networks with multiple layers to automatically learn features from large datasets. |
| Event Detection | A process in gait analysis where specific gait events (e.g., heel strike, toe off) are identified within a walking cycle, typically using peaks or features in acceleration or gyroscope signals. |
| F1-Score | The harmonic mean of precision and recall, used to balance false positives and false negatives in classification. |
| Gait Analysis | Study of walking patterns primarily by using kinematics, kinetics, and/or spatiotemporal variables to describe gait. |
| Gait Segmentation | The process of identifying periods of walking from continuous sensor data, often used to isolate walking bouts. |
| Heuristic Model | A rule-based approach that applies fixed thresholds or conditions to make predictions, often used as a baseline method. |
| Inertial Measurement Unit | A small sensor device that includes accelerometers and gyroscopes to capture motion and orientation of body segments. |
| Kinematics | The study of motion in terms of displacement, velocity, and acceleration, without considering the forces that cause it. |
| Kinetics | The study of forces and torques that cause movement, such as ground reaction forces during walking. |
| Machine Learning | A branch of artificial intelligence where algorithms learn patterns from data and improve performance without being explicitly programmed. |
| Markerless Motion Capture | A system that estimates joint and body segment positions from video without using physical markers attached to the subject. |
| Osteoarthritis | A degenerative joint disease where cartilage wears down over time, leading to pain, stiffness, and decreased mobility, especially in weight-bearing joints. |
| Precision | The proportion of correctly predicted positive observations among all predicted positives, indicating model exactness. |
| Principal Component Analysis | A dimensionality reduction technique that transforms data into principal components to highlight variation and simplify analysis. |
| Recall | The proportion of actual positives correctly identified by the model, indicating its sensitivity. |

| | |
|---|---|
| Spatiotemporal Parameters | Gait metrics that involve both space (e.g., stride length) and time (e.g. stride time), used to assess walking performance. |
| Stance Time | The period when the foot is in contact with the ground during a gait cycle. |
| Stride Time | The duration between successive contacts of the same foot, typically measured from heel strike to heel strike. |
| Swing Time | The portion of the gait cycle when the foot is off the ground and moving forward. |
| Total Knee Arthroplasty | A surgical procedure in which a damaged knee joint is replaced with artificial components. |
| Varus Thrust | A visible and abrupt inward movement of the knee during the stance phase of walking, commonly observed in knee OA. |

# Chapter 1: Literature Review

## 1. Introduction

Osteoarthritis (OA) is a debilitating joint disease characterized by the loss of cartilage and structural changes to the surrounding bone and soft tissue. These changes can lead to increased pain, reduced function, and a decline in overall quality of life (Moskowitz, 2009). The knee is the most affected joint, with the end-stage treatment being a total knee arthroplasty (TKA). OA affects millions of older adults, and the burden is growing. With Canada's aging population, TKA procedures have been steadily increasing over the past two decades, approaching 100,000 annual surgeries and placing intense pressure on healthcare systems (Canadian Institute for Health Information, 2023).

Knee OA is associated with altered lower limb movement patterns and joint loading during walking (Kumar et al., 2013), making gait analysis a valuable tool for monitoring disease progression (Elbaz et al., 2014). Traditional gait analysis typically involves placing markers on a participant and having them walk several times across a room or on a treadmill. While these systems do provide a non-invasive way to support personalized treatment and management of OA, their use is not widespread and the potential for implementation as a standard of care is limited. This is due to that fact that these systems are prohibitively expensive for most clinics with respect to the required equipment, experienced operators, and time constraints (Simon, 2004). Furthermore, they also only provide a small snapshot of a patient's gait throughout the day, and drawing conclusions

from these limited number of gait cycles in a highly constrained environment may not be representative of their real-world gait and function (Benedetti et al., 2013). Consequently, while gait analysis research has identified key factors that could inform clinical decision making, such as knee adduction angles and moments, these insights are not yet routinely translated into clinical practice due to barriers in implementation, accessibility, and clinician training (Outerleys et al., 2021; Young-Shand et al., 2020).

Wearable sensors offer a promising alternative to traditional laboratory-based gait analysis. Compared to conventional systems, they are more affordable and easier to deploy in clinical and real-world settings, with inertial measurement units (IMU) being the most used. Additionally, these sensors have consistently been shown to provide accurate and reliable data on spatiotemporal and kinematic data in controlled settings (Kobsar et al., 2020a). The ability to collect these data continuously in a variety of environments, including more uncontrolled, free-living conditions, offer the potential to gather more ecological validity data that may better represent patients' daily lives (Kobsar et al., 2020b, 2017).

While free-living gait data collected from wearable sensors can provide a richer and more continuous stream of information, the volume and variability of these data also present significant challenges for analysis, particularly in identifying walking bouts. Simple event detection algorithms often fail outside controlled environments, especially in clinical populations with pathological gait. To address this, machine learning methods have been increasingly applied to free-living gait analysis in populations such as older adults

(Nouredanesh et al., 2021), chronic obstructive pulmonary disease (Buekers et al., 2023), Parkinson's disease (Martindale et al., 2021), and multiple sclerosis (Shema-Shiratzky et al., 2020). However, these models are rarely designed for knee OA patients, who often walk more slowly and with altered patterns, nor are they typically trained on data from lower limb sensors mounted near the knee. This represents a clear opportunity to develop a deep learning-based model tailored for OA populations to identify periods of gait from periods of non-gait.

Overall, with recent advancements in technology, many of the components necessary to support more advanced and responsive treatment approaches for OA are now available. However, these tools have either not been extensively examined within OA populations (e.g., the reliability of gait metrics in free-living environments) or have not been applied in an integrated manner (e.g., using biomechanical models to monitor treatment-related gait changes over time). Although much of the existing work has demonstrated the utility of wearable sensors and analytical models in controlled laboratory settings, their greatest potential may lie in free-living applications, where real-world movement patterns offer richer insight into functional status. Fully realizing this potential will require the incorporation of machine learning methods to reliably identify walking bouts in unstructured, real-world data. This, in turn, would allow for a more comprehensive evaluation of these tools in clinically relevant settings and facilitate a better understanding of their sensitivity to change and their value relative to conventional, in-lab gait

assessments. Therefore, the broad and overarching questions which drive my doctoral dissertation are:

1. Can deep learning approaches improve identification of walking gait over current threshold-based methods in free-living data, particularly in older adults with OA?
2. Do wearable sensors, when worn for an extended collection period prior to surgery, provide a different representation of a patient's gait than those collected from a laboratory-based motion capture system?
3. How does the sensitivity of free-living gait metrics derived from wearable sensors compare to laboratory-based measures in detecting gait changes across pre- and post-operative timepoints following surgical intervention?

## 1.1 Knee Osteoarthritis

Osteoarthritis (OA) is a serious and incurable joint disease, with the underlying causes and etiology are largely unknown (March, 2016). While OA can affect many different joints, the most common site for OA is at the knee (Chehab et al., 2014). Knee OA is characterized by structural changes in the soft tissues surrounding the joints, which over time lead to joint narrowing and pain (Palmer et al., 2020). Specifically, reductions of cartilage at the knee joint over time lead to additional changes in the bones (Figure 1.1). Initial diagnoses for knee OA are done radiographically using the Kellgren-Lawrence scale, with grades rating the structure in the knee from 0 (no structural damage), 1 (doubtful joint narrowing, possible tissue degradation), 2 (possible joint narrowing with definitive tissue degradation), 3 (definite joint narrowing with tissue damage) to 4 (severe joint narrowing and tissue damage) (Kellgren and Lawrence, 1957). Patients that eventually receive the end-stage treatment, total knee arthroplasty (TKA), can take up to 6 months to fully heal (Paravlic et al., 2022). Overall, the procedure results in significant pain relief and improvement in walking, though individual outcomes can greatly vary (Ritter et al., 2008). OA affects millions of older adults and has grown substantially with annual number of arthroplasty procedures over 11000, further straining healthcare systems (Canadian Institute for Health Information, 2023).

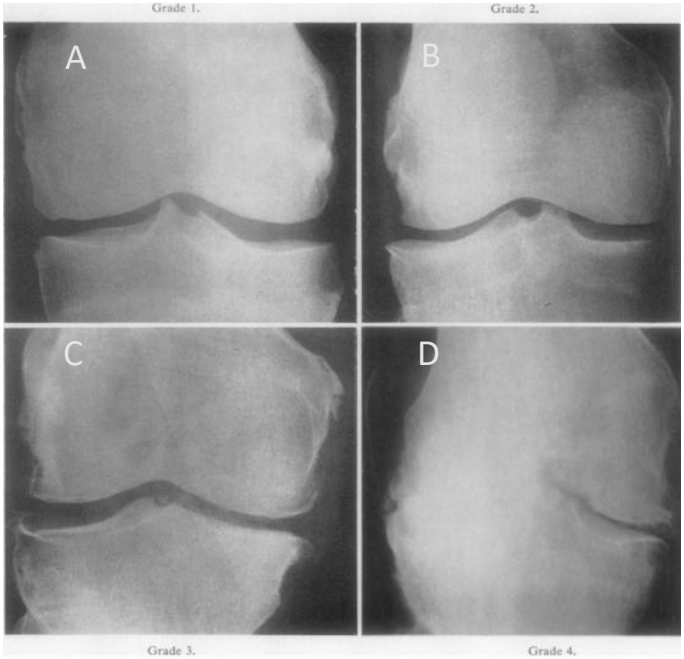



**Figure 1.1:** Radiographs of osteoarthritic knee Grade 1 to Grade 4, from Kellgren and Lawrence (Kellgren and Lawrence, 1957). This figure demonstrated grades rating the structure in the knee. A. KL Grade 1, where there is doubtful joint narrowing, with possible tissue degradation. B. KL Grade 2, where there is possible joint narrowing with definitive tissue degradation. C. KL Grade 3, where there is definite joint narrowing with tissue damage. D. KL Grade 4, where there is severe joint narrowing and tissue damage.

## 1.2 Gait Analysis Overview

Gait analysis is a critical tool for evaluating movement in both healthy and pathological populations. It enables the characterization of an individual's gait through several biomechanical domains: kinetics, kinematics, and spatiotemporal parameters. Kinetics refer to the forces and moments acting on joints or segments, such as ground

reaction forces during walking. Kinematics describe the motion of joints or segments in terms of position, velocity, or acceleration, such as knee flexion angles during stance. Spatiotemporal parameters include timing- and distance-related metrics such as step length, step time, and gait speed, which offer functional insight into overall gait performance. Gait analysis may also incorporate electromyography to assess muscle activation patterns and their influence on joint motion. Another commonly used approach for deriving joint kinetics is inverse dynamics, which estimates internal joint moments by combining kinematic data with external forces, such as those measured by force plates.

On a basic level, clinical gait analysis is looking for the presence of aberrant movements during periods of walking. Clinicians often use these visual assessments during pre- and post-surgical visits. Early research-based gait analysis moved from visual assessments to video collection. As technology progressed, researchers increasingly collected with optoelectronic motion capture systems, using cameras and markers to track the subject through a calibrated collection area. Recent advances with markerless motion capture also allows for gait analysis without markers. Wearable sensors have become integral to gait analysis both within and outside the laboratory, enhancing our understanding of OA progression and the effects of various interventions. A recent systematic review concluded that many gait biomechanical parameters are associated with increased odds of OA onset and progression in the disease at the knee, with those in the frontal plane and sagittal plane measures being the most meaningful (D'Souza et al., 2022). Overall, 91% (21/23) of the studies analyzed showed association with at least one

biomechanical variable. The following sections will highlight the utility and limitations of each approach for gait analysis, with a particular focus on OA populations.

### 1.2.1 Traditional Laboratory-based Gait Analyses

Using force platforms as an analysis of kinetics via ground reaction forces (GRF) offer a way to better understand the loading environment of the knee using inverse dynamics. Numerous studies have analyzed knee adduction moment (KAM), which is the external torque through the knee, as a primary proxy for distribution of frontal plane loading through the knee. KAM is calculated via an inverse dynamics approach that takes into account the dynamic alignment of the knee and location of the GRF during stance. With respect to OA, KAM has been linked to severity (Mündermann et al., 2005), progression (Miyazaki, 2002), pain (Hurwitz et al., 2002), and treatment (Shull et al., 2013). Additionally, a meta-analysis showed increased odds of OA progression with greater baseline peak KAM (D'Souza et al., 2022). Research has also shown that the KAM may be a modifiable biomechanical marker of OA. A recent study by Seagers, et al. found that modifying foot progression angle can reduce KAM while not increasing hip moments (Seagers et al., 2022). A 10-degree toe-out gait resulted in reductions of 7.6% and 11.0% for the first and second KAM peaks, respectively. Nevertheless, assessing KAM requires multi-axis, in-ground force plates which make it highly inaccessible, especially in clinical settings.

Kinematic measures are particularly important in the gait analysis of individuals with knee OA. For example, knee flexion at initial contact and toe-off is typically increased

in patients with severe OA compared to healthy controls, accompanied by reduced knee flexion excursion during stance (Favre et al., 2016). In a five-year longitudinal study, Favre et al. (2016) found that baseline knee flexion angle at heel strike, reflecting a more vertical shank, was associated with cartilage thinning. This underscores the importance of sagittal plane kinematics in tracking OA progression.

Another potential marker of disease severity is varus thrust, defined as a sudden lateral movement of the knee into a varus (i.e. bow-legged) position during the stance phase of gait (Figure 1.2). This abrupt motion has been associated with a four-fold increase in the risk of OA progression (Chang et al., 2004). Clinicians often rely on visual observation to detect varus thrust, but efforts to objectively quantify a proxy measure of varus thrust by using kinematic metrics, such as peak knee varus angle or varus angular velocity, have yielded inconsistent definitions. The most common definition calculates the frontal plane joint angle excursion at initial foot contact to a subsequence point later in stance (Takigami et al., 2000). However, the window for excursion can vary from the first 10% of gait cycle (Deie et al., 2014) to a broadly defined "early stance" (Shimada et al., 2024) to the entire stance phase (Mahmoudian et al., 2017). Despite inconsistency in definition, some studies have related a proxy measure for varus thrust to severity. For instance, Kuroyanagi, et al., used marker-based motion capture to examine peak varus angles in patients with K-L grades 2, 3, and 4, reporting mean angles of 2.4°, 2.8°, and 7.2°, respectively (Kuroyanagi et al., 2012). Although suggestive of increasing lateral movement with disease progression, further research is needed to establish validated thresholds. Taken together, these findings

highlight the importance of kinematic indicators, such as knee flexion and varus thrust, in understanding how OA alters gait. However, these measures are more complex to assess and interpret than spatiotemporal parameters, which may also serve as valuable indicators of functional gait health.

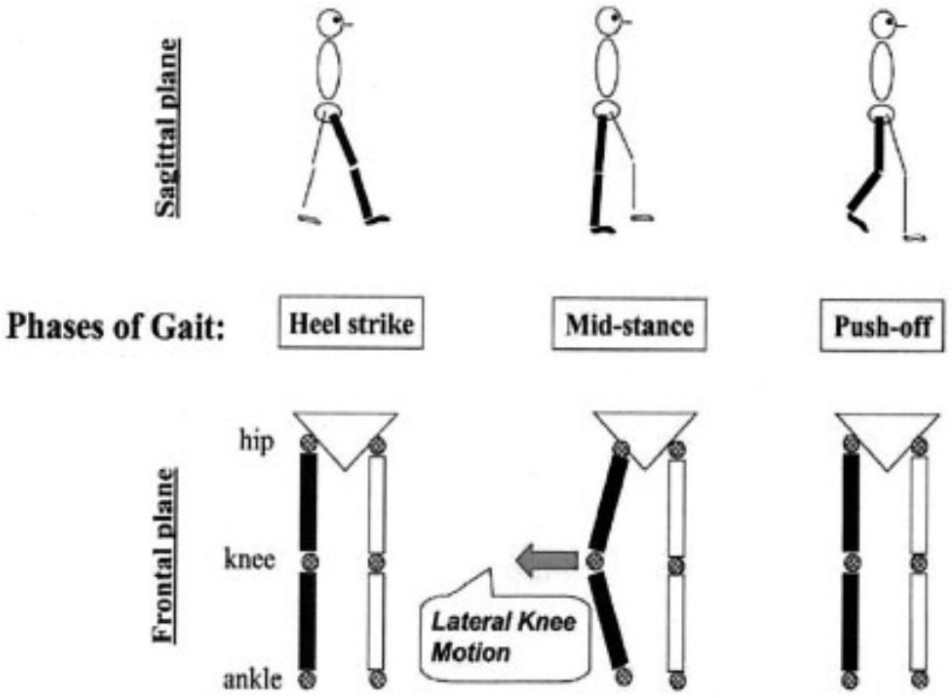



**Figure 1.2:** Simplified stance phase in sagittal and frontal planes, from Chang, et al. (2004) depicting sudden lateral knee motion, often termed varus thrust. Gait analysis was typically done using subjective visual identification of aberrations in gait, such as abrupt lateral knee motion during stance depicted in the lower panel. Gait analysis has increasingly been assessed using objective measures, such as laboratory-based optoelectronic motion capture and wearable sensors but have not arrived at a consensus on defining proxies for this motion.

Spatiotemporal parameters can reflect a patient's functional capacity through measures such as gait speed, step length, and step time, making them simple yet important indicators in OA. Spatiotemporal parameters can be used as a method of quantifying function in OA patients, with a review by Dong, et al., finding differences in walking speed and stride length in patients with unicompartmental knee arthroplasties and TKA (Dong et al., 2023). Pressure mats have been effective in determining these parameters to classify the severity of gait deficits in OA populations (Elbaz et al., 2014). Specifically, the Elbaz et al. study found that spatiotemporal parameters could objectively classify patients with knee OA by severity with 90% accuracy and misclassifications of only one grade. This classification also correlated well with other factors such as pain and function. Variability of spatiotemporal measures could serve as a marker of disease severity in knee OA. For example, a study by Kiss found that gait complexity decreases in patients with knee OA, with greater variability in kinematic parameters. Side-to-side differences ranged from 15 degrees in moderate cases to over 30 degrees in severe cases, which was also associated with increased variability in spatiotemporal parameters (Kiss, 2011). Even simple metrics such as the standard deviation of spatiotemporal parameters may capture gait fluctuations.

Recently, advances in markerless motion capture could greatly improve the access to gait analyses. Early markerless motion capture studies have characterized it as valid and reliable as markered motion capture, while eliminating the time consuming placement of markers from gait analysis and greatly improving the accessibility of these measures (Kanko et al., 2021a, 2021b). Furthermore, work by Keller, et al., found negligible effects

on reliability of markerless motion capture by clothing worn, even over multiple sessions, allowing subjects to wear their "normal" clothes (Keller et al., 2022). These findings further increase the ecological validity of this research tool and value as a potential clinical tool. Additionally, concurrent assessments of gait kinematics between markered and markerless systems showed comparable results between the two systems (Kanko et al., 2021b). Important sagittal plane measures, such as knee flexion angle, were seen to have the greatest similarity, with a root mean squared difference of 3.3 degrees which is within the error of most systems. Consequently, this technology also potentially unifies biomechanical analyses between lab spaces, as variability in data due to differences in lab marker sets and marker placements are eliminated and allow biomechanics researchers to pool data more effectively (Kanko et al., 2021a). Markerless motion capture does have potential issues with the underlying manually labeled data used as training sets, with joint centers at the hip and knee up to 50 mm. This may be mitigated as some companies have proprietary biomechanically labeled datasets (Kanko et al., 2021c; Needham et al., 2021).

Taken together, there is great utility and promise within using gait analyses for monitoring OA. Kinetics, kinematics, and spatiotemporal variables characterize gait, and each provide valuable information on mechanisms of OA-related gait changes, as well as factors that can describe progression and inform treatment options. Markerless motion capture also offers a more accessible way to collect gait analyses. However, there are important considerations with respect to limitations that must be discussed.

### 1.2.2 Limitations of Laboratory-based Gait Analysis

Despite being considered the reference standard for gait analysis, laboratory-based motion capture has several notable limitations. For markered motion capture, the main issues relate to marker placement and skin artifact (i.e., where markers move with the skin rather than the underlying bony landmarks). A study by Miranda et al. comparing markered motion capture to biplanar fluoroscopy found joint center position errors as high as 30 mm (mean: 9-19 mm), and rotational errors up to 14 degrees (average: 2.5-5.5 degrees) (Miranda et al., 2013). Moreover, markered motion capture can be uncomfortable or restrictive for participants, potentially altering their natural walking patterns (Chen et al., 2016; Robles-García et al., 2015). In a study by Robles-García et al., patients with Parkinson's disease and healthy participants were recorded walking along a predetermined path while knowingly being observed, and then covertly recorded on their return walk after being told the trial had ended. Gait speed, cadence, and step length differed significantly between overt and covert recordings, with participants walking faster and with longer strides when aware of being observed. A systematic review of laboratory-based gait analyses also raised concerns regarding the validity and reliability of kinematic measures and emphasized the need for larger sample sizes (Ornetti et al., 2010). Taken together, these factors could limit the applicability of laboratory-based gait analysis in both research and clinical settings.

Despite these limitations, kinetic, kinematic, and spatiotemporal variables derived from laboratory-based gait data provides valuable information for knee OA with respect to

severity, progression, and treatment. However, an additional, but critical limitation could be how well laboratory-based patient gait data reflects their everyday walking. Hillel, et al., sought to characterize differences in laboratory-based and free-living data through use of dual-task walking (Hillel et al., 2019). Dual-task walking, which is walking while completing a secondary cognitive task, is one common approach to simulate "real-world", more "normal" walking in-lab. For Hillel's study, older adults were observed in-lab during regular (single-task) walking and during dual-task walking and then observed out-of-lab over a week using a wearable sensor. These sensors allow continuous collection and potentially provide a much more comprehensive picture of a participant's gait. Indeed, Hillel's study supported the idea that laboratory-based gait is unique from out-of-lab gait, as the laboratory-based measures of gait speed, step regularity, and stride regularity during regular walking were different from both dual-task in-lab walking and out-of-lab walking. Simultaneously, while dual-task walking was more similar to out-of-lab walking, they were still not equivalent. In summary, while dual-tasking gait in-lab is close to real-world walking, neither laboratory-based measures of gait parameters reliably reflected out-of-lab gait parameters. As a result of these factors, wearable sensors have become increasingly prevalent in gait analysis, with a growing number of studies focused on free-living gait outside controlled laboratory environments. The next section will discuss the current state of gait analysis research with wearable sensors as well as the limitations of the technology.

### 1.2.3 Wearable Sensor-based Gait Analysis

Wearable sensors, and particularly inertial measurement units (IMU), which typically include accelerometers, gyroscopes, and sometimes magnetometers, are a cost-effective alternative to motion capture. There is also substantial interest in characterizing OA gait with wearable sensors, as the number of studies investigating OA gait with wearable sensors have gone up considerably in the past decade (Figure 1.3A). Wearable sensors biggest strength is that they can be collected for extended periods, trading the high-fidelity, full body kinematics and/or kinetics for high volume but noisy measurements of segment motion. As previously noted, laboratory-based systems require substantial time to place markers accurately before completing some number of activities, and then ultimately drawing conclusions from a limited number of motion trials (e.g., 5-10 total gait cycles). Wearable sensors can collect hundreds to thousands of gait cycles outside of a lab setting for analysis, which could be potentially more representative, but still ultimately require precise placement.

In general, there is a wide range of the numbers, types, and locations of wearable sensors used in gait research. Studies range from individual accelerometer sensors to an integrated system of sensors in an IMU (e.g., accelerometer, gyroscope, and/or magnetometer), and further to pressure insoles or newer flexible strain sensors (Bolam et al., 2021; Chen et al., 2016; Gholami et al., 2020; Zijlstra and Hof, 2003). Wearable sensors are also flexible in terms of the numbers of sensors used, ranging from a single sensor to an array of multiple sensors (Abe and Nagamune, 2021; Buckley et al., 2019), and can be

placed at different locations (e.g., back, thigh, shank, foot) to obtain a variety of gait metrics (Figure 1.3B) (Kobsar et al., 2020b). While foot or back placements are most common overall, the shank placement is most common for motion measurements (via acceleration signals). The shank placement also would more effectively measure proxies of potential kinematic gait markers, such shank inclination (or possibly knee flexion) angle or varus thrust, while also able to measure spatiotemporal variables effectively. Depending on attachment location, the shank could have reliability issues but these could potentially be overcome by using an alignment algorithm (Hafer et al., 2020). However, the improvement in reliability from using an alignment algorithm has not been documented.

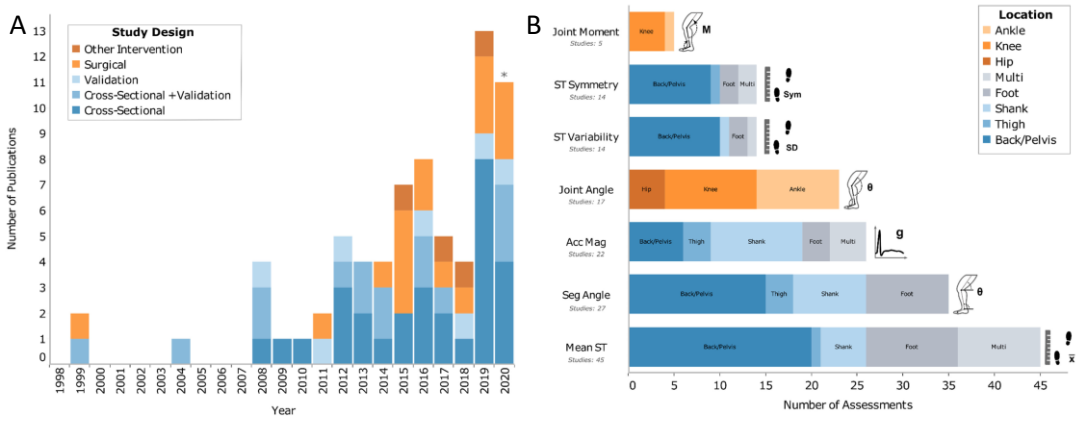



**Figure 1.3:** Rise in publications using wearable sensors in OA, with the placements used to obtain common gait variables, from Kobsar, et al. (2020). A. The number of published studies investigating OA gait with wearable sensors has increased considerably as the sensors have become more commercially available and a viable clinical tool. B. Number of studies using different variables obtained from different wearable sensor placements for gait analysis assessments in osteoarthritis populations. Kinematic variables, such as joint moment, joint angle, segment angle, acceleration magnitude (Acc Mag) can describe the segment (e.g. knee) motion. Spatiotemporal (ST) variables can refer to mean stride or step times as well as the asymmetry or variability between sides.

Applying wearable sensors in a clinical environment with OA patients has the potential to transform how the disease is assessed and managed. To realize this potential, it is essential to evaluate how well wearable-derived gait metrics correspond to reference-standard laboratory-based measurements. As previously discussed, there is a large body of literature on laboratory-based gait metrics (e.g., KAM, knee flexion angle, spatiotemporal

parameters), and a critical step toward clinical translation is understanding how wearable sensor-derived data compares to these reference standards. From a kinetic perspective, for example, He et al. used an array of pressure sensors and a foot-mounted IMU to estimate KAM, demonstrating a strong correlation (median r = 0.90) with values from a laboratory-based motion capture system during foot progression angle modification (Z. He et al., 2019). Five out of six patients successfully reduced their first KAM peak. However, it is important to note that KAM is difficult to estimate from IMUs alone and typically requires additional force-related inputs.

Wearable sensors show particularly strong agreement with laboratory-based systems for kinematic group differences and spatiotemporal parameters. From a kinematic standpoint, Hafer et al. compared wearable IMU-derived gait data to motion capture across young adults, older asymptomatic individuals, and those with knee OA, finding no significant tool × group interaction ($p = 0.67$-$0.98$), suggesting both systems captured group-level differences consistently (Hafer et al., 2020). Costello et al. similarly demonstrated that a single IMU could estimate varus thrust, where increases in peak adduction velocity were significantly associated with increased walking speed, although the strength of correlation was not formally reported (Costello et al., 2020). Spatiotemporal variables have shown the strongest alignment between wearable sensors and laboratory-based systems. In the Hafer study, wearable-derived stride length and walking speed had low mean square errors (<0.07 m and <0.05 m/s, respectively). Systematic reviews further support good-to-excellent validity and reliability for wearable-derived stride time (ICC =

0.91-0.94) and stride length (ICC = 0.81-0.94) in OA populations (Kobsar et al., 2020b, 2020a, 2016).

In addition to the well-studied kinetic, kinematic, and spatiotemporal parameters, wearable sensors can also provide access to additional variables such as cumulative impact load and inter-limb load symmetry. These metrics have been widely explored in athletic settings, where IMUs are used to quantify acceleration peaks and monitor external workload during practices and games. Impact load metrics have shown good-to-excellent reliability in sport-specific tasks like cutting and directional changes (ICC = 0.58-0.97) (Burland et al., 2021) and are often correlated with physiological responses such as perceived exertion and heart rate (Helwig et al., 2023). Because these data can be captured in ecologically valid contexts (i.e. outside laboratory constraints), they have been incorporated into return-to-play protocols and injury prevention strategies (Kupperman et al., 2021). These same strengths suggest potential for translation into clinical populations, particularly for tracking rehabilitation progress and asymmetry in gait.

The application of impact load metrics in clinical OA populations remains limited, but emerging studies suggest meaningful potential. For example, Ren, et al., used load asymmetry during jump tasks to assess recovery following hip resurfacing arthroplasty, finding that asymmetries present before and at 3 months post-surgery had resolved by 6 months (Ren et al., 2023). While not gait-based, this framework could be adapted for ambulatory monitoring in knee OA patients. A more direct application is the study by Bolam, et al., which demonstrated the feasibility of using wearable sensors to monitor

cumulative impact load, asymmetry, and knee flexion during early recovery following total knee arthroplasty (Figure 1.4) (Bolam et al., 2021). Patients wore sensors weekly up to 6 weeks post-operatively, and the data which was collected over a 12-hour window once a week, revealed a 371% increase in cumulative impact load and a trend toward symmetric loading between limbs. These findings underscore the potential for wearable sensors to capture clinically relevant recovery metrics in real-world contexts.

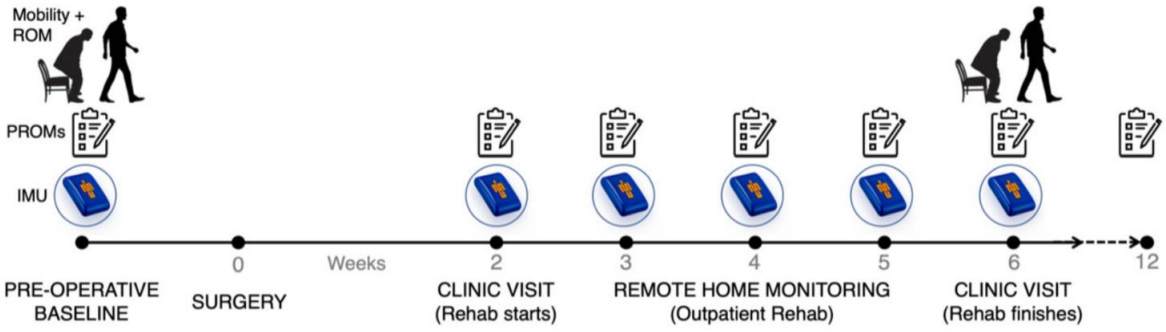



**Figure 1.4:** Longitudinal study design by Bolam, et al. (2021). While Bolam successfully measured OA gait following TKA, each collection only lasted 12 hours and may not capture the full range of gait metrics a patient may experience during a typical week. The study measured kinematic improvements from a pre-operative timepoint to post-TKA in cumulative impact load, asymmetry, and knee flexion.

The use of wearable sensors for gait analysis represents a significant opportunity for clinicians to improve patient care. Sensor-based monitoring has been shown to relate

well to laboratory-based measures, with wearable-derived gait metrics correlating with both self-reported quality of life and traditional functional tests (e.g., 6-minute walk, timed up and go) (Youn et al., 2020). For example, step time has been shown to correlate with the three subscales of the Knee Injury and Osteoarthritis Outcome Score: pain ($r = 0.64$), symptoms ($r = 0.76$), and quality of life ($r = 0.77$). Step time also showed strong correlations with functional tests, including the 6-minute walk ($r = 0.77$), timed up and go ($r = 0.75$), and 30-second chair stand ($r = 0.80$). An earlier study by the same group demonstrated that IMUs could estimate kinetic and kinematic parameters post-operatively in TKA patients, with Pearson correlations ranging from 0.51 to 0.79 between IMU-derived variables and ground reaction forces (Youn et al., 2018). Providing contextual information by classifying recognizable activities has also been shown to improve the interpretability of gait metrics for clinicians and patients alike (Chen et al., 2020). Several studies have used wearable sensor-based metrics as the foundation for research aimed at guiding clinical decision making, ranging from gait rehabilitation to predicting treatment response (Z. He et al., 2019; Kobsar et al., 2017). However, the processing and management of raw sensor data, especially for kinematic and spatiotemporal metrics, remains a critical but often overlooked step in improving their clinical utility especially in free-living collections.

### 1.2.4 Limitations of Wearable Sensor-based Gait Analysis

Despite their promise, wearable sensors present important limitations. While there are encouraging validation studies, there are notable differences remain that highlight the need for continued refinement of wearable-based gait analysis. A recent 2023 study by

Hafer, et al., found notable differences in knee flexion angle, with motion capture showing greater flexion than IMUs at specific gait phases (0-38% and 58-91% of the stride) (Hafer et al., 2023). While these differences did not alter clinical interpretation, they highlight limitations in the fidelity of IMU-based kinematic estimates under certain conditions. Additionally, achieving one-to-one equivalence with lab-based metrics is challenging, particularly in free-living environments (Hillel et al., 2019). Factors such as sensor placement variability, signal drift, and the absence of ground reaction force data can reduce accuracy (Kobsar et al., 2020a, 2020b, 2016). While sensors enable the collection of large volumes of ecologically valid data, this often comes at the cost of measurement precision. Standardized processing pipelines, improved alignment algorithms, and context-aware models may help mitigate these issues (Chen et al., 2016; Halilaj et al., 2018). Researchers and clinicians must therefore weigh the trade-offs between volume, ecological validity, and fidelity.

Ultimately, the use of using wearable sensors comes down to the trade-off between volume and detail that often accompanies free-living sensor data. A study like the one by Bolam, et al., benefited from repeated, ecologically valid measurements from wearable sensors. However, there was no motion capture conducted, nor it did not include standard gait metrics such as spatiotemporal parameters or peak kinematic events from the free-living collection. It also did not link sensor-derived outcomes with patient-reported or functional measures. Additionally, the 12-hour collection window each week still may not have captured fluctuations in pain or gait variability across days (Allen et al., 2009; Parry

et al., 2017), and the six-week follow-up may have been too short to reflect the full recovery trajectory, which can extend to six months or more (Paravlic et al., 2022). There also remain research gaps regarding how well any of the free-living metrics relate to motion capture-derived metrics.

Overall, the literature illustrates both alignment in spatiotemporal metrics but also potential divergence in kinetic and kinematic measurements between motion capture and wearable sensors. Nonetheless, wearable inertial sensors remain a promising tool for extending gait monitoring beyond the lab, where they can complement insights gained from traditional motion capture systems. Improvement in wearable sensors data quality, particularly in unstructured, real-world conditions where traditional methods of extracting gait characteristics often fail, could also improve agreement between methods. Machine learning offers a promising path forward by improvement of identification of periods of gait. The next section will outline machine learning approaches that may enhance the accuracy, consistency, and interpretability of wearable sensor-derived gait metrics in clinical practice. The next section will outline machine learning approaches that may enhance the accuracy, consistency, and interpretability of wearable sensor-derived gait metrics in clinical practice.

## 1.3 Machine Learning Applications for Gait Analysis

### 1.3.1 Machine Learning Overview

Machine learning (ML) is a subset of computational methods within the field of artificial intelligence that has rapidly advanced in recent years. This growth has been fueled

by increasing computational power, reduced hardware costs, and the availability of large-scale datasets. As a result, ML techniques are now widely applied across fields ranging from computer science to healthcare. Within the realm of gait analysis and knee OA, ML has a growing number of promising applications. For instance, markerless motion capture, based on pose estimation algorithms, has the potential to make biomechanical assessment more accessible in clinical settings. Wearable sensors, which generate large volumes of continuous data from a single user across multiple days, also present further opportunities for ML applications. Several studies have shown that biomechanical variables such as the knee adduction moment (KAM) and knee joint angle can be estimated using ML models (Seagers et al., 2022; Tan et al., 2022). Other applications include predicting post-treatment outcomes (Bini et al., 2019) and identifying responders versus non-responders to rehabilitation interventions (Kobsar and Ferber, 2018). Together, these developments highlight the growing potential for ML to enhance clinical decision-making and improve long-term management of OA.

A major challenge in analyzing free-living data collected with wearable sensors is the lack of contextual cues about the activity being performed at any given time. As such, a foundational step in processing this data is gait segmentation, which is the process of identifying periods of walking, or gait bouts, from continuous streams of sensor data. Accurate gait segmentation is essential for ensuring that downstream gait metrics reflect true ambulatory behavior. In structured lab settings, segmentation is straightforward due to the presence of trial markers and controlled environments. In contrast, free-living data is

noisy, unstructured, and interspersed with non-walking activities, making segmentation significantly more complex.

Traditionally, heuristic segmentation methods, also known as rules-based approaches, have been used to infer walking periods based on predefined thresholds such as signal magnitude, periodicity, or orientation. While interpretable and computationally efficient, these methods often fail in real-world conditions, especially among individuals with gait impairments such as those with knee (Mariani et al., 2013; Ullrich et al., 2020). The rise of wearable sensor research has coincided with the growth of ML in healthcare, creating new opportunities to apply data-driven approaches to segmentation. ML models, particularly those that learn from temporal data, are well-suited to handle the variability and scale of free-living recordings.

As these models are developed and evaluated, their performance is typically quantified using classification metrics such as accuracy, precision, recall, and the F1 score. *Accuracy* refers to the percentage of total predictions the model gets correct. *Precision* measures how many of the segments identified as true positives are actually true positives, while *recall* (also called sensitivity) captures how many true positives the model successfully identifies. The *F1-score* combines precision and recall into a single value, balancing false positives and false negatives. These metrics are especially important when working with imbalanced data, which could be common in free-living recordings where walking may be relatively infrequent. Within the context of using a ML model for gait

segmentation, the periods identified as walking could then use the rules-based methods for identifying individual steps or strides.

Improving gait segmentation could substantially enhance the reliability of gait metrics used as inputs to predictive models, enabling more ecologically valid assessments of mobility and treatment response. However, before these downstream applications can be realized, gait segmentation models must first be rigorously evaluated in clinical populations. The following subsections will provide an overview of current machine learning approaches for gait segmentation, with a focus on their use in populations with pathological gait, such as individuals with OA.

### 1.3.2 Machine Learning for Gait Segmentation

Enhancing gait segmentation is pivotal research, as its improvements can amplify the predictive capabilities of other ML models, such as those forecasting outcome scores following exercise intervention (Kobsar et al., 2017) or TKA (Bini et al., 2019). However, determining which approach or approaches are ideal is difficult, as many of the models were created primarily on younger healthy subjects. As noted in a review by Halilaj, et al., many ML models in biomechanics are developed and evaluated inconsistently (Halilaj et al., 2018). Each framework has been shown to have certain strengths and weaknesses. Heuristic methods offer simple approaches of segmenting data based on gait events from raw sensor data or frequency components (Mariani et al., 2013; Ullrich et al., 2020). While these are adequate for data collected in a controlled lab setting, they may not be as robust when collected remotely, especially in populations with pathological gaits.

The complexity of machine learning models can vary widely depending on their structure and training requirements. Supervised learning refers to models that are trained on labeled datasets, where each input is paired with a known output or class. Common supervised algorithms include support vector machines (SVMs), decision trees, and neural networks. In the context of gait segmentation, supervised models are typically trained to classify segments of sensor data as "walking" or "not walking" based on annotated ground truth. These models can achieve high accuracy, but their performance often depends on the quality and diversity of the training data. Because many supervised models are developed using controlled laboratory data, they may overfit to these conditions and struggle to generalize to free-living environments (Andaur Navarro et al., 2021), generating the labeled datasets required for training can be time-consuming and labor-intensive. In contrast, unsupervised learning involves algorithms that analyze unlabeled data to identify inherent patterns, groupings, or structures without prior knowledge of the output categories. These models can be particularly useful when labeled data is scarce or unavailable. In gait analysis, unsupervised techniques may be used to cluster movement patterns or detect walking bouts based on shared features. While unsupervised models reduce the need for manual annotation, they are often more complex to develop, interpret, and validate. They may also require more computational resources and careful parameter tuning to achieve robust performance.

Many ML gait models are created on young healthy adults. For example, a study by Chen, et al., using a SVM model as part of a larger study on activity recognition built

their model on 10 healthy adults. (Chen et al., 2020). With the model built on such limited data, it may be highly susceptible to overfitting and with deficiencies in robustness (Andaur Navarro et al., 2021; Halilaj et al., 2018). For example, it may perform well on healthy subjects, but not older, clinical populations. Another example by Li, et al., was developed on 5 young adults, and validated on a public dataset of 9 young adults. Activity recognition is achieved using a supervised model called a bidirectional long short-term memory (BiLSTM), where extracted spatial features are obtained from IMU sensors using a residual convolutional neural network (CNN) before obtaining additional dependencies of feature sequences (Figure 1.5) (Li and Wang, 2022). This model performed well in terms of identifying walking, both within the test set data (recall: 1.00; precision: 0.99), as well as on a separate public dataset (recall: 0.98; precision: 0.99). Similar to the Chen study, while this model performs well on a healthy population, it is unclear whether either of these models would generalize to individuals with slower, pathological gait, such as Parkinson's disease or severe OA.

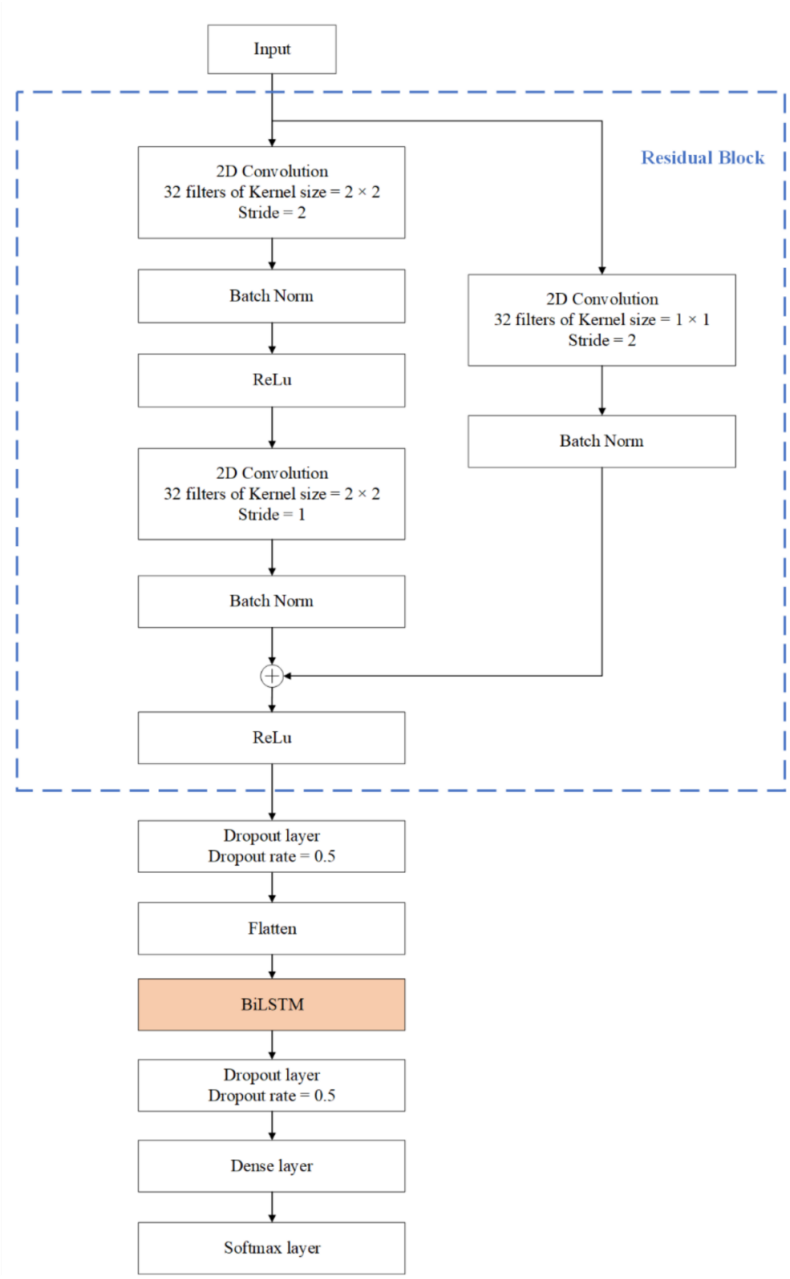



**Figure 1.5**: Deep learning framework utilized by Li and Wang (2022) which combines a residual block and bidirectional long short-term memory (BiLSTM). The residual black uses convolutional neural networks to extract spatial features from the signal, while the

BiLSTM captures the forward and backward temporal information from time sequences A framework like this could be used to segment OA gait data into walking bouts.

There are also examples of ML gait segmentation models which incorporate higher numbers of healthy participants. An approach by Martindale, et al., combined three public datasets to build a large robust model with over 100 subjects evaluated while using one dataset in the training and testing and the other two datasets for validation (Martindale et al., 2021). This model can find edges of a gait segment using a CNN and uses the recurrent neural network (RNN) to model the temporal dependencies of the data (i.e., stride time and swing time durations of detected strides). All data was collected laboratory-based and evaluated across several different activities (walking, sitting, jogging, running, stairs, cycling, jumping, resting). The model resulted in high F1-scores for activity recognition and phase identification across all activities. Walking resulted in an F1-score of 95.7 for activity recognition and 98.2 for phase identification, to predict the exact timing and duration of activities. Nevertheless, while models such as the one described by Martindale are larger and potentially more robust, they are still created with laboratory-based data, primarily from healthy participants. This remains an important limitation when applying these models to free-living, clinical gait data, as the model will likely will not perform as well as the performance metrics indicate.

### 1.3.3 Limitations of Machine Learning in Gait Segmentation

While machine learning offers promising improvements over heuristic approaches, several limitations must be considered when applying these methods to gait segmentation. The review by Halilaj, et al., notes a number of challenges regarding applying ML models to biomechanical data, primarily regarding generalizability and (Halilaj et al., 2018). As also noted in the previous section, ML models are trained on data collected in structured environments using healthy populations, which may not reflect the variability and irregularities seen in free-living data or clinical populations such as those with knee OA. This is also potentially exacerbated if the model is trained on a small number of participants, which as noted by Andaur Navarro, et al., makes models more prone to overfitting (Andaur Navarro et al., 2021). As a result, models may fail to perform reliably outside the conditions under which they were trained. Another concern is model transparency and interpretability. As noted by Halilaj, high-performing models, particularly deep learning architectures, operate as "black boxes," making it difficult for clinicians or researchers to understand how classification decisions are made. This lack of interpretability may limit clinical trust and adoption, especially in contexts where explainability is crucial for decision-making.

### 1.3.4 Applying a Machine Learning Framework for Osteoarthritis Gait Segmentation

Clinical gait by nature is likely to be more complex and dissimilar between patients as compared to healthy control subjects. However, despite this potentially added complexity, it doesn't necessarily mean the gait is unable to be modeled. Unlike healthy

gait models that perform well based on minimal numbers of subjects, clinical gait models will likely require more data to produce robust models that generalize well to data they haven't seen before (Halilaj et al., 2018). Because of the often slower and pathological gait of severe OA patients, heuristic-based gait segmentation may not properly identify walking bouts, especially in free-living collection. Similar pathological gaits, such as Parkinson's disease, are a well-studied area using ML models to segment gait from wearable sensors (Martindale et al., 2021; Roth et al., 2021). While there are a wide range of potential applications for ML frameworks within OA, improving gait segmentation is of particular importance. Improvements of the data quality from gait segmentation can positively affect everything downstream that relies on it (e.g., metrics, other models, etc.). This section will primarily discuss research that has been deployed, or could be via transfer learning, for gait segmentation on an OA population.

Several ML models have been developed for gait analysis with the basis of analyzing pathological gait. These studies typically analyze populations with known pathological gait, such as Parkinson's disease. One unsupervised approach, known as Hidden Markov Modeling (HMM), was demonstrated by Roth, et al. Specifically, this model was used to make predictions of human gait based on IMU data by combining two models to segment data into strides and transitions (Roth et al., 2021). The combined model was trained on both laboratory-based and free-living data from 28 patients with Parkinson's disease with sensors placed on the instep of each foot. The models developed resulted in high evaluation metrics. Laboratory-based data segmentation performed by the HMM were

comparable to a separate heuristic-based dynamic time warping algorithm (F1-scores: 96.2% vs 94.6%, respectively). However, on out-of-lab data, both HMM trained using laboratory-based and free-living data both considerably outperformed the dynamic time warping (F1-scores: 92.2, 92.4, and 85.1, respectively). ML modeling on pathological gait segmentation such as this could potentially be applied to similar clinical populations and out-perform heuristic-based models.

Thus far, studies using ML techniques to segment gait in an OA population are limited. Most studies are focused on predicting outcomes or other parameters such as knee joint angle. A study by Renani, et al., analyzed segmentation of OA gait using a deep neural network framework, while also analyzing the optimal sensor combination for prediction accuracy of different outcome variables (Sharifi Renani et al., 2020). Interestingly, in terms of prediction accuracy, the shank location was consistently ranked highest, while the pelvis placement, which is a common sensor placement location in gait research, ranked lowest. However, this study was completed laboratory-based and did not do an out-of-lab assessment using sensors, which, as has been noted, is a common limitation in many studies.

Regardless of a lack of current research in the OA population, there has been progress in ML-based gait segmentation models on other pathological gait (e.g., Parkinson's disease) patients that could be utilized on OA gait via transfer learning. The previously discussed Roth and Martindale studies offer promise in terms of developing advanced gait segmentation models in pathological gait populations, which could include

OA (Martindale et al., 2021; Roth et al., 2021). The Roth HMM framework was trained on laboratory-based data but appears to have been successfully transferred to free-living data. However, other HMM frameworks, such as done by Attal, et al., were not as accurate, achieving around 80% accuracy (Attal et al., October 25-29). Similarly, the Martindale model was developed on a large dataset of primarily healthy adults, even though it did include a small number of pathological gait subjects (e.g., four Parkinson's disease patients) in their study. Additionally, the Martindale model only involved laboratory-based activities and was not developed nor evaluated on free-living data. While the BiLSTM approach shown by Li and Wang was initially created on a sample of 5 subjects, it's high performance on public datasets is particularly of interest. The model framework similarly could be retrained to be deployed on an OA population. Taken all together, while there is substantial need for ML-based gait segmentation models developed on truly free-living gait data from OA patients, the surrounding research is promising but it has yet to be done.

### 1.4 Gaps in literature

There is a growing and promising body of research supporting the use of both laboratory-based and wearable systems for clinical gait analysis. Laboratory systems, such as optical motion capture and force plates, remain the reference standard due to their high spatial and temporal precision. Meanwhile, wearable sensor systems offer distinct advantages in terms of portability, scalability, and the ability to collect ecologically valid data outside of controlled settings (Muro-de-la-Herran et al., 2014). However, there is a persistent disconnect exists between conventional laboratory-based gait analysis and

wearable sensor-based approaches, particularly when data are collected in free-living environments. Most gait analysis research continues to be conducted in structured lab settings, which, while highly controlled, may not fully capture the variability or functional relevance of gait in daily life. This may limit the generalizability of findings to real-world clinical populations. Although wearable sensors have demonstrated potential to bridge this gap, few studies have fully leveraged their capabilities for longitudinal ambulatory monitoring, especially in individuals with knee OA. As a result, the clinical relevance of wearable-derived gait metrics, such as stride time, in free-living conditions remain insufficiently characterized in OA populations.

Similarly, while machine learning has shown promise for tasks such as gait event detection and segmentation, many existing models have important limitations. These include reliance on small, homogenous datasets, sometimes with as few as five participants, and training primarily on healthy individuals. When applied to clinical populations or data collected outside the lab, model performance often declines substantially. Moreover, there is significant opportunity to develop more robust, patient population-specific models that integrate the volume and ecological validity of wearable sensor data with the biomechanical precision of laboratory systems.

Another notable limitation across the field is the relative lack of longitudinal data tracking individuals throughout the rehabilitation continuum. Most studies offer only a snapshot view, making it difficult to understand how gait patterns evolve over time or in response to interventions. Longitudinal, real-world data could provide a much richer

foundation for predictive modeling and personalized rehabilitation strategies. Together, these gaps highlight the need for more integrated approaches that combine wearable sensors, machine learning, and laboratory-based methods. Doing so may allow for more comprehensive, scalable, and clinically meaningful assessment of gait in populations with movement impairments such as knee OA.

### 1.5 Study Objectives

Wearable sensors and deep learning hold promise for advancing gait analysis, particularly for free-living assessment. While studies like Bolam et al. have begun exploring this potential, they often collect limited free-living data and lack integration with lab-based metrics or patient-reported outcomes. Similarly, although various machine learning models have been applied to gait segmentation, few focus on individuals with osteoarthritis, despite the promise of deep learning models (e.g. BiLSTM) for improving segmentation accuracy. To date, much of this research has been conducted in isolation (e.g. on healthy populations, in lab environments, or at small scale) without a unified framework for clinical application. This thesis aimed to fill that gap by developing a clinically relevant gait analysis framework that uses wearable sensors and deep learning to detect changes in gait in older adults with knee OA. Additionally, this work examined the sensitivity of wearable-derived gait metrics compared to traditional lab-based assessments. A key focus is on the use of shank-mounted sensors, given their demonstrated ability to capture both spatiotemporal and kinematic gait metrics (e.g., stride time, lateral knee motion), as well as a supporting work that reflects a realistic, sparse sensor deployment.

To achieve these goals, the thesis is structured around four research aims designed to build and validate a framework for clinical use:

- **Aim 1**: Develop a basic framework for collecting free-living gait data using wearable sensors that identifies walking bouts and extracts key kinematic (e.g., peak impact accelerations, peak angular velocities, etc.) and spatiotemporal gait parameters (e.g., stride time, stance time, swing time).
- **Aim 2**: Refine walking gait segmentation methods for free-living wearable sensor data using a previous deep learning approach (Li and Wang, 2022) trained on a more diverse dataset featuring participants with OA.
- **Aim 3** Assess the agreement of gait metrics (i.e., stride time, stance time, swing time, and peak angular velocity) in-lab and free-living against a reference-standard system in end-stage OA patients prior to surgery.
- **Aim 4:** Examine longitudinal sensitivity of the wearable sensor framework to detect changes in gait following end-stage surgical treatment, as compared to a conventional, laboratory-based gait analysis system.

In summary, the goals of this thesis are to not only develop a framework for using wearable sensors for free-living gait analysis, but to evaluate framework on adults with late-stage knee OA. The development of the framework, from conception to creating a deep learning model to evaluating agreement between in-lab gait to free-living gait, is discussed in Chapter 2, Chapter 3, and Chapter 4, with the evaluation of the framework addressed in Chapter 5.

# Chapter 2: Development of a Data-Informed, Clinically Viable Gait Framework

## Preamble

As discussed in the literature review, there is growing interest in using wearable sensors for clinical gait analysis. However, a central challenge remains: how best to analyze data collected in free-living environments, particularly in ways that are both clinically meaningful and adaptable across diverse use cases. Previous studies have introduced processing frameworks tailored to specific clinical populations (e.g. ACL rehabilitation) or sensor configuration (e.g., thigh sensor) (Gurchiek et al., 2019), which enhances internal validity but limits broader applicability. Conversely, some frameworks take a fully agnostic approach, offering modular tools applicable across populations and sensor placements, such as the open-source platform introduced by Beyer, et al. (Beyer et al., 2024). While this platform is a major advancement, its current scope is restricted to wrist and ankle placements and does not yet fully address the needs of lower-limb pathologies like knee osteoarthritis (OA). The method relies solely on the mediolateral gyroscope for gait identification; as noted earlier, simple heuristic approaches are often less robust in clinical populations. Incorporating additional IMU channels could improve detection accuracy, but this requires further study.

This chapter briefly and broadly describes the development of a gait analysis framework that aims to balance these extremes: specific enough to be validated in a target clinical population (knee OA) using a strategic sensor placement (shank), yet modular and

extensible enough to be adapted for broader clinical applications. The framework was implemented in Python and serves as both a benchmark for future enhancements (e.g., deep learning-based models) and a foundation for real-world clinical deployment, as outlined in future chapters. By developing a more general gait framework that is sensor and clinical population agnostic, there should be broader applicability when being deployed to clinical use. The sections that follow describe the rationale, design considerations, and standardized processing pipeline of this prototype, which serves as the benchmark against which later refinements will be evaluated and, if warranted, incorporated into the finalized clinical pipeline.

## 2.1 Framework Considerations

Several key factors guided the development of the proposed gait analysis framework, including sensor placement, sensor location, number of sensors, the impact of and repeatability in a knee OA population. These decisions were informed by both practical deployment constraints and scientific considerations regarding signal quality and clinical relevance.

### 2.1.1 Placement and Location

Sensor placement (e.g., self-placed vs. researcher-placed) and anatomical location (e.g., back, thigh, shank, ankle, or foot) directly affect data quality, patient compliance, and clinical interpretability. Unsupervised self-placement offers scalability, enabling remote deployment via mail and easy replacement if a sensor is dislodged. This approach, used by

Beyer, et al., requires rigorous wear-time validation, since patients may remove sensors daily (Beyer et al., 2024). Researcher placement, while more labor-intensive, ensures consistent positioning and has been more commonly used in the literature. However, few studies have directly compared the reliability of researcher- versus self-placed sensors, and fewer still have evaluated methods to mitigate placement variability.

### 2.1.2 Sensor Location

A range of sensor locations have been explored in prior work. According to a review by Kobsar, et al., OA patients have been studied using sensors on the back, thigh, shank, ankle, and foot, with the back being the most common (Kobsar et al., 2020b). While back placement is convenient, it provides limited information about joint-specific motion. Foot sensors can aid in estimating spatiotemporal metrics like stride time and length but similarly offer limited insight into joint dynamics. The shank, by contrast, was selected in this framework for its ability to capture both stride-level timing and joint-relevant metrics such as peak acceleration and angular velocity. Its bony attachment point just below the knee also reduces skin motion artifact compared to placements like the thigh, which has been found to have larger movement artifact even after alignment correction (Mihy et al., 2022).

### 2.1.3 Number of Sensors

Many in-lab studies deploy multiple sensors to estimate joint kinematics or evaluate sensor fusion approaches (Hafer et al., 2023; Kobsar et al., 2016). However, multi-sensor setups are less feasible for long-term, free-living use due to burden on participants. Recent

studies suggest that meaningful gait analysis is possible with a reduced number of sensors (e.g., 2-3), which improves compliance and supports clinical translation (Bolam et al., 2021). This framework was designed to enable interlimb comparisons (e.g., symmetry) while limiting patient burden. The determination of using self-placed sensors also plays a role in this decision, as each sensor self-placed could introduce error from incorrect placements and potentially decrease patient compliance.

### 2.1.4 Repeatability and Population Considerations

Repeatability of gait metrics in OA populations remains understudied, particularly outside the lab. While healthy adults have been evaluated extensively under controlled conditions (Kobsar et al., 2020a), free-living reliability in clinical populations has received less attention. Therefore, there was a need to assess repeatability of both waveforms and discrete peak variables from accelerations and angular velocities in an OA population outside of the lab.

## 2.2 Foundational Methodological Studies

This section summarizes two pre-thesis studies that informed the development of the gait framework. These studies, published in the *Journal of Biomechanics* in 2022 and 2023, examined (1) the validity of self-placed sensors and (2) the repeatability of gait metrics in an OA population.

### 2.2.1 Researcher-Placed vs Self-Placed Sensors

The first study (Ruder et al., 2022) evaluated whether self-placed sensors could produce gait data comparable in validity and reliability to those placed by a researcher. It

also tested whether a principal component analysis (PCA)-based alignment correction, adapted from Hafer, et al., (Hafer et al., 2020) could mitigate placement variability. Young, healthy adults self-placed a sensor on their left tibia, while a researcher placed a second sensor on the right tibia. Participants completed two in-lab walking sessions using a marker-based optical motion capture system as the reference standard. Gait metrics were extracted from both sensors before and after PCA-based alignment and evaluated using intraclass correlation coefficients (ICCs) for discrete peak variables and coefficients of multiple correlation (CMCs) for waveform data.

After alignment correction, both self-placed and researcher-placed sensors showed good-to-excellent validity and test-retest reliability for key gait metrics, including vertical, anterior-posterior, and resultant acceleration peaks. Although placement consistency was lower for self-placed sensors (ICC = 0.55 vs. 0.85), corrected waveform data remained highly reliable (CMC $\geq$ 0.93). Slight reductions in validity were observed for mediolateral acceleration and frontal plane angular velocity in the self-placed sensors, likely due to placement variability and lower signal amplitudes. Overall, these results support the feasibility of self-placed sensors when alignment correction is applied, potentially enabling their use in decentralized settings such as remote monitoring or clinical trials.

### 2.2.2 Repeatability of Gait Metrics from Out-of-Lab Gait Assessment in OA Patients

The second study (Ruder et al., 2023) assessed the between-day repeatability of gait metrics derived from wearable sensors in a clinical population. Nine adults with moderate-to-severe knee OA completed four out-of-lab gait assessments, approximately one week

apart, during a standardized six-minute walk test. Sensors were placed on the tibia and aligned using the same PCA-based method from the previous study.

Gait cycles were extracted, and both waveform-level and discrete peak variables (e.g., stride time, vertical acceleration) were evaluated for test-retest reliability. Results showed good-to-excellent reliability for most metrics, particularly when multiple steps were averaged. Ensemble waveform reliability improved with increasing step counts. ICC values for key discrete metrics exceeded 0.75, particularly in the more affected limb. As in the first study, slightly lower reliability was observed in mediolateral acceleration and frontal plane angular velocity. The findings indicate that with proper preprocessing and alignment, wearable sensors can provide stable and clinically meaningful gait metrics in free-living environments, supporting their use in longitudinal monitoring and intervention tracking in OA populations.

### 2.2.3 Implications of Findings

These two studies played a central role in shaping the design of the proposed framework. Although the first study demonstrated that PCA alignment could effectively correct for placement variability in self-placed sensors, pilot testing revealed practical limitations: older adult participants often struggled with placing on the correct limb, forgot to charge or activate the sensors, or inadvertently left them unplaced. Of 10 patients recruited during early testing, usable data were obtained from only two. Consequently, the decision was made to adopt researcher placement for all subsequent data collections. Nevertheless, PCA-based alignment remained a key component of the framework to

account for variation introduced by different researchers. Similarly, the use of two sensors, one on each shank, was chosen to balance measurement richness with participant burden, while enabling assessment of interlimb symmetry. The second study confirmed that tibial sensors yield reliable waveform and discrete metric data in OA patients under free-living conditions. These findings provided essential evidence for the clinical validity of the framework and helped identify which gait metrics (e.g., peak sagittal angular velocity) are most robust for longitudinal tracking in this population.

With these foundational studies complete, the framework was ready for broader implementation. The next section provides a high-level overview of its structure and processing steps.

### 2.3 Description of Gait Framework

Initial development of the gait framework was informed by both laboratory and free-living datasets collected during pilot studies. The design emphasized modularity, enabling different segmentation methods and analyses to be integrated as the framework evolves. Although this chapter does not provide formal validation metrics, subjective evaluation of algorithm performance in identifying walking bouts and gait events guided initial development. Subsequent chapters will quantify the framework's performance relative to other segmentation methods (e.g., deep learning; Chapter 3) and use this foundation for further analyses (Chapter 4 and 5). Wearable sensor data are processed using a standardized pipeline implemented in Python, designed to support both in-lab and free-

living applications. The pipeline consists of five core stages: preprocessing, gait segmentation, event detection, outlier rejection, and metric extraction (Figure 2.1).

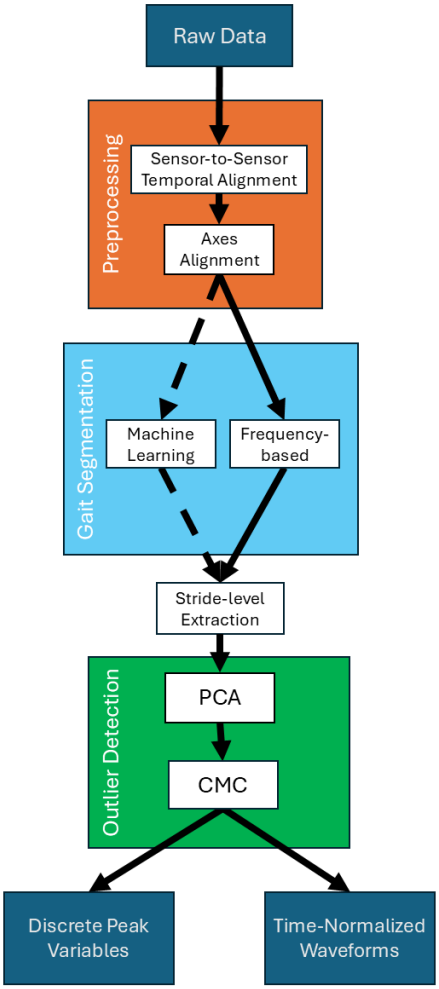



**Figure 2.1:** Simplified workflow for processing pipeline, run in parallel for both sensors. Raw data is preprocessed and then undergoes gait segmentation to identify walking bouts. Walking bouts are then evaluated using event detection to further segment into individual strides and evaluated for outliers before outputting metrics and averaging parameters.

### 2.3.1 Preprocessing

Raw data from bilateral tibial sensors are automatically loaded based on sensor naming conventions. The framework supports reading files from both the binary .cwa format used by the Axivity AX6 devices (Axivity AX6, 100Hz, Axivity Ltd., Newcastle, UK) and .csv data from alternative sensors (e.g. IMeasureU). Sensor data are temporally aligned to the later-starting device to minimize clock drift at the beginning of recordings. Orientation correction adapted from Hafer, et al., is performed in two stages (Hafer et al., 2020). First, the vertical axis is identified and aligned to gravity using a static acceleration window. Then, PCA-based optimization is applied during walking to align the mediolateral gyroscope axis. This ensures that axes are consistently oriented across sessions and participants. Physical activity levels are then estimated using shank-specific cut points developed by Gafoor, et al., producing both absolute and normalized activity counts across predefined intensity bins (Gafoor et al., 2024). These outputs support physical activity monitoring of interest to clinicians alongside gait analysis. The physical activity counts can also be used for wear time detection and removal of non-wear sections of data after the sensor is removed.

### 2.3.2 Gait Segmentation

Gait segmentation is performed independently for each sensor using a user-defined bout detection method. A more detailed description is included in Chapter 3, but as noted, this step is designed to be flexible to run different methods in parallel for evaluation. The default implementation is a frequency-based heuristic method developed by Ullrich, et al.,

originally evaluated for individuals with Parkinson's disease. (Ullrich et al., 2020). This method was chosen for its simplicity and demonstrated effectiveness in altered gait populations, having been originally developed for free-living gait detection in individuals with Parkinson's disease. In short, this method analyzes 10-second windows of the mediolateral gyroscope signal (e.g., sagittal plane swinging motion of the leg). If the signal energy exceeds the defined threshold, the dominant frequency is estimated using an autocorrelation. A fast Fourier transform is then applied, and the presence of peaks at two or more of the first four harmonics of the estimated dominant frequency indicates cyclical gait for the window of interest. Detected walking windows are stored as Boolean masks, allowing for efficient reuse and facilitating downstream processing. The modular structure of this step allows for easy comparison between segmentation methods, such as heuristic versus deep learning approaches.

### 2.3.3 Event Detection and Stride Segmentation

Within each walking bout, individual strides are identified using an event detection method adapted from Mariani et al. (2013). Mid-swing peaks are first located in the mediolateral angular velocity signal. The interval between successive mid-swing peaks is divided in half: the largest resultant acceleration peak in the first half is classified as heel strike, and the largest in the second half as toe off. This approach was selected based on pilot testing and prior literature indicating consistent alignment with ground truth. Discrete variables were calculated between each during each stride (heel strike to heel strike), stance (heel strike to toe off), and swing (toe off to subsequent heel strike) as needed. Specifically,

from these events, spatiotemporal gait metrics (i.e., stride time, stance time, and swing time) are computed, as well as peak angular velocities and accelerations are extracted for each phase (stride, stance, swing). All strides are time-normalized to 101 points to allow for ensemble averaging and comparison across participants. These candidate strides are then subjected to a multi-stage outlier detection protocol.

### 2.3.4 Outlier Detection

After event detection, each candidate stride is time-normalized and evaluated through a multi-step outlier detection framework designed to ensure signal fidelity and biomechanical consistency. First, principal component analysis (PCA) is applied to a stride-level signal (typically the mediolateral angular velocity). Strides are projected into a 15-component score space to capture at least 90% of variance, and multivariate z-scores are calculated. Strides with z-scores exceeding ±2.5 in any component are flagged as statistical outliers and removed. This step captures signal-level anomalies such as abrupt noise spikes, artifacts, or irregular gait patterns. Next, the remaining strides are evaluated using the coefficient of multiple correlation (CMC), which quantifies biomechanical similarity between each individual stride and a reference waveform (defined as the mean of all in-lab strides). Strides with a CMC value below 0.5 are excluded, reflecting insufficient waveform similarity. This added layer filters biomechanically inconsistent strides that may have passed the statistical threshold but diverge from typical gait patterns. Only strides passing both filters are retained for metric extraction.

### 2.3.5 Metric Extraction

Final outputs include both stride-wise peak metrics and time-normalized waveforms. These outputs support condition comparisons, longitudinal tracking, and group-level analysis.

## 2.4 Conclusions

The development of this gait analysis framework involved a series of practical and methodological decisions intended to support both research utility and clinical translation. While not itself a standalone study, this work represents a critical foundation for the broader dissertation. The design was informed by two foundational methodological studies that addressed key barriers to real-world implementation: the impact of sensor placement variability and the reliability of gait metrics in an OA population. These studies demonstrated that with proper alignment correction and careful deployment, wearable sensors can provide valid and repeatable gait measures in both lab-based and free-living settings.

The framework is deliberately modular, supporting integration of future segmentation and event detection methods while maintaining a standardized core for repeatable analysis. For the current work, a researcher-placed, dual-shank configuration was selected based on pilot testing, participant compliance, and metric reliability. PCA-based alignment and multi-step outlier detection further enhance the robustness of the extracted gait metrics, making the framework suitable for clinical monitoring and research applications.

As wearable technology continues to advance, this framework offers a platform for expansion. Additional modules, such as sleep tracking or activity classification, could be added in future iterations. Likewise, as new gait segmentation methods emerge, including those based on deep learning, they can be integrated into the existing pipeline. One such method is explored in Chapter 3.

# Chapter 3: Augmenting a ResNet + BiLSTM deep learning model with clinical mobility data outperforms heuristic frequency-based model for walking bout segmentation (Study 1)

## Preamble

This chapter presents my first formal dissertation study, in which I trained, tested, and evaluated a machine learning activity classification model using data from both healthy adults and patients with knee osteoarthritis to identify gait bouts. The previous chapter described a baseline gait detection framework built on a heuristic frequency model. In the following chapter, I extend that framework by retraining a promising machine learning model and assessing its performance independently and against the heuristic approach. I systematically expanded the training dataset in stages, retraining and re-evaluating the model after each increment to track performance gains. Overall, this work advances the development of robust classifiers tailored to individuals with altered gait.

This work was prepared for submission to *Sensors*, with the following coauthors:

Matthew C. Ruder, Vincenzo E. Di Bacco, Kushang Patel, Rong Zheng, Kim Madden, Anthony Adili, and Dylan Kobsar

**Abstract**

Wearable sensors have become valuable tools for assessing gait in both laboratory and free-living environments. However, detection of walking in free-living environments remains challenging, especially in clinical populations. Machine learning models may offer more robust gait identification, but most are trained on healthy participants which limits their generalizability to other populations. To extend a previously validated machine learning model, an updated model was trained using an open dataset (PAMAP2), before progressively including training datasets with additional healthy participants and a clinical osteoarthritis population. The performance of the model identifying walking was also evaluated using a common heuristic, frequency-based gait detection algorithm. Results showed that the model trained with all three datasets performed best in terms of activity classification, ultimately achieving a high accuracy of 96% on held out test data. When the model was used for identification of walking activity classification, the model generally performed on par with the heuristic, frequency-based method. However, for patients with slower gait speeds (<0.8 m/s), the machine learning model maintained high recall (>0.89), while the heuristic method performed poorly with recall as low as 0.38. This study demonstrated the potential of training with clinical data to improve model robustness for pathological gait. Although further validation, especially in free-living environments, is needed, this study provides a framework for enhancing existing model architecture using diverse datasets and highlights the importance of dataset diversity when developing models for clinical applications.

## 3.1 Introduction

Wearable inertial measurement unit (IMU) systems have emerged as powerful, cost-effective tools for evaluating walking gait in clinical contexts. These evaluations are essential for understanding disease status, progression, and response to treatments (Caramia et al., 2018; Carcreff et al., 2020; García-de-Villa et al., 2025). In knee osteoarthritis (OA), one of the most prevalent musculoskeletal conditions (Chapple et al., 2011), IMUs have been used to assess patient function (Hafer et al., 2020; Ismailidis et al., 2020), track disease progression (Costello et al., 2020), and monitor rehabilitation response (Z. He et al., 2019). While these assessments can be invaluable in clinical environments, free-living daily life offers the potential for deeper insights into functional capacity, disease state, or response to treatment that may not be captured in controlled settings (Bolam et al., 2021; Hillel et al., 2019). However, working with free-living IMU data presents significant challenges due to the large volume of data across various activities, many of which may not be of interest. Accurately identifying activities, or simply when a participant is walking, remains a key hurdle in translating IMU technology for real-world applications outside of controlled laboratory environments.

Traditional heuristic, or rules-based, methods are often used to identify periods of walking, referred to as walking bouts. One of the simplest methods is identifying peak acceleration impacts that exceed a certain threshold, which correspond to heel strikes during walking (Selles et al., 2005). While this method is easy to implement, it lacks robustness and is poorly suited for free-living data due the variability, not only between

individuals' gait but also within an individual's gait in uncontrolled environments (Hillel et al., 2019). Alternatively, the cyclical nature of walking gait can be effectively incorporated into walking bout identifier algorithms by examining the signal's frequency content. For example, the harmonic frequencies of the IMU signals can be analyzed using a fast Fourier transform (FFT) to identify walking behavior by detecting harmonics that are indicative of gait (Ullrich et al., 2020). This approach has demonstrated strong performance in laboratory settings (sensitivity = 0.98), with only minimal drop-off when deployed on semi-structured, unsupervised, and remotely collected validation data (sensitivity = 0.97). While this improvement over simple peak heuristics is promising, there are still challenges related to thresholding estimates for harmonics, which can limit the generalizability and robustness for walking classification. Participants also generally walk slower in free-living assessments, which may complicate gait detection in populations with altered gait that already move slower than healthy populations (Alkjaer et al., 2015; Hillel et al., 2019). Therefore, developing gait identification algorithms that are robust and agnostic to these variations must be accounted for before deployment on free-living data.

Machine learning (ML) methods have emerged alongside wearable IMUs and offer improved predictive power over traditional heuristic models. Techniques like support vector machines (Gurchiek et al., 2019) and Hidden Markov Models (Roth et al., 2021) have been used to identify walking bouts using handcrafted features from three-dimensional acceleration and angular velocity data. While effective, these methods rely on feature engineering and may not capture the full complexity of human movement. More

advanced deep learning approaches, such as Convolutional Neural Networks, can automatically extract relevant features from raw IMU data, enabling more effective gait analysis (Martindale et al., 2021). However, given the temporal dependence of human movement signals, utilizing a deep learning model that also integrates the temporal structure of the signal can be critical to identifying walking bouts in free-living data.

Long Short-Term Memory networks, a type of recurrent neural network, are deep learning models designed to capture long-term dependencies in time-series data. These advanced networks excel in tasks where temporal information is essential, such as sequence prediction or activity recognition, because they can retain information over long sequences and mitigate the vanishing gradient problem often encountered in traditional recurrent neural networks. A recent study by Li and Wang tested the use of this model to identify activities, including walking, from an open dataset (PAMAP2) (Li and Wang, 2022). The architecture consists of two main components: 1) the Residual Convolutional 2D block (ResNet), which extracts features from a window of IMU data, and 2) the Bidirectional Long Short-Term Memory (BiLSTM) network, which captures long-range dependencies and reduces information loss in sequential data. This model architecture is particularly intriguing because it has demonstrated strong performance across various datasets, ranging from those with multiple sensor placements to a single sensor positioned just below the knee. Specifically, the authors found that the model identified activities with over 95% accuracy trained independently across three datasets: the full PAMAP2 dataset, another open dataset (WISDM), and a lab-created dataset with a single proximally placed sensor

on the shank. While these findings are promising, they share limitations common to many models, as highlighted in a recent review (Halilaj et al., 2018). Specifically, the models were trained and validated exclusively with healthy subjects, with no evaluation on individuals with altered gait, limiting the generalizability of the findings. Moreover, these results were obtained without a specific, separate test set and were not compared to traditional heuristics, making it difficult to assess the model's performance relative to simpler, more established methods.

Rather than developing a new ML architecture, this work extends Li and Wang's (2022) approach to create a clinically viable system that uses a single lower-limb sensor capable of identifying activities in both healthy individuals and those with knee OA, with a particular focus on walking bouts due to their significance in these populations. A similar model architecture to that of Li and Wang (2022) was implemented, along with additional training data (from both healthy individuals and those with knee OA) to enhance classification performance. Additionally, the performance of this trained deep learning model was evaluated against a common heuristic gait detection algorithm (Ullrich et al., 2020) to compare ML performance with a heuristic approach, using a more comprehensive test set across a variety of gait speeds. Specifically, we hypothesized that (i) the deep learning model, when trained with additional healthy and clinical data, would outperform the base open dataset model on a clinical test set; (ii) it would outperform the heuristic gait detection algorithm; and (iii) this advantage would be particularly evident in individuals with lower gait speeds. By doing so, this study aims to develop a clinical gait detection

model using only a single sensor located on the shank.

## 3.2 Methods

### 3.2.1 Description of Datasets

Data used in the current study were sourced from three distinct datasets: (a) the PAMAP2 dataset (Reiss, 2012), (b) a new healthy adult dataset, and (c) a new clinical dataset of adults with knee and hip OA. Each dataset contributed to the development, validation, and testing of the model, with data obtained from sensors placed at varying locations on the shank. The PAMAP2 dataset, which includes a variety of activities and sensor placements (chest, wrist, and ankle), served as the base dataset for model development. In this application, only the ankle sensor data were used to simplify the sensor array and facilitate the extraction of meaningful gait metrics from OA patients using this location, extending the scope of current work (Kobsar et al., 2020b). The inclusion of (b) a secondary healthy dataset, in which sensors were placed more proximally on the shank, was intended to promote model development that is more robust to variations in sensor placement along the shank (Figure 3.1). Next, the limited number of healthy participants in the base dataset may restrict its generalizability to older adults with gait impairments. To address this, (c) a clinical dataset comprised of adults with knee OA, collected approximately two weeks prior to joint replacement surgery, was included to increase population diversity. This dataset represents end-stage OA and provides valuable data on individuals exhibiting a range of gait speeds and condition-related gait alterations (Alkjaer et al., 2015). The datasets and activity labels are described in full below, with the

full protocols for the healthy and clinical datasets available in Section 3.6 Supplemental Material.

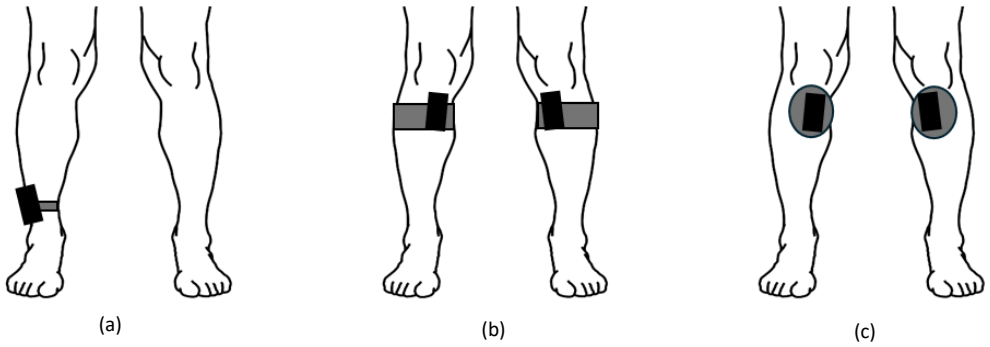



**Figure. 3.1** Sensor placements for PAMAP2 (a), healthy dataset (b), and clinical dataset (c). The PAMAP2 sensor was attached to the dominant side lateral ankle with a strap. The healthy dataset was attached with a semi-elastic strap with sensor placement medial and inferior to the knee. The clinical dataset used the same attachment location as the healthy dataset but instead attached with medical grade tape. Please note that sensors are not to scale, as increased size is used to clearly show placements.

1) *PAMAP2 Dataset*

This open dataset was created as a benchmark dataset for physical activity monitoring (Reiss, 2012). The dataset contained 9 participants (8M, 1F; age: 27.2 ± 3.3 years; height: 179.4 ± 8.4 cm, weight: 80.9 ± 10.3 kg) wearing multiple IMU sensors (Colibri wireless IMUs, Trivisio) sampled at 100 Hz while completing 18 varying activities.

For the purposes of developing a clinical gait detection model for this study, the only sensor utilized was the IMU attached on the dominant side ankle (Figure 3.1a). There are 12 activities that were completed during the protocol (lying, sitting, standing, walking, running, cycling, Nordic walking, ironing, vacuuming, jumping rope, ascending and descending stairs) and up to six optional activities (watching TV, computer work, driving a car, folding laundry, and playing soccer). Data from Subject 109 in the dataset only contained labeled data for jumping rope and was therefore excluded. All remaining subjects were included. A total of 13 activity labels were identified in the datasets used and updated during subsequent preprocessing.

2) *Healthy Lab Dataset*

This dataset was collected as part of a concurrently collected study. The McMaster Research Ethics Board approved the study (MREB 6120), and all participants provided their informed consent prior to enrollment. In addition to the primary study, 14 healthy participants (10M, 4F; age: 25.0 ± 4.4 years; height: 180.1 ± 9.1 cm; weight: 77.1 ± 13.4 kg) consented to additionally wearing one sensor (IMeasureU Blue Trident; Vicon Ltd., Oxford, UK), sampled at 1600 Hz, on each leg at the anterior-medial aspect of the proximal tibia and attached using semi-elastic straps (Figure 3.1b), while completing five different activities. The activities included three walking trials (i.e. walking in a straight line, walking with toe-out, and walking in a slalom pattern, all at a self-selected speeds), a static trial, and non-gait ambulation. The static trial included data where the participants were standing or sitting, with limited to no movement. The non-gait ambulation trial simulated slower movements (e.g.,

sweeping) that might be completed during activities of daily living but should not be considered classified as gait for gait analysis purposes. Following collection, the sensors were removed and downloaded using the CaptureU software (Version 1.3.1) and downsampled to 100 Hz. Data from left and right sensors were concatenated together into a single structure, with non-labeled data removed.

3) *Clinical Dataset*

The clinical dataset was created as a subset (n=32) from a larger study utilizing wearable sensors to longitudinally monitor gait in a cohort of older adults (12M, 20F; age: 65.7 ± 7.8 years; 168.0 ± 9.9 cm; 92.3 ± 23.9 kg) with OA awaiting knee (n=25) or hip arthroplasty (n=7). The Hamilton Integrated Research Ethics Board (HiREB 16236) approved this study, and all participants provided their informed consent prior to enrollment. Patients wore one sensor (Axivity AX6; UK), sampled at 100 Hz, on each leg at the anterior-medial aspect of the proximal tibia (Figure 3.1c), attached using waterproof medical grade tape (Simpatch Adhesive Patch). Sensors were worn for one week to capture free-living gait and activity data, as well as in-clinic performance-based functional tasks recorded with markerless motion capture (Theia3D, Kingston, ON). For the current study, only quiet standing and a 60-second self-selected walking speed tasks were manually labeled since not every patient completed all functional tasks and represents the minimum dataset across all patients. While markerless motion capture video data is not included in the current study, the data were used to calculate gait speed as described in Outerleys, et al. (Outerleys et al., 2024a). To improve robustness across gait speeds for the final model, patients' gait

speeds were classified as "slow" (<0.8 m/s), "average" (0.8-1.2 m/s), and "fast" (> 1.2 m/s) based on previously reported mean gait speeds for OA patients (Marcum et al., 2014; Wiik et al., 2017). Data from left and right sensors were concatenated together to form a single data structure.

### 3.2.2 Preprocessing

All subsequent data processing was completed in Python 3.9. Only the acceleration and gyroscope signals were utilized given these were available in all three datasets. To account for different sensor placements and orientations used, a calibration procedure was used to have consistent axes across all datasets as described by Mihy et al. (Mihy et al., 2022). For each subject in all datasets, data were first aligned with gravity in the vertical direction using a segment of quiet standing. Next, using the labeled data, a walking section in the data was used to rotate the sensor around the vertical direction to be fully aligned with the anteroposterior direction. To ensure all units were consistent across all datasets, units were converted as needed so the accelerometer units used were meters per second squared ($m/s^2$) and angular velocity was expressed in degrees per second (deg/s).

Given the goal of using this model for gait detection, the activity labels from each dataset were modified during preprocessing to be consistent across all datasets. The PAMAP2 dataset was most affected given that there were 18 total activities, but since not all activities were required (e.g. playing soccer, jumping rope) or differentiable using a single lower limb sensor (sitting vs standing), activities were reclassified as outlined in Supplemental Table I at the end of the chapter. The final relabeled activities for the

PAMAP2 dataset resulted in 7 total activities for classification: static, walking, running, cycling, stair ascent, stair descent, and other. Static may describe any condition where the subject is not moving, including sitting, standing, or laying down. The "other" classification was created to capture activities of daily living that do not represent gait behavior, such as vacuuming or house cleaning, that would not be completely static but also not true gait. These activities were selected to be representative of potential activities in the clinical population and utilized as needed for the other datasets for consistency. For the healthy dataset, the "static" label was used, while all three gait tests (i.e., straight walking, toe out walking, and slalom walking) were labeled as "walking". The cone task and sweeping were both classified as "other". For the clinical dataset, the quiet standing task was labeled as "static" and self-selected walking speed task was labeled as "walking". For all datasets, unlabeled data (including previously labeled activities that were unused) were removed. For PAMAP2 data, 500 samples from the beginning and end of each activity were removed to account for potential inaccuracies in starting and ending a data label, ensuring only steady state data was included from each activity (e.g., walking labels often start with static or transitional data). To account for differences in participants and sensor placement between datasets, data were scaled using the StandardScaler function (SciKit-Learn v1.6.1). Since the PAMAP2 dataset included running as an activity, the walking section from each participant was used to generate a consistent scaling transform for each participant across all datasets. The scaling transform was then applied to all data for that participant.

Following event relabeling, the sensor data was then segmented into sequential n-second sliding windows of length n·sampling frequency (termed window_sz), with a 50% overlap, with a corresponding activity label for that window. For static-labeled sections only, the window was evaluated to check that it was a true static window by ensuring the root mean squared acceleration was less than 1 $m/s^2$. While the original Li and Wang paper uses a one second window (window_sz = 100), a leave-one-subject-out analysis found minimal changes in accuracy up to five seconds (window_sz = 500). Given that the target population for this model are clinical patients who may walk slower, the window was extended to five seconds. A five second window also aligns well with a heuristic frequency-based model (Ullrich et al., 2020) used in the secondary analysis.

### 3.2.3 ResNet + BiLSTM Model Framework

The model used in this study was structured identically to the architecture outlined by Li and Wang (Li and Wang, 2022). In short, there are two blocks within the model: a ResNet and BiLSTM. The ResNet is used to extract spatial features from the signal, while the BiLSTM captures the forward and backward temporal information from time sequences (Figure 3.2). The following section will provide a high-level overview of model architecture.

The input to the model consisted of segmented, sequential IMU data, represented as (window_sz, n_channels,1), where window_sz represents number of samples in the current sequence, n_channels represents each individual IMU channel included (i.e. accelerometer and gyroscope axes). For the current study, window_sz = 500 indicating a

length of 500 samples. The n_channels parameter was set to 6 since all channels were used. To extract spatial features from the input data, residual block is introduced. The residual block is comprised of two convolutional layers and an additional convolutional shortcut connection added. Each convolutional layer is composed of 32 kernels of 2x2 size. The first convolutional layer has a stride length of 2 for the convolutional window. Following this layer, a batch normalization (batch norm) layer is used to accelerate training and re-center the data before passing to the next layer. Next, ReLU is used as the activation function before passing to the second convolutional layer. Following the second convolutional layer, which has a stride length of 1, another batch norm layer is used. At this point, the shortcut connection, featuring the third convolutional layer with stride length of 2, is added to mitigate vanishing gradients, enabling deeper networks. The residual block concludes with a second ReLU activation function before a dropout layer with 0.5 dropout is used. Before being passed to the BiLSTM layer, the resulting output to this point is passed to a flatten layer, which collapses the dimensions into a 1-dimensional array.

The output from the residual block was then passed to the BiLSTM. While the residual block extracted the spatial features and local patterns of the signals, the BiLSTM captured the temporal relationships of the time sequences from the output of the residual block. The BiLSTM used is a special type of recurrent neural network that analyzes the forward and backward temporal relationships of the signals. Within this BiLSTM, there are forward and backward layers because a standard Long Short-Term Memory network would only be based on previous data. The bidirectional nature of this layer allows for

better learning with respect to human activity recognition. Following the BiLSTM layer, there is an additional dropout layer with 0.5 dropout. A dense layer is to connect every neuron of the previous layer, which helps prevent overfitting. The output of the data is extracted through a softmax activation layer, which finally predicts the probability of each activity. The predicted probability can then be used to predict the most likely activity for a given window.

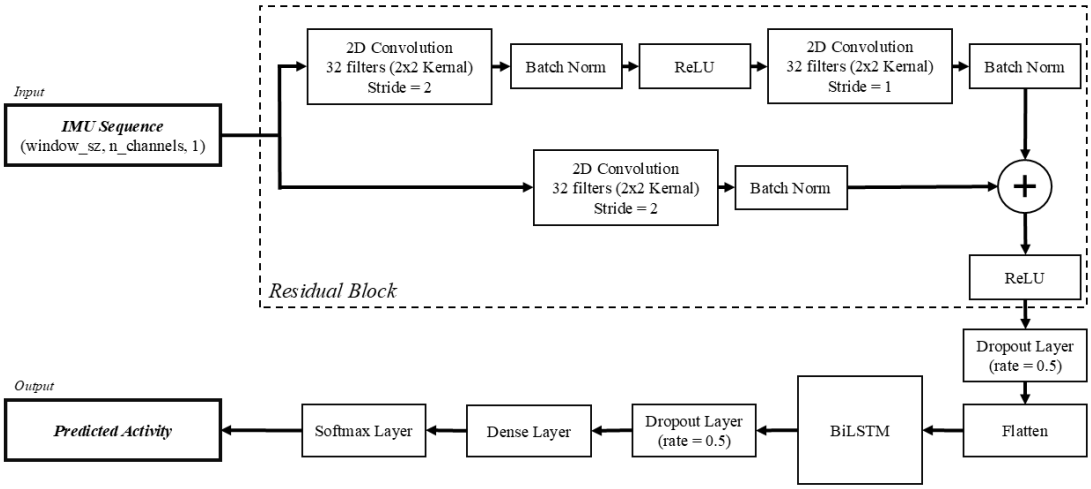



**Figure 3.2.** Model architecture of the ResNet + BiLSTM model, adapted from Li and Wang, 2022.

### 3.2.4 Model Training and Performance Analysis

The models were trained on a desktop computer equipped with a 3.5 GHz 16-Core CPU, with 32 GB of RAM, and a graphics processor (NVIDIA GeForce RTX 2070 Super). The algorithm was implemented using Python 3.9 using TensorFlow 2.10.1, using Spyder

as the integrated development environment on a 64-bit version of Windows 10. Three models were independently trained with progressive combinations of the PAMAP2 dataset, healthy dataset, and clinical dataset, starting with only the PAMAP2 dataset, then adding the healthy dataset, and finally adding the clinical dataset (Table 3.1). Participants in each dataset were randomly divided between training, validation and testing sets so at least 75% used for training. Validation and testing sets were randomly and evenly divided, except for the clinical dataset, where testing participants were specifically selected to ensure a range of gait speeds. The same hyperparameters in Li and Wang were used, except for training time. The models were trained by minimizing sparse categorical cross entropy using the Adam optimizer. A batch size of 64 was used to train each model. Whereas the Li and Wang used a training time of 80 to train each model, early stopping was used to prevent model overfitting, with a patience of 10. The best model based on minimized loss was retained from each model training and saved as the final model.

**Table 3.1:** Description of each model iteration in terms of training, validation, and testing data used in model development and evaluation, with the number of n subjects included from each dataset. Additional data was added for training and validation during model development, but the testing test was consistent across all model iterations.

| Model Iteration | Training Data (n) | Validation Data (n) | Testing Data (n) |
|---|---|---|---|
| PAMAP2 Only | PAMAP2 (6) | PAMAP2 (1) | PAMAP2 (1)<br>Healthy (2)<br>Clinical (6) |
| PAMAP2 +<br>Healthy | PAMAP2 (6)<br>Healthy (10) | PAMAP2 (1)<br>Healthy (2) | PAMAP2 (1)<br>Healthy (2)<br>Clinical (6) |
| PAMAP2 +<br>Healthy +<br>Clinical | PAMAP2 (6)<br>Healthy (10)<br>Clinical (26) | PAMAP2 (1)<br>Healthy (2)<br>Clinical (3) | PAMAP2 (1)<br>Healthy (2)<br>Clinical (6) |

Metrics were calculated for each final model to describe the overall performance across training, validation, and testing. For each classification, true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) were the following metrics using functions from the SciKit-Learn package in Python (SciKit-Learn v1.6.1). Specifically, performance from accuracy (1), precision (2), recall (3), and F1-score (4) were calculated during training, validation, and testing, resulting in training, validation, and testing performance metrics, using the following equations:

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \, (1)$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \, (2)$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP+FN}} (3)$$

$$F1 - score = \frac{2 * \mathrm{TP}}{2*\mathrm{TP+FP+FN}} (4)$$

Test performance metrics were calculated at the conclusion of training on the held-out test sets, comprised of randomly selected participants from each dataset as follows: one PAMAP2 participant, two healthy participants, and six clinical patients. Regardless of the included datasets used for each of the three trained models, the test set was comprised of all participants from all datasets to better characterize how generalizable the overall model architecture is through inclusion of additional training data.

Since this model was planned to be primarily used as a gait detection model, a secondary analysis to compare performance with a heuristic model was completed. In short, the frequency-based method uses harmonic frequencies to detect gait or non-gait by analyzing a 10 second sliding window on a given signal. Following a fast Fourier transform of the window, the harmonic frequency was found, and if peaks were detected in at least two of the first four harmonics, then it was deemed "gait". Conversely, if less than 2 peaks were found, then it was classified as "not gait". The output of this method results in a Boolean array (i.e., True or False) describing if gait is occurring. Because the original study used foot-mounted sensors, the algorithm was evaluated during pilot testing to determine the optimal signal to use as well as the optimal peak prominence (i.e., how much a peak stands out relative to the signal baseline). It was found that for sensors with the shank

attachment used in this study, the mediolateral gyroscope with a peak prominence of 5 best captured slower walking.

Following training and evaluation of all models (i.e., PAMAP2 only, PAMAP2 + Healthy, and PAMAP2 + Healthy + Clinical), the best-performing model (i.e. PAMAP2 + Healthy + Clinical) was tested for sensitivity in detecting walking, comparing it to a heuristic frequency-based model (Ullrich et al., 2020). To do this, the activity predictions from the ML model were converted into a Boolean array, where each five-second segment was labeled as either "walking" or "not walking," to match the output format of the frequency-based method. These predictions, with a 50% overlap between windows, were then mapped back to the original data length. A new array, initialized to zeros, was created. As the function processed the overlapping windows, ones were added to indicate walking periods. The proportion of windows predicting walking was calculated for each sample. Finally, this array was converted into a final Boolean output, where any value greater than 0.5 was considered a "True" walking window. The analysis was performed only on labeled sections of the test sets, generating performance metrics for both the heuristic and ML models.

## 3.3 Results

### 3.3.1 Dataset composition

Following preprocessing (i.e., removing unnecessary data and relabeling activities), the datasets provided varying amounts of data for training, validation, and testing. The PAMAP2 dataset provided the largest amount of data, with 50.3% of the data (1,466,715

total samples), followed by the healthy dataset with 26.1% (760,308 total samples), and then the clinical dataset with 23.6% (689,341 total samples). The breakdown of each activity in each dataset is provided in Table 3.2, both in terms of the individual datasets as well as the combined datasets used for augmenting model training.

**Table 3.2:** Activity breakdown by percentage for each dataset, both individually as well as total breakdown by dataset combinations (i.e., PAMAP2 + Healthy, PAMAP2 + Healthy + Clinical). Percentages are in terms of combined labeled activity data in training, validation, and test sets.

| Datasets | Static | Walking | Running | Cycling | Stair Ascent | Stair Descent | Other |
|---|---|---|---|---|---|---|---|
| PAMAP2 | 38.7% | 16.3% | 6.7% | 11.2% | 8.0% | 7.2% | 12.0% |
| Healthy | 20.3% | 53.6% | 0.0% | 0.0% | 0.0% | 0.0% | 26.1% |
| Clinical | 47.8% | 52.2% | 0.0% | 0.0% | 0.0% | 0.0% | 0% |
| PAMAP+ Healthy | 32.4% | 29.0% | 4.4% | 7.4% | 5.3% | 4.7% | 16.8% |
| PAMAP+ Healthy + Clinical | 36.0% | 34.5% | 3.4% | 5.6% | 4.0% | 3.6% | 12.8% |

### 3.3.2 Model Training, Validation, and Testing

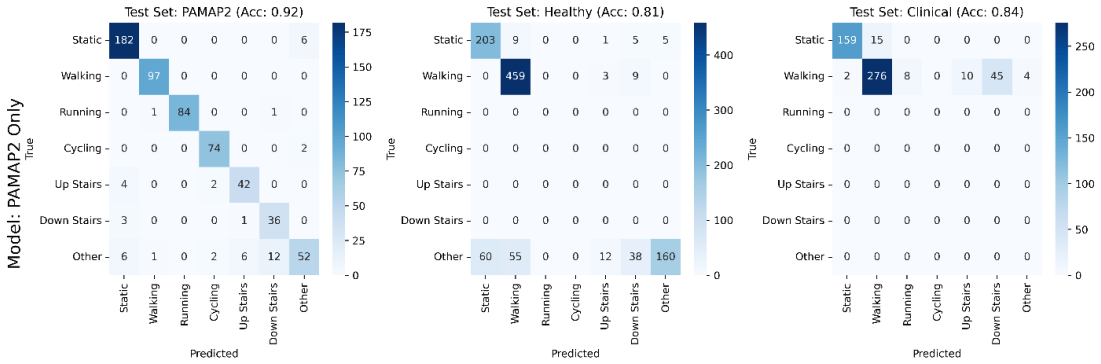
General model performance for each model across all datasets used in training, validation, and testing is detailed in Tabel 3.3. Model training took 46, 49, and 56 epochs before early stopping for PAMAP2 only, PAMAP2 + Healthy, and PAMAP2 + Healthy + Clinical training datasets, respectively. In general, model performance was relatively stable in terms of training, with performance metrics (i.e., accuracy, precision, recall, and F1-

score) all ranging from 0.96 to 0.98. There was a drop in model performance in the validation set, with performance metrics ranging from 0.91 to 0.93. Similar trends were seen in performance metrics when evaluating the test sets featuring held out data from each dataset. Overall model performance on the test sets for PAMAP2-only model achieved accuracy of 0.85, while the PAMAP2 + Healthy, and PAMAP2 + Healthy + Clinical models both achieved accuracy of 0.94

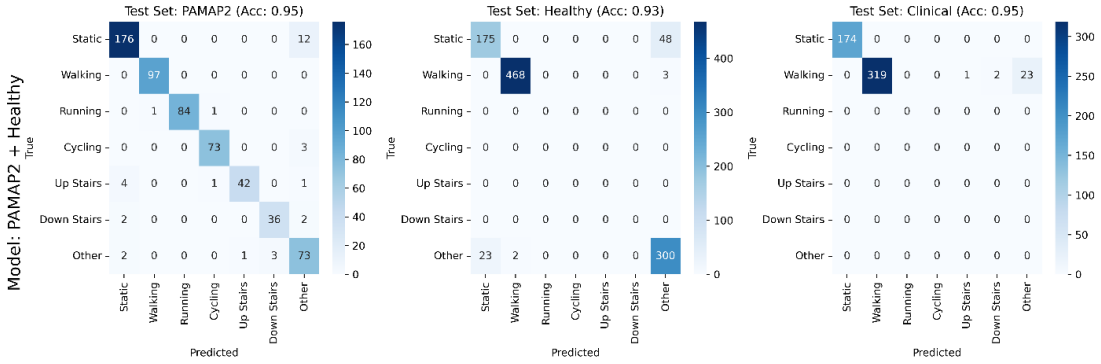
**Table 3.3:** Overall performance on training, validation, and test sets for each model.

| **Model Datasets** | **Performance Metric** | **PAMAP2** | **PAMAP2 + Healthy** | **PAMAP2 + Healthy + Clinical** |
|---|---|---|---|---|
| **Training** | Accuracy | 0.96 | 0.97 | 0.98 |
| | Precision | 0.96 | 0.97 | 0.98 |
| | Recall | 0.96 | 0.97 | 0.98 |
| | F1-score | 0.96 | 0.97 | 0.98 |
| **Validation** | Accuracy | 0.91 | 0.93 | 0.92 |
| | Precision | 0.93 | 0.93 | 0.92 |
| | Recall | 0.91 | 0.93 | 0.92 |
| | F1-score | 0.92 | 0.93 | 0.92 |
| **Testing** | Accuracy | 0.85 | 0.94 | 0.94 |
| | Precision | 0.89 | 0.94 | 0.94 |
| | Recall | 0.85 | 0.94 | 0.94 |
| | F1-score | 0.85 | 0.94 | 0.94 |

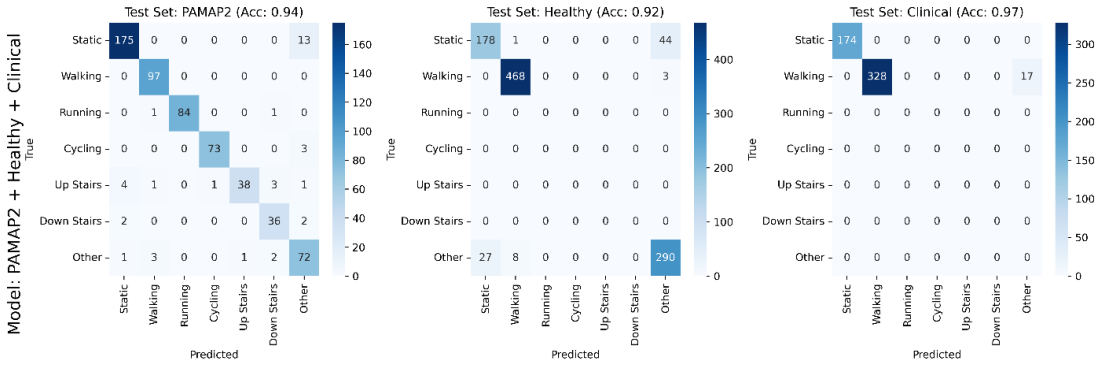
Further details of model performance with respect to performance on individual test sets from each dataset can be found in Figure 3.3a-c. The PAMAP2 test set performance improved slightly from an accuracy of 0.92 with just PAMAP2 training data, to 0.95 and 0.94 with models trained with additional healthy and clinical data, respectively. Gait-related misclassifications decreased with additional data from the healthy and clinical datasets. Healthy test set accuracy increased substantially from the PAMAP2-only trained model to the PAMAP2 + Healthy and PAMAP2 + Healthy + Clinical trained models, rising from 0.81 to 0.93 and 0.92, respectively. Clinical testing performance contained a similar trend, with accuracy increasing from the PAMAP2-only trained model to the PAMAP2 + Healthy trained model, from 0.84 to 0.95, before slightly increasing to 0.97 when clinical data was incorporated into the final PAMAP2 + Healthy + Clinical trained model.




(a)




(b)




(c)

**Figure 3.3.** Confusion matrices broken out individually for PAMAP2, healthy, and clinical testing sets, with accuracy for each testing set shown for each model. Model performance generally improved on test sets as additional data was incorporated into training.

### 3.3.3 Model Comparison to Heuristic Gait Detection

As the combined PAMAP2, healthy, and clinical trained model demonstrated the best performance, this model was then used to compare against the heuristic frequency-based model. The median performance on each test set is shown in Table 3.4. With the exception of the PAMAP2 testing participant, the frequency-based model performed on par with the ML-based model. The PAMAP2 testing only achieved an accuracy of 0.80, compared to the ML model accuracy of 0.97. The ML-based model (accuracy range: 0.97 to 0.98) performed slightly better than the frequency-based model (accuracy range: 0.94 to 0.95) on healthy and clinical datasets. Overall, there were surprisingly no other notable differences.

**Table 3.4:** Median performances on individual test sets from each data, based on the top performing model containing PAMAP2 + Healthy + Clinical datasets when used for gait detection and the heuristic frequency-based method.

| Dataset (n) | Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| PAMAP2 (1) | Frequency | 0.80 | 0.41 | 0.97 | 0.58 |
| | ML | 0.98 | 0.91 | 0.98 | 0.94 |
| Healthy (4) | Frequency | 0.95 | 0.97 | 0.95 | 0.96 |
| | ML | 0.99 | 1.00 | 0.99 | 1.00 |
| Clinic (6) | Frequency | 0.94 | 0.99 | 0.94 | 0.96 |
| | ML | 0.97 | 1.00 | 0.98 | 0.97 |

### 3.3.4 Effect of Gait Speed

Further performance comparisons between the trained ML model and the heuristic frequency-based method were evaluated individually on the clinical test set (Table 3.5). There were two patients for each gait speed previously defined gait speed range of slow (0.30 and 0.75 m/s), average (1.00 and 0.99 m/s), and fast (1.32 and 1.42 m/s). The machine model performed much better than the heuristic frequency on the patients with slower gait speeds, with the frequency-based method only having accuracies of 0.53 and 0.81, compared to the ML-based model with 0.96 and 0.93. For normal and fast gait speed patients, the overall performance of the models on the healthy and clinical datasets, the accuracies were very similar between the frequency-based method (accuracy range: 0.94-0.98) and ML model (accuracy range: 0.97-0.99).

**Table 3.5:** Method performance on participants in clinic test set by speed. Individual performance metrics for each patient that are below 0.90 are bold and italicized.

| Patient | Gait Speed (m/s) | Speed Type | Method | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|---|---|---|
| 1 | 0.30 | Slow | Frequency | ***0.53*** | ***0.87*** | ***0.38*** | ***0.53*** |
| | | | ML | 0.96 | 0.99 | 0.96 | 0.97 |
| 2 | 0.75 | Slow | Frequency | ***0.81*** | 1.00 | **0.72** | ***0.84*** |
| | | | ML | 0.93 | 1.00 | ***0.89*** | 0.94 |
| 3 | 1.00 | Average | Frequency | 0.96 | 0.99 | 0.94 | 0.97 |
| | | | ML | 0.97 | 0.96 | 0.99 | 0.98 |
| 4 | 0.99 | Average | Frequency | 0.95 | 0.99 | 0.94 | 0.97 |
| | | | ML | 0.98 | 0.99 | 0.98 | 0.99 |
| 5 | 1.32 | Fast | Frequency | 0.94 | 0.95 | 0.94 | 0.95 |
| | | | ML | 0.97 | 0.97 | 0.98 | 0.98 |
| 6 | 1.42 | Fast | Frequency | 0.98 | 1.00 | 0.96 | 0.98 |
| | | | ML | 0.99 | 0.98 | 0.99 | 0.99 |

### 3.4 Discussion

The purpose of this study was to (a) enhance the training of a previously developed deep learning architecture, which combines a ResNet and BiLSTM, by incorporating additional healthy and clinical participants into model training; (b) compare the sensitivity of this model to a heuristic frequency-based model; and (c) evaluate its performance across lower gait speeds. The results demonstrated that the base model, trained only on the PAMAP2 dataset, performed surprisingly well on clinical data, achieving 91% accuracy (Figure 3.3a). However, augmenting the model with additional clinical training data improved its performance, reaching 97% accuracy (Figure 3.3c). Additionally, the deep learning model generally outperformed the heuristic frequency-based method, with the greatest improvement observed at slower walking speeds in patients with end-stage OA. Although further validation is needed, particularly in free-living environments, this study presents a promising framework for enhancing gait detection models for healthy and pathological populations using wearable sensors.

In terms of model evaluation using the original architecture proposed by Li and Wang, the overall model performance of the current study is comparable to the original study. Li and Wang reported a validation accuracy of 97% on the PAMAP2 dataset, as well as 97% on their own human activity dataset. In comparison, the validation accuracy in this study was lower, particularly for the PAMAP2-only model (91%) and higher but still not reaching Li and Wang's accuracy in the PAMAP2 + Healthy trained model (0.93) or the PAMAP2 + Healthy + clinical trained model (92%). Several methodological factors may explain these differences. Li and Wang's lab-created dataset, which used a shank sensor

placement and consisted of five young adults, split the data into 70% training and 30% validation, but did not appear to separate the data by subject. This could have led to memory leakage between training and validation sets, potentially inflating their model's performance. As noted in the review by Haliaj, a common pitfall in ML is not fully separating training, validation, and testing datasets (Halilaj et al., 2018). In contrast, the current study ensured that subjects were only included in one of the datasets: training, validation, or testing. Additionally, while both studies used similar activities for training and validation on the PAMAP2 dataset, this study focused on activities more likely performed by older adults, such as walking, cycling, and stairs, while excluding those less likely (e.g., rope jumping, Nordic walking). Finally, the number of sensors used differed between studies. Li and Wang used the full set of sensors in the PAMAP2 dataset, while this study used a single sensor placed below the knee, limiting the model's ability to distinguish between activities like sitting, standing, and lying, which were grouped as "static". These differences in model training, data preprocessing, and sensor placement likely account for the variations in model performance between the two studies.

The current study aimed to further characterize the model architecture by augmenting the PAMAP2 dataset with additional healthy and clinical datasets and evaluating the model on held-out test sets. Li and Wang trained and validated their models using only their lab-created dataset and two open datasets. In contrast, this current study used the PAMAP2 dataset as a base and augmented the training and validation data with additional static and walking data from both healthy and clinical populations, with a slightly different sensor placement. The additional data slightly improved the model's

performance on the PAMAP2 test set, which achieved accuracy of 92%, 95%, and 94% across the three trained models. With the base PAMAP2-only trained model, most misclassifications in the clinical dataset involved walking data being classified as “stairs” or “other.” This could be due to the slower or reduced magnitude IMU data in the clinical group, which may resemble stair walking or shuffling gait patterns observed in the healthy PAMAP2 dataset. However, when both healthy and clinical data were incorporated into model training, the model performance on the clinical test set improved accuracy to 97%, with only 5% of gait windows misclassified as “other”, a significant improvement over the PAMAP2-only model, which misclassified 20% of walking data.

One of the main objectives of this study was to develop a ML model capable of accurately identifying gait bouts in a clinical OA population using only one lower-limb sensor, with performance surpassing traditional heuristic models. On the surface, the current deep learning, augmented with clinical data, appears only slightly better than the heuristic frequency-based method (94% vs. 97%). However, when analyzed individually by gait speed, the deep learning model shows no performance drop, while the heuristic model performs poorly on slower gait speeds. Recall for the two patients with slower gait speeds was 0.38 and 0.72 for the frequency-based model, compared to 0.96 and 0.89 for the deep learning model. This ability to identify slower walking speeds is critical for both clinical and free-living applications. For instance, healthy older adults typically walk at speeds between 1.1 and 1.2 m/s (Andrews et al., 2023), while patients with OA often have slower speeds, around 1.0 m/s (Marcum et al., 2014; Wiik et al., 2017) or slower, as seen in this study. Additionally, walking speeds in both healthy and clinical populations can be

further reduced in free-living conditions outside of controlled testing environment (Foucher et al., 2010; Takayanagi et al., 2019). Therefore, it is important for a model to accurately detect walking bouts at speeds below 1.0 m/s for real-world applications. The current model's ability to perform well, even with a patient walking at an extremely slow speed of 0.3 m/s, demonstrates its robustness in identifying a wide range of gait types. However, future studies should evaluate its performance in identifying clinical OA gait in free-living environments.

The study aligns well within the broader literature. While there is not an exact one-to-one comparison, several studies have attempted to characterize slower gait in older adults, both with and without deep learning. A similar study incorporating gait data from individuals with Parkinson's disease, stroke, multiple sclerosis, and chronic low back pain used a temporal convolutional neural network to identify gait events, validated against optical motion capture, at three different speeds (Romijnders et al., 2022). This model performed well across both ankle and shank sensor locations, with high recall (>95%), precision (>98%), and F1-score (94%) for both initial and terminal contact events. Additionally, the previously mentioned study by Roth, et al., which used a Hidden Markov Model to identify gait in patients with Parkinson's disease achieved high performance metrics (F1-score = 92.1%), especially for longer bouts (F1-score = 96.2%). The current study appears to largely be in line with the performance of these models in different populations, but extends this work to a broader focus on osteoarthritis populations where such work is lacking (Kokkotis et al., 2020). Overall, within the current study the model trained with PAMAP2 + Healthy + clinical data achieved high performance metrics on an

osteoarthritis population, which will enable future gait analyses outside of controlled laboratory environments.

This study has several notable limitations. Ideally, a dataset of older adults, both with and without pathological gait, would include a wide range of activities for model training. However, in clinical populations with gait impairment, this is not often feasible. Patients who can complete a broad range of activities are typically higher functioning and exhibit less pathological gait. As a result, using only static and walking activities for evaluation may overstate the model's performance. While the high performance on the held-out PAMAP2 test set suggests that satisfactory results across multiple activities are achievable, other activity classifications in clinical populations remain unvalidated and require further research. That being said, it is important to clarify that the current study was primarily focused on developing a model to specifically identify walking, which allowed for greater emphasis on accurate walking labels. Moreover, one of the limitations surrounding many other models is that they are built on a single data collection source (Halilaj et al., 2018), whereas the current study demonstrated excellent results in identifying walking bouts across different datasets with different sensors, placements, and populations, highlighting the robustness of the model. Additionally, the study did not validate the models on free living data. Given the lack of a gold standard for activity classification in free-living environments, validation would have been challenging. Nonetheless, as noted previously, future studies will seek to assess the performance of these models in free-living data settings.

In summary, this study provides additional evidence that the ResNet+BiLSTM

model is a valid approach to activity classification using IMU data. It also demonstrates the effectiveness of using open datasets, like PAMAP2, as a base that can be augmented with data from different sensor locations or populations. The ability to incorporate larger training datasets significantly improves model performance, as shown in this study. Furthermore, the deep learning model outperformed a heuristic frequency-based method when trained with data from a slower, pathological population. This is particularly relevant for future studies on free-living gait, as both healthy and clinical populations tend to walk slower and will require more robust gait identification.

### 3.5 Bridge Section – Comparison of Free-Living Bout Identification from Frequency Method and ML Method

Although the present study concentrates on identifying walking bouts from controlled laboratory settings, deploying the resulting model in free-living environments is the ultimate goal. In free-living data, where no ground-truth labels exist, comparisons can only be qualitative in nature. To obtain an initial sense of relative performance, both the frequency-based and ML algorithms were applied to the same participants and compared the number and length of bouts and the resulting number of strides from those bouts (Table 3.6).

Overall, the ML method detected fewer, more continuous bouts with more strides compared to the frequency method. On average, median bout length was 141.3 s with the ML method versus 16.1 s with the frequency method. The maximum bout length averaged

3471.1 s for ML versus 355.3 s for the frequency method. These results suggest that the frequency method tends to fragment extended periods of walking into multiple short bouts, whereas the ML method is capable of detecting them naturally as longer bouts. Which behaviour is preferable depends on the research question, but in free-living studies, capturing longer continuous bouts is often advantageous, because any stray non-walking strides can be excluded later during outlier detection.

For the participants in Study 3, the ML method therefore yielded a more complete picture of free-living gait by identifying more strides across longer bouts. Accordingly, Chapters 4 and 5 use gait parameters derived from the ML method only.

**Table 3.6:** Performance of the frequency-based and machine-learning bout-detection methods in free-living data.

| Patient | Frequency | | | | ML | | | |
|---|---|---|---|---|---|---|---|---|
| | **Bouts (n)** | **Median Length (s)** | **Max Length (s)** | **Total Strides (n)** | **Bouts (n)** | **Median Length (s)** | **Max Length (s)** | **Total Strides (n)** |
| 1 | 1092 | 15.5 | 141.8 | 2095 | 777 | 56.1 | 1463.5 | 4789 |
| 2 | 1180 | 16.4 | 158.4 | 7385 | 750 | 57.4 | 1235.3 | 9397 |
| 3* | 153 | 15.1 | 229.6 | 1518 | 75 | 106 | 1015 | 1981 |
| 4 | 2318 | 18.6 | 307.5 | 20939 | 220 | 540.3 | 14376.6 | 28773 |
| 5 | 1436 | 15.3 | 319.5 | 13611 | 1007 | 50.2 | 758.1 | 17621 |
| 6 | 1598 | 15.9 | 974.9 | 14981 | 1187 | 38 | 1978 | 18414 |
| Mean | 1296.2 | 16.1 | 355.3 | 10088.2 | 669.3 | 141.3 | 3471.1 | 13495.8 |

* Patient removed sensors after 1 day.

## 3.6 Supplemental Material

### 3.6.1 Healthy Lab Dataset

As part of a concurrently collected study, 14 healthy subjects (10M, 4F; age: 25.0 ± 4.4 years; height: 180.1 ± 9.1 cm; weight: 77.1 ± 13.4 kg) had gait recorded during three walking trials at a self-selected speed. The primary study collected gait data from another motion capture system, which were not analyzed for the current study. Participants wore one sensor (IMeasureU Blue Trident; Vicon Ltd., Oxford, UK), sampled at 1600 Hz, on each leg at the anterior-medial aspect of the proximal tibia and attached using semi-elastic straps. Following set up and calibration, the following gait tasks were collected on an indoor walking track:

1) 4 x 10 m straight line walk.

2) 2 x 10 m "toe-out" straight line walk.

3) 2 x 10 m slalom walk, where the subjects navigated around cones in a serpentine pattern.

Following the completion of the walking trials, subjects stood or sat down while they waited to begin the final sensor collection and kept the IMeasureU sensors on the shanks on and recording during this process to collect static and semi-static data, approximately for 2 minutes. Afterwards, subjects completed common activities of daily living that are ambulation but not typical walking, to simulate slower gait that might be completed during activities of daily living but should not be considered classified as gait

for gait analysis purposes. These trials consisted of:

1) Picking up 12 cones arranged in a 3x4 grid, spaced approximately one meter apart, in any order they wished.

2) Broom sweeping the area around where the grid had been arranged for 1 minute.

Sensors were removed and downloaded using the CaptureU software (1.3.1) and downsampled to 100 Hz. Data from each sensor were manually labeled using a custom MATLAB script (The Mathworks, Natick, MA), identifying each active section as "walking", "static" or "other", while non-classified data were removed. Data from left and right sensors were concatenated together into a single structure.

### 3.6.2 Clinical Lab Dataset

A subset of data from a larger study utilizing wearable sensors to longitudinally monitor gait in a cohort of older adults (12 M, 20 F; age: 65.7 ± 7.8 years) with OA awaiting knee (n=25) or hip arthroplasty (n=7) were included. With respect to gait, OA patients typically walk slower with an atypical gait compared to healthy control subjects, although level of function can vary greatly with pain [13]. Algorithms, both ML and heuristic, may not be able to accurately classify gait cycles due, especially when patients have a slower gait speed. Wearable sensor data from the patients' preoperative visit, approximately two weeks before a scheduled surgery, was used. Patients were asked to wear one sensor (Axivity AX6; Vicon Ltd., Oxford, UK), sampled at 100 Hz, on each leg at the anterior-medial aspect of the proximal tibia and attached using waterproof medical grade tape (Simpatch Adhesive Patch) to be worn for one week to capture free-living gait and activity

data. Following sensor placement, patients were asked to complete a battery of functional tasks recorded with 10 cameras (Sony RX0-II, 60 Hz, Sony Corporation) for post-processing with markerless motion capture software (Theia3D, Kingston, ON):

1) 30 seconds of quiet standing
2) 60 seconds of self-selected pace walking
3) 30 seconds of faster walking,
4) Five repetition sit-to-stand
5) Two steps up and down.

Following the one-week period, patients mailed the sensors back using provided envelopes. Data were downloaded using the OmGui software (1.0.0.43) and manually labeled using a custom Python script in a similar way as described above for the healthy dataset. For the current study, only quiet standing and 60-second self-selected walking tasks were utilized since not every patient completed all five functional tasks and represents the minimum dataset across all patients. Additionally, to the interest in identifying slower gait, faster walking was not included. The step up and down task, while similar to stair ascent and decent, was ultimately not long enough of a trial to be included for potential classification. Data from left and right sensors were concatenated together to form a single data structure.

Markerless motion capture video data was processed in Theia3D and then processed into gait cycles using Visual 3D. Heel strike events calculate spatiotemporal

variables, including gait speed, as described in Outerleys, et al. To ensure robustness across gait speeds for the final model, patients' gait speeds were classified as "slow" (<0.8 m/s), "average" (1.0 m/s), and "fast" (> 1.2 m/s) based on previous reported mean gait speeds for OA patients.

Data labels were updated across all datasets to be consistent during preprocessing. The full original and updated labels for each are displayed in Supplemental Table 1.

**Supplemental Table 1:** Original labels and updated labels for all datasets. Updated labels were used to reflect the same activities in different datasets.

| Dataset | Original Label | Updated Label |
|---|---|---|
| PAMAP2 | 1 – Lying<br>2 – Sitting<br>3 – Standing | 0 – Static |
| | 4 – Walking | 1 – Walking |
| | 5 – Running | 2 – Running |
| | 6 – Cycling | 3 – Cycling |
| | 12 – Stair ascent | 4 – Stair ascent |
| | 13 – Stair descent | 5 – Stair descent |
| | 16 – Vacuum cleaning<br>19 – House cleaning | 6 – Other |
| | 7 – Nordic walking<br>17 – Ironing<br>24 – Rope jumping | Removed |
| Healthy | 1 – Straight walking<br>2 – Toe-out walking<br>3 – Serpentine walking | 1 – Walking |
| | 4 – Static (sitting and/or standing) | 0 – Static |
| | 5 – Cone Task<br>6 - Sweeping | 6 – Other |
| Clinical | 1 – Quiet Standing | 0 - Static |
| | 2 – Self-selected walking (60 seconds) | 1 – Walking |
| | 3 – Fast walking (30 seconds)<br>4 – 5 repetition sit-to-stand.<br>5 – Stairs | Removed |

# Chapter 4: Comparing Gait Metrics from In-Lab Gait Analyses to Free-Living Assessment from Wearable Sensors in End-Stage Osteoarthritis Patients (Study 2)

## Preamble

Chapter 4, my second study, examines gait data collected simultaneously with in-lab motion capture, wearable sensors, and extended free-living recordings. Although many studies have shown strong agreement between motion capture and wearable sensor measurements, few have explored how these metrics compare with longer, real-world gait samples that may better reflect everyday movement. Here, I analyze several key variables common to both systems. The in-lab results replicate previous findings, while the free-living analyses extend them by showing how the same metrics behave outside the laboratory. Also, while not directly evaluated, this study also utilized the gait detection model developed in Study 1. Together, these findings support the use of wearable sensor-derived gait measures in free-living assessments and set the stage for Chapter 5, which evaluates them longitudinally.

This work was prepared for submission to *Journal of Biomechanics*, with the following coauthors:

Matthew C. Ruder, Vincenzo E. Di Bacco, Kim Madden, Anthony Adili, and Dylan Kobsar

**Abstract**

Gait analysis provides objective metrics to evaluate mobility in populations such as individuals with knee osteoarthritis (OA). However, in-lab assessments may not reflect real-world gait. Wearable inertial sensors offer a promising alternative, but few studies have directly compared concurrent gait measures from motion capture and wearable sensor-based gait analysis with free-living data collections. This study collected gait data preoperatively from 45 older adults with end-stage knee OA using both in-lab markerless motion capture and wearable sensors, as well as up to seven days of free-living wearable sensor recordings. Stride time and stance time demonstrated excellent agreement and reliability between motion capture and in-lab wearable data. Reliability between in-lab and free-living measures was lower, though peak angular velocity retained moderate-to-good reliability. These findings suggest that certain spatiotemporal gait metrics, particularly peak angular velocity, may be viable for use in extended free-living assessments. The study supports the integration of wearable sensors into long-term gait monitoring for clinical populations.

## 4.1 Introduction

Osteoarthritis (OA) is a common and debilitating joint disease characterized by cartilage degradation and alterations to in bone morphology, often leading to increased pain, reduced mobility, and diminished quality of life (Andriacchi and Favre, 2014). OA affects millions of adults worldwide, most commonly involving the knee joint, and has

risen substantially in prevalence (Li et al., 2024). Gait analysis using motion capture enables highly accurate and objective quantification of various gait parameters (e.g. gait speed, stride length, knee flexion, frontal plane alignment), providing valuable insights into functional status and disease progression (Birmingham et al., 2017; Duffell et al., 2017; Ornetti et al., 2010) .

Laboratory-based gait analysis, using marked or markerless motion capture, remains the gold standard for high-fidelity gait measurement (Ornetti et al., 2010). However, despite the importance of assessing gait health in populations such as those with OA, gait analysis is not commonly used in clinical care. These systems are costly, both monetarily and in terms of physical space, requiring specialized camera equipment and dedicated facilities. Moreover, gait data collected in laboratory settings may not reflect real-world gait patterns, particularly in individuals with knee OA, where functional limitations and fluctuations in pain may lead to a disconnect between laboratory-based assessments and everyday walking (Allen et al., 2009; Asay et al., 2013; Hillel et al., 2019).

Wearable sensors, such as inertial measurement units (IMUs), offer a cost-effective way to continuously collect data in real-world settings. Within laboratory environments, gait metrics from IMUs have been shown to closely align with those obtained from motion capture across a range of populations, including individuals with gait impairments (Prisco et al., 2024). In knee OA specifically, spatiotemporal (e.g., stride time (Hafer et al., 2020)) and kinematic variables (e.g., sagittal knee range of motion (Hafer et al., 2023; Seel et al., 2014)), consistently demonstrate high reliability and strong agreement with motion capture data. Nonetheless, it remains uncertain how well these validated metrics capture the

complexity and variability of gait that occurs during daily life outside controlled laboratory conditions.

Although IMUs offer the potential to capture real-world walking patterns, relatively few studies have directly compared gait metrics collected in laboratory settings to those measured in free-living environments. One notable study by Hillel, et al., found that gait parameters collected from older adults in the lab differed significantly from those collected in free-living environments (Hillel et al., 2019). The authors concluded that in-lab measurement may not accurately reflect everyday walking. The findings are supported by studies involving other clinical population, such as individuals with COPD (Buekers et al., 2023), Parkinson's disease (Del Din et al., 2016), and multiple sclerosis (Shema-Shiratzky et al., 2020), where gait assessed in free-living conditions typically shows slower gait speeds and longer strides times compared to the controlled laboratory environment. Unfortunately, these studies often lack concurrent comparisons to conventional motion capture systems for the in-lab assessments and have yet to directly examine these differences in individuals with knee OA.

Therefore, the purpose of this study was to quantify agreement between in-lab motion capture and IMU-derived gait variables, as well as free-living gait variables, in adults with knee OA. Specifically, biomechanical variables derived from wearable IMUs during in-lab walking were compared to those collected concurrently from a markerless motion capture system, and to a week of free-living IMU data. In line with previous studies, we hypothesized that the two in-lab systems would show strong agreement, whereas the free-living data would diverge from in-lab IMU data in a manner consistent with findings

from other populations (e.g., longer stride times). A secondary purpose of this work was to examine whether IMU-derived gait variables, obtained in-lab, during free-living, or as a difference metric between the two settings, were significantly correlated with patient-reported outcome measures for knee OA. We hypothesized that free-living gait variables would show stronger correlations with patient-reported outcomes.

## 4.2 Materials and Methods

### 4.2.1. Participants

This work is part of a larger study cohort study following individuals with knee OA as they undergo a surgical arthroplasty procedure (i.e., joint replacement surgery). The Hamilton Integrated Research Ethics Board (HiREB 16236) approved this study, and all participants provided their informed consent prior to enrollment. For this sub-study, participants also consented to wearing two small wearable sensors on their knees for up to seven days. Only data collected prior approximately two weeks prior to their scheduled arthroplasty procedure were examined. Exclusion criteria included any significant disability that prevented ambulation or the inability to provide informed consent. To be included in the current analysis, participants were required to wear both sensors for at least four days.

### 4.2.2 Data Collection Protocol

Following consent, participants completed surveys assessing patient-reported outcome measures related to pain, function, and depression. Pain was assessed with a visual analog scale, asking participants “Rate your average pain in your knee in the last week.” Function was evaluated using the Oxford Knee Score (OKS, Dawson et al., 1998)), which

yields a score from 0 (worst outcome) to 48 (best outcome). Depressive symptom severity was measured using the Patient Health Questionnaire (PHQ-8, Kroenke et al., 2009)), which produces a score from 0 to 24. A score of 10 or more is considered major depression, while 20 or more indicates severe major depression.

Sensor placement began following the completion of the surveys. Two IMU sensors (Axivity AX6, 100 Hz, Axivity Ltd., Newcastle, UK) were affixed just below each knee, medial and inferior to each tibial tuberosity, using adhesive tape (Simpatch) by a researcher (Figure 4.1). The sensors were configured to record continuously for seven days. Participants were instructed to maintain their normal daily routine to remove them only at the end of the seven-day period, unless removal was required earlier due to discomfort. After removal, participants returned the sensors to the research team using a pre-addressed mailing envelope.

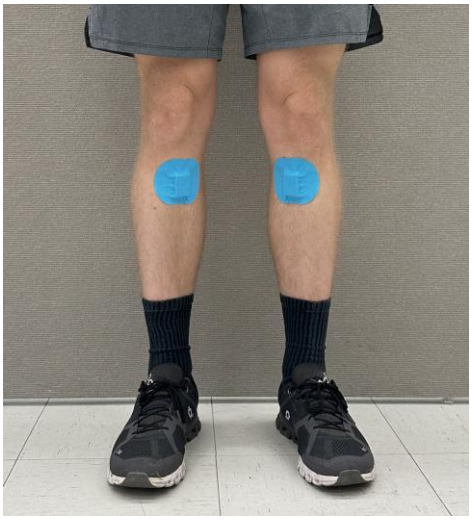


**Figure 4.1.** Sensor locations on shank, located medial and inferior to the tibial tuberosity.

Following the placement of the sensors, participants completed a series of functional tasks for movement analysis using markerless motion capture. Ten cameras (Sony RX0-II, 60 Hz, Sony Corporation) were used to record participants as they performed movements representative of activities of daily living. Specifically, the recorded tasks included: 30 seconds of quiet standing, 60 seconds of walking at a self-selected pace, 30 seconds of faster walking, a five-repetition sit-to-stand test, and a two-step ascent and descent task on moveable stairs. The walking tasks were done along an outlined 7.6 m walkway. Due to functional variability among participants, some did not attempt or complete all movement tasks. However, all participants completed at least quiet standing and 60 seconds of walking at a self-selected pace, with the latter serving as the primary data source for this study. Following the in-lab movement analysis, participants left the lab with the sensors remaining in place to capture one week of free-living data.

### 4.2.3. Data Analysis

#### 4.2.3.1 In-lab Motion Capture

In-lab motion capture data from each participant were processed into kinematic variables using Theia3D (V2023.1.0.3161, Theia Markerless, Kingston, ON). Kinematic data were then analyzed in Visual3D (HAS-Motion, Kingston, ON) to segment walking data into individual strides (Outerleys et al., 2024b). Spatiotemporal and kinematic variables were calculated for each stride and for each limb. Spatiotemporal variables of interest included stride time, stance time, and swing time. Shank angular orientation was

differentiated to calculate shank angular velocity, which served as an additional comparator to IMU-derived variables.

#### 4.2.3.2 Wearable Sensor

Upon receipt of the wearable IMUs, data were downloaded using the OMGui program (V1.0.0.43). All subsequent processing was conducted in Python 3.9 using Spyder as the integrated development environment. A custom Python script was used to align sensor data, identify walking bouts, segment the signals into individual strides, and calculate the gait variables.

First, using an alignment procedure described by Mihy et al., sensor data were aligned to account for slight misalignments and standardize to ensure consistent axes across participants (Mihy et al., 2022). The in-lab portion of data was manually identified to isolate the quiet standing and self-selected walking sections. The quiet standing trial was used to align the vertical axis of the sensor to the gravity vector. A segment of the self-selected walking trial was then used to virtually rotate the sensor data about the mediolateral axis, aligning the local coordinate system to sagittal plane of the shank during walking. Accelerometer data were converted from gravitational units (*g*'s) to meters per second squared ($m/s^2$), while angular velocity was expressed in degrees per second (deg/s). The in-lab sections from each sensor were exported to separate files for further analysis.

After extracting the in-lab subset from IMU data, walking bouts were identified in the full dataset for each sensor. A previously developed machine learning model architecture, trained using data from adults with knee OA, was used to generate the activity classification model required to identify walking bouts. Sensor data were segmented into

five-second windows with 50% overlap and scaled to the walking section using StandardScaler from Scikit-Learn (SciKit-Learn v1.6.1). The model classified each five-second window into one of several activity categories: “static” (including sitting, standing, and lying down), “walking”, “running”, “cycling”, “stair ascent”, “stair descent”, and “other” (e.g., other dynamic activities such as cleaning that would like not be correctly classified otherwise). This output was then converted to a binary array, with ones representing “walking” and zeros as “not walking.” Within each walking bout, gait events were detected using previously defined methods (Mariani et al., 2013), and stride-level variables were calculated using the same approach as the motion capture data. These variables included stride time, stance time, swing time, and peak angular velocity during free-living gait.

To minimize the inclusion of erroneously detected strides, a two-step outlier detection procedure was implemented. First, a principal component analysis was conducted on all in-lab, time-normalized (0-100%) mediolateral angular velocity data. The resulting principal component analysis model was then applied to all initially detected strides, and any stride with a z-score exceeding 2.5 on any principal component was flagged as an outlier. Next, a stride similarity check was performed by computing the correlation between each remaining stride and the average of all in-lab strides (Ferrari et al., 2010). Strides with a correlation coefficient below 0.5 were additionally considered outliers and excluded from further analysis. Together, these steps provided an approach based on intentionally broad thresholds, aimed at removing only strides with clearly atypical patterns that were not representative of a gait cycle, thereby limiting the influence of false walking data on the

resulting discrete variables. In-lab walking data was manually identified for each participant and labeled.

### 4.2.4 Statistical Analysis

To address the primary purpose of the study, each variable was compared between motion capture and in-lab IMU data, as well as in-lab IMU data and free-living IMU data. For each participant, the median and standard deviation were calculated for each variable. Agreement between motion capture and in-lab IMU data was assessed using Pearson correlation coefficients and Bland-Altman plots with 95% limits of agreement. Additionally, agreement was further evaluated with a two-way random effects intraclass coefficient (i.e., ICC(2,1)) for both motion capture vs. in-lab IMU and in-lab IMU vs. free-living IMU data comparisons. ICC values were interpreted as follows: <0.50 = poor agreement , 0.50-0.75 = moderate agreement , 0.75-0.90 = good agreement , and >0.90 indicate excellent agreement (Koo and Li, 2016). Similarly, Cohen's *d* was calculated to quantify the effect size between each variable for both motion capture vs. in-lab IMU and in-lab IMU vs. free-living IMU data.

For the secondary purpose, examining the relationship between gait measures and patient-reported outcomes, Pearson correlation coefficients were calculated. Specifically, correlations between each gait variable (stride time, stance time, swing time, peak angular velocity) and each patient-reported outcome measures (pain, PHQ-8, OKS). Additionally, correlations were computed for the differences between in-lab and free-living gait measures to determine whether discrepancies between settings were related to the patient-reported outcome measures. Finally, although gait speed was not estimated from the IMUs

in this study, its the relationship to clinical outcomes was explored by correlating motion capture–derived gait speed with each patient-reported measure, given its prevalence in the literature.

### 4.3 Results

Participant characteristics are displayed in Table 4.1. A total of 45 participants were included in the current analysis, consisting of 14 males and 31 females. Overall, the sample was representative of a typical knee OA population, with moderate levels of pain and reduced function.

The total number of strides available for gait analysis varied by measurement technique. Data were not synchronized between systems, as this was outside the scope of the current study, resulting in a greater number of strides captured by the IMU (73 ± 13) compared to motion capture (30 ± 5) during the in-lab assessment. This difference was primarily due to limited capture volume of cameras along the walkway in the gait lab. Additionally, with up to seven days collected out-of-lab, free-living IMU data yielded a large number of strides available for analysis (24,541 ± 12,360).

**Table 4.1.** Descriptive statistics for the sample.

| Statistical Value | Age (yrs) | Height (cm) | Mass (kg) | Pain (0-10) | PHQ-8 (0-24) | OKS (0-48) |
|---|---|---|---|---|---|---|
| **Mean (SD)** | 65 (8) | 167 (9) | 94 (21) | 7 (2) | 5 (4) | 23 (7) |
| **Min, Max** | 52, 83 | 152, 188 | 48, 154 | 1, 10 | 0, 21 | 4, 37 |

With respect to concurrent validity between motion capture and in-lab IMU data, there was moderate to excellent agreement across all variables of interest (Table 4.2). The Bland-Altman plots illustrate the bias and limits of agreement between the two systems (Figure 4.2). Stride time showed particularly strong agreement, with a low mean difference (0.003 s), RMSE (0.04 s), and high correlation of ($r$ = 0.98). Stance time had slightly higher values for both mean difference (0.05 s) and RMSE (0.08 s) but still demonstrated excellent correlation ($r$ = 0.91). Swing time exhibited a small negative bias toward the motion capture system (-0.05 s) but had narrower limits of agreement (-0.15 to 0.06) than stance time. However, it showed only a moderate correlation ($r$ = 0.71), which was notably lower than for stride and stance times. Peak angular velocity had larger mean difference (20.2 deg/s) and RMSE (43.0 deg/s) but still showed moderate correlation ($r$ = 0.74), suggesting a consistent bias between the systems.

**Table 4.2.** Concurrent validity between motion capture and in-lab sensor data.

| | **Motion Capture Mean (SD)** | **In-Lab Sensors Mean (SD)** | **Mean Difference (Limits of Agreement)** | **RMSE** | **Pearson *r* (95% CI)[1]** |
|---|---|---|---|---|---|
| **Stride Time (s)** | 1.23 (0.19) | 1.23 (0.17) | 0.003 (-0.08, 0.08) | 0.04 | 0.98 (0.96, 0.99) |
| **Stance Time (s)** | 0.85 (0.17) | 0.80 (0.14) | 0.05 (-0.09, 0.18) | 0.08 | 0.91 (0.84, 0.95) |
| **Swing Time (s)** | 0.38 (0.03) | 0.43 (0.07) | -0.05 (-0.15, 0.06) | 0.07 | 0.71 (0.52, 0.83) |
| **Peak Angular Velocity (deg/s)** | 308.6 (50.1) | 288.3 (53.7) | 20.2 (-54.1, 94.6) | 43.0 | 0.74 (0.56, 0.85) |

[1] All p-values were < 0.001.

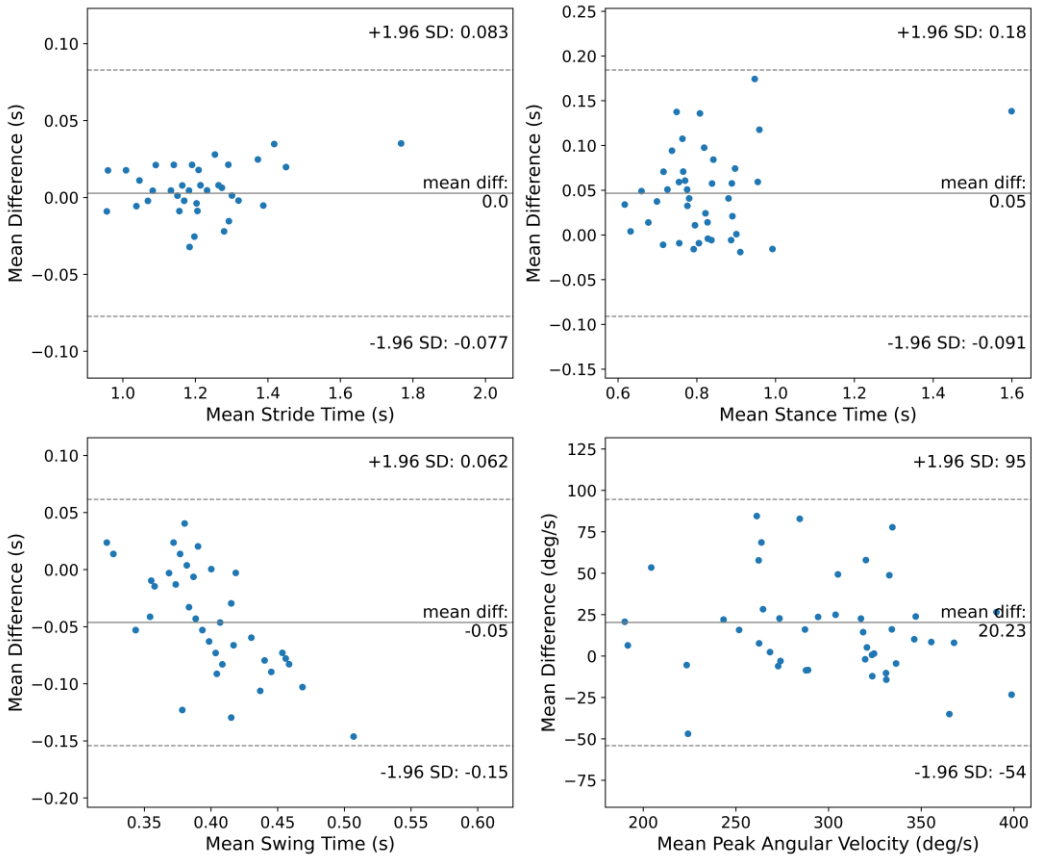



**Figure 4.2.** Bland-Altman plots for each variable between motion capture and in-lab IMU data.

Descriptive statistics and additional metrics for the gait variables of interest are presented in Table 4.3. When examining the in-lab and free-living ecological validity, all variables, except swing time, showed only moderate agreement. However, swing time just crossed the threshold for good agreement, with an ICC of 0.76. In general, gait measures

from free-living data reflected patterns of slower walking, with longer stride time and stance times. Peak angular velocity, derived from the mediolateral angular velocity during swing, was lower in free-living data, suggesting reduced limb movement intensity, even though swing time remains consistent.

**Table 4.3.** Ecological validity of gait variables from in-lab sensors and free-living sensors

| Variable | In-Lab Sensors Mean (SD) | Free-Living Sensors Mean (SD) | ICC(2,1) | Cohen's *d* | *p*-value |
|---|---|---|---|---|---|
| **Stride Time (s)** | 1.23 (0.17) | 1.26 (0.12) | 0.66 | -0.23 | 0.07 |
| **Stance Time (s)** | 0.80 (0.14) | 0.82 (0.11) | 0.66 | -0.14 | 0.17 |
| **Swing Time (s)** | 0.43 (0.07) | 0.42 (0.05) | 0.76 | 0.019 | 0.44 |
| **Peak Angular Velocity (deg/s)** | 288.3 (53.7) | 251.3 (40.8) | 0.71 | 0.77 | < 0.001 |

The histogram for stride time from a representative participant, whose median free-living stride time was within 5% of the overall sample median, is shown in Figure 4.3. This participant demonstrated nearly identical stride times between motion capture and in-lab IMU data (median: 1.08 ± 0.03 s vs. 1.08 ± 0.05 s). In contrast, the median stride time during free-living was 1.24 ± 0.31 s, which was 13.8% slower and substantially more variable. As shown in the histogram, the free-living stride time distribution was noticeably wider than the relatively narrow range observed in-lab, with a greater number of slower

strides. Based on this distribution, the participant's in-lab stride time (1.08 s) was faster than 92% of their free-living stride times. Stride time histograms for all participants are provided in Figure S1 at end of text.

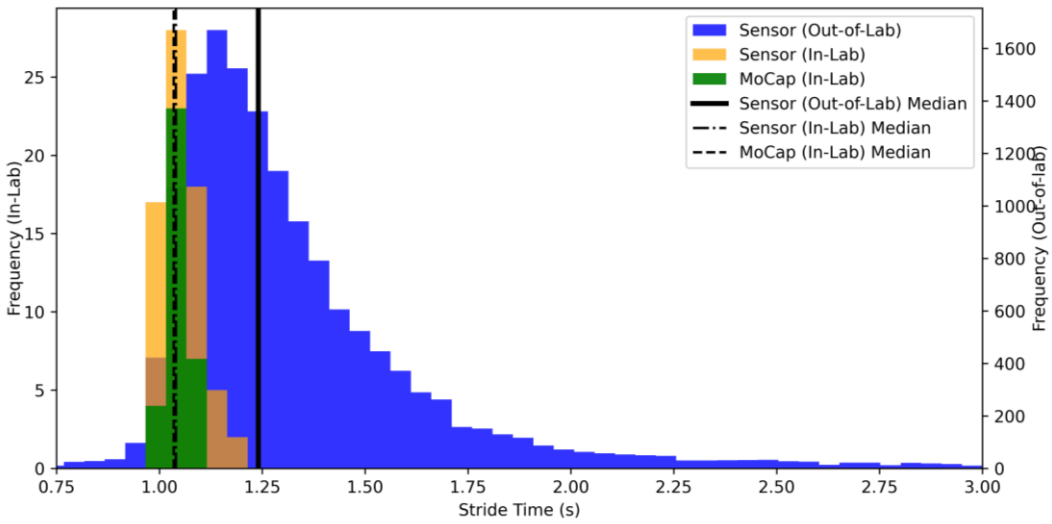



**Figure 4.3.** Representative plot of stride time distribution for in-lab data and free-living IMU data.

Gait variables and patient-reported outcome measures demonstrated several significant correlations relating to function (OKS), depressive symptoms (PHQ-8), and self-reported pain in the last week (Table 4.4). OKS and PHQ-8 showed similar patterns of associations with gait variables, reflecting their inverse scoring structures, where higher OKS scores indicate better function, while higher PHQ-8 scores indicate more depressive symptoms. Specifically, longer stride times and stance times were significantly associated

with worse self-reported function and greater depressive symptoms. Although these relationships were present for both in-lab and free-living measures, the strongest correlations were observed with in-lab data ($r$ = 0.37-0.48), compared to free-living ($r$ = 0.21-0.25). In contrast, the only IMU variable associated with self-reported pain was free-living peak angular velocity ($r$ = -0.31; $p$ = 0.04). Interestingly, this association was comparable to the relationship between pain and gait speed measured via in-lab motion capture. Additionally, peak angular velocity, whether assessed in-lab or during free-living (Figure 4.4a), showed some of the strongest correlations with OKS ($r$ = 0.44-0.45), highlighting its potential utility as a quantitative marker of functional, similar to gait speed (Figure 4.4b).

**Table 4.4.** Pearson correlations between gait metrics from in-lab, free-living, difference between metrics and patient reported outcome measures relating to function (OKS; higher scores indicate better function), depressive symptoms (PHQ-8; lower scores indicate fewer depressive symptoms), and pain (last week, visual analog scale; higher scores indicate more pain).

| | | OKS | | PHQ-8 | | Pain | |
|---|---|---|---|---|---|---|---|
| | | **Pearson *r*** | ***p*** | **Pearson *r*** | ***p*** | **Pearson *r*** | ***p*** |
| **Stride Time (s)** | **In-lab** | -0.46 (CI: -0.66, -0.19) | <0.001 | 0.43 (CI: 0.16, 0.64) | <0.001 | 0.15 (CI: -0.15, 0.42) | 0.34 |
| | **Free-living** | -0.21 (CI: -0.47, 0.09) | 0.17 | 0.18 (CI: -0.12, 0.45) | 0.24 | 0.09 (CI: -0.21, 0.38) | 0.54 |
| **Stance Time (s)** | **In-lab** | -0.48 (CI: -0.68, -0.22) | <0.001 | 0.37 (CI: 0.09, 0.6) | 0.01 | 0.26 (CI: -0.04, 0.51) | 0.09 |
| | **Free-living** | -0.25 (CI: -0.51, 0.04) | 0.09 | 0.21 (CI: -0.09, 0.48) | 0.16 | 0.11 (CI: -0.19, 0.39) | 0.47 |
| **Swing Time (s)** | **In-lab** | -0.12 (CI: -0.4, 0.18) | 0.43 | 0.29 (CI: -0.01, 0.53) | 0.06 | -0.16 (CI: -0.44, 0.14) | 0.29 |
| | **Free-living** | 0.05 (CI: -0.24, 0.34) | 0.72 | -0.07 (CI: -0.36, 0.23) | 0.63 | -0.11 (CI: -0.39, 0.19) | 0.49 |
| **Peak Angular Velocity (deg/s)** | **In-lab** | 0.44 (CI: 0.16, 0.65) | <0.001 | -0.24 (CI: -0.5, 0.05) | 0.11 | -0.16 (CI: -0.43, 0.14) | 0.31 |
| | **Free-living** | 0.45 (CI: 0.18, 0.66) | <0.001 | -0.26 (CI: -0.52, 0.03) | 0.08 | -0.31 (CI: -0.56, -0.02) | 0.04 |
| **Gait Speed (m/s)** | **Motion Capture** | 0.57 (CI: 0.34, 0.74) | <0.001 | -0.28 (CI: -0.53, 0.02) | 0.07 | -0.33 (CI: -0.57, -0.04) | 0.03 |

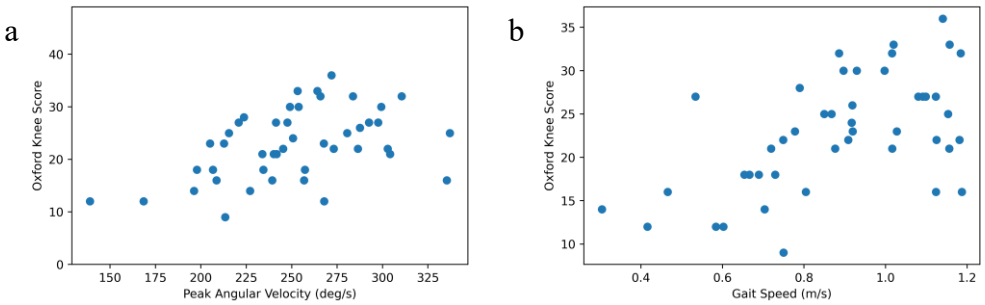



**Figure 4.4.** Scatter plots for (a) free-living peak angular velocity vs OKS and (b) gait speed (from motion capture) vs OKS. Both indicate higher velocities correlate with higher levels of self-reported function.

## 4.4 Discussion

The purpose of this study was to better understand the relationships between in-lab and free-living gait analyses in a cohort of knee OA patients. There was moderate to excellent agreement between systems measuring during in-lab assessments, particularly with stride time and stance time. However, this level of agreement did not extend to comparisons between in-lab and free-living gait, suggesting that even adults with knee OA preparing for joint arthroplasty surgery tend to walk faster and demonstrate higher functional performance in the lab than in daily life. Several gait metrics were also significantly associated with patient-reported outcomes, as expected. Notably, peak angular velocity, reflecting the intensity of limb swing during gait, was significantly related to self-reported function and served as a strong proxy for gait speed, which is often more

computationally complex to estimate. Taken together, these findings provide a clearer picture of the relationships between motion capture, in-lab IMU data, and gait collected over longer free-living periods.

The findings of this study generally align with previous research comparing motion capture and IMUs for spatiotemporal gait assessment. Although the current analysis presents median-level comparisons rather than stride-to-stride agreement, the results indicate that the systems remain closely aligned overall. In particular, these findings are consistent with prior studies demonstrating excellent agreement for stride time and stance time, along with the commonly observed reduction in agreement for swing time assessment (Jakob et al., 2021; Kobsar et al., 2020a, 2020b).

Additionally, the current findings align with previous research showing that individuals in both healthy older and clinical populations tend to walk more slowly outside of laboratory environments (Buekers et al., 2023; Del Din et al., 2016; Hillel et al., 2019; Shema-Shiratzky et al., 2020). This was also observed in the present study, where only moderate agreement was found between in-lab and free-living gait metrics in a sample of older adults with knee OA. These results are consistent with those reported by Hillel et al., where step time was greater, but not significantly different, in free-living compared to in-lab settings, while gait speed was significantly slower. When examining a representative participant, stride time measured via motion capture and in-lab sensors was identical, whereas free-living stride time was slower and exhibited a broader distribution. As shown in Figure 4.3, this greater variability is reflected in the histogram, with the in-lab stride time corresponding to the 92nd percentile of the free-living distribution. For this participant, the

difference between in-lab and free-living stride time was substantial, and similar patterns were observed in some additional participants (Figure S1). However, at the group level, the difference in stride time was small and not statistically significant (1.23 s vs. 1.26 s; Cohen's $d$ = 0.23). Interestingly, a more notable discrepancy was found in peak angular velocity (288.3 deg/s in-lab vs. 251.3 deg/s free-living), with an effect size of 0.77, highlighting this metric as a potentially sensitive marker of functional change. While environmental differences in stride time were less pronounced, the preoperative status of the cohort may have blunted free-living variability, suggesting that larger differences could emerge postoperatively.

A notable finding from this study is the level of agreement observed for peak angular velocity in the mediolateral axis (i.e., swing velocity), a native measure of IMUs placed on the shank. Given the long-standing use of motion capture in gait analysis, measures such as joint angles, stride length, and gait speed are often prioritized, leading researchers to use IMU data to estimate these variables rather than to focus on direct, native outputs. Our findings suggest that this metric not only shows good agreement with motion capture but may also offer clinical relevance. Although systematic differences were observed between the motion capture and IMU-derived values, the two remained well-correlated, suggesting these differences likely stem from inherent system characteristics and processing methods, rather than errors that undermine the utility of the metric. In fact, IMUs may provide a more direct measure of limb swing, potentially offering a better ground truth than optical motion capture in certain contexts. Notably, peak angular velocity may also serve as a practical proxy for gait speed, capturing movement intensity without

requiring step length estimation. Its clinical relevance is further supported by the significant correlations with self-reported pain and function (OKS) in free-living conditions. Nevertheless, further research is needed to understand its sensitivity to change relative to conventional gait metrics, particularly in response to treatment or recovery following surgery.

This study has limitations that should be noted. First, this study examined only a limited set of gait parameters. Most notably, the current study did not include gait metrics such as joint angles, stride length, or gait speed, which are prevalent for an OA population as markers of disease progression. However, these measures are difficult to estimate accurately using a single sensor at the shank. Additionally, as noted, IMU-derived estimates of these measures may be prone error, particularly over extended data collection periods such as a week. This highlights the potential value of native IMU measures as alternatives, though further research is needed to fully establish their viability as a clinically relevant gait metric. Finally, this study focused exclusively on individuals with late-stage knee OA. As such, the observed relationships and levels of agreement may not generalize to other populations, though the findings do provide a useful framework for evaluating gait measures across different contexts.

### 4.5 Conclusions

This study quantified the agreement between gait measures obtained from motion capture and wearable sensors, both in laboratory and free-living environments, in an OA population. These findings support the extension of gait analysis to longer-term, real-world data collection and highlight the potential utility of a native IMU-based measure of gait

intensity in future studies. With previous systematic reviews have noted that many "free-living" gait analyses are actually not truly free-living (Boekesteijn et al., 2022; Kobsar et al., 2020b), this study demonstrates the ability of these measures to be used in longer-term, truly free-living collections.

## 4.6 Supplemental Figure

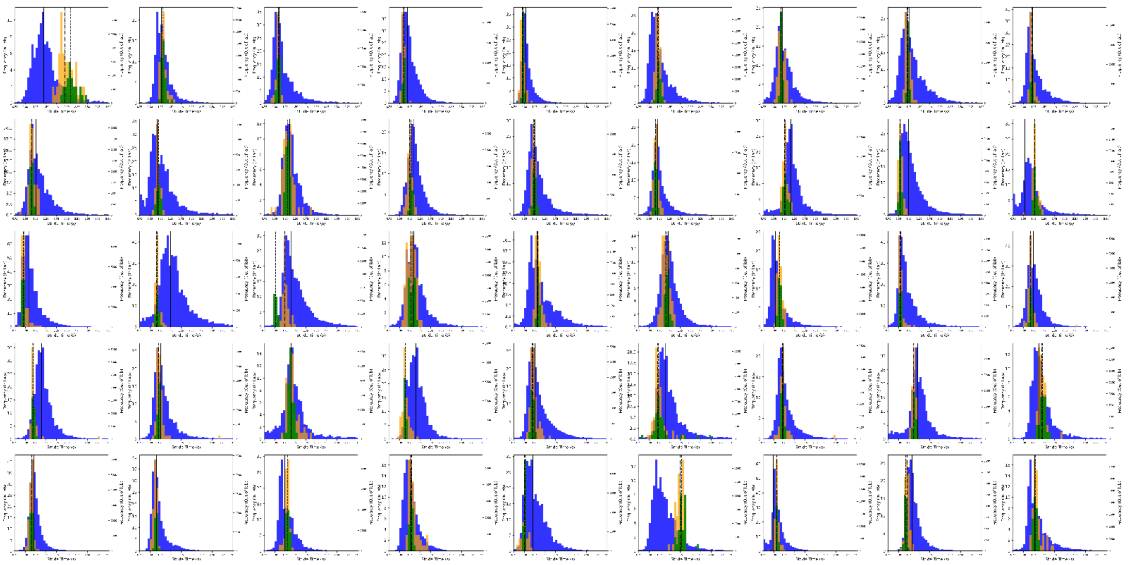


**Figure S1.** Stride time histograms for all study participants.

# Chapter 5: Evaluating the Longitudinal Sensitivity of In-Lab and Free-Living Gait Assessments in Knee Osteoarthritis (Study 3)

## Preamble

In my final study, I assessed the longitudinal sensitivity of gait metrics obtained from both in-lab motion capture and free-living wearable sensors. While in-lab recordings provide highly precise measurements, they capture only a brief snapshot of movement. Wearable sensors, in contrast, deliver longer, more representative, yet more variable, datasets. Previous research, summarized in Chapter 4, has largely focused on the agreement between these two modalities. I extended that work by examining how each set of measures changes over time. The findings indicate that in-lab metrics are more sensitive to functional changes, whereas wearable-sensor metrics may better reflect day-to-day activity.

This work was prepared for submission to *Osteoarthritis and Cartilage*, with the following coauthors:

Matthew C. Ruder, Vincenzo E. Di Bacco, Kim Madden, Anthony Adili, and Dylan Kobsar

## Abstract

Objective: To determine how markerless motion capture and free-living wearable inertial sensors jointly characterise gait in knee osteoarthritis (OA).

Methods: Gait data from a clinically adjacent hallway-based markerless system was compared with up to seven days of free-living data from bilateral shank-mounted sensors in 42 arthroplasty candidates. Spatiotemporal and kinematic outcomes were evaluated longitudinally from pre-operative assessment to at least one postoperative assessment up to 12-month follow-up.

Results: Laboratory-derived gait speed, stride time, and knee flexion as well as abduction/adduction excursion angles each showed clear, statistically significant improvements from the preoperative visit through every post-operative follow-up (all $p < 0.001$). These in-lab variables yielded the largest effect sizes, demonstrating high sensitivity to change. Analogous free-living inertial sensor metrics demonstrated the similar directional trends but reached statistical significance less often, typically only at the 12-month assessment. This likely reflects the greater day-to-day variability and smaller effect magnitudes that characterise real-world walking.

Conclusion: Markerless motion capture maximises sensitivity to surgical change, whereas wearable sensors may contextualise everyday mobility. When used together, they may yield a more complete picture of functional status than either modality alone. This work demonstrates the strengths and limitations of each system.

Impact: Findings inform design of pragmatic trials and clinical decision-support tools that triage care to those most likely to benefit.

## 5.1 Introduction

Gait impairment is a hallmark of many chronic diseases, yet objective gait assessments are rarely incorporated into routine clinical care. These assessments require dedicated laboratory space, specialized equipment, and additional time, making them difficult for both patients and clinicians to accommodate. Consequently, clinicians frequently rely on patient-reported outcome measures, which may not fully reflect a patient's real-world function (Braaksma et al., 2020). In conditions such as osteoarthritis (OA), gait is further influenced by pain and structural changes to the joint (Astephen Wilson et al., 2011). With OA prevalence steadily increasing among older adults, accessible tools to objectively quantify gait could enhance clinical assessment, guide treatment decisions, and ultimately improve patient outcomes (Kobsar et al., 2017; Outerleys et al., 2021).

Emerging technologies such as markerless motion capture and wearable sensors are making gait analysis far more accessible for clinical populations. Markerless motion capture eliminates the need for time consuming marker placement, enabling the collection of gait data during routine visits clinically feasible. Recent studies have demonstrated that joint angles (Kanko et al., 2021b) and spatiotemporal gait variables (Kanko et al., 2021c) obtained from markerless motion capture systems are comparable to those from traditional marker-based systems. Markerless motion capture systems also exhibit high intersession repeatability, with less than 3 degrees of variability across all lower limb joints (Kanko et al., 2021a). Wearable sensors offer additional advantages, including portability, cost-effectiveness, and the ability to assess gait over extended periods and in real-world

environments. Gait can be evaluated using single or multiple sensors placed at locations ranging from the lower back to the foot (Kobsar et al., 2020b). Their kinematic outputs, including joint angles and segment orientations, show strong agreement with motion capture data. Likewise, systematic reviews report good-to-excellent validity and reliability for sensor-derived spatiotemporal metrics (ICC = 0.81–0.94) in both healthy and osteoarthritis cohorts (Kobsar et al., 2020b, 2020a, 2016). Collectively, these advancements support the expanded integration of objective gait assessment into everyday clinical practice.

A key barrier to integrating gait analysis into routine care is understanding how sensitive common gait metrics are to change over time. Much of the existing research has focused on comparing motion capture and wearable sensor-derived gait metrics in cross-sectional studies. Specifically, several studies have reported differences in gait between healthy young adults (Lee et al., 2019), older adults (Werner et al., 2020), and clinical populations such as Parkinson's disease (Jakob et al., 2021; Serrao et al., 2019) and OA (Hafer et al., 2023; Seel et al., 2014) using both systems. While these cross-sectional findings provide valuable insights into their sensitivity to group-level differences, they do not address how well variables from these systems capture longitudinal change, an essential requirement for clinical monitoring.

While fewer studies have examined longitudinal changes in gait, initial research using both motion capture and wearable sensor systems show promising results. Solomonow-Avnon, et al., followed patients for one year after hip arthroplasty and

demonstrated that gait speed and several sagittal plane kinematic variables, measured with motion capture, predicted functional improvement (Solomonow-Avnon et al., 2017). Bolam, et al., also leveraged wearable sensors, monitoring cumulative load and impact asymmetry during the first six weeks of rehabilitation following total knee arthroplasty (Bolam et al., 2021). However, they did not evaluate the sensitivity of these measures to change, and the short follow-up period may have missed key recovery milestones, which often extend beyond six months (Paravlic et al., 2022). Collectively, these studies underscore the potential of objective gait metrics for monitoring change over time, but further work is needed to establish their sensitivity to longitudinally change.

Despite the growing interest in objective gait assessments for clinical populations, most prior research has either (1) compared motion capture and wearable sensor data at a single timepoint or (2) tracked longitudinal change using only one system. Consequently, the relative longitudinal sensitivity of the two approaches remains unclear, particularly in populations undergoing interventions such as total knee arthroplasty (TKA). To address this gap, we concurrently and longitudinally monitored gait in older adults with knee OA scheduled for TKA, collecting gait data in both in-lab motion capture and free-living collection with inertial sensors. Additionally, rather than analyzing only overlapping metrics (e.g., stride time), we assessed a broader set of variables unique to each modality alongside common measures. This design allowed us to characterize each system's sensitivity to change and clarify the strengths and limitations of motion capture versus wearable inertial sensors for clinical use.

## 5.2 Materials and Methods

### 5.2.1 Participants

Participants were recruited as part of a larger longitudinal study monitoring gait and functional changes in individuals scheduled for hip or knee arthroplasty. The Hamilton Integrated Research Ethics Board approved the study (HiREB 16236), and all participants provided written informed consent. All individuals had end-stage hip or knee OA, but only participants with knee OA were included in the present analysis. Those enrolled in this sub-study also agreed to wear two small wearable sensors for up to seven days at each assessment. Exclusion criteria included any significant disability that prevented ambulation or inability to provide informed consent. Gait assessments were scheduled preoperatively approximately two weeks before surgery, and again at approximately 3, 6, and 12 months postoperatively.

### 5.2.2 Data Collection Protocol

The data collection protocol was identical at every timepoint, with a baseline (preoperative; T00) and postoperative follow-ups at approximately 3 (T03), 6 (T06), and 12 months (T12). Patients first completed electronic patient-reported outcome measures on an iPad. Pain over the past week was rated on a visual analogue scale (VAS; 0-10, with 0 being no pain and 10 being the worst pain). Knee function was assessed with the Oxford Knee Score (OKS; scored 0–48, where higher scores indicate better function) (Dawson et al., 1998). At postoperative follow-up visits, participants were asked “With respect to your osteoarthritis, how would you rate yourself now compared to before the surgery?” to assess

surgical satisfaction, using an 11-point Likert scale, with -5 (much worse) and 5 (much better) as anchors. All patient-reported outcomes were collected at every assessment.

After completing the surveys, two inertial sensors (Axivity AX6, Axivity Ltd., Newcastle, UK) were attached to participants just below their knees. Each sensor was positioned on the medial aspect of the shank, slightly inferior to the tibial tuberosity, and secured with medical grade adhesive tape (Simpatch). The sensors were programmed to record linear accelerations (±8g at 100 Hz) and angular velocity data (±1,000 degrees/second at 100 Hz) continuously for seven days. Participants were provided with a prepaid envelope to return the devices by mail at the end of the recording period.

After affixing sensors, participants completed a series of functional tasks while recorded by markerless motion capture. A synchronized 10-camera system (Sony RX0-II, 60 Hz, Sony Corporation) installed within a clinically adjacent hallway captured each task. Functional tasks included a quiet standing trial, two walking trials (self-selected speed and faster than self-selected), a five repetition sit-to-stand, and a two-step ascent and descent. Walking tasks were completed along a 7.6 m walkway, outlined on the floor using a decal. For the self-selected walking task, participants were instructed to walk at their typical walking speed for 60 seconds. For the fast-walking task, they were asked to walk “faster than their typical walking speed” for 30 seconds. Participants were encouraged to complete each task to the best of their ability but could opt out of tasks they did not feel comfortable performing. Upon completion of functional tasks, participants left with sensors still attached and were instructed to maintain their typical routines.

### 5.2.3. Data Analysis

#### 5.2.3.1 In-lab Motion Capture

After each session, video files from all cameras were transferred to a local computer for processing and analysis. Markerless trajectories were generated using Theia3D (v2023.1.0.3161, Theia Markerless, Kingston, ON). Walking trials were then segmented into individual strides in Visual3D using a previously described algorithm (Outerleys et al., 2024b), and stride-level spatiotemporal and kinematic data were exported. The primary spatiotemporal variables were stride time, stance time, swing time, and gait speed. Knee flexion and knee abduction/adduction excursions during stance (i.e., the joint's range of motion) were selected as kinematic variables. These metrics are well-established indicators of gait health in laboratory studies (Huang et al., 2023; Kobsar et al., 2020b; Marcum et al., 2014; McCarthy et al., 2013; Solomonow-Avnon et al., 2017). Each variable was calculated for the operative ($Side_{Op}$) and nonoperative ($Side_{NonOp}$) limbs. Symmetry, expressed as a percentage relative to the operative limb, was defined as:

$$2 * \frac{Side_{Op} - Side_{NonOp}}{Side_{Op} + Side_{NonOp}} * 100 \quad \textit{(Eq. 1)}$$

#### 5.2.3.2 Free-Living Wearable Sensor

Wearable sensor processing was completed following the return of the inertial sensors at each timepoint. Upon delivery of the sensors, raw data were downloaded using OMGui (V1.0.0.43, Axivity Ltd, Newcastle, UK) in CWA file format. Processing was performed in Python 3.9 using Spyder as the integrated development environment to functionally align sensor data, identify walking bouts using the previously developed ML

gait detection algorithm (i.e., from Chapter 3), and segment each inertial sensor signal to the stride level for variable extraction.

For each sensor, data were first aligned using a combination of standing and walking data. Following the approach described by Mihy, et al., a section of acceleration data from the quiet standing trial was used to align the vertical axis with the gravity vector. Next, a section of angular velocity data rotating around the mediolateral axis during the self-selected walking trial was used to virtually rotate the sensor data, aligning the local segment coordinate system to the sagittal plane of the shank. The Axivity sensors are calibrated to output acceleration in gravitational units (*g*) and angular velocity in degrees per second (degrees/second). For analysis, acceleration data were converted to metres per second squared ($m/s^2$), while angular velocity was maintained as degrees per second (degrees/second).

Next, walking bouts were identified using a previously developed machine learning gait model (Li and Wang, 2022) and segmented into stride-level data. In short, the machine learning model, trained on healthy participants and individuals with knee OA, identified each five second window of data as one of eight activities, including walking (See Chapter 3). For the purposes of the current study, all non-walking segments were labeled as 0 and all walking segments as 1. Within each identified walking bout, initial contact and terminal contact were estimated using previously described methods (Mariani et al., 2013) before calculating stride-level variables (e.g. stride time, peak mediolateral angular velocity). Finally, a two-step outlier detection process was applied to minimize the inclusion of erroneous strides. First, principal component analysis was conducted on time-normalized

mediolateral angular velocity data, flagging strides with a z-score greater than 2.5 as outliers. Second, a stride similarity check was performed using the multiple correlation coefficient relative to a previously identified in-lab average stride section for each individual. Strides with a correlation coefficient below 0.5 were excluded. This process aimed to eliminate clearly atypical strides, ensuring that only representative gait cycles influenced the analysis.

For wearable sensor data, variables of interest were temporal and kinematic in nature. The temporal metrics (i.e., stride time, stance time, and swing time) paralleled those derived from motion capture but capture the wider variability present in free-living recording. Kinematic measures were taken directly from native sensor outputs to avoid the errors inherent in estimating joint angles or stride length. Peak resultant acceleration during stance is common biomechanical variable relating impact during gait (James et al., 2023; Lafortune, 1991). Peak frontal plane acceleration has been used as a proxy measure for sudden lateral knee movement (i.e., varus thrust) (Misu et al., 2022). Peak mediolateral angular velocity during swing, as shown in Chapter 4, may provide similar functional information to gait speed. Each variable was calculated separately for the operative and nonoperative limbs, and limb symmetry was evaluated using Eq. 1.

### 5.2.4 Statistical Analysis

Longitudinal change in each variable of interest, including patient-reported outcomes, was evaluated using a linear mixed-effects model (LMM):

$$Variable\ of\ interest \sim Timepoint + Age + Sex + (1 \mid Participant)$$

Timepoint included four ordered levels: preoperative (T00, reference), and postoperative assessments at 3 months (T03), 6 months (T06), and 12 months (T12). Age and sex (0 = female, 1 = male) were mean-centered, so the intercept reflects the expected value for an average aged participant of average sex at T00. A random intercept for each participant accounted for baseline differences. LMMs use all available observations under the missing-at-random assumption, accommodating participants with incomplete follow-up.

Estimated marginal means (EMMs) for each timepoint were generated from the fitted LMMs by setting covariates to their sample means. The EMM were used to calculate effect sizes, where Cohen's *d* was derived from EMM differences divided by the model-based pooled baseline standard deviation (SD). Residual normality and homoscedasticity were evaluated with Q-Q plots and residual-versus-fitted plots with no substantive violations were observed. For each postoperative timepoint comparted to the preoperative baseline, raw beta coefficients (i.e., change in original units) with 95% confidence intervals, *p*-values, standardized betas (i.e., change in SD units), and Cohen's *d* were reported. Bonferoni corrected *p*-values were also calculated between sequential visits (e.g. T00 to T03, T03 to T06, etc.). All analyses were conducted in Python 3.9 using statsmodels 0.14.0. Statistical significance was set at two-tailed $\alpha = 0.05$.

**5.3 Results**

In total, 42 patients (age: 66 ± 8 yrs, height = 168 ± 10 cm, weight = 91 ± 22 kg, 29 females, 13 males) consented to the study at their preoperative visit and completed at least one postsurgical follow up at 3-months (n=29), 6-months (n=25), or 12-months (n=22). Eight patients completed all four timepoints, and 26 patients attended at least two

postoperative follow-ups. An additional 32 patients had consented for the sub-study at their preoperative appointment but declined to return or were unreachable during postoperative appointment scheduling.

### 5.3.1 Patient-Reported Outcomes

The results from the LMM for the patient-reported outcomes are displayed in Table 5.1. Preoperatively, the cohort reflected typical knee OA presentation, with reduced self-reported function and moderate-to-high levels of pain. Postoperatively, pain was significantly reduced on average (estimated mean T00: 6.22; estimated mean T12: 1.43). Self-reported function initially improving markedly (estimated mean T00: 24.07; estimated mean T03: 36.61) before leveling off at T06 and T12 (39.32 and 40.07, respectively). Correspondingly, surgical satisfaction increased from three to twelve months, although inter-patient variability was slightly greater at the final visit.

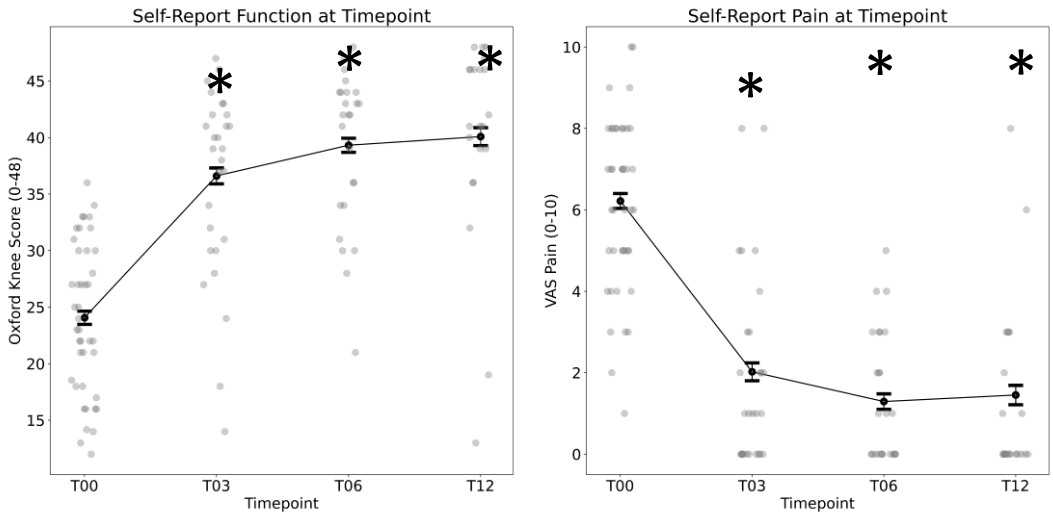



**Figure 5.1:** Self-reported function via Oxford Knee Score (left) and visual analog scale (VAS) pain. Preop = T00; postop = T03, T06, T12. After Bonferoni correction, all postoperative timepoints were different than T00 (indicated with * on plot).

**Table 5.1:** Results of linear mixed-effects models for self-reported outcomes at each study timepoint. Beta estimate includes a confidence interval.

| Variable | Visit | Marginal Mean | Beta Estimate | *p*-value | Beta (Stand.) | Cohen's *d* |
|---|---|---|---|---|---|---|
| **OKS** | T00 | 24.07 | - | - | - | - |
| | T03 | 36.61 | 12.64 (10.02 - 15.26) | < 0.001 | 0.52 | 1.69 |
| | T06 | 39.32 | 14.76 (12.02 - 17.50) | < 0.001 | 0.58 | 2.05 |
| | T12 | 40.07 | 15.91 (13.07 - 18.74) | < 0.001 | 0.61 | 2.15 |
| **Pain** | T00 | 6.22 | - | - | - | - |
| | T03 | 2.02 | -4.22 (-4.99 - -3.45) | < 0.001 | -0.59 | -2.06 |
| | T06 | 1.29 | -4.77 (-5.57 - -3.96) | < 0.001 | -0.64 | -2.41 |
| | T12 | 1.45 | -4.76 (-5.59 - -3.93) | < 0.001 | -0.61 | -2.34 |

### 5.3.2 Motion Capture

Results from the LMMs for each in-lab motion capture-derived variable of interest are presented in Table 5.2, with Bonferoni corrected *p*-values in Supplemental Table 5.1. Overall, these variables demonstrate changes indicative of improvements in function following the surgical intervention. Gait speed, displayed longitudinally in Figure 5.2 (left) increased significantly over time, with patients not only improving their gait speed immediately after surgery (T00: 0.92 m/s; T03: 1.03 m/s), but continuing to progress throughout the first postoperative year (T03: 1.03 m/s; T12: 1.18 m/s). Temporal variables relating to the phases of gait (stride, stance, and swing times) all showed significant

improvements (decreases) over the perioperative period. Stride time is displayed in Figure 5.2 (right), demonstrated a similar trend from lower (faster) stride times after surgery (T00: 1.23 s; T03: 1.16) and decreasing at each postoperative timepoint, albeit lesser improvement from 6 months to 12 months postoperatively (T03: 1.16 s; T06: 1.12 s; T12: 1.10 s).

Kinematic gait variables from motion capture also indicate functional gains. Knee flexion excursion increased from 12.5 degrees at T00 to 16.4 degrees at T12 (Figure 5.2, left). This is consistent with time-normalized, ensemble averaged knee flexion angle at each timepoint from all patients completing all timepoints (Figure 5.3, right). Similarly, frontal knee excursion decreased from 3.8 to 2.9 degrees between T00 to T06, before increasing to 3.53 degrees at T12 (Figure 5.3, left). The pronounced varus/valgus excursion seen preoperatively is greatly reduced after surgery and remains stable through the later follow-ups (Figure 5.3, right). Although frontal plane traces are inherently more variable than sagittal traces, the post-operative curves cluster more tightly. These patterns reinforce the temporal findings, with sagittal motion recovering more gradually and frontal plane stability improving immediately after surgery and then plateaus.

For measures of asymmetry, stride, stance, and swing time all showed general improvements over time. However, only swing time demonstrated consistent, statistically significant improvements in symmetry from T00 to T12. While knee flexion excursion asymmetry also showed a trend toward improvement, this did not reach statistical significance any timepoint. In contrast, abduction/adduction excursion asymmetry was

significantly greater at all postoperative timepoints compared to baseline. This suggests potential differences in mediolateral alignment between a knee with an implant versus one without.

**Table 5.2:** Results of linear mixed-effects models for variables derived from markerless motion capture. Asymmetry is relative to surgical side (i.e. positive = operative side values are higher, negative = nonoperative side values are higher). Beta estimates include a confidence interval.

| Variable | Visit | Marginal Mean | Beta Estimate | *p*-value | Beta (Stand.) | Cohen's *d* |
|---|---|---|---|---|---|---|
| **Gait Speed (m/s)** | T00 | 0.92 | - | - | - | - |
| | T03 | 1.03 | 0.11 (0.06 - 0.17) | < 0.001 | 0.19 | 0.49 |
| | T06 | 1.1 | 0.18 (0.12 - 0.23) | < 0.001 | 0.29 | 0.78 |
| | T12 | 1.18 | 0.25 (0.19 - 0.31) | < 0.001 | 0.38 | 1.12 |
| **Stride Time (s)** | T00 | 1.23 | - | - | - | - |
| | T03 | 1.16 | -0.07 (-0.11 - -0.04) | < 0.001 | -0.17 | -0.44 |
| | T06 | 1.12 | -0.1 (-0.14 - -0.06) | < 0.001 | -0.23 | -0.62 |
| | T12 | 1.10 | -0.14 (-0.18 - -0.10) | < 0.001 | -0.29 | -0.79 |
| **Stride Time Asymmetry (%)** | T00 | 0.16 | - | - | - | - |
| | T03 | -0.01 | -0.17 (-0.70 - 0.36) | 0.53 | -0.07 | -0.16 |
| | T06 | 0.01 | -0.15 (-0.70 - 0.41) | 0.60 | -0.06 | -0.14 |
| | T12 | -0.04 | -0.21 (-0.80 - 0.37) | 0.48 | -0.08 | -0.19 |
| | T00 | 0.84 | - | - | - | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Stance Time (s)** | T03 | 0.79 | -0.06 (-0.09 - -0.03) | < 0.001 | -0.16 | -0.39 |
| | T06 | 0.76 | -0.08 (-0.11 - -0.05) | < 0.001 | -0.21 | -0.57 |
| | T12 | 0.73 | -0.11 (-0.15 - -0.07) | < 0.001 | -0.27 | -0.75 |
| **Stance Time Asymmetry (%)** | T00 | -1.44 | - | - | - | - |
| | T03 | -0.67 | 0.85 (-0.23 - 1.92) | 0.12 | 0.14 | 0.31 |
| | T06 | -0.28 | 1.01 (-0.13 - 2.16) | 0.08 | 0.16 | 0.46 |
| | T12 | -0.43 | 0.88 (-0.34 - 2.09) | 0.16 | 0.13 | 0.40 |
| **Swing Time (s)** | T00 | 0.39 | - | - | - | - |
| | T03 | 0.37 | -0.01 (-0.03 - -0.00) | 0.01 | -0.17 | -0.40 |
| | T06 | 0.36 | -0.02 (-0.03 - -0.01) | < 0.001 | -0.25 | -0.64 |
| | T12 | 0.36 | -0.03 (-0.04 - -0.01) | < 0.001 | -0.27 | -0.70 |
| **Swing Time Asymmetry (%)** | T00 | 3.52 | - | - | - | - |
| | T03 | 1.37 | -2.27 (-4.40 - -0.14) | 0.04 | -0.18 | -0.42 |
| | T06 | 0.26 | -3.04 (-5.28 - -0.79) | 0.01 | -0.23 | -0.63 |
| | T12 | 0.65 | -2.67 (-5.07 - -0.28) | 0.03 | -0.19 | -0.56 |
| **Flexion Excursion (degrees)** | T00 | 3.78 | - | - | - | - |
| | T03 | 3.09 | -0.69 (-1.08 - -0.30) | 0.001 | -0.29 | -0.69 |
| | T06 | 2.92 | -0.86 (-1.27 - -0.46) | < 0.001 | -0.35 | -0.87 |
| | T12 | 3.53 | -0.24 (-0.68 - 0.20) | 0.29 | -0.09 | -0.25 |
| **Flexion Excursion Asymmetry (%)** | T00 | 7.07 | - | - | - | - |
| | T03 | -17.93 | -25.46 (-39.51 - -11.41) | < 0.001 | -0.33 | -0.79 |
| | T06 | -13.87 | -19.88 (-34.59 - -5.18) | 0.008 | -0.25 | -0.67 |
| | T12 | -9.28 | -15.89 (-31.65 - -0.13) | 0.05 | -0.18 | -0.52 |
| **Abduction/** | T00 | 12.47 | - | - | - | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Adduction Excursion (degrees)** | T03 | 13.04 | 0.62 (-0.80 - 2.04) | 0.39 | 0.06 | 0.15 |
| | T06 | 14.19 | 1.66 (0.17 - 3.15) | 0.03 | 0.17 | 0.45 |
| | T12 | 16.40 | 3.78 (2.17 - 5.39) | < 0.001 | 0.35 | 1.02 |
| **Abduction/ Adduction Excursion Asymmetry (%)** | T00 | -10.83 | - | - | - | - |
| | T03 | -11.5 | -0.42 (-11.29 - 10.46) | 0.94 | -0.01 | -0.02 |
| | T06 | -4.29 | 5.62 (-5.80 - 17.03) | 0.34 | 0.08 | 0.23 |
| | T12 | 0.43 | 11.68 (-0.61 - 23.97) | 0.06 | 0.16 | 0.39 |

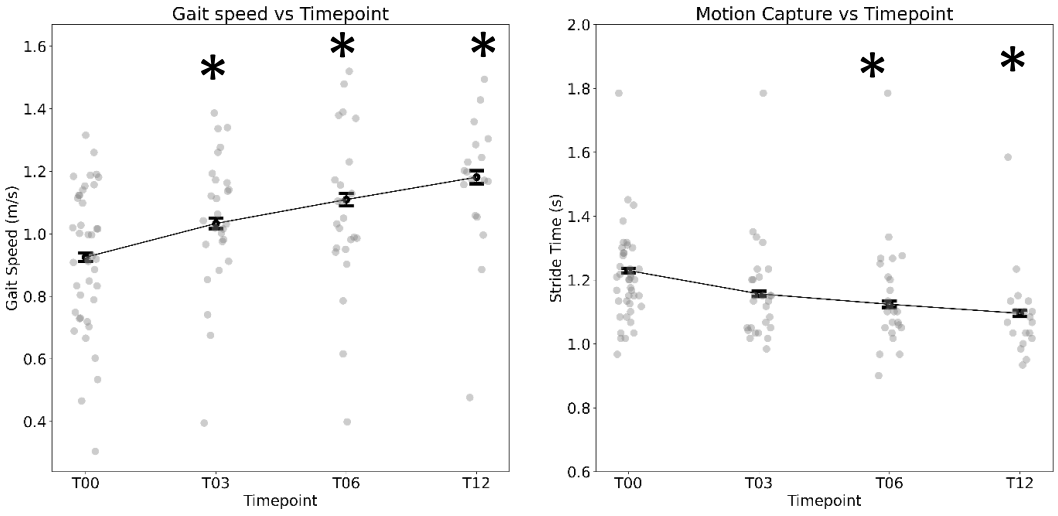



**Figure 5.2:** Gait speed (left) and stride time (right), derived from motion capture, plotted for each timepoint. When Bonferoni corrected was applied, there were only significant differences between T00 and all postoperative visits for gait speed (indicated with * on plot), and for stride time between T00 and T06 as well as T00 and T12 (indicated with * on plot).

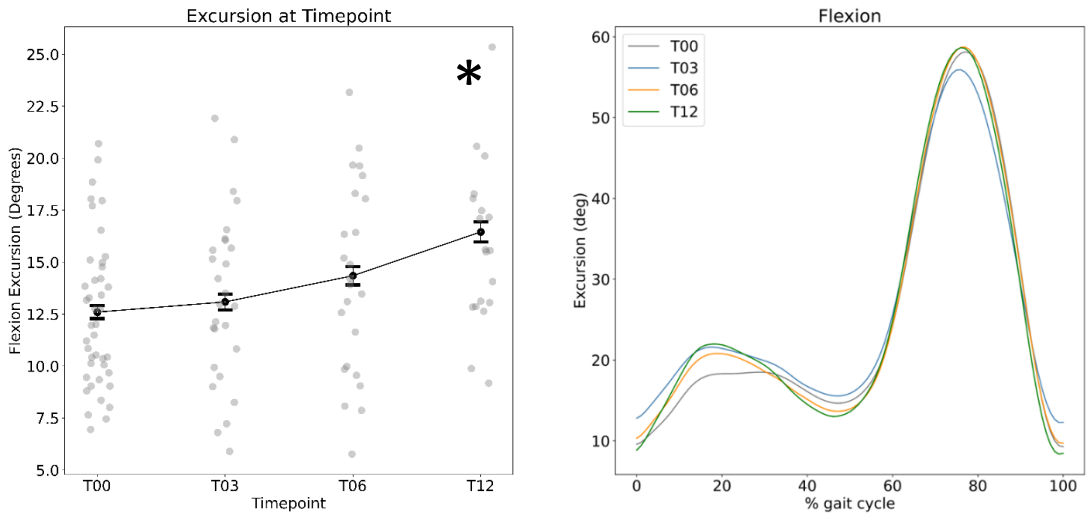



**Figure 5.3:** Knee joint flexion excursion at each timepoint (left). Normalized knee joint flexion angles averaged from participants who completed all each timepoints (right). Preop = T00; postop = T03, T06, T12. After Bonferoni correction, only T12 was different than T00 (indicated with * on plot).

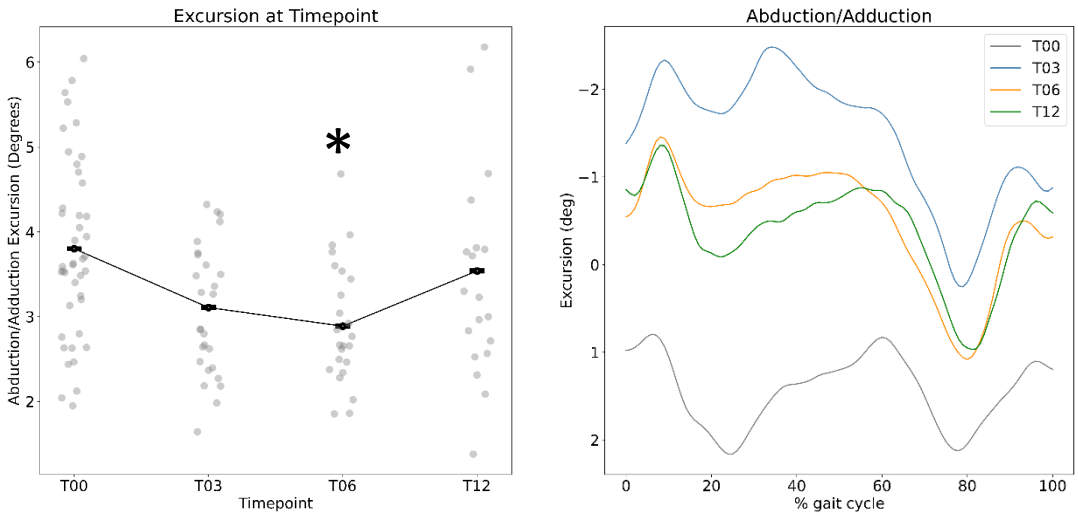



**Figure 5.4:** Frontal plane excursion at each timepoint (left). Normalized flexion and abduction/adduction knee joint angles averaged from participants who completed all each timepoints (right). Preop = T00; postop = T03, T06, T12. After Bonferoni correction, only T06 was different than T00 (indicated with * on plot).

### 5.3.3 Wearable Sensors

Similar trends were observed in the LMM results for variables derived from wearable sensors (Table 5.3, with Bonferoni corrected *p*-values in Supplemental Table 5.2). Consistent with the temporal results from motion capture, stride time, stance time, and swing time tended to decrease at postoperative timepoints, indicating potentially faster gait. However, unlike the in-lab motion capture results, these changes were not statistically significant.

Few significant changes were seen in the kinematic measures from inertial sensor. Peak sagittal angular velocity slightly, not significantly, increased postoperatively from T00 (229.27 degrees/second) to T03 (248.35 degrees/second) as might be expected with potentially faster gait, but leveled off at T06 (240.47 degrees/second) and T12 (248.30 degrees/second). There was little change preoperatively to postoperatively in peak resultant acceleration and peak stance frontal plane acceleration.

Asymmetry measures trends toward more symmetrical marginal means, but few of these changes reached significance. Notably, asymmetry in peak sagittal angular velocity (Figure 5.5, right) showed a moderate reduction (effect sizes: T06 = 0.50, $p$=0.03; T12 = 0.55; $p$=0.04). Peak frontal plane acceleration asymmetry (Figure 5.5, left) showed similar reductions (effect sizes: T06 = 0.48, $p$=0.03; T12 = 0.52; $p$=0.04). Large confidence intervals likely reflect greater variability in real-world asymmetries compared to in-lab assessments. Both asymmetry measures trended towards being more symmetrical further in the postoperative timeline.

**Table 5.3:** Results of linear mixed-effects models for variables derived from wearable sensors. Asymmetry is relative to surgical side (i.e. positive = operative side values are higher, negative = nonoperative side values are higher).

| Variable | Visit | Marginal Mean | Beta Estimate | *p*-value | Beta (Stand.) | Cohen's *d* |
|---|---|---|---|---|---|---|
| **Stride Time (s)** | T00 | 1.20 | - | - | - | - |
| | T03 | 1.25 | 0.05 (-0.04 - 0.14) | 0.30 | 0.10 | 0.24 |
| | T06 | 1.19 | -0.01 (-0.11 - 0.09) | 0.84 | -0.07 | -0.2 |
| | T12 | 1.15 | -0.06 (-0.16 - 0.04) | 0.27 | -0.11 | -0.26 |
| **Stride Time Asymmetry (%)** | T00 | 2.06 | - | - | - | - |
| | T03 | -2.15 | -4.15 (-12.14 - 3.83) | 0.31 | -0.12 | -0.27 |
| | T06 | 0.51 | -1.69 (-9.97 - 6.59) | 0.69 | -0.05 | -0.1 |
| | T12 | 2.95 | 0.83 (-8.34 - 10.0) | 0.86 | 0.02 | 0.06 |
| **Stance Time (s)** | T00 | 0.77 | - | - | - | - |
| | T03 | 0.80 | 0.03 (-0.03 - 0.08) | 0.37 | 0.09 | 0.22 |
| | T06 | 0.76 | -0.01 (-0.07 - 0.05) | 0.78 | -0.07 | -0.19 |
| | T12 | 0.74 | -0.04 (-0.10 - 0.02) | 0.22 | -0.12 | -0.3 |
| **Stance Time Asymmetry (%)** | T00 | -1.68 | - | - | - | - |
| | T03 | -4.61 | -2.88 (-11.19 - 5.44) | 0.50 | -0.08 | -0.18 |
| | T06 | -1.45 | 0.07 (-8.52 - 8.65) | 0.99 | 0.00 | 0.01 |
| | T12 | 0.89 | 2.09 | 0.66 | 0.05 | 0.16 |
| **Swing Time (s)** | T00 | 0.42 | - | - | - | - |
| | T03 | 0.44 | 0.02 (-0.02 - 0.07) | 0.19 | 0.10 | 0.25 |
| | T06 | 0.42 | 0.00 (-0.05 - 0.04) | 0.28 | -0.06 | -0.17 |
| | T12 | 0.41 | -0.01 | 0.58 | -0.06 | -0.11 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | (-0.06 - 0.03) | | | |
| **Swing Time Asymmetry (%)** | T00 | 10.84 | - | - | - | - |
| | T03 | 2.63 | -8.09 (-20.12 - 3.94) | 0.27 | -0.15 | -0.35 |
| | T06 | 4.07 | -7.03 (-19.64 - 5.59) | 0.93 | -0.12 | -0.29 |
| | T12 | 6.06 | -3.82 (-17.18 - 9.54) | 0.58 | -0.06 | -0.21 |
| **Peak Sagittal Angular Velocity (degrees/ second)** | T00 | 229.27 | - | - | - | - |
| | T03 | 248.35 | 18.71 (-5.10 - 42.53) | 0.12 | 0.14 | 0.33 |
| | T06 | 240.47 | 13.83 (-11.54 - 39.19) | 0.29 | 0.05 | 0.08 |
| | T12 | 248.30 | 16.56 (-10.36 - 43.48) | 0.23 | 0.11 | 0.32 |
| **Peak Sagittal Angular Velocity Asymmetry (%)** | T00 | -10.63 | - | - | - | - |
| | T03 | -5.99 | 4.74 (-5.87 - 15.35) | 0.38 | 0.09 | 0.21 |
| | T06 | 0.17 | 10.53 (-0.43 - 21.49) | 0.06 | 0.19 | 0.48 |
| | T12 | 1.54 | 11.74 (-0.12 - 23.61) | 0.05 | 0.19 | 0.52 |
| **Peak Stance Frontal Plane Acceleration ($m/s^2$)** | T00 | 0.59 | - | - | - | - |
| | T03 | 0.55 | -0.04 (-0.12 - 0.06) | 0.457 | -0.06 | -0.14 |
| | T06 | 0.61 | 0.04 (-0.06 - 0.13) | 0.47 | 0.06 | 0.10 |
| | T12 | 0.67 | 0.07 (-0.03 - 0.17) | 0.18 | 0.11 | 0.36 |
| **Peak Stance Frontal Plane Acceleration Asymmetry (%)** | T00 | -14.02 | - | - | - | - |
| | T03 | -18.96 | -4.87 (-20.89 - 11.15) | 0.55 | -0.05 | -0.13 |
| | T06 | 4.90 | 18.61 (2.02 - 35.22) | 0.03 | 0.19 | 0.50 |
| | T12 | 7.00 | 18.84 (0.90 - 36.78) | 0.04 | 0.18 | 0.55 |
| **Peak Resultant Acceleration ($m/s^2$)** | T00 | 2.19 | - | - | - | - |
| | T03 | 2.16 | -0.04 (-0.19 - 0.12) | 0.64 | -0.04 | -0.07 |
| | T06 | 2.14 | -0.01 (-0.18 - 0.15) | 0.87 | -0.05 | -0.21 |
| | T12 | 2.32 | 0.11 | 0.23 | 0.08 | 0.25 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | (-0.07 - 0.29) | | | |
| **Peak Resultant Asymmetry (%)** | T00 | -3.97 | - | - | - | - |
| | T03 | -7.32 | -3.36 (-11.12 - 4.40) | 0.40 | -0.10 | -0.22 |
| | T06 | 1.30 | 5.30 (-2.70 - 13.30) | 0.19 | 0.15 | 0.35 |
| | T12 | -0.70 | 3.12 (-5.49 - 11.73) | 0.48 | 0.08 | 0.22 |

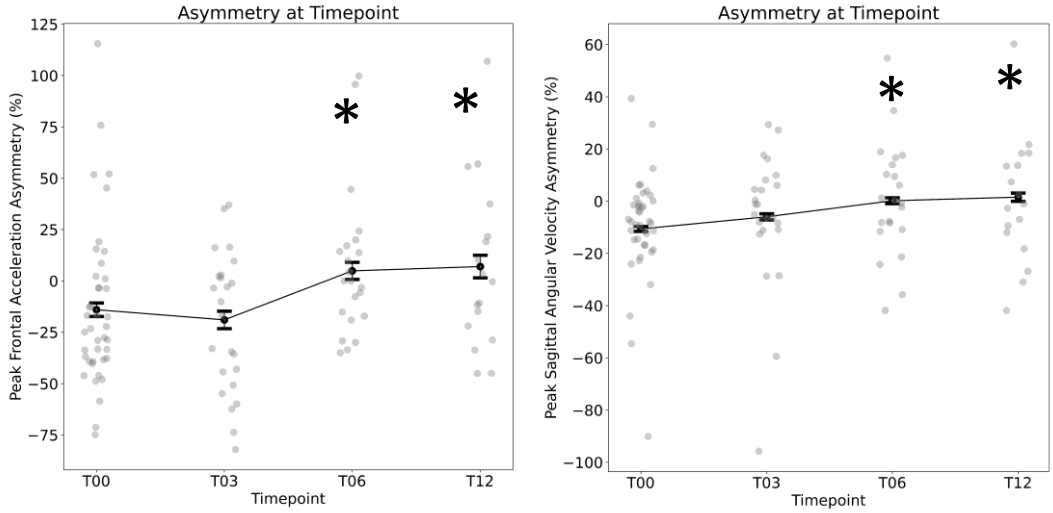



**Figure 5.5**: Peak frontal acceleration asymmetry (left) and peak sagittal angular velocity asymmetry (right), derived from free-living wearable sensor data, plotted for each timepoint. When Bonferoni corrected was applied, there were only significant differences between T00 and T06 as well as T00 and T12 postoperative visits for both variables (indicated with * on plot).

## 5.4 Discussion

This study evaluated how well gait metrics capture change when measured in-lab using markerless motion capture and in free-living conditions with multi-day inertial sensor recordings. A practical protocol was implemented for arthroplasty patients, combining a clinically adjacent hallway-based motion capture system with a minimal two sensor wearable setup. Laboratory-derived variables, including gait speed, stride time, and knee flexion and abduction/adduction excursion angles, demonstrated clear, statistically significant improvements from the preoperative visit through postoperative follow-ups. Free-living inertial sensor metrics exhibited similar directional trends but reached statistical significance less frequently, likely reflecting higher environmental variability and smaller effect sizes. While further research is needed to better understand the relationships between in-lab and free-living measures, as well as how to optimize the use of wearable sensor-derived metrics, this study establishes a realistic framework for integrating objective gait assessment into routine clinical practice.

Most validation studies compare wearable sensors directly with motion capture, treating the latter as the reference standard. These investigations typically report good agreement. For example, Hafer, et al., found that while absolute joint angle values differed between systems, the relative patterns were consistent, supporting comparable clinical interpretation even when the raw numbers diverged. Such work is essential for establishing the construct validity of sensor-derived metrics. However, the current study differed in approach. Rather than seeking one-to-one equivalence, we examined the sensitivity of each

modality to longitudinal change: the brief but detailed laboratory snapshot versus the multi-day, free-living record captured by inertial sensors. The results suggest that motion capture remains the more sensitive tool for detecting postoperative improvements, whereas wearable sensors, though showing similar directional trends, displayed weaker signal-to-noise under real-world conditions. From a clinical perspective, motion capture appears to provide a more sensitive measures of change than comparable metrics derived from wearable sensors.

The reduced sensitivity of free-living measures from wearable sensors may stem from several sources. In-lab walking can be different than free-living walking. In Hillel's week-long monitoring of 150 older adults, free-living gait was usually slower than what the same participants achieved in the lab. Specifically, median laboratory usual-walking speed exceeded the speed observed in approximately 64% of all 30-s daily-living bouts (63.8 ± 23.3%), while even the more demanding laboratory dual-task speed was faster than about half of those bouts (50.9 ± 25.4%). Put simply, the "snapshot" captured in the lab, whether single- or dual-task, tends to overestimate the pace at which people walk in their everyday lives. Consequently, while free-living measures may provide a more representative picture of everyday gait, it may not provide a more sensitive measure of change following arthroplasty.

Alternatively, some of the dissonance between in-lab and free-living measures may stem from the study population itself. A systematic review by Arnold, et al., reported minimal changes in physical activity, as measured by pedometers, in individuals following

total hip and knee arthroplasties. The review found mixed results at 6 months postoperatively, with larger improvements at 12 months in some studies, although activity levels remained lower than those of healthy controls. These findings are supported by a subsequent systematic review by Hammett, et al., which reported no improvements in physical activity at 6 months and only small to moderate improvement at 12 months. Despite consistent improvements in quality of life, pain, and physical function, physical activity in these patient populations did not improve substantially (Hammett et al., 2018). In the current study, similar improvements in pain and function were seen in-lab but not clearly reflected in free-living measures. This may reflect a true lack of change in everyday gait patterns, but it may also highlight fundamental differences in what each measurement system capture.

The observed discordance between measurement systems may be attributed to the different aspects of gait that each system captures. In the structured lab setting, patients may feel more motivated to walk faster or perform differently due to the awareness of being observed. This effect has been well-documented across both healthy and clinical gait populations (Ardestani and Hornby, 2020; Jeon et al., 2023; Robles-García et al., 2015). As a result, motion capture assessments may primarily reflect functional capacity, the best performance a patient is capable of under observation. Indeed, gait speed has consistently been shown to be a strong predictor of functional capacity across multiple populations (Ilgin et al., 2011; Nakamoto et al., 2015; Pophal Da Silva et al., 2021). In contrast, the lack of sensitivity to change in wearable sensors may reflect that patients, although

improving functionally, have not yet translated this increased capacity into their everyday activity patterns. As noted, physical activity levels tend to not increase substantially postoperatively, suggesting that free-living data may better represent actual day-to-day behavior rather than peak functional performance.

While this study produced several novel findings, there are also important limitations to consider. It is possible that more sensitive gait measurements could be derived from alternative sensor placement locations, such as the trunk or foot. However, the shank remains a commonly used placement for gait assessments in OA populations (Kobsar et al., 2020b), and provides data related to knee kinematics (e.g. peak frontal angular velocity) that other placements cannot. The shank placement is also less obstructive than trunk-mounted sensors and does depend on wearing shoes, as would be for a foot or insole mounted sensor. Ultimately, the research question should guide placement decisions in any study. Another limitation is the number of follow-up assessments completed by participants. Only 8 of 42 participants completed all four research visits, and only 26 of 42 participants completed at least two of the three postoperative follow-ups. While some degree of attrition is expected in longitudinal studies, higher follow-up rates could improve sensitivity of both measurement systems. It is important to note that the standard of care for postoperative follow-ups at the clinical partner includes only 6-week and 12-month visits. As such, the 3- and 6-month gait assessment follow-ups were not tied to receiving clinical care and required additional visits to the hospital, which not all participants were willing or able to attend. Although participants received a parking pass and a CAD $20 gift

card for these extra visits, these incentives only partly mitigated attrition. The statistical methods used were selected to account for missing data. However, it should also be noted that the study is ongoing, and some participants have not yet reached their final postoperative follow-ups, which may improve these numbers. Regardless, future researchers are encouraged to align research assessments with routine clinical visits to maximize study attendance.

## 5.5 Conclusions

This study sought to better characterize the sensitivity to change in gait measures obtained from both motion capture and wearable sensors in a cohort of knee OA patients. The findings support the use of motion capture to quantify gait before and after surgical intervention, with gait speed emerging as a particularly valuable indicator of functional capacity. While free-living gait measures were found to be less sensitive to change longitudinally, they demonstrated similar directional trends to those observed with motion capture. Taken together, these findings provide a practical foundation for integrating objective gait assessment, both in-lab and in daily life, into the clinical management of individuals with knee OA.

## 5.6 Supplemental Tables

**Supplemental Table 5.1**: Bonferoni corrected *p*-values between each timepoint for motion capture variables.

| Outcome | From | To | Difference | SE | t | df | $p_{raw}$ | $p_{Bonferoni}$ |
|---|---|---|---|---|---|---|---|---|
| **Gait Speed (m/s)** | T00 | T03 | 0.11 | 0.03 | 4.26 | 102 | < 0.001 | 0.03 |
| | T00 | T06 | 0.18 | 0.03 | 6.39 | 102 | < 0.001 | < 0.001 |
| | T00 | T12 | 0.25 | 0.03 | 8.23 | 102 | < 0.001 | < 0.001 |
| | T03 | T06 | 0.07 | 0.03 | 2.07 | 102 | 0.04 | 1.00 |
| | T03 | T12 | 0.14 | 0.04 | 3.91 | 102 | < 0.001 | 0.09 |
| | T06 | T12 | 0.07 | 0.04 | 2.03 | 102 | 0.05 | 1.00 |
| **Stride Time (s)** | T00 | T03 | -0.07 | 0.02 | -3.80 | 102 | < 0.001 | 0.14 |
| | T00 | T06 | -0.10 | 0.02 | -4.98 | 102 | < 0.001 | < 0.001 |
| | T00 | T12 | -0.14 | 0.02 | -6.23 | 102 | < 0.001 | < 0.001 |
| | T03 | T06 | -0.03 | 0.02 | -1.21 | 102 | 0.23 | 1.00 |
| | T03 | T12 | -0.06 | 0.03 | -2.52 | 102 | 0.01 | 1.00 |
| | T06 | T12 | -0.04 | 0.03 | -1.42 | 102 | 0.16 | 1.00 |
| **Stride Time Asymmetry (%)** | T00 | T03 | -0.17 | 0.27 | -0.64 | 102 | 0.53 | 1.00 |
| | T00 | T06 | -0.15 | 0.28 | -0.52 | 102 | 0.60 | 1.00 |
| | T00 | T12 | -0.21 | 0.30 | -0.71 | 102 | 0.48 | 1.00 |
| | T03 | T06 | 0.02 | 0.31 | 0.08 | 102 | 0.94 | 1.00 |
| | T03 | T12 | -0.04 | 0.32 | -0.13 | 102 | 0.90 | 1.00 |
| | T06 | T12 | -0.07 | 0.33 | -0.20 | 102 | 0.84 | 1.00 |
| **Stance Time (s)** | T00 | T03 | -0.06 | 0.02 | -3.56 | 102 | < 0.001 | 0.33 |
| | T00 | T06 | -0.08 | 0.02 | -4.73 | 102 | < 0.001 | < 0.001 |
| | T00 | T12 | -0.11 | 0.02 | -6.12 | 102 | < 0.001 | < 0.001 |
| | T03 | T06 | -0.02 | 0.02 | -1.19 | 102 | 0.24 | 1.00 |
| | T03 | T12 | -0.05 | 0.02 | -2.62 | 102 | 0.01 | 1.00 |
| | T06 | T12 | -0.03 | 0.02 | -1.54 | 102 | 0.13 | 1.00 |
| **Stance Time Asymmetry (%)** | T00 | T03 | 0.85 | 0.55 | 1.54 | 102 | 0.13 | 1.00 |
| | T00 | T06 | 1.01 | 0.58 | 1.73 | 102 | 0.09 | 1.00 |
| | T00 | T12 | 0.88 | 0.62 | 1.42 | 102 | 0.16 | 1.00 |
| | T03 | T06 | 0.17 | 0.64 | 0.26 | 102 | 0.80 | 1.00 |
| | T03 | T12 | 0.03 | 0.69 | 0.05 | 102 | 0.96 | 1.00 |
| | T06 | T12 | -0.14 | 0.70 | -0.19 | 102 | 0.85 | 1.00 |
| | T00 | T03 | -0.01 | 0.01 | -2.54 | 102 | 0.01 | 1.00 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Swing Time (s)** | T00 | T06 | -0.02 | 0.01 | -3.87 | 102 | < 0.001 | 0.11 |
| | T00 | T12 | -0.03 | 0.01 | -4.15 | 102 | < 0.001 | 0.04 |
| | T03 | T06 | -0.01 | 0.01 | -1.29 | 102 | 0.20 | 1.00 |
| | T03 | T12 | -0.01 | 0.01 | -1.68 | 102 | 0.10 | 1.00 |
| | T06 | T12 | 0.00 | 0.01 | -0.50 | 102 | 0.62 | 1.00 |
| **Swing Time Asymmetry (%)** | T00 | T03 | -2.27 | 1.09 | -2.08 | 102 | 0.04 | 1.00 |
| | T00 | T06 | -3.04 | 1.15 | -2.65 | 102 | 0.01 | 1.00 |
| | T00 | T12 | -2.67 | 1.22 | -2.19 | 102 | 0.03 | 1.00 |
| | T03 | T06 | -0.77 | 1.27 | -0.60 | 102 | 0.55 | 1.00 |
| | T03 | T12 | -0.41 | 1.37 | -0.30 | 102 | 0.77 | 1.00 |
| | T06 | T12 | 0.36 | 1.39 | 0.26 | 102 | 0.80 | 1.00 |
| **Flexion Excursion (degrees)** | T00 | T03 | 0.62 | 0.72 | 0.86 | 102 | 0.39 | 1.00 |
| | T00 | T06 | 1.66 | 0.76 | 2.19 | 102 | 0.03 | 1.00 |
| | T00 | T12 | 3.78 | 0.82 | 4.60 | 102 | < 0.001 | 0.01 |
| | T03 | T06 | 1.04 | 0.85 | 1.22 | 102 | 0.22 | 1.00 |
| | T03 | T12 | 3.16 | 0.93 | 3.40 | 102 | < 0.001 | 0.55 |
| | T06 | T12 | 2.12 | 0.93 | 2.27 | 102 | 0.03 | 1.00 |
| **Flexion Excursion Asymmetry (%)** | T00 | T03 | -0.42 | 5.55 | -0.08 | 102 | 0.94 | 1.00 |
| | T00 | T06 | 5.62 | 5.82 | 0.96 | 102 | 0.34 | 1.00 |
| | T00 | T12 | 11.68 | 6.27 | 1.86 | 102 | 0.07 | 1.00 |
| | T03 | T06 | 6.03 | 6.50 | 0.93 | 102 | 0.36 | 1.00 |
| | T03 | T12 | 12.10 | 7.07 | 1.71 | 102 | 0.09 | 1.00 |
| | T06 | T12 | 6.07 | 7.14 | 0.85 | 102 | 0.40 | 1.00 |
| **Abduction/ Adduction Excursion (degrees)** | T00 | T03 | -0.69 | 0.20 | -3.46 | 102 | < 0.001 | 0.46 |
| | T00 | T06 | -0.86 | 0.21 | -4.14 | 102 | < 0.001 | 0.04 |
| | T00 | T12 | -0.24 | 0.23 | -1.05 | 102 | 0.30 | 1.00 |
| | T03 | T06 | -0.18 | 0.23 | -0.75 | 102 | 0.45 | 1.00 |
| | T03 | T12 | 0.45 | 0.25 | 1.79 | 102 | 0.08 | 1.00 |
| | T06 | T12 | 0.63 | 0.26 | 2.45 | 102 | 0.02 | 1.00 |
| **Abduction/ Adduction Excursion Asymmetry (%)** | T00 | T03 | -25.46 | 7.17 | -3.55 | 102 | < 0.001 | 0.33 |
| | T00 | T06 | -19.88 | 7.50 | -2.65 | 102 | 0.01 | 1.00 |
| | T00 | T12 | -15.89 | 8.04 | -1.98 | 102 | 0.05 | 1.00 |
| | T03 | T06 | 5.58 | 8.32 | 0.67 | 102 | 0.50 | 1.00 |
| | T03 | T12 | 9.57 | 8.92 | 1.07 | 102 | 0.29 | 1.00 |
| | T06 | T12 | 4.00 | 9.08 | 0.44 | 102 | 0.66 | 1.00 |

**Supplemental Table 5.2**: Bonferoni corrected *p*-values between each timepoint for motion capture variables.

| Outcome | From | To | Difference | SE | t | df | $p_{raw}$ | $p_{Bonferoni}$ |
|---|---|---|---|---|---|---|---|---|
| **Stride Time (s)** | T00 | T03 | 0.05 | 0.05 | 1.05 | 98 | 0.30 | 1.00 |
| | T00 | T06 | -0.01 | 0.05 | -0.20 | 98 | 0.84 | 1.00 |
| | T00 | T12 | -0.06 | 0.05 | -1.11 | 98 | 0.27 | 1.00 |
| | T03 | T06 | -0.06 | 0.06 | -1.08 | 98 | 0.29 | 1.00 |
| | T03 | T12 | -0.11 | 0.06 | -1.83 | 98 | 0.07 | 1.00 |
| | T06 | T12 | -0.05 | 0.06 | -0.80 | 98 | 0.43 | 1.00 |
| **Stride Time Asymmetry (%)** | T00 | T03 | -4.15 | 4.07 | -1.02 | 92 | 0.31 | 1.00 |
| | T00 | T06 | -1.69 | 4.22 | -0.40 | 92 | 0.69 | 1.00 |
| | T00 | T12 | 0.84 | 4.68 | 0.18 | 92 | 0.86 | 1.00 |
| | T03 | T06 | 2.46 | 4.66 | 0.53 | 92 | 0.60 | 1.00 |
| | T03 | T12 | 4.99 | 5.08 | 0.98 | 92 | 0.33 | 1.00 |
| | T06 | T12 | 2.53 | 5.31 | 0.48 | 92 | 0.64 | 1.00 |
| **Stance Time (s)** | T00 | T03 | 0.03 | 0.03 | 0.90 | 98 | 0.37 | 1.00 |
| | T00 | T06 | -0.01 | 0.03 | -0.28 | 98 | 0.78 | 1.00 |
| | T00 | T12 | -0.04 | 0.03 | -1.23 | 98 | 0.22 | 1.00 |
| | T03 | T06 | -0.03 | 0.03 | -1.02 | 98 | 0.31 | 1.00 |
| | T03 | T12 | -0.07 | 0.04 | -1.81 | 98 | 0.07 | 1.00 |
| | T06 | T12 | -0.03 | 0.04 | -0.85 | 98 | 0.40 | 1.00 |
| **Stance Time Asymmetry (%)** | T00 | T03 | -2.88 | 4.24 | -0.68 | 92 | 0.50 | 1.00 |
| | T00 | T06 | 0.07 | 4.38 | 0.02 | 92 | 0.99 | 1.00 |
| | T00 | T12 | 2.09 | 4.74 | 0.44 | 92 | 0.66 | 1.00 |
| | T03 | T06 | 2.95 | 4.84 | 0.61 | 92 | 0.54 | 1.00 |
| | T03 | T12 | 4.97 | 5.16 | 0.96 | 92 | 0.34 | 1.00 |
| | T06 | T12 | 2.02 | 5.26 | 0.39 | 92 | 0.70 | 1.00 |
| **Swing Time (s)** | T00 | T03 | 0.02 | 0.02 | 1.11 | 98 | 0.27 | 1.00 |
| | T00 | T06 | 0.00 | 0.02 | -0.09 | 98 | 0.93 | 1.00 |
| | T00 | T12 | -0.01 | 0.02 | -0.55 | 98 | 0.59 | 1.00 |
| | T03 | T06 | -0.03 | 0.03 | -1.02 | 98 | 0.31 | 1.00 |
| | T03 | T12 | -0.04 | 0.03 | -1.38 | 98 | 0.17 | 1.00 |
| | T06 | T12 | -0.01 | 0.03 | -0.40 | 98 | 0.69 | 1.00 |
| **Swing Time Asymmetry (%)** | T00 | T03 | -8.09 | 6.14 | -1.32 | 92 | 0.19 | 1.00 |
| | T00 | T06 | -7.03 | 6.44 | -1.09 | 92 | 0.28 | 1.00 |
| | T00 | T12 | -3.82 | 6.82 | -0.56 | 92 | 0.58 | 1.00 |
| | T03 | T06 | 1.06 | 7.09 | 0.15 | 92 | 0.88 | 1.00 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | T03 | T12 | 4.27 | 7.46 | 0.57 | 92 | 0.57 | 1.00 |
| | T06 | T12 | 3.21 | 7.77 | 0.41 | 92 | 0.68 | 1.00 |
| **Peak Sagittal Angular Velocity (degrees/ second)** | T00 | T03 | 18.72 | 12.15 | 1.54 | 98 | 0.13 | 1.00 |
| | T00 | T06 | 13.83 | 12.94 | 1.07 | 98 | 0.29 | 1.00 |
| | T00 | T12 | 16.56 | 13.74 | 1.21 | 98 | 0.23 | 1.00 |
| | T03 | T06 | -4.89 | 14.45 | -0.34 | 98 | 0.74 | 1.00 |
| | T03 | T12 | -2.15 | 15.43 | -0.14 | 98 | 0.89 | 1.00 |
| | T06 | T12 | 2.74 | 15.97 | 0.17 | 98 | 0.86 | 1.00 |
| **Peak Sagittal Angular Velocity Asymmetry (%)** | T00 | T03 | 4.74 | 5.41 | 0.88 | 92 | 0.38 | 1.00 |
| | T00 | T06 | 10.53 | 5.59 | 1.88 | 92 | 0.06 | 1.00 |
| | T00 | T12 | 11.74 | 6.05 | 1.94 | 92 | 0.06 | 1.00 |
| | T03 | T06 | 5.79 | 6.29 | 0.92 | 92 | 0.36 | 1.00 |
| | T03 | T12 | 7.00 | 6.81 | 1.03 | 92 | 0.31 | 1.00 |
| | T06 | T12 | 1.22 | 6.91 | 0.18 | 92 | 0.86 | 1.00 |
| **Peak Stance Frontal Plane Acceleration (m/s2)** | T00 | T03 | -0.04 | 0.05 | -0.74 | 98 | 0.46 | 1.00 |
| | T00 | T06 | 0.04 | 0.05 | 0.72 | 98 | 0.47 | 1.00 |
| | T00 | T12 | 0.07 | 0.05 | 1.35 | 98 | 0.18 | 1.00 |
| | T03 | T06 | 0.07 | 0.06 | 1.26 | 98 | 0.21 | 1.00 |
| | T03 | T12 | 0.11 | 0.06 | 1.77 | 98 | 0.08 | 1.00 |
| | T06 | T12 | 0.04 | 0.06 | 0.58 | 98 | 0.57 | 1.00 |
| **Peak Stance Frontal Plane Acceleration Asymmetry (%)** | T00 | T03 | -4.87 | 8.18 | -0.60 | 92 | 0.55 | 1.00 |
| | T00 | T06 | 18.62 | 8.47 | 2.20 | 92 | 0.03 | 1.00 |
| | T00 | T12 | 18.84 | 9.15 | 2.06 | 92 | 0.04 | 1.00 |
| | T03 | T06 | 23.48 | 9.58 | 2.45 | 92 | 0.02 | 1.00 |
| | T03 | T12 | 23.71 | 10.40 | 2.28 | 92 | 0.03 | 1.00 |
| | T06 | T12 | 0.22 | 10.53 | 0.02 | 92 | 0.98 | 1.00 |
| **Peak Resultant Acceleration ($m/s^2$)** | T00 | T03 | -0.04 | 0.08 | -0.47 | 98 | 0.64 | 1.00 |
| | T00 | T06 | -0.01 | 0.09 | -0.17 | 98 | 0.87 | 1.00 |
| | T00 | T12 | 0.11 | 0.09 | 1.21 | 98 | 0.23 | 1.00 |
| | T03 | T06 | 0.02 | 0.10 | 0.24 | 98 | 0.81 | 1.00 |
| | T03 | T12 | 0.15 | 0.10 | 1.42 | 98 | 0.16 | 1.00 |
| | T06 | T12 | 0.12 | 0.11 | 1.16 | 98 | 0.25 | 1.00 |
| **Peak Resultant Asymmetry (%)** | T00 | T03 | -3.36 | 3.96 | -0.85 | 92 | 0.40 | 1.00 |
| | T00 | T06 | 5.30 | 4.08 | 1.30 | 92 | 0.20 | 1.00 |
| | T00 | T12 | 3.12 | 4.39 | 0.71 | 92 | 0.48 | 1.00 |
| | T03 | T06 | 8.66 | 4.51 | 1.92 | 92 | 0.06 | 1.00 |
| | T03 | T12 | 6.48 | 4.78 | 1.36 | 92 | 0.18 | 1.00 |
| | T06 | T12 | -2.18 | 4.89 | -0.45 | 92 | 0.66 | 1.00 |

# Chapter 6: General Discussion

## 6.1 Research Implications

As noted throughout this dissertation, the rising prevalence of osteoarthritis (OA) is expected to place increasing strain on healthcare systems as populations in many developed countries continue to age (Canadian Institute for Health Information, 2023; Li et al., 2024). Effectively prioritizing care for patients with the greatest need will therefore become essential. Objective gait metrics are poised to play a key role within a triage framework for OA management. The research presented in this dissertation lays important groundwork for advancing the use of objective gait assessment to support the health and treatment of older adults with OA.

Motion capture is a well-established technique for objectively quantifying gait by measuring spatiotemporal and kinematic parameters during walking. Beyond healthy populations, it has been widely applied to diverse clinical groups, including individuals with Parkinson's disease (Pistacchi, 2017), multiple sclerosis (Chua et al., 2014), and OA (Chehab et al., 2014), among others. The emergence of markerless systems is now making motion capture more accessible and enabling the development of large, multicentre datasets across sites (Horsak et al., 2024; Kanko et al., 2021b; Outerleys et al., 2021). Researchers have further enhanced traditional laboratory-based gait analyses by integrating complementary tools such as electromyography (Bovi et al., 2011) and force platforms (Chehab et al., 2014; Hurwitz et al., 2002). Wearable sensors, used independently or in

concert with motion capture, represent the next technological advance in objectively assessing gait.

This dissertation introduced a framework for collecting gait data using both motion capture and wearable sensor systems. By pairing a clinically adjacent hallway-based markerless motion capture setup with a sparse two sensor configuration on the shanks, clinics can collect high fidelity snapshots of gait alongside ecologically valid free-living records, all without prohibitive costs or space demands. Unlike earlier studies that examined sensitivity at a single in-lab time point across systems (Hafer et al., 2023) or tracked longitudinal change within one system (Bolam et al., 2021), this work helps clarify the trade-offs between the two modalities. Although the wearable sensor system proved less sensitive than motion capture, the results provide a clearer understanding of how the measures from each approach relate to one another.

Developing both the overall framework and the machine learning gait segmentation model were central to this dissertation. The framework was deliberately designed to be modular, allowing future extensions and refinements. Rather than focusing solely on derived laboratory metrics such as stride length or joint angles (Seel et al., 2014), this work emphasized “native” sensor variables (e.g., peak resultant acceleration, peak shank angular velocity) that are often overlooked in the gait literature. While simple measures like stride time can be calculated reliably from event detection, more complex variables risk compounding error in the inherently noisier free-living environment (Kobsar et al., 2020a, 2020b). The ResNet + BiLSTM model, trained on combined healthy and OA data,

outperformed a frequency-based heuristic method, particularly at the slower walking speeds common in late-stage OA. Clinically, this results in fewer fragmented walking bouts and more representative stride data for subsequent analyses.

By analyzing in-lab motion capture data alongside free-living wearable sensor data and tracking both longitudinally, this dissertation clarifies how these two measurement approaches interact and what each truly captures. At a single time point, gait metrics from wearable sensors and motion capture correlate well, echoing earlier validation studies (Hafer et al., 2023; Prisco et al., 2024; Seel et al., 2014). However, far fewer studies have compared those in-lab metrics with free-living data. Available work (e.g., Hillel et al., 2019) shows that gait characteristics outside the lab differ from those recorded within it (Hillel et al., 2019). Chapter 4 of this dissertation confirms both patterns within a single study: (1) wearable sensor gait measures generally agree with motion capture outputs in the lab, and (2) those same wearable measures reflect different gait behavior in free-living contexts. Longitudinally, markerless motion capture proved more sensitive than wearable inertial sensors for detecting postoperative recovery, whereas metrics from inertial sensors showed similar directional trends but greater variability. These findings suggest that the two technologies serve complementary rather than competing roles: in-lab sessions can detect subtle clinical changes, while wearables contextualize everyday gait behaviour. The results also highlight the trade-offs associated with relying on a single system, offering guidance to researchers and clinicians on when each modality is most appropriate.

## 6.2 Contributions to Existing Knowledge

The central thread of this dissertation is a baseline gait analysis framework introduced in Chapter 2. Designed to be modular and sensor agnostic, the framework balances clinical realism, using researcher-placed tibial sensors, with broad generalizability. It unifies sensor alignment, segmentation, event detection, and outlier handling within a single open platform. Although similar pipelines exist (Beyer et al., 2024), they typically focus on wrist and ankle sensors and rely on a single gyroscope channel, an approach that is less robust for clinical populations. In contrast, the present framework employs bilateral shank sensors, enabling symmetry assessments without overburdening patients and allowing for future extensions. Its versatility is demonstrated throughout this dissertation: Chapter 3 integrates a ResNet + BiLSTM model for gait detection, Chapter 4 uses the pipeline to compare motion capture and wearable data at a single time point, and Chapter 5 applies it to longitudinal monitoring in a clinical cohort. While additional refinements are possible, the current pipeline offers a robust, extensible foundation for future researchers and development.

In the first study (Chapter 3), an existing machine learning framework was retrained with additional data from both healthy adults and individuals with OA. Many gait analysis studies introduce new models trained on small datasets of young, healthy participants, leaving their performance on other sensor placements or clinical populations untested. Rather than creating a new architecture, we adapted a promising existing framework (Li and Wang, 2022) that had not previously been trained or evaluated on OA data to build an

OA-specific deep learning segmentation model. The resulting model achieved approximately 97% accuracy on clinical test data and significantly improved bout continuity compared with heuristic methods, particularly at the slower walking speeds typical of late-stage OA. This highlights a broader limitation in the literature, as models trained exclusively on small, healthy cohorts often generalize poorly to clinical populations (Halilaj et al., 2018).

The second study (Chapter 4) demonstrates that key inertial sensor-derived gait metrics are generally in agreement with those from outside the laboratory in knee OA patients, supporting their use for longitudinal monitoring. Previous research has primarily validated wearable sensor measures against motion capture, showing that both systems capture similar trends despite differing absolute values (Hafer et al., 2023). This agreement was replicated and extended in Chapter 4 to incorporate up to one week of free-living data. Moreover, peak shank angular velocity, a variable rarely studied in free-living settings, was evaluated and found to correlate with the Oxford Knee Score, paralleling trends observed for gait speed. Given its computational simplicity, this metric may offer a practical remote proxy for functional status.

The final study (Chapter 5) represents, to our knowledge, the first head-to-head, longitudinal comparison of sensitivity to change between markerless motion capture and multi-day inertial sensor recordings in arthroplasty patients. Laboratory-based metrics remain the gold standard for detecting postoperative change, whereas inertial sensors capture a broader behavioural context. Prior work assessed sensitivity either at a single

time point or within a single modality over time, leaving open questions about whether sensitivity holds across measurement systems and during extended free-living monitoring. The present results align with and extend previous research by examining these comparisons simultaneously within a single framework. While detailed lab assessments reflect functional capacity (Ilgin et al., 2011; Nakamoto et al., 2015), wearable sensor data provide a more representative picture of day-to-day activity (Boekesteijn et al., 2022; Buekers et al., 2023; Hillel et al., 2019). When used together, the two approaches offer a more comprehensive view of pre- and postoperative function than either can achieve alone, while also providing richer context than physical activity estimates alone.

### 6.3 Limitations and Considerations

This dissertation has several limitations, the most prominent being the sensor configuration. Chapter 2 outlined the rationale for using two shank-mounted sensors on the medial side, highlighting that this widely adopted placement captures valuable spatiotemporal features and knee-related kinematics. However, alternative sensor locations, such as the lower back or foot, could offer greater sensitivity and enable direct estimation of variables like stride length (Kobsar et al., 2020b). Additionally, while adding more sensors would increase setup complexity, the potential gains in sensitivity and data richness may justify the extra effort in certain research or clinical context. That being said, this work does not provide data to directly suggest that alternative placements, such as the back or foot, would be superior. For example, given the importance of gait speed as a

measure, a foot sensor placement would provide reliable estimates of this measure (Kobsar et al., 2020a). However, this remains an open question for future research.

This research also only looked at several gait-related variables independently from motion capture and wearable sensors. While many variables are available for analysis, it is not feasible to analyze them all, as such a subset was selected. As a result, there may be additional variables, such as gait variability or complexity, that could better leverage the unique strengths of each measurement system. Ultimately, there is a need to use motion capture and wearable sensors not independently, but as complementary tools, with each providing information inaccessible to the other. Integrating both measurement systems efficiently will be a key area of research in the future. This integration could range from straightforward data complementation, more in line with the current work, to more advanced approaches where the distinct contributions of each system are combined to reconstruct a more comprehensive view of patient function. For instance, biomechanical models might use lab-derived joint kinematics to more directly contextualize wearable-derived movement and loading patterns in real-world settings, offering a more holistic understanding of function and disease progression across environments. Additionally, this research also concentrated on kinematic and spatiotemporal variables, rather than kinetic, as these variables are both more clinically viable and more easily interpreted by patients and clinicians. While measures such as knee adduction moment are well-studied, this emphasis may lead to underutilizing of kinematic and spatiotemporal variables. Given their applicability in both laboratory and real-world environments, kinematic and spatiotemporal

variables may offer a more versatile and scalable basis for monitoring patient function, particularly when compared to kinetic measures that are typically constrained to controlled settings.

These findings may not generalize to all populations. The deep learning model was trained exclusively on healthy adults and individuals with end-stage OA. Although it outperformed a heuristic frequency-based method for gait segmentation, its performance may not translate to other clinical groups with altered gait patterns. Moreover, the algorithm was validated only in laboratory settings. Free-living validation remains largely qualitative in the absence of synchronized video recordings. Similarly, the sensitivities observed in the comparisons between motion capture and wearable sensors may differ in other clinical populations. Future studies should replicate and extend these motion capture versus wearable sensor comparisons in groups with more diverse or pathological gait characteristics.

Finally, the small sample size and participant attrition during follow-ups limit the strength of the findings. As noted in Chapter 5, only 42 patients completed at least one follow-up visit, and just 8 attended all four gait assessments. Although linear mixed models accommodate some missing data, the study would benefit from more complete follow-up data. Offering larger incentives or providing patients with personalized gait analysis reports could improve retention. Whenever possible, aligning study assessments with routine clinic visits, particularly in clinically integrated research, may further enhance follow-up attendance.

### 6.4 Recommendations for Future Work

Several recommendations emerge from this work. First, the framework should continue evolving toward its goal of being sensor-agnostic and modular. Although it has been validated in individuals with end-stage knee OA, adapting it to other populations may require alternative sensor placements that are more sensitive to specific pathologies (e.g., trunk or foot sensors for Parkinson's disease). Second, free-living validation of gait segmentation algorithms remains essential, despite the inherent challenges of obtaining ground-truth labels. Future efforts could draw on approaches that have succeeded in similar contexts. For example, in a study by Hickey, et al., participants wore an inertial sensor and a body mounted camera in free-living settings to evaluate their step counting algorithm for the sensor against gold standard video (Hickey et al., 2017). Using this approach, they found high relative (rho > 0.99) and absolute (ICC(2,1) > 0.94) agreement in identifying walking bouts and step counting. Finally, evaluating the machine learning architecture in additional clinical cohorts would clarify its robustness to other presentations of slow or altered gait.

Large, multi-site studies that track OA patients over longer periods, and across a broader range of disease severities and rehabilitation timelines, are essential to advancing gait-based clinical care. Like many areas of biomechanics, this field remains fundamentally limited by small, fragmented datasets. The framework developed in this dissertation represents a step toward addressing that gap, providing a foundation for scalable, standardized data collection. However, substantial work remains to achieve the large,

diverse datasets needed to establish normative gait values at each stage of OA progression. Prospective studies that follow early-stage patients with both motion capture and wearable sensors would help to identify the most relevant gait features to monitor, guide testing of nonsurgical interventions, and quantify functional decline until surgery becomes necessary. Such datasets would also support the development of robust, clinically relevant models that link gait characteristics to personalized rehabilitation plans. Although gait patterns have previously guided exercise interventions (Kobsar et al., 2017), larger and more diverse cohorts are needed to substantially improve predictive accuracy and clinical utility.

For gait analysis tools to gain traction in clinical practice, they must be both clinically viable and user-friendly enough to maintain patient engagement. Clinicians require succinct, easily interpreted gait reports that can inform treatment decisions, while patients need clear, accessible feedback on their progress and remaining deficits. This dissertation contributes to that goal by demonstrating the sensitivity of multiple gait measures across different measurement systems. However, despite examining a wide range of spatiotemporal and kinematic variables, this work represents only a subset of the many potential gait metrics that could be explored. Other metrics, such as gait complexity, haver shown promise for detecting subtle changes in movement patterns but were beyond the scope of this work (Bisi and Stagni, 2016). Likewise, additional integrated technologies, such as pressure sensing insoles, could further enrich gait data and deliver real-time postoperative feedback to support recovery (J. He et al., 2019). Integrating data from all these systems and using them in complementary roles will likely provide the clearest

picture of gait health. Exploring the utility of these additional variables and technologies represents a possible next step with the potential to reveal valuable insights along the longer journey of advancing and translating clinical gait assessments for OA.

### 6.5 Conclusions

This dissertation advances the field of clinical gait assessment by (i) developing a scalable processing framework, (ii) validating a deep learning segmentation model tailored to OA gait, (iii) demonstrating agreement free-living sensor metrics, and (iv) showing that markerless motion capture currently offers superior sensitivity to surgical change, while inertial sensors add ecological breadth. Together, these contributions lay the groundwork for personalised, data-driven management of knee OA that bridges the lab-clinic-home divide and set a roadmap for extending objective gait analytics to broader patient populations.

A central theme throughout this work is the recognition that no single technology captures the full picture of human gait. By leveraging the complementary strengths of motion capture and wearable sensors, this research illustrates how combining laboratory precision with free-living context provides a more comprehensive understanding of mobility, recovery, and function. While challenges remain, including the need for larger datasets, broader validation, and streamlined clinical tools, this work demonstrates that scalable, clinically meaningful gait assessment is both achievable and an essential next step toward improving OA management and mobility research more broadly.

## Chapter 7: References

Abe, T., Nagamune, K., 2021. Development of a Measuring System for Lateral Translation of the Knee Joint During Walking Using 6-axis Inertial Sensors, in: 2021 World Automation Congress (WAC). Presented at the 2021 World Automation Congress (WAC), IEEE, Taipei, Taiwan, pp. 121–124. https://doi.org/10.23919/WAC50355.2021.9559460

Alkjaer, T., Raffalt, P.C., Dalsgaard, H., Simonsen, E.B., Petersen, N.C., Bliddal, H., Henriksen, M., 2015. Gait variability and motor control in people with knee osteoarthritis. Gait & Posture 42, 479–484. https://doi.org/10.1016/j.gaitpost.2015.07.063

Allen, K.D., Coffman, C.J., Golightly, Y.M., Stechuchak, K.M., Keefe, F.J., 2009. Daily pain variations among patients with hand, hip, and knee osteoarthritis. Osteoarthritis and Cartilage 17, 1275–1282. https://doi.org/10.1016/j.joca.2009.03.021

Andaur Navarro, C.L., Damen, J.A.A., Takada, T., Nijman, S.W.J., Dhiman, P., Ma, J., Collins, G.S., Bajpai, R., Riley, R.D., Moons, K.G.M., Hooft, L., 2021. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. BMJ n2281. https://doi.org/10.1136/bmj.n2281

Andrews, A.W., Vallabhajosula, S., Boise, S., Bohannon, R.W., 2023. Normal gait speed varies by age and sex but not by geographical region: a systematic review. Journal of Physiotherapy 69, 47–52. https://doi.org/10.1016/j.jphys.2022.11.005

Andriacchi, T.P., Favre, J., 2014. The Nature of In Vivo Mechanical Signals That Influence Cartilage Health and Progression to Knee Osteoarthritis. Curr Rheumatol Rep 16, 463. https://doi.org/10.1007/s11926-014-0463-2

Ardestani, M.M., Hornby, T.G., 2020. Effect of investigator observation on gait parameters in individuals with stroke. Journal of Biomechanics 100, 109602. https://doi.org/10.1016/j.jbiomech.2020.109602

Asay, J.L., Boyer, K.A., Andriacchi, T.P., 2013. Repeatability of gait analysis for measuring knee osteoarthritis pain in patients with severe chronic pain: REPEATABILITY OF OA GAIT. J. Orthop. Res. 31, 1007–1012. https://doi.org/10.1002/jor.22228

Astephen Wilson, J.L., 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 19, 186–193. https://doi.org/10.1016/j.joca.2010.10.020

Attal, F., Amirat, Y., Chibani, A., Mohammed, S., October 25-29. Human Gait Phase Recognition using a Hidden Markov Model Framework. Presented at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV.

Benedetti, M.G., Merlo, A., Leardini, A., 2013. Inter-laboratory consistency of gait analysis measurements. Gait & Posture 38, 934–939. https://doi.org/10.1016/j.gaitpost.2013.04.022

Beyer, K.B., Weber, K.S., Cornish, B.F., Vert, A., Thai, V., Godkin, F.E., McIlroy, W.E., Van Ooteghem, K., 2024. NiMBaLWear analytics pipeline for wearable sensors: a

modular, open-source platform for evaluating multiple domains of health and behaviour. BMC Digit Health 2, 8. https://doi.org/10.1186/s44247-024-00062-3

Bini, S.A., Shah, R.F., Bendich, I., Patterson, J.T., Hwang, K.M., Zaid, M.B., 2019. Machine Learning Algorithms Can Use Wearable Sensor Data to Accurately Predict Six-Week Patient-Reported Outcome Scores Following Joint Replacement in a Prospective Trial. The Journal of Arthroplasty 34, 2242–2247. https://doi.org/10.1016/j.arth.2019.07.024

Birmingham, T.B., Moyer, R., Leitch, K., Chesworth, B., Bryant, D., Willits, K., Litchfield, R., Fowler, P.J., Giffin, J.R., 2017. Changes in biomechanical risk factors for knee osteoarthritis and their association with 5-year clinically important improvement after limb realignment surgery. Osteoarthritis and Cartilage 25, 1999–2006. https://doi.org/10.1016/j.joca.2017.08.017

Bisi, M.C., Stagni, R., 2016. Complexity of human gait pattern at different ages assessed using multiscale entropy: From development to decline. Gait & Posture 47, 37–42. https://doi.org/10.1016/j.gaitpost.2016.04.001

Boekesteijn, R.J., van Gerven, J., Geurts, A.C.H., Smulders, K., 2022. Objective gait assessment in individuals with knee osteoarthritis using inertial sensors: A systematic review and meta-analysis. Gait & Posture 98, 109–120. https://doi.org/10.1016/j.gaitpost.2022.09.002

Bolam, S.M., Batinica, B., Yeung, T.C., Weaver, S., Cantamessa, A., Vanderboor, T.C., Yeung, S., Munro, J.T., Fernandez, J.W., Besier, T.F., Monk, A.P., 2021. Remote

Patient Monitoring with Wearable Sensors Following Knee Arthroplasty. Sensors 21, 5143. https://doi.org/10.3390/s21155143

Bovi, G., Rabuffetti, M., Mazzoleni, P., Ferrarin, M., 2011. A multiple-task gait analysis approach: Kinematic, kinetic and EMG reference data for healthy young and adult subjects. Gait & Posture 33, 6–13. https://doi.org/10.1016/j.gaitpost.2010.08.009

Braaksma, C., Wolterbeek, N., Veen, M.R., Prinsen, C.A.C., Ostelo, R.W.J.G., 2020. Systematic review and meta-analysis of measurement properties of the Hip disability and Osteoarthritis Outcome Score - Physical Function Shortform (HOOS-PS) and the Knee Injury and Osteoarthritis Outcome Score - Physical Function Shortform (KOOS-PS). Osteoarthritis and Cartilage 28, 1525–1538. https://doi.org/10.1016/j.joca.2020.08.004

Buckley, C., Micó-Amigo, M.E., Dunne-Willows, M., Godfrey, A., Hickey, A., Lord, S., Rochester, L., Del Din, S., Moore, S.A., 2019. Gait Asymmetry Post-Stroke: Determining Valid and Reliable Methods Using a Single Accelerometer Located on the Trunk. Sensors 20, 37. https://doi.org/10.3390/s20010037

Buekers, J., Megaritis, D., Koch, S., Alcock, L., Ammour, N., Becker, C., Bertuletti, S., Bonci, T., Brown, P., Buckley, E., Buttery, S.C., Caulfied, B., Cereatti, A., Chynkiamis, N., Demeyer, H., Echevarria, C., Frei, A., Hansen, C., Hausdorff, J.M., Hopkinson, N.S., Hume, E., Kuederle, A., Maetzler, W., Mazzà, C., Micó-Amigo, E.M., Mueller, A., Palmerini, L., Salis, F., Scott, K., Troosters, T., Vereijken, B., Watz, H., Rochester, L., Del Din, S., Vogiatzis, I., Garcia-Aymerich, J., 2023. Laboratory and free-living gait performance in adults with COPD and healthy

controls. ERJ Open Res 9, 00159–02023. https://doi.org/10.1183/23120541.00159-2023

Burland, J.P., Outerleys, J.B., Lattermann, C., Davis, I.S., 2021. Reliability of wearable sensors to assess impact metrics during sport-specific tasks. Journal of Sports Sciences 39, 406–411. https://doi.org/10.1080/02640414.2020.1823131

Canadian Institute for Health Information, 2023. CJRR annual report: Hip and knee replacements in Canada, 2021–2022. CIHI, Ottawa, ON.

Caramia, C., Torricelli, D., Schmid, M., Munoz-Gonzalez, A., Gonzalez-Vargas, J., Grandas, F., Pons, J.L., 2018. IMU-Based Classification of Parkinson's Disease From Gait: A Sensitivity Analysis on Sensor Location and Feature Selection. IEEE J. Biomed. Health Inform. 22, 1765–1774. https://doi.org/10.1109/JBHI.2018.2865218

Carcreff, L., Gerber, C.N., Paraschiv-Ionescu, A., De Coulon, G., Newman, C.J., Aminian, K., Armand, S., 2020. Comparison of gait characteristics between clinical and daily life settings in children with cerebral palsy. Sci Rep 10, 2091. https://doi.org/10.1038/s41598-020-59002-6

Chang, A., Hayes, K., Dunlop, D., Hurwitz, D., Song, J., Cahue, S., Genge, R., Sharma, L., 2004. Thrust during ambulation and the progression of knee osteoarthritis. Arthritis Rheum 50, 3897–3903. https://doi.org/10.1002/art.20657

Chapple, C.M., Nicholson, H., Baxter, G.D., Abbott, J.H., 2011. Patient characteristics that predict progression of knee osteoarthritis: A systematic review of prognostic studies. Arthritis Care Res 63, 1115–1125. https://doi.org/10.1002/acr.20492

Chehab, E.F., Favre, J., Erhart-Hledik, J.C., Andriacchi, T.P., 2014. Baseline knee adduction and flexion moments during walking are both associated with 5 year cartilage changes in patients with medial knee osteoarthritis. Osteoarthritis and Cartilage 22, 1833–1839. https://doi.org/10.1016/j.joca.2014.08.009

Chen, D., Cai, Y., Qian, X., Ansari, R., Xu, W., Chu, K.-C., Huang, M.-C., 2020. Bring Gait Lab to Everyday Life: Gait Analysis in Terms of Activities of Daily Living. IEEE Internet Things J. 7, 1298–1312. https://doi.org/10.1109/JIOT.2019.2954387

Chen, S., Lach, J., Lo, B., Yang, G.-Z., 2016. Toward Pervasive Gait Analysis With Wearable Sensors: A Systematic Review. IEEE J. Biomed. Health Inform. 20, 1521–1537. https://doi.org/10.1109/JBHI.2016.2608720

Chua, M.C., Hyngstrom, A.S., Ng, A.V., Schmit, B.D., 2014. Movement strategies for maintaining standing balance during arm tracking in people with multiple sclerosis. Journal of Neurophysiology 112, 1656–1666. https://doi.org/10.1152/jn.00598.2013

Costello, K.E., Eigenbrot, S., Geronimo, A., Guermazi, A., Felson, D.T., Richards, J., Kumar, D., 2020. Quantifying varus thrust in knee osteoarthritis using wearable inertial sensors: A proof of concept. Clinical Biomechanics 80, 105232. https://doi.org/10.1016/j.clinbiomech.2020.105232

Dawson, J., Fitzpatrick, R., Murray, D., Carr, A., 1998. Questionnaire on the perceptions of patients about total knee replacement. The Journal of Bone and Joint Surgery. British volume 80-B, 63–69. https://doi.org/10.1302/0301-620X.80B1.0800063

Deie, M., Hoso, T., Shimada, N., Iwaki, D., Nakamae, A., Adachi, N., Ochi, M., 2014. Differences between opening versus closing high tibial osteotomy on clinical outcomes and gait analysis. The Knee 21, 1046–1051. https://doi.org/10.1016/j.knee.2014.04.007

Del Din, S., Godfrey, A., Galna, B., Lord, S., Rochester, L., 2016. Free-living gait characteristics in ageing and Parkinson's disease: impact of environment and ambulatory bout length. J NeuroEngineering Rehabil 13, 46. https://doi.org/10.1186/s12984-016-0154-5

Dong, M., Fan, H., Yang, D., Sun, X., Yan, C., Feng, Y., 2023. Comparison of spatiotemporal, kinematic, and kinetic gait characteristics in total and unicompartmental knee arthroplasty during level walking: A systematic review and meta-analysis. Gait & Posture 104, 58–69. https://doi.org/10.1016/j.gaitpost.2023.06.005

D'Souza, N., Charlton, J., Grayson, J., Kobayashi, S., Hutchison, L., Hunt, M., Simic, M., 2022. Are biomechanics during gait associated with the structural disease onset and progression of lower limb osteoarthritis? A systematic review and meta-analysis. Osteoarthritis and Cartilage 30, 381–394. https://doi.org/10.1016/j.joca.2021.10.010

Duffell, L.D., Jordan, S.J., Cobb, J.P., McGregor, A.H., 2017. Gait adaptations with aging in healthy participants and people with knee-joint osteoarthritis. Gait & Posture 57, 246–251. https://doi.org/10.1016/j.gaitpost.2017.06.015

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 22, 457–463. https://doi.org/10.1016/j.joca.2013.12.015

Favre, J., Erhart-Hledik, J.C., Chehab, E.F., Andriacchi, T.P., 2016. Baseline ambulatory knee kinematics are associated with changes in cartilage thickness in osteoarthritic patients over 5 years. Journal of Biomechanics 49, 1859–1864. https://doi.org/10.1016/j.jbiomech.2016.04.029

Ferrari, A., Cutti, A.G., Cappello, A., 2010. A new formulation of the coefficient of multiple correlation to assess the similarity of waveforms measured synchronously by different motion analysis protocols. Gait & Posture 31, 540–542. https://doi.org/10.1016/j.gaitpost.2010.02.009

Foucher, K.C., Thorp, L.E., Orozco, D., Hildebrand, M., Wimmer, M.A., 2010. Differences in Preferred Walking Speeds in a Gait Laboratory Compared With the Real World After Total Hip Replacement. Archives of Physical Medicine and Rehabilitation 91, 1390–1395. https://doi.org/10.1016/j.apmr.2010.06.015

Gafoor, F., Ruder, M., Kobsar, D., 2024. Validation of physical activity levels from shank-placed Axivity AX6 accelerometers in older adults. PLoS ONE 19, e0290912. https://doi.org/10.1371/journal.pone.0290912

García-de-Villa, S., Ruiz, L.R., Neira, G.G.-V., Álvarez, M.N., Huertas-Hoyas, E., del-Ama, A.J., Rodríguez-Sánchez, M.C., Seco, F., Jiménez, A.R., 2025. Validation of an IMU-Based Gait Analysis Method for Assessment of Fall Risk Against

Traditional Methods. IEEE J. Biomed. Health Inform. 29, 107–117. https://doi.org/10.1109/JBHI.2024.3434973

Gholami, M., Napier, C., Patiño, A.G., Cuthbert, T.J., Menon, C., 2020. Fatigue Monitoring in Running Using Flexible Textile Wearable Sensors. Sensors 20, 5573. https://doi.org/10.3390/s20195573

Gurchiek, R.D., Choquette, R.H., Beynnon, B.D., Slauterbeck, J.R., Tourville, T.W., Toth, M.J., McGinnis, R.S., 2019. Open-Source Remote Gait Analysis: A Post-Surgery Patient Monitoring Application. Sci Rep 9, 17966. https://doi.org/10.1038/s41598-019-54399-1

Hafer, J.F., Mihy, J.A., Hunt, A., Zernicke, R.F., Johnson, R.T., 2023. Lower Extremity Inverse Kinematics Results Differ Between Inertial Measurement Unit- and Marker-Derived Gait Data. Journal of Applied Biomechanics 39, 133–142. https://doi.org/10.1123/jab.2022-0194

Hafer, J.F., Provenzano, S.G., Kern, K.L., Agresta, C.E., Grant, J.A., Zernicke, R.F., 2020. Measuring markers of aging and knee osteoarthritis gait using inertial measurement units. Journal of Biomechanics 99, 109567. https://doi.org/10.1016/j.jbiomech.2019.109567

Halilaj, E., Rajagopal, A., Fiterau, M., Hicks, J.L., Hastie, T.J., Delp, S.L., 2018. Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. Journal of Biomechanics 81, 1–11. https://doi.org/10.1016/j.jbiomech.2018.09.009

Hammett, T., Simonian, A., Austin, M., Butler, R., Allen, K.D., Ledbetter, L., Goode, A.P., 2018. Changes in Physical Activity After Total Hip or Knee Arthroplasty: A Systematic Review and Meta-Analysis of Six- and Twelve-Month Outcomes. Arthritis Care & Research 70, 892–901. https://doi.org/10.1002/acr.23415

He, J., Lippmann, K., Shakoor, N., Ferrigno, C., Wimmer, M.A., 2019. Unsupervised gait retraining using a wireless pressure-detecting shoe insole. Gait & Posture 70, 408–413. https://doi.org/10.1016/j.gaitpost.2019.03.021

He, Z., Liu, T., Yi, J., 2019. A Wearable Sensing and Training System: Towards Gait Rehabilitation for Elderly Patients With Knee Osteoarthritis. IEEE Sensors J. 19, 5936–5945. https://doi.org/10.1109/JSEN.2019.2908417

Helwig, J., Diels, J., Röll, M., Mahler, H., Gollhofer, A., Roecker, K., Willwacher, S., 2023. Relationships between External, Wearable Sensor-Based, and Internal Parameters: A Systematic Review. Sensors 23, 827. https://doi.org/10.3390/s23020827

Hickey, A., Del Din, S., Rochester, L., Godfrey, A., 2017. Detecting free-living steps and walking bouts: validating an algorithm for macro gait analysis. Physiol. Meas. 38, N1–N15. https://doi.org/10.1088/1361-6579/38/1/N1

Hillel, I., Gazit, E., Nieuwboer, A., Avanzino, L., Rochester, L., Cereatti, A., Croce, U.D., Rikkert, M.O., Bloem, B.R., Pelosin, E., Del Din, S., Ginis, P., Giladi, N., Mirelman, A., Hausdorff, J.M., 2019. Is every-day walking in older adults more analogous to dual-task walking or to usual walking? Elucidating the gaps between gait performance in the lab and during 24/7 monitoring. Eur Rev Aging Phys Act 16, 6. https://doi.org/10.1186/s11556-019-0214-5

Horsak, B., Prock, K., Krondorfer, P., Siragy, T., Simonlehner, M., Dumphart, B., 2024. Inter-trial variability is higher in 3D markerless compared to marker-based motion capture: Implications for data post-processing and analysis. Journal of Biomechanics 166, 112049. https://doi.org/10.1016/j.jbiomech.2024.112049

Huang, C., Chan, P.-K., Chiu, K.-Y., Yan, C.-H., Yeung, S.-S., Lai, C.W.-K., Leung, A.K.-L., Fu, S.N., 2023. The association between tibial torsion, knee flexion excursion and foot progression during gait in people with knee osteoarthritis: a cross-sectional study. BMC Sports Sci Med Rehabil 15, 110. https://doi.org/10.1186/s13102-023-00726-z

Hurwitz, D.E., Ryals, A.B., Case, J.P., Block, J.A., Andriacchi, T.P., 2002. The knee adduction moment during gait in subjects with knee osteoarthritis is more closely correlated with static alignment than radiographic disease severity, toe out angle and pain. J. Orthop. Res. 20, 101–107. https://doi.org/10.1016/S0736-0266(01)00081-X

Ilgin, D., Ozalevli, S., Kilinc, O., Sevinc, C., Cimrin, A., Ucan, E., 2011. Gait speed as a functional capacity indicator in patients with chronic obstructive pulmonary disease. Ann Thorac Med 6, 141. https://doi.org/10.4103/1817-1737.82448

Ismailidis, P., Nüesch, C., Kaufmann, M., Clauss, M., Pagenstert, G., Eckardt, A., Ilchmann, T., Mündermann, A., 2020. Measuring gait kinematics in patients with severe hip osteoarthritis using wearable sensors. Gait & Posture 81, 49–55. https://doi.org/10.1016/j.gaitpost.2020.07.004

Jakob, V., Küderle, A., Kluge, F., Klucken, J., Eskofier, B.M., Winkler, J., Winterholler, M., Gassner, H., 2021. Validation of a Sensor-Based Gait Analysis System with a Gold-Standard Motion Capture System in Patients with Parkinson's Disease. Sensors 21, 7680. https://doi.org/10.3390/s21227680

James, K.A., Corrigan, P., Lanois, C., Huang, C.-H., Davis, I.S., Stefanik, J.J., 2023. Association of tibial acceleration during walking to pain and impact loading in adults with knee osteoarthritis. Clinical Biomechanics 109, 106097. https://doi.org/10.1016/j.clinbiomech.2023.106097

Jeon, J., Kwon, S., Lee, Y., Hong, J., Yu, J., Kim, J., Kim, S.-G., Lee, D., 2023. Influence of the Hawthorne effect on spatiotemporal parameters, kinematics, ground reaction force, and the symmetry of the dominant and nondominant lower limbs during gait. Journal of Biomechanics 152, 111555. https://doi.org/10.1016/j.jbiomech.2023.111555

Kanko, R.M., Laende, E., Selbie, W.S., Deluzio, K.J., 2021a. Inter-session repeatability of markerless motion capture gait kinematics. Journal of Biomechanics 121, 110422. https://doi.org/10.1016/j.jbiomech.2021.110422

Kanko, R.M., Laende, E.K., Davis, E.M., Selbie, W.S., Deluzio, K.J., 2021b. Concurrent assessment of gait kinematics using marker-based and markerless motion capture. Journal of Biomechanics 127, 110665. https://doi.org/10.1016/j.jbiomech.2021.110665

Kanko, R.M., Laende, E.K., Strutzenberger, G., Brown, M., Selbie, W.S., DePaul, V., Scott, S.H., Deluzio, K.J., 2021c. Assessment of spatiotemporal gait parameters using a

deep learning algorithm-based markerless motion capture system. Journal of Biomechanics 122, 110414. https://doi.org/10.1016/j.jbiomech.2021.110414

Keller, V.T., Outerleys, J.B., Kanko, R.M., Laende, E.K., Deluzio, K.J., 2022. Clothing condition does not affect meaningful clinical interpretation in markerless motion capture. Journal of Biomechanics 141, 111182. https://doi.org/10.1016/j.jbiomech.2022.111182

Kellgren, J.H., Lawrence, J.S., 1957. Radiological Assessment of Osteo-Arthrosis. Annals of the Rheumatic Diseases 16, 494–502. https://doi.org/10.1136/ard.16.4.494

Kiss, R.M., 2011. Effect of severity of knee osteoarthritis on the variability of gait parameters. Journal of Electromyography and Kinesiology 21, 695–703. https://doi.org/10.1016/j.jelekin.2011.07.011

Kobsar, D., Charlton, J.M., Tse, C.T.F., Esculier, J.-F., Graffos, A., Krowchuk, N.M., Thatcher, D., Hunt, M.A., 2020a. Validity and reliability of wearable inertial sensors in healthy adult walking: a systematic review and meta-analysis. J NeuroEngineering Rehabil 17, 62. https://doi.org/10.1186/s12984-020-00685-3

Kobsar, D., Ferber, R., 2018. Wearable Sensor Data to Track Subject-Specific Movement Patterns Related to Clinical Outcomes Using a Machine Learning Approach. Sensors 18, 2828. https://doi.org/10.3390/s18092828

Kobsar, D., Masood, Z., Khan, H., Khalil, N., Kiwan, M.Y., Ridd, S., Tobis, M., 2020b. Wearable Inertial Sensors for Gait Analysis in Adults with Osteoarthritis—A Scoping Review. Sensors 20, 7143. https://doi.org/10.3390/s20247143

Kobsar, D., Osis, S.T., Boyd, J.E., Hettinga, B.A., Ferber, R., 2017. Wearable sensors to predict improvement following an exercise intervention in patients with knee osteoarthritis. J NeuroEngineering Rehabil 14, 94. https://doi.org/10.1186/s12984-017-0309-z

Kobsar, D., Osis, S.T., Phinyomark, A., Boyd, J.E., Ferber, R., 2016. Reliability of gait analysis using wearable sensors in patients with knee osteoarthritis. Journal of Biomechanics 49, 3977–3982. https://doi.org/10.1016/j.jbiomech.2016.11.047

Kokkotis, C., Moustakidis, S., Papageorgiou, E., Giakas, G., Tsaopoulos, D.E., 2020. Machine learning in knee osteoarthritis: A review. Osteoarthritis and Cartilage Open 2, 100069. https://doi.org/10.1016/j.ocarto.2020.100069

Koo, T.K., Li, M.Y., 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine 15, 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kroenke, K., Strine, T.W., Spitzer, R.L., Williams, J.B.W., Berry, J.T., Mokdad, A.H., 2009. The PHQ-8 as a measure of current depression in the general population. Journal of Affective Disorders 114, 163–173. https://doi.org/10.1016/j.jad.2008.06.026

Kumar, D., Manal, K.T., Rudolph, K.S., 2013. Knee joint loading during gait in healthy controls and individuals with knee osteoarthritis. Osteoarthritis and Cartilage 21, 298–305. https://doi.org/10.1016/j.joca.2012.11.008

Kupperman, N., DeJong, A.F., Alston, P., Hertel, J., Saliba, S.A., 2021. Athlete Workloads During Collegiate Women's Soccer Practice: Implications for Return to Play. Journal of Athletic Training 56, 321–330. https://doi.org/10.4085/205-20

Kuroyanagi, Y., Nagura, T., Kiriyama, Y., Matsumoto, H., Otani, T., Toyama, Y., Suda, Y., 2012. A quantitative assessment of varus thrust in patients with medial knee osteoarthritis. The Knee 19, 130–134. https://doi.org/10.1016/j.knee.2010.12.007

Lafortune, M.A., 1991. Three-dimensional acceleration of the tibia during walking and running. Journal of Biomechanics 24, 877–886. https://doi.org/10.1016/0021-9290(91)90166-K

Lee, J., Shin, S.Y., Ghorpade, G., Akbas, T., Sulzer, J., 2019. Sensitivity comparison of inertial to optical motion capture during gait: implications for tracking recovery, in: 2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR). Presented at the 2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR), IEEE, Toronto, ON, Canada, pp. 139–144. https://doi.org/10.1109/ICORR.2019.8779411

Li, X., Cao, F., Zhao, C., Ge, M., Wei, H., Tang, J., Xu, W., Wang, S., Gao, M., Wang, P., Pan, H., 2024. Global burden of osteoarthritis: Prevalence and temporal trends from 1990 to 2019. Int J of Rheum Dis 27, e15285. https://doi.org/10.1111/1756-185X.15285

Li, Y., Wang, L., 2022. Human Activity Recognition Based on Residual Network and BiLSTM. Sensors 22, 635. https://doi.org/10.3390/s22020635

Mahmoudian, A., Van Dieën, J.H., Bruijn, S.M., Baert, I.A.C., Faber, G.S., Luyten, F.P., Verschueren, S.M.P., 2017. Dynamic and static knee alignment at baseline predict structural abnormalities on MRI associated with medial compartment knee

osteoarthritis after 2 years. Gait & Posture 57, 46–51. https://doi.org/10.1016/j.gaitpost.2017.05.024

March, L., 2016. Osteoarthritis: A Serious Disease: Submitted to the U.S. Food and Drug Administration.

Marcum, Z.A., Zhan, H.L., Perera, S., Moore, C.G., Fitzgerald, G.K., Weiner, D.K., 2014. Correlates of Gait Speed in Advanced Knee Osteoarthritis. Pain Med 15, 1334–1342. https://doi.org/10.1111/pme.12478

Mariani, B., Rouhani, H., Crevoisier, X., Aminian, K., 2013. Quantitative estimation of foot-flat and stance phase of gait using foot-worn inertial sensors. Gait & Posture 37, 229–234. https://doi.org/10.1016/j.gaitpost.2012.07.012

Martindale, C.F., Christlein, V., Klumpp, P., Eskofier, B.M., 2021. Wearables-based multi-task gait and activity segmentation using recurrent neural networks. Neurocomputing 432, 250–261. https://doi.org/10.1016/j.neucom.2020.08.079

McCarthy, I., Hodgins, D., Mor, A., Elbaz, A., Segal, G., 2013. Analysis of knee flexion characteristics and how they alter with the onset of knee osteoarthritis: a case control study. BMC Musculoskelet Disord 14, 169. https://doi.org/10.1186/1471-2474-14-169

Mihy, J.A., Wagatsuma, M., Cain, S.M., Hafer, J.F., 2022. A Functional Sensor-to-Segment Orientation Method to Reduce the Effects of Varied Sensor Placement: Implications for IMU Best Practices. https://doi.org/10.1101/2022.11.29.22282894

Miranda, D.L., Rainbow, M.J., Crisco, J.J., Fleming, B.C., 2013. Kinematic differences between optical motion capture and biplanar videoradiography during a jump–cut

maneuver. Journal of Biomechanics 46, 567–573. https://doi.org/10.1016/j.jbiomech.2012.09.023

Misu, S., Tanaka, S., Ishihara, K., Asai, T., Nishigami, T., 2022. Applied Assessment Method for Varus Thrust during Walking in Patients with Knee Osteoarthritis Using Acceleration Data Measured by an Inertial Measurement Unit. Sensors 22, 6460. https://doi.org/10.3390/s22176460

Miyazaki, T., 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

Moskowitz, R.W., 2009. The burden of osteoarthritis: clinical and quality-of-life issues. Am J Manag Care 15, S223-229.

Mündermann, A., Dyrby, C.O., Andriacchi, T.P., 2005. Secondary gait changes in patients with medial compartment knee osteoarthritis: Increased load at the ankle, knee, and hip during walking. Arthritis Rheum 52, 2835–2844. https://doi.org/10.1002/art.21262

Muro-de-la-Herran, A., Garcia-Zapirain, B., Mendez-Zorrilla, A., 2014. Gait Analysis Methods: An Overview of Wearable and Non-Wearable Systems, Highlighting Clinical Applications. Sensors 14, 3362–3394. https://doi.org/10.3390/s140203362

Nakamoto, M., Otsuka, R., Yuki, A., Nishita, Y., Tange, C., Tomida, M., Kato, Y., Ando, F., Shimokata, H., Suzuki, T., 2015. Higher gait speed and smaller sway area decrease the risk for decline in higher-level functional capacity among middle-aged and

elderly women. Archives of Gerontology and Geriatrics 61, 429–436. https://doi.org/10.1016/j.archger.2015.08.001

Needham, L., Evans, M., Cosker, D.P., Wade, L., McGuigan, P.M., Bilzon, J.L., Colyer, S.L., 2021. The accuracy of several pose estimation methods for 3D joint centre localisation. Sci Rep 11, 20673. https://doi.org/10.1038/s41598-021-00212-x

Nouredanesh, M., Godfrey, A., Howcroft, J., Lemaire, E.D., Tung, J., 2021. Fall risk assessment in the wild: A critical examination of wearable sensor use in free-living conditions. Gait & Posture 85, 178–190. https://doi.org/10.1016/j.gaitpost.2020.04.010

Ornetti, P., Maillefert, J.-F., Laroche, D., Morisset, C., Dougados, M., Gossec, L., 2010. Gait analysis as a quantifiable outcome measure in hip or knee osteoarthritis: A systematic review. Joint Bone Spine 77, 421–425. https://doi.org/10.1016/j.jbspin.2009.12.009

Outerleys, J., Mihic, A., Keller, V., Laende, E., Deluzio, K., 2024a. Markerless motion capture provides repeatable gait outcomes in patients with knee osteoarthritis. Journal of Biomechanics 168, 112115. https://doi.org/10.1016/j.jbiomech.2024.112115

Outerleys, J., Mihic, A., Keller, V., Laende, E., Deluzio, K., 2024b. Markerless motion capture provides repeatable gait outcomes in patients with knee osteoarthritis. Journal of Biomechanics 168, 112115. https://doi.org/10.1016/j.jbiomech.2024.112115

Outerleys, J.B., Dunbar, M.J., Richardson, G., Hubley-Kozey, C.L., Astephen Wilson, J.L., 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, 130–138. https://doi.org/10.1123/jab.2020-0051

Palmer, J.S., Jones, L.D., Monk, A.P., Nevitt, M., Lynch, J., Beard, D.J., Javaid, M.K., Price, A.J., 2020. Varus alignment of the proximal tibia is associated with structural progression in early to moderate varus osteoarthritis of the knee. Knee Surg Sports Traumatol Arthrosc 28, 3279–3286. https://doi.org/10.1007/s00167-019-05840-5

Paravlic, A.H., Meulenberg, C.J., Drole, K., 2022. The Time Course of Quadriceps Strength Recovery After Total Knee Arthroplasty Is Influenced by Body Mass Index, Sex, and Age of Patients: Systematic Review and Meta-Analysis. Front. Med. 9, 865412. https://doi.org/10.3389/fmed.2022.865412

Parry, E., Ogollah, R., Peat, G., 2017. Significant pain variability in persons with, or at high risk of, knee osteoarthritis: preliminary investigation based on secondary analysis of cohort data. BMC Musculoskelet Disord 18, 80. https://doi.org/10.1186/s12891-017-1434-3

Pistacchi, M., 2017. Gait analysis and clinical correlations in early Parkinson's disease. FN 32, 28. https://doi.org/10.11138/FNeur/2017.32.1.028

Pophal Da Silva, L., Boneti Moreira, N., Barbosa De Freitas, P., Pereira, G., Rodacki, A.L.F., 2021. Gait Parameters of Older Adults according to Their Fall History and

Functional Capacity While Walking at Different Speeds. Gerontology 67, 532–543. https://doi.org/10.1159/000513601

Prisco, G., Pirozzi, M.A., Santone, A., Esposito, F., Cesarelli, M., Amato, F., Donisi, L., 2024. Validity of Wearable Inertial Sensors for Gait Analysis: A Systematic Review. Diagnostics 15, 36. https://doi.org/10.3390/diagnostics15010036

Reiss, A., 2012. PAMAP2 Physical Activity Monitoring. https://doi.org/10.24432/C5NW2H

Ren, R., Spaan, J., Jordan, A., Shafiro, A., Su, E.P., 2023. Novel Use of an Accelerometer to Assess Load Asymmetry Over Time After Hip Resurfacing Arthroplasty. The Journal of Arthroplasty 38, S58–S64. https://doi.org/10.1016/j.arth.2023.04.015

Ritter, M.A., Wing, J.T., Berend, M.E., Davis, K.E., Meding, J.B., 2008. The Clinical Effect of Gender on Outcome of Total Knee Arthroplasty. The Journal of Arthroplasty 23, 331–336. https://doi.org/10.1016/j.arth.2007.10.031

Robles-García, V., Corral-Bergantiños, Y., Espinosa, N., Jácome, M.A., García-Sancho, C., Cudeiro, J., Arias, P., 2015. Spatiotemporal Gait Patterns During Overt and Covert Evaluation in Patients With Parkinson's Disease and Healthy Subjects: Is There a Hawthorne Effect? Journal of Applied Biomechanics 31, 189–194. https://doi.org/10.1123/jab.2013-0319

Romijnders, R., Warmerdam, E., Hansen, C., Schmidt, G., Maetzler, W., 2022. A Deep Learning Approach for Gait Event Detection from a Single Shank-Worn IMU: Validation in Healthy and Neurological Cohorts. Sensors 22, 3859. https://doi.org/10.3390/s22103859

Roth, N., Küderle, A., Ullrich, M., Gladow, T., Marxreiter, F., Klucken, J., Eskofier, B.M., Kluge, F., 2021. Hidden Markov Model based stride segmentation on unsupervised free-living gait data in Parkinson's disease patients. J NeuroEngineering Rehabil 18, 93. https://doi.org/10.1186/s12984-021-00883-7

Ruder, M.C., Hunt, M.A., Charlton, J.M., Tse, C.T.F., Kobsar, D., 2022. Validity and reliability of gait metrics derived from researcher-placed and self-placed wearable inertial sensors. Journal of Biomechanics 142, 111263. https://doi.org/10.1016/j.jbiomech.2022.111263

Ruder, M.C., Masood, Z., Kobsar, D., 2023. Reliability of waveforms and gait metrics from multiple outdoor wearable inertial sensors collections in adults with knee osteoarthritis. Journal of Biomechanics 160, 111818. https://doi.org/10.1016/j.jbiomech.2023.111818

Seagers, K., Uhlrich, S.D., Kolesar, J.A., Berkson, M., Kaneda, J.M., Beaupre, G.S., Delp, S.L., 2022. Changes in foot progression angle during gait reduce the knee adduction moment and do not increase hip moments in individuals with knee osteoarthritis. Journal of Biomechanics 141, 111204. https://doi.org/10.1016/j.jbiomech.2022.111204

Seel, T., Raisch, J., Schauer, T., 2014. IMU-Based Joint Angle Measurement for Gait Analysis. Sensors 14, 6891–6909. https://doi.org/10.3390/s140406891

Selles, R.W., Formanoy, M.A.G., Bussmann, J.B.J., Janssens, P.J., Stam, H.J., 2005. Automated estimation of initial and terminal contact timing using accelerometers; development and validation in transtibial amputees and controls. IEEE Trans.

Neural Syst. Rehabil. Eng. 13, 81–88. https://doi.org/10.1109/TNSRE.2004.843176

Serrao, M., Chini, G., Caramanico, G., Bartolo, M., Castiglia, S.F., Ranavolo, A., Conte, C., Venditto, T., Coppola, G., Di Lorenzo, C., Cardinali, P., Pierelli, F., 2019. Prediction of Responsiveness of Gait Variables to Rehabilitation Training in Parkinson's Disease. Front. Neurol. 10, 826. https://doi.org/10.3389/fneur.2019.00826

Sharifi Renani, M., Myers, C.A., Zandie, R., Mahoor, M.H., Davidson, B.S., Clary, C.W., 2020. Deep Learning in Gait Parameter Prediction for OA and TKA Patients Wearing IMU Sensors. Sensors 20, 5553. https://doi.org/10.3390/s20195553

Shema-Shiratzky, S., Hillel, I., Mirelman, A., Regev, K., Hsieh, K.L., Karni, A., Devos, H., Sosnoff, J.J., Hausdorff, J.M., 2020. A wearable sensor identifies alterations in community ambulation in multiple sclerosis: contributors to real-world gait quality and physical activity. J Neurol 267, 1912–1921. https://doi.org/10.1007/s00415-020-09759-7

Shimada, N., Shimada, M., Toriyama, M., Ishikawa, M., Hirata, K., Kono, Y., Ushio, K., Mikami, Y., Adachi, N., 2024. Functional electrostimulation therapy for vastus medialis decreases the varus thrust during gait. J Phys Ther Sci 36, 190–194. https://doi.org/10.1589/jpts.36.196

Shull, P.B., Silder, A., Shultz, R., Dragoo, J.L., Besier, T.F., Delp, S.L., Cutkosky, M.R., 2013. Six-week gait retraining program reduces knee adduction moment, reduces pain, and improves function for individuals with medial compartment knee

osteoarthritis: GAIT RETRAINING REDUCES KNEE OSTEOARTHRITIS PAIN. J. Orthop. Res. 31, 1020–1025. https://doi.org/10.1002/jor.22340

Simon, S.R., 2004. Quantification of human motion: gait analysis—benefits and limitations to its application to clinical problems. Journal of Biomechanics 37, 1869–1880. https://doi.org/10.1016/j.jbiomech.2004.02.047

Solomonow-Avnon, D., Herman, A., Levin, D., Rozen, N., Peled, E., Wolf, A., 2017. Positive outcomes following gait therapy intervention for hip osteoarthritis: A longitudinal study. Journal Orthopaedic Research 35, 2222–2232. https://doi.org/10.1002/jor.23511

Takayanagi, N., Sudo, M., Yamashiro, Y., Lee, S., Kobayashi, Y., Niki, Y., Shimada, H., 2019. Relationship between Daily and In-laboratory Gait Speed among Healthy Community-dwelling Older Adults. Sci Rep 9, 3496. https://doi.org/10.1038/s41598-019-39695-0

Takigami, H., Koshino, T., Saito, T., 2000. An evaluation of the velocity of the lateral thrust in osteoarthritic knees with a varus deformity. The Knee 7, 11–15. https://doi.org/10.1016/S0968-0160(99)00028-9

Tan, J.-S., Tippaya, S., Binnie, T., Davey, P., Napier, K., Caneiro, J.P., Kent, P., Smith, A., O'Sullivan, P., Campbell, A., 2022. Predicting Knee Joint Kinematics from Wearable Sensor Data in People with Knee Osteoarthritis and Clinical Considerations for Future Machine Learning Models. Sensors 22, 446. https://doi.org/10.3390/s22020446

Ullrich, M., Kuderle, A., Hannink, J., Del Din, S., Gassner, H., Marxreiter, F., Klucken, J., Eskofier, B.M., Kluge, F., 2020. Detection of Gait From Continuous Inertial Sensor Data Using Harmonic Frequencies. IEEE J. Biomed. Health Inform. 1–1. https://doi.org/10.1109/JBHI.2020.2975361

Werner, C., Heldmann, P., Hummel, S., Bauknecht, L., Bauer, J.M., Hauer, K., 2020. Concurrent Validity, Test-Retest Reliability, and Sensitivity to Change of a Single Body-Fixed Sensor for Gait Analysis during Rollator-Assisted Walking in Acute Geriatric Patients. Sensors 20, 4866. https://doi.org/10.3390/s20174866

Wiik, A.V., Aqil, A., Brevadt, M., Jones, G., Cobb, J., 2017. Abnormal ground reaction forces lead to a general decline in gait speed in knee osteoarthritis patients. WJO 8, 322. https://doi.org/10.5312/wjo.v8.i4.322

Youn, I.-H., Leutzinger, T., Youn, J.-H., Zeni, J.A., Knarr, B.A., 2020. Self-Reported and Performance-Based Outcome Measures Estimation Using Wearables After Unilateral Total Knee Arthroplasty. Front. Sports Act. Living 2, 569932. https://doi.org/10.3389/fspor.2020.569932

Youn, I.-H., Youn, J.-H., Zeni, J., Knarr, B., 2018. Biomechanical Gait Variable Estimation Using Wearable Sensors after Unilateral Total Knee Arthroplasty. Sensors 18, 1577. https://doi.org/10.3390/s18051577

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

Zijlstra, W., Hof, A.L., 2003. Assessment of spatio-temporal gait parameters from trunk accelerations during human walking. Gait & Posture 18, 1–10. https://doi.org/10.1016/S0966-6362(02)00190-X