Identifying the Trajectory of Normal Recovery Following Knee Arthroplasty via Physical Activity Monitoring

A dissertation submitted to the University of Oxford in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Musculoskeletal Sciences



Scott Richard Small

St Edmund Hall

Hilary Term 2022

Abstract

Current orthopaedic practice lacks low-burden tools for objectively quantifying patient recovery in the first weeks following knee replacement surgery. Postoperative functional tests are burdensome to the clinical care team and unsuited for assessment of at-home recovery, while most patient reported outcome measures are not validated for perioperative use and may not reflect objective improvement in postoperative mobility. The aim of this thesis is to establish a methodology for objectively assessing the trajectory of early postoperative recovery following knee arthroplasty.

A scoping review was conducted to evaluate the use of wearable sensors in knee arthroplasty clinical care, finding wearable-based gait analysis and physical activity monitoring as the most common applications of wearables in this area. This review highlights a significant missed opportunity to leverage physical activity monitoring for longitudinal assessment of patient recovery in the early postoperative period.

Initially, fifty-four healthy volunteers were recruited for 24 hours of free-living activity monitoring. Reduced sampling rate was found to underestimate overall physical activity at both the wrist and hip body placements using the Axivity AX3 accelerometer. Transformations were subsequently generated to enable the direct comparison of accelerometer data collected at 25 and 100 Hz. Ground truth steps of thirty-nine healthy volunteers were recorded during 1 hour of free-living activity. A hybrid machine learning model was developed to count steps from a wrist-based accelerometer. In external validation, this model predicted steps with a 0.8% mean bias and 5.1% mean absolute percent error, compared to >20% error common to

most current wrist and hip-based sensors. When step counts were derived for UK Biobank participants, daily steps and peak cadence were significant factors differentiating participants with lower-limb arthritis and non-arthritic matched controls.

In a clinical study, patients undergoing primary knee arthroplasty were recruited from the Nuffield Orthopaedic Centre. Baseline preoperative physical activity data was collected from 141 patients, while 105 patients contributed up to 42 days of postoperative activity data, with high wear compliance during both preoperative and postoperative monitoring periods. Increase in week-to-week postoperative physical activity was documented, culminating in a median step count and acceleration of 3,983 steps/day and 19.5 m*g*, respectively, during the sixth postoperative week. Significantly increased recovery of physical activity was observed in patients undergoing unicompartmental versus total knee arthroplasty.

Postoperative physical activity monitoring was highly acceptable in the clinical population and demonstrated utility for tracking recovery at the individual level and differentiating outcomes between surgical cohorts. Further implementation of activity monitoring in the clinical orthopaedic setting can be a useful means for assessing early postoperative patient recovery while additionally serving as an objective clinical trial outcome measure.

Dedication

To Katie: my love, my partner, and the bedrock of our family.

Acknowledgements

I would like to express my deepest gratitude to my three supervisors: Professor Andrew Price, Dr. Sara Khalid, and Associate Professor Aiden Doherty. To Andrew, for taking chance on me from the start, for your enthusiastic support of this project, and for your continual guidance toward conducting high quality clinical research. To Sara, for your supervision and partnership in this work beginning a year before this project got off the ground. And to Aiden, for welcoming me into your research group, for your kind encouragement and mentorship, and for helping me plot the course of this work through COVID lockdowns and interruptions.

This work would not have been possible without the Clarendon Scholarship Fund. Thank you for supporting my studies and my family during my time at Oxford. An additional thank you to the Medical Sciences Division and the COVID-19 Scholarship Extensions Fund for further supporting this last year of my work.

I am deeply grateful for my examiners during this course of study. To Professor Dominic Furniss and Professor Scott Strath, thank you for taking time out of your full schedules to examine this body of work. Thank you to Professor David Beard for your work as a transfer of status examiner, and Dr. Chris Hinds for serving as a confirmation examiner for this study.

Thank you to the POAC clinical team at the Nuffield Orthopaedic Centre for welcoming me into the team and for your support as I recruited patients in the clinic. A special thank you to the research nurses for showing me the ins-and-outs of clinical research, most specifically Gail Lang, who has been an invaluable source of wisdom and encouragement.

Thank you to my manuscript co-authors, Mr. Dan Jackson, Dr. Shing Chan, Dr. Paula Dhiman and Dr. Marialena Trivella for your help, advice, and training while conducting research and publishing manuscripts from the work in Chapters 2 and 3 of this thesis. An additional thank you to Dr. Garrett Bullock for serving as my second reviewer for the systematic scoping review, and to Lennart von Fritsch for your many hours spent watching participant videos and manually annotating step counts as the second annotator of ground truth labelling in Chapter 4.

I owe a deep debt to the 226 individuals who volunteered as study participants to support this thesis; participants who wore monitors, answered questionnaires, and participated in discussion groups in selfless benefit of this work. Thank you for sharing of yourselves to advance this research.

A special thank you to the Headington Baptist Church community, who welcomed our family with open arms and quickly made us feel at home in Oxford.

I am deeply grateful to my mentor and friend, Dr. Michael Berend, who introduced me to orthopaedics and clinical research in 2006. You were the biggest advocate of our work together in 10 years at the JRSI Foundation and have been an even greater support to my family. Thank you for your encouragement to undertake this PhD journey and thank you for the endless kindness and generosity you have shown us through the years.

Most importantly, thank you my family for your boundless love and support. To my parents and my in-laws, thank you for encouragement, even while our time in Oxford has kept your grandchildren further away than you would have preferred. Thank you, Dad, for your willingness to read through and proofread this thesis. To my children, you are incredible, resilient world travelers, and I am so proud of the young people you are becoming. And to my wife, Katie, your dedication and selflessness are an inspiration. You have worked harder and sacrificed more than anyone to keep us going and to make this Oxford dream a reality. Thank you for being by my side through it all.

"*Whatever you do, in word or in deed, do all in the name of the Lord Jesus, giving thanks to God the Father, through him.*" Colossians 3:17 (World English Bible)

COVID-19 Pandemic Interruptions

Patient recruitment for the primary clinical study took place from February 2020 to July 2021. During this time, emergency redistribution of healthcare resources led to a reduction in hospital capacity for non-urgent knee arthroplasty and periods of time when all non-essential research activities were stopped within the Oxford University Hospitals Trust. As a result, recruitment for the primary clinical study of this thesis was necessarily paused from 17 March to 27 July 2020 and from 28 December 2020 to 22 March 2021. Reduced clinical capacity for knee arthroplasty patients limited recruitment opportunities extending into May 2021. Personal interruptions included caring responsibilities for three primary school aged children during at-home schooling from 18 March 2020 to 20 July 2020 and from 5 January 2021 to 5 March 2021.

Abbreviations

ASTP	Active-to-Sedentary Transition Probability
BMI	Body Mass Index
CCI	Charslon Comorbidity Index
CHAMPS	Community Health Activities Model Program for Seniors
CI	Confidence Interval
ENMO	Euclidean Norm Minus One
EQ-5D	Euroqol 5 Dimension, 3 Level Index Score
GPPAQ	General Practice Physical Activity Questionnaire
HES	Hospital Episode Statistics
HFEN	Highpass Filtered Euclidean Norm
ICD	International Classification of Diseases
IMU	Inertial Measurement Unit
IPAQ	International Physical Activity Questionnaire
IQR	Interquartile Range
KSS	Knee Society Score
LFENMO	Lowpass Filtered Euclidean Norm Minus One
LOESS	Locally Estimated Scatterplot Smoothing
m*g*	Milli-Gravitational unit (0.00981 meters per second^2)
MAD	Mean Amplitude Deviation
MAPE	Mean Absolute Percent Error

MUA	Manipulation Under Anaesthesia
MVPA	Moderate-to-Vigorous Physical Activity
NHS	National Health Service
NOC	Nuffield Orthopaedic Centre
OKS	Oxford Knee Score
OPCS	Office of Population Censuses and Surveys
POAC	Preoperative Assessment Clinic
PROMs	Patient Reported Outcome Measures
ProPASS	Prospective Physical Activity, Sitting and Sleep
PPI	Patient and Public Involvement
RMSE	Root Mean Square Error
SD	Standard deviation
TKA	Total Knee Arthroplasty
UCLA	University of California Los Angeles
UKA	Unicompartmental Knee Arthroplasty
VAS	Euroqol Visual Analogue Score
WOMAC	Western Ontario and McMaster Universities Osteoarthritis Index

Contents

Chapter 1 : Introduction 1

1.1 Thesis Purpose 1

1.2 Knee Arthroplasty Overview 1

1.3 Patient Reported Outcome Measures 2

1.4 Functional Assessment 4

1.5 Wearables for Functional Analysis 6

1.6 Physical Activity Questionnaires 7

1.7 Wearables for Physical Activity Monitoring 9

1.7.1 Body Placement 11

1.7.2 Device Selection 13

1.7.3 Sampling Rate 15

1.7.4 Wear time 16

1.7.5 Monitoring Period 16

1.7.6 Data Analysis 17

1.8 Thesis Scope and Structure 20

1.9 Contributions 22

Chapter 2 : Wearables for the Assessment of Outcome Following Knee Arthroplasty: A Systematic Scoping Review 23

2.1 Copyright Notice 23

2.2 Introduction 23

2.3 Methods 24
2.4 Results 28
2.4.1 Wearables for Functional Assessment 30
2.4.2 Wearables for Objective Physical Activity Measurement 35
2.4.3 Wearables for Joint Instability 40
2.5 Discussion 41
2.6 Conclusion 47
Chapter 3 : Assessment of Free-Living Activities of Daily Living using Wearables: The Role of Sampling Rate on Valid Interstudy Comparison 49
3.1 Copyright Notice 49
3.2 Introduction 49
3.3 Methods 51
3.3.1 Ethical Approval 51
3.3.2 Study Design 51
3.3.3 Primary Analysis with the Biobank Accelerometer Analysis Tool 53
3.3.4 Secondary Analysis with the GGIR R Package 55
3.3.5 Machine Learning Activity Classification 56
3.3.6 Statistical Analysis 57
3.4 Results 58
3.4.1 Participant Demographics 58
3.4.2 24-Hour Physical Activity Assessment 59
3.4.3 Intensity Measurement of Self-Reported Activities 64

3.4.4 Machine Learning Activity Classification 64

3.4.5 Linear and Nearest Neighbour Resampling Methods..... 67

3.4.6 Data Conversion for Interstudy Comparison 69

3.5 Discussion 73

3.5.1 Strengths and Limitations..... 76

3.6 Conclusion..... 77

Chapter 4 : Development of a Hybrid Machine Learning Peak Detection Model for Step Counting of Wrist-based Accelerometry Datasets..... 79

4.1 Introduction..... 79

4.2 Methods..... 82

4.2.1 Datasets..... 82

4.2.2 Step Detection Methods..... 91

4.2.3 External Model Validation..... 95

4.2.4 Sample Rate Comparison and Adjustment 96

4.3 Results..... 97

4.3.1 Participant Demographics and Ground Truth..... 97

4.3.2 Hybrid Model Refinement..... 98

4.3.3 Internal Model Validation..... 100

4.3.4 External Validation and Final Model Selection 105

4.3.5 Model Implementation into the Capture 24 Dataset 112

4.3.6 Sample Rate Comparison and Adjustment 114

4.4 Discussion 116

4.4.1 Strengths and Limitations 120

4.5 Conclusions 121

Chapter 5 : Introduction of Step Count as an Activity Metric in the UK Biobank: Difference in Physical Activity between Healthy, End-Stage Arthritic, and Postoperative Hip and Knee Arthroplasty Patients 123

5.1 Introduction 123

5.2 Methods 125

5.2.1 Standard Accelerometer Data Processing 125

5.2.2 Implementation of Step Detection into the UK Biobank 127

5.2.3 Participant Cohorts Based on Clinical Status 130

5.2.4 Statistical Analysis 134

5.3 Results 135

5.3.1 Physical Activity in the Full UK Biobank Physical Activity Cohort 135

5.3.2 Activity Assessment of Knee and Hip Arthroplasty Patients 141

5.4 Discussion 150

5.4.1 Strengths and Limitations 154

5.5 Conclusions 155

Chapter 6 : Implementation of Physical Activity Monitoring in Clinical Orthopaedic Practice: Study Setup, Preoperative Analysis, Participant Compliance and Feedback 158

6.1 Introduction 158

6.2 Methods 158

6.2.1 Ethical Review 158
6.2.2 Recruitment and Enrolment 159
6.2.3 Baseline Data Collection 160
6.2.4 Postoperative Data Collection 162
6.2.5 Data Processing and Analysis 165
6.2.6 Patient and Public Involvement Informing Future Clinical Direction 168
6.3 Results 169
6.3.1 Recruitment and Enrolment 169
6.3.2 Participant Demographics 170
6.3.3 Accelerometer Wear Time Compliance 171
6.3.4 Baseline Physical Activity and PROMs 173
6.3.5 Regression Analysis of Preoperative Physical Activity 177
6.3.6 Patient and Public Involvement Informing Future Clinical Deployment 179
6.4 Discussion 181
6.4.1 Strengths and Limitations 185
6.5 Conclusions 186
Chapter 7 : Assessing Knee Arthroplasty Postoperative Recovery Through Physical Activity Monitoring 187
7.1 Introduction 187
7.2 Methods 187

7.2.1 Missing Data 187

7.2.2 Statistical Analysis 189

7.3 Results 191

7.3.2 Physical Activity Imputation Performance Analysis 193

7.3.3 Wearable Physical Activity Measurement for Postoperative Monitoring: Assessment of the Trajectory in Recovery Following Knee Arthroplasty 195

7.3.4 Wearable Physical Activity Monitoring for the Comparison of Unicompartmental and Total Knee Arthroplasty Patients 219

7.3.5 Case Studies 225

7.4 Discussion 231

7.5 Conclusion 235

Chapter 8 : Summary and Conclusions 236

8.1 Summary of Findings 237

8.2 Study Limitations 241

8.3 Clinical Implementation and Future Directions 242

8.3.1 Near-term Next Steps 243

8.3.2 Clinical Implementation 243

8.3.3 Other Considerations 246

8.3.4 Broadened Clinical Application 247

8.4 Conclusions 248

Appendix 249

References 268

Figures

Figure 2-1: PRISMA literature review flow diagram 29

Figure 2-2: The number of publications within this review by publication year (*2019 through June). 30

Figure 3-1: Diagram of accelerometer placement and flow chart of participants with valid wear time at each body location. (Created with BioRender) 52

Figure 3-2: Schematic of primary data analysis to inform conversion of accelerometer data collected at 25 Hz to enable comparison with studies collected at 100 Hz using linear and nearest neighbour resampling interpolation methods. 54

Figure 3-3: Schematic of secondary activity metric data analysis with the GGIR R package to enable comparison of 25 Hz and 100 Hz collected data in 5 s epochs and in alternate activity metrics. 56

Figure 3-4: Plot of mean (95%CI) 24-hour activity in 1-hour epochs as measured by ENMO [GGIR] with linear interpolation resampling 60

Figure 3-5: Bland-Altman plots comparing the mean 24-hour ENMO and differences between the 25 Hz and 100 Hz sensors at the wrist and hip 62

Figure 3-6: Bland-Altman plots comparing the mean 24-hour ENMO and differences between the downsampled 25 Hz acceleration and 100 Hz raw acceleration from sensors recording at 100 Hz at the wrist and hip 63

Figure 3-7: Bland-Altman plots comparing the mean 24-hour ENMO and differences between the 25 Hz and 100 Hz sensors at the wrist 68

Figure 3-8: Bland-Altman plots comparing the mean 24-hour ENMO and differences between the 25 Hz and 100 Hz sensors at the hip 69

Figure 4-1: Two Axivity AX3 accelerometers fitted onto a participant's dominant wrist, two clipped at the dominant-sided hip, and a foot-facing action camera clipped at the participant's waistline. 83
Figure 4-2: Screenshot of step annotation with the Elan software package, showing the foot-facing ground truth video image 86
Figure 4-3: Flowchart for the development and tuning of the hybrid machine learning and peak detection step counting algorithm. 93
Figure 4-4: Schematic of the process for generating step count from raw accelerometer data using the final hybrid step counting model 95
Figure 4-5: Bland-Altman plot of ground truth step counts as annotated by the two annotators in 33 participants. 98
Figure 4-6: Heatmap of step counting error across iterated peak detection prominence and distance parameters using the 5 second epoch, 2 steps per epoch walking model. 99
Figure 4-7: Bland-Altman plots for the comparison of step counting models in the primary, internal validation dataset 104
Figure 4-8: Bland-Altman plots for the comparison of step counting models in the Clemson external validation dataset 111
Figure 4-9: Scatterplot of overall daily acceleration versus daily step count in the Capture 24 dataset. 114
Figure 4-10: Bland-Altman plots comparing mean 24 hour step count in 100 Hz and 25 Hz data 115
Figure 5-1: Measures of bulk physical activity in the UK Biobank Physical Activity Cohort, summarised by age and sex 138

Figure 5-2: Measures of activity intensity in terms of step cadence in the UK Biobank Physical Activity Cohort, summarised by self-reported walking pace and sex 139
Figure 5-3: Measures of activity intensity in terms of step cadence in the UK Biobank Physical Activity Cohort, summarised by age and sex 140
Figure 5-4: Study flowchart mapping cohort selection from the UK Biobank Physical Activity Cohort. 142
Figure 6-1: Participant recruitment and interaction flow chart 159
Figure 6-2: Recruitment and data collection flow chart 169
Figure 6-3: Histogram of postoperative wear time in 105 participants 172
Figure 6-4: Preoperative correlation matrix using Spearman's Rank Test including objectively measured activity parameters, patient reported outcome measures, and self-reported activity questionnaires collected during the baseline monitoring period 176
Figure 7-1: Quantile regression of daily postoperative [0-10] pain scores 193
Figure 7-2: Quantile regression plot of increasing daily step count during the six weeks following knee arthroplasty 199
Figure 7-3: Quantile regression plot of increasing daily acceleration during the six weeks following knee arthroplasty 200
Figure 7-4: Quantile regression plot of increasing peak cadence during the six weeks following knee arthroplasty 202
Figure 7-5: Quantile regression plot of decreasing activity fragmentation during the six weeks following knee arthroplasty 203
Figure 7-6: Quantile regression plot of increasing light activity during the six weeks following knee arthroplasty 205

Figure 7-7: Quantile regression plot of increasing MVPA during the six weeks following knee arthroplasty 206
Figure 7-8: Quantile regression plot of decreasing sedentary activity during the six weeks following knee arthroplasty 207
Figure 7-9: Quantile regression plot of sleep during the six weeks following knee arthroplasty 208
Figure 7-10: Postoperative correlation matrix 210
Figure 7-11: Correlation plots comparing Spearman's rank correlation coefficient of activity and pain measurements during postoperative Weeks 1 and 6. 211
Figure 7-12: Correlation plots comparing Spearman's rank correlation coefficient of activity and pain measurements during postoperative Weeks 2 and 6. 212
Figure 7-13: Boxplot of daily step count by operative intervention compared with preoperative baseline in participants contributing postoperative data. 221
Figure 7-14: Boxplot of acceleration by operative intervention compared with preoperative baseline in all participants contributing postoperative data. 221
Figure 7-15: Case Study Participant 1: Individual daily step counts 226
Figure 7-16: Case Study Participant 1: Individual daily acceleration 227
Figure 7-17: Case Study Participant 2: Individual daily step counts 229
Figure 7-18: Case Study Participant 2: Individual daily activity fragmentation . 230

Tables

Table 2-1: Literature Review Inclusion and Exclusion Criteria 25

Table 2-2: Database Search Strategy 26

Table 2-3: Population and Methods Summary of Functional Assessment Studies 33

Table 2-4: Sensor Summary for Functional Assessment Studies 34

Table 2-5: Population and Methods Summary for Physical Activity Studies 37

Table 2-6: Technology Summary for Physical Activity Studies 39

Table 2-7: Population and Technology Summary for Instability Studies 41

Table 3-1: Participant Demographics 59

Table 3-2: Accelerometer Measured Physical Activity for Overall and Diary-Reported Activities Over a Single 24-Hour Period (30 Second Epoch) 61

Table 3-3: Secondary Analysis of Alternative Physical Activity Metrics as Calculated on Wrist Sensors Over 5 Second Epochs in the GGIR R Package .. 62

Table 3-4: Machine Learned Activity Classification for Wrist Accelerometers in a Convenience Sample of 51 Adults Over a Single 24-Hour Period (Willetts Model) 65

Table 3-5: Machine Learned Activity Classification for Wrist Accelerometers in a Convenience Sample of 51 Adults Over a Single 24-Hour Period (Walmsley Model) 65

Table 3-6: Comparison of Rotation-Invariant Machine Learning Features Between 100 Hz and 25 Hz Data Collection 66

Table 3-7: Linear Regression for Comparison of 25 Hz and 100 Hz 24-hour AX3 Accelerometer Data 71

Table 3-8: Linear Regression to Convert Wrist-Based Machine Learning Activity Classification Collected at 25 Hz for Comparison to Data Collected at 100 Hz . 72
Table 4-1: Walking Annotation Defined as Number of Steps Per Epoch 87
Table 4-2: Summary of Datasets 90
Table 4-3: Participant Demographics for Step Detection 98
Table 4-4: Results of Step Count Parameter Iteration Including Peak Detection Settings and Associated Minimum Step Count Error Across Held Out Test Data 99
Table 4-5: Results of Step Count Parameter Iteration Including Peak Detection Settings and Associated Minimum Step Count Error Across Held Out Test Data, Using a 5 Hz Lowpass Filtered Timeseries Signal 100
Table 4-6: Overall Step Counts as Calculated Through Moving Average Peak Detection with Open Movement Software in Comparison with Ground Truth .. 101
Table 4-7: Overall Step Counts as Calculated Through Random Forest Regression in Comparison with Ground Truth 101
Table 4-8: Walking Classification Results and Performance Metrics Based on Random Forest Classification Model 102
Table 4-9: Step Count Results and Performance Metrics Based on Random Forest Classification and Peak Detection Step Counting 103
Table 4-10: Side-by-Side Comparison on Best Performing Models for Each Step Detection Method During Internal Validation with 100 Hz Wrist-based Acceleration 105
Table 4-11: Random Forest Regression Step Counting in Clemson Shimmer3 Dataset Participants Over All Combined Gait Subtypes 106

Table 4-12: Random Forest Regression Step Counting in Clemson Shimmer3 Dataset During Three Different Gait Subtypes 107
Table 4-13: Step Counts in Clemson Shimmer3 Dataset Participants as Predicted by Hybrid Walking Classification and Step Counting Algorithm Combined Over All Gait Subtypes 108
Table 4-14: Hybrid Step Counting Models Applied to Clemson Shimmer3 Dataset During Three Different Gait Subtypes 109
Table 4-15: Comparison of Candidate Final Models for During External Validation in the Clemson Shimmer3 Dataset Participants Including All Gait Subtypes ... 110
Table 4-16: Confusion Matrix of Predicted Walking and Non-Walking Epochs Compared to the Willetts Ground Truth Annotations in the Capture24 Dataset 112
Table 4-17: Predicted Walking and Non-Walking Epochs Compared to the Doherty Ground Truth Annotations 113
Table 4-18: Predicted Walking and Non-Walking Epochs Compared to the Binary Annotation Categories of Potential Walking and Likely Not Walking 113
Table 4-19: 24-Hour Step Counts in Participants Wearing AX3 Accelerometers on the Dominant Wrist, Recording at Both 25 Hz and 100 Hz 115
Table 5-1: OPCS-4 Standalone Codes of Operation and Location 131
Table 5-2: OPCS-4 Combination Codes of Operation and Location 132
Table 5-3: Demographics of Full UK Biobank Physical Activity Cohort 136
Table 5-4: Overall Acceleration and Step Count Activity Metrics Across All UK Biobank Participants Aged 45-79 Years with Valid Accelerometer Wear Time 137
Table 5-5: Peak Cadence as Calculated Across All UK Biobank Participants Aged 45-79 Years with Valid Accelerometer Wear Time 140

Table 5-6: Demographics of Knee Arthroplasty Cohorts with Matched Controls 143
Table 5-7: Demographics of Hip Arthroplasty Cohorts with Matched Controls 144
Table 5-8: Physical Activity Metrics of UK Biobank Participants Wearing an Accelerometer Less Than 12 Months Prior to Unilateral Primary Knee Arthroplasty, Compared to Non-Arthritic Controls 146
Table 5-9: Physical Activity Metrics of UK Biobank Participants Wearing an Accelerometer Greater Than 12 Months After Unilateral Primary Knee Arthroplasty, Compared to Non-Arthritic Controls 146
Table 5-10: Physical Activity Metrics of UK Biobank Participants Wearing an Accelerometer Less Than 12 Months Prior to Unilateral Primary Hip Arthroplasty, Compared to Non-Arthritic Controls 148
Table 5-11: Physical Activity Metrics of UK Biobank Participants Wearing an Accelerometer Greater Than 12 Months After Unilateral Primary Hip Arthroplasty, Compared to Non-Arthritic Controls 148
Table 5-12: Results of Linear Regressions Comparing End-Stage Arthritis and Postoperative Knee Arthroplasty Cohorts 149
Table 5-13: Results of Linear Regressions Comparing End-Stage Arthritis and Postoperative Hip Arthroplasty Cohorts 150
Table 6-1: Preoperative Participant Demographics 170
Table 6-2: Dropout Multivariable Logistic Regression 173
Table 6-3: Baseline Patient Reported Outcome Measures and Self-Reported Activity 174
Table 6-4: Baseline Objectively Measured Physical Activity 174
Table 6-5: Preoperative Step Count Regression Table 177

Table 6-6: Preoperative Acceleration Regression Table 178
Table 6-7: Questions and Answer Options for Patient and Public Involvement Event 180
Table 7-1: Postoperative Cohort Clinical Characteristics 191
Table 7-2: Postoperative (Week 6) Self-Reported Activity and Patient Reported Outcome Measures 192
Table 7-3: Physical Activity Imputation Methods Performance Analysis 194
Table 7-4: Physical Activity Metrics in the Full Valid Postoperative Cohort (n=95) 196
Table 7-5: Table of Wilcoxon Rank Test P-Values of Week-to-Week Change in Activity Metrics Across the Full Valid Postoperative Cohort (n=95) 197
Table 7-6: Regression Analysis of Postoperative Step Count without Including Preoperative Baseline Step Count as a Predictor 214
Table 7-7: Regression Analysis of Postoperative Step Count Including Preoperative Baseline Step Count as a Predictor 215
Table 7-8: Regression Analysis of Postoperative Acceleration without Including Preoperative Baseline Acceleration as a Predictor 217
Table 7-9: Regression Analysis of Postoperative Acceleration Including Preoperative Baseline Acceleration as a Predictor 218
Table 7-10: Comparison of Physical Activity Between UKA and TKA Patients with Valid Wear During Postoperative Recovery (Baseline – Week 3) 222
Table 7-11: Comparison of Physical Activity Between UKA and TKA Patients with Valid Wear During Postoperative Recovery (Week 4 – Week 6) 223
Table 7-12: Wilcoxon Rank Test P-Values of Week-to-Week Change in Activity Metrics 224

Chapter 1: Introduction

1.1 Thesis Purpose

The purpose of this thesis is to develop methodology for, and to evaluate the utility of wearable sensors for the evaluation of early recovery following knee arthroplasty. The overarching aims of this thesis are:

1) To establish methodology for the effective implementation of wearable-based data collection in the orthopaedic clinical pathway
2) To evaluate the ability for wearable-derived metrics to differentiate functional ability in knee arthritis and knee arthroplasty clinical populations
3) To ascertain if wearable-based patient monitoring can differentiate early trends in postoperative recovery

The hypothesis for this study is that wearable sensors will be well tolerated by knee arthroplasty patients and that the patterns of activity derived from continuous longitudinal monitoring will provide a useful metric by which early postoperative functional recovery can be assessed.

1.2 Knee Arthroplasty Overview

Knee arthroplasty is a surgical intervention for the treatment of end-stage loss of the articular cartilage at the tibiofemoral and patellofemoral joints. Osteoarthritis of the knee is the most common site for symptomatic osteoarthritis with a population-based incidence of 10-15 cases per 1,000 person years in people over the age of sixty-five[1]. The risk factors of knee osteoarthritis are multifactorial, with obesity, age, female sex, and prior knee injury the primary drivers of cartilage degradation necessitating surgical intervention and joint replacement[1,2]. Key symptoms of end-stage arthritis include severe joint pain, joint stiffness, and

reduced or impaired mobility. There is no established threshold of symptoms to indicate surgical intervention, thus the decision to perform a knee replacement is based on shared decision making between a patient and their surgeon after discussion of the risks, benefits, and alternative treatment options[3].

Total knee arthroplasty (TKA) is the most common procedure for primary joint replacement for at the knee[4]. During TKA the entirety of both the tibiofemoral and patellofemoral articular surfaces are resected and replaced with prosthetic components. The operation is typically performed as an inpatient procedure with patients staying in hospital for one to four days for monitoring and postoperative evaluation prior to discharge home[5]. TKA is a highly successful intervention with good long-term outcomes and implant survivorship across many device designs[6,7]. When only a portion of the articular cartilage of the knee is affected by osteoarthritic damage, unicompartmental knee arthroplasty (UKA) can be performed to replace the affected articular surface while leaving intact the undamaged cartilage. UKA is a less invasive surgery associated with lower morbidity, increased function, shorter hospital stays and improved recovery when compared to TKA[5,8].

1.3 Patient Reported Outcome Measures

Historically, outcome metrics for joint arthroplasty have focused on implant survival and morbidity[7,9]. In a move to more fully understand tangible postoperative benefit from the patient perspective, patient reported outcome measures (PROMs) have been developed. PROMs are commonly tuned for specific joints or procedures and can measure a variety of different domains, including health-related quality of life, pain, joint function, perceived activity, and participation in activities of daily living[10]. A wide range of PROMs have been

developed to assess outcomes following knee arthroplasty[11,12]. The Knee Society Score (KSS), Oxford Knee Score (OKS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), and EQ-5D are the four most commonly used PROMs for knee arthroplasty in European orthopaedic practice[12]. Harris et al. found the OKS and WOMAC to be the two of the best performing PROMs in terms of validity and reliability[13]. These self-reported questionnaires are important instruments in clinical orthopaedics to quantify change in patient experience and function following joint arthroplasty. PROMs can be useful for measuring improvements in function and pain from baseline, but have inherent limitations based on sensitivity levels, minimal clinically important differences, ceiling and floor effects, non-responder bias, and the lack of a clear and consistent definition of what qualifies as a successful knee intervention[14,15]. While PROMs frequently emphasise the quantification of pain and functional deficit, few elucidate a patient's return to their expected participation in preferred activities of daily living[16]. Further, in many cases, PROMs only moderately correlate with patient satisfaction following orthopaedic surgical intervention[17]. If true effectiveness of surgical intervention is to be judged based on a synthesis of the patient experience in terms of pain, function, and overall satisfaction, PROMs provide a starting point, though perhaps not an endpoint for thorough assessment of postoperative patient outcomes.

Most PROMs are designed to be administered preoperatively, and then again at six or more months postoperatively to quantify end-point patient recovery. In a systematic review, Strickland et al[18] found no patient reported outcome measures meeting quality assessment for valid evaluation of early near-term postoperative recovery after hip or knee arthroplasty. The authors noted the need for tools

developed within an orthopaedic population to serve as a gold standard for assessing interventions and recovery pathways in the early postoperative period. Subsequently, Strickland et al. recently published a first-of-its-kind PROM validated for assessing functional recovery of lower limb arthroplasty patients within the first six postoperative weeks[5,19]. This new recovery-based patient reported outcome measure includes pain, fatigue/sleep, nausea, and functional mobility domains. Recent development of outcome measures focused on the early postoperative period highlight the importance and continued interest in understanding patient recovery in the first few postoperative weeks as a means for early feedback on the pathway for long-term functional rehabilitation.

1.4 Functional Assessment

Following knee arthroplasty, a wide range of tests and protocols are deployed to assess pre- and postoperative physical status, including evaluation of walking speed, endurance, stair negotiation, and sit-to-stand function[20]. Contrary to self-reported questionnaires and PROMs identifying patient-perceived function and physical abilities, functional assessments are designed to characterise a patient's true capacity for task performance. Functional tests are routinely performed by a physiotherapist to ensure adequate mobility and independence prior to postoperative hospital discharge. The Osteoarthritis Research Society International recommends five primary functional tests for perioperative assessment of knee and hip arthroplasty patients[20]. The timed up and go test is performed to assess a patient's ability to get in and out of a chair. In this test, the patient is directed to stand from a chair, walk three meters, turn around, walk back to the chair, and sit down again. The amount of time taken to perform this task allows for quantification of function and longitudinal monitoring of change in

function over time[21]. The 30 second chair stand test quantifies the number of times a patient can completely rise from a chair and then sit back down in 30 seconds, with an increasing number of cycles correlating with higher sit-to-stand function[20]. Stair negotiation is typically assessed by a timed four, nine, or twelve stairstep test[20]. For walking speed assessment, the 40 meter fast paced walk test is conducted wherein patients walk up and down a 10 meter walkway four times as fast as possible. Walking endurance is quantified with the six minute walk test, wherein patients walk as far as possible over a flat walkway for six minutes[22].

Knee range of motion is an additional important functional assessment prior to and following knee arthroplasty. Alongside pain relief, adequate improvement in knee range of motion may be closely related to long-term patient satisfaction[23]. Knee stiffness is common after surgery and inadequate knee range of motion can limit functional mobility in the short- and long-term following knee arthroplasty. Both active and passive knee range of motion are routinely collected during clinical evaluation and typically increase steadily within the first eight postoperative weeks[24]. Knee range of motion is generally measured by a trained clinician using a goniometer, with a flexion angle of greater than 110 degrees typically required for adequate performance of common activities of daily living[25].

Standardised functional tests provide a valuable tool for the assessment of clinical populations and are broadly adopted in patient evaluation during perioperative care[20]. These assessments, however, are burdensome to the patient, require interaction with a caregiver or clinician, and are not able to be routinely embedded into offsite patient evaluation or monitoring. While they are designed to replicate common activities of daily living, laboratory-based gait may significantly differ from a patient's gait in the free-living environment[26]. Therefore, while useful for onsite

clinical assessment of function, these tests alone cannot fully describe the degree in which orthopaedic patients are able to functionally perform within the free-living environment or how that performance changes on a short-term basis during early postoperative recovery.

1.5 Wearables for Functional Analysis

Optical motion capture systems are the gold standard for functional gait analysis. These high accuracy systems provide detailed data regarding joint kinetics and kinematics and can be used to assess nuanced changes in whole body or joint-specific motion following a surgical intervention[27]. Optical motion capture systems require specialised equipment, laboratory space, and dedicated staff for implementation into a clinical care setting. For this reason, broad clinical utilisation of motion capture gait analysis is typically reserved for paediatric evaluation of conditions such as spastic cerebral palsy and clubfoot[28,29]. In an effort to increase the accessibility of gait analysis for broader clinical usage, and to enable gait analysis in the free-living environment without the need for dedicated facilities, methods for clinical functional analysis using wearable sensors have rapidly developed over the past decade[30]. Recent studies have adopted inertial measurement unit (IMU)-based functional analysis for a wide range of applications, though most frequently for the assessment of upper and lower limb mobility[30]. In a recent review, Kobsar et al.[31] described 72 papers implementing wearable sensors for the assessment of patients with any type of osteoarthritis with focus primarily on gait-specific spatial-temporal parameters, joint moments, and joint angles[31]. In a review of physical function evaluation using wearable sensors, Granat et al.[32] described numerous studies implementing IMUs to collect functional data in elderly patients during a wide range of activities and functional

tests similar to those described in Section 1.4. While IMU-assessments allow for functional assessment outside of laboratory-based settings, they require expert setup and typically rely of the use of two to four sensors spread across a number of potential body locations[31]. Functional analysis using IMUs provide additional, more granular data when used while performing standard functional assessments, but the clinical utility of that data in comparison with standard functional tests has yet to be fully demonstrated[33]. Longitudinal analysis of changes in function as a result of disease progression or postoperative recovery remain absent in the literature[31]. This may be due to the necessary expert setup of multiple sensors alongside the relatively high battery and storage requirements of IMUs, limiting the convenience of repeated or long-term measurement. A scoping review of the current clinical use of wearable-based functional analysis in perioperative knee arthroplasty patients will be reported in Chapter 2.

1.6 Physical Activity Questionnaires

Historically, physical activity research has been conducted via validated self-reported questionnaires with physical activity guidelines and recommendations based on these studies of self-reported activity[34]. The World Health Organization publishes annual physical activity guidelines for health and welfare and currently recommends at least 150 minutes of moderate-to-vigorous physical activity per day in adults and 60 minutes of active play per day in children over the age of five[35]. The use of self-reported activity questionnaires is inexpensive and relatively simple to implement into a study protocol, and is thus frequently used as a functional outcome measure. Several self-reported physical activity questionnaires have been developed, with the International Physical Activity Questionnaire (IPAQ)[36] and the University of California Los Angeles

(UCLA) Physical Activity Questionnaire[37] being the two of the most commonly utilised[38]. The IPAQ is a validated questionnaire with short and long versions that aims to determine the amount of time the participant spends across sedentary, low, moderate, and high intensity activities over the previous week. The UCLA physical activity questionnaire is a simple form with a single question posted to the study participant, to "Check one box that best describes current activity level", with responses ranging from 1 to 10 representing a scale from a person who is completely inactive up to a person who is a frequent participant in impact sports[37]. The UK National Health Services (NHS) currently utilises its own validated physical activity questionnaire, the General Practice Physical Activity Questionnaire (GPPAQ) which is aimed at creating an index for reference to clinicians implementing intervention to increase patient activity levels.

Due in part to their ease of implementation, self-reported physical activity questionnaires are frequently used in assessment of lower limb arthroplasty[39]. Several prior systematic reviews have found that the UCLA activity score has been used extensively to evaluated physical activity in joint replacement populations[39–45]. As a short, simple questionnaire, the UCLA score also showed the highest completion rate and no floor effects compared to other metrics[46]. However, in at least one study, the UCLA activity score is not well correlated with patient satisfaction[17]. The Community Health Activities Model Program for Seniors (CHAMPS) activity questionnaire is a self-reported questionnaire for use with older populations and has shown high test-retest reliability and good correlation with device-based activity measurement[47,48]. This questionnaire should be well suited for use in the joint replacement population, and has been occasionally implemented[49]; however, this questionnaire consists of 41 questions and requires

four weeks of activity recall, and thus may be excessively burdensome to the patient in deployment.

1.7 Wearables for Physical Activity Monitoring

Within the last decade, a shift in methodology has seen the adoption of accelerometer-based activity tracking for objective measures of physical activity in clinical and epidemiological studies[34]. Self-reported activity questionnaires have well documented issues of measurement error that can be problematic when used as a basis for epidemiological or clinical research[50]. Accelerometer-based physical activity monitoring allows for the objective measurement of overall activity, quantification of activity intensity, and activity classification through machine learning models[51]. Objective activity measurement has been successfully integrated into large prospective cohorts in the United States[52] and United Kingdom[53], allowing for epidemiological analysis of physical activity as it relates to a wide range of medical conditions. Beyond population based epidemiological studies, objective measurement of physical activity is growing across a range of clinical fields and can be used as an outcome measure in randomised controlled trials for treatments and interventions, including patients with osteoarthritis[44,54–56]. An assessment of the current clinical use of wearable-based physical activity monitoring of perioperative knee arthroplasty patients will be reported in Chapter 2.

The use of wearables for physical activity monitoring is appealing as the primary mode of patient monitoring, as physical activity is a key marker of function and patient return to activities of daily living. By monitoring recovery of physical activity, we can monitor changes in the overall functional state of the patient during recovery. This monitoring can be conducted with a single accelerometer, with

minimal burden to the patient, and no need for additional office visits or dependence on interaction with research and clinical teams. I chose not to pursue wearable monitoring of functional gait, as this use of wearables is primarily aimed at reproducing optical gait analysis measurements in the free-living environment. Producing these measurements requires the use of multiple sensors, generate large amounts of data, and utilise resource-expensive gyroscopes and magnetometers. The use of these sensors for longitudinal monitoring of recovery during the early postoperative period would likely require either multiple data transfers and battery charges or multiple on-site visits for the participants. While there are interesting applications for the clinical implementation of wearables sensors for gait and functional analysis, the increased participant burden outweighs the potential benefit when compared to physical activity monitoring in my view. Further, in a systematic review of wearable-based gait analysis, Mills et al.[57] found that a significant decrease in cadence is the best distinguishing feature between patients with severe end-stage OA and healthy controls. Cadence is a metric that can be derived from a single accelerometer during routine physical activity monitoring, thus further reducing the need for resource expensive IMU-based functional analysis.

While device measured physical activity removes recall bias and other issues of self-report, device expense and a lack of consensus on body location, standardised data processing and device-based activity metrics remain obstacles needing addressed in future large-scale physical activity research[34]. The following sections detail key methodological parameters for the implementation of wearable-based physical activity monitoring in a clinical research protocol.

1.7.1 Body Placement

A key factor in the implementation of wearables for functional monitoring of a clinical orthopaedic population is the body location where the sensors are to be worn. The location of the sensor on the body has implications in terms of patient acceptability, wear time expectations and compliance, functional metrics able to be measured, and data quality.

Hip or waist placement of accelerometers has historically been the most common location for physical activity tracking[58]. Compared to other body locations, hip placement provides acceleration near to the center of body mass, while reducing signal noise associated with non-ambulatory activities. Hip-based accelerometry was implemented within the US-based National Health and Nutrition Examination Study, as an objectively measured physical activity study in a large longitudinal study cohort[52]. The primary limitation to hip-based accelerometry lies in the need for the participant to wear either a belt strap, or for the device to be clipped to the participant's clothing.

Thigh device placement is a popular choice for physical activity tracking, particularly when posture and sedentary activity is a primary area of focus[59]. When placed on the thigh, activities such as cycling, sitting, and lying down can be distinguished based on the direction of gravity relative to the orientation of a triaxial accelerometer's primary axes[60]. Thigh positioning usually relies on physically affixing a device to the participant's leg with an adhesive and is likely less comfortable for the participant compared to wrist or hip placement. The ProPASS (Prospective Physical Activity, Sitting and Sleep) Consortium is an international collaboration focused on pooling physical activity studies to relate health outcomes to sleep, activity, and postural data collected from thigh-worn datasets.

Physical activity monitoring with a wearable sensor at the ankle is less common than hip, thigh, or wrist-based measurement. A primary motivating factor for ankle placement is for increased accuracy in step detection[61]. The StepWatch is an ankle-based research grade wearable which has the highest step detection accuracy when compared to other off-the-shelf devices[61]. While this device is a research-grade device, the metrics and software are proprietary, thus potentially limiting broad use. Ankle placement may also suffer from some level of cultural stigma, as ankle-based location monitoring devices are common in judicial and penal systems. Other body locations have been used in research studies as well. In prior studies, physical activity has been collected from sensors placed at the ear, chest, and around the neck, among other locations[62]. Each of these locations have related benefits and limitations but have not been widely adopted within the physical activity research community.

Wrist-based accelerometry has gained popularity in physical activity research over the past decade. Wrist-based commercial wearables dominate the personal fitness tracking devices market and reflect high acceptability and preference of this body position over all others[63]. The primary advantage of wrist sensor location over any other body location relates to wear time and participant compliance[64]. Because a wrist-based sensor does not need to be removed for sleeping, bathing, or changing clothes, several studies using wrist-based measurement implement 24-hour monitoring protocols. The shift of wear time expectations from waking hours to twenty-four hours increases overall participant wear time by reducing the number of opportunities a participant removes and potentially forgets to replace the device. Additionally, comfortable sensor wear at the wrist during sleep enables analysis of sleep time, sleep efficiency, and other metrics inaccessible at other

body placements[65]. A historical limitation of physical activity quantification at the wrist is the prevalence of non-ambulatory noise in the accelerometry data from extraneous arm movement. Lower accuracy in step detection is frequently presumed as limitation of device placement at the wrist; however, recent studies have shown similar accuracy of more recent wrist and hip based commercial step trackers[61]. Wrist-based accelerometry has been validated against gold standard measures of energy expenditure and have facilitated successful activity classification through machine learning methods[66,67]. A primary consideration that must be made when selecting wrist-based accelerometry is whether to place the sensor on the dominant or non-dominant arm. A dominant arm placement allows for most participants to keep wearing a watch on the non-dominant arm. Dominant arm sensor location is associated with approximately 10% higher measured acceleration than non-dominant positioning and must be accounted for when comparing metrics between studies[68].

1.7.2 Device Selection

Decisions made regarding expected wear time, length of data collection, and body position all play a role in device selection for patient monitoring. A primary consideration to be made is the selection of a commercial or research-grade activity monitor for ambulatory monitoring of physical activity. Commercial activity trackers, made by companies including Apple, Fitbit, Garmin, and Polar frequently include elegant user interfaces through smartphone and web applications. Commercial devices have traditionally reported physical activity in terms of daily step count and energy expenditure calculated from accelerometer data, but more recently have begun to incorporate measures of heartrate, blood oxygenation, galvanic skin response, and GPS location. While commercial devices can be

purchased for many body locations, most commercial activity monitors are designed for wear at the wrist through integration into a smartwatch or as a wristband[63]. To preserve storage space and to enable fast data transfer, commercial activity monitors typically convert raw data to summary measures on the device, limiting the ability for the generation of bespoke post hoc analysis of raw accelerometer data. Third party services, such as Fitabase, aim to help facilitate the use of commercial monitors for research, allowing groups to capitalise on the user-friendly and economic advantages of commercial devices[69].

As an alternative to commercial activity trackers, a wide range of research-grade accelerometers and IMUs are also available to researchers. Commonly used research-grade devices for activity tracking include products from ActiGraph, Axivity, Pal Technologies, Activinsights, Modus, and Shimmer, among others. The difference between commercial devices and research-grade devices lies primarily in data storage, transfer, and processing. Research-grade devices frequently come with software for off-the-shelf data processing; however, these devices allow for direct access of raw accelerometer data for bespoke analysis and metric generation. Because these devices collect raw accelerometer data instead of computing on-board summary statistics, data is typically logged directly onto the device instead of transferred regularly via a Bluetooth smartphone connection. Data logging necessitates physical transfer of the device between patient and research center prior to data analysis, and thus limits real time reporting or feedback of activity metrics in most instances. To enable real-time data transfer and feedback, some newer research-grade devices have begun to implement Bluetooth syncing as a solution for raw acceleration data transfer. These devices

significantly increase the cost of data collection in large studies but may add significant flexibility and functionality in long-term study protocols.

1.7.3 Sampling Rate

When collecting acceleration data from an accelerometer, adjustment of device sampling rate, the number of observations recorded per second, significantly affects the use of data storage capacity as well as battery life of during the desired monitoring period. A higher sampling rate will allow for increased granularity of data and finer detail in the timeseries during which the data was collected. Each day of data collection from a triaxial accelerometer at 100 Hz requires approximately 34 MB of storage, resulting in approximately 14 days of data collection in a typical 512 MB device storage setup. If a monitoring period of longer than 14 days is desired, a reduction in sampling rate would be required to extend storage capacity. Subsequently, collecting accelerometer data at 25 Hz would theoretically extend the monitoring window up to 56 days if battery charge was maintained. According to technical specifications of the Axivity AX3 accelerometer, data collection at 100 Hz leads to a battery life of 14 days, while a reduction in sampling rate to 12.5 Hz allows for 30 days of continuous monitoring under a single battery charge[70]. A sampling rate of 80-100 Hz is common in physical activity studies and has been implemented in some of the largest studies to date[53,71]. While some authors suggest accelerometer data should be captured at as high a sampling rate as possible[71], others suggest that most studies capture more data than necessary. Khan et al.[72] reason that typical sampling rates are up to 57% higher than necessary for analysis of human motion. Human gait typically occurs with motion repeating at 3 Hz or less. According to the Nyquist theorem, a minimum sampling rate, therefore, would be on the order of 5-10 Hz to allow for

full characterisation of the activity. Ultimately, study design and methodological requirements will dictate the sampling rate necessary for data collection.

1.7.4 Wear time

The expected daily wear time within a physical activity study is highly dependent on the body location selected for device placement. Two common wear time protocols are waking-hours-only and 24-hour wear time protocols. Studies assessing physical activity from a hip-worn device frequently incorporate waking-hour wear time protocols, as hip-worn devices must be taken off while bathing, when changing clothes, and can cause discomfort during sleep. Conversely, studies incorporating wrist-based devices are frequently designed with a 24-hour wear protocol, as wrist-based devices are comfortable for many people while sleeping and are not attached to clothing, thus can be worn during bathing or while changing clothes.

1.7.5 Monitoring Period

In most physical activity studies, a seven day monitoring period is the standard length of data collection across academic disciplines and clinical fields using research-grade accelerometers[58]. A length of seven days allows for data collection at a relatively high sampling rate while limiting power consumption to a point where the batteries of most devices will not need to be recharged by the participant. The minimum monitoring period to assess habitual physical activity is quoted as a range between just a few days to a week, depending on target population[73–75]. Weekend activity bias can be significant, so many studies aim to include measures during the work week and weekend[75]. Some studies investigating the change in physical activity over time, namely after an intervention or disease diagnosis, have elected monitoring periods beyond the typical seven days[76,77]. In these cases,

some investigators elect to reduce sampling rate and/or swap accelerometers on a regular basis in order to maintain storage and memory capacity over an extended period of time. With the use of Bluetooth syncing via smartphone app or data hub, monitoring periods of indefinite length are possible, but place a regular battery-charging burden on the participant.

1.7.6 Data Analysis

As mentioned in Section 1.7.2, a major advantage to the use of research-grade devices in physical activity studies is the ability to retain full control over raw acceleration data and subsequent analysis of that data. Wherein commercial activity trackers deploy proprietary algorithms for generation of summary activity metrics and activity classifications, the ability to employ bespoke and open-sourced analytical software packages to accelerometer data sets is a powerful advantage for the use of research-grade devices in epidemiological and clinical studies.

Summary measures of physical activity are perhaps the most commonly used physical activity metrics in clinical and epidemiological studies[51]. Summary measures commonly used include daily step count and variations of overall acceleration magnitude including Euclidean Norm Minus One, Lowpass Filtered Euclidean Norm Minus One, Highpass Filtered Euclidean Norm, and Mean Amplitude Deviation[78]. The Actigraph “count” is a proprietary summary metric that is commonly reported in studies using ActiGraph branded devices. Daily step count is a common summary measure wherein the number of steps or gait cycles taken per day are estimated. Step count is a highly intuitive measure of overall daily physical activity yet can be difficult to accurately quantify depending on sensor location and detection algorithm used[61]. Acceleration magnitude summary

metrics provide a mechanism to aggregate raw acceleration from all three axes of a triaxial accelerometer. These metrics are validated against doubly-labelled water for estimating energy expenditure and serve as a basis for time usage behaviour based on cut-point analysis[66,79].

Time-use cut-point analysis is a frequently used tool to determine time spent at predefined levels of activity intensity. Cut-points for moderate-to-vigorous physical activity (MVPA), are highly dependent on device location and the population of interest, and thresholds for sedentary, light, and moderate-to-vigorous activities can differ widely between children, healthy adults, and older adult populations[80]. A limitation of cut-point analysis is the all-or-nothing nature of dichotomising activity levels based on an acceleration threshold. For example, a person may be participating in ten minutes of continuous exercise wherein their overall average acceleration fluctuates between 99 and 101 m*g*. In a typical cut-point analysis with a 100 m*g* threshold for MVPA, half of the exercise (the portion spent above 100 m*g*) would be classified as time spent in MVPA, whereas the other half would be considered light intensity activity, with nothing to indicate this light intensity activity time as nearly identical to the time spent in MVPA.

Activity recognition through machine learning classification is a process in which signal features are extracted from raw accelerometry data to classify time spent in various physical behaviours. Activity classification can take the form of predicting times of walking, running, cycling, riding in a vehicle, and other specific forms of activity, or in more general activity categories including light, moderate, sedentary, or sleep[67,81,82]. Early implementation of machine learning activity classification relied on laboratory-based model training, however more recent model improvements are based on labelled accelerometer data captured during

times of unscripted free-living and are shown to outperform laboratory-trained models[82,83]. Random forests, artificial neural networks, and support-vector machine models have been most commonly deployed as behavioural recognition classifiers[84].

A selection of popular open-source software packages exist for aiding researchers extract meaningful outcome measures from raw accelerometer data. The R Package, GGIR enables autocalibration of accelerometer data with options to calculate non-wear time, overall acceleration metrics, time-based cut-point analysis, and sleep detection[85]. The OpenMovement software https://github.com/digitalinteraction/openmovement/tree/master/Software/AX3/omconvert) is used to handle device programming and data download from Axivity devices, but also provides basic processing features for Axivity raw data or from any other device with a .csv outputted filetype. Both Pampro (https://github.com/Thomite/pampro) and the BiobankAccelerometerAnalysis Tool (https://github.com/activityMonitoring/biobankAccelerometerAnalysis) are open-source, python-based tools for accelerometer analysis. Pampro enables the import of multiple proprietary accelerometer filetypes in addition to .csv accelerometer data for the calculation of summary acceleration and cut-point metrics following calibration and non-wear detection. The BiobankAccelerometerAnalysis Tool allows for the import of Axivity, GeneActiv, and .csv data files, offering calibration, non-wear detection, summary and cut-point based activity measures, and additionally machine learning-based activity classification.

1.8 Thesis Scope and Structure

- In Chapter 1, a background knowledgebase is established around the clinical and methodological considerations for using wearable sensors to evaluate early recovery following knee arthroplasty.
- A systematic scoping review is presented in Chapter 2 to evaluate the current state of wearable sensor use to assess functional and clinical outcomes following knee arthroplasty. Commonly used sensor body placement, outcome metrics, and experimental protocols are reported, with the gaps in current implementation of wearable sensors identified and discussed.
- The effect of accelerometer sampling rate on the validity of interstudy comparisons of objectively measured physical activity is presented in Chapter 3. This chapter, along with the next, explore key methodological factors for objective activity monitoring using wearable sensors, as evaluated with the collection of primary data from a cohort of healthy volunteers.
- In Chapter 4, primary data is used to evaluate multiple approaches for estimating step count from raw acceleration time series data collected from the wrist by a research-grade accelerometer. The development and validation of a novel, open-source step detection algorithm based on a hybrid machine learning and heuristic methodology is presented.
- Data from the UK Biobank physical activity cohort is used in Chapter 5 to evaluate the utility of a novel step counting algorithm to generate step count and cadence metrics as measures of physical activity. This chapter additionally presents baseline activity metrics for healthy, non-arthritic

participants alongside cohorts of participants with end-stage osteoarthritis or a documented history of lower limb arthroplasty.

- A prospective, observational, longitudinal cohort study of physical activity in patients following unicompartmental or total knee arthroplasty is initiated in Chapter 6. This chapter focuses on the methodology and implementation of the clinical study. Baseline preoperative physical activity is reported in addition to patient acceptability, accelerometer wear time and retention, and the results from a patient and public involvement event.
- In Chapter 7, postoperative data from the continuous monitoring of knee arthroplasty patients is presented. The utility of continuous longitudinal postoperative monitoring following knee arthroplasty is assessed for identifying the trajectory of patient recovery, and to compare early recovery between unicompartmental knee arthroplasty and total knee arthroplasty.
- Finally, Chapter 8 presents a summary of key results of the preceding chapters. This summary and discussion explores next steps in future research including the potential for full implementation of wearable activity monitoring into the clinical orthopaedic setting.

1.9 Contributions

At the time of submission, two journal manuscripts have been published from the work of this thesis:

1. Small SR, Bullock GS, Khalid S, Barker K, Trivella M, Price AJ. Current Clinical Utilisation of Wearable Motion Sensors for the Assessment of Outcome Following Knee Arthroplasty: A Scoping Review. *BMJ Open*, 2019; 9(12).
2. Small SR, Khalid S, Dhiman P, Chan S, Jackson D, Doherty AR, Price AJ. Impact of Reduced Sampling Rate on Accelerometer-based Physical Activity Monitoring and Machine Learning Activity Classification. *Journal for the Measurement of Physical Behaviour*, 2021; 4(4):298-310.

Chapter 2: Wearables for the Assessment of Outcome Following Knee Arthroplasty: A Systematic Scoping Review

2.1 Copyright Notice

This thesis chapter reuses content published in the BMJ Open (CC BY-NC) under the following citation:

Small SR, Bullock GS, Khalid S, Barker K, Trivella M, Price AJ. Current Clinical Utilisation of Wearable Motion Sensors for the Assessment of Outcome Following Knee Arthroplasty: A Scoping Review. *BMJ Open*, 2019; 9(12).

2.2 Introduction

Several methodological considerations must be made during the implementation of wearable motion sensors in the study of functional outcomes following a clinical intervention. Body location, sampling rate, wear time, monitoring length, and testing protocols are each key aspects of sensor-based research which may affect the quality and reliability of data being collected[58,86–88]. Wearable technology is increasingly being used to perform gait analysis and to assess patient mobility in clinical patient populations[31]. The manner in which wearable sensors have been deployed in research and clinical care of knee arthroplasty patients has not been fully examined. As a first step in this thesis, I conducted a scoping review to assess the clinical application of wearable sensors to evaluate functional outcomes following knee arthroplasty. Specifically, this chapter has two primary aims:

1) To evaluate the wearable sensor methodology in the clinical setting to assess standard procedures and to identify opportunities for, or barriers to interstudy comparison.

2) To establish the current utilisation of wearable sensors during the early postoperative recovery period to understand gaps in current clinical practice and opportunities for improved perioperative monitoring in this patient population.

2.3 Methods

I conducted a scoping review following the Joanna Briggs Institute Methodology for JBI Scoping Reviews[89] and consulted the PRISMA Extension for Scoping Reviews checklist for reporting[90]. The protocol is registered within the Open Science Framework (https://osf.io/rzg9q/). For inclusion within this review, manuscripts must report the use of wearable sensor technology for monitoring patient function, mobility, surgery effectiveness, or joint stability, or serve as the basis for the assessment of a questionnaire or other measure of physical activity and function. Included studies could be conducted in a hospital or home setting and must include postoperative data collection following any total or partial knee arthroplasty for end-stage osteoarthritis. Explicit validation studies of wearable sensor technology and studies implementing the use of inertial measurement units (IMUs) for robotic-assisted rehabilitation or surgical navigation were excluded, as those studies do not meet the primary aims of evaluating basic patient outcome. Inclusion and exclusion screening criteria are presented in Table 2-1.

Table 2-1: Literature Review Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
- Studies including the use of any combination of accelerometers, pedometers, or inertial measurement units for postoperative patient assessment - Studies reporting sensor-based data for patients having undergone total or partial knee replacement at any timepoint post-op - Studies including a minimum of n = 5 knee arthroplasty patients - Studies published between 2008 and June 2019	- Conference abstracts - Systematic reviews - Studies that do not differentiate/stratify data results between knee arthroplasty patients and those having undergone other treatments - Studies focusing on the use of accelerometers/IMUs for use in surgical navigation - Studies utilising cadaveric, robotic, or animal models - Studies using sensors as part of an active feedback assistive technology - Published study protocols - Validation studies with the primary aim of evaluating technology and limited clinical inquiry

To identify relevant studies, an electronic database literature search was conducted with assistance from outreach librarian Nia Roberts. Databases searched were: CINAHL (1982 to present), EMBASE (1974 to present), MEDLINE (1950 to present), and Web of Science Core Collection (1945-Present). Search criteria are presented in Table 2-2. Because wearable sensor technology is an emerging field benefiting from developments in miniaturisation and wireless data transfer, the literature search was designed to primarily capture the current use of modern wearable instrumentation. Thus, manuscripts published between January

2008 and June 2019 were considered. Additionally, reference lists of relevant articles were searched for additional appropriate inclusions. Systematic reviews were not included in the review but were also checked for potentially relevant references. Upon completion of the comprehensive searches, duplicate entries were removed through manual verification within Mendeley reference management software.

Table 2-2: Database Search Strategy

1	Osteoarthritis, Knee/
2	Osteoarthritis.ti,ab
3	Knee/
4	Knee*.ti,ab
5	Wearable Electronic Devices/
6	Wearable*.ti,ab
7	Monitoring, Ambulatory/
8	(IMU* OR inertial measurement unit*).mp
9	(fitbit* OR jawbone* OR iwatch* OR activity tracker* OR step counter* OR axivity* OR tracpatch* OR geneactive* OR actipal* OR pedometer* OR actigraph*).mp
10	(accelerometer* OR gyroscope*).mp
11	(arthroplast* OR replace*).mp
12	Arthroplasty, Replacement, Knee/
13	1 OR 2
14	3 OR 4
15	13 OR 14
16	5 OR 6 OR 7 OR 8 OR 9 OR 10
17	11 OR 12
18	15 AND 16 AND 17

Screening of manuscript titles and abstracts were conducted by two independent, blinded reviewers, myself and fellow student Garrett Bullock, through the use of the Rayyan systematic review web app[91]. Prior to screening, reviewers discussed inclusion and exclusion criteria, while working through a calibration exercise of example titles and abstracts to ensure consistency between individuals. Following

title and abstract review, both reviewers evaluated the subsequent full-text manuscripts for inclusion in the final review. Any disagreements following full-text assessment were discussed between reviewers. A third reviewer was available for consultation if no consensus could be reached.

Following the screening process, I collected data from all included full-text manuscripts within a data extraction spreadsheet, which was subsequently verified by Garrett Bullock. Any inconsistencies were resolved by consensus. Data extracted from the manuscripts were categorised by patient demographics, technology, testing protocol, and sensor-based outcome variables. Specifically, data were extracted to include the number of male and female participants in knee arthroplasty, end-stage osteoarthritis, or healthy control populations. The type of wearable sensor was recorded, including number of axes of motion recorded, as well as the use of accelerometers, gyroscopes, magnetometers, and global positioning systems, and the reported data sampling rate for all sensor measurements. The number of sensors and body location was also recorded. For data charting purposes, manuscripts were divided into three distinct subgroups based on the primary aims of each study as follows: 1) studies with the primary aim of assessing functional or gait parameters 2) physical activity studies with the primary aim of quantifying change in overall activity as a result of knee arthroplasty, and 3) instability studies with the primary aim of assessing postoperative joint stability. The duration of free-living activity monitoring was extracted for all physical activity studies, as well as the pre- and postoperative timepoints when data was collected. Finally, sensor-based outcome variables were recorded to assess the specific metrics reported in each study.

2.4 Results

A PRISMA flow diagram[92] detailing the results of the literature search and review is presented in Figure 2-1. The searches yielded 346 records for title and abstract screening, of which 69 were assessed as full texts. Of these, 45 were included in the review. The 24 excluded studies lacked knee arthroplasty patients or stratified data, included too few study participants, were protocol or validation studies, or deemed unrelated to the review question. While any type of knee arthroplasty for end-stage osteoarthritis was included in the search, no studies evaluating patients following bi-compartmental, patellofemoral, or other partial knee arthroplasty were identified.

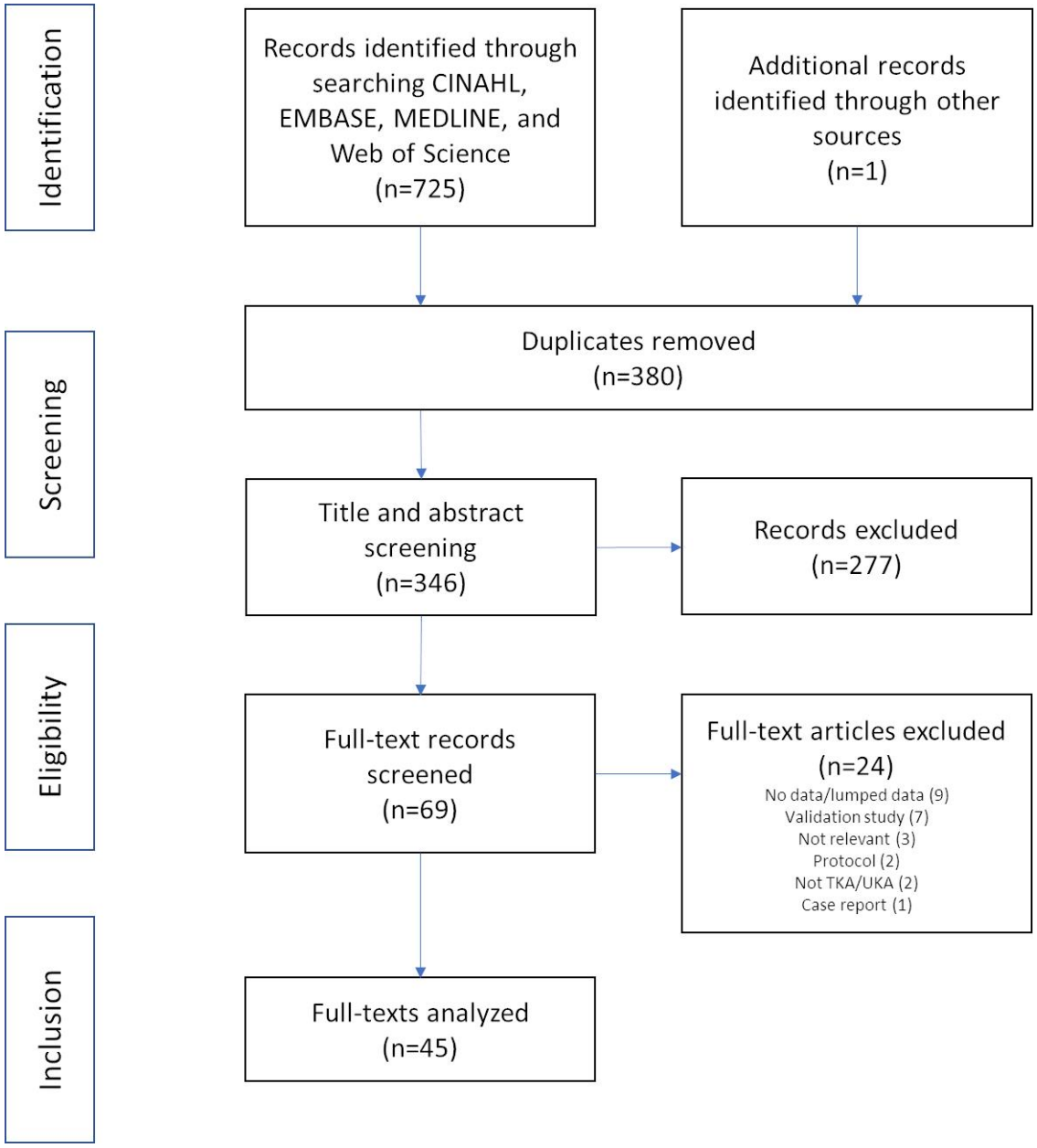


Figure 2-1: PRISMA literature review flow diagram

The use of wearable technology to assess outcome and effectiveness of TKA and UKA has increased over the period evaluated in this review. As shown in Figure 2-2, an average of just 1 to 2 studies per year were published between 2008 and 2012. More recently there has been continued growth in the number of studies implementing wearable motion sensors entering the public record, with 10 wearable-sensor based studies focused on knee replacement in 2018.

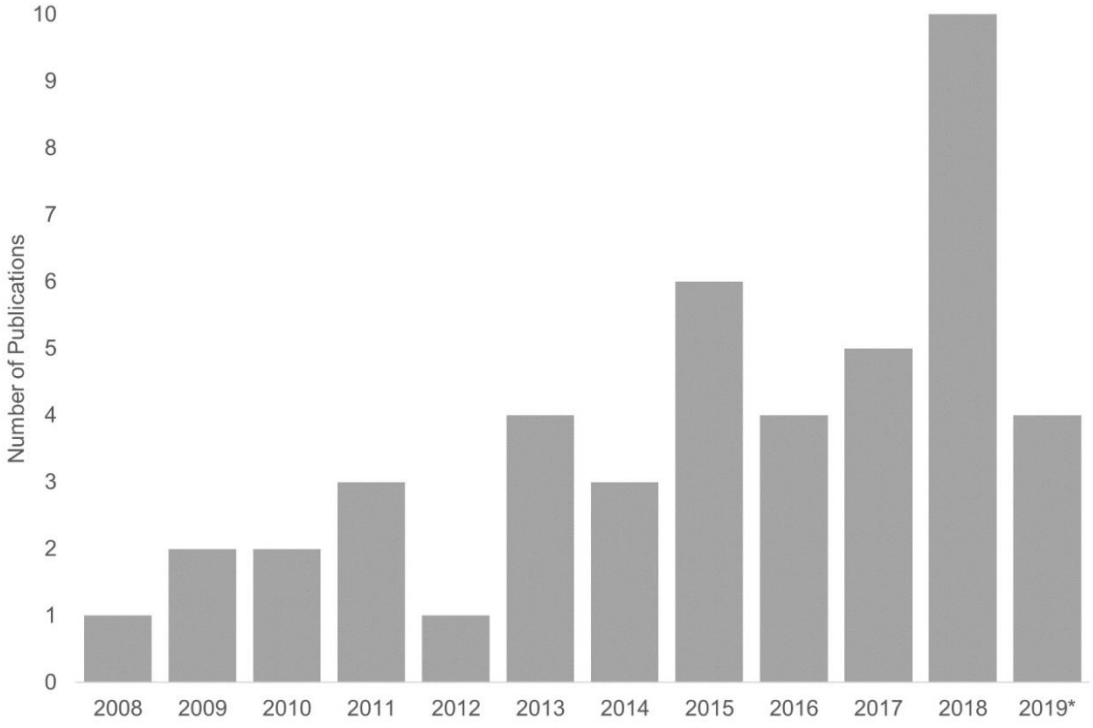


Figure 2-2: The number of publications within this review by publication year (*2019 through June).

2.4.1 Wearables for Functional Assessment

Sixteen studies were included in the functional assessment subgroup (Table 2-3), comprising data collection on 448 TKA patients (57% female), mirroring the 43:57 male-to-female ratio within primary TKA across the UK[93]. A single study included functional comparisons with 29 end-stage OA patients without knee replacement,[94] while eight studies included comparative assessments with a total of 141 healthy control populations[95–102].

A triaxial accelerometer was the most commonly used sensor incorporated in the functional assessment subgroup, with 15 of 16 studies (94%)[95,96,105–109,97–104] using the sensor alone or in combination with other instrumentation. Seven of the

functional studies exclusively utilised triaxial accelerometers as the means for data collection during functional analysis[95–97,99,101,105,108]. Three studies used IMUs with an integrated triaxial accelerometer and triaxial gyroscope[98,104,106], while three others recorded gait with a 9 degree-of-freedom IMU integrating triaxial accelerometer, gyroscope, and magnetometer[102,103,107]. Storey et al.[100] used a triaxial accelerometer for indoor gait analysis, but included two GPS watches for outdoor walking tests. Two studies by van Hemmert et al.[109,110] incorporated the DynaPort ADL sensor system consisting of 6 uniaxial accelerometers, with one of those studies supplementing the DynaPort system with an additional triaxial accelerometer[109]. Sensor sampling rate for data recording was varied across studies, ranging from below 50 Hz up to 500 Hz and not directly reported in a third of studies[98,101,107,109,110]. Body location for sensor placement was equally varied, with no more than seven studies sharing any one specific sensor placement.

Testing protocols within the function studies were generally consistent and primarily comprised gait analysis, with patients walking on a 10-20 m walkway or treadmill. Four studies[95,97,109,110] used the DynaPort knee test protocol[111] which includes a series of walking, stepping, and lifting movements to mimic activities of daily living. Bolink et al.[103] tested patients through a sit-to-stand protocol, while both Bolink and Christiansen et al.[96] tested patients in stepping exercises. Most studies (13 studies, 81%)[95,96,106,107,110,97–99,101–105] included preoperative data collection with repeated tests at one or more timepoints up to 1-year postoperatively, most frequently up to 6-weeks after surgery. Three studies collected data at an unspecified point any time greater than 6 months postoperatively[100,108,109].

As presented in Table 2-4, the evaluation of gait symmetry and variability was the most common sensor-based outcome metric (8 studies, 50%)[96,99,101,103,105,106,108,109], while dynamic range of motion was also commonly reported (6 studies, 38%)[98,102–104,106,107]. A wide variety of spatiotemporal gait parameters were also reported, including varied combinations of walking speed[99–101,103,106,109], cadence[99,103,109], stride length[99–101,103,106,109], stride time[98,99,106], stance time[101,106], vertical displacement[99,109], swing power[101], and foot fall[101]. Three studies reported functional performance primarily based on a stand-alone function score[97,109,110]. Other studies reported on gait function through a variety of other parameters including acceleration magnitudes[96,103,108], activity transitions[95], multiple aspects of dynamic range of motion[98,103,104,106], and varied measures of gait symmetry, inefficiency, and variability[96,99,101,103,105,106,108,109].

Sensor wear location was highly heterogeneous in these studies. The thigh (7 studies, 44%)[98,101,102,104,107,109,110] and proximal tibia (7 studies, 44%)[96,98,102,104,107,109,110] were the most common wear locations; however, foot (4 studies, 25%)[101,105,106,108], hip (3 studies, 19%)[100,109,110], back (3 studies, 19%)[103,109,110], sternum (3 studies, 19%)[101,109,110], and ear (2 studies, 13%)[95,97] locations were also used in multiple studies. Only one study utilised a wrist-based sensor for functional assessment of knee arthroplasty patients[100].

Table 2-3: Population and Methods Summary of Functional Assessment Studies

	Population				Testing Protocol					Timepoints					
	Knee Arthroplasty Patients	% Female	Healthy Control	Control % Female	Dynaport ADL Protocol[111]	10-20 m walkway	Outdoor course	Stair stepping	Other	Pre-operative	0-6 weeks post-op	7-12 weeks post-op	24-26 weeks post-op	1+ year post-op	Unclear
Ali et al. 2012[95]	5	40%	10	30%	✓					✓	✓	✓	✓		
Bolink et al. 2015[103]	20	65%				✓		✓	✓	✓				✓	
Chapman et al. 2019[102]	20	41%	10	50%					✓	✓	✓			✓	
Chiang et al. 2017[104]	18								✓	✓	✓				
Christiansen et al. 2015[96]	24	54%	19	53%			✓	✓		✓	✓		✓		
Hiyama et al. 2015[105]	57	61%				✓			✓	✓	✓				
Kluge et al. 2018[106]	24	67%				✓				✓				✓	
Kwasnicki et al. 2015[97]	14	57%	15		✓					✓	✓	✓	✓		
Rahman et al. 2015[98]	45		29	59%		✓				✓		✓		✓	
Ramkumar et al. 2019[107]	25	56%							✓	✓	✓	✓			
Senden et al. 2011[99]	24	54%	24	54%		✓				✓	✓	✓			
Storey et al. 2013[100]	16		22				✓							✓	
van Hemert et al. 2009a[109]	53				✓	✓								✓	
van Hemert et al. 2009b[110]	76				✓					✓		✓	✓	✓	
Youn et al. 2018[108]	18	50%							✓						✓
Zhang et al. 2016[101]	12	58%	12	50%		✓				✓	✓		✓		

Table 2-4: Sensor Summary for Functional Assessment Studies

	Technology						Sampling Rate				Body Location								Sensor Metric					
	1-axis accelerometer(s)	3-axis accelerometer(s)	3-axis gyroscope(s)	3-axis magnetometer(s)	GPS	Total # of sensors	Not reported	≤ 50 Hz	100-200 Hz	>200 Hz	Ear	Sternum	Arm/Wrist	Lower back	Hip/Waist	Thigh	Proximal Tibia	Foot/Ankle	Acceleration Magnitude	Activity Transitions	Gait Symmetry/Variability	Multivariate sensor-based function score	Range of Motion	Spatiotemporal Gait Parameters
Ali et al. 2012[95]		✓				1		✓			✓									✓				
Bolink et al. 2015[103]		✓	✓	✓		1			✓					✓					✓		✓		✓	✓
Chapman et al. 2019[102]		✓	✓	✓		2		✓								✓	✓						✓	
Chiang et al. 2017[104]		✓	✓			2		✓								✓	✓						✓	
Christiansen et al. 2015[96]		✓				2				✓							✓		✓		✓			
Hiyama et al. 2015[105]		✓				1												✓			✓			
Kluge et al. 2018[106]		✓	✓			2			✓									✓			✓		✓	✓
Kwasnicki et al. 2015[97]		✓				1		✓			✓											✓		
Rahman et al. 2015[98]		✓	✓			4	✓									✓	✓						✓	✓
Ramkumar et al. 2019[107]		✓	✓	✓		2	✓									✓	✓						✓	
Senden et al. 2011[99]		✓				1			✓												✓			✓
Storey et al. 2013[100]		✓			✓	3			✓				✓		✓									✓
van Hemert et al. 2009a[109]	✓	✓				7	✓	*				✓		✓	✓	✓	✓				✓	✓		✓
van Hemert et al. 2009b[110]	*					6*	✓	*				*		*	*	*	*					✓		
Youn et al. 2018[108]		✓				2				✓								✓	✓		✓			
Zhang et al. 2016[101]		✓				5	✓					✓				✓		✓			✓			✓

2.4.2 Wearables for Objective Physical Activity Measurement

Twenty-six studies were identified as primarily focusing on physical activity following knee arthroplasty (Table 2-5), including a total of 1,550 patients (58% female). Five of those studies included comparisons with a total of 620 end-stage osteoarthritis patients without a knee replacement[112–115], though, 458 of those patients came from a single report[114]. Three studies compared physical activity levels of knee replacement patients to a total of 272 healthy controls[112,116,117]. Four studies also included physical activity from measures from total hip replacement patients for comparison to knee patients[118–121].

Physical activity studies were consistent in the number of sensors implemented, with twenty-one studies (81%) using a single sensor for activity measurements[112,113,124–133,114,134,115–117,119,120,122,123]. As technology and availability of wearable accelerometers has progressed, the type of sensor used has also changed. In studies published between 2008 and 2016, 9 studies (69%)[113,114,117,119,121,122,135–137] reported using a uniaxial accelerometer versus just 3 studies (23%)[123,132,133] using a triaxial accelerometer. Conversely, in studies published in 2017-2019, 8 studies (62%)[112,115,124,126,127,130,131,134] reported using a triaxial accelerometer, while only 3 (23%)[116,120,122] used a uniaxial accelerometer. Overall, three studies did not report the number of axes recorded[118,125,128], and most studies within this subgroup (73%)[113,114,127–132,134,137,117–122,124,125] did not report the sampling rate at which data was collected. No standard frequency was observed within studies reporting sampling rate, as two studies collected data at 10 Hz[116,133], two at or near 30 Hz[126,135], and three at 100 Hz[112,115,123]. Accelerometers were mounted at hip or waist level, most commonly with an elastic strap, in half of the included physical activity studies[113,114,130,133,135,115,119,120,122–

[124,126,127]. Seven studies positioned sensors on the thigh[112,116,118,121,125,135,137], four on the sternum[118,121,136,137], four on the lower leg[117,129,131,136], one on the foot[136], and two on the wrist[128,134].

All studies investigating physical activity of knee arthroplasty patients collected data in the free-living environment. Most commonly, free-living data was collected in 7-day sessions (13 studies, 50%)[113,114,130,134,135,115,116,119,122–124,126,127], with 10 studies (43%)[112,117,118,121,125,129,131,135–137] collecting data over shorter sessions (e.g., 2-4 days). Four studies reported longitudinal activity metrics for longer time periods, with Hayashi et al.[120] and Tsuji et al.[133] reporting physical activity during postoperative hospitalisation for 8 and 14 days, respectively, while Luna et al.[128] reported basic accelerometer counts for 20 days post-op and Taniguchi et al.[132] asked patients maintain a daily step count log for 6 months following surgery. Two primary methodologies exist for sensor wear time. Fifteen studies (58%)[112,113,130–132,135,136,114,115,119,120,122–125] instructed patients to wear the accelerometers during waking hours only, while ten studies (43%)[116–118,121,126,128,129,133,134,137] instructed patients to wear the accelerometers 24 hours per day.

Table 2-5: Population and Methods Summary for Physical Activity Studies

	Population						Session Length			Wear Time			Timepoints						
	Knee Arthroplasty Patients	TKA/UKA %Female	End-Stage OA	OA (%Female)	Healthy Control	Control % Female	< 7 days	= 7 Days	> 7 days	Waking Hours Only	24 hours per day	Not Reported	Pre-op	0-6 weeks post-op	7-12 weeks post-op	6 months post-op	1 year post-op	> 1 year post-op	Unspecified Post-op
Agarwal et al. 2017[122]	32	47%						✓		✓			✓						✓
Bolszak et al. 2014[123]	50	50%						✓		✓									
Brandes et al. 2011[135]	53	64%					✓	✓					✓		✓	✓	✓		
Christiansen et al. 2019[124]	43	53%						✓		✓			✓	✓		✓	✓		
Cooper et al. 2017[116]	62	58%			62	42%		✓			✓		✓	✓		✓			
Daugaard et al. 2018[112]	52	50%	54	54%	171	44%	✓			✓								✓	
De Groot et al. 2008[118]	44	55%					✓						✓		✓	✓			
Fenten et al. 2018[125]	80	54%					✓			✓				✓					
Frimpong et al. 2018[126]	45	93%						✓			✓		✓	✓		✓			
Harding et al. 2014[119]	44	64%						✓		✓			✓			✓			
Hayashi et al. 2018[120]	40	85%							✓	✓				✓					
Hayes et al. 2011[136]	52	52%	13	54%			✓			✓			✓	✓	✓	✓	✓		
Kahn et al. 2015[113]	60	50%	63	49%				✓		✓									✓
Kahn et al. 2016[114]	60	50%	458	51%				✓		✓									✓
Losina et al. 2018[127]	150	57%						✓				✓	✓			✓			
Luna et al. 2019[128]	41	59%							✓		✓		✓	✓					
Lützner et al. 2014[117]	97	46%			39	59%	✓				✓		✓				✓		
Lützner et al. 2016[129]	221	57%					✓				✓		✓				✓		
Pellegrini et al. 2018[130]	16	69%						✓		✓			✓		✓	✓			
Schotanus et al. 2017[131]	20	35%					✓			✓			✓	✓					
Taniguchi et al. 2016[132]	81	90%							✓	✓				✓	✓	✓			
Tsuji et al. 2010[133]	20	70%							✓		✓		✓	✓					
Twiggs et al. 2018[134]	91	51%						✓			✓		✓	✓					
Vissers et al. 2010[137]	44	55%					✓				✓		✓			✓			
Vissers et al. 2013[121]	14						✓						✓			✓		✓	
Webber et al. 2017[115]	38	58%	32					✓		✓								✓	

Physical activity was measured at different timepoints primarily dependent upon the type of studies being conducted. The longitudinal studies reviewed most frequently assessed physical activity within 1-2 weeks preoperatively, and then subsequently near 6 weeks and 6 months postoperatively. Three longitudinal studies assessed physical activity up to 1-year post-op[124,129,135], while Vissers et al.[121] reported physical activity up to 4 years follow-up. Cross-sectional studies were more varied, with some reporting at exact postoperative timepoints, while others evaluated patient activity at non-specific or broadly defined timepoints[113,114,122,123]. The most commonly reported timepoints for activity analysis across all physical activity studies were preoperative (18, 69%)[102,116,129–131,133–137,117–119,121,122,126–128] and at 6-month postop (12, 46%)[116,118,136,137,119,121,124,126,127,130,132,135].

Sensor-based outcome variables within the physical activity studies varied considerably between studies (Table 2-6), though not as widely as in the functional assessment subgroup. Reporting of step counts is the most commonly utilised measure of physical activity within the knee arthroplasty population and has been increasingly common within these studies, included in 9 of 12 (75%) studies[112,115,116,120,124,126,127,131,134] from 2017-2019, compared to just 4 of 13 (31%) studies[117,129,132,135] from 2008-2016. The amount of active and sedentary time (12 studies, 46%)[115,117,136,137,118,119,121,125,126,129–131], quantification of time spent at different activity intensity levels (12 studies, 46%)[113,114,130,135,115,117,122–124,126,127,129], and approximations of energy expenditure (10 studies, 38%)[112–114,123,126,128,130,131,133,136] are also commonly used measures which provide another level of detail beyond basic step counting techniques. Three studies reported transitions between activities and bouts of activities[112,118,131].

Table 2-6: Technology Summary for Physical Activity Studies

	Technology					Sampling Rate				Body Location							Sensor Metric					
	1-axis accelerometer(s)	2-axis accelerometer(s)	3-axis accelerometer(s)	Not Reported/Unclear	Total # of sensors	10 Hz	30-32 Hz	100 Hz	Not reported/unclear	Sternum	Arm/Wrist	Hip/Waist	Thigh	Ant. Tibia/Lower Leg	Foot/Ankle	Not reported	Activity Intensity/Time	Active/Sedentary Time	Counts/METS/Energy	Steps/Gait Cycles	Sit-to-stand transfers	Walking Bouts
Agarwal et al. 2017[122]	✓				1				✓			✓					✓					
Bolszak et al. 2014[123]			✓		1			✓				✓					✓		✓			
Brandes et al. 2011[135]	✓				4		✓					✓	✓				✓			✓		
Christiansen et al. 2019[124]			✓		1				✓			✓					✓			✓		
Cooper et al. 2017[116]	✓				1	✓							✓							✓		
Daugaard et al. 2018[112]			✓		1			✓					✓							✓	✓	✓
De Groot et al. 2008[118]				✓	3				✓	✓			✓					✓			✓	
Fenten et al. 2018[125]				✓	1				✓				✓					✓				
Frimpong et al. 2018[126]			✓		1		✓					✓					✓	✓	✓	✓		
Harding et al. 2014[119]	✓				1				✓			✓						✓	✓			
Hayashi et al. 2018[120]	✓				1				✓			✓								✓		
Hayes et al. 2011[136]	✓				5				✓	✓				✓	✓			✓	✓			
Kahn et al. 2015[113]	✓				1				✓			✓					✓		✓			
Kahn et al. 2016[114]	✓				1				✓			✓					✓		✓			
Losina et al. 2018[127]			✓		1				✓			✓					✓			✓		
Luna et al. 2019[128]				✓	1				✓		✓								✓			
Lützner et al. 2014[117]	✓				1				✓					✓			✓	✓		✓		
Lützner et al. 2016[129]	✓				1				✓					✓			✓	✓		✓		
Pellegrini et al. 2018[130]			✓		1				✓			✓					✓	✓	✓			
Schotanus et al. 2017[131]			✓		1				✓					✓				✓	✓	✓	✓	✓
Taniguchi et al. 2016[132]			✓		1				✓							✓				✓		
Tsuji et al. 2010[133]			✓		1	✓						✓							✓			
Twiggs et al. 2018[134]			✓		1				✓		✓									✓		
Vissers et al. 2010[137]	✓				4				✓	✓			✓					✓				
Vissers et al. 2013[121]	✓				4				✓	✓			✓					✓				
Webber et al. 2017[115]			✓		1			✓				✓					✓	✓		✓		

2.4.3 Wearables for Joint Instability

Three studies from the literature review used wearable sensors affixed to the operative tibia to examine overall joint instability following TKA[138–140]. A total of 124 patients (69% female) were examined (Table 2-7) with two studies including comparisons to a total of 54 healthy control patients[138,139]. For instability measurement, Khan et al.[138] and Soeno et al.[140] used triaxial accelerometers, while Roberts et al.[139] used an integrated IMU with triaxial gyroscope and triaxial magnetometer. Both Khan et al.[138] and Roberts et al.[139] recorded joint motion in patients at a minimum of 6 months post-op at a sampling rate of 100 Hz while patients performed stepping, sit-to-stand, and walking protocols. Similarly, both studies compared joint stability between TKA patients and healthy controls. Khan et al.[138] reported instability based on mean acceleration and distribution of acceleration frequencies, while Roberts et al.[139] quantified means and ranges of acceleration and jerk parameters. Soeno et al.[140] collected data on patients more than 1-year post-op at a sampling rate of 250 Hz while quantifying instability directional root mean square acceleration, comparing patients with and without self-reported subjective instability.

Table 2-7: Population and Technology Summary for Instability Studies

	Population				Technology				Sampling Rate		Sensor Location	Testing Protocol			Timepoints	
	Knee Arthroplasty Patients	% Female	Healthy Controls	Control % Female	3-axis accelerometer(s)	3-axis gyroscope(s)	3-axis magnetometer(s)	Total # of sensors	100 Hz	250 Hz	Proximal Tibia	Walk	Sit-to-stand	Block/Stair Step	Minimum 6 months post-op	> 1-year post-op
Khan et al. 2013[138]	38	59%	18	61%	✓			1-2	✓		✓	✓	✓	✓	✓	
Roberts et al. 2013[139]	27	59%	18	81%	✓	✓	✓	1-2	✓		✓	✓	✓	✓	✓	
Soeno et al. 2018[140]	70	77%			✓			1-2		✓	✓	✓				✓

2.5 Discussion

Wearable motion sensors are increasingly being used to assess patient function, physical activity, and joint stability following knee arthroplasty. The use of IMUs and accelerometers allow clinicians and researchers the ability to track a broad scope of outcome measures as they relate to free-living activities outside of the clinical setting, with greater objectivity when compared to activity questionnaires. In this scoping review, I found that wearable technology has been most commonly used for preoperative and postoperative comparisons at 6 to 12 months of follow-up, with less focus on early recovery and rehabilitation. No standard outcome measure or testing methodology has been established in wearable-based gait analysis or physical activity monitoring.

The use of wearable motion sensors for gait analysis is an attractive alternative to optical gait analysis for functional assessment of knee arthroplasty patients. Whereas optical gait analysis requires significant overhead with dedicated equipment and facilities, IMU gait analysis offers the potential of quick setup and data capture in a wide variety of gait protocols and environments. As is evident

within the context of this review, however, current knee arthroplasty literature lacks consensus for testing method and performance measures of spatiotemporal gait parameters using wearable technology. Within optical gait analysis, standard reflective marker positions and testing protocols are well established in TKA research[141,142]. Similar progress is needed in the implementation of standard measurements and performance-based tests for characterisation of patient function and recovery in sensor-based analysis. Testing protocols incorporating multiple activities, including those in which patients with knee osteoarthritis find more challenging, may be more indicative measures of patients' ability to return to participation in normal activities and overall function[143]. In my collaborators' clinical experience, many knee patients find activities such as navigating through a crowd of people to be one of the more difficult daily activities. Evaluation of more complex, real-world activities is where the benefit to integrating IMU-based gait analysis may be most valuable.

Aside from full gait analysis, simple functional measures (e.g., joint range of motion) can also be derived using inertial sensors and are candidate applications for IMU-based longitudinal monitoring in this patient population. Ramkumar et al.[144] recently reported a longitudinal study using two IMUs for monitoring knee flexion during daily exercise sessions up to 12 weeks postoperatively. Their method requires the patient to set up one IMU on the thigh and another on the proximal tibia across the operative joint, whereupon the patient performs daily regimented movements. Take-home IMU-based systems for functional assessment are less prevalent than accelerometer-based systems for general activity monitoring, as gyroscopes in 9 degree-of-freedom IMUs have substantial power and data storage requirements[145]. As a result, longitudinal functional

assessment via IMU can be logistically difficult and requires a level of active participation by the patient or caregiver for battery charging and initiation of data collection and transfer. While, Ramkumar et al.[144] demonstrate the feasibility of implementing longitudinal functional assessment of knee range of motion, this process requires significant burden to be transferred onto the patient for daily data collection. Conversely, physical activity monitoring through accelerometry is a more established technique than IMU-based gait analysis. Accelerometers are less battery and memory intensive than IMUs and offer the possibility of extended data collection without active involvement by the patient or caregiver.

One of the more surprising findings of this review is the lack of research using wrist-worn activity monitors for evaluation of patients' physical activity. To date, sensors have been typically deployed near the patient's centre of mass for activity analysis, at the hip or back, to most accurately monitor patients' overall movement[88]. While networks of IMUs may provide improved data quality when performing gait analysis with wearable sensors, a single wrist-based triaxial accelerometer is the most common method of personal, at-home activity tracking. Based on current market research, wrist-based monitors and smart watches account for 95% of commercial wearables sales[63]. Outside of orthopaedics, a movement towards wrist-based sensors for physical activity monitoring has developed[88]. This is likely to be a result of widespread public adoption of commercial fitness trackers and smartwatches[63,146], concerns of patient compliance with hip-based monitors[64,88,147], improved accuracy in wrist-based accelerometry[148–151], and development of techniques for analysis of raw accelerometer data output[88,152]. The largest deployment of population-based physical activity monitoring was the assessment of over 100,000 volunteers

enrolled within the UK Biobank[53]. The study used a triaxial accelerometer (AX3, Axivity) mounted on the participant's dominant wrist to collect activity data continuously for 7 days[53]. When I conducted this study in 2019, no studies had analysed physical activity from knee arthroplasty patients within the UK Biobank; however, significant research potential exists across a variety of fields and disciplines.

Within this review, most studies using wearable sensors in longitudinal cohort studies report incremental changes in physical activity from a preoperative baseline to 6 weeks, 6 months, or 12 months post-op. Within these studies, there has been conflicting evidence as to the level of improvement in physical activity levels of knee arthroplasty patients beyond their preoperative baseline[153,154]. Conversely, limited research has been conducted regarding the pattern of recovery based on physical activity and how the trajectory towards return to participation may inform clinicians of an individual patient's early rehabilitation. Mapping early patient recovery may be a particularly valuable clinical application of activity monitoring after TKA, as patient questionnaires and self-reported patient function may not accurately represent true functional abilities in near-term follow-up[22,155,156]. Only three identified studies employed a method for monitoring patients over a continuous period for more than two weeks following surgery. Recently orthopaedic manufacturer Zimmer Biomet along with corporate partner Apple reported results from their prospective randomised controlled trial examining patient outcomes with personalised coaching through their Apple Watch myMobility platform[157,158]. This multicentre study relied on iPhone and Apple Watch integration to continuously track patient step count, while providing exercise coaching following TKA, UKA, and total hip arthroplasty in an aim provide

motivation during the recovery process. The primary and secondary outcomes of the study were based on 90-day postoperative questionnaire responses, incidence of manipulation under anaesthesia, standard functional assessments, satisfaction, and healthcare costs. The study results showed a reduction of physiotherapy visits, but no difference in functional outcomes in participants with mobile-based coaching. Unfortunately, no physical activity data was reported from participants wearing the Apple Watch during the study period, thus the value in active coaching for increased overall functional mobility remains undescribed in this cohort[158]. Nevertheless, this study demonstrates the potential for activity monitoring and feedback through near-continuous data collection, as well as the continued emergence of wrist-worn devices as a preferred data collection site in postoperative activity studies.

Current literature lacks depth as to the type of sensor-based outcome metric best suited for evaluation of postoperative recovery, and how to best provide feedback to the patient. As Thompson et al.[159] have noted, opportunity exists for physical activity monitors to provide multidimensional metrics to finetune activity guidelines for specific patient populations. Because physical activity is multidimensional, improved monitoring during early rehabilitation can be utilised to explore the rate of recovery, overall mobility, and patient satisfaction based not only on the amount of activity achieved but also the nature and timing of that activity within early postoperative period. The establishment of an evidence-based target recovery curve could be used as a baseline during the entire rehabilitation process from which to identify positively or negatively trending patient outcomes. Feedback mechanisms could be established to encourage more, or less, of a specific type of activity based on the expected activity levels at the specific patient's

postoperative time. In 2019, Panda et al. [160] reported the ability to differentiate oncological patients suffering postoperative complications through passive physical activity monitoring. As they have demonstrated, real-time monitoring of physical activity by means of a sensor paired with a data network could direct the clinical care team to patients exhibiting a suboptimal recovery pattern who might benefit from additional evaluation or follow-up. As Ramkumar et al.[107] explored in their recent pilot study, remote monitoring of patient recovery also has the potential to reduce unnecessary in-person clinical follow-up visits if the clinical care team can remotely identify patients tracking along an established recovery trajectory[107].

The use of accelerometers in instability studies seems to be a niche research area, but one that demonstrates the versatility of wearable motion sensors within postoperative care and research. The development of research techniques, such as the quantitative measurement of the subjective patient experience of joint instability, could be a valuable tool in the refinement of overall evaluation of surgical intervention. The application of wearable technology in the assessment of two competing implant designs, as Soeno et al.[161] have done, adds another tool available in the effort to improve outcomes and overall patient satisfaction.

There are limitations to this scoping review. The use of wearable sensors for the assessment of knee arthroplasty patients is accelerating. Subsequently, several studies will likely be published between the completion of the review and publication of this thesis. Because the use of wearable technology continues to be an emerging technique in knee arthroplasty patient assessment, particularly within the field of gait analysis, exclusion of validation studies based on their level of clinical inquiry was subject to the authors' best judgement. However, this review

focuses on utilisation of sensor-based patient care and reflects emerging techniques practically implemented into current clinical practice.

2.6 Conclusion

The use of wearable sensors in knee arthroplasty research is increasing, with gait analysis and physical activity the two primary modes of investigation. Wearable sensors offer potential for the advancement of patient care to include bespoke rehabilitation and follow-up schedules based on remotely collected ambulatory monitoring of physical activity, range of motion, and gait parameters within the patient's free-living environment. In practice, this could result in reduced clinical visits for patients with on-target recovery markers and opportunities for early intervention in patients following a suboptimal recovery curve. The technology also offers a mechanism for patient coaching and encouragement for patients to reach targeted goals during rehabilitation.

Relative to the first stated aim of this chapter, this review underscores the need for increased interstudy consistency through improved technical reporting and standardisation of sensor-based outcome metrics for the application of wearable sensors in orthopaedical clinical practice and research. Considering this review, Chapter 3 will investigate the role of accelerometer sampling rate on physical activity outcome measures in an effort to enable harmonisation of data collected in studies with differing methodologies. Further, the prevalence of step count reporting in many current studies, paired with the lack of wrist-based data collection among studies identified in this review informed the need to generate a mechanism for validated wrist-based step counting. Chapter 4 seeks to address this need.

The second aim of this chapter was to identify gaps in the orthopaedic literature with the use of wearable sensors. The current findings highlight a lack of focus on early recovery using wearable sensors following knee arthroplasty. The absence of continuous, longitudinal physical activity data in postoperative patients is a missed opportunity for gaining insight into recovery pathways and for the identification and targeting of poorly recovering patients for early clinical intervention. As a result, the primary clinical investigation in this thesis, as presented in Chapter 6 and Chapter 7, utilises methodology tailored to capture and analyse patient physical activity data during the early postoperative recovery period.

Chapter 3: Assessment of Free-Living Activities of Daily Living using Wearables: The Role of Sampling Rate on Valid Interstudy Comparison

3.1 Copyright Notice

With permission from the publisher, this thesis chapter reuses content published in the Journal for the Measurement of Physical Behaviour under the following citation:

Small SR, Khalid S, Dhiman P, Chan S, Jackson D, Doherty AR, Price AJ. Impact of Reduced Sampling Rate on Accelerometer-based Physical Activity Monitoring and Machine Learning Activity Classification. *Journal for the Measurement of Physical Behaviour*, 2021; 4(4):298-310.

3.2 Introduction

Accelerometers are increasingly used to obtain exposure measures in large observational physical activity healthcare studies[52,53,162] and to obtain outcome measures in randomised controlled trials[163]. Seven days of physical activity measurement is a common measurement window across a variety of accelerometer studies[75,164–166]; however, in the case of activity monitoring in clinical populations recovering from trauma or a surgical intervention, extended monitoring well beyond a standard seven day protocol may be desired[76]. Extended longitudinal protocols incorporating measurement windows beyond a few weeks require frequent participant interaction for battery charging, which can be burdensome and undesirable. In these cases, a reduction of sampling rate would offer the potential for longer continuous physical activity measurement, while reducing patient and caregiver interactions with the monitoring device.

Until very recently, Brønd and Arvidsson offered the only side-by-side comparison of accelerometers recording at different sampling rates, recording laboratory-based activities with hip-mounted Actigraph GTX3+ devices[86]. In a newly published manuscript, Clevenger et al.[167] also investigate the effect of sampling rate on physical activity from side-by-side hip-worn ActiGraph wGT3X-BT activity monitors. Other studies comparing sampling rates have generated downsampled data from a single accelerometer to assess the effects of varying sampling rates, while relying on the untested assumption that this postprocessing accurately reproduces the recording behaviour of the underlying hardware[150,168]. In addition, studies investigating sampling rate have been limited to ActiGraph and GENEA triaxial accelerometers[86,150,168,169]. Sampling rate has not been investigated in data collected with the Axivity AX3 accelerometer, which was used in the UK Biobank physical activity cohort study of 100,000 participants[53]. The UK Biobank physical activity cohort is a valuable resource used in Chapter 5 of this thesis to establish reference cohorts for comparison to the clinical cohort described in Chapter 6 and Chapter 7. While the UK Biobank collected accelerometer data at 100 Hz, the extended longitudinal monitoring of my clinical population requires a reduced 25 Hz sampling rate to enable long-term continuous data collection. This chapter was conducted to enable the validated comparison of physical activity data between the 100 Hz UK Biobank accelerometry dataset and datasets integrating reduced, non-standard sampling rates.

Machine learning activity classification is another valuable tool for understanding activity behaviours of a participant cohort. In recent years, two machine learning classification models have been trained using 100 Hz data from Axivity AX3

accelerometers[67,81]. No assessment has been performed on these models using 25 Hz Axivity AX3 data, and thus the use of these models in classifying acceleration data of reduced sampling rate is currently unvalidated.

The aim of this study was to validate data collected at a reduced sampling rate (25 Hz) in comparison to data collected at the commonly used 100 Hz with the AX3 accelerometer. The objectives were to: 1) identify any effect of sampling rate on device-measured activity for both overall activity and specific free-living behaviours; 2) characterise the effect of reduced sampling rate on machine learning activity classification; 3) develop a transformation to enable direct comparison between data collected at standard and reduced sampling rates.

3.3 Methods

3.3.1 Ethical Approval

Ethical approval for participant recruitment was obtained from the Central University Research Ethics Committee of the University of Oxford (Ref: R63137/RE001). Written informed consent was obtained from adult volunteers (aged 18 and above) with no lower limb injury within the previous 6 months and who were able to walk without an assistive device. Participants were recruited through advertising via departmental newsletters, flyers, and word-of-mouth within the University of Oxford, the local community and in senior citizen groups.

3.3.2 Study Design

Participants were instructed to wear four triaxial accelerometers (AX3, Axivity, Newcastle, UK) for 24 hours, except during bathing or swimming activities. Participants were permitted to remove sensors at night if they disrupted sleep. Two accelerometers were placed side-by-side on the dominant wrist using a

wristband, and two on the dominant-side hip, waist level at the anterior-posterior midline via a belt clip (Figure 3-1).

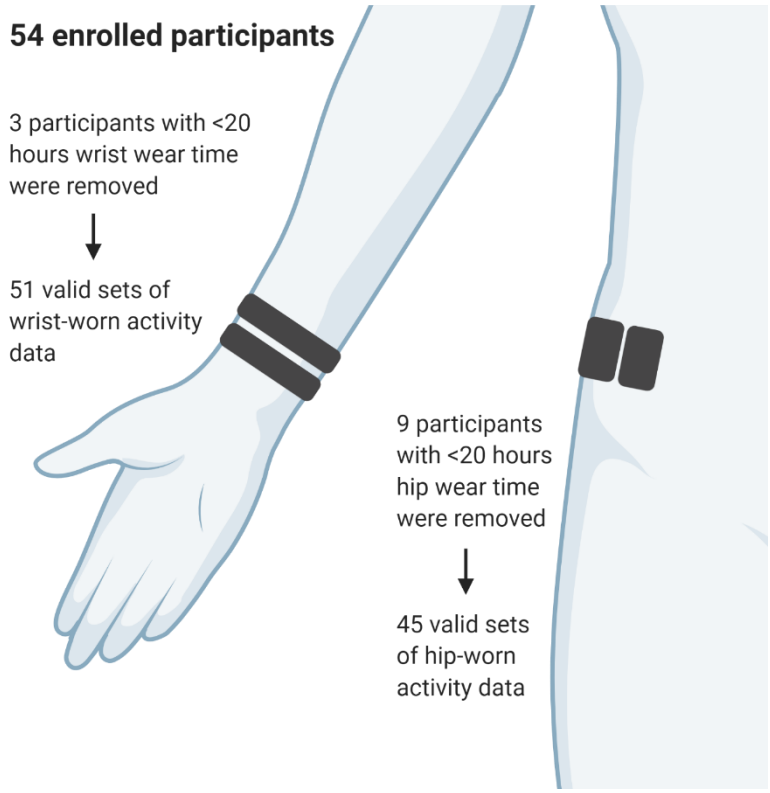


Figure 3-1: Diagram of accelerometer placement and flow chart of participants with valid wear time at each body location. (Created with BioRender)

Accelerometers were synchronised and programmed via the Open Movement software (v1.0.0.42) (https://openmovement.dev). At each body location, one sensor was programmed to collect data at a sampling rate of 25 Hz while the other collected data at 100 Hz, both with a dynamic range of ±8 *g*. The same four unique sensors were used on all participants. Accelerometer orientation and axis alignment within the hip and wrist straps were consistent with manufacturer guidance and verified prior to the start of each recording session. The assignment of hip or wrist body location and sampling rate of each specific accelerometer was randomised by serial number so that the four accelerometers contributed to measurements at both body locations and sampling rates. Participants were

additionally asked to fill out an activity diary noting the times during which they slept, cycled, walked more than 100 meters, ate a meal, participated in self-defined exercise, or removed the accelerometer.

3.3.3 Primary Analysis with the Biobank Accelerometer Analysis Tool

A schematic of the primary study analysis is presented in Figure 3-2. Following the monitoring period, triaxial acceleration data was processed using the open source Biobank Accelerometer Analysis Tool (https://github.com/activityMonitoring/biobankAccelerometerAnalysis) (v3.1, commit 704ea71)[53]. Acceleration data was automatically calibrated by identifying stationary points within the data to reduce sensor-based measurement bias and offset error[53,170]. Overall wear time was calculated and stationary non-wear episodes of 60 minutes or greater wherein all three axes had a standard deviation of less than 13 m*g* were automatically removed from analysis[53]. Data nominally collected at 25 Hz and 100 Hz were resampled to eliminate device sampling errors and generate datasets with precisely 25 and 100 samples per second. A duplicate dataset from the 100 Hz recording was downsampled to 25 Hz for an additional comparison, as implemented in previous studies[168,169]. Overall acceleration was calculated as the Euclidean norm of the three accelerometer axes with 1 *g* subtracted to account for gravity and negative values truncated to zero (ENMO). A fourth-order Butterworth lowpass filter (20 Hz cutoff) was applied to the 100 Hz data to eliminate machine noise. Acceleration metrics were collected into 30 second epochs for analysis and overall 24-hour activity was calculated with data reported in milli-gravitational units (m*g*) following the same processing methodology as was used in the UK Biobank physical activity cohort[53]. ENMO was also calculated for all diary-logged bouts of extended walking, cycling, sleeping,

exercising, and eating. Time and duration of these activities were extracted from participant activity diaries. Beginning and ending times for diary labels were manually verified by inspecting and cross-referencing raw accelerometer data. If no end time was given for eating activities, a conservative estimate of 15 minutes was designated per meal. Overall moderate-to-vigorous physical activity (MVPA) was additionally quantified, defined as time spent at ENMO ≥ 100 m*g* at the wrist and ≥ 70 m*g* at the hip[171].

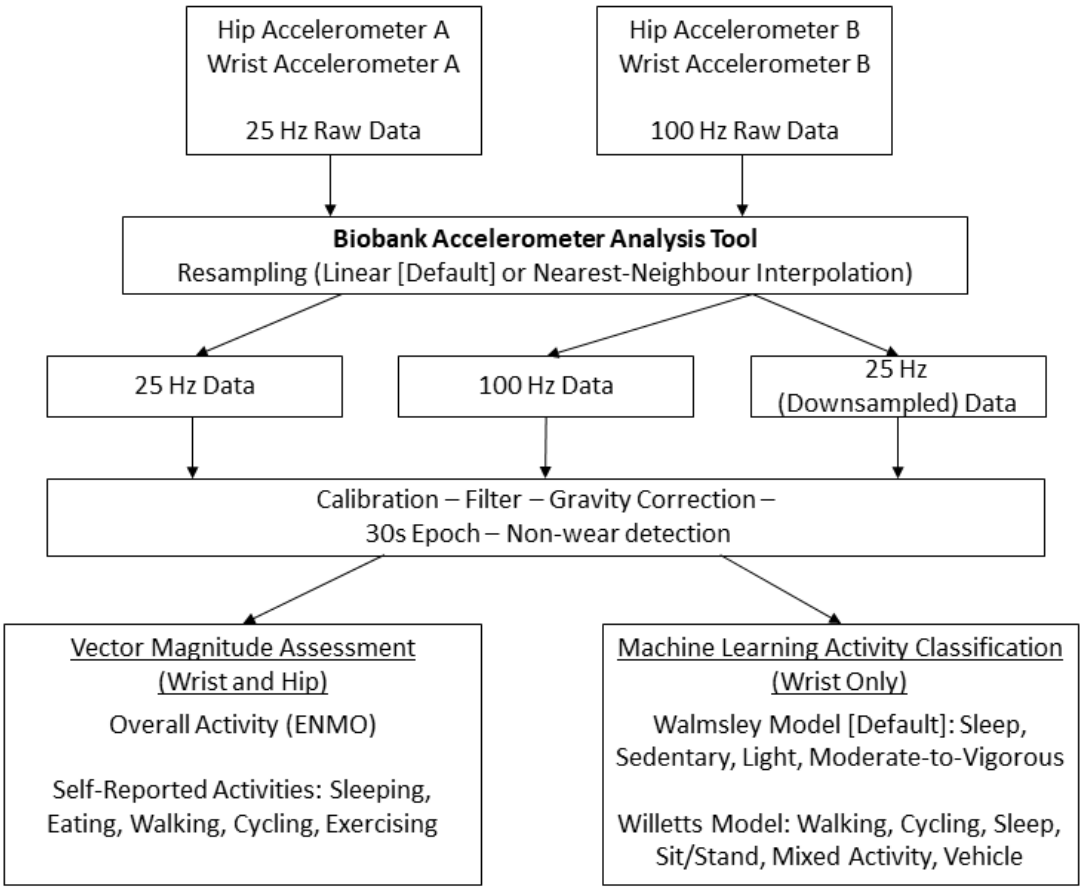


Figure 3-2: Schematic of primary data analysis to inform conversion of accelerometer data collected at 25 Hz to enable comparison with studies collected at 100 Hz using linear and nearest neighbour resampling interpolation methods.

3.3.3.1 Analysis of Resampling Methodology

In the primary analysis, nearest neighbour interpolation was used during resampling to minimize any potential down-smoothing of the raw accelerometer data. However, linear interpolation during resampling of raw accelerometer data is common practice in physical activity data analysis as the methodology in both the UK Biobank physical activity cohort and in the popular GGIR R Package (https://cran.r-project.org/web/packages/GGIR/vignettes/GGIR.html). Additional data analysis was performed using linear interpolation while resampling datasets to assess the effect of interpolation method on activity metric comparability between sampling rates.

3.3.4 Secondary Analysis with the GGIR R Package

A secondary analysis was performed to assess the effect of reduced sampling rate on shorter epoch lengths and on other commonly used activity metrics in the highly popular GGIR R package[85]. This analysis was necessary to ensure any effects attributed to sampling rate were consistent between different analysis packages. A schematic of the workflow for this analysis is presented in Figure 3-3. In this analysis, the GGIR (v.2.3-0) R Package was used to resample and calibrate data, followed by 5 s epoch calculations of ENMO, Lowpass Filtered Euclidean Norm (LFENMO), Highpass Filtered Euclidean Norm (HFEN) and Mean Amplitude Deviation (MAD). Signal filtering utilised a 4th-order Butterworth filter, with a 0.2 Hz cutoff for HFEN and 20 Hz cutoff for LFENMO. Within this manuscript, all data calculated through the Biobank Accelerometer Analysis tool is referred to as ENMO, and all data calculated through GGIR is designated by either ENMO [GGIR], LFENMO, HFEN, or MAD. It should be noted that LFENMO and ENMO

generate very similar metrics and are analogous to the lowpass filtered ENMO as calculated within the Biobank Accelerometer Analysis tool.

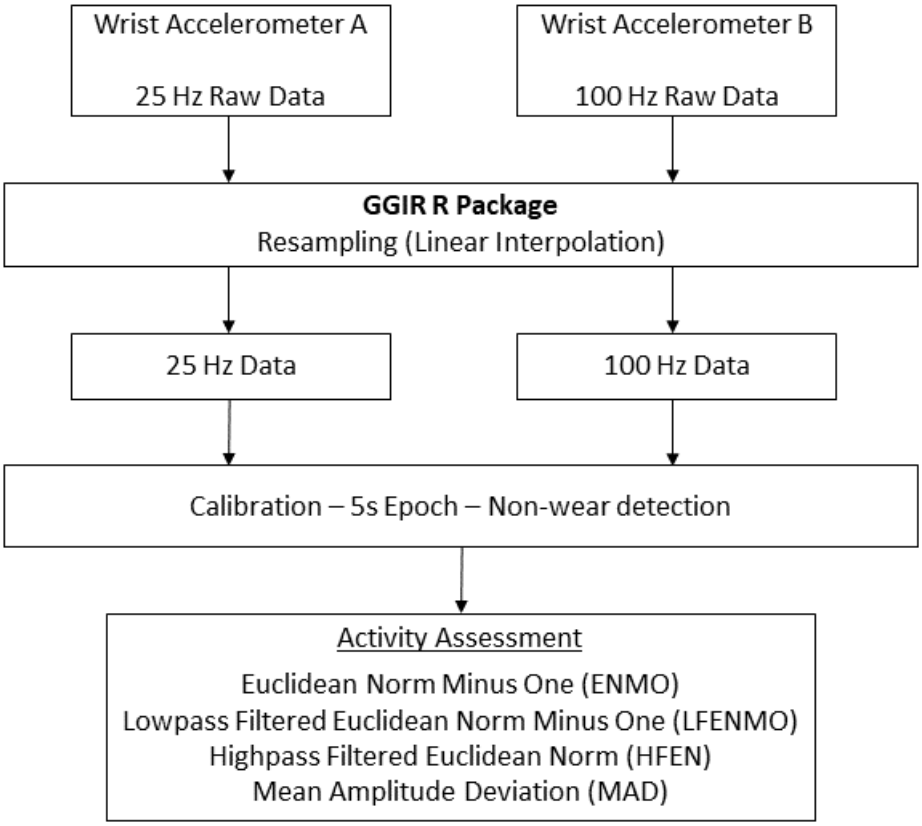


Figure 3-3: Schematic of secondary activity metric data analysis with the GGIR R package to enable comparison of 25 Hz and 100 Hz collected data in 5 s epochs and in alternate activity metrics.

3.3.5 Machine Learning Activity Classification

In addition to the generation of summary activity metrics, accelerometer data can also be used to generate signal features for machine learning activity classification. Machine learning methods for activity classification of wrist-worn accelerometer data were performed using a two-stage machine learning model of balanced random forest and hidden Markov models[67]. The classification models used in the current study are open-source models implemented within the Biobank Accelerometer Analysis tool, having been previously developed and validated for

activity classification of the UK Biobank physical activity cohort[67,81]. The models were trained using data from Axivity AX3 accelerometer recording at 100 Hz (±8 g) on the dominant wrist, with concurrent ground truth images recorded every 20 seconds for a total of 160,000 minutes in 134 participants. Linear resampling during feature extraction was performed in the development of both models and was retained in processing in the current study. Activities were classified in 30 s epochs as time spent cycling, mixed activity, sit/stand, sleep, vehicle, and walking within the Willetts model, and in time spent in sleep, sedentary, light, and moderate-to-vigorous activity using rotation-invariant features in the Walmsley model[83,88,151,172,173].

3.3.6 Statistical Analysis

A minimum sample size of 30 participants was calculated to detect a conservative correlation coefficient estimate of 0.6 between sampling rate and body location measurements with 80% power and alpha set at 0.05. This minimum sample size allowed for 30% data loss for dropout, protocol non-compliance, or poor data quality. Participants with less than 20 hours of wear were excluded from analysis. Summary and descriptive statistics were calculated for participant demographics, overall activity levels, and activity levels performed during self-reported activity designations. Mean ENMO was plotted in 1 hour increments over a 24-hour day to visualise differences in activity patterns recorded at either sampling rate. Two-sided Spearman's rank correlation was calculated to compare activity levels, time in MVPA, and time in classified activities between accelerometer sampling rates and body locations. Mean difference and relative percentage difference in activity level between sampling rates was calculated with reference to 100 Hz measurements. Bland-Altman plots were used to assess fixed or proportional bias

and limits of agreement between sampling rates across the full range of participant activity metrics and between resampling interpolation methods. Linear and 3rd order polynomial regression was used to determine the association between sampling rates and to estimate a conversion factor between acceleration metrics and activity classification recorded at the two sampling rates, with 25 Hz measurements modelled as the predictor and 100 Hz measurements as the regression outcome. Leave-one-participant-out cross validation (*n*-1) was repeated *n* times across each dataset to assess linear model generalisability, with the corresponding root mean square error (RMSE) and R^2 reported. Statistical analysis was performed in R (v.4.0.3) and RStudio (v1.4.1103).

3.4 Results

3.4.1 Participant Demographics

Fifty-four healthy adults (33 female, 21 male) with a mean age of 43.4 years (SD 17.6; range 19.5 - 81.2 years) participated in free-living physical activity analysis (Table 3-1). Acceleration measurement error following self-calibration was <2.6 m*g* across all recordings in the Biobank Accelerometer Analysis tool and <4.1 m*g* in GGIR processing, with no instances of calibration failure or device malfunction. Median hip and wrist accelerometer wear time was 23.7 hours (IQR 0.6) and 24.0 hours (IQR 0.5), respectively. Three sets of wrist-worn and nine sets of hip-worn sensors were removed from analysis as they were worn for less than 20 hours. Participants self-reported an overall mean of 7.4 hours of sleep (SD 1.0), 46.8 minutes of eating (SD 37.8), 28.4 minutes of general exercise (SD 56.7), 58.2 minutes of extended walking (SD 64.0), and 34.1 minutes of cycling (SD 27.8) over the 24-hour assessment period as calculated in participants who logged non-zero time in those activities.

Table 3-1: Participant Demographics

	Mean	SD	Sex	n	%	Handedness	n	%
Age (years)	43.4	17.3	Female	33	61.1%	Left	8	14.8%
Height (cm)	170.5	10.8	Male	21	38.9%	Right	46	85.2%
Weight (kg)	70.5	13.3						
BMI (kg/m^2)	24.2	3.6						

3.4.2 24-Hour Physical Activity Assessment

Mean physical activity across all valid measurements according to time of day is presented in Figure 3-4 and 30 s epoch ENMO for wrist and hip accelerometers are presented in Table 3-2. Strong correlation was observed between 25 Hz and 100 Hz recorded data in overall ENMO and time spent in MVPA at both the hip (ρ = 0.99 to 1.00) and the wrist (ρ = 0.98 to 0.99). At the hip, ENMO collected at 25 Hz resulted in consistently lower overall activity compared to the 100 Hz data, with an absolute mean difference in ENMO (SD) of 0.5 (1.4) m*g* and 4.9 (10.0) minutes of MVPA per 24 hours. This represents a -3.1% relative difference in 24-hour activity and -5.9% difference in MVPA time when 25 Hz measurements when compared to 100 Hz measurements at the hip. At the wrist, 25 Hz data collection also resulted in consistently lower overall activity compared to the 100 Hz data, with an absolute mean difference (SD) of 1.9 (2.4) m*g* in ENMO and 13.1 (10.5) minutes of MVPA, a -5.7% and -11.5% relative difference in 25 Hz measurements at the wrist, respectively.

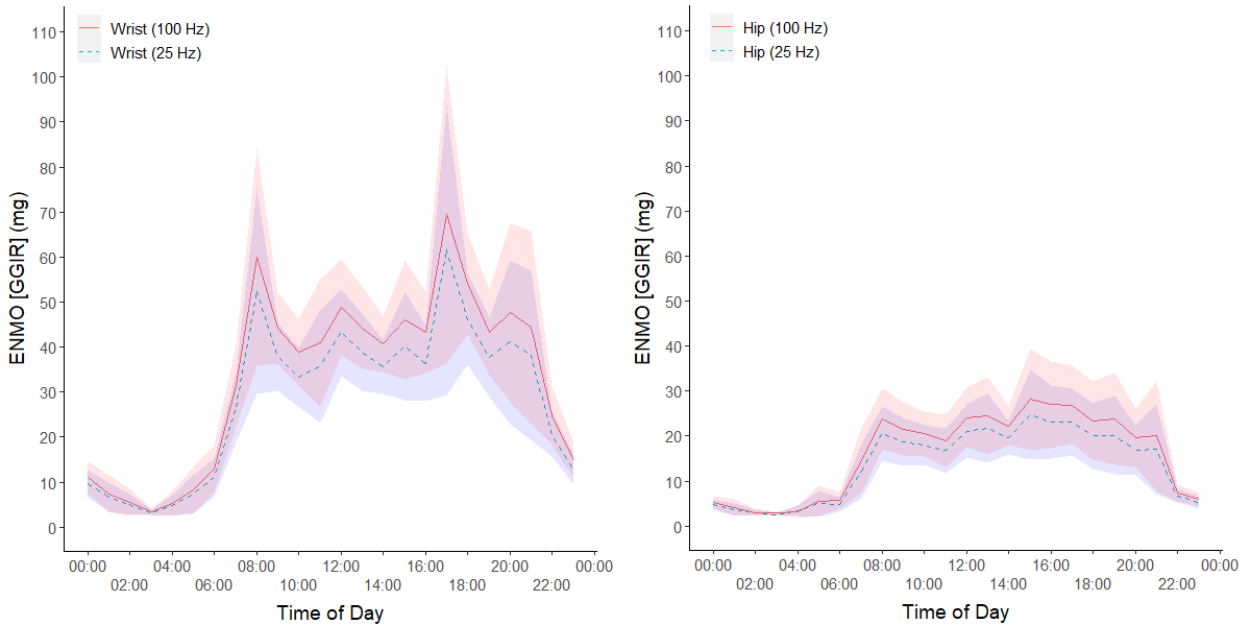


Figure 3-4: Plot of mean (95%CI) 24-hour activity in 1-hour epochs as measured by ENMO [GGIR] with linear interpolation resampling across a convenience sample of 51 participants at each of the four sensor location/rate combinations.

Table 3-2: Accelerometer Measured Physical Activity for Overall and Diary-Reported Activities Over a Single 24-Hour Period (30 Second Epoch)

	ENMO (m*g*)	MVPA (min)	Cycling (m*g*)	Eating (m*g*)	Exercising (m*g*)	Sleeping (m*g*)	Walking (m*g*)
Wrist Placement	*n = 51*	*n = 51*	*n = 13*	*n = 44*	*n = 16*	*n = 51*	*n = 44*
25 Hz Mean [SD]	31.3 [12.0]	100.9 [52.1]	91.9 [28.3]	29.6 [13.5]	317.9 [321.7]	3.7 [0.8]	134.1 [55.4]
100 Hz Mean [SD]	33.2 [12.1]	114.0 [54.3]	126.9 [58.0]	31.8 [13.9]	331.2 [327.5]	3.2 [0.8]	136.3 [58.7]
Mean difference [SD] 25 Hz -100 Hz	-1.9 [2.4]	-13.1 [10.5]	-35.0 [17.9]	-2.2 [2.9]	-13.3 [114.8]	0.5 [0.2]	-2.2 [12.2]
Spearman's correlation	0.98	0.98	0.98	0.97	1.00	0.92	0.98
ICC	0.99	0.98	0.74	0.99	1.00	0.95	0.98
	ENMO (m*g*)	MVPA (min)	Cycling (m*g*)	Eating (m*g*)	Exercising (m*g*)	Sleeping (m*g*)	Walking (m*g*)
Hip Placement	*n = 45*	*n = 45*	*n = 11*	*n = 40*	*n = 14*	*n = 45*	*n = 39*
25 Hz Mean [SD]	15.6 [6.8]	77.6 [47.1]	57.4 [18.7]	10.2 [7.0]	109.0 [115.6]	2.8 [0.4]	114.0 [48.0]
100 Hz Mean [SD]	16.1 [7.2]	82.5 [47.7]	60.6 [18.8]	10.7 [7.6]	119.6 [128.9]	2.3 [0.3]	121.7 [51.3]
Mean difference [SD] 25 Hz -100 Hz	-0.5 [1.4]	-4.9 [10.0]	-3.2 [7.4]	-0.5 [1.6]	-10.6 [43.3]	0.1 [0.1]	-7.7 [10.6]
Spearman's correlation	0.99	0.99	0.99	0.98	0.99	0.55	0.99
ICC	0.99	1.00	0.99	0.99	0.99	0.61	1.00

MVPA reported in minutes calculated by 30s epochs with ENMO ≥ 100 mg at the wrist and ≥70 mg at the hip

Bland-Altman plots exhibiting the difference between 25 Hz and 100 Hz 24-hour ENMO are presented in Figure 3-5. A secondary analysis of ENMO [GGIR], LFENMO, HFEN, and MAD calculated in 5 s epochs in GGIR are presented in Table 3-3. Differences between these 25 Hz and 100 Hz activity metrics calculated with linear interpolation resampling at 5 s epochs with the GGIR package are more pronounced than in the primary analysis. In this secondary analysis, 25 Hz data capture resulted in 24-hour relative differences of -13.9% in ENMO [GGIR], 10.6% in LFENMO, -9.6% in HFEN, and -13.3% in MAD metrics.

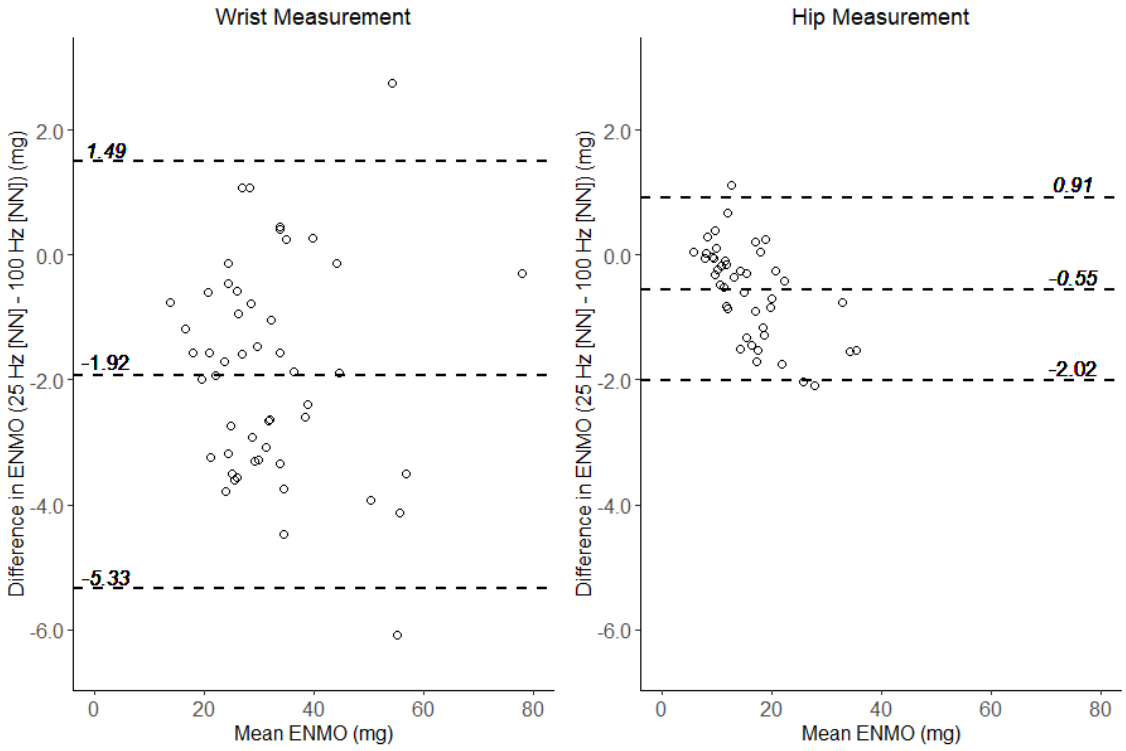


Figure 3-5: Bland-Altman plots comparing the mean 24-hour ENMO and differences between the 25 Hz and 100 Hz sensors at the wrist and hip in a convenience sample of 51 adults. Dashed lines indicated mean bias and 95% limits of agreement. Negative bias on the y-axis indicates lower values in the 25 Hz sensor. Both 25 Hz and 100 Hz data calculated using nearest neighbour interpolation resampling.

Table 3-3: Secondary Analysis of Alternative Physical Activity Metrics as Calculated on Wrist Sensors Over 5 Second Epochs in the GGIR R Package

	ENMO (mg)	LFENMO (mg)	HFEN (mg)	MAD (mg)
25 Hz Mean [SD]	28.6 [12.2]	28.7 [12.2]	90.7 [24.4]	41.6 [13.0]
100 Hz Mean [SD]	33.2 [12.7]	32.1 [12.6]	100.4 [25.8]	48.0 [14.4]
Mean difference [SD] 25 Hz -100 Hz	-4.6 [2.7]	-3.4 [2.7]	-9.7 [5.5]	-6.4 [3.0]
Spearman's correlation	0.97	0.97	0.98	0.98
ICC	0.96	0.97	0.95	0.94

At both hip and wrist locations, effectively perfect correlation (ρ = 1.000) was observed for overall ENMO between the recorded 100 Hz and the downsampled measurements from the same sensor. Comparison of sampling rate via the downsampling of 100 Hz data to 25 Hz in the same sensor resulted in less than 0.1% difference in overall activity and MVPA across both the hip and wrist. Bland-Altman plots comparing 100 Hz data to downsampled 25 Hz data (Figure 3-6) demonstrate no mean bias between data and limits of agreement ranging no greater than ± 0.05 m*g* at either the wrist or hip sensor locations.

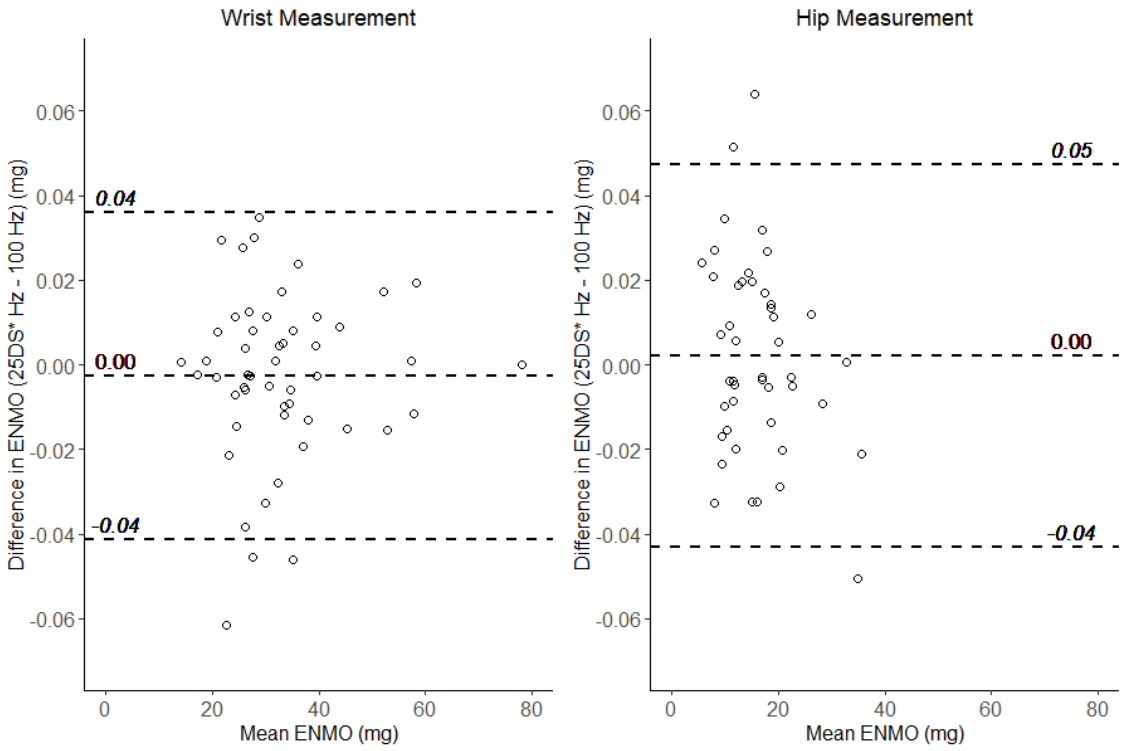


Figure 3-6: Bland-Altman plots comparing the mean 24-hour ENMO and differences between the downsampled 25 Hz acceleration and 100 Hz raw acceleration from sensors recording at 100 Hz at the wrist and hip. Dashed lines indicated mean bias and 95% limits of agreement. Negative bias on the y-axis indicates lower values in the downsampled 25 Hz data.

3.4.3 Intensity Measurement of Self-Reported Activities

Correlation between acceleration ENMO measured at 25 Hz and 100 Hz sampling rates was high during all self-reported non-sleeping activities at the wrist (ρ = 0.97 to 1.00) and hip (ρ = 0.94 to 0.99) (Table 3-2). With nearest neighbour resampling, the 25 Hz sampling rate at the hip resulted in a mean relative difference in ENMO of -4.7% to -8.9% when compared to the 100 Hz sampling rate across all non-sleeping activities. The greatest difference in 25 Hz and 100 Hz data at the wrist was observed during cycling activity, where the 25 Hz accelerometer recorded relative difference of -27.6% when compared to the 100 Hz recording. Other non-sleeping activities resulted in a -1.6% to -6.9% relative difference when recorded at 25 Hz at the wrist. ENMO during sleep was predictably low, with lower correlation between sampling rates when compared to other activities (ρ = 0.55 to 0.95). Actual differences are small due to the limited activity measured during sleep intervals.

3.4.4 Machine Learning Activity Classification

Results of the machine learning activity classification are presented in Table 3-4 and Table 3-5. In Willetts classifications (Table 3-4), periods of walking, sleep, sit/stand activities, and mixed activities were identified in all participants, whereas periods of cycling and vehicle activities were identified in 13 and 41 participants, respectively. In all participants who had cycling or vehicle activities classified within the 100 Hz 24-hour recording, that activity was also identified within the 25 Hz data. Notably, reduced sampling rate resulted in a 7.2% increase in estimated sit/stand time and a 10.1% decrease in estimated vehicle time. Within the more robust, rotation-invariant Walmsley model (Table 3-5), 25 Hz data collection resulted in increased classification of sedentary activity at the expense of time

classified in both sleep and light activities. An 18.2% increase in sedentary time classification in 25 Hz data was countered by a 12.0% reduction in classified sleep and a 14.8% reduction in classified time in light activities. Classification of MVPA was consistent between the two sampling rates.

Table 3-4: Machine Learned Activity Classification for Wrist Accelerometers in a Convenience Sample of 51 Adults Over a Single 24-Hour Period (Willetts Model)

	Cycling (min) *n = 13*	Mixed (min) *n = 51*	Sit/Stand (min) *n = 51*	Sleep (min) *n = 51*	Vehicle (min) *n = 41*	Walking (min) *n = 51*
25 Hz Mean Classified Time [SD]	28.6 [19.1]	161.9 [121.3]	595.7 [142.2]	452.8 [64.4]	70.4 [56.2]	161.5 [87.1]
100 Hz Mean Classified Time [SD]	28.5 [19.3]	164.1 [127.9]	555.2 [148.5]	475.9 [73.9]	78.3 [55.9]	170.4 [86.7]
Mean difference [SD] 25 Hz -100 Hz	0.1 [7.5]	-2.2 [24.7]	40.5 [28.8]	-23.1 [13.7]	-7.9 [12.4]	-8.9 [17.2]
Spearman's correlation	0.99	0.91	0.93	0.86	0.92	0.89
ICC	0.99	0.96	0.96	0.84	0.92	0.92

Table 3-5: Machine Learned Activity Classification for Wrist Accelerometers in a Convenience Sample of 51 Adults Over a Single 24-Hour Period (Walmsley Model)

	Sleep (min) *n = 51*	Sedentary (min) *n = 51*	Light (min) *n = 51*	MVPA (min) *n = 48*
25 Hz Mean Classified Time [SD]	415.1 [63.0]	683.4 [129.6]	280.8 [133.1]	60.8 [41.5]
100 Hz Mean Classified Time [SD]	471.6 [61.5]	578.2 [143.9]	329.7 [147.3]	60.5 [42.4]
Mean difference [SD] 25 Hz -100 Hz	-56.5 [12.3]	105.2 [27.1]	-48.9 [27.8]	0.3 [8.6]
Spearman's correlation	0.82	0.94	0.96	0.98
ICC	0.85	0.93	0.96	0.99

A comparison of rotation-invariant machine learning features as calculated from 25 Hz and 100 Hz wrist data is presented in Table 3-6. Mean values across all participants and epochs demonstrate 19.8% lower maximum vector magnitude, 34.8% lower skewness, 0.2% to 3.5% lower Fourier transform coefficients corresponding to 1-11 Hz and 4.7% to 6.5% lower values in the first twelve Fourier

coefficients in 25 Hz collected data. Epoch minimum and mean power deviation values were calculated 34.3% and 80.9% higher in the 25 Hz compared to 100 Hz collected data.

Table 3-6: Comparison of Rotation-Invariant Machine Learning Features Between 100 Hz and 25 Hz Data Collection

Feature	100 Hz Mean [SD]	25 Hz Mean [SD]	% Difference (25 Hz-100 Hz) [SD]
ENMO (Truncated)	0.03 [0.08]	0.03 [0.07]	-4.6% [1.4%]
Mean	0.24 [0.28]	0.24 [0.27]	-1.7% [5.2%]
Standard deviation	0.13 [0.14]	0.13 [0.14]	-3.5% [2.7%]
Coefficient of variation	0.70 [0.46]	0.67 [0.41]	-3.7% [8.4%]
Median	0.21 [0.25]	0.20 [0.25]	-1.4% [4.8%]
Minimum	0.03 [0.05]	0.04 [0.06]	34.3% [1.1%]
Maximum	1.09 [1.40]	0.87 [1.00]	-19.8% [23.4%]
25th percentile	0.16 [0.19]	0.16 [0.19]	-1.1% [3.7%]
75th percentile	0.30 [0.34]	0.29 [0.34]	-1.7% [6.6%]
1 second autocorrelation	0.29 [0.28]	0.29 [0.28]	0.8% [5.4%]
Dominant frequency	1.14 [3.02]	1.12 [2.46]	-2.2% [53.0%]
Power of dominant frequency	-10.41 [4.53]	-10.44 [4.52]	0.4% [87.1%]
Dominant frequency (0.3 to 5 Hz)	0.79 [0.63]	0.87 [0.72]	10.5% [12.9%]
Dominant frequency power (0.3 to 5 Hz)	-12.26 [3.97]	-12.18 [3.81]	-0.6% [74.8%]
Spectral entropy	0.56 [0.21]	0.60 [0.20]	7.5% [3.9%]
Fourier transform coefficient (1 Hz)	-3.38 [1.98]	-3.37 [1.91]	-0.2% [37.4%]
Fourier transform coefficient (2 Hz)	-4.04 [1.98]	-3.99 [1.90]	-1.3% [37.3%]
Fourier transform coefficient (3 Hz)	-6.05 [1.93]	-5.84 [1.72]	-3.5% [35.1%]
Fourier transform coefficient (4 Hz)	-6.25 [1.82]	-6.07 [1.60]	-2.9% [32.9%]
Fourier transform coefficient (5 Hz)	-6.34 [1.70]	-6.20 [1.51]	-2.1% [30.9%]
Fourier transform coefficient (6 Hz)	-6.43 [1.61]	-6.31 [1.44]	-1.8% [29.4%]
Fourier transform coefficient (7 Hz)	-6.56 [1.60]	-6.41 [1.39]	-2.3% [28.8%]
Fourier transform coefficient (8 Hz)	-6.65 [1.57]	-6.49 [1.34]	-2.5% [28.1%]
Fourier transform coefficient (9 Hz)	-6.71 [1.54]	-6.55 [1.31]	-2.4% [27.5%]
Fourier transform coefficient (10 Hz)	-6.77 [1.51]	-6.60 [1.30]	-2.4% [26.9%]
Fourier transform coefficient (11 Hz)	-6.84 [1.47]	-6.65 [1.25]	-2.7% [26.2%]
Mean absolute deviation	0.05 [0.09]	0.05 [0.08]	-5.8% [1.6%]
Mean power deviation	0.00 [0.00]	0.00 [0.00]	80.9% [0.0%]
Skewness	1.39 [3.21]	0.91 [2.71]	-34.8% [57.2%]
Kurtosis	-2.99 [0.03]	-2.98 [0.06]	-0.4% [0.9%]
Dominant frequency (0.3 to 15 Hz)	3.13 [3.10]	3.01 [2.90]	-3.9% [57.7%]
Dominant frequency power (0.3 to 15 Hz)	-13.75 [3.74]	-13.65 [3.55]	-0.8% [70.1%]
Second dominant frequency (0.3 to 5 Hz)	3.32 [3.17]	3.17 [2.96]	-4.8% [59.0%]

Dominant frequency (0.6 to 2.5 Hz)	1.47 [0.55]	1.48 [0.56]	1.2% [10.7%]
Dominant frequency power (0.6 to 2.5 Hz)	-14.00 [3.77]	-13.86 [3.57]	-1.0% [70.7%]
Total power	-10.46 [3.91]	-10.24 [3.57]	-2.2% [72.0%]
First Fourier coefficient	0.01 [0.02]	0.008 [0.019]	-4.7% [0.4%]
Second Fourier coefficient	0.00 [0.01]	0.04 [0.01]	-4.9% [0.2%]
Third Fourier coefficient	0.00 [0.00]	0.00 [0.00]	-6.5% [0.1%]
Fourth Fourier coefficient	0.00 [0.00]	0.00 [0.00]	-6.3% [0.1%]
Fifth Fourier coefficient	0.00 [0.00]	0.00 [0.00]	-6.1% [0.0%]
Sixth Fourier coefficient	0.00 [0.00]	0.00 [0.00]	-6.0% [0.0%]
Seventh Fourier coefficient	0.00 [0.00]	0.00 [0.00]	-5.8% [0.0%]
Eighth Fourier coefficient	0.00 [0.00]	0.00 [0.00]	-5.4% [0.0%]
Ninth Fourier coefficient	0.00 [0.00]	0.00 [0.00]	-5.2% [0.0%]
Tenth Fourier coefficient	0.00 [0.00]	0.00 [0.00]	-5.1% [0.0%]
Eleventh Fourier coefficient	0.00 [0.00]	0.00 [0.00]	-5.0% [0.0%]
Twelfth Fourier coefficient	0.00 [0.00]	0.00 [0.00]	-4.5% [0.0%]

Italicised features are calculated using a lowpass filtered 3-axis acceleration vector magnitude, $\sqrt{x^2 + y^2 + z^2}$ *, as described by Ellis et al.* [83]

3.4.5 Linear and Nearest Neighbour Resampling Methods

Bland-Altman plots comparing sampling rate interpolation methods are presented in Figure 3-7 and Figure 3-8. Figure 3-7 demonstrates the difference in mean bias at the wrist when calculating ENMO of 25 Hz data using either nearest neighbour or linear interpolation resampling in comparison to the standard 100 Hz linear interpolation resampling. In this instance the nearest neighbour interpolation of the 25 Hz data reduces the mean bias between sampling rates from -4.39 m*g* to -1.84 m*g*. A parallel comparison between interpolation methods of data collected at the hip is presented in Figure 3-8, with nearest neighbour interpolation resampling reducing mean bias between 25 Hz and 100 Hz data from -2.04 m*g* to -0.53 m*g*.

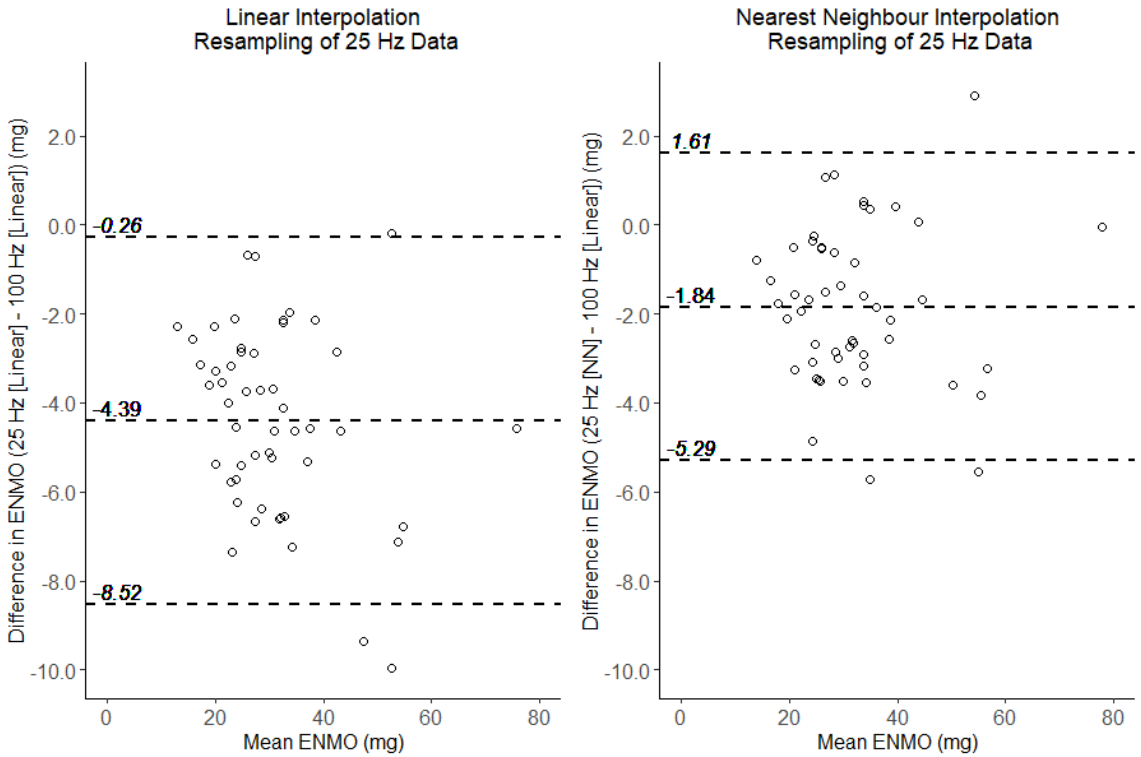


Figure 3-7: Bland-Altman plots comparing the mean 24-hour ENMO and differences between the 25 Hz and 100 Hz sensors at the wrist. Dashed lines indicated mean bias and 95% limits of agreement. Negative bias on the y-axis indicates lower values in the 25 Hz sensor. 25 Hz data was resampled using linear interpolation on the left plot and nearest neighbour [NN] interpolation on the right plot. In both instances 100 Hz data calculated using standard linear interpolation resampling, as would be the case when comparing to a reference 100 Hz dataset.

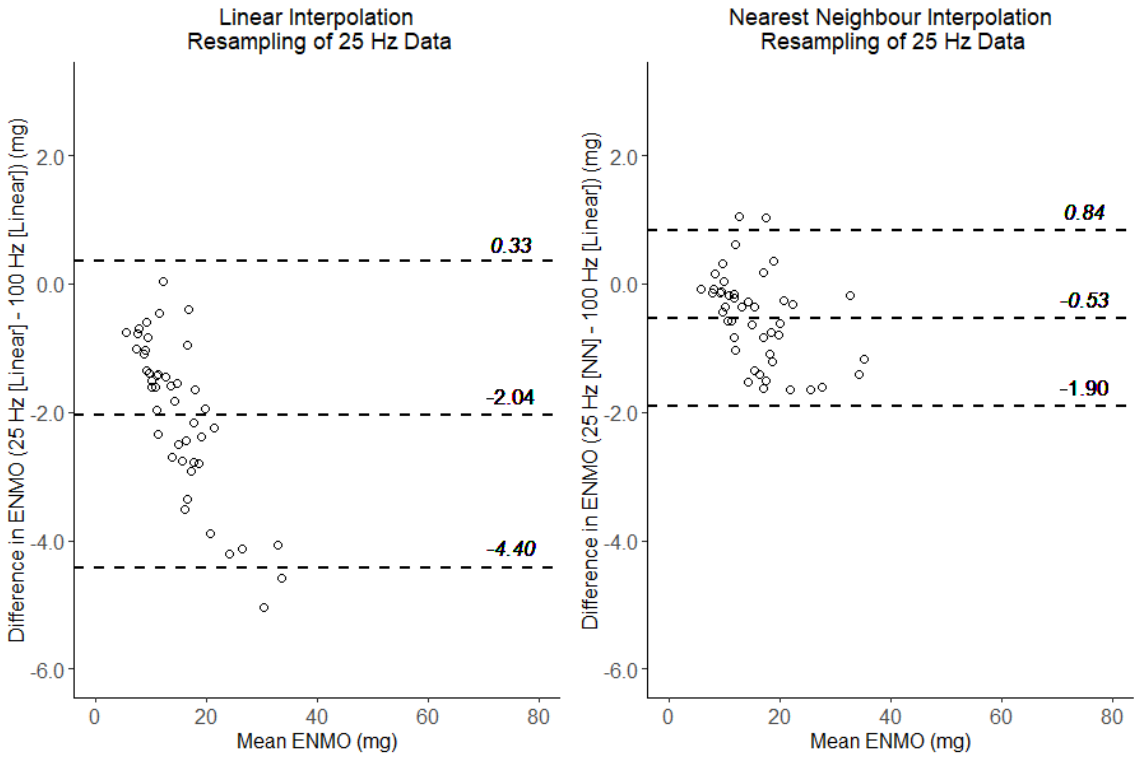


Figure 3-8: Bland-Altman plots comparing the mean 24-hour ENMO and differences between the 25 Hz and 100 Hz sensors at the hip. Dashed lines indicated mean bias and 95% limits of agreement. Negative bias on the y-axis indicates lower values in the 25 Hz sensor. 25 Hz data was resampled using linear interpolation on the left plot and nearest neighbour [NN] interpolation on the right plot. In both instances 100 Hz data calculated using standard linear interpolation resampling, as would be the case when comparing to a reference 100 Hz dataset.

3.4.6 Data Conversion for Interstudy Comparison

Results from the linear regression and cross validation relating sampling rates based on overall 24-hour activity metrics and activity classification models are presented in Table 3-7 and Table 3-8. Coefficients beyond the 1st order in all activity metric polynomial regression models were statistically insignificant ($p>0.05$), thus only linear transformations are reported. At the hip and wrist, the linear model of sampling rate accounted for between 97% and 99% of the adjusted variability in the linearly interpolated ENMO as described by R^2. N-1 cross validation

resulted in RSME values of 6.5% and 4.7% of the mean wrist and hip ENMO, respectively. Transformations for ENMO and combinations of nearest neighbour and linear interpolation, as well as for alternate metrics in 5 s epochs as calculated in GGIR are additionally reported with high cross-validated R^2.

Transformations from 25 Hz to 100 Hz data were generated for classifications in both Willetts and Walmsley models. In wrist-based activity classification, 69% to 97% of the variability between sampling rates was accounted for in the Willetts model, with cross validation RMSE representing 7.5% of the mean time in the sit/stand classification and up to 29.9% the mean vehicle time. Within the Walmsley model incorporating rotation-invariant features, 70.7% to 92.4% of variability was accounted for, with RMSE representing 7.0%, 8.8%, 12.2%, and 10.7% of mean sleep, sedentary, light, and MVPA time.

Table 3-7: Linear Regression for Comparison of 25 Hz and 100 Hz 24-hour AX3 Accelerometer Data

	Coefficient [SE]	95% CI	Intercept [SE]	95% CI	R^2	RMSE
ENMO (BiobankAA) [30 s epoch] (m*g*) 25 Hz LI to 100 Hz LI Resampling						
Wrist	1.038 [0.026]	(0.986, 1.090)	3.310 [0.798]	(1.706, 4.914)	0.967	2.155
Hip	1.158 [0.019]	(1.120, 1.195)	-0.180 [0.283]	(-0.751, 0.391)	0.988	0.757
ENMO (BiobankAA) [30 s epoch] (m*g*) 25 Hz NNI to 100 Hz LI Resampling						
Wrist	0.989 [0.021]	(0.947, 1.031)	2.179 [0.699]	(0.775, 3.584)	0.976	1.832
Hip	1.042 [0.014]	(1.013, 1.071)	-0.121 [0.245]	(-0.615, 0.372)	0.991	0.659
ENMO (BiobankAA) [30 s epoch] (m*g*) 25 Hz NNI to 100 Hz NNI Resampling						
Wrist	0.997 [0.021]	(0.956, 1.039)	2.000 [0.693]	(0.608, 3.392)	0.977	1.824
Hip	1.062 [0.014]	(1.034, 1.090)	-0.415 [0.237]	(-0.892, 0.063)	0.992	0.637

	Coefficient [SE]	95% CI	Intercept [SE]	95% CI	R^2	RMSE
ENMO (GGIR) [5 s epoch] - 25 Hz to 100 Hz (m*g*)						
Wrist	1.025 [0.032]	(0.961, 1.089)	3.684 [0.988]	(1.685, 5.683)	0.960	2.536
LFENMO (GGIR) [5 s epoch] - 25 Hz to 100 Hz (m*g*)						
Wrist	1.014 [0.030]	(0.954, 1.074)	2.988 [0.927]	(1.113, 4.862)	0.964	2.368
HFEN (GGIR) [5 s epoch] - 25 Hz to 100 Hz (m*g*)						
Wrist	1.042 [0.031]	(0.980, 1.104)	5.862 [2.873]	(0.051, 11.673)	0.964	4.851
MAD (GGIR) [5 s epoch] - 25 Hz to 100 Hz (m*g*)						
Wrist	1.096 [0.025]	(1.036, 1.156)	2.365 [1.292]	(-0.274, 4.978)	0.969	2.488

R^2 and RMSE values are reported from leave-one-participant-out cross validation
BiobankAA: Biobank Accelerometer Analysis tool
NNI: Nearest Neighbor Interpolation
LI: Linear Interpolation

Table 3-8: Linear Regression to Convert Wrist-Based Machine Learning Activity Classification Collected at 25 Hz for Comparison to Data Collected at 100 Hz

	Coefficient [SE]	95% CI	Intercept [SE]	95% CI	R^2	RMSE
Willetts Classification Model (min) 25 Hz to 100 Hz						
Cycling	0.996 [0.023]	(0.950, 1.042)	-0.26 [0.51]	(-1.28, 0.76)	0.972	3.14
Mixed Activity	1.013 [0.042]	(0.929, 1.097)	0.12 [8.42]	(-16.79, 17.03)	0.919	36.1
Sit/Stand	1.005 [0.040]	(0.927, 1.095)	-43.41 [24.75]	(-99.88, 2.45)	0.920	41.5
Sleep	0.974 [0.087]	(0.840, 1.250)	34.82 [39.63]	(-73.65, 100.33)	0.687	40.3
Vehicle	0.914 [0.063]	(0.769, 1.049)	13.87 [5.66]	(5.09, 29.34)	0.822	23.4
Walking	0.916 [0.056]	(0.804, 1.028)	22.4 [10.18]	(1.94, 42.86)	0.831	35.3
	Coefficient [SE]	95% CI	Intercept [SE]	95% CI	R^2	RMSE
Walmsley Classification Model (min) 25 Hz to 100 Hz						
Sleep	0.835 [0.073]	(0.629, 0.962)	124.94 [30.45]	(72.3, 210.8)	0.707	33.0
Sedentary	1.043 [0.054]	(0.952, 1.161)	-134.4 [37.79]	(-216.5, -70.1)	0.873	50.7
Light	1.066 [0.042]	(1.005, 1.138)	30.35 [13.07]	(12.40, 50.17)	0.924	40.1
MVPA	1.011 [0.022]	(0.967, 1.055)	-0.923 [1.61]	(-4.16, 2.32)	0.976	6.5

R^2 and RMSE values are reported from leave-one-participant-out cross validation.

Linear interpolation resampling was used in both 25 Hz and 100 Hz datasets

3.4.6.1 Example Interstudy Acceleration Transformation

An example of activity metric transformation between 25 Hz collected data and a 100 Hz reference dataset is as follows:

A 68-year-old female participant has 7-day mean overall acceleration magnitude, ENMO, of 24.8 m*g* as recorded at 25 Hz on the dominant wrist with an AX3 accelerometer (Acc_{25}). Researchers wish to compare this individual to the baseline 65-74 year-old female in the UK Biobank database 26.6 (SD 7.1) m*g*, recorded at 100 Hz with an AX3 accelerometer (Acc_{100})[53]. Without sampling rate conversion, a 24.8 m*g* activity level is below average for her age group, and places this patient in the 40th percentile of activity in the 65-74 year-old female cohort. With 25 Hz to 100 Hz conversion following the wrist-based regression model with default linear interpolation resampling, $Acc_{100} = 1.038*Acc_{25} + 3.310$, the participant's comparable acceleration magnitude is 29.1 m*g*, placing her in the 64th percentile of the appropriate demographic cohort. Conversely, if 25 Hz data is resampled using nearest neighbour interpolation (Acc_{25nn}), the appropriate conversion for comparison to the linearly resampled 100 Hz reference database would be $Acc_{100} = 0.989*Acc_{25nn} + 2.179$.

3.5 Discussion

Outcome measures based on acceleration magnitude, cut-point thresholds, and activity classification are commonly used in clinical and population studies of physical activity[53,62,71]. To validate the use of reduced sampling rate during physical activity measurement, acceleration metrics and activity classification were assessed across body location and self-reported activity types. Reduction in recorded sampling rate from 100 Hz to 25 Hz resulted in small, but consistently lower measured overall activity at both hip and wrist body locations across all

generated activity metrics in both the Biobank Accelerometer Analysis and GGIR processing tools. Linear interpolation produced a larger discrepancy between sampling rates, particularly when applied to the 25 Hz collected data. Reduced sampling rate resulted in large differences in several rotation-invariant signal features used for machine learning activity classification, and subsequently resulted in increased pooling of data into sedentary activity classification at the expense of sleep and light activity classifications. While both acceleration magnitude and activity classification are highly correlated between sampling rates, subtle, yet real differences in outcome measures were observed. Transformation of 25 Hz data can enable extended activity monitoring protocols while retaining comparability to data collected at 100 Hz.

In the current study of healthy adults in the free-living environment, I found a repeatable difference in activity metrics and MVPA in a side-by-side comparison of data collection at 25 Hz and 100 Hz. When using a reduced sampling rate, a 3.1% to 13.9% lower 24-hour activity was observed when compared to 100 Hz data collection at the hip or wrist. Consistently lower ENMO was also observed in the 25 Hz recording across diary-logged free-living activities. Similar to the current study, Brønd and Arvidsson[86] found ActiGraph GT3x counts demonstrated small differences between sampling rates (30 Hz, 40 Hz, and 100 Hz) during walking, yet large differences between sampling rates during running, with up to 3,000 counts per minute greater intensity recorded in the faster sampling rates. Two other prior investigations of accelerometer sampling rate relied on downsampling accelerometer data from a single sensor, with mixed results and frequently little difference between sampling rates when using raw acceleration-based metrics[168,169]. In this study I conducted a secondary analysis by downsampling

data initially collected at 100 Hz, which also resulted in no difference in 24-hour activity and MVPA at a lower sampling rate, even while raw data collection at 25 Hz demonstrated a reduction in the activity intensity captured during acceleration analysis. Subsequently, basic downsampling of data from a single accelerometer is probably insufficient to determine the full effect of different sampling rates. As a measurement device, the AX3 stores the output of the underlying ADXL345 accelerometer with no material difference in how the device behaves at a lower sampling rate. A contributor to the observed differences in data collected at different sampling rates may be the result of the internal ADXL345 single-pole IIR digital filter or otherwise related to the data capture of the MEMs accelerometer. Subsequently, a similar effect would be observed in the GENEActiv activity monitor, which uses the same underlying hardware.

The results of the current study demonstrate an increased sensitivity to interpolation method in the lower sampling rate, with little difference between methods in the 100 Hz data. The resampling of raw accelerometer data is a common processing step in order to eliminate inconsistent data timing in raw datasets. Both GGIR and the Biobank Accelerometer Analysis Tool utilise linear interpolation methods by default. Because published bulk data has likely been processed via linear interpolation resampling (e.g., 100 Hz activity data from the UK Biobank physical activity cohort), I have generated transformations to enable direct data comparison between combinations of nearest neighbour and linear interpolations at 25 Hz and 100 Hz sampling rates. I recommend the use of nearest neighbour resampling going forward, particularly in the instance of reduced sampling rate, to avoid unintended smoothing of the slower sampled data.

When using machine learning models of activity classification, I found high correlation between different sampling rates across classified activities. Ideally, machine learning activity classification models would be retrained on 25 Hz data to allow for classification on reduced sampling rate data; however, in most cases this is impractical. Transformations are provided for activity classifications of time spent in sit/stand, sleep, vehicle, sedentary, and light activity captured at a 25 Hz sampling rate when comparing to a 100 Hz dataset. Due to the small difference in classification of mixed activity, cycling, and MVPA, transformation for these activities is not recommended. Most notably, a reduced sampling rate resulted in increased classification of sedentary activity with decreased contributions of sleep and light activity within the Walmsley classification model. The 15% reduction in classified light activity is likely due in part to the general underestimation of ENMO throughout the 25 Hz timeseries. This finding is contrary to a study published by Zhang et al.[150] which reported no significant reduction in activity classification accuracy when comparing downsampled data to the original 80 Hz sequence. It is likely the use of downsampling versus a side-by-side comparison has contributed to the opposing findings.

3.5.1 Strengths and Limitations

A major strength of this chapter is that it includes a side-by-side comparison of standard and reduced accelerometer sampling rates at multiple body locations, without dependency on downsampling. This methodology includes the comprehensive use of acceleration-based activity metrics, cut-point, and machine learning variables to facilitate direct comparison of data collected at 25 Hz with existing large physical activity cohorts that have collected data at 100 Hz. However, data was collected in a healthy convenience sample, and there may be

unquantified effects of measurement in disease populations. While individual devices were randomised to body location (hip/wrist), a potential limitation is that the exact position of devices at each location was not randomised by design. For example, when sensors were placed side-by-side on participants' wrists, it could be possible that the 100 Hz sensor was more frequently placed nearer to the wrist and thus experience slightly higher accelerations. However, this alternative explanation is unlikely to account for the differences found in this and other studies[86]. The analysis of additional intermediate sample rates was not conducted in this study but could further shape the understanding of the relationship between sampling rates and activity metrics. Additionally, the given linear transformations may not be well suited for activity transformation during extended near-zero acceleration activity levels. Epoch-based comparisons can have sensitivity to exact epoch boundaries, as all individual sensor clocks drift slightly over time. Comparing two epochs between devices that report the same boundaries may be comparing two slightly different periods.

3.6 Conclusion

To facilitate comparisons between studies, researchers should consider the effect of both sampling rate and resampling interpolation method on study protocol, data analysis, and fully report all accelerometer parameters in the study methodology. The generation of validated transformations between 25 Hz and 100 Hz collected data when using the Axivity AX3 accelerometer allows for the direct comparison of long-term longitudinal acceleration data and the week-long physical activity data collected at 100 Hz in the UK Biobank Physical Activity Cohort. This is specifically of interest when comparing baseline population activity levels to clinical cohorts with transient activity levels—in the rehabilitation setting when monitoring

functional gains, or when monitoring a clinical population with potential long-term functional decline.

Chapter 4: Development of a Hybrid Machine Learning Peak Detection Model for Step Counting of Wrist-based Accelerometry Datasets

4.1 Introduction

Measures of physical activity using wearable sensors in academic research frequently involve activity metrics aggregated from raw accelerometer data, including cut-point defined time-use behaviours, intensity gradients, activity counts, and overall daily acceleration[51,174]. Use of these common accelerometer-based metrics are typically validated against the gold standards of indirect calorimetry or doubly labelled water, and provide reasonable estimates for energy expenditure for use in clinical and epidemiological public health research[79,174]. While these metrics provide an objective means for quantification of overall activity, they are not easily interpretable by a lay audience. The use of step count as an alternative activity metric is common, particularly in clinical research areas not typically specialising in the objective measurement of physical activity. As demonstrated in Section 2.4.2, step count from hip-mounted activity monitors is the most commonly used activity metric in orthopaedic research to date. Step count is an intuitive and readily accessible outcome measure, as steps are routinely provided as the primary metric in commercial activity trackers and offer a measure of physical activity which is highly relatable to the general public[175].

A major limitation of step counting for research purposes is the difficulty in accurately detecting steps from raw acceleration measurements. Most step counting devices and algorithms are validated in laboratory settings, with wearers

participating in scripted activities and walking or running on a treadmill for a predetermined distance. ANSI/CTA standards for step count accuracy are based on 5 minutes of walking and 5 minutes of running on a treadmill at a self-selected comfortable speed[176]. To meet the standard, device-measured step count error must fall below 10% in a cohort of at least 20 participants in walking and running trials. Treadmill-based testing does not reflect usual daily activities, however, which include a wide variety of intermittent gait patterns and variable gait speed which complicate step counting and are not replicated in laboratory conditions[26]. Many devices perform well when step counting during moderately paced walking[177], yet reduced accuracy is common during slow or assisted walking[178–181]. A recent consensus guideline statement for determining step count validity in consumer wearables recommends testing should culminate with validation in the free-living environment[182], wherein mean step counting errors are regularly above 20% in commercial and research grade activity monitors[61]. This includes hip or thigh-mounted activPal and Actigraph monitors used in many of the currently published assessments of post-operative physical activity in knee arthroplasty patients.

Body placement is a significant consideration when performing any research study assessing physical activity with wearable accelerometers. Accuracy of step counting devices can be highly dependent on body placement, with ankle and hip accelerometers generally producing more accurate step counts than wrist-based accelerometers[61,183]. Placement of an accelerometer at the wrist is more susceptible to error due to upper body motion associated with non-walking activities, leading to further difficulty in accurately assessing daily step counts. While wrist placement increases the burden in accurately assessing step count,

most commercial activity trackers purchased annually are wrist-based devices, indicating a popular preference for wrist-based devices in the general public[63]. General comfort with wrist-based devices is further supported by the high wear time compliance in large wrist accelerometery studies[53]. With the ability to gather sleep data and limit device non-wear time in study participants, the use of wrist-based activity is increasing in popularity in clinical and physical activity studies.

While commercial wearable manufacturers protect proprietary step counting algorithms and their associated validation, the availability of open-source ground truth data for the validation of step counting in research-grade accelerometers is also limited, particularly in free-living conditions. As such, published development of step counting algorithms routinely rely on small or unpublished datasets[184], scripted or treadmill activity[185], or are without a true gold standard of reference[186]. Subsequently, the research community currently lacks a well-validated, open-source algorithm designed for step counting of wrist-based accelerometry from research-grade devices.

Therefore, the purpose of this study was twofold: 1) to create an open-source dataset with ground-truth labelled steps in a free-living population, and 2) to create an open-source algorithm tuned for valid step count quantification using the research-grade AX3 accelerometer for wrist-based physical activity monitoring.

4.2 Methods

4.2.1 Datasets

4.2.1.1 Internal Ground Truth Dataset

The primary dataset in this study is a novel, unscripted and free-living dataset generated in conjunction with the data collected in Chapter 3.

4.2.1.1.1 Ethical Review and Approval

Ethical approval for participant recruitment was obtained from the Central University Research Ethics Committee of the University of Oxford (Ref: R63137/RE001) as an arm of the study described in Chapter 3.

4.2.1.1.2 Experimental Setup

As described in Section 3.3.2, participants were instructed to wear four triaxial accelerometers (AX3, Axivity, Newcastle, UK) for 24 hours, except during bathing or swimming activities. Two accelerometers were placed side-by-side on the dominant wrist using a wristband, and two on the dominant-side hip, waist level at the anterior-posterior midline via a belt clip (Figure 4-1). At the time of study enrolment, participants were given the option to enrol in an additional study arm, in which they would provide a ground truth video recording for up to one hour during the 24-hour monitoring period. The interval at which this 1-hour period began was arranged at the convenience of the participant and did not necessarily coincide with the exact start of the overall monitoring period.



Figure 4-1: Two Axivity AX3 accelerometers fitted onto a participant's dominant wrist, two clipped at the dominant-sided hip, and a foot-facing action camera clipped at the participant's waistline.

Accelerometers were synchronised and programmed as presented in Section 3.3.2. Ground truth video was captured using a commercially available wide-angle action camera (Action Camera CT9500, Crosstour, Shenzhen, China). The camera was contained within a clip mount accessory and clipped at the participant's beltline in a comfortable position. The LCD display was used to ensure that the motion of both feet was adequately captured within the camera field of view. Participants were instructed to wear the camera for a 1-hour period. Participants were also instructed to remove the camera and place it face down at any time they felt uncomfortable or participated in any activities they wished not to be recorded, including the use of the lavatory. Audio recording was disabled during all recording so as not to infringe on participant privacy. At the beginning of the video recording, the participant was instructed to strike together their wrist and hip

accelerometers in four successive, forceful blows within view of the camera to enable video and accelerometer timestamp synchronisation during video annotation.

4.2.1.1.3 Video Stitching

As is common in commercial action cameras, the Crosstour Action Camera records videos in 14-to-16-minute segments to protect against file corruption. Prior to annotation, these segments were stitched together to produce one continuous video file for each participant using Windows Video Editor. Upon video inspection a one-frame dropout was discovered at each segment seam in the full-length video (a loss of approximately 0.17 seconds of video 60 frames per second). This data dropout was adjusted for as noted in Section 4.2.1.1.5.

4.2.1.1.4 Ground Truth Step Annotation

Ground truth step annotation was independently performed by two annotators, Annotator #1: Lennart von Fritsch, and Annotator #2: Scott Small. All files were annotated by Annotator #2, and a subset of files annotated by Annotator #1. Prior to annotation, the annotators conferred over a sample of videos and came to an agreement on the definition of a qualifying step. Steps were subsequently defined as a foot strike in which the participant's centre of mass was judged to have shifted along the floor in an intentional manner. Examples of foot motion not qualifying as a step included foot shuffling, changing foot alignment without overall body displacement, and shifting of weight from one foot to the other. Annotation was conducted within an open-source video annotation software package (Elan 6.0, The Language Archive, Nijmegen, Netherlands) wherein the raw accelerometer CSV data file from the 100 Hz collected wrist AX3 accelerometer and participant ground truth video were imported and collated. Video and raw accelerometer data

were synchronised through the software package's Media Synchronization Mode by identifying the initial video frame of the first of four accelerometer strikes as recorded during setup and aligning that image to the first of the four distinguishable peaks observed in the raw z-axis accelerometer data. Following synchronisation, annotation was performed within the Segmentation Mode of the Elan software, wherein an "Enter" keystroke was made during video playback at heel strike of every identified step within the video file (Figure 4-2). After the ground truth video annotation was complete, a tab-delimited text file was exported from the software containing a single timestamp of the beginning and ending of each step-defining keystroke. Following annotation of all subjects, intraclass correlation coefficient was used as a measure of interrater reliability based on the total number of ground-truth steps counted for each participant.



Figure 4-2: Screenshot of step annotation with the Elan software package, showing the foot-facing ground truth video image (top), raw accelerometer data (centre), and heel strike annotations for each individual step (bottom).

4.2.1.1.5 Data Cleaning and Epoch Annotation

In preparation for walking classification and step counting analysis, exported annotation timestamps were cleaned and processed using a series of custom-written scripts using R (v4.0.3) and RStudio (v1.4.1103). Raw accelerometer data was extracted from the compressed cwa AX3 file using the Pywear (https://github.com/activityMonitoring/actipy) accelerometer processing tool during which linear resampling to the nominal accelerometer sampling rate was performed, followed by local gravity calibration and the application of a 20 Hz

Butterworth low pass filter to remove machine noise[53]. This data was processed in the bespoke R script alongside the step annotation data exported from the Elan software package. Step annotations were aligned with the raw accelerometer files based on initial offsets as determined during initial synchronisation, with a subsequent labelling of each 0.01 second data point with a binary 0 or 1 label if a step annotation was defined during that time window. Timestamps of step annotations were shifted forward by 0.17 seconds following each video chapter seam to compensate for frame dropout as described in Section 4.2.1.1.3. Step count data was subsequently collected into 5, 15, and 30 second epochs with the total number of annotated steps summed per epoch. Additional annotations of "walk" or "nonwalk" for each epoch were added for future model classification training based on the minimum number of steps per epoch chosen to define a walking segment as presented in Table 4-1. A range of walking definitions from 0.33 to 1 steps per second were used for each epoch length to facilitate the optimisation of classification sensitivity within a machine learning walking classification model.

Table 4-1: Walking Annotation Defined as Number of Steps Per Epoch

Name	Epoch Length	Minimum Steps for Defined Walking
Model 5/2	5 seconds	2 steps per epoch
Model 5/3	5 seconds	3 steps per epoch
Model 5/5	5 seconds	5 steps per epoch
Model 15/5	15 seconds	5 steps per epoch
Model 15/10	15 seconds	10 steps per epoch
Model 15/15	15 seconds	15 steps per epoch
Model 30/10	30 seconds	10 steps per epoch

4.2.1.2 Clemson Shimmer3

An open-source step-annotated dataset from Clemson University (http://cecas.clemson.edu/~ahoover/pedometer/)[187] was used as a reference dataset for external model validation. Within this external dataset of 30 participants (15 F, mean age 21.9 (SD 2.4)), researchers recorded a mean 37 minutes of scripted activity split between periods of defined regular, semiregular, and irregular walking for each participant. Participants were video recorded throughout scripted activities with steps timestamp annotated while wearing Shimmer3 (Shimmer, Dublin, Ireland) inertial measurement units (IMU) recording at 15 Hz. Sensors included ankle, hip, and wrist accelerometers always placed on the participant's left. Researchers annotated steps as well as "shifts", foot movement not necessarily tied to a change in body position, though these annotated shifts were not included in the current analysis. In the current study, calibrated triaxial accelerometer data and step annotation files from the Clemson dataset were cleaned and formatted to match the required CSV input format of the Biobank Accelerometer Analysis tool for data processing.

4.2.1.3 Capture24

A second open-source dataset was used in this study for further model assessment. The Capture24[67,188] dataset consists of Axivity AX3 accelerometer data from 152 participants [aged 18-91, (98 F, 54 M)] with waking time activity annotations generated from first-person point of view wearable cameras for 24 hours. In the Capture 24 dataset, accelerometer data was collected at 100 Hz with ground truth still images captured approximately every 20 seconds. Annotations for all still images were performed by at least two trained annotators with 206 different activity labels and then subsequently collated into categories including a

"Willetts" annotation set, comprised of: bicycling, mixed activity, sit-stand, sleep, vehicle time, and walking, and a "Doherty" annotation set, comprised of: moderate activity, sedentary activity, sleep, light tasks, and walking. In the current study, annotations were additionally collected into two bespoke binary categories of potential walking/stepping and likely not walking/stepping based on original annotations. The final Capture24 dataset is presented with activity annotations labelled in 30 second epochs. This dataset does not feature ground truth step counts but can be used for a general comparison of the binary walking classification with another independent activity classification system.

4.2.1.4 Summary of Primary and External Validation Datasets

In total, two external datasets were used in conjunction with data collected in the current study for use in step counting model validation and assessment. A summary of these open-source datasets alongside the primary dataset are presented in Table 4-2.

Table 4-2: Summary of Datasets

Characteristic	Current Study	Clemson Shimmer3[187]	Capture24[67,188]
Institution	University of Oxford	Clemson University	University of Oxford
Number of Participants	39	30	152
Sensor	Axivity AX3 triaxial accelerometer	Shimmer3 9-axis IMU	Axivity AX3 triaxial accelerometer
Sampling Rate	25 Hz and 100 Hz	15 Hz	100 Hz
Body Location	Dominant wrist and dominant hip	Left wrist, left hip, left ankle	Dominant wrist
Activity Protocol	Unscripted free-living	Laboratory-setting, scripted regular gait, semiregular gait, irregular gait	Unscripted free-living
Measurement Window	1 hour	37 minutes	24 hours
Ground Truth Capture	Foot-facing video camera worn at waist	Participant-facing video camera held by researcher	First-person still images from lanyard every 20-60 seconds
Annotation of data	Timestamped steps by two annotators	Timestamped steps and shifts	206 unique activity labels
Label granularity	Timestamps to 0.001 s precision	Timestamps to 0.001 s precision	1 for label every 30 seconds of accelerometer data
Participant Age	38.5 (SD 14.0) years Range: 19 - 81	21.9 (SD 2.4) years	41.6 (SD 15.7) years Range: 18-91
Participant Sex	19 Female, 20 Male	15 Female, 15 Male	98 Female, 54 Male

4.2.2 Step Detection Methods

4.2.2.1 Baseline Step Detection Model: Open Movement Moving Average Peak Detection

A preliminary step count was performed using the open source Open Movement OMConvert command line software (OMConvert, OpenLab, Newcastle University, Newcastle, UK) (https://github.com/digitalinteraction/openmovement/tree/master/Software/AX3/omconvert). The Open Movement OmConvert software package generates a step count from raw .cwa files of Axivity AX3 devices through a moving average peak detection based on an algorithm designed by Cho et al[189]. Raw accelerometer files were processed with the step counting feature of the Open Movement OMConvert software, with steps exported in a csv as detected in 60 second epochs through the entire 24-hour measurement period. The exported step count data for the entire monitoring period was subsequently trimmed to retain only the calculated steps during video recording for ground truth comparison. Root mean square error (RMSE) was calculated for each participant and Bland-Altman plots were created for step counting compared to ground truth.

4.2.2.2 Random Forest Regression

A first pass in generating a novel step detection model was completed using a random forest regression prediction method to directly estimate the number of steps taken during a given time epoch from a series of calculated accelerometer signal features without relying on peak detection. Raw accelerometer data from each participant was processed using the Biobank Accelerometer Analysis software package to extract a series of rotation-invariant time and frequency domain signal features, listed in Table 3-6[81,83,151,172,173], as calculated in 5, 15, and 30 second epochs. Random forest regression was performed in Python (v3.7.9)

using the scikit-learn machine learning Python package. Random forest regression incorporating leave-one-participant-out cross validation was performed across all participants, with predictions of epoch-level step count based on extracted features in the 5, 15, and 30 second epochs. Predictions of steps counts in individual epochs were allowed to be fractional; however, the total step count per participant was rounded to the nearest integer value. Predictive model quality was determined based on the RMSE comparing predicted to ground truth step counts through the measurement period for each individual participant.

4.2.2.3 Hybrid Activity Classification and Peak Detection

Building from the prior two step counting models, a hybrid machine learning and peak detection algorithm was additionally developed. In building this model, leave-one-participant-out cross validation was used to develop and validate a machine learning walking classification model, fine-tune parameters for peak detection step detection, and then evaluate the overall performance of a combined walking classification and step count function. A flowchart of the leave-one-participant-out model training and tuning process is presented in Figure 4-3. Firstly, the participant dataset was divided into n folds wherein the group of n-1 participants underwent a randomised 80%-20% split. Data from the randomly selected 80% of participants was used to train a machine learning classification model for walking detection. A balanced random forest model with 1,000 threads was used, alongside integrated smoothing via a hidden Markov model, calculated from signal features to predict walk or non-walk labelling as previously defined.

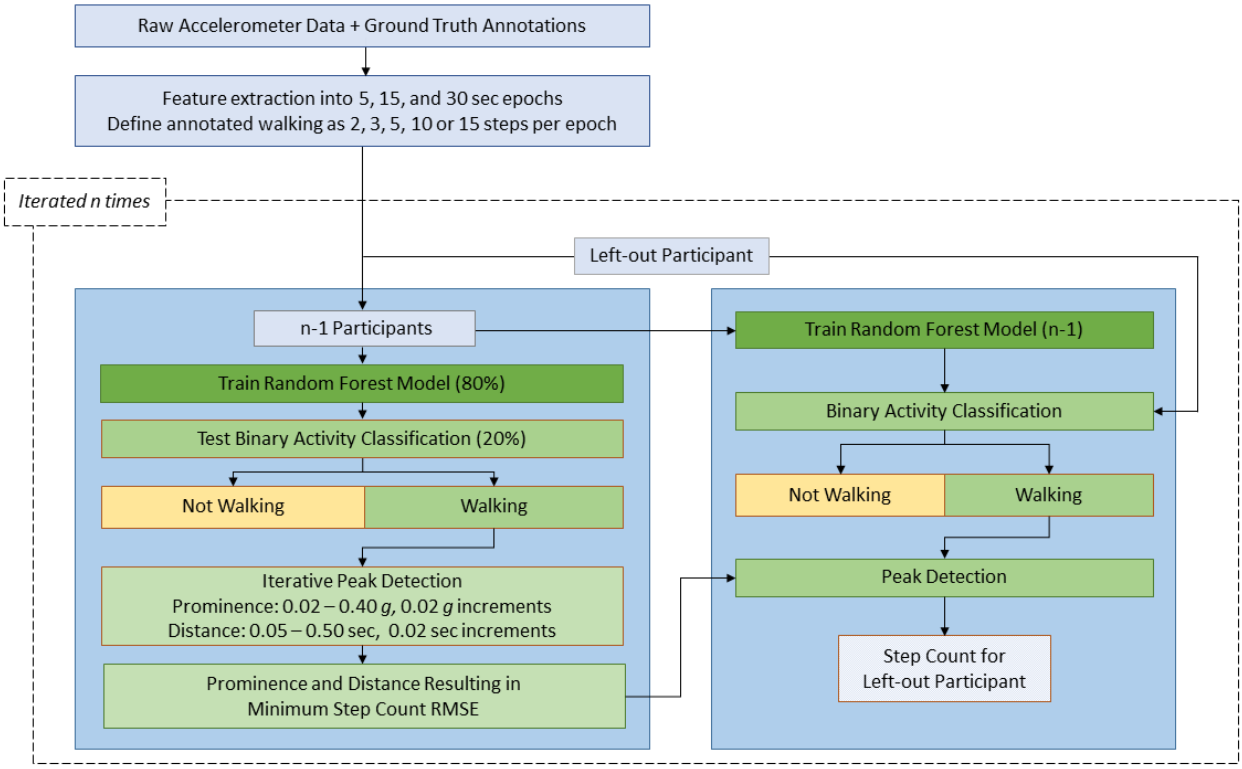


Figure 4-3: Flowchart for the development and tuning of the hybrid machine learning and peak detection step counting algorithm.

Following model development in each fold, label predictions for the held-out 20% of data were generated. Raw accelerometer data, along with walking classification labels were then imported into a bespoke Python script which applied the "signal.find_peaks" method from the Python scipy package to count the number of local signal peaks, or steps, within those epochs labelled as walking epochs from the classification model. During step counting, two primary parameters, peak prominence (the degree in which a peak differs from the surrounding baseline signal) and peak distance (the minimum allowable time between consecutively measure peaks) were iterated across a pre-selected range of values. Preliminary iterations were conducted with a peak prominence was iterated across the values from 0.02 to 0.50 *g* in increments of 0.02 *g*, with peak distance iterated across the values 0.05 to 1.00 seconds in increments of 0.05 seconds. Final analysis was

conducted with narrowed windows of peak prominence (0.02 to 0.40 *g* in increments of 0.02 *g*) and peak distance (0.2 to 0.5 s in increments of 0.02 s). The step count RMSE for each combination of peak prominence and peak distance was averaged across all folds to determine the combination of step count parameter values generating the lowest overall predictive RMSE.

Finally, the walking classification model was retrained across n-1 participants, applied to the held-out participant, and final step counting was performed on that participant in time epochs classified as walking, using the tuned step count parameters generating the lowest RMSE in the previous model development stage. This process was repeated over n folds, with step count RMSE averaged across the entire participant population to determine step counting accuracy. F1 and mean accuracy was additionally calculated to evaluate walking classification performance. The overall procedure was then iterated across all 9 combinations of epoch length and defined walking (Table 4-3) to select a final hybrid step counting model. A secondary analysis of this procedure was performed with accelerometer data filtered with a 5 Hz Butterworth low pass filter to assess the impact of aggressive timeseries smoothing when used for peak detection. A schematic for the generation of step count data from the hybrid step detection model is shown in Figure 4-4.

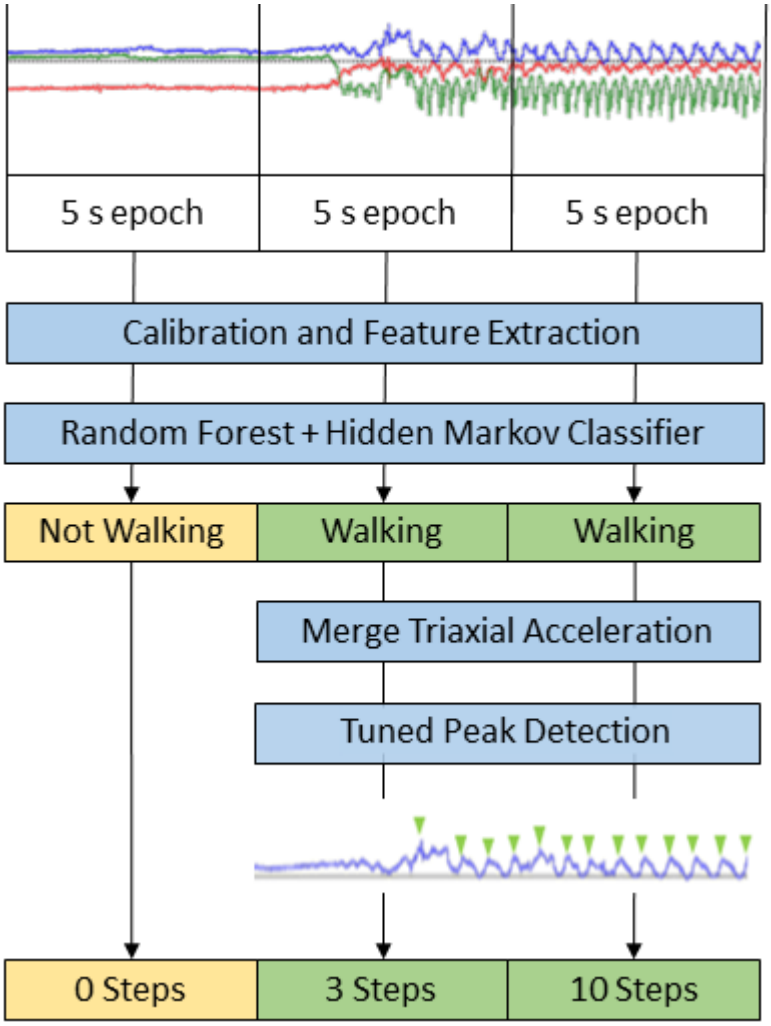


Figure 4-4: Schematic of the process for generating step count from raw accelerometer data using the final hybrid step counting model

4.2.3 External Model Validation

4.2.3.1 Clemson Shimmer3

Upon completion of preliminary model development, external model validity was assessed using the Clemson Shimmer3 open-source dataset described in Section 4.2.1.2. Two step counting analyses were performed on the external Clemson dataset with models generated from the primary study data: random forest regression, and hybrid classification and peak detection. Performance evaluations were applied to the Clemson dataset predictions in the same manner as previously

described in Section 4.2.2.3, with the calculation of F1 and accuracy metrics in walking classification and step count RMSE and mean absolute percent error (MAPE) in step detection for all developed models.

To facilitate direct comparison with the implementation of wrist, hip, and ankle step count algorithms as applied to this dataset by Mattfeld et al.[187], running count accuracy was additionally calculated for the best performing step counting model. This accuracy metric was defined by the authors as the average ratio of detected steps to ground truth steps across each individual participant.

4.2.3.2 Capture 24

Following model training and implementation into the Clemson external validation dataset, the best performing model was selected and applied to the Capture24 dataset for secondary assessment. Accelerometer data from the Capture24 dataset was processed using the Biobank Accelerometer Analysis tool while implementing the final walking and step counting model for activity classification and step count estimation. Comparisons in wearable camera annotations and walk/non-walk activity classifications were compared to assess 24-hour implementation of the final generated algorithm. Additionally, the overall accelerometer acceleration was compared to step count detection via linear regression to define an approximate relationship between average daily overall acceleration and daily step count.

4.2.4 Sample Rate Comparison and Adjustment

Following final model validation, a duplicate step counting model of the same epoch length and step definition as the final model was created using accelerometer data collected at the wrist with a 25 Hz sampling rate. Subsequently, step counting was conducted on Chapter 3 participants who

elected not to participate in ground truth video collection. The final step counting model, trained on 100 Hz data, was used to assess 24-hour step counts in these participants using both 100 Hz and 25 Hz collected data. Step counts of the 25 Hz data was then reassessed using the new model version trained on 25 Hz data. Bland-Altman plots were used to assess the validity of using the original 100 Hz model on 25 Hz collected data, compared to using the model specifically trained on 25 Hz data.

4.3 Results

4.3.1 Participant Demographics and Ground Truth

Accelerometer data and matched ground truth camera data was collected from 39 participants (19 F, 20 M) with a mean age of 38.5 years (range 19.5 to 81.2 years) (Table 4-3). A total of 37.9 hours of video was annotated for an average of 58 minutes of data per participant. Two annotators viewed and marked steps in this study, with 33 participants annotated by Annotator 1 and 39 participants annotated by Annotator 2. Annotated step counts for all participants are presented in Figure 4-5. A mean bias of 26 fewer steps were counted by Annotator 1 in the 33 participants annotated by both annotators, representing an RMSE of 3% between annotators, with an intraclass correlation coefficient of 1.0. The amount of annotated walking time for each participant was dependent on the definition of walking in terms of steps per epoch. For walking classification, the step counts from Annotator 2 were defined as the ground truth for all participants. Overall, participants took a median [IQR] 863 [312 - 2,123] steps, ranging from 94 to 5,942 within the hour measurement period. A median 10.1 [4.8 - 26.7] minutes of walking was identified in participants, with a range of 1.3 to 54.4 minutes across all participants, when defining walking as a 5 second epoch with at least 2 steps.

Table 4-3: Participant Demographics for Step Detection

	Mean	SD	Sex	n	%	Handedness	n	%
Age (years)	38.5	14.0	Female	19	48.7%	Left	7	17.9%
Height (cm)	171.9	11.6	Male	20	51.3%	Right	32	82.1%
Weight (kg)	73.8	13.7						
BMI (kg/m^2)	24.9	3.9						

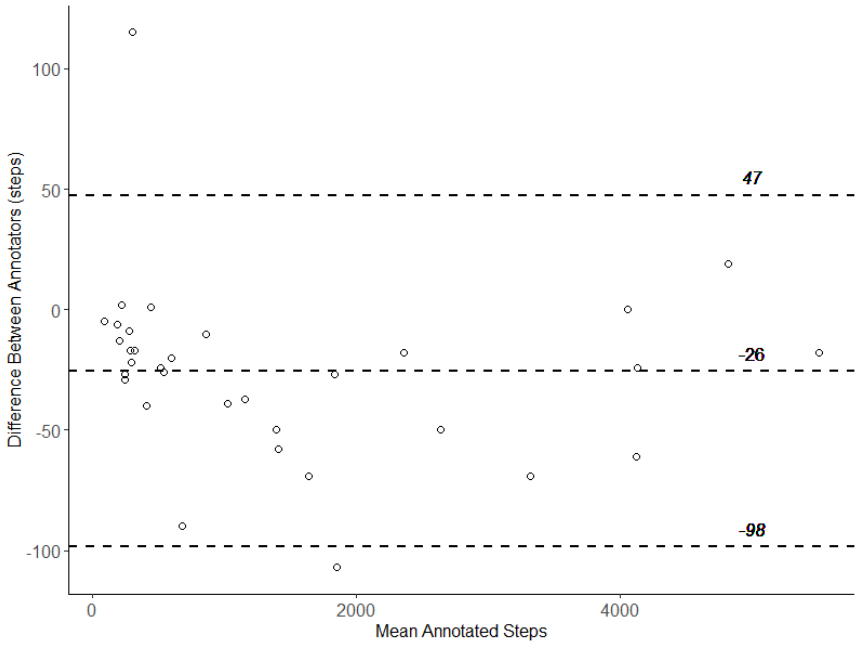


Figure 4-5: Bland-Altman plot of ground truth step counts as annotated by the two annotators in 33 participants.

4.3.2 Hybrid Model Refinement

4.3.2.1 Step Count Parameter Optimisation

The results summary from step count parameter optimisation are presented in Table 4-4, with the combination of peak prominence and peak distance producing the minimum overall step count RMSE for each combination of epoch length and steps per epoch walking definition. An example heatmap of step count error across parameter iteration is presented in Figure 4-6. Minimum preliminary step count

RMSE ranged from 182 steps to 329 steps across all models. The lowest RMSE observed in the 30 second epoch, 10 steps per epoch classification model (Model 30/10) when step detecting using a peak distance of 0.36 seconds and a peak prominence of 0.18 *g*.

Table 4-4: Results of Step Count Parameter Iteration Including Peak Detection Settings and Associated Minimum Step Count Error Across Held Out Test Data

Walking Model	Peak Distance (s)	Peak Prominence (g)	RMSE (steps)
Model 5/2	0.32	0.14	217
Model 5/3	0.24	0.20	329
Model 5/5	0.26	0.08	293
Model 15/5	0.36	0.16	210
Model 15/10	0.32	0.10	217
Model 30/10	0.36	0.18	182
Model 30/15	0.36	0.12	215

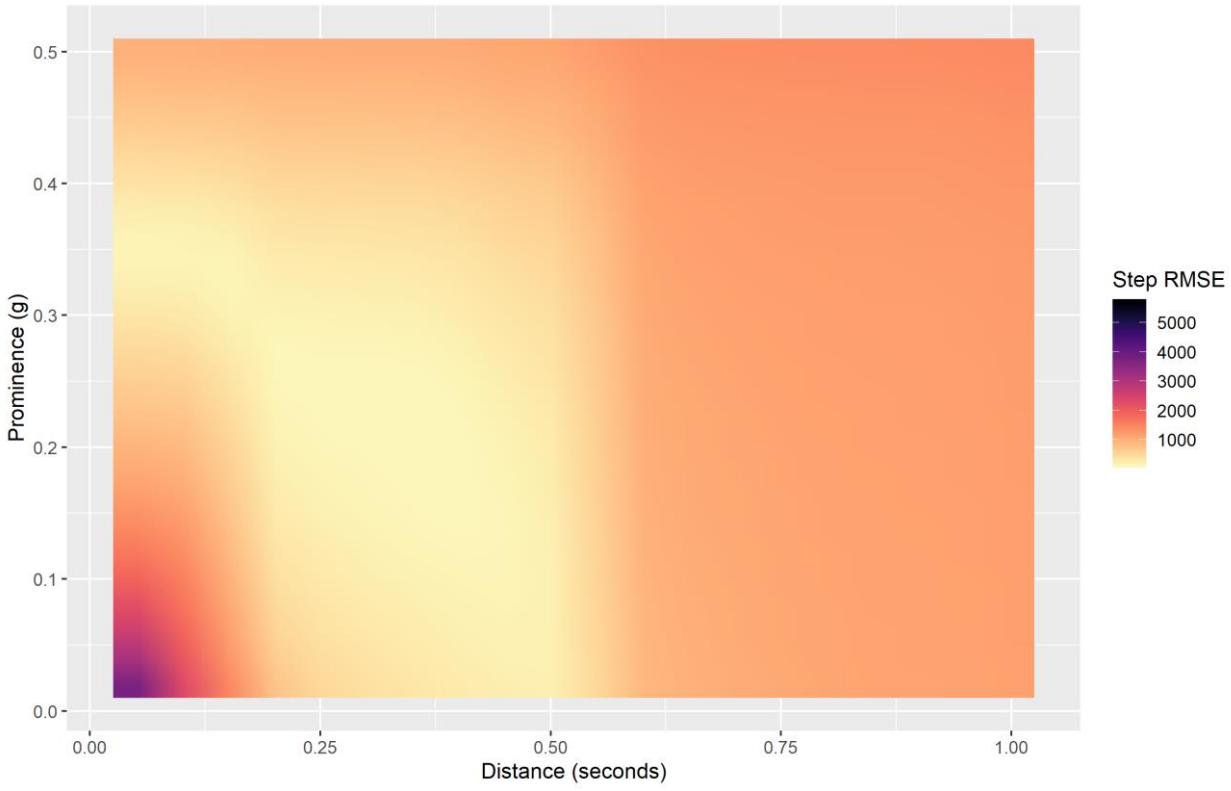


Figure 4-6: Heatmap of step counting error across iterated peak detection prominence and distance parameters using the 5 second epoch, 2 steps per epoch walking model.

The results of aggressive low pass filtering for timeseries smoothing for peak detection are presented in Table 4-5. The increasingly smoothed timeseries provided no added benefit in terms of reduced step measurement error in comparison to the 20 Hz filtered data and was subsequently dropped from further analysis to avoid unnecessary timeseries data loss.

Table 4-5: Results of Step Count Parameter Iteration Including Peak Detection Settings and Associated Minimum Step Count Error Across Held Out Test Data, Using a 5 Hz Lowpass Filtered Timeseries Signal

Walking Model	Peak Distance (s)	Peak Prominence (g)	RMSE (steps)
Model 5/2	0.32	0.06	228
Model 5/3	0.24	0.12	303
Model 5/5	0.26	0.04	298
Model 15/5	0.34	0.10	228
Model 15/10	0.32	0.04	222
Model 30/10	0.32	0.12	202
Model 30/15	0.32	0.08	223

4.3.3 Internal Model Validation

4.3.3.1 Baseline Step Counting

Summary results from step counting through baseline moving average peak detection via the Open Movement software system are presented in Table 4-6. Hip-based step detection, as calculated from 100 Hz data collection, resulted in an overall underestimation of 6.3% and an RMSE of 10.2%, relative to the participant mean ground truth step count. Of primary interest in this study, wrist-based step detection, as calculated from 100 Hz data collection, demonstrated an overall overestimation of step count by 18.6% with an RMSE of 32.3% relative from the mean ground truth step counts.

Table 4-6: Overall Step Counts as Calculated Through Moving Average Peak Detection with Open Movement Software in Comparison with Ground Truth

		Hip Sensor		Wrist Sensor	
Step Count	Ground Truth	25 Hz	100 Hz	25 Hz	100 Hz
Total Steps	62,520	58,686	58,798	74,152	74,437
Mean Steps	1,603	1,505	1,508	1,901	1,909

	Hip Sensor		Wrist Sensor	
Performance	25 Hz	100 Hz	25 Hz	100 Hz
% Ground Truth Steps Counted	93.7%	94.0%	118.6%	119.1%
RMSE (Steps)	163	162	517	538
% RMSE	10.2%	10.1%	32.3%	33.6%

4.3.3.2 Random Forest Regression

Summary results from step counting via random forest regression of wrist-based, 100 Hz accelerometer are presented in Table 4-7. Differences in overall epoch length used in step calculation were minimal, with a 0.9%, 0.6%, and 0.8% underestimation of overall steps in the 5 second, 15 second, and 30 second epoch lengths, respectively. RMSE to mean ground truth steps was highest at the 5 s epoch level and lowest, at 15.7% of mean step count when the regressor was applied over 30 s epochs.

Table 4-7: Overall Step Counts as Calculated Through Random Forest Regression in Comparison with Ground Truth

Step Count	Ground Truth	5 s Epoch	15 s Epoch	30 s Epoch
Total Steps	62,520	61,966	62,138	61,989
Mean Steps	1,603	1,589	1,593	1,589

Performance	5 s Epoch	15 s Epoch	30 s Epoch
% Ground Truth Steps Counted	99.1%	99.4%	99.2%
RMSE (Steps)	339	266	252
% RMSE	21.3%	16.6%	15.7%

4.3.3.3 Hybrid Models for Walking Classification and Step Detection

Results from leave-one-participant-out walking classification are presented in Table 4-8. In all instances, an increased ratio of steps per second to define a walking epoch resulted in lower total labelled and classified walking time. Walking classification based on epoch length and steps per epoch walking definition demonstrate similar performance in random forest regression predictions across all iterations tested, with F1 metrics ranging between 0.85 and 0.88 and mean accuracy between 92% and 95% when classifying walking time. RMSE ranged between 12% and 19% of the mean ground truth walking time, with lower error with increased steps-per-epoch ratio and longer epoch windows.

Table 4-8: Walking Classification Results and Performance Metrics Based on Random Forest Classification Model

Classification	Model 5/2	Model 5/3	Model 5/5	Model 15/5	Model 15/10	Model 30/10	Model 30/15
Total Labelled Walking (min)	670.7	625.8	541.8	710.5	602.3	736.5	658.5
Mean Labelled Walking (min)	17.2	16.0	13.9	18.2	15.4	18.9	16.9
Total Predicted Walking (min)	606.3	568.1	509.1	699.0	586.8	731.0	640.0
Mean Predicted Walking (min)	15.5	14.6	13.1	17.9	15.0	18.7	16.4

Performance	Model 5/2	Model 5/3	Model 5/5	Model 15/5	Model 15/10	Model 30/10	Model 30/15
% Labelled Walking Predicted	90.4%	90.8%	94.0%	98.4%	97.4%	99.3%	97.2%
RMSE (min)	3.3	3.0	1.8	2.5	1.8	2.2	2.5
% RMSE	19.2%	18.7%	13.0%	13.7%	11.7%	11.6%	14.8%
Mean Accuracy	91.6%	92.8%	94.9%	91.6%	93.9%	91.2%	92.9%
F1	0.85	0.86	0.89	0.86	0.88	0.87	0.88

Step detection results from leave-one-out walking classification and step detection are presented in Table 4-9 cross the range of epoch lengths and walking-defined steps per epoch. Contrary to walking classification, highest accuracy in step counts was observed when steps per epoch walking definitions were lowest. RMSE relative to mean ground truth step counts ranged from 11% in Model 30/10, to 22% in Model 5/5.

Table 4-9: Step Count Results and Performance Metrics Based on Random Forest Classification and Peak Detection Step Counting

Step Counts	Ground Truth	Model 5/2	Model 5/3	Model 5/5	Model 15/5	Model 15/10	Model 30/10	Model 30/15
Total Steps	62,520	62,486	59,450	58,506	64,473	61,748	63,240	61,620
Mean Steps	1,603	1,602	1,524	1,500	1,653	1,583	1,622	1,580

Performance	Model 5/2	Model 5/3	Model 5/5	Model 15/5	Model 15/10	Model 30/10	Model 30/15
% Ground Truth Counted	99.9%	95.1%	93.6%	103.1%	98.8%	101.2%	98.6%
RMSE (steps)	210	246	288	193	208	173	221
% RMSE	13.1%	15.3%	18.0%	12.0%	13.0%	10.8%	13.8%

4.3.3.4 Model Comparison in the Primary Internal Dataset

Bland-Altman plots comparing the baseline step detection method alongside the best performing random forest regression and hybrid models are presented in Figure 4-7. Within the internal dataset, the baseline model resulted in a mean bias of 306 steps with limits of agreement ranging from -594 to 1,205 steps. Much tighter limits of agreement are observed in the random forest regression and hybrid models, with near-zero mean bias in each.

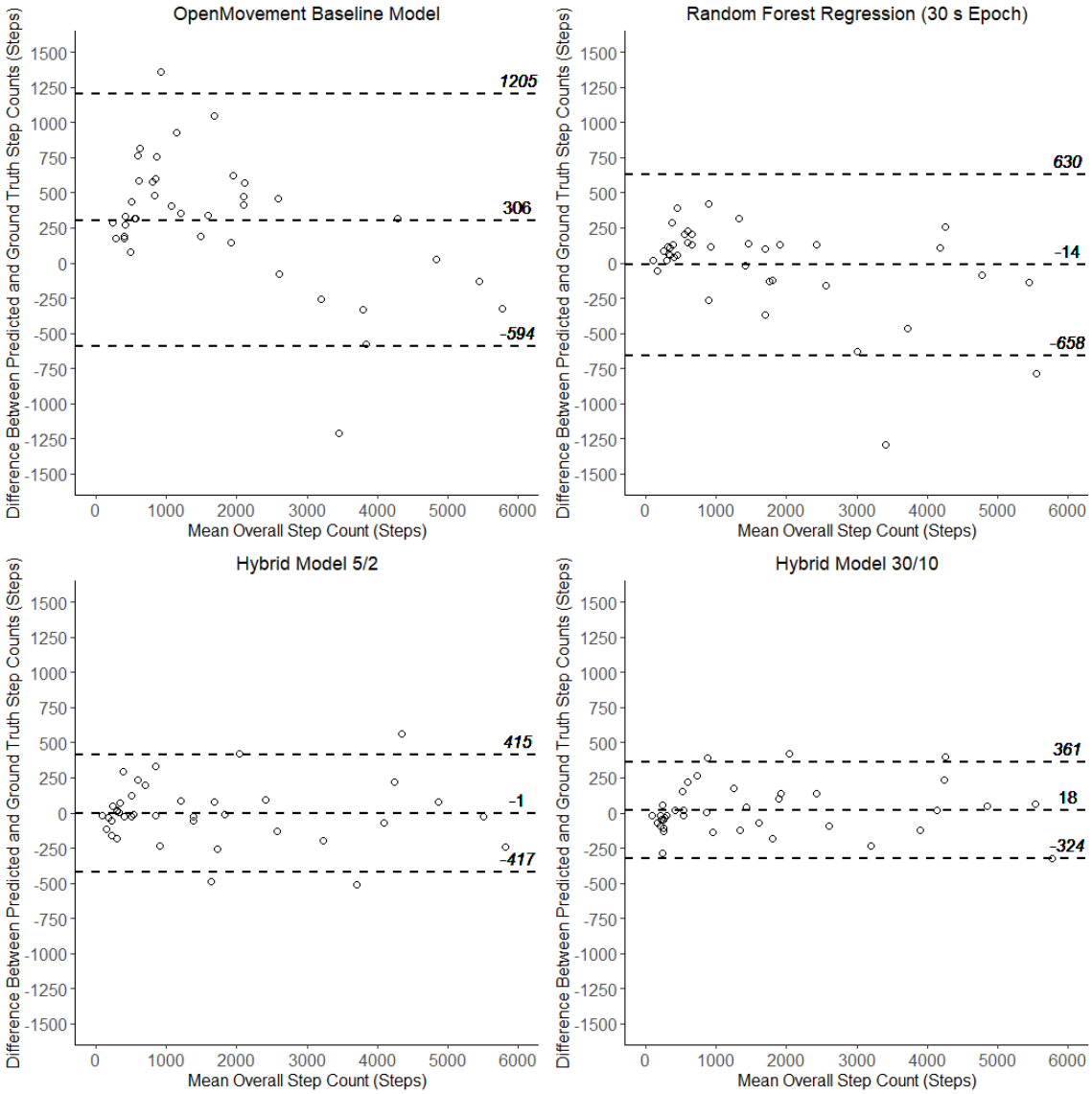


Figure 4-7: Bland-Altman plots for the comparison of step counting models in the primary, internal validation dataset, (Top left) baseline step counting model, (Top Right) random forest regression with 30 s epoch, (Bottom left) hybrid Model 5/2, and (Bottom right) hybrid Model 30/10.

A full comparison of the best performing candidate models for each step count detection method is presented in Table 4-10. The random forest regression and hybrid step count models substantially outperformed the baseline model in terms of wrist-based step counting. The baseline model overestimated step counts, with a total population overestimate of +19.1% and an RMSE of 33.6%. The 30 s epoch

random forest regression demonstrated much improved performance versus baseline, with a slight -1.5% underestimation of population step count and an RMSE of 15.7%. Two hybrid models are presented for comparison, with minimal over estimation (+0.1% and +1.2%) and an RMSE of 13.1% and 10.8% in the 5/2 and 30/10 models, respectively.

Table 4-10: Side-by-Side Comparison on Best Performing Models for Each Step Detection Method During Internal Validation with 100 Hz Wrist-based Acceleration

Comparator	Ground Truth	Baseline (Wrist)	Random Forest Regression (30 s Epoch)	Hybrid (Model 5/2)	Hybrid (Model 30/10)
Total Steps	62,520	74,437	61,989	62,486	63,240
Mean Steps	1,603	1,909	1,589	1,602	1,622
% Ground Truth Counted		119.1%	98.5%	99.9%	101.2%
RMSE (Steps)		538	252	210	173
% RMSE		33.6%	15.7%	13.1%	10.8%

RMSE = Root mean square error

4.3.4 External Validation and Final Model Selection

Following internal validation and tuning, random forest regression and hybrid step detection models were implemented into the Clemson Shimmer 3 dataset for external validation. Each of the three random forest regression epoch lengths were tested in external validation, as well as each of the seven hybrid model combinations (listed in Table 4-1). However, based on the results from internal validation, the 30 s epoch random forest regression and hybrid Model 5/2 and Model 30/10 were considered primary candidate models for final model selection. The baseline model was not carried over into external validation due to poor internal validation performance and incompatibility between the OpenMovement software and the Shimmer 3 data files.

4.3.4.1 External Validation in the Clemson Dataset

Overall step counting performance by random forest regression is presented in Table 4-11, with regression performance across each gait subtype reported in Table 4-12. As presented in Table 4-11, when incorporating all gait types, step count by random forest regression resulted in moderate overestimation of step counts (4.7% to 7.7% overestimation) with per-participant RMSE between 8.3% and 10.6%. However, random forest regression performance was highly variable by gait subtype (Table 4-12). During regular gait, the regression model tended to underestimate step count, with step predictions between 93.8% and 95.8% of the true step count, albeit with low percent RMSE (6.4% to 7.9%). Poorest step prediction performance was observed during irregular walking, when participants were engaged in intermittent walking while participating in manual construction of building block kits. Step prediction during irregular gait resulted in 202.4% to 210.3% of true steps predicted, with percent RMSE over 100% in each epoch length tested.

Table 4-11: Random Forest Regression Step Counting in Clemson Shimmer3 Dataset Participants Over All Combined Gait Subtypes

Step Count	Ground Truth	5 s Epoch	15 s Epoch	30 s Epoch
Total Steps	56,682	59,348	61,038	60,475
Mean Steps per Participant	1,889	1,978	2,035	2,016
Performance		**5 s Epoch**	**15 s Epoch**	**30 s Epoch**
% Ground Truth Counted		104.7%	107.7%	106.7%
RMSE (Steps)		157	199	191
% RMSE		8.3%	10.6%	10.1%
MAPE		6.7%	8.9%	8.6%

Table 4-12: Random Forest Regression Step Counting in Clemson Shimmer3 Dataset During Three Different Gait Subtypes

	Regular Gait			
Step Count	Ground Truth	5 s Epoch	15 s Epoch	30 s Epoch
Total Steps	31,443	29,481	30,109	29,502
Mean Steps	1,048	983	1004	983
Performance		5 s Epoch	15 s Epoch	30 s Epoch
% Ground Truth Counted		93.8%	95.8%	93.8%
RMSE (Steps)		83	67	83
RMSE %		7.9%	6.4%	7.9%
MAPE		6.3%	4.3%	6.2%
	Semiregular Gait			
Step Count	Ground Truth	5 s Epoch	15 s Epoch	30 s Epoch
Total Steps	19,920	19,099	19,903	19,785
Mean Steps	664	637	663	660
Performance		5 s Epoch	15 s Epoch	30 s Epoch
% Ground Truth Counted		95.9%	99.9%	99.3%
RMSE (Steps)		72	72	80
RMSE %		10.9%	10.9%	12.0%
MAPE		9.5%	9.2%	10.2%
	Irregular Gait			
Step Count	Ground Truth	5 s Epoch	15 s Epoch	30 s Epoch
Total Steps	5,319	10,768	11,026	11,188
Mean Predicted Steps	177	359	368	373
Performance		5 s Epoch	15 s Epoch	30 s Epoch
% Ground Truth Counted		202.4%	207.3%	210.3%
RMSE (Steps)		190	197	204
RMSE %		107.1%	110.9%	115.0%
MAPE		106.8%	111.2%	114.5%

Step counting performance of the novel hybrid models is presented in Table 4-13, with step counting performance across each gait subtype reported in Table 4-14. Across all combined gait subtypes (Table 4-13), hybrid models predicted 92.6% to 101.4% of true step counts, with percent RMSE falling between 6.8% and 11.3%. The best performing step prediction model, when applied across all participants and gait types was Model 5/2, demonstrating an RMSE of 6.8%, while predicting a step count of 100.8% of all ground truth steps.

Table 4-13: Step Counts in Clemson Shimmer3 Dataset Participants as Predicted by Hybrid Walking Classification and Step Counting Algorithm Combined Over All Gait Subtypes

Step Count	Ground Truth	Model 5/2	Model 5/3	Model 5/5	Model 15/5	Model 15/10	Model 30/10	Model 30/15
Total Steps	56,682	57,138	52,466	57,477	55,235	56,995	54,866	55,720
Mean Steps	1,889	1,905	1,749	1,916	1,841	1,900	1,829	1,857

Performance	Model 5/2	Model 5/3	Model 5/5	Model 15/5	Model 15/10	Model 30/10	Model 30/15
% Ground Truth Counted	100.8%	92.6%	101.4%	97.4%	100.6%	96.8%	98.3%
RMSE	128	214	151	169	138	189	176
RMSE %	6.8%	11.3%	8.0%	8.9%	7.3%	10.0%	9.3%
MAPE	5.1%	9.1%	5.7%	7.3%	5.3%	8.2%	7.3%

As demonstrated in Table 4-14, step prediction was most accurate during regular gait, during which the hybrid models estimated step count with a RMSE between 2.4% and 9.9%. Step counting performance was most variable during irregular gait, with hybrid models producing an RMSE of 35.5% and 88.8%, and underestimating step count in most models. The hybrid Model 5/2 had a RMSE of 3.8% during regular gait and performed the best of all models during irregular gait, with an RMSE of 35.3% during the interrupted gait sequence.

Table 4-14: Hybrid Step Counting Models Applied to Clemson Shimmer3 Dataset During Three Different Gait Subtypes

	Regular Gait							
Step Count	Ground Truth	Model 5/2	Model 5/3	Model 5/5	Model 15/5	Model 15/10	Model 30/10	Model 30/15
Total Steps	31,443	30,865	30,305	32,719	29,653	31,255	29,367	30,144
Mean Steps	1,048	1,029	1,010	1,091	988	1,042	979	1,005
Performance		Model 5/2	Model 5/3	Model 5/5	Model 15/5	Model 15/10	Model 30/10	Model 30/15
% Ground Truth Counted		98.2%	96.4%	104.1%	94.3%	99.4%	93.4%	95.9%
RMSE (Steps)		40	76	72	93	25	104	74
RMSE %		3.8%	7.3%	6.9%	8.9%	2.4%	9.9%	7.1%
MAPE		2.1%	4.2%	4.3%	5.6%	1.3%	6.4%	4.3%
	Semiregular Gait							
Step Count	Ground Truth	Model 5/2	Model 5/3	Model 5/5	Model 15/5	Model 15/10	Model 30/10	Model 30/15
Total Steps	19,920	21,678	19,424	23,146	20,448	23,293	19,689	21,953
Mean Steps	664	723	647	772	682	776	656	732
Performance		Model 5/2	Model 5/3	Model 5/5	Model 15/5	Model 15/10	Model 30/10	Model 30/15
% Ground Truth Counted		108.8%	97.5%	116.2%	102.7%	116.9%	98.8%	110.2%
RMSE (Steps)		103	90	143	78	141	77	104
RMSE %		15.5%	13.6%	21.5%	11.8%	21.2%	11.7%	15.6%
MAPE		11.9%	11.4%	17.8%	9.2%	18.3%	9.2%	12.8%
	Irregular Gait							
Step Count	Ground Truth	Model 5/2	Model 5/3	Model 5/5	Model 15/5	Model 15/10	Model 30/10	Model 30/15
Total Steps	5,319	4,595	2,737	1,612	5,134	2,447	5,810	3,623
Mean Steps	177	153	91	54	171	82	194	121
Performance		Model 5/2	Model 5/3	Model 5/5	Model 15/5	Model 15/10	Model 30/10	Model 30/15
% Ground Truth Counted		86.4%	51.5%	30.3%	96.5%	46.0%	109.2%	68.1%
RMSE (Steps)		63	97	131	89	129	130	157
RMSE %		35.3%	54.7%	73.7%	50.1%	73.0%	73.1%	88.8%
MAPE		29.8%	47.8%	69.3%	45.8%	67.5%	64.3%	72.0%

4.3.4.2 Model Comparison and Final Model Selection

A comparison the external validation performance of candidate final models, as selected in Section 4.3.3.4, is presented in Table 4-15. While the random forest regression models performed well in internal validation, performance was poorer in terms of step count overestimation during external validation, with the candidate 30 s epoch model demonstrating a population-wide step count overestimation of 6.7% and an RMSE of 8.6%. Conversely, performance of the hybrid models remained consistent during external validation, with the 5/2 hybrid model demonstrating only a 0.8% overestimation, with an associated RMSE of 6.8%.

Table 4-15: Comparison of Candidate Final Models for During External Validation in Clemson Shimmer3 Dataset Participants Including All Gait Subtypes

Comparator	Ground Truth	Random Forest Regression (30 s Epoch)	Hybrid (Model 5/2)	Hybrid (Model 30/10)
Total Steps	56,682	60,475	57,138	54,866
Mean Steps	1,889	2,016	1,905	1,829
% Ground Truth Counted		106.7%	100.8%	96.8%
RMSE (Steps)		191	128	189
% RMSE		10.1%	6.8%	10.0%
MAPE		8.6%	5.1%	8.2%

Bland-Altman plots comparing final model candidates using random forest regression and hybrid models are presented in Figure 4-8. As is clear in this figure, the hybrid Model 5/2 produced a mean bias of 15 steps with tighter limits of agreement when compared to the random forest regression and the hybrid Model 30/10.

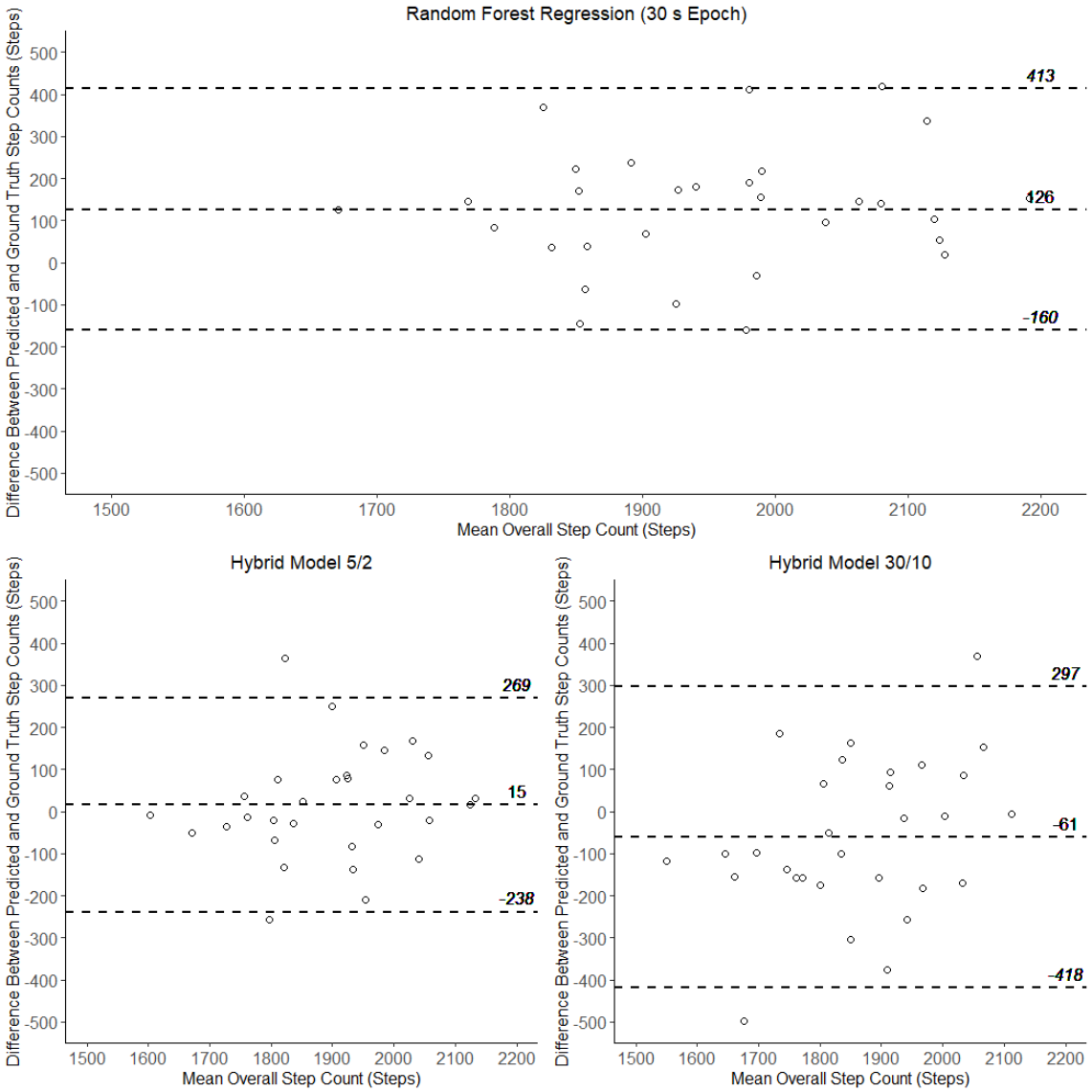


Figure 4-8: Bland-Altman plots for the comparison of step counting models in the Clemson external validation dataset, (Top) random forest regression, (Bottom left) hybrid Model 5/2, and (Bottom right) hybrid Model 30/10.

Subsequently, the final step counting model to be used for the remainder of this thesis was selected as Model 5/2. During leave-one-participant-out validation on the primary dataset, this model predicted 99.9% of ground truth steps with a 13.1% RMSE during free-living step monitoring. When applied to the external Clemson validation dataset, this model demonstrated both the lowest overall step count percent RMSE and lowest error during irregular walking. An overall running count accuracy of 0.99 (SD 0.23) was observed in the Clemson dataset with this model, encompassing an accuracy of 0.99 (SD 0.03) during regular gait, 1.10 (SD 0.16) during semiregular gait, and 0.87 (SD 0.33) in irregular, or interrupted gait.

4.3.5 Model Implementation into the Capture 24 Dataset

Walking classification of accelerometer data from the Capture 24 dataset are presented in Table 4-16, Table 4-17, and Table 4-18. When comparing binary walking classification to the Willetts and Doherty annotations, high levels of non-walking were predicted during bicycling, sit-stand, sleep, sedentary, and vehicle annotations. Conversely, epochs labelled as walking with Willetts and Doherty annotations were subsequently classified as walking at a rate of 51.3% and 53.2%, respectively. As presented in Table 4-18, epochs labelled as potential walking were classified as walking in 41.8% of cases, whereas 97.6% of likely non-walking epochs were labelled as not walking.

Table 4-16: Confusion Matrix of Predicted Walking and Non-Walking Epochs Compared to the Willetts Ground Truth Annotations in the Capture24 Dataset

	Willetts Labels (5 s epochs)						
	Bicycling	Mixed	Sit-Stand	Sleep	Vehicle	Walking	Not Labelled
Walking	912 (4.9%)	76,669 (31.8%)	15,774 (2.2%)	1,899 (0.3%)	4,318 (6.2%)	63,769 (51.3%)	86,698 (9.3%)
Not Walking	17,550 (95.1%)	164,706 (68.2%)	712,097 (97.8%)	697,771 (99.7%)	65,674 (93.8%)	60,435 (48.7%)	847,742 (90.7%)

Table 4-17: Predicted Walking and Non-Walking Epochs Compared to the Doherty Ground Truth Annotations

	Doherty Labels (5 s epochs)					
	Moderate	Sedentary	Sleep	Tasks-Light	Walking	Not Labelled
Walking	23,665 (18.9%)	13,698 (1.9%)	1,899 (0.3%)	25,132 (18.3%)	98,947 (53.2%)	86,698 (9.3%)
Not Walking	101,367 (81.1%)	719,589 (98.1%)	697,771 (99.7%)	112,467 (81.7%)	87,039 (46.8%)	847,742 (90.7%)

Table 4-18: Predicted Walking and Non-Walking Epochs Compared to the Binary Annotation Categories of Potential Walking and Likely Not Walking

	Binary Walking Labels (5 s epochs)		
	Potential Walking	Likely Not Walking	Not Labelled
Walking	126,033 (41.8%)	37,308 (2.4%)	86,698 (9.3%)
Not Walking	175,799 (58.2%)	1,542,434 (97.6%)	847,742 (90.7%)

A mean of 13,263 daily steps (6,205 SD) were counted across all Capture 24 participants, with a mean daily overall acceleration of 31.6 m*g* (10.6 SD). A plot of acceleration and step count is presented in Figure 4-9. A single outlying data point was removed from the dataset prior to linear regression based on a Cook's distance of 48 times the mean. Following removal of the single outlier, linear regression assumptions were met. Linear regression of overall acceleration and step count resulted in an adjusted R^2 of 0.58 with 476 steps (CI: 410 to 542) for every 1 *mg* increase in daily overall acceleration in this participant population.

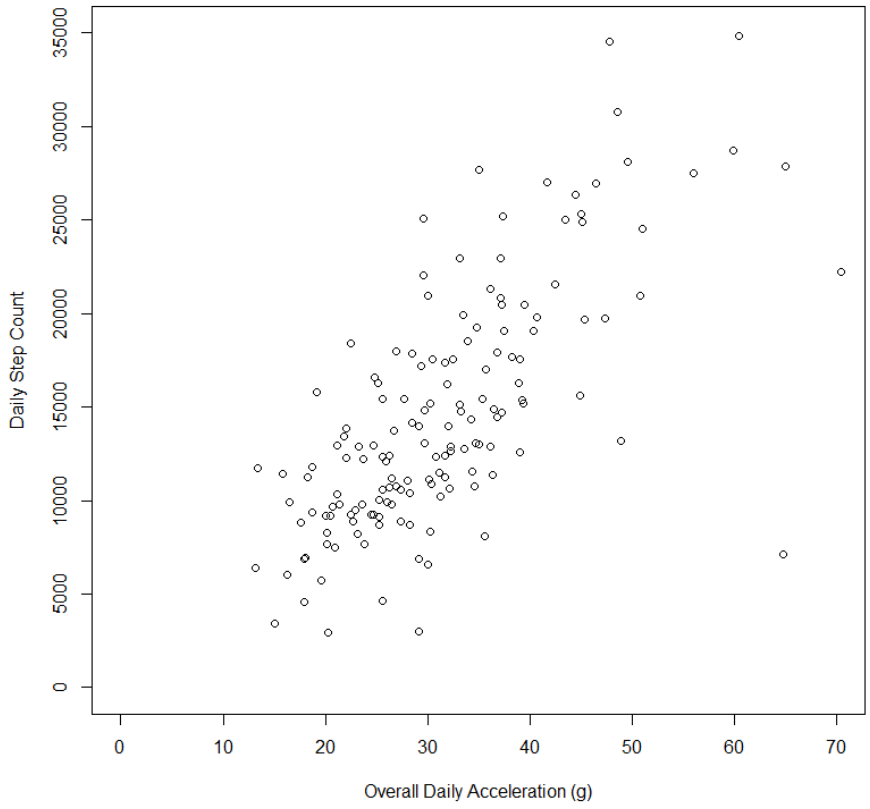


Figure 4-9: Scatterplot of overall daily acceleration versus daily step count in the Capture 24 dataset.

4.3.6 Sample Rate Comparison and Adjustment

Wrist accelerometer data was collected from a subset of 21 participants, (17 F, 4 M) mean 56.5 years of age [24 to 81 years], who did not complete ground truth video recording during the study described in Chapter 3. Results from step counting of this dataset with both the Model 5/2, and a duplicate model trained on 25 Hz data (Model 5/2-25) are presented in Table 4-19. Bland-Altman plots (Figure 4-10) demonstrate a mean bias of 2718 steps and limits of agreement approximately ± 4800 steps in 25 Hz data when assessed using the 100 Hz trained Model 5/2. This bias is a primarily a result of a systematic overestimation of 25 Hz step count in increasingly active participants. Conversely, when the 25 Hz trained model, Model 5/2-25, was applied to the 25 Hz data, step count of the 25 Hz data

a mean bias of -495 steps was observed, with tighter limits of agreement of approximately ±2700 steps and no systematic deviation from the mean bias in more active participants.

Table 4-19: 24-Hour Step Counts in Participants Wearing AX3 Accelerometers on the Dominant Wrist, Recording at Both 25 Hz and 100 Hz

	100 Hz Data	25 Hz Data	
	Model 5/2	Model 5/2	Model 5/2-25
Mean Steps [SD]	14,630 [5,803]	17,348 [6,943]	14,135 [5,766]

Step count was estimated using models trained on either 100 Hz or 25 Hz data

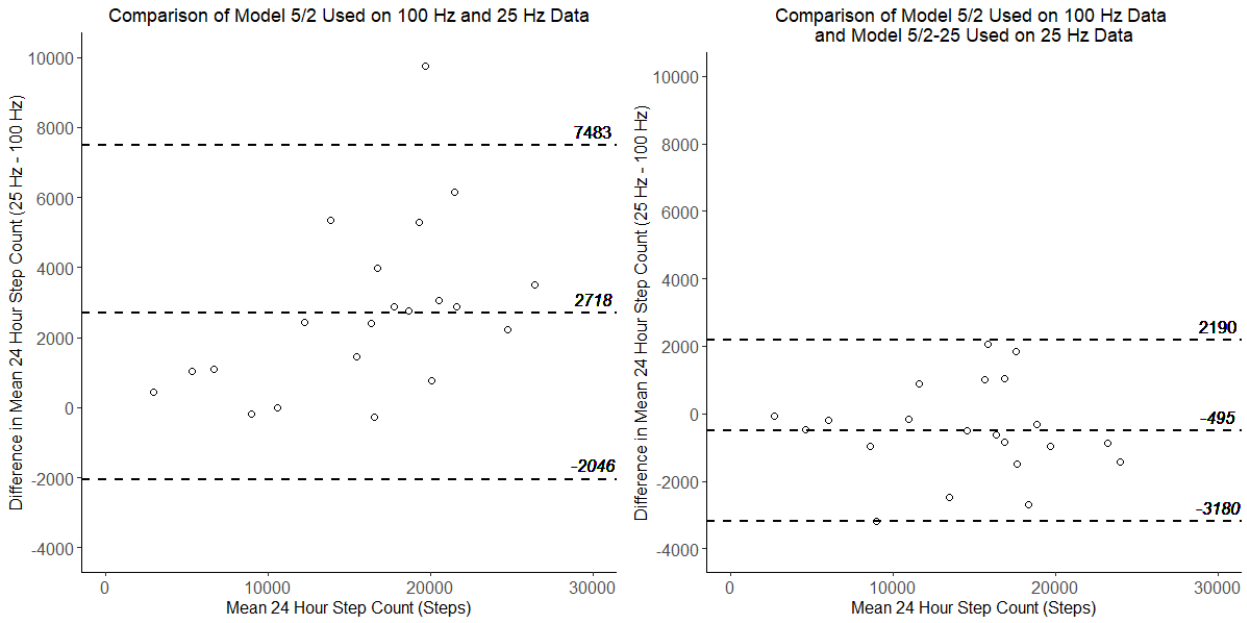


Figure 4-10: Bland-Altman plots comparing mean 24 hour step count in 100 Hz and 25 Hz data, where (top) the primary hybrid model trained on 100 Hz data was applied to both 100 Hz and 25 Hz recordings, and (bottom) the primary 100 Hz trained model was applied to 100 Hz data, while a secondary 25 Hz trained model was applied to 25 Hz recordings.

4.4 Discussion

Daily step count is a popular metric for monitoring physical activity in the general population and is frequently used in clinical studies of physical activity, particularly when commercial wearables are selected in study methodology. Research grade accelerometers are frequently selected preferentially in epidemiological studies of physical activity for their improved access to raw accelerometer data and the ability to control and fine tune data analysis for specific study needs. Currently, few ground-truth step labelled datasets are available to the research community for generation of validated step count of wrist-worn research-grade accelerometer data. In this study we generated a labelled dataset and created an externally validated step counting algorithm, demonstrating a walking classification F1 metric of 0.85, predicting 99.9% of ground truth step counts with a RMSE of 13.1% in internal validation, while predicting 100.8% of all ground truth step counts with a MAPE of 5.1% in an external validation dataset collected by a different measurement device across a combination of gait subtypes. This model demonstrates the utility of a hybrid machine learning and peak detection methodology for training a step counting algorithm for use in large scale epidemiological physical activity studies.

Three different step counting algorithms were assessed in this study: a moving average peak detection, random forest regression, and a combination of random forest activity classification and threshold-based peak detection. Prior studies have implemented variations of peak detection, frequency domain analysis, and machine learning classification models for step classification across body placements of wearables and smartphones with varied accuracy [184,186,189–195]. The majority of these studies have relied on training data from a very small number of

participants, collected data exclusively during treadmill, or performed comparisons of relative agreement in step counts between devices in the absence of a true gold standard.

As an initial evaluation of time domain based peak detection, I performed step count detection using the moving average based peak detection algorithm first described by Cho et al, implemented in the command line version of the Open Movement software created for use with the Axivity AX3 accelerometer. This step counting model, initially designed for wrist-based step detection, represents an improvement over routine threshold-based peak detection routinely used in early step counting algorithms, which can be sensitive to external noise[189]. While the authors of this algorithm reported an average step counting error of 1-2% in their dataset, we observed significantly greater error against ground truth steps in the current study. As a reference, the algorithm was tested on both hip and wrist collected accelerometer data at 25 Hz and 100 Hz. Hip based step detection was shown to have good results with this method of step detection, with far poorer performance in wrist-based data collection.

Random forest regression provides a simple and efficient machine learning alternative to peak detection to estimate step count based on accelerometer-based features. To our knowledge, no other publicly available step detection algorithm has attempted to implement random forest regression for estimation of step counts at the epoch level. This method provided reasonable levels of accuracy within internal and external validations, particularly in terms of estimating total number of steps across the entire sample. The regression performed best over longer epoch periods, which lends toward increased computational efficiency versus shorter analysis windows. However, random

forest regression was highly susceptible to overestimating step counts in participants with low activity volume, and underestimating step counts in participants with high levels of activity, as demonstrated by the 4 to 6% underestimate of step counts during normal gait and 102 to 110% overestimation of steps in irregular gait within the Clemson external validation dataset.

Hybrid machine learning and peak detection for step counting was first described by Bagui et al[184]. In their study, the authors attempted binary walking classification with k-nearest neighbour, support vector machine, and random forest machine learning algorithms, finding best classification performance with the random forest model. Following classification, the authors used double autocorrelation to smooth time series data for subsequent step counting. In the current study, I used a random forest classification algorithm with hidden Markov model for binary walking classification. In a secondary analysis, timeseries smoothing was implemented with a 5 Hz lowpass filter, which did not significantly impact the overall peak detection performance of our model. Bagui et al.[184] trained their model on a small, presumably proprietary developmental dataset in partnership with Actigraph. Fourteen datasets of an unspecified length were included from six participants during walking or running. No information is provided as to the nature of the training data as free-living or laboratory based. In the current model, I trained the classifier with 1-hour free-living datasets from 39 participants and performed subsequent external validation on a mean 37 minutes of activity from each of 30 participants in a varied gait, laboratory-based dataset. The best overall model for step counting prediction in this study utilised a combination of machine learning activity classification alongside peak detection step counting. This hybrid methodology allowed for the removal of classified times of unlikely walking, to

allow for peak step detection without extraneous data from wrist movement and activity not associated with gait. Subsequently, this method did not demonstrate the overestimation of steps during low intensity activity, as seen in moving average peak detection.

In the development of the Clemson Shimmer3 dataset, Mattfeld et al.[187] applied representative peak detection, threshold crossing, and autocorrelation step counting algorithms to assess algorithm accuracy across sensor body location. The authors reported moderately improved step count accuracy at the hip and ankle in comparison to wrist step counting measurement. In that study, model accuracy was combined across wrist, hip, and ankle locations, and observed best performance in autocorrelation-based step counting methods, with running count accuracy of 0.95 (SD 0.24) in regular gait, 0.93 (SD 0.17) in semiregular gait, and 1.11 (SD 0.44) in irregular gait averaged across the three body positions[187]. In comparison, we report running count accuracy only in wrist-based data, with an accuracy of 0.99 (SD 0.03) during regular gait, 1.10 (SD 0.16) during semiregular gait, and 0.87 (SD 0.33) during irregular gait. The current hybrid machine learning wrist-based model outperformed autocorrelation model during both regular and semiregular walking, with an absolute 2% reduction in accuracy versus the autocorrelation method in irregular walking. In all cases, interparticipant variability in step count accuracy was substantially improved using the hybrid machine learning model. Subsequently, a machine learning based step counting algorithm at the wrist can provide similar, if not improved, step counting accuracy to data collected at other more traditional sensor locations.

In Chapter 2 the effect of data collection sampling rate in the AX3 was demonstrated, in that a reduced sampling rate significantly affects overall activity

metrics and machine learning features. For this reason, I tested the 100 Hz trained model on 25 Hz data, in preparation for data analysis for long-term longitudinal activity monitoring when a lower sampling rate is required. Accordingly, I found that a reduced AX3 sampling rate did impact the performance of the step detection model. Because sampling rate did not affect the step detection accuracy of the peak detection model described in 4.3.2, it is likely that the difference in machine learning features due to sampling rate is driving a difference in the walking classification portion of the hybrid step counting model. To account for this difference and enable accurate step counting in 25 Hz data, I retrained a secondary step counting model to be used exclusively with 25 Hz data collected from the AX3. This model demonstrated good step counting performance and is the basis for step counting in the following analysis of clinical data in Chapter 6 and Chapter 7.

During this study, step counting data with ground truth video was additionally collected at 100 Hz and at 25 Hz at the hip. This data was not analysed in the current study but will be available for use for further research following data archival. This hip data is an added resource for the physical activity community and can be used to generate or validate hip-based step detection models with the AX3 accelerometer in future studies.

4.4.1 Strengths and Limitations

Recently published recommendations for the optimal validation of wearable step counting including free-living validation of step counts during all waking hours of a minimum 15 participants over a week of data with ground-truth video labelled by two annotators. The authors acknowledge this recommendation is logistically daunting and may not be feasible and was not targeted with the current

methodology. The current study resulted in a step counting algorithm with good accuracy based on the largest dataset of free-living wrist-based accelerometer data with ground truth video step annotation. Ground truth video was labelled by two independent annotators and was collected from a group of participants ranging in age from 19 to 81 years of age. The ANSI/CTA standard for assessing step count accuracy requires devices to exhibit less than 10% MAPE in more than 20 participants during treadmill walking and running. While the current study did not assess algorithm accuracy in accordance with the ANSI/CTA standard, external validation of the final walking model in the Clemson dataset demonstrated a 2.1% MAPE during approximately 12 minutes of regular walking for 30 study participants. While improvements in the methodology could be made in terms of increased time of data collection and a more specifically targeted demographic, the current model has demonstrated good step counting performance during both internal and external validations. There continues to be room for improvement of the model, and further work could be done to improve activity classification and more finely tune peak detection. Additional work is needed to improve generalisability of the algorithm across different devices, sampling rates, and paediatric or targeted clinical populations.

4.5 Conclusions

While the aggregation of daily step count as a metric for overall physical activity levels does not provide the versatility and level of insight that other accelerometer-based metrics offer, step count remains a popular and accessible outcome metric in public health and clinical research. This study demonstrates the utility of machine learning approaches for improving step count detection in wrist-worn accelerometer data and enables the derivation of step count from raw

accelerometer data to a reasonable degree of accuracy during studies of free-living physical activity. Implementation of a hybrid machine learning step counting at the wrist enables step detection at a comparable, if not improved, accuracy compared to standard step counting algorithms for hip-based step detection, while maintaining the benefits of increased acceptability and wear time compliance in wrist-based physical activity measurement protocols.

Chapter 5: Introduction of Step Count as an Activity Metric in the UK Biobank: Difference in Physical Activity between Healthy, End-Stage Arthritic, and Postoperative Hip and Knee Arthroplasty Patients

5.1 Introduction

Physical activity and mobility following lower limb arthroplasty is an area of focus for rehabilitation and perioperative care in an effort to improve joint function and patient quality of life. Recent randomised controlled trials have implemented behaviour change interventions to increase physical activity in end-stage arthroplasty patients[196], and in the first 6 to 12 months following surgery[197–201]. In many current studies, most orthopaedic patients report perceived postoperative functional improvement on patient reported outcome measures, yet these gains may be a result of pain relief associated with joint arthroplasty and not a true increase in activity and function[202]. Overall, studies using objective activity measurement in relatively small or poorly matched cohorts have led to conflicting evidence regarding postoperative levels of physical activity following joint replacement surgery[39]. Use of the UK Biobank provides the opportunity to examine physical activity in large clinical cohorts of orthopaedic patients within a well-controlled accelerometer-based protocol with linkage to a reliable and regularly updated electronic medical record database. Subsequently, this database allows for investigation into the association between physical activity, end-stage osteoarthritis, and lower limb arthroplasty, while addressing the challenges of small sample size and poorly controlled confounders of physical

activity which hinder generalisability of several current studies in the orthopaedic literature.

While epidemiologists and clinical trialists most frequently report physical activity via validated measures of acceleration magnitude, accelerometer counts or machine learning activity classification, daily step count is a popular outcome measure across scientific disciplines[203] due its ease of interpretability. In 2020, Saint-Maurice et al. published a study associating step count and all-cause mortality in nearly 5,000 participants in the NHANES database[204]. In 2022, Paluch et al. presented a meta-analysis of step count and all-cause mortality in 15 studies and 47,471 participants[205]. Similarly, Hall et al.[206] identified 17 studies including 30,000 participants investigating the association between step count and a range of health outcomes. Notably, a large percentage of current physical activity studies of knee arthroplasty patients report step count as a primary activity metric, noted in Section 2.4.2. Despite the popularity of wrist-based step counting, a step counting algorithm has never been applied for use in the UK Biobank—the world's largest wrist-mounted physical activity dataset. The application of validated step-based activity metrics into the UK Biobank offers the potential to provide insight into physical activity-related health outcomes alongside electronic health record linkage and genomic sequencing using an intuitive measure of physical activity easily consumed by the lay public.

Using the UK Biobank Physical Activity Cohort, I sought to compare activity profiles of lower limb arthroplasty patients with matched non-arthritic peer groups using both acceleration and step-based physical activity metrics. While extensive work has been completed validating acceleration-based metrics in this cohort[53], no similar work has been performed using daily step count. Therefore, the current

chapter was designed with two primary aims: 1) to evaluate the use of step-based metrics for the purpose of assessing physical activity in a large-scale cohort, and 2) to determine the difference in physical activity between participants with end-stage knee or hip arthritis, participants having undergone knee or hip arthroplasty, and matched non-arthritic controls.

5.2 Methods

Accelerometer-based physical activity data for this study was gathered from the UK Biobank Physical Activity Cohort. UK Biobank participants were invited to wear an Axivity AX3 accelerometer on their dominant wrist for 24 hours a day over a period of 7 days. Over 100,000 of the 500,000 UK Biobank participants agreed to participate and contributed a single week of physical activity data between 2013 and 2016.

5.2.1 Standard Accelerometer Data Processing

5.2.1.1 Overall Acceleration

To process this data, I followed methods described by Doherty et al.[53] to generate the primary acceleration metrics used in the current study. Acceleration data was collected in three axes at a sampling rate of 100 Hz and a range of ± 8 m*g*. Overall acceleration and wear time metrics from participant accelerometer data were processed by curators of the UK Biobank Physical Activity Cohort and are made available within the standard UK Biobank dataset. The metric of overall acceleration was derived using raw accelerometer data processed through a series of resampling, calibration, and filtering. Firstly, linear resampling was performed to eliminate time-errors in the raw accelerometer data. Data was calibrated to local gravity based on the identification of 10-second stationary periods within the raw data when all three accelerometer axes have a standard

deviation of less than 13 m*g.* A lowpass Butterworth 20 Hz filter was applied to eliminate any machine noise and subsequently, acceleration from all three axes was merged into a single summary metric by calculating the Euclidean norm of all three axes, equivalent to the LFENMO metric as described in Section 3.3.4. Finally, periods of non-wear time greater than 60 minutes were removed from the data and imputed with average acceleration values from the valid data during the same time from other days in the participant's dataset. Valid wear time was defined as a dataset including at least 72 hours of wear over the 7-day monitoring period, in addition to wear time in each of the 24 one-hour periods during the week[53].

Overall acceleration, valid wear time, calibration and data quality, and time spent under a range of acceleration cut-point thresholds are all standard metrics calculated through the process described above and are made available in the standard UK Biobank data showcase.

5.2.1.2 Machine Learning Activity Classification

To assess time use of physical activity behaviours in this dataset, I used a previously-developed, open-source machine learning activity classification model, applied to all UK Biobank participants with valid accelerometer wear[81]. The physical activity classification model was previously trained using the Capture24 dataset, described in Section 4.2.1.2, and was deployed using the Biobank Accelerometer Analysis Tool, introduced in Section 3.3.3. This machine learning model incorporates a combined random forest and hidden Markov model activity classification procedure, predicting physical activity behaviours using the rotation-invariant feature set described in Section 3.3.5 and Table 3-6, and

includes classification labels for time spent in sleep, sedentary activity, light activity, and moderate-to-vigorous activities.

5.2.1.3 Activity Fragmentation

Active-To-Sedentary-Probability (ASTP) is a simple metric developed by Schrack et al[207,208]. to assess how fragmented a person's daily bouts of physical activity tend to be. This metric is designed to assess the likelihood that during one minute of activity, the participant's next minute of activity will be one of inactivity or rest. ASTP is calculated as the number of active bouts during a 24-hour day divided by the total number of minutes in that day the participant spent above a predefined, population- and device-specific acceleration cut-point for sedentary behaviour. For the purpose of this study, an accelerometer cut-point of 62.5 m*g* served as a dividing line between active and sedentary behaviour, based on the recent work of Fraysse et al[209] defining sedentary cut-points in older adults using the GENEActiv accelerometer. Overall participant ASTP was defined as the median ASTP across valid participant wear days.

5.2.2 Implementation of Step Detection into the UK Biobank

5.2.2.1 Calculation of Daily Step Count

Participant daily step counts were derived from all raw accelerometer datasets using the UK Biobank Accelerometer Analysis Tool in conjunction with the final 100 Hz step detection algorithm (Model 5/2) developed in Chapter 4 and described in Section 4.3.4.2. As a first step, a binary random forest classifier with hidden Markov Model smoother was deployed to classify raw triaxial accelerometer data into walking or non-walking segments. Accelerometer files for each participant were reprocessed using the Biobank Accelerometer Analysis Tool using standard processing steps, as described in Section 5.2.1.1 with summary data and machine

learning feature data collated into 5 second epochs. The binary machine learning walking classifier was then applied to rotation-invariant features (Table 3-6) to produce walking and non-walking labels during all non-imputed accelerometer wear time.

Following machine learning activity classification, a bespoke Python script incorporating iteratively tuned peak detection parameters (Table 4-5, Model 5/2) was implemented for step detection during all walking-labeled time points within the raw accelerometer data. Daily step count and the individual timestamps for each detected step were collected for all participants. Overall daily step count was reported as the median number of steps taken all days for each participant. Valid wear was defined in parallel with Doherty et al.[53], as participants with at least 72 hours of wear time and with some valid wear contributions in each of the 24 diurnal hours.

5.2.2.2 Calculation of Peak Step Cadence

Peak walking speed for each participant was estimated by calculating mean peak cadence as previously described by Saint-Maurice et al[204]. For each day of wear, individual step timestamps were used to calculate the number of minute-level steps taken. One minute peak cadence was then defined as the 1-minute interval each day with the greatest number of steps detected. Thirty-minute cadence was defined as the mean number of steps taken in the thirty 1-minute intervals (not required to be consecutive minutes) with the greatest number of steps for each day.

5.2.2.3 Comparison of Age and Sex-Related Trends of Step Count and Overall Acceleration

To compare daily step count to the standard and previously validated overall acceleration metric[66], daily steps were first assessed across all biobank participants between the ages of 45 and 79 with valid accelerometer wear time. Overall acceleration and daily step count were calculated for female and male participants with age groups defined as 45-54, 55-64, 65-74, and 75-79 (age categorisation of participants in agreement with Doherty et al.[53]). Age and sex-related changes in physical activity were compared for both acceleration and daily step count activity metrics.

5.2.2.4 Comparison of Peak Cadence and Self-Reported Usual Walking Pace

As a second assessment in the validity of step detection with the UK Biobank, 1-minute and 30-minute peak walking cadence was compared to self-reported usual walking pace across all participants. Upon enrolment in the biobank cohort, participants answered the question: “How would you describe your usual walking pace”, with the response options given as Slow (less than 3 miles per hour), Steady (between 3-4 miles per hour), or Brisk (more than 4 miles per hour)[210]. Comparisons between calculated peak cadence and self-reported walking speed were made across participant sex.

Similar to the analysis performed in Section 5.2.2.3, measures of 1-minute and 30-minute cadence speed were also compared across age and sex groups for an assessment of similar trends in age and sex-based changes.

5.2.3 Participant Cohorts Based on Clinical Status

5.2.3.1 Clinical Data Linkage

The UK Biobank provides regularly updated linkage to Hospital Episode Statistics (HES) data for all participants with clinical coding from hospital inpatient admissions as managed by NHS Digital. Data linkage dates back from 1997 for participants in England, 1998 for participants in Wales, and 1981 for participants in Scotland[211]. Operative codes are listed in the OPCS-4 coding classification system, while ICD-9 and ICD-10 codes are used for in-hospital admission episode diagnoses.

5.2.3.1.1 Charlson Comorbidity Index

A modified Charlson Comorbidity Index (CCI) was used to quantify the multi-morbidity status of all participants[212,213]. CCI was calculated based on ICD-10 data linkage for each participant within a 5-year lookback window from the start date of the participant's accelerometer wear period. This modified version of the CCI produces a scoring range from 0-50, with a higher score indicating a greater risk of one-year mortality.

5.2.3.1.2 Operative Coding

OPCS-4 code definitions from the National Joint Registry[214] were used to identify participants with a medical history of primary or revision knee or hip arthroplasty. Standalone codes OPCS-4 for primary and revision knee and hip arthroplasty are presented in Table 5-1 and relevant combination codes are presented in Table 5-2. ICD-9 and ICD-10 diagnostic codes were used to identify diagnoses of hip or knee osteoarthritis at any timepoint. ICD-9 codes for hip and knee osteoarthritis included 71515 and 71516. ICD-10 codes for diagnosed knee arthritis included M170, M171, M172, M173, M174, M175, M179, and associated

sublevels. ICD-10 codes for diagnosed hip arthritis included M160, M161, M162, M163, M164, M165, M166, M167, M169 and associated sublevels.

Table 5-1: OPCS-4 Standalone Codes of Operation and Location

Primary Knee Arthroplasty	Revision Knee Arthroplasty	Primary Hip Arthroplasty	Revision Hip Arthroplasty
O181	O180	W371	W370
O188	O182	W378	W372
O189	O183	W379	W373
W401	O184	W381	W374
W408	W400	W388	W380
W409	W402	W389	W382
W411	W403	W391	W383
W418	W404	W398	W384
W421	W410	W399	W392
W428	W412	W931	W393
W429	W413	W938	W395
	W414	W939	W462
	W420	W941	W472
	W422	W948	W482
	W432	W949	W932
	W425	W951	W933
		W958	W940
		W959	W942
			W943
			W952
			W953
			W954

Adapted from the National Joint Registry[214]

Table 5-2: OPCS-4 Combination Codes of Operation and Location

Primary Knee Arthroplasty	Revision Knee Arthroplasty	Primary Hip Arthroplasty	Revision Hip Arthroplasty
W521 (with: Z846, Z765, Z845, Z844, Z774, or Z787)	W424 (with: Y032 or Y037)	W521, W531, W541, or W581 (with: Z843, Z761, or Z756)	W394 (with: Y032 or Y037)
	W522, W523, W532, W533, W542, W543, W544, W553, W564, W574, W582, W603, W613, W641, or W642 (with: Z846, Z765, Z845, Z844, Z774, or Z787)		W522, W523, W533, W542, W543, W544, W572, W574, W582 (with: Z843, Z761, or Z756)

Adapted from the National Joint Registry[214]

5.2.3.2 Knee Arthroplasty Cohorts

Participants with any medical history of knee arthroplasty, primary or revision, were identified through a medical record search using the appropriate coding combinations as described in Section 5.2.3.1.2. To create a clean, well-defined cohort for investigation, participants with only a single, unilateral primary knee arthroplasty at any point in their medical records were selected. Participants who had one or more subsequent revisions of the primary knee arthroplasty were also included. Once knee arthroplasty patients were identified, the cohort was further split into two final cohorts based on operative date. The first cohort, defined as the End-Stage Knee Arthritis cohort, was created from participants who wore an accelerometer less than 12 months before they underwent primary knee arthroplasty. Thus, these participants participated in the UK Biobank Physical Activity Cohort with an imminent need for knee arthroplasty. A second cohort, the

Post-Operative Knee cohort, was created from participants who wore an accelerometer more than 12 months after undergoing a single primary knee arthroplasty.

5.2.3.3 Hip Arthroplasty Cohorts

Hip arthroplasty cohorts were defined in the same manner as knee arthroplasty cohorts, described in Section 5.2.2 above. Clinical coding as defined for primary and revision hip arthroplasty in Table 5-1 and Table 5-2 along with associated episode dating was used to define End-Stage Hip Arthritis and Post-Operative Hip Arthroplasty cohorts.

5.2.3.4 Non-Arthritic Healthy Control

A cohort of healthy, non-arthritic controls was selected for comparison with knee and hip arthroplasty cohorts. Controls were selected as participants with valid accelerometer wear time, no medical record of hip or knee arthroplasty, no hip or knee osteoarthritis diagnosis, and no self-report of chronic (3+ months) of hip pain or knee pain at any follow-up appointment or interaction with the UK Biobank.

5.2.3.4.1 Wear Season

To facilitate participant matching in accordance with potential physical activity seasonality, participant data was categorised by season with December, January, and February labelled as winter, March, April, and May labelled as spring, June, July, and August labelled as summer, and September, October, and November labelled as autumn.

5.2.4 Statistical Analysis

All statistical analysis was performed using installations of RStudio (v1.2.5042) and R (v3.6.2) on the Biomedical Research Cluster at the University of Oxford Big Data Institute.

5.2.4.1 Comparison of Step Count and Overall Acceleration

Statistical methods for the analysis of step count across the entire UK Biobank Physical Activity cohort (Section 5.2.2) were as follows: Kruskal-Wallis tests were performed to evaluate the change in activity metrics across age levels in both female and male groups. Sub-group analyses of activity metrics between female and males at each age level were performed using the Wilcoxon Sum Rank test. Boxplots are presented for activity metrics across age and sex subgroups with median and interquartile range reported in all cases. Spearman's rank correlation was performed between acceleration and median daily step count for each sex and age group, as well collectively across all participants. A linear regression model was performed to assess the overall relationship between acceleration and daily step count in this cohort, for direct comparison to the relationship between acceleration and steps in the Capture 24 dataset presented in Section 4.3.5.

5.2.4.2 Step Detection for Calculation of Peak Walking Cadence

Kruskal-Wallis tests were performed to assess differences in cadence across categories of self-reported walking pace and across differences across age and sex categories. Sub-group analyses were performed using the Wilcoxon Sum Rank test.

5.2.4.3 Clinical Cohort Comparisons

To compare physical activity metrics between selected clinical cohorts (Section 5.2.2), matched analysis was performed using the "MatchIt" R package (v.4.3.0).

Four-to-one matching was performed for all comparisons between healthy non-arthritic controls and clinical cohorts, implementing nearest neighbor matching with a caliper value of 0.2 standard deviations. Covariates for matching were selected from variables having a significant correlation with physical activity in other UK Biobank studies. These covariates included age at the time of wear, sex, ethnic background, education level, smoking status, alcohol consumption status, Townsend Deprivation Index, wear season, and CCI[53,81,215]. Comparisons of physical activity metrics between cohorts was performed using the Wilcoxon Sum Rank test, with median and interquartile range reported for all activity metrics.

A final comparison between preoperative and postoperative cohorts was performed using a linear regression model in both hip and knee arthroplasty cohorts. Physical activity covariates used in matching were retained for the linear regression. Separate models were generated to assess each physical activity metric between end-stage arthritis and postoperative cohorts.

5.3 Results

5.3.1 Physical Activity in the Full UK Biobank Physical Activity Cohort

5.3.1.1 Population Demographics

Demographics for all participants in the UK Biobank physical activity cohort with valid wear is presented in Table 5-3. A total of 54,291 female (56%) and 42,126 (44%) male participants between the ages of 45 and 79 are included in this analysis. The cohort is heavily skewed with respect to ethnicity, with 97% of participants self-describing as white.

Table 5-3: Demographics of Full UK Biobank Physical Activity Cohort

	Full UK Biobank Activity Cohort	
Characteristic	Female	Male
Overall, N (%)	54,291 (56)	42,126 (44)
Age, Median [IQR]	62.7 [55.8–68.0]	64.5 [57.1–69.3]
Ethnicity, N (%)		
Asian	561 (1)	552 (1)
Black	493 (1)	311 (1)
Other	841 (2)	534 (1)
White	52,396 (97)	40,729 (97)
Body Mass Index, Median [IQR]	25.3 [22.9–28.6]	26.7 [24.6–29.4]
CCI, Median [IQR]	0 [0–4]	0 [0–5]
Wear Season, N (%)		
Spring	12,452 (23)	9,442 (22)
Summer	14,330 (26)	11,001 (26)
Autumn	16,145 (30)	12,621 (30)
Winter	11,364 (21)	9,062 (22)
Education, N (%)		
School Leaver	14,167 (26)	8,939 (21)
Further Education	17,357 (32)	14,522 (34)
Higher Education	22,767 (42)	18,665 (44)
Smoking Status, N (%)		
Never	33,021 (61)	21,878 (52)
Former	17,933 (33)	16,665 (40)
Current	3,199 (6)	3,463 (8)
Alcohol Consumption, N (%)		
Never	3,568 (7)	1,902 (5)
≤ 3 Days Per Week	27,470 (51)	16,287 (39)
4+ Days Per Week	23,204 (43)	23,905 (57)
Townsend Deprivation, N (%)		
Least Deprived (<-3.8)	13,420 (25)	11,000 (26)
Second Least Deprived (-3.8 to -2.5)	12,832 (24)	10,089 (24)
Second Most Deprived (-2.5 to -0.2)	14,000 (26)	10,730 (25)
Most Deprived (>-0.2)	13,979 (26)	10,257 (24)

5.3.1.2 Comparison of Age- and Sex-Related Trends with Overall Acceleration

Summary physical activity metrics overall acceleration and daily step count, grouped by age and sex, are presented in Table 5-4. Similar age and sex-related trends in physical activity were mirrored in both acceleration and daily step count. A statistically significant difference across age groups was seen in both men and women in both metrics ($p<0.001$). Similarly, a significant difference in activity between females and males was seen at each age range ($p<0.001$). A strong positive correlation of 0.74 was observed between acceleration and median daily steps across all participants. Increasingly positive correlation between step count and acceleration was seen with increased age in both sexes. Linear regression of acceleration to steps resulted in an adjusted R^2 of 0.51 with 363 (CI: 361 to 365) steps for every 1 m*g* increase in daily overall acceleration in this cohort.

Table 5-4: Overall Acceleration and Step Count Activity Metrics Across All UK Biobank Participants Aged 45-79 Years with Valid Accelerometer Wear Time

	Female		
Age, years	Acceleration (m*g*)	Daily Steps	Correlation Coefficient
45-54 *(n=12,206)*	30.1 [25.3-35.8]	10,399 [7,981-13,316]	0.71
55-64 *(n=20,233)*	28.1 [23.7-33.3]	10,111 [7,631-12,962]	0.74
65-74 *(n = 20,839)*	25.8 [21.6-30.6]	9,210 [6,712-11,926]	0.77
75-79 *(n = 1,013)*	23.0 [19.2-27.5]	7,589 [5,087-10,037]	0.76
	Male		
Age, years	Acceleration (m*g*)	Daily Steps	Correlation Coefficient
45-54 *(n=8,208)*	29.6 [24.6-35.9]	9,730 [7,446-12,493]	0.67
55-64 *(n=13,732)*	27.5 [22.7-33.1]	9,572 [7,083-12,428]	0.72
65-74 *(n=19,052)*	24.6 [20.3-29.5]	8,775 [6,305-11,469]	0.76
75-79 *(n=1,134)*	22.1 [18.2-26.8]	7,347 [5,157-10,117]	0.79

Activity reported as median [interquartile range]

Boxplots of acceleration and daily step count are presented in Figure 5-1. Acceleration decreased 6.6%, 8.2%, and 10.9% in females and 7.1%, 10.5%, and 10.2% in males each additional decade after the 45-54 year age group. Similarly, step count decreased 2.8%, 8.9%, and 17.6% in females, and 1.6%, 8.3%, and 16.3% in males in each additional 10 years of age.

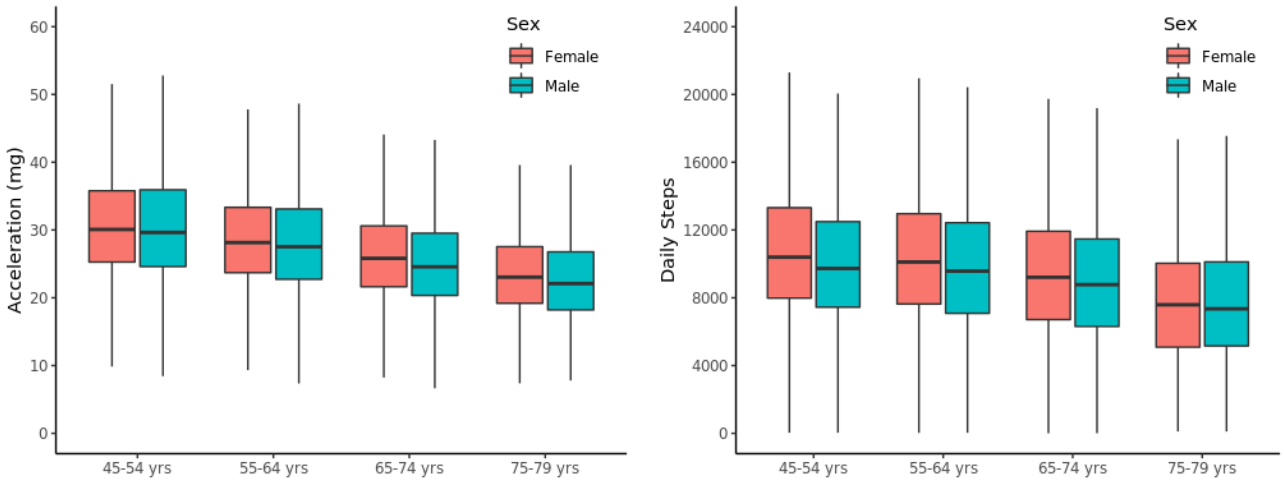


Figure 5-1: Measures of bulk physical activity in the UK Biobank Physical Activity Cohort, summarised by age and sex. (Left) Overall acceleration measured in m*g* as reported as the standard objective activity metric in this cohort, analysed to replicate Figure 4 from Doherty et al.[53] (Right) Daily step count showing the same trends and activity patterns as overall acceleration.

5.3.1.3 Comparison of Step Cadence with Self-Reported Usual Walking Pace

Boxplots of peak 1-minute and peak 30-minute cadence relative to sex and self-reported walking pace are presented in Figure 5-2. Kruskal-Wallis tests demonstrated a statistically significant difference in both cadence metrics across all self-reported pace categories as well as across sex, with significantly higher peak cadence correlating with increasing self-reported usual walking pace ($p<0.001$). The median [IQR] 1-minute peak cadence of males was 103.3 [94.8-110.1] steps per minute slow walkers, 109.9 [103.9-115.1] in steady walkers,

and 112.6 [107.0-117.9] in brisk walkers. Median [IQR] 1-minute peak cadence was of females was 105.5 [97.1-112.8] steps per minute slow walkers, 113.0 [106.9-118.9] in steady walkers, and 117.0 [111.0-122.8] in brisk walkers. Female participants had a higher measured step cadence than male participants in all comparisons (p<0.001).

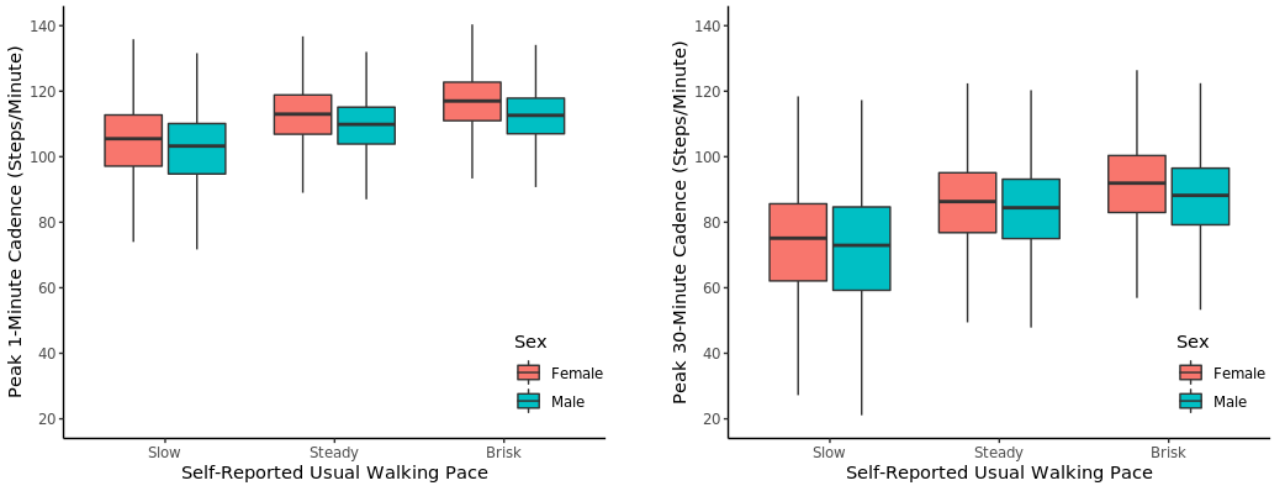


Figure 5-2: Measures of activity intensity in terms of step cadence in the UK Biobank Physical Activity Cohort, summarised by self-reported walking pace and sex. (Left) Peak 1-minute step cadence speed (Right) Peak 30-minute step cadence speed.

5.3.1.4 Age- and Sex-related Trends in Peak Step Cadence

Boxplots of peak 1-minute and 30-minute step cadence relative to sex and age groups are presented in Figure 5-3. Step cadence also decreased with increasing age in both sexes, but at a lower rate than the overall activity metrics of acceleration and step count, averaging a 2% decrease in 1-minute and 4% decrease in 30-minute peak cadence per decade of age (Table 5-5). A statistically significant difference across age groups was seen in both men and women in all calculated activity metrics (p<0.001). Similarly, a significant difference in activity between females and males was seen at each age range (p<0.001) apart from daily steps and 30-minute cadence in the 75-79 age group.

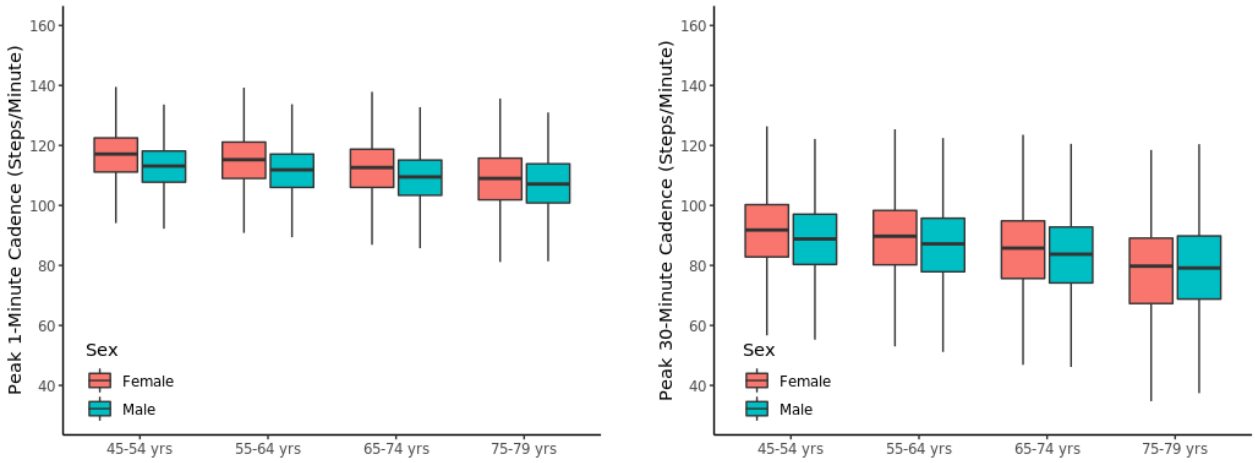


Figure 5-3: Measures of activity intensity in terms of step cadence in the UK Biobank Physical Activity Cohort, summarised by age and sex. (Left) Peak 1-minute step cadence speed (Right) Peak 30-minute step cadence speed.

Table 5-5: Peak Cadence as Calculated Across All UK Biobank Participants Aged 45-79 Years with Valid Accelerometer Wear Time

	Female	
Age, years	1-Min Cadence (Steps/min)	30-Min Cadence (Steps/min)
45-54 *(n=11,457)*	117.1 [111.1-122.5]	91.8 [82.8-100.3]
55-64 *(n=19,060)*	115.3 [109.0-121.1]	89.7 [80.2-98.3]
65-74 *(n = 19,851)*	112.6 [106.0-118.8]	85.8 [75.6-94.8]
75-79 *(n = 968)*	109.0 [101.9-115.8]	79.8 [67.3-89.1]
	Male	
Age, years	1-Min Cadence (Steps/min)	30-Min Cadence (Steps/min)
45-54 *(n=7,580)*	113.1 [107.8-118.1]	88.8 [80.3-97.1]
55-64 *(n=12,837)*	111.9 [106.0-117.1]	87.2 [77.9-95.7]
65-74 *(n=18,095)*	109.5 103.4-115.1]	83.7 [74.1-92.8]
75-79 *(n=1,085)*	107.1 [100.8-113.9]	79.1 [68.8-89.8]

Cadence reported as median [interquartile range]

5.3.2 Activity Assessment of Knee and Hip Arthroplasty Patients

5.3.2.1 Matched Analysis

A flow chart for selection of knee and hip arthroplasty cohorts is shown in Figure 5-4. For all selected participants, data analysis was performed with a median 6.4 years (a range of 5.2 to 7.8 years) of follow-up in the Hospital Episode Statistics database after accelerometer wear time, updated through 28 February 2021. Demographics of the selected cohorts are listed in Table 5-6 for knee arthroplasty cohorts and their matched controls. A total of 173 participants with end-stage knee arthritis and valid accelerometer wear a median 7.4 [4.1-9.7] months prior to primary knee arthroplasty were matched with 686 healthy controls. The post-knee arthroplasty cohort consisted of 605 participants evaluated at a median 48.5 [28.3-75.6] months after surgery, matched with 2401 healthy controls. Table 5-7 presents demographic data for hip arthroplasty cohorts and their matched controls. For participants with end-stage unilateral hip arthritis, 186 participants wore an accelerometer a median 6.5 [3.5-9.0] months prior to surgery and were matched to 744 healthy controls. A total of 904 participants wore an accelerometer a median 53.8 [30.8-95.5] months after single, unilateral hip arthroplasty and were matched with 3,610 healthy controls.

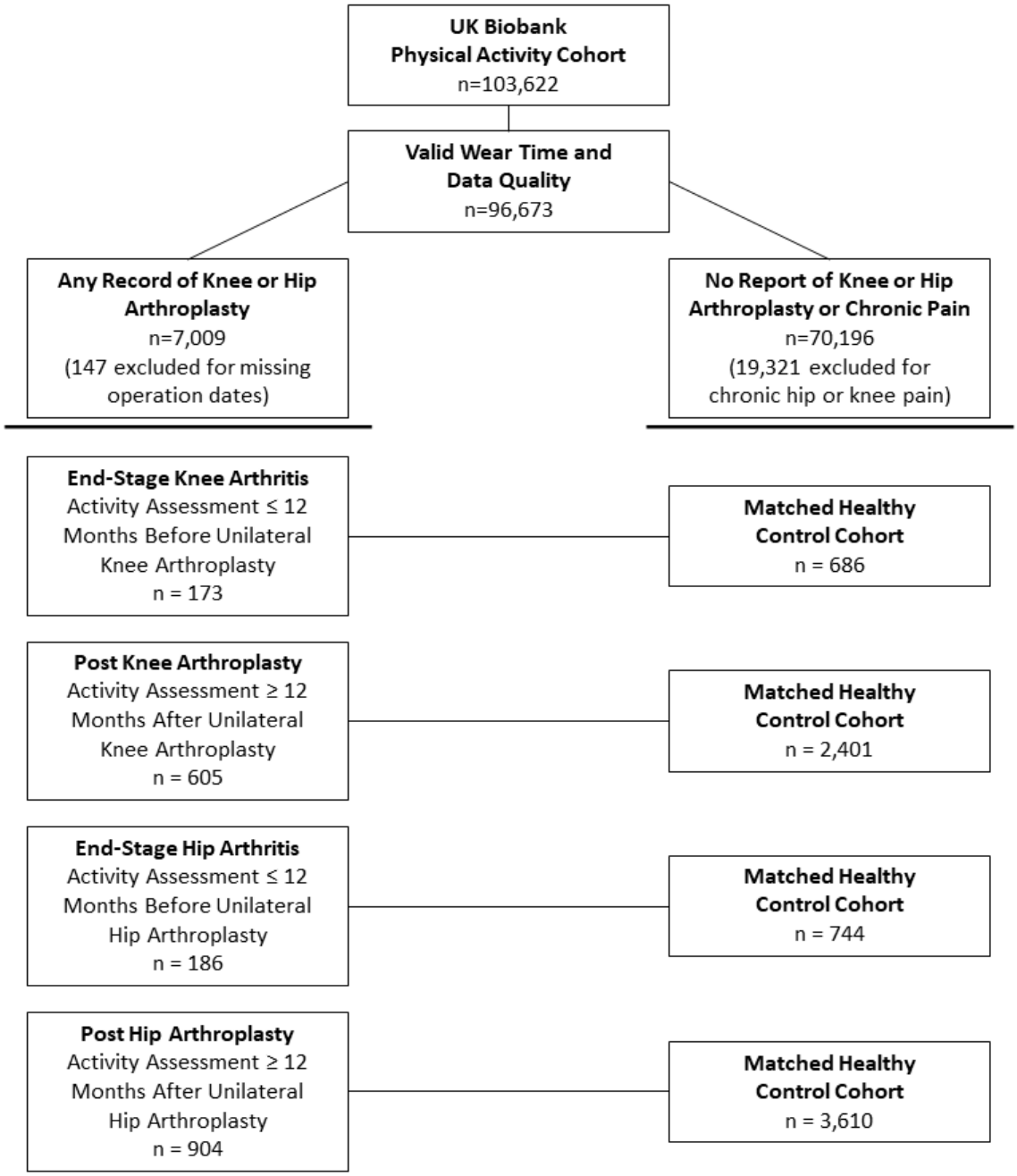


Figure 5-4: Study flowchart mapping cohort selection from the UK Biobank Physical Activity Cohort.

Table 5-6: Demographics of Knee Arthroplasty Cohorts with Matched Controls

	Preoperative Knee Cohort		Postoperative Knee Cohort	
Characteristic	≤ 12 Months Preop	Matched Controls	≥ 12 Months Postop	Matched Controls
Overall, N (%)	173 (100)	686 (100)	605 (100)	2,401 (100)
Sex, N (%)				
Female	96 (55)	392 (57)	318 (53)	1,220 (51)
Male	77 (45)	294 (43)	287 (47)	1,181 (49)
Age, Median [IQR]	68.2 [64.8–71.5]	69.1 [65.6–72.5]	69.3 [65.5–72.3]	69.6 [66.0–72.7]
Ethnicity, N (%)				
Asian	1 (1)	5 (1)	3 (0)	16 (1)
Black	1 (1)	1 (0)	3 (0)	10 (0)
Other	3 (2)	11 (2)	8 (1)	24 (1)
White	168 (97)	669 (98)	591 (98)	2,351 (98)
Body Mass Index, Median [IQR]	28.7 [26.0–32.2]	28.0 [25.4–31.6]	29.2 [26.4–33.0]	28.6 [25.8–32.3]
CCI, Median [IQR]	4 [0–11]	0 [0–9]	3 [0–9]	0 [0–8]
Wear Season, N (%)				
Spring	37 (21)	144 (21)	132 (22)	517 (22)
Summer	50 (29)	183 (27)	157 (26)	635 (26)
Autumn	48 (28)	196 (29)	202 (33)	803 (33)
Winter	38 (22)	163 (24)	114 (19)	446 (19)
Education, N (%)				
School Leaver	51 (29)	194 (28)	225 (37)	916 (38)
Further Education	65 (38)	264 (38)	225 (37)	864 (36)
Higher Education	57 (33)	228 (33)	155 (26)	621 (26)
Smoking Status, N (%)				
Never	88 (51)	350 (51)	294 (49)	1,149 (48)
Former	71 (41)	282 (41)	279 (46)	1,134 (47)
Current	13 (8)	50 (7)	31 (5)	115 (5)
Alcohol Consumption, N (%)				
Never	11 (6)	44 (6)	40 (7)	168 (7)
≤ 3 Days Per Week	81 (47)	319 (47)	288 (48)	1,088 (45)
4+ Days Per Week	81 (47)	323 (47)	275 (45)	1,143 (48)
Townsend Deprivation, N (%)				
Least Deprived (<-3.8)	33 (19)	124 (18)	141 (23)	578 (24)
Second Least Deprived (-3.8 to -2.5)	50 (29)	213 (31)	144 (24)	522 (22)
Second Most Deprived (-2.5 to -0.2)	51 (29)	197 (29)	159 (26)	685 (29)
Most Deprived (>-0.2)	39 (23)	152 (22)	161 (27)	643 (27)
Months After Surgery, Median [IQR]	-7.4 [-4.1– -9.7]		48.5 [28.3–75.6]	

Table 5-7: Demographics of Hip Arthroplasty Cohorts with Matched Controls

	Preoperative Hip Comparison		Postoperative Hip Comparison	
Characteristic	≤ 12 Months Preop	Matched Controls	≥ 12 Months Postop	Matched Controls
Overall, N (%)	186	744	904	3610
Sex, N (%)				
Female	111 (60)	428 (58)	523 (58)	2,140 (59)
Male	75 (40)	316 (42)	381 (42)	1,470 (41)
Age, Median [IQR]	67.6 [62.7–71.4]	68.1 [64.1–71.3]	68.6 [64.1–71.6]	68.3 [64.1–71.9]
Ethnicity, N (%)				
Asian	0 (0)	0 (0)	1 (0)	1 (0)
Black	0 (0)	0 (0)	5 (1)	18 (0)
Other	1 (1)	6 (1)	10 (1)	41 (1)
White	185 (99)	738 (99)	888 (98)	3,550 (98)
Body Mass Index, Median [IQR]	26.7 [24.6–29.7]	26.6 [24.2–30.0]	26.7 [24.2–29.8]	26.5 [24.0–29.4]
CCI, Median [IQR]	0 [0–8]	0 [0–8]	0 [0–8]	0 [0–8]
Wear Season, N (%)				
Spring	48 (26)	175 (24)	197 (22)	774 (21)
Summer	51 (27)	187 (25)	246 (27)	988 (27)
Autumn	53 (28)	235 (32)	281 (31)	1,098 (30)
Winter	34 (18)	147 (20)	180 (20)	750 (21)
Education, N (%)				
School Leaver	61 (33)	237 (32)	267 (30)	1,062 (29)
Further Education	44 (24)	174 (23)	292 (32)	1,168 (32)
Higher Education	81 (44)	333 (45)	345 (38)	1,380 (38)
Smoking Status, N (%)				
Never	93 (50)	374 (50)	484 (54)	1,937 (54)
Former	83 (45)	341 (46)	365 (40)	1,451 (40)
Current	9 (5)	29 (4)	55 (6)	222 (6)
Alcohol Consumption, N (%)				
Never	10 (5)	40 (5)	52 (6)	218 (6)
≤ 3 Days Per Week	77 (41)	304 (41)	376 (42)	1,489 (41)
4+ Days Per Week	99 (53)	400 (54)	474 (52)	1,901 (53)
Townsend Deprivation, N (%)				
Least Deprived (<-3.8)	38 (20)	153 (21)	261 (29)	1,024 (28)
Second Least Deprived (-3.8 to -2.5)	50 (27)	210 (28)	221 (24)	906 (25)
Second Most Deprived (-2.5 to -0.2)	49 (26)	187 (25)	221 (24)	870 (24)
Most Deprived (>-0.2)	49 (26)	194 (26)	201 (22)	810 (22)
Months After Surgery, Median [IQR]	-6.5 [-3.5– -9.0]		53.8 [30.8–95.5]	

Results from the comparison of healthy controls and participants with end-stage knee arthritis are presented in Table 5-8. No statistically significant difference between the two cohorts was observed when comparing overall acceleration, ASTP, time spent in light or sedentary activity, or time spent sleeping. Statistically significant differences between the two cohorts were observed when measuring daily steps, peak cadence, and time spent in MVPA. Relative to healthy controls, the end-stage knee arthritis cohort took a median 905 fewer steps per day, a 10.5% reduction in functional activity ($p=0.019$). Arthritic participants spent a median 4 minutes per day fewer in MVPA, a 20% reduction in comparison with their healthy peers ($p=0.002$). In terms of peak step frequency, participants with knee arthritis took 3.7 fewer steps per minute in their fastest-walking minute of the day (3.3%, $p<0.001$), and 5.4 steps per minute fewer in their 30 fastest-walking minutes per day (6.4%, $p<0.001$).

Activity results from postoperative knee participants against matched controls are presented in Table 5-9. Contrary to the end-stage arthritis cohort, no significant difference in daily step count was observed between the postoperative knee cohort and their matched peers ($p=0.181$). Significant differences between the postoperative cohort and matched controls remained in time spent in MVPA (5 minutes per day fewer [26.3%, $p<0.001$]), time spent in sedentary time (7 minutes per day fewer [1.2%, $p=0.026$]), peak 1-minute cadence (2.2 steps per minute fewer [2.0%, $p<0.001$]), and peak 30-minute cadence (3.3 steps per minute fewer [3.9%, $p<0.001$]).

Table 5-8: Physical Activity Metrics of UK Biobank Participants Wearing an Accelerometer Less Than 12 Months Prior to Unilateral Primary Knee Arthroplasty, Compared to Non-Arthritic Controls

Activity Metric	End-Stage Knee Arthritis	Non-Arthritic Controls	p
Daily Steps	7,709 [5,399–10,589]	8,614 [6,148–11,613]	**0.019**
Overall Acc (mg)	24.4 [19.2–27.8]	24.4 [20.1–29.8]	0.192
MVPA (min/day)	16 [6–28]	20 [9–35]	**0.002**
Light (min/day)	326 [249–393]	326 [267–398]	0.350
Sedentary (min/day)	551 [486–639]	547 [477–620]	0.182
Sleep (min/day)	521 [475–573]	518 [476–570]	0.879
ASTP	0.333 [0.282–0.378]	0.327 [0.276–0.376]	0.367
1 min Cadence (steps/min)	107.4 [98.6–112.1]	111.1 [103.5–117.1]	**<0.001**
30 min Cadence (steps/min)	78.9 [69.3–89.1]	84.3 [72.7–94.5]	**<0.001**

All metrics reported as median [IQR]
P-values calculated from Wilcoxon Sum Rank test used to compare cohorts

Table 5-9: Physical Activity Metrics of UK Biobank Participants Wearing an Accelerometer Greater Than 12 Months After Unilateral Primary Knee Arthroplasty, Compared to Non-Arthritic Controls

Activity Metric	Post Knee Arthroplasty	Non-Arthritic Controls	p
Daily Steps	8,146 [5,652–11,204]	8,370 [5,995–11,242]	0.181
Overall Acc (mg)	24.2 [20.1–28.9]	24.1 [20.1–29.1]	0.887
MVPA (min/day)	14 [5–28]	19 [8–36]	**<0.001**
Light (min/day)	330 [266–396]	319 [257–385]	**0.030**
Sedentary (min/day)	558 [483–628]	565 [497–639]	**0.026**
Sleep (min/day)	521 [479–564]	516 [476–562]	0.060
ASTP	0.328 [0.280–0.386]	0.324 [0.277–0.379]	0.111
1 min Cadence (steps/min)	108.4 [100.5–114.3]	110.6 [103.4–116.6]	**<0.001**
30 min Cadence (steps/min)	80.3 [68.8–89.9]	83.6 [71.8–93.0]	**<0.001**

All metrics reported as median [IQR]
P-values calculated from Wilcoxon Sum Rank test used to compare cohorts

Results from the comparison of healthy controls and participants with end-stage hip arthritis are presented in Table 5-10. Like the end-stage knee arthritis cohort, no statistically significant difference between the hip arthritic and control cohorts was observed in overall acceleration, ASTP, time spent in light or sedentary activity, or time spent sleeping. Significant differences between the two cohorts were observed when measuring daily steps, peak 30-minute cadence, and time spent in MVPA. Relative to healthy controls, the end-stage hip arthritis cohort took a median 836 fewer steps per day, a 9.1% reduction in functional activity (p=0.011) and spent a median 7 minutes per day fewer in MVPA, a 30.4% reduction in comparison with their healthy peers (p=0.001). In terms of peak step cadence, participants with hip arthritis took 2.2 steps per minute fewer in their 30 fastest-walking minutes per day (2.6%, p=0.012).

Activity results from postoperative hip participants against matched controls are presented in Table 5-11. Unlike the end-stage hip arthritis cohort, no significant difference in daily step count or cadence was observed between the postoperative knee cohort and their matched peers. Significant differences between the postoperative cohort and matched controls were observed only in time spent in MVPA, with postoperative hip participants spending 1 minute less per day (4.5%, p=0.034).

Table 5-10: Physical Activity Metrics of UK Biobank Participants Wearing an Accelerometer Less Than 12 Months Prior to Unilateral Primary Hip Arthroplasty, Compared to Non-Arthritic Controls

Activity Metric	End-Stage Hip Arthritis	Non-Arthritic Controls	p
Daily Steps	8,328 [5,881–10,698]	9,164 [6,566–11,894]	**0.011**
Overall Acc (mg)	24.8 [21.3–29.8]	25.5 [20.6–30.4]	0.985
MVPA (min/day)	16 [8–35]	23 [10–40]	**0.001**
Light (min/day)	329 [279–380]	325 [265–392]	0.751
Sedentary (min/day)	564 [498–630]	556 [492–630]	0.803
Sleep (min/day)	516 [481–560]	514 [480–556]	0.897
ASTP	0.313 [0.264–0.366]	0.313 [0.267–0.365]	0.789
1 min Cadence (steps/min)	111.0 [103.4–117.4]	112.0 [105.9–117.6]	0.070
30 min Cadence (steps/min)	83.5 [72.2–91.6]	85.7 [75.9–94.7]	**0.012**

All metrics reported as median [IQR]
P-values calculated from Wilcoxon Sum Rank test used to compare cohorts

Table 5-11: Physical Activity Metrics of UK Biobank Participants Wearing an Accelerometer Greater Than 12 Months After Unilateral Primary Hip Arthroplasty, Compared to Non-Arthritic Controls

Activity Metric	Post Hip Arthroplasty	Non-Arthritic Controls	p
Daily Steps	9,216 [6,626–12,010]	9,122 [6,634–11,870]	0.682
Overall Acc (mg)	25.6 [21.2–30.3]	25.7 [21.4–30.8]	0.594
MVPA (min/day)	21 [9–39]	22 [10–40]	**0.034**
Light (min/day)	332 [275–397]	334 [274–401]	0.953
Sedentary (min/day)	553 [483–622]	547 [480–617]	0.333
Sleep (min/day)	515 [471–560]	516 [478–560]	0.272
ASTP	0.311 [0.270–0.360]	0.312 [0.267–0.363]	0.806
1 min Cadence (steps/min)	111.9 [105.3–117.8]	112.1 [105.5–118.1]	0.447
30 min Cadence (steps/min)	85.6 [75.2–94.7]	86.0 [75.4–95.3]	0.453

All metrics reported as median [IQR]
P-values calculated from Wilcoxon Sum Rank test used to compare cohorts

5.3.2.2 Linear Regression Comparison of End-Stage Arthritis and Postoperative Cohorts

Coefficients for the linear regression comparison of end-stage knee arthritis with the postoperative cohort are presented in Table 5-12. A difference in operative status (end-stage knee arthritis vs post-arthroplasty) was not found to be a significant predictor in any physical activity metric when adjusting for all demographic variables listed in Table 5-6.

Table 5-12: Results of Linear Regressions Comparing End-Stage Arthritis and Postoperative Knee Arthroplasty Cohorts

Activity Metric	Coefficient	p
Daily Steps	466 [-163–1094]	0.146
Overall Acc (mg)	0.52 [-0.54–1.58]	0.336
MVPA (min/day)	0.5 [-3.0–4.0]	0.300
Light (min/day)	8.2 [-7.3–23.8]	0.299
Sedentary (min/day)	-9.1 [-27.1–8.9]	0.321
Sleep (min/day)	0.4 [-13.4–14.1]	0.958
ASTP	0.001 [-0.012–0.015]	0.834
1 min Cadence (steps/min)	1.3 [-0.7–3.3]	0.202
30 min Cadence (steps/min)	1.1 [-1.7–3.9]	0.441

Separate regressions for each activity metric were performed

Coefficients of linear regression for daily step count in end-stage hip arthritis and hip arthroplasty patients are found in Table 5-13. Regression of daily steps, time spent in MVPA, and 30-minute peak cadence were all significantly influenced by hip arthroplasty status. In this regression, after adjusting for demographic variables listed in Table 5-7, postoperative hip arthroplasty status was significantly associated with 1,034 more daily steps (p=0.002), 4 more minutes spent in MVPA (p=0.044), and 2.6 more steps per minute in peak 30-minute cadence (p=0.038).

Table 5-13: Results of Linear Regressions Comparing End-Stage Arthritis and Postoperative Hip Arthroplasty Cohorts

Activity Metric	Coefficient	p
Daily Steps	1,034 [384–1684]	**0.002**
Overall Acc (mg)	0.3 [-0.9–1.4]	0.659
MVPA (min/day)	3.8 [0.1–7.5]	**0.044**
Light (min/day)	4.9 [-9.7–19.4]	0.514
Sedentary (min/day)	-4.9 [-21.2–11.4]	0.556
Sleep (min/day)	-3.8 [-15.2–7.6]	0.515
ASTP	-0.005 [-0.016–0.007]	0.422
1 min Cadence (steps/min)	1.3 [-0.3–2.9]	0.116
30 min Cadence (steps/min)	2.6 [0.1–5.1]	**0.038**

Separate regressions for each activity metric were performed

5.4 Discussion

In this study I implemented metrics of step count and peak step cadence into the UK Biobank Physical Activity Cohort by processing the participants' raw accelerometer files using the novel hybrid machine learning step detection algorithm developed and validated in Chapter 4. To assess the epidemiological value and face validity of daily step-based parameters within the UK Biobank, I compared these metrics with established activity metrics and self-reported responses within the biobank dataset. Daily step count demonstrated strong correlation with patterns in overall acceleration in terms of age-related activity decline, with slightly higher levels of overall activity in women compared to men. I additionally observed agreement between self-reported walking pace and objectively measured peak cadence. Secondly, this chapter aimed to assess the impact of unilateral joint replacement on physical activity by profiling activity in patients with end-stage lower limb arthritis and those having undergone knee or hip arthroplasty. Through comparisons of fully matched non-arthritic controls and

in direct comparison between participants by surgical history, a significant positive association was observed between lower limb arthroplasty and higher functional mobility as measured by daily step count and peak walking cadence.

In the largest physical activity study to date, Doherty et al.[53] demonstrated a population-level decrease in objectively measured physical activity across age groups in terms overall measured acceleration with a mean 7.5% decrease in physical activity per decade past the age of 45. In the current study, a mean 6-11% decrease in daily step count per decade past 45 years was observed, though this reduction increased dramatically in the older age groups. Strong positive correlation between daily step count and overall acceleration was observed, with increasing correlation coefficients between steps and acceleration in higher age groups. This increased correlation may suggest a progressive shift of physical activity consisting of relatively more walking and less of vigorous activity as participants age. Cadence, the measure of steps taken per minute during walking, captures walking intensity of the participant during free-living. Prior studies have shown that walking cadence is highly correlated with walking speed, and negatively associated with age[216]. Like step count and acceleration, peak cadence also decreased with age in this study while closely matching trends of higher cadence with increased self-reported walking pace.

Joint replacement is a highly effective surgical intervention for relieving pain, restoring functional mobility, and improving patient quality of life[217]. While the effectiveness of joint replacement in indisputable, there is mixed evidence as to whether physical activity increases for the average patient following surgery. Many patients report increased physical activity and functional improvement on self-reported patient reported outcome measures; however, most studies using

objectively measured activity describe minimal gains in postoperative activity and step count in small samples of patients[39,218]. Several current longitudinal studies of objectively-measured physical activity following lower limb arthroplasty have described minimal to no improvement relative to preoperative activity, even with marked improvement in patient reported outcome measures[39,44,219]. A few studies have compared physical activity levels between postoperative hip or knee arthroplasty patients and non-arthritic controls, reporting significantly lower activity in these patients compared to their peers[220–225]. However, most current studies include a small number of participants, and those making comparisons with healthy controls have matched only on age and sex, neglecting several other physical activity covariates known to influence patient physical activity, including BMI, education, alcohol and smoking status, and other medical comorbidities[39,81,222].

An increase of 1,000 steps per day has been repeatedly attributed to lower risk of cardiovascular disease and associated complications, in addition to lower risk of all-cause mortality[206]. In both knee and hip end-stage arthritis cohorts, participants took nearly 1,000 fewer steps per day than their peers and spent significantly less time in MVPA. When comparing postoperative cohorts directly to end-stage arthritic cohorts, a significantly higher step count was observed in postoperative hip arthroplasty participants, with regression models of both hip and knee arthroplasty indicating a 1,034 step per day increase in daily step count associated with operative intervention. A 466 increase in daily step count was associated with knee arthroplasty intervention, though this lacked statistical significance. These findings correlate well with the most recently published longitudinal study of postoperative activity from Matsungaga-Myoji et al.[226], who reported a 1,020 step

increase in arthroplasty patients at 6 months and a 1,145 step increase at 2 years postoperatively. While some studies have reported postoperative patient activity may still be up to 20% less than healthy controls[223,225], the current study found that postoperative patients are similarly active to their non-arthritic peers when matching for BMI, socioeconomic, behavioral, and comorbidity status. Both postoperative hip and knee cohorts had daily step counts and overall acceleration equivalent to controls. The postoperative hip cohort participated in only 1 minute less of MVPA per day than their peers and were otherwise equivalently active across all other metrics. The postoperative knee cohort were slightly less sedentary and exhibited 11 minutes more light activity per day than their peers. However, this cohort participated in 5 fewer minutes of MVPA per day, 26% less than their matched peers. While overall activity was equivalent in terms of step count and overall acceleration, time spent in activities of elevated intensity may be a needed area of focus for postoperative rehabilitation programs for knee arthroplasty patients.

Step count differed significantly between end-stage arthritic and control populations and was also found to be significantly higher in postoperative vs. preoperative hip cohorts, yet these relationships were not observed in overall acceleration. Because the current step count metric is trained to ignore periods of non-walking activity and is insensitive to varied intensity levels, the measure of step count may highlight functional mobility in a manner potentially unobserved in the acceleration-based metrics. No differences between end-stage arthritis and matched healthy cohorts in time spend in sedentary, light, or sleep activities were observed, nor were any differences observed in activity fragmentation as measured by ASTP. This would imply that end-stage arthritis affects overall

mobility, but do not greatly affect observed behavioural patterns of sedentary and light activity. Similarly, the average duration and fragmentation of active bouts remains similar to peers whether or not the patient has end-stage arthritis or has had joint replacement surgery.

5.4.1 Strengths and Limitations

This is the largest study of physical activity in hip or knee arthroplasty patients to date. Electronic medical record linkage between the UK Biobank and NHS Digital provides detailed clinical records enabling the identification of over 7,000 participants who have contributed valid, objectively measured physical activity, and have, at some point, had at least one hip or knee arthroplasty. This study also incorporates a large control group for matching across a large set of well-defined confounders in physical activity assessment. Additionally, this study incorporates a validated step-counting algorithm from a research-grade wrist-based accelerometer. Wrist-based accelerometry leads to improved wear time compliance, while the use of a research-grade device avoids the potential limitations of commercial activity trackers, including the use of black-box step detection algorithms and limited access to raw accelerometer data. This is a cross-sectional study of physical activity, relying on a single 7-day monitoring period of physical activity. No follow-up physical activity data is available to assess change in physical activity over time. Additionally, many of the variables used in this analysis, including BMI, Townsend deprivation index, and self-reported walking pace were recorded upon enrolment in the UK Biobank. This data was potentially collected several years prior to physical activity assessment and may not be accurate for some participants at the time of accelerometer wear.

5.4.1.1 Walking Aid Bias

The current step detection algorithm was trained on a large age range of healthy participants and did not include participants using walking aids or other assistance. Some caution may need to be taken when applying a step detection algorithm to certain patient populations; step count underestimation has been documented in some commercial step counters in older persons[227], those with very slow gait[228], or with the use of walking aids[180]. According to Singh et al.[229], as many as 16% knee replacement patients have at least some dependence on walking aids at 5-years follow-up, while 25-28%[230,231] of patients may use a walking aid preoperatively. Minimal assistance is most common in knee arthroplasty patients[231,232], with a single stick making up 93% of assistive device use[231]. Because assistance is generally minimal, the effect on step count accuracy should be small in comparison to the use of crutches or stationary and rolling walkers[180]. In the current study, higher postoperative step count in arthroplasty patients may include some bias in preoperative use of walking aids. Even so, this increased postoperative step count retains and supports the overall conclusion of improved functional mobility (increased activity and/or reduced levels of aid) in postoperative arthroplasty patients. To fully quantify sources of walking aid bias on step count, further validation is needed in these patient groups.

5.5 Conclusions

The first aim of this chapter was to evaluate the use of step count and peak cadence metrics as a measure of functional activity in a large physical activity dataset. The results of this study have demonstrated the parallel trends in reduction of physical activity with increasing age in both men and women. This has been well documented in the standard, validated overall acceleration metrics

commonly implemented in epidemiological research in the UK Biobank, and is mirrored when daily step count is applied to the cohort. Peak step cadence additionally demonstrates a reduction in step intensity with increasing age, and an increasing step intensity coinciding with increased self-reported walking pace.

The second and chief aim of this chapter was to evaluate physical activity in a cross-sectional analysis in knee and hip arthroplasty patients based on preoperative or postoperative status at the time of objective activity measurement. When applied to clinical cohorts of interest, differences in step count, peak cadence, and time spent in MVPA between patients before and after surgical intervention were observed, even when differences were not seen in overall acceleration or other machine learning activity classifications. In these cohorts, participants with end-stage knee arthritis presented as less functionally mobile than their peers. In participants following a knee arthroplasty, functional mobility more closely matched their peer group; however, lower time spent in activities of elevated intensity remained. Participants with end-stage hip arthritis also demonstrated lower functional mobility than their peers; however, participants in the postoperative hip arthroplasty cohort had very similar activity profiles to their matched peers.

This study demonstrates the potential utility of step-based activity measures to provide insights into the physical activity profiles of UK Biobank participants. As a more easily interpretable metric than overall acceleration, daily step count may provide additional insight into functional mobility in clinical populations, as observed in postoperative knee and hip arthroplasty cohorts. Future studies may find integration of step-based measures an effective way to assess functional

mobility, alongside other metrics including overall acceleration and machine learning based measures of activity.

Chapter 6: Implementation of Physical Activity Monitoring in Clinical Orthopaedic Practice: Study Setup, Preoperative Analysis, Participant Compliance and Feedback

6.1 Introduction

The primary aim of this chapter is to establish the methodology for implementing a physical activity monitoring study into clinical orthopaedic practice. This chapter will: 1) lay out methods used for implementing patient activity monitoring, 2) report preoperative baseline metrics for self-reported and objectively measured physical activity, 3) assess accelerometer wear time compliance and participant retention during pre and postoperative monitoring, and 4) describe patient and public involvement (PPI) regarding expectations of remote monitoring in the perioperative setting for future implementation into routine clinical practice.

6.2 Methods

6.2.1 Ethical Review

Ethical review of this study was conducted by the South West-Frenchay Research Ethics Committee under the reference number 19/SW/0151 with initial approval received from the NHS Health Research Authority on 15 October 2019. Additional substantial amendments to the protocol received approval on 16 January 2020 and 16 May 2020. Per ethical requirements, a protocol for this study is publicly registered at clinicaltrials.gov, listed under the identifier: NCT04240769 (https://clinicaltrials.gov/ct2/show/NCT04240769).

6.2.2 Recruitment and Enrolment

A flowchart of potential participant identification, recruitment, enrolment, and subsequent study interactions is presented in Figure 6-1. Potential research participants were identified from the preoperative assessment clinic (POAC) schedule lists at the Nuffield Orthopaedic Centre (NOC). Patients scheduled for an appointment at POAC for an impending total knee arthroplasty (TKA) or unicompartmental knee arthroplasty (UKA) with the participating consultant surgeons were posted an approved invitation letter and participant information sheet 1-2 weeks prior to their appointment. Potential participants were approached during their POAC appointment, assessed for eligibility, and asked if they were interested in further discussion of the study. Any patient interested in further discussing the study was given a description of the study, the opportunity to ask any questions, asked for their consent to participate, and enrolled at that time.

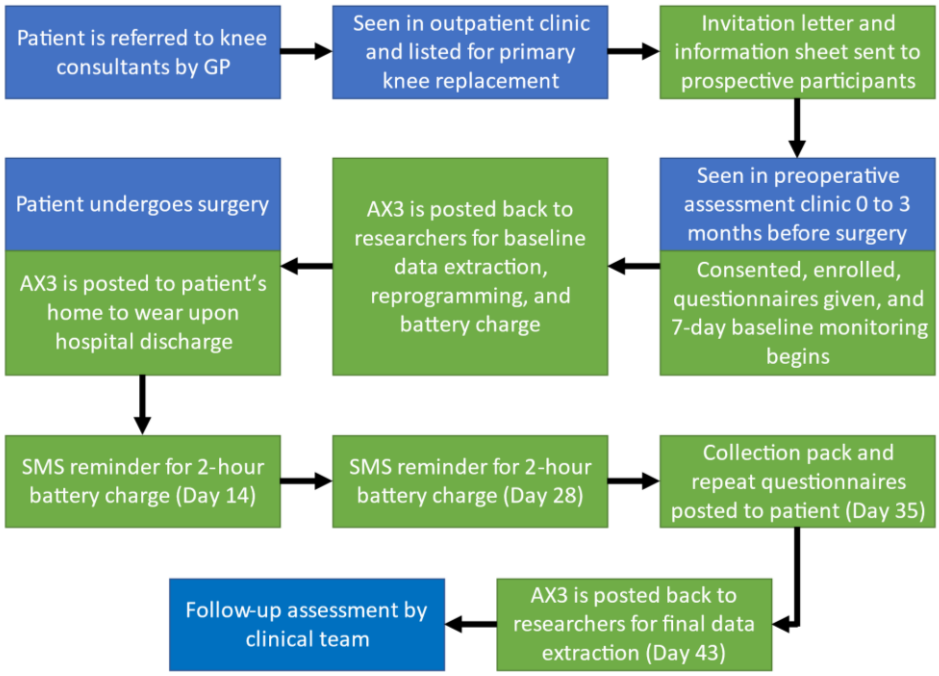


Figure 6-1: Participant recruitment and interaction flow chart. Clinical interactions are shown in blue, research interactions are shown in green.

6.2.3 Baseline Data Collection

6.2.3.1 Participant Demographics

Participant height, weight, body mass index (BMI), sex, ethnicity, hand dominance, and prior medical history were collected upon study enrolment. Contact information was additionally confirmed, including telephone and mobile numbers, and mailing address.

6.2.3.2 Preoperative Self-Reported Measures

Patient reported outcome measures (PROMs) and self-reported physical activity questionnaires are discussed in Section 1.3 and Section 1.6. At the time of enrolment, self-reported activity levels and PROMs were collected to determine preoperative baseline levels. As secondary outcome measures for comparison of knee function and health with objectively measured physical activity data, the Oxford Knee Score (OKS)[233], the EQ-5D index score, as calculated from the Euroqol 5 Dimension, 3 Level instrument[234], and the Euroqol Visual Analogue Score (VAS) were selected as measures for surgical outcome. These measures are routinely collected by the National Joint Registry, apply relatively low burden to the patient, and allow for easy interstudy comparisons[235].

The UCLA Activity Score[37] and the short form version of the International Physical Activity Questionnaire (IPAQ)[36] were selected as secondary self-reported measures of physical activity, as they have an extensive history of use in and out of orthopaedics allowing for robust interstudy comparisons, while adding minimal burden to participants[40,236].

6.2.3.3 Baseline Physical Activity Monitoring

Section 1.7 details the key primary considerations for implementation of wearables in a physical activity study, including body placement, device selection, sampling

rate, wear time, monitoring period, and data analysis. For physical activity monitoring, the research-grade Axivity AX3 (Axivity, Newcastle, UK) was selected as the device to be used in all data collection. At approximately £110 each, the AX3 is a relatively low-cost research grade device that provides access to raw triaxial acceleration at a range of sampling rates. An added benefit to using the AX3 is the ability to make direct comparisons to the UK Biobank physical activity cohort, which also used the Axivity AX3 to measure physical activity in 100,000 participants[53].

For longitudinal monitoring of patient physical activity data, a 7-day preoperative monitoring window was selected, fitting well within standard activity monitoring protocols[58]. Baseline physical activity monitoring was conducted using the AX3 accelerometer deployed in a silicone wristband. AX3 accelerometers were programmed using the Open Movement software (v1.0.0.42) (https://openmovement.dev) to begin recording at 17:00 on the day of the POAC appointment, continuing until 8:00 am eight days later. This recording schedule allows for the capture of seven full 24-hour days and eight full nights of sleep. Participants were instructed to wear the accelerometer on their dominant wrist for as many hours as possible at day or night over the course of the collection monitor. The accelerometer was set to record at a sampling rate of 25 Hz and a resolution of ±8 *g*.

6.2.3.4 Preoperative Activity Diary

Outside of monetary compensation, the keeping of a daily journal has been found as the best method for enhancing compliance in physical activity studies[237]. Therefore, patients were additionally asked to complete an activity diary during

their baseline activity monitoring period. The baseline activity diary included the following questions for each day:

1. At what times, if any, did you take an extended walk?
2. At what times, if any, did you ride a bicycle?
3. Was there any time in the past 24 hours that you removed your activity monitor?

6.2.3.5 Baseline Monitor Collection

Upon enrolment in the study, patients received a prepaid envelope addressed to the Botnar Research Centre for monitor return. An SMS message was programmed using the Oxford University Chorus SMS service and sent to the participant at the time of baseline monitoring completion, informing the participant of the completion of the monitoring period and serving as a reminder to post the accelerometer, wristband, and activity diary in the provided packaging as soon as possible. Upon receipt of the returned packages, activity data was downloaded from the device, memory cleared, and the device and wristband reserved for the participant until the time of surgery. Devices and wrist straps were cleaned with 70% isopropyl alcohol wipes during handling prior to distribution and upon collection.

6.2.4 Postoperative Data Collection

Patient surgery schedules were tracked using the Oxford University Hospitals Trust electronic medical records platform. Upon confirmation of the participant's completed surgery, a postoperative monitoring pack was distributed to the participant. This postoperative pack included an AX3 activity monitor and wristband, and 3-prong USB wall-charger, a USB micro charging cable, a postoperative activity diary, AX3 monitor charging instructions, and a reference

card for the 0-10 pain scale. Prior to 17 March 2020, this postoperative pack was delivered in-person to the patient at the NOC on the day of surgery. Following 17 March 2020, the postoperative packet was posted to each participant's home, to reduce patient contacts during the COVID-19 pandemic.

At the time of operation, clinical data was collected from each patient from their electronic health record, including confirmation of surgical technique (TKA or UKA), length of hospital stay, postoperative knee range of motion as measured by the physiotherapy team, and the type of walking aid used upon hospital discharge.

6.2.4.1 Longitudinal Postoperative Physical Activity Monitoring

A 42-day (6 week) postoperative monitoring window was selected to gather physical activity data over the full course of early postoperative recovery[5]. Subsequently, a reduced device sampling rate was necessary to enable adequate device data storage capacity for a full six weeks of postoperative data collection. Supported by the validation study undertaken in Chapter 3 which facilitates comparisons between reduced (25 Hz) and standard (100 Hz) AX3 sampling rates, the AX3 accelerometers were programmed to record at 25 Hz with a resolution of ±8 *g*. For all participants, the same serial number AX3 monitor was used for both baseline and postoperative activity monitoring. Prior to 17 March 2020, monitors set to begin recording at 20:00 the evening of surgery (20:00 the day following surgery after 17 March 2020) and to cease recording at 8:00 in the morning 42 days after surgery. Patients were required to charge the monitor twice during the postoperative monitoring period. SMS reminders were programed in the Oxford University Chorus SMS system to be sent at 18:00, 14 days and 28 days after surgery, asking participants to charge their device for at least 2-3 hours

in the evening on that day. Additional reminders were placed within the postoperative activity diary on postoperative days 14 and 28 (Figure A-1).

6.2.4.2 Postoperative Activity Diary

Each day, participants were asked to complete a daily diary including a repeat of the preoperative activity diary questions (Section 6.2.3.4), with the additional question:

1. Using the pain scale provided (0-10), how would you rate your average knee pain over the past 24 hours?

6.2.4.3 Postoperative Monitor and Questionnaire Collection

At five weeks following surgery, a postoperative collection pack was posted to each participant. This pack included the postoperative self-reported physical activity and PROMs questionnaires to be filled out by the participants, as well as a postage-paid envelope addressed to the research team at the Botnar Research Centre for the return of the monitor after the completion of the monitoring period. An SMS text message was sent to the participant at 09:00, 43 days following surgery, reminding the participant that their monitoring period was now complete, and to return the activity monitor and completed questionnaires within the provided packaging as soon as possible. Upon receipt, activity data was downloaded from the device, monitors and wristbands were cleaned with alcohol wipes, quarantined for 72 hours, recharged, and then released for reuse in subsequent activity monitoring of additional study participants.

6.2.4.4 Week 6 Postoperative Self-Reported Measures

A repeat of the IPAQ and UCLA self-reported activity questionnaires were filled out by study participants six weeks postoperatively to assess postoperative

activity levels at the time of activity monitoring completion. Repeats of the OKS, EQ-5D, and VAS questionnaires were also completed by study to assess postoperative patient function at the time of activity monitoring completion.

6.2.4.5 Week 6 Clinical Discharge Appointment

Most patients return to the NOC at 6 weeks following surgery to be examined by a physiotherapist and/or consultant surgeon for evaluation and discharge from hospital care. Study data collected from this clinical appointment included information regarding mobility and wound complications, recommendation for ongoing care, knee range of motion, and current use of walking aids. During national COVID-19 lockdowns, patients were consulted by the clinical team over the telephone. In these cases, the clinical team estimated knee range of motion based on conversation with the patient.

6.2.5 Data Processing and Analysis

6.2.5.1 Accelerometer Data Processing

As discussed in Section 1.7.6, it is possible to calculate a wide range of activity metrics from raw accelerometer data. Because the most suitable metric for differentiating activity and function in the perioperative knee arthroplasty clinical setting is unknown, I selected several objectively measured activity metrics for comparison in this study. Processing of accelerometer-based physical activity data was conducted using the Biobank Accelerometer Analysis Tool following standard methodology as described in Section 5.3.1. The primary physical activity metrics used for both preoperative and postoperative data analysis included daily step count and overall acceleration (equivalent to Euclidean norm minus one with negative values truncated to zero, (ENMO)). Secondary activity metrics in this analysis included active-to-sedentary transition probability (ASTP)[207,208], one

minute and thirty minute peak cadence[204], and time spent in light activity, moderate-to-vigorous activity (MVPA), sedentary activity, and sleep[81]. Because a 25 Hz sampling rate was used in both preoperative and postoperative data collection, acceleration and machine learning activity classification metrics were adjusted using the 25 Hz to 100 Hz transforms generated in Chapter 3 and detailed in Section 3.4.6.

Valid preoperative wear for the assessment of acceleration-based metrics was defined as at least three days (72 hours) of monitor wear time with data contribution from each of the 24 diurnal hours over the course of the 7-day preoperative monitoring period[53]. Valid preoperative wear for the assessment of step-based metrics was defined as at least three days of more than 14 hours of wear each day[204]. For all metrics, valid postoperative days were considered any day with at least 14 hours of accelerometer wear time.

6.2.5.2 Statistical Analysis

Data analysis for postoperative physical activity monitoring was conducted using R Studio (version 1.4.1103) and R 4.0.3. Physical activity metrics were reported throughout analysis in terms of median and presented alongside interquartile range. Normality of data distribution was assessed using the Shapiro-Wilk test. Correlation between objectively measured and self-reported physical activity and PROMs was calculated using the non-parametric Spearman Rank Order Correlation coefficient and presented in correlation plots, with participants missing PROMs data eliminated from correlation analysis.

Linear regression was used to assess the relationship between the primary outcome variables of daily step count and daily acceleration, with the independent variables including participant clinical characteristics, PROMs, and self-reported

activity questionnaires. Multiple chained equations imputation with the MICE package (v.3.13.0) was used to impute missing data from self-reported data prior to regression, with five imputation iterations pooled into final regression analysis. Covariates in the regression of physical activity included continuous measures of age, BMI, self-reported activity and preoperative PROMs, and categorical measures of sex, ethnicity, and use of walking aids. Additionally, a multivariable logistic regression model was created to assess the participant characteristics associated with a participant dropping out of the study between the preoperative and postoperative data collection periods. Dropout was defined as a participant electing not to continue following preoperative monitoring, not charging the device, or not returning a device during postoperative monitoring. Covariates selected for this model included sex, BMI, age, median daily preoperative step count, preoperative acceleration, preoperative walking aid usage, total preoperative wear time, and self-reported preoperative OKS, UCLA, IPAQ, VAS, and EQ-5D.

6.2.5.3 Sample Size Determination

Sample size estimation was calculated using an online tool at: https://www.imim.cat/ofertadeserveis/software-public/granmo/. Using participant acceleration from the UK Biobank as a basis, I estimated a minimum difference of 6 m*g* acceleration (SD 9.7 m*g*) between the least and most active participants six weeks after knee arthroplasty[53]. Assuming a post-enrolment dropout rate of 25%, alpha risk of 0.05, beta risk of 0.20, and a cohort split of 80/20 “good” and “poor” recoverers[238], a sample size of 175 subjects would be needed to identify a statistically significant difference of 6 m*g* difference between surgical cohorts.

During initial planning stages for this study, sample size was also targeted from the accessible population of primary knee arthroplasty patients at the NOC, where

876 knee arthroplasty surgeries were performed in 2018. An estimated available patient population of 400 patients per year from our four cooperating surgical consultants and a 50% enrolment rate from eligible participants led to an enrolment target of 200 participants per year of recruitment at a 50-50 TKA-UKA distribution.

6.2.6 Patient and Public Involvement Informing Future Clinical Direction

To understand patient acceptability and expectations of perioperative monitoring and digital health interventions in the orthopaedic patient population, a PPI meeting through the Open Arms program was organised with help from Natalie Ford in the Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences Department Outreach and Public Engagement office. On 28 June 2021, I presented a high-level, 20-minute talk on the current study and the utilisation of remote monitoring in postoperative clinical orthopaedic practice. This “Meet the Researcher” event was advertised to patients in the NOC with flyers and leaflets, to members of the Oxford Medical Sciences Division community, and to the broader community through email and social media posts. As a part of this presentation, the audience was asked to anonymously respond to poll questions encompassing their current use of personal activity trackers and their preferences for feedback from a hypothetical physical activity monitoring program in routine orthopaedic practice. The goal of this PPI session was to inform considerations for future implementation of this study into standard clinical practice.

6.3 Results

6.3.1 Recruitment and Enrolment

A flowchart of participant recruitment and enrolment is presented in Figure 6-2. From February 2020 to July 2021, invitation letters and information packets were distributed to 293 potential study participants, 249 potential participants approached, and 216 deemed eligible for enrolment. A total of 156 patients from the NOC were enrolled into this study, an enrolment rate of 83% of eligible, approached patients. After accounting for surgery cancellations, withdrawals, and sensor non-return, 90.4% of enrolled participants contributed valid baseline accelerometer wear. In total, I met face-to-face with 208 eligible patients who ultimately progressed to primary TKA or UKA, leading to an overall baseline data collection rate of 68% of the targeted patient population.

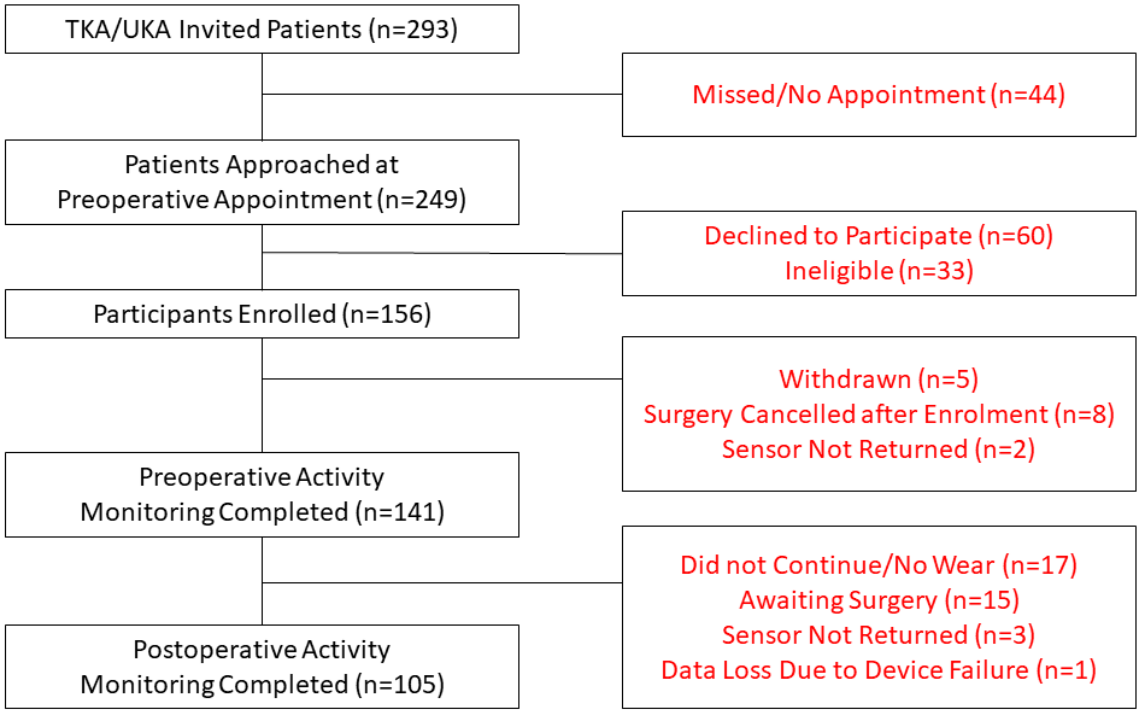


Figure 6-2: Recruitment and Data Collection Flow Chart

6.3.2 Participant Demographics

Demographics for participants completing baseline monitoring are presented in Table 6-1.

Table 6-1: Preoperative Participant Demographics

Characteristic	Female	Male	Full Preoperative Cohort
Overall, N	73	68	141
Age, Median [IQR]	71.6 [61.9-76.8]	69.1 [59.3-74.7]	70.4 [61.6–76.4]
Ethnicity, N (%)			
Asian	1 (1%)	1 (1%)	2 (1%)
Black	1 (1%)	0 (0%)	1 (1%)
Other/Not Stated	3 (4%)	14 (20%)	17 (12%)
White	68 (93%)	53 (78%)	121 (86%)
Body Mass Index, Median [IQR]	30.9 [26.9-34.9]	29.7 [26.9-33.1]	30.1 [26.9–33.9]
Handedness, N (%)			
Left	7 (10%)	13 (19%)	20 (14%)
Right	66 (90%)	55 (81%)	121 (86%)
Usual Walking Aid, N (%)			
Unaided	43 (59%)	56 (82%)	99 (70%)
Stick(s)	19 (26%)	8 (12%)	27 (19%)
Crutch(es)	6 (8%)	3 (4%)	9 (5%)
Wheeled Frame	5 (7%)	1 (2%)	6 (4%)
Prior Knee Arthroplasty, N (%)			
Contralateral Knee	22 (30%)	18 (26%)	40 (28%)
None	51 (70%)	50 (74%)	101 (72%)
Prior Hip Arthroplasty, N (%)			
One Hip	5 (7%)	6 (9%)	11 (8%)
Both Hips	3 (4%)	3 (4%)	6 (4%)
None	65 (89%)	59 (87%)	124 (88%)
Self-Identify as Disabled , N (%)			
Yes	29 (40%)	20 (29%)	43 (30%)
No	27 (37%)	35 (51%)	68 (48%)
No Response	17 (23%)	13 (19%)	30 (21%)
Consultant Surgeon, N (%)			
A	17 (23%)	14 (21%)	31 (22%)
B	13 (18%)	20 (29%)	33 (23%)
C	25 (34%)	13 (19%)	38 (27%)
D	18 (25%)	21 (31%)	39 (28%)

The participant population was evenly balanced by sex; however, ethnic background was primarily Caucasian. Most patients (70%) walked unaided at the time of enrolment, while 26% regularly utilised sticks or crutches, and only 4% relying on a wheeled frame for mobility. Participants were recruited from the practices of four consultant surgeons, with a balanced enrolment between the four consultants. A total of 28% of participants had previously undergone primary knee replacement of the contralateral knee prior to enrolment, while 12% had prior hip arthroplasty in one or both hips.

6.3.3 Accelerometer Wear Time Compliance

6.3.3.1 Baseline Accelerometer Wear

From the 141 eligible participants who returned physical activity monitors, wear compliance was high. Participants contributed a median 7.4 [IQR: 6.7-7.6] days of wear out of a possible 7.7 days. When assessing valid wear for acceleration-based activity metrics, a total of 129 participants (91.5%) contributed at least one hour of data for each of the 24 diurnal hours with a total of at least 72 hours of valid wear. For assessing daily step-based activity metrics, 132 participants (93.6%) contributed at least three days of greater than 14 hours of wear per day.

6.3.3.2 Postoperative Accelerometer Wear

As presented in Figure 6-2, from the 141 eligible participants who returned physical activity monitors, 17 participants (12%) elected not to continue with postoperative monitoring, 15 participants (11%) were still waiting for surgery at the time of this report, 3 devices (2%) were never returned or lost in the post, and one device malfunctioned during postoperative wear (1%). Subsequently, a total of 105 participants contributed a median 896 hours, or 37.3 days [IQR: 28.5-39.5] of accelerometer wear (Figure 6-3). Of these 105 participants who contributed

postoperative data, 95 contributed valid postoperative wear defined as at least 10 days of >14 hours of daily wear. One participant contributed a total of 348 hours or 14.5 days of wear, but no days of greater than 14 hours of wear per day. Subsequently, all postoperative statistical analysis was performed on 95 participants.

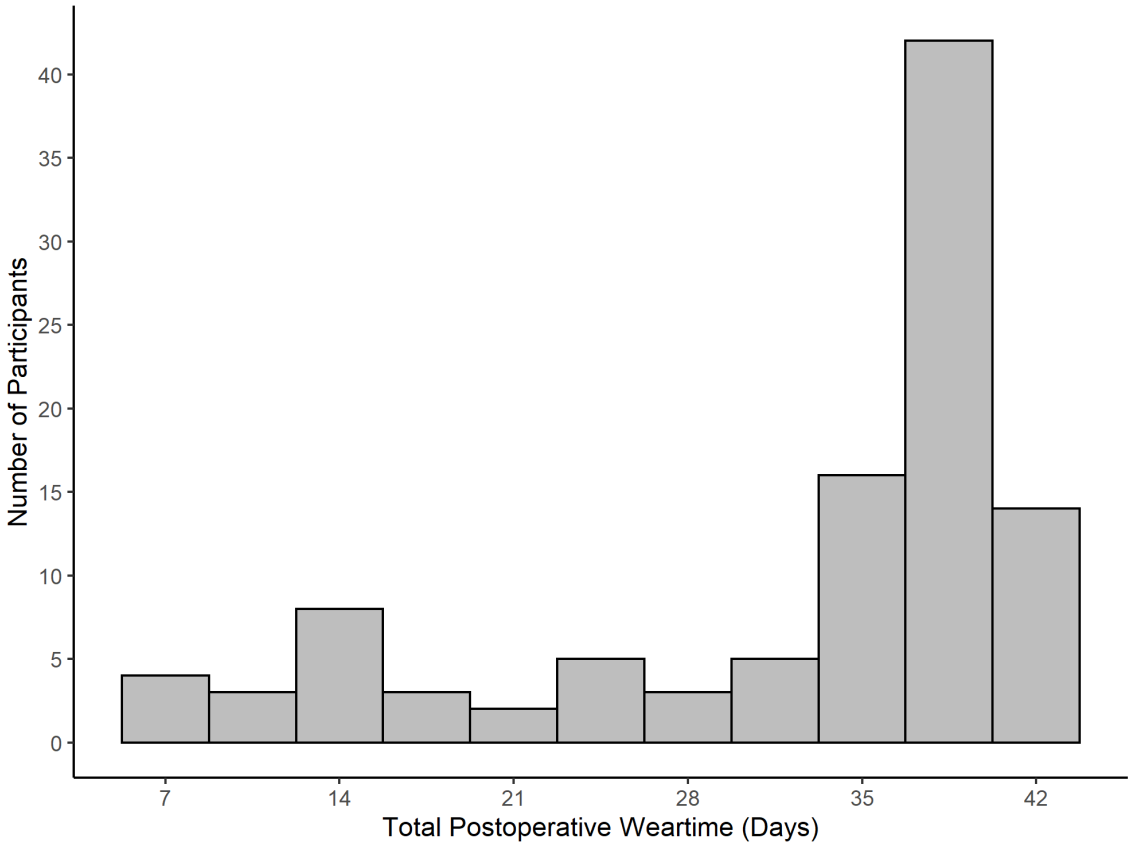


Figure 6-3: Histogram of Postoperative Wear Time in 105 Participants

6.3.3.3 Dropout Analysis

Results from the participant dropout model are presented in Table 6-2. From the logistic regression, preoperative wear time was the only significant predictor of study dropout, with an odds ratio of 0.62 (0.44-0.88, 95% CI). Thus, for every additional 24 hours of non-wear during preoperative monitoring was significantly

associated with a 38% greater likelihood of dropping out prior to postoperative monitoring (p=0.007).

Table 6-2: Dropout Multivariable Logistic Regression

Predictor	Odds Ratio	Confidence Interval	p
Age	1.017	(0.950, 1.089)	0.624
Body Mass Index	1.102	(0.973, 1.249)	0.125
EQ-5D	27.15	(0.297, 2,484.2)	0.150
IPAQ Weekly MET Minutes	1.000	(1.000, 1.000)	0.449
Oxford Knee Score	0.919	(0.813, 1.038)	0.171
Preoperative Accelerometer Wear Time (Days)	0.622	(0.442, 0.875)	**0.007**
Sex, *(Female)*			
Male	1.900	(0.536, 6.732)	0.317
UCLA Activity Score	1.094	(0.711, 1.685)	0.680
VAS	0.979	(0.944, 1.015)	0.251
Walking Aid, *(Unaided)*			
Aided	1.783	(0.384, 8.270)	0.457

Baseline conditions for categorical variables listed in italics

6.3.4 Baseline Physical Activity and PROMs

Self-reported physical activity and PROMs scores for this study population are presented in Table 6-3. The median UCLA activity score was 4 out of 10, a response indicating, “Regular participation in mild activities, such as walking, limited housework, and limited shopping”[236]. Using the IPAQ questionnaire, patients most commonly scored themselves as highly active (38%). Median PROMs scores were 22 out of a possible score of 48 on the OKS, 0.583 out of a possible 1.0 on the EQ-5D, and 70 out of a possible 100 on the VAS.

Table 6-3: Baseline Patient Reported Outcome Measures and Self-Reported Activity

Self-Reported Metric/PROM	Female	Male	Full Cohort
Oxford Knee Score (OKS) (n=132, F:69, M:63)	21 [12-28]	22 [19-29]	22 [15–28]
EQ-5D (n=120, F:62, M:58)	0.569 [0.294-0.659]	0.569 [0.357-0.659]	0.583 [0.339–0.659]
UCLA Activity Score (n=141, F:73, M:68)	4 [3-5]	5 [3-6]	4 [3–6]
VAS (n=118, F:59, M:59)	67 [50-80]	70 [50-85]	70 [50–80]
IPAQ (n=136, F:70, M:66)			
IPAQ Weekly MET-min	717 [240-1746]	1,543 [594-3083]	1,071 [391–2,600]
Low, N (%)	31 (44%)	17 (26%)	48 (35%)
Moderate, N (%)	21 (30%)	16 (24%)	37 (27%)
High, N (%)	18 (26%)	33 (50%)	51 (38%)

** All metrics reported as median [IQR] unless otherwise noted*
IPAQ = International Physical Activity Questionnaire

Table 6-4: Baseline Objectively Measured Physical Activity

Activity Metric	Female	Male	Full Cohort
Daily Steps[+]	6,348 [3,090–8,640]	7,619 [3,828–,380]	6,855 [3,201–9,003]
1 min Cadence (steps/min) [+]	102.6 [92.0–107.6]	105.1 [98.2–.9]	103.8 [93.8–109.1]
30 min Cadence (steps/min) [+]	74.5 [53.2–85.4]	81.5 [64.4–93.0]	78.1 [59.2–87.4]
Overall Acc (mg) [#]	21.1 [17.9–25.6]	21.8 [17.1–26.0]	21.5 [17.4–25.9]
Moderate-to-Vigorous Physical Activity (min/day) [#]	3 [1–9]	6 [1–20]	4 [1–12]
Light (min/day) [#]	347 [290–441]	338 [261–399]	340 [280–420]
Sedentary (min/day) [#]	476 [420–534]	478 [446–557]	477 [432–541]
Sleep (min/day) [#]	577 [518–627]	573 [527–619]	574 [524–623]
Active-to-Sedentary Transition Probability[#]	0.395 [0.301–0.455]	0.383 [0.315–0.472]	0.388 [0.305–0.466]

All metrics reported as median [IQR]
** Daily time classification may not add to 1,440 minutes due to rounding and 25 Hz to 100 Hz adjustment*
[+] Participants with 3+ days of 14-hour wear (n=132, F:68, M:64)
[#]Participants with 24 diurnal hours and >72 hours or wear (n=129, F:66, M:63)

Objectively measured preoperative physical activity for valid participant data across all computed metrics is presented in Table 6-4. Participants in this preoperative cohort took a median 6,855 steps per day during the monitoring period, with low levels of time spent in MVPA. A correlation plot between self-reported physical activity, PROMs, and objectively measured physical activity is presented in Figure 6-4. The IPAQ activity score was slightly more correlated to objectively measured activity when compared to the UCLA activity score in both acceleration (0.31 vs. 0.29) and step count (0.42 vs. 0.38). Preoperative PROMs showed weak correlation with objectively measured acceleration (EQ-5D: -0.04, OKS: 0.02, VAS: -0.04) or step count (EQ-5D: 0.04, OKS: 0.08, VAS: 0.09). However, PROMs did demonstrate moderate correlation with the IPAQ (OKS: 0.23, EQ-5D: 0.08, VAS: 0.26) and UCLA (OKS: 0.45, EQ-5D: 0.23, VAS: 0.29) self-reported physical activity measures in this preoperative cohort.

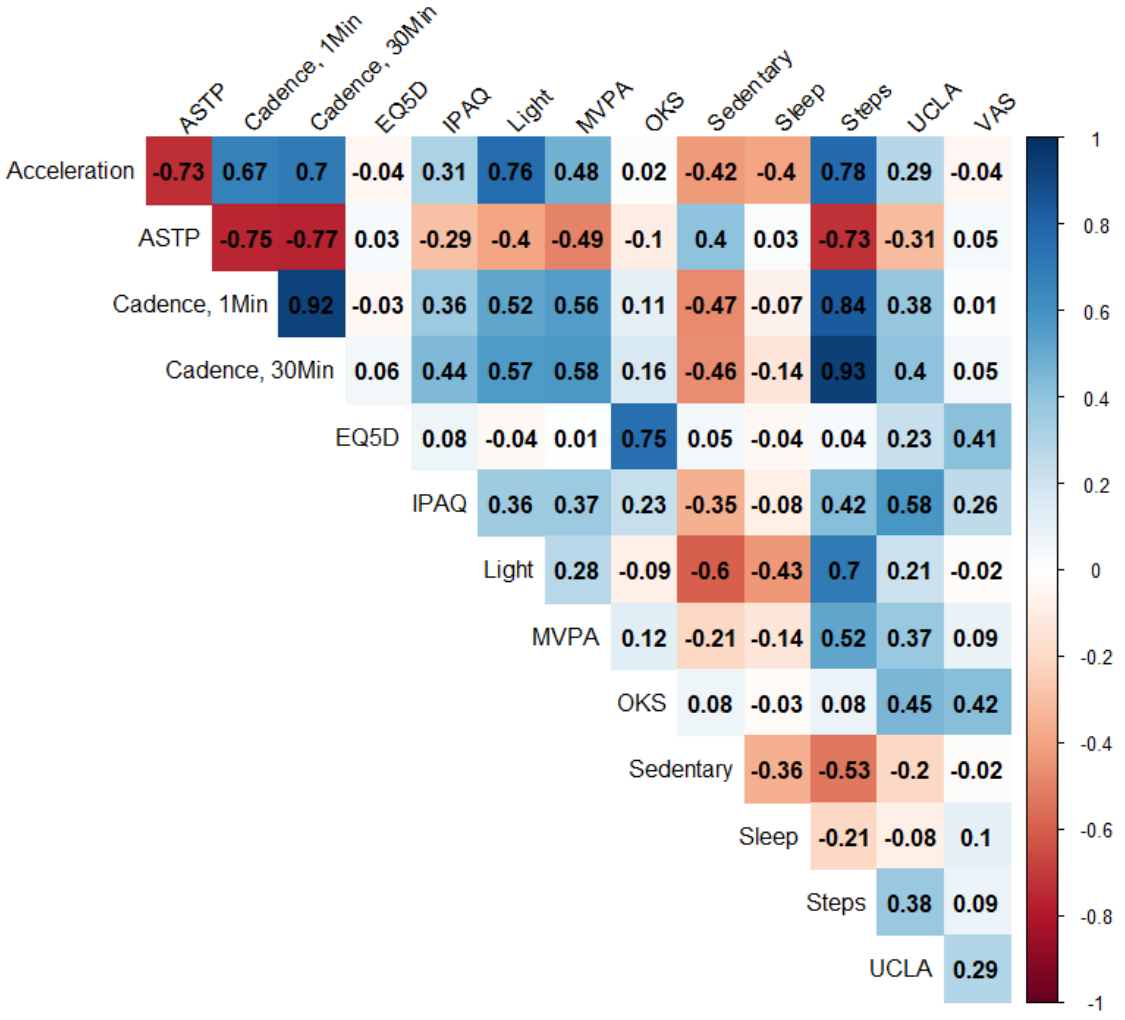


Figure 6-4: Preoperative correlation matrix using Spearman's Rank Test including objectively measured activity parameters, patient reported outcome measures, and self-reported activity questionnaires collected during the baseline monitoring period. *ASTP = Active-to-Sedentary Transition Probability, EQ5D = EQ-5D-3L Index Score, IPAQ = International Physical Activity Questionnaire, MVPA = Moderate-to-Vigorous Activity, OKS = Oxford Knee Score, VAS = EQ-5D Visual Analogue Score*

6.3.5 Regression Analysis of Preoperative Physical Activity

The results of a linear regression model of baseline preoperative step count over participant demographics and clinical characteristics are presented in Table 6-5, while the regression model of overall acceleration is presented in Table 6-6. In the step count model, increased age (p<0.001), increased BMI (p<0.001), and walking aid usage (p=0.008) were all statistically associated with a decrease in baseline step counts. Prior unilateral hip arthroplasty was associated with an increase of 2,943 daily steps (p=0.003). The UCLA activity score was not a statistically significant predictor of step count; however, each unit increase of IPAQ weekly MET-min was associated with 0.5 more preoperative steps (p=0.006).

Table 6-5: Preoperative Step Count Regression Table

Predictor	Estimate (Steps)	Confidence Interval	p
Age	-240	(-313, -166)	**<0.001**
Body Mass Index	-266	(-410, -122)	**<0.001**
Sex, *(Female)*			
Male	-1029	(-2,458, 400)	0.156
Ethnicity, *(White)*			
Asian	-1,616	(-7,223, 3,990)	0.569
Black	1,232	(-4,441, 6,904)	0.668
Other/Not Stated	-826	(-3,057, 1,405)	0.465
Prior Hip Arthroplasty, *(None)*			
Unilateral	2,943	(371, 5,514)	**0.003**
Bilateral	-350	(-3,521 , 2,820)	0.827
Prior Knee Arthroplasty, *(None)*			
Contralateral	1037	(-503 , 2,577)	0.185
Disabled, Self-Reported, *(No)*			
Yes	521	(-1,344, 2,385)	0.579
Walking Aid, *(Unaided)*			
Aided	-2,729	(-4,730 , -727)	**0.008**
EQ-5D	-5,323	(-10,509, 2,645)	0.103
IPAQ Weekly MET Minutes	0.5	(0.1, 0.8)	**0.006**
Oxford Knee Score	120	(-32, 272)	0.121
UCLA Activity Score	-17	(-505, 471)	0.945
VAS	20	(-24, 64)	0.373

Baseline conditions for categorical variables listed in italics

In the acceleration regression model (Table 6-6), increased age (p<0.001), increased BMI (p<0.001), walking aid usage (p=0.032), increased EQ-5D (p=0.029), and male sex (p=0.001) were all statistically associated with lower baseline acceleration. As in the step count model, the UCLA activity score was not a statistically significant predictor of activity; however, each unit increase of IPAQ weekly MET-min resulted in an increase of 0.0007 m*g* of daily baseline acceleration (p=0.004). Prior unilateral hip arthroplasty was additionally associated with a 4.7 m*g* increase in baseline acceleration (p=0.018).

Table 6-6: Preoperative Acceleration Regression Table

Predictor	Estimate (m*g*)	Confidence Interval	p
Age	-0.4	(-0.5,-0.3)	**<0.001**
Body Mass Index	-0.5	(-0.7, -0.3)	**<0.001**
Sex, *(Female)*			
Male	-3.6	(-5.7, -1.4)	**0.001**
Ethnicity, *(White)*			
Asian	1.6	(-6.8, 10.0)	0.704
Black	1.8	(-6.6, 10.2)	0.672
Other/Not Stated	1.5	(-1.8, 4.8)	0.388
Prior Hip Arthroplasty, *(None)*			
Unilateral	4.7	(0.8, 8.5)	**0.018**
Bilateral	3.5	(-1.2, 8.3)	0.141
Prior Knee Arthroplasty, *(None)*			
Contralateral	1.2	(-1.1, 3.5)	0.314
Disabled, Self-Reported, *(No)*			
Yes	1.3	(-1.5, 4.0)	0.354
Walking Aid, *(Unaided)*			
Aided	-3.6	(-6.6, -0.6)	**0.019**
EQ-5D	-10.1	(-19.2, -1.1)	**0.029**
IPAQ Weekly MET Minutes	0.0007	(0.0002, 0.0013)	**0.004**
Oxford Knee Score	0.2	(-0.0, 0.4)	0.081
UCLA Activity Score	-0.4	(-0.8, 0.7)	0.919
VAS	0.0	(-0.1, 0.1)	0.671

Baseline conditions for categorical variables listed in italics

6.3.6 Patient and Public Involvement Informing Future Clinical Deployment

Fourteen individuals participated in the Meet the Researcher PPI event associated with this study, including one study participant and 13 other NOC patients and/or members of the public. Answers from the group poll questions are presented in Table 6-7. Nine of fourteen participants (64%) indicated that they already track step count or physical activity on their own through an activity band, smart watch, or mobile device. Most participants indicated an expectation of step count updates and clinician feedback on physical activity if asked to wear a monitor postoperatively. Participants indicated a preference to using an app installed on their own mobile device for feedback and data transfer and indicated they would be happy to answer questionnaires in addition to activity monitoring during recovery. Meeting participants indicated the desire for regular activity updates in addition to clinician evaluation of physical activity during recovery. In the event of a normal recovery, 50% of participants indicated the desire for positive encouragement feedback, while 36% desired physiotherapy reminders or instructional videos. If an abnormal recovery was flagged during monitoring, all participants would expect a check-up call from a clinician, 64% additionally desired an alert from the device, and 57% would want to receive a questionnaire from the clinician regarding their recovery status.

Table 6-7: Questions and Answer Options for Patient and Public Involvement Event

Do you currently actively track your step count or physical activity via a wrist band, smart watch, or smart phone? (Select all that apply)
(3) Activity band (e.g. Fitbit)
(0) Pedometer
(2) Smart watch (e.g. Apple Watch)
(4) Smart phone (e.g. Android/iPhone)
(5) None
If you were loaned a wrist-based activity monitor as part of routine care after surgery, how would you feel?
(14) Happy to wear the monitor
(0) Reluctant to wear the monitor, but probably wear it anyway
(0) Reluctant to wear the monitor, and probably not wear it
What feedback would you expect from wearing an activity monitor after surgery? (Select all that apply)
(13) Daily/Weekly Activity Report (step count, etc.)
(12) Clinician's evaluation of recovery progress
(9) Physiotherapy reminders
(1) Other
In addition to wearing an activity monitor, what type of feedback system would work best for you?
(10) Interact with an app you install on your current smartphone
(4) You are provided with a phone/device for dedicated feedback/data transfer
How often would you be willing to answer questions regarding your recovery? (Select any that apply)
(6) 1 or 2 short questions every day
(6) 10-12 questions once a week
(0) No questions, wearing a monitor is enough
(5) 10-15 minute survey at beginning and/or end of 6 weeks
If an AI algorithm flagged your recovery progress as suboptimal or abnormal, what would you want to happen? (Select all that apply)
(14) Receive a check-up call from a clinician
(4) Receive pre-recorded videos/reminders of physiotherapy exercises
(9) Receive an alert from the device
(8) Receive a questionnaire to be reviewed by a clinician
(1) Other
If an AI algorithm flagged your recovery as normal, what would you want to happen? (Select all that apply)
(7) Receive an alert/encouragement from the device
(5) Receive pre-recorded videos/reminders of physiotherapy exercises
(4) No special notice given

Number of responses per answer given in parentheses

6.4 Discussion

The aim of this chapter was to establish methodology and initiate the study of perioperative physical activity in patients with end-stage knee osteoarthritis progressing to primary knee arthroplasty. Participant recruitment began in late February 2020 and, through two study pauses due to national lockdowns, culminated with the enrolment of 156 participants through July 2021. While recruitment efforts were severely hampered by the COVID-19 pandemic, 68% of eligible patients elected to participate and contributed valid baseline accelerometer data, significantly greater than the expected 50% enrolment rate. This level of successful recruitment reveals high acceptance of perioperative wearable-based monitoring in this clinical population. Wear compliance was high in baseline monitoring, with most participants contributing beyond the minimum 72 hours of accelerometer wear. While few physical activity studies extend beyond a 7-day monitoring protocol, an extended protocol is necessary for tracking longitudinal changes in physical activity in a postoperative clinical cohort. In postoperative monitoring, 14% of participants declined to continue or did not return a sensor; however, wear compliance in continuing patients was high with a median wear time of greater than 5 weeks.

To evaluate participant retention following preoperative monitoring, a logistic regression dropout analysis was performed based on those patients who declined to contribute wear for the extended 42-day monitoring period. Contrary to the belief that older patients may be technology-hesitant, age was not a significant predictor of patient retention. Only accelerometer wear time during the 7-day preoperative monitoring period was a significant predictor of dropout, indicating that once enrolled, patients quickly decided whether they were comfortable and happy with

the monitor overall. No added retention efforts or incentives were offered in this study as others have done[127]; nevertheless, participant retention and acceptability remained high in this study design.

Participants in the current study demonstrated physical activity within the range of other similar studies. In comparison to UK Biobank participants with end-stage knee osteoarthritis, as presented in Section 5.3.2, the current study cohort is more closely balanced in participant sex, is slightly older (70.4 vs 68.2 years), and has a slightly higher median BMI (30.1 vs 28.7). These factors, likely confounded by intermittent national lockdowns, resulted in lower baseline activity compared to the UK Biobank reference end-stage knee arthritis population. Nuffield Orthopaedic Centre patients presented with 854 fewer median daily steps and 2.9 m*g* lower median overall acceleration. While this cohort was less physically active than the UK Biobank comparison group, baseline activity was higher than those in other similarly sized studies of preoperative knee arthroplasty patients. Compared to the current study median of 6,855 daily steps, Losina et al.[127] reported 5,345 mean daily steps in 150 preoperative knee patients, Lützner et al.[239] reported 5,371 mean daily steps in 221 pre-TKA patients. In a recent study by Lyman et al.[240], published during the course of the current work, authors utilise smart phone accelerometers to track patient step count in 128 THA patients and 139 TKA patients continuously for 6 months postoperatively. While significant benefits are clear in the use of mobile phones in terms of data collection and participant communication, a major limitation to using smart phones accelerometry for data collection is the limited ability to control and monitor wear time for reliable activity classification or quantification of summary activity measures. While 24-hour wear time protocols are typical in wrist-worn devices, reliance on smart phones for step

counting may lead to significant underestimation of participant activity. This effect is clearly evident in the median 1861 steps per day in preoperative TKA participants reported by Lyman et al.[240], four to five thousand daily steps fewer than this and other comparable studies using body-worn sensors.

Self-reported physical activity questionnaires are commonly used in orthopaedic research; the UCLA activity score was first used to assess physical activity in orthopaedic populations in 1984[236]. In this study similar, correlation of the UCLA and IPAQ activity to preoperative objectively measured physical activity was observed, with slightly higher correlation in both measures to step count (UCLA, 0.38, IPAQ: 0.42) compared to acceleration (UCLA: 0.29, IPAQ: 0.31). This level of association is in range with comparisons made by other studies[241]. In this study, preoperative EQ-5D, VAS and OKS scores demonstrated no correlation with preoperative acceleration or daily step count. Interestingly, the OKS demonstrated a correlation coefficient of 0.45 with self-reported UCLA scores, but coefficients of 0.02 with acceleration and 0.08 with daily step count. This finding supports the assertion that PROMs and objective activity metrics target two distinct aspects of the patient experience, a breadth of insight not captured with self-reported physical activity measures. Thus, PROMs and ground truth activity measurements should be seen as complementary approaches for the assessment of surgical intervention outcomes, patient mobility, and quality of life.

Linear regression was used to assess the association of demographic and clinical variables on the summary-level physical activity metrics of daily step count and acceleration. In agreement with population based epidemiological studies[53,242], age and BMI were both highly associated with physical activity levels in this population, while male sex was associated with acceleration, but not daily step

count. As noted in Section 5.4.1.1, the use of walking aids may result in underreporting of step counts and wrist-based activity, most problematically during the use of wheeled support frames. Walking aid use was associated with lower physical activity in both models; however, the direction of causality is unclear and is likely reflective of both lower activity in gait-aided individuals and some underreporting of activity in those individuals requiring the greatest degree of walking support. In both models, self-reported IPAQ activity was associated with overall physical activity, though no association was observed with UCLA activity score.

According to a review by Bergmann et al.[243], patient desires for body-worn medical sensors most prominently include limited disruption to normal daily behaviour, simple operation and maintenance, and retained precedence of clinician-based care and decision-making. In a series of studies, Papi et al. described patient[244] and clinician[245] preferences for wearable technologies in patients with knee osteoarthritis. According to these studies, patients were open to monitoring during recovery in order to encourage compliance to therapy protocols, increase virtual supervision, and provide recovery information to the clinical care team. From the clinician perspective, remote monitoring was seen as an opportunity to evaluate and inform treatment and decision-making, while helping to manage patient care more efficiently. In line with these reviews, broad acceptance of wearable-based monitoring and the desire for feedback from both the device and the clinician were also reflected in our PPI group. More than half of the participants in our PPI group already tracked personal physical activity to some degree, and all considered themselves willing to wear a monitor as part of routine clinical care. When presented with the scenario of having an abnormal recovery, all PPI participants

desired contact with the clinical care team, via call, prerecorded message, or query. In the event of a normal recovery, only 4 of 14 desired no additional feedback via encouragement or alert.

6.4.1 Strengths and Limitations

A major strength of this study was the use of a protocol incorporating research grade, wrist-worn devices inducing high accelerometer wear time with direct comparability to the UK Biobank physical activity cohort. The study design proved to be minimally intrusive to both daily participant behaviour and to the clinical care pathway. Amid the COVID-19 pandemic, this study protocol was highly flexible in terms of patient interaction and was able to be quickly restarted with minimal face-to-face patient interaction.

A primary limitation of this study was the potential confounding of physical activity due to government-enforced limitation of social interaction and societal mobility during national lockdowns.[246] Additionally, walking aid bias may play a role in measurement error when recording physical activity at the wrist. In this cohort, 30% of participants relied on some type of walking aid; however, only 4% of participants used a wheeled frame, while 26% of participants used crutch or stick walking support, both of which are less prone to underreporting bias[180]. Machine learning physical activity classification and walking detection was performed using a random forest model trained on non-clinical populations. Osteoarthritis of the knee significantly affects gait, and subsequently, bias in activity classification has been introduced by using models not trained on the specific population of interest. Future work is needed to validate the accuracy of machine learning classification models in osteoarthritic patients.

Further, inherent risk of bias is associated with missing self-reported data. The volume of missing self-reported data in this study was low and assessed likely to be missing-at-random. While no formal sensitivity analysis of the imputation procedure is presented, a preliminary complete case analysis was performed, resulting in no difference in the overall study findings.

6.5 Conclusions

The results presented in this chapter demonstrate the ability to implement wearable-based physical activity monitoring in a perioperative clinical environment using methods shown widely acceptable to the patient population. High study enrolment, wear time compliance, low dropout, and enthusiastic desire for clinical feedback during the early postoperative period indicate a potential avenue for further development and integration within the standard orthopaedic care pathway. These results demonstrated limited correlation between self-reported activity, PROMs, and true patient mobility and activity, suggesting the inclusion of physical activity monitoring alongside standard PROMs is necessary to provide further insight into patient function.

Chapter 7: Assessing Knee Arthroplasty Postoperative Recovery Through Physical Activity Monitoring

7.1 Introduction

To date, most physical activity studies of knee arthroplasty populations consist of short assessments of activity beyond six months after surgical intervention[62]. The previous chapters of this thesis have established the methodology for extended longitudinal physical activity monitoring of an orthopaedic patient population in a manner allowing for valid and direct comparison to other epidemiological physical activity studies, using standard acceleration-based metrics alongside more interpretable and mobility-focused step count metrics. The goal of this chapter is to assess the use of wearable physical activity monitors in the clinical orthopaedic setting. The two primary opportunities for clinical integration of wearable monitors are: 1) the use of ambulatory physical activity monitoring for assessment of individual patient perioperative recovery, and 2) the use of objectively measured physical activity to compare recovery between unicompartmental knee arthroplasty (UKA) and total knee arthroplasty (TKA).

7.2 Methods

Methodology for study setup and data collection for both preoperative baseline and postoperative data is presented in Section 6.2. The following section addresses analysis methods of postoperative data. Additionally, an exploratory cluster analysis is presented in Appendix A.2.

7.2.1 Missing Data

As presented in Section 6.2.5.1, a valid postoperative wear day was defined as a day during which the participant wore the activity monitor for at least 14 hours.

Subsequently a valid postoperative time series was considered a patient who contributed at least 10 postoperative days of valid wear. In patients contributing a valid postoperative timeseries, missing data for daily activity metrics were imputed. Prior to full data analysis, a performance analysis for imputation of daily postoperative physical activity metrics was performed on a subset of study participants with complete self-reported data and postoperative wear time greater than 38 days (912 hours). In these participants with exceptional wear time, designated segments of the true daily acceleration and step counts were systematically removed, replaced with imputed values, with a subsequent comparison made between the true and imputed values with root mean square error (RMSE) reported. Imputation methods were compared in how well they imputed missing data with data from only Week 1 was removed, with only Week 3 data was removed, with only Week 6 data was removed, and finally with every odd numbered postoperative day removed. Four imputation methods were considered for imputing activity metrics of postoperative days: linear interpolation, Kalman filter interpolation, single imputation from chained equations and random forest imputation. Linear and Kalman filter imputation was performed at the individual participant level using the imputeTS R-package (v.3.2). Random forest imputation was performed with the missForest R-package (v1.4) using 100 trees a maximum of 10 iterations, with activity data predicted on all preoperative clinical data, demographic variables, postoperative day (day of surgery designated as Day 0), and preoperative summary activity measures. Single chained equations imputation was performed using the MICE R-package (v.3.13.0), predicting on the same variables as the random forest imputation, limited to those with a minimum spearman correlation of 0.2 with the primary outcome measure. Following

performance analysis, imputation of objectively measured physical activity from missing and non-valid days was performed using the selected imputation method. Independently, any missing patient reported outcome measures (PROMs) and self-reported data was imputed using random forest imputation following time series activity imputation, with the same set of clinical and preoperative variables, using the missForest package initiating 100 trees and a maximum of 10 iterations.

7.2.2 Statistical Analysis

7.2.2.1 Time Series Analysis

For week-to-week comparisons of physical activity, the Shapiro-Wilk test was used to assess normal distribution of data, and paired Wilcoxon Rank-Sum tests were used to assess weekly change in activity levels, with a p-value of 0.05 or less considered statistically significant. Median and interquartile range were calculated weekly across all participants in each physical activity metric. Timeseries quantile plots using additive quantile regression smoothing was used in the ggplot R-package (v.3.3.5) to plot collective recovery trends over the 42-day postoperative measurement period. These curves were additionally plotted to compare subgroups of participants based on early perioperative activity. Decile groups for comparison were derived by taking the median activity of each metric during the first postoperative week at 10% increments, and then plotting quantile curves to compare trends of six-week recovery relative to activity in the first postoperative week.

7.2.2.2 Correlation of Self-Reported Data with Objectively Measured Acceleration

Correlation between objectively measured and self-reported physical activity and PROMs was calculated using the non-parametric Spearman Rank Correlation coefficient. Correlation was calculated between median Week 6 activity, median

pain, and postoperative self-reported activity and PROMs. Correlation was also calculated between Week 1 to Week 6 physical activity, and Week 2 to Week 6 physical activity.

7.2.2.3 Regression Analysis

To model the association of step count and acceleration with postoperative day, demographic and clinical factors, mixed effects models were developed using the lmerTest R package (3.1-4). Because the recovery trajectory in observed to be non-linear through smoothed scatter plotting, day number was included as a linear and second-order polynomial predictor. In a first iteration, a mixed effect model was generated with participant as the fixed effect and day number, clinical and demographic factors as random effects. In a second iteration, the model was repeated with preoperative physical activity in terms of preoperative step count (or acceleration) as an additional predictor.

7.2.2.4 Comparison of Total and Unicompartmental Knee Arthroplasty Patients

A second analysis was conducted using objectively measured physical activity to compare unicompartmental knee arthroplasty and total knee arthroplasty surgical interventions. Baseline and weekly physical activity metrics were collected for each of the two surgical groups, with Wilcoxon Sum Rank tests performed to compare activity metrics at baseline and then again at weekly intervals during the first six postoperative weeks. Week-to-week within-cohort change was evaluated using paired Wilcoxon Sum Rank tests. Box plots of baseline and median weekly physical activity metrics for each surgical group were also created.

7.3 Results

Demographics and primary clinical data from all participants returning a postoperative activity monitor are presented in Table 7-1. Characteristic of the patient population and care protocols at the Nuffield Orthopaedic Centre (NOC), UKA (57, 54%) was performed more frequently than TKA (48, 46%). Further, nearly half (50, 48%) of all participants underwent a day case protocol with no overnight stay following surgery.

Table 7-1: Postoperative Cohort Clinical Characteristics

Characteristic	Completed TKA Cohort	Complete UKA Cohort	Completed Postoperative Cohort
Overall, N (%)	48	57	105 (100%)
Age, Median [IQR]	71.0 [61.8-75.3]	69.8 [58.9-74.6]	70.4 [60.9–74.8]
Body Mass Index, Median [IQR]	30.2 [26.8-34.0]	29.6 [26.7-33-6]	30.0 [26.7–33.8]
Sex, N (%)			
Female	24 (50%)	30 (53%)	54 (51%)
Male	24 (50%)	27 (47%)	51 (49%)
Ethnicity, N (%)			
Asian	0 (0%)	1 (2%)	1 (1%)
Black	0 (0%)	1 (2%)	1 (1%)
Other/Not Stated	7 (15%)	6 (11%)	13 (12%)
White	41 (85%)	49 (86%)	90 (86%)
Hospital Length of Stay, N (%)			
Outpatient	11 (23%)	39 (68%)	50 (48%)
1 Night	17 (35%)	12 (21%)	29 (28%)
2 Nights	4 (8%)	4 (7%)	8 (8%)
3 Nights	5 (10%)	2 (4%)	7 (7%)
4+ Nights	11 (23%)	0 (0%)	11 (11%)

Scoring from Week 6 self-reported questionnaires are presented in Table 7-2. Oxford Knee Score (OKS) and EQ-5D scores were collected from 92% of participants, whereas 82% of participants completed UCLA activity scores and 81% completed valid IPAQ-SF scores. Compared to a baseline median OKS of

23 OKS, an EQ-5D of 0.596, and a VAS of 70, these participants reported an increase of 6, 0.094, and 10 points for each of the PROMs scores, respectively, at six postoperative weeks. The median preoperative and postoperative UCLA activity scores remained stationary at 4, while the median self-reported weekly IPAQ fell from 1266 MET-min at baseline to 891 MET-min.

Table 7-2: Postoperative (Week 6) Self-Reported Activity and Patient Reported Outcome Measures

Self-Reported Metric	TKA Cohort	UKA Cohort	Baseline Cohort
Oxford Knee Score (OKS) (n=97, TKA: 46, UKA: 51)	30 [24-34]	29 [26-36]	29 [25–34]
EQ-5D (n=97, TKA: 46, UKA: 51)	0.675 [0.567-0.761]	0.690 [0.596-0.761]	0.690 [0.596–0.761]
UCLA Activity Score (n=86, TKA: 39, UKA: 47)	4 [3-5]	4 [4-6]	4 [3–6]
VAS (n=92, TKA: 44, UKA: 48)	80 [65-85]	80 [66-88]	80 [65–85]
International Physical Activity Questionnaire (IPAQ) (n=85, TKA: 38, UKA: 47)			
IPAQ Weekly MET-min	743 [344-1782]	924 [432-1908]	891 [396–1782]
Low, N (%)	15 (39%)	15 (32%)	30 (35%)
Moderate, N (%)	11 (29%)	19 (40%)	30 (35%)
High, N (%)	12 (32%)	13 (28%)	25 (29%)

All metrics reported as median [IQR] unless otherwise noted

7.3.1.1 Postoperative Pain Log

Of the 105 participants contributing postoperative data, 90 completed daily pain logs in addition to accelerometer wear. A quantile regression plot including all participant reported data points is presented in Figure 7-1, indicating a gradual decrease in joint-specific pain during the postoperative period.

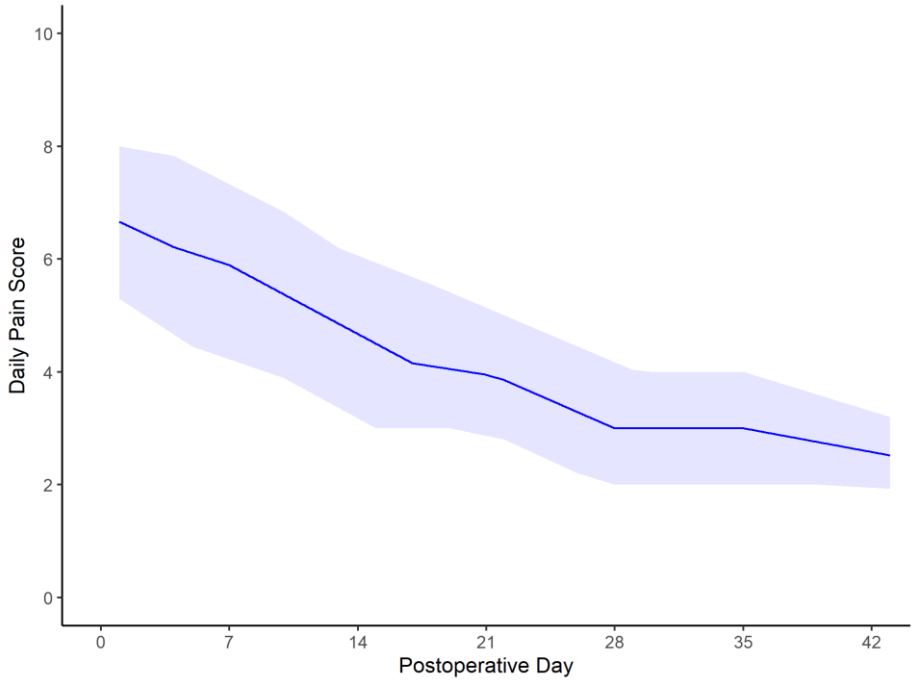


Figure 7-1: Quantile regression of daily postoperative [0-10] pain scores (n=90), with solid line representing the median and shaded region representing the interquartile range.

7.3.2 Physical Activity Imputation Performance Analysis

Prior to full data analysis, a performance analysis was conducted to assess the appropriate method for dealing with missing timeseries data in daily objectively measured physical activity data. Of the 95 participants completing a minimum 10 days of >14 hours of monitoring, a total of 29 participants were selected for inclusion in the performance analysis based on having at least 38 days of valid postoperative wear and no missing self-reported data. Root mean square error

(RMSE) values for imputation of missing step counts and acceleration are presented in Table 7-3. In both step count and acceleration measures, within-participant linear and Kalman imputation methods performed better than chained equation or random forest imputation when Week 1 and Week 3 data were removed. Random forest imputation resulted in the lowest RMSE when data from alternating days were removed. Based on a judgement call with preference for within-participant imputation and best performance during Week 1, when missing data was most common overall, I selected linear interpolation for subsequent imputation of missing objectively measured physical activity in subsequent data analysis.

Table 7-3: Physical Activity Imputation Methods Performance Analysis

	Linear	Kalman	Chained Equations	Random Forest
Step Count (Steps)				
Week 1 Removed	**1183**	1338	6583	1829
Week 3 Removed	**1392**	1401	5794	1482
Week 6 Removed	**2517**	2300	5417	2294
Odd Days Removed	**1166**	1041	2228	571
Acceleration (m*g*)				
Week 1 Removed	**2.93**	3.10	10.30	3.65
Week 3 Removed	**2.50**	2.32	6.87	2.51
Week 6 Removed	**3.49**	3.39	7.63	3.41
Odd Days Removed	**1.57**	1.46	3.57	1.16

Root mean square error reported in all instances for error between true and interpolated activity values

7.3.3 Wearable Physical Activity Measurement for Postoperative Monitoring: Assessment of the Trajectory in Recovery Following Knee Arthroplasty

7.3.3.1 Overview of Week-to-Week Physical Activity Measures

Median and interquartile range (IQR) of physical activity measures for all contributing participants are presented in Table 7-4, and p-values from week-to-week comparisons are presented in Table 7-5. Significantly lower physical activity and mobility was demonstrated in the first postoperative week when compared to baseline across all metrics. A statistically significant increase in week-to-week activity was observed across all measures and all weeks with the single exception of time spent in MVPA from Week 5 to Week 6. While activity steadily increased on a weekly basis, all metrics demonstrated significantly lower activity or mobility at postoperative Week 6 when compared to baseline preoperative levels ($p<0.05$). A key difference in overall summary metrics is demonstrated when relating postoperative step count and acceleration to baseline values. During the sixth postoperative week, participants had recovered to 86.6% of the median preoperative acceleration, yet only recovered 55.3% of the median preoperative daily step count.

Table 7-4: Physical Activity Metrics in the Full Valid Postoperative Cohort (n=95)

Activity Metric	Preop Baseline	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6+
Daily Steps	7,198 [4,205-9,301]	653 [258-1,141]	1,482 [681-3,064]	2,338 [1,166-4,370]	3,082 [1,453-5,192]	3,605 [1,670-6,017]	3,983 [2,026-6,390]
Overall Acceleration (mg)	22.5 [18.5-26.3]	14.6 [12.3-18.2]	16.9 [13.9-20.1]	17.3 [14.5-21.4]	18.2 [15.0-21.5]	19.8 [15.8-22.9]	19.5 [16.2-24.0]
MVPA (min/day)	4 [1-13]	0 [0-0]	0 [0-3]	0 [0-5]	1 [0-6]	3 [0-8]	3 [0-10]
Light (min/day)	347 [312-450]	148 [93-214]	208 [133-307]	256 [170-322]	276 [209-350]	289 [223-366]	303 [236-380]
Sedentary (min/day)	463 [421-518]	663 [562-761]	669 [583-747]	602 [519-679]	557 [504-639]	543 [482-603]	524 [464-595]
Sleep (min/day)	575 [527-616]	582 [497-636]	542 [474-580]	553 [505-610]	569 [522-607]	559 [523-608]	554 [525-600]
ASTP	0.372 [0.305-0.446]	0.621 [0.533-0.751]	0.576 [0.470-0.670]	0.489 [0.411-0.631]	0.485 [0.397-0.621]	0.452 [0.371-0.563]	0.435 [0.363-0.561]
1 min Cadence (steps/min)	104.9 [97.5-110.1]	69.0 [48.3-82.2]	78.6 [61.1-89.1]	88.1 [73.1-97.7]	91.0 [77.6-100.9]	94.4 [79.6-101.4]	96.1 [83.7-103.0]
30 min Cadence (steps/min)	81.2 [65.0-87.4]	23.9 [10.2-33.7]	35.9 [22.9-52.0]	45.1 [33.2-64.6]	55.5 [37.7-70.6]	58.5 [40.7-74.2]	62.4 [45.8-80.0]

All values presented as median [interquartile range]

Table 7-5: Table of Wilcoxon Rank Test P-Values of Week-to-Week Change in Activity Metrics Across the Full Valid Postoperative Cohort (n=95)

Activity Metric	Baseline to Week 1	Week 1 to Week 2	Week 2 to Week 3	Week 3 to Week 4	Week 4 to Week 5	Week 5 to Week 6	Week 1 to Week 6	Baseline to Week 6
Daily Steps	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**
Overall Acceleration (mg)	**<0.001**	**<0.001**	**<0.001**	**0.010**	**<0.001**	**0.018**	**<0.001**	**<0.001**
MVPA (min/day)	**<0.001**	**<0.001**	**0.008**	**0.003**	**<0.001**	0.689	**<0.001**	**<0.001**
Light (min/day)	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.025**	**<0.001**	**<0.001**
Sedentary (min/day)	**<0.001**	0.297	**<0.001**	**<0.001**	**<0.001**	0.069	**<0.001**	**<0.001**
Sleep (min/day)	0.631	**<0.001**	**<0.001**	0.075	0.898	0.860	0.087	**0.014**
ASTP	**<0.001**	**<0.001**	**<0.001**	**0.022**	**<0.001**	**0.004**	**<0.001**	**<0.001**
1 min Cadence (steps/min)	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.004**	**0.004**	**<0.001**	**<0.001**
30 min Cadence (steps/min)	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**

7.3.3.2 Trajectory of Recovery Using Step Count and Acceleration Summary Metrics

Postoperative recovery in terms of daily step count is presented in Figure 7-2 as a quantile regression plot representing the median daily step count of the study population. A clear, non-linear increase in step count is observed with a slight levelling inflection seen near Postoperative Day 21. Postoperative data is additionally presented in terms of deciles of participants based on Week 1 step count. As highlighted in the figure, a wide range of step counts were observed throughout the preoperative recovery period, with differentiation of groups evident early in the postoperative period. The top 10% most active participants at Week 1 ended with a median [IQR] 7,194 [5,004-8,988] steps per day during Week 6, compared to a median 492 [216-970] steps per day in the least active 10% of Week 1 participants.

Similar to daily postoperative step count, a non-linear, yet less dramatic increase in daily postoperative acceleration is presented in Figure 7-3 with Week 6 acceleration levels significantly below median baseline and median UK Biobank acceleration levels. The top 10% most active participants at Week 1 ended with a median 24.1 [21.2-31.6] m*g* acceleration during Week 6, compared to a median 15.5 [12.2-17.7] m*g* acceleration in the least active 10% of Week 1 participants.

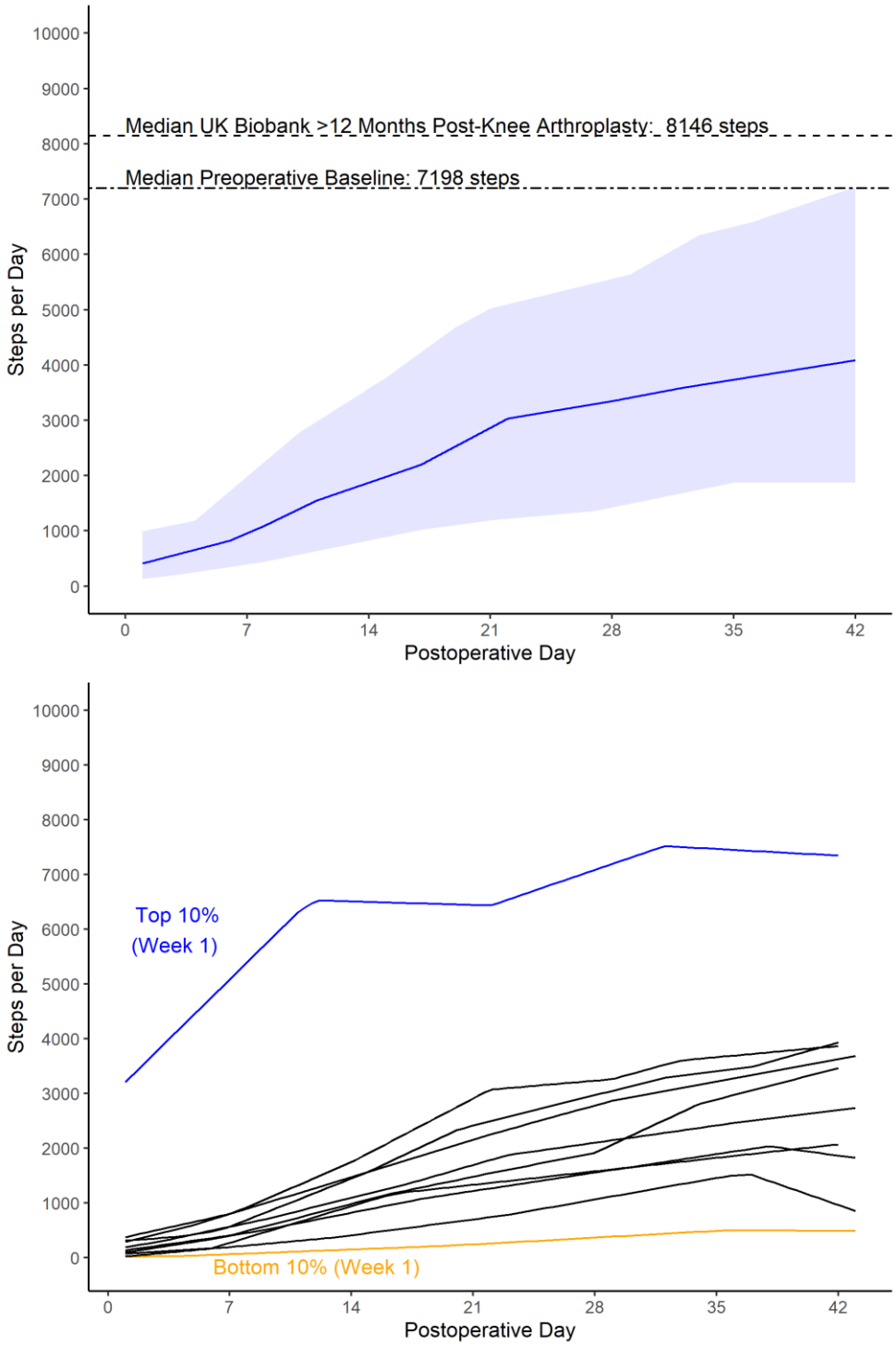


Figure 7-2: (Top) Quantile regression plot of increasing daily step count during the six weeks following knee arthroplasty, with solid line representing the median and shaded region representing the interquartile range. (Bottom) Median quantile regression plots of participant deciles based on median step count during the first postoperative week. The most active and least active deciles are highlighted, while deciles making up the middle 80% of participants are shown in black.

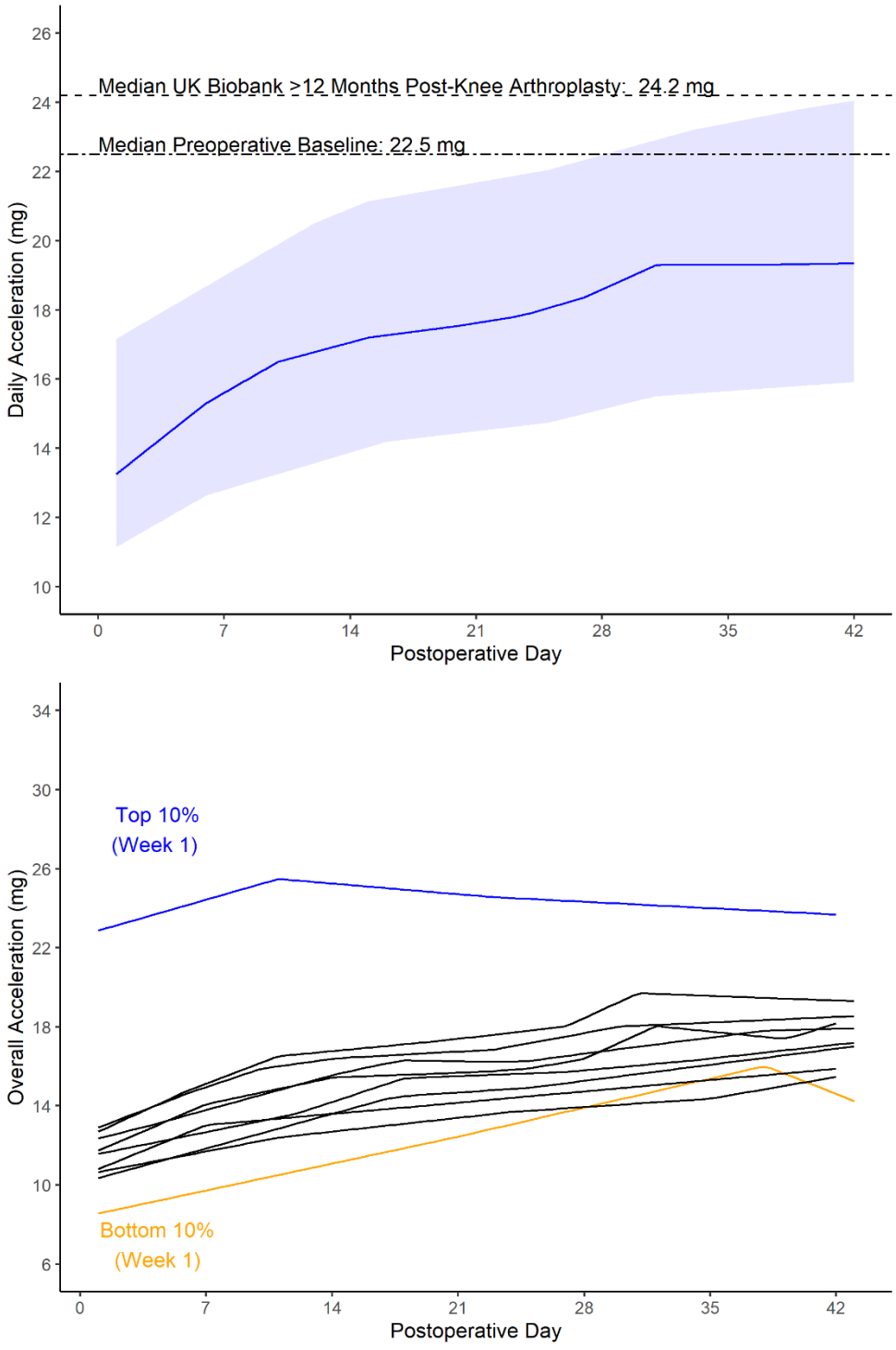


Figure 7-3: (Top) Quantile regression plot of increasing daily acceleration during the six weeks following knee arthroplasty, with solid line representing the median and shaded region representing the interquartile range. (Bottom) Median quantile regression plots of participant deciles based on median acceleration during the first postoperative week. The most active and least active deciles are highlighted, while deciles making up the middle 80% of participants are shown in black.

7.3.3.3 Trajectory of Recovery Using Peak Cadence and ASTP Metrics

As a measure of ease of mobility, postoperative recovery of peak 1-minute walking cadence is presented in Figure 7-4. Similar to daily step count, peak 1-minute cadence increases consistently throughout the monitoring period, with a wide range in peak cadence across study participants. The slowest walking decile of participants had a median Week 6 peak cadence of 64 [41-83] steps per minute, while the fastest decile walked with a median peak cadence of 104 [94-108] steps per minute during Week 6.

Postoperative change in active bout fragmentation, in terms of ASTP, is presented in Figure 7-5. ASTP decreases steadily during the postoperative period, correlating with longer, less interrupted active bouts as participants progressed in postoperative recovery. The least fragmented decile had a median ASTP of 0.314 [0.238-0.365] at postoperative Week 6, compared to 0.603 [0.481-0.735] in the most fragmented decile.

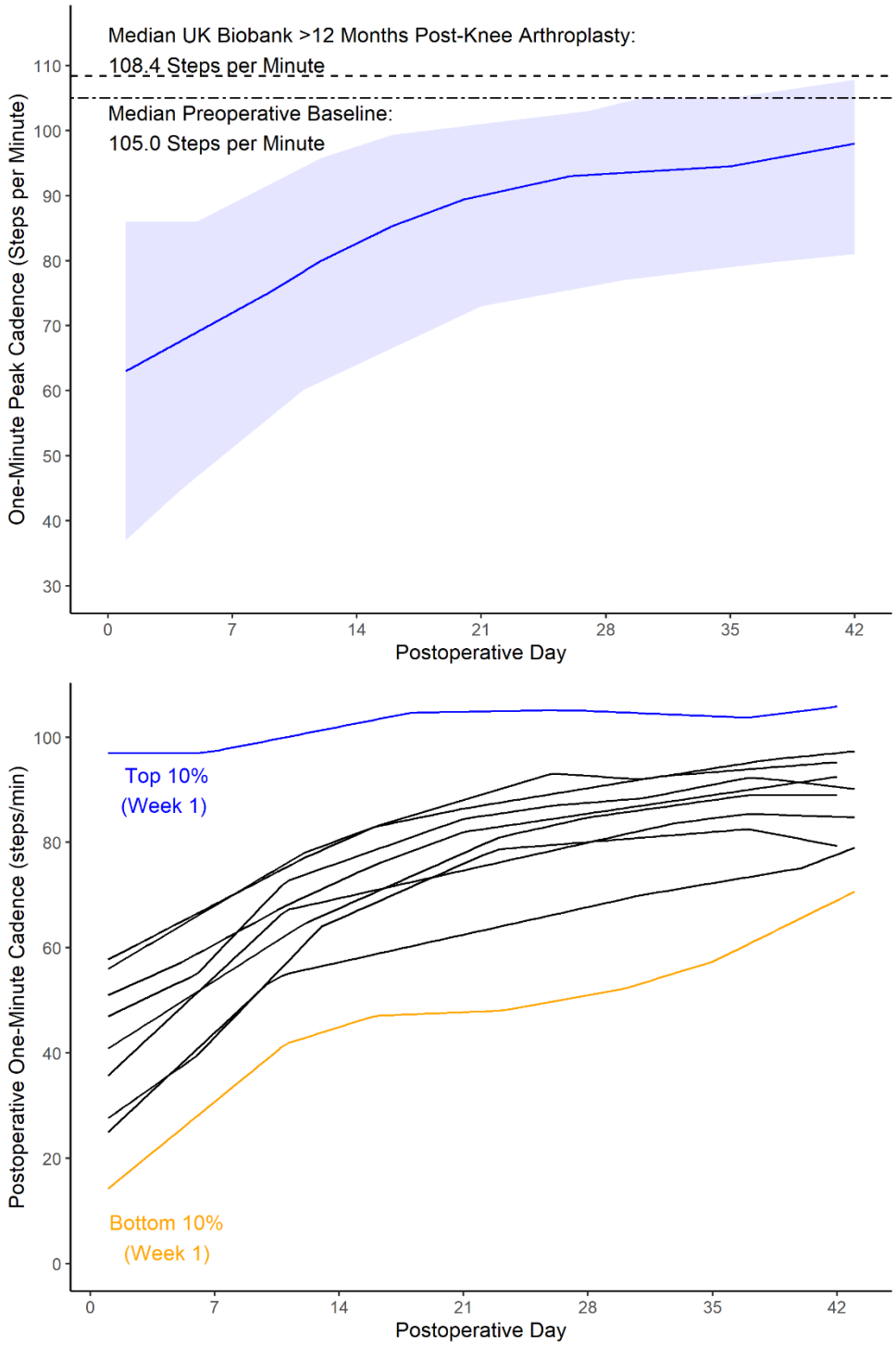


Figure 7-4: (Top) Quantile regression plot of increasing peak cadence during the six weeks following knee arthroplasty, with solid line representing the median and shaded region representing the interquartile range. (Bottom) Median quantile regression plots of participant deciles based on median peak cadence during the first postoperative week. The most active and least active deciles are highlighted, while deciles making up the middle 80% of participants are shown in black.

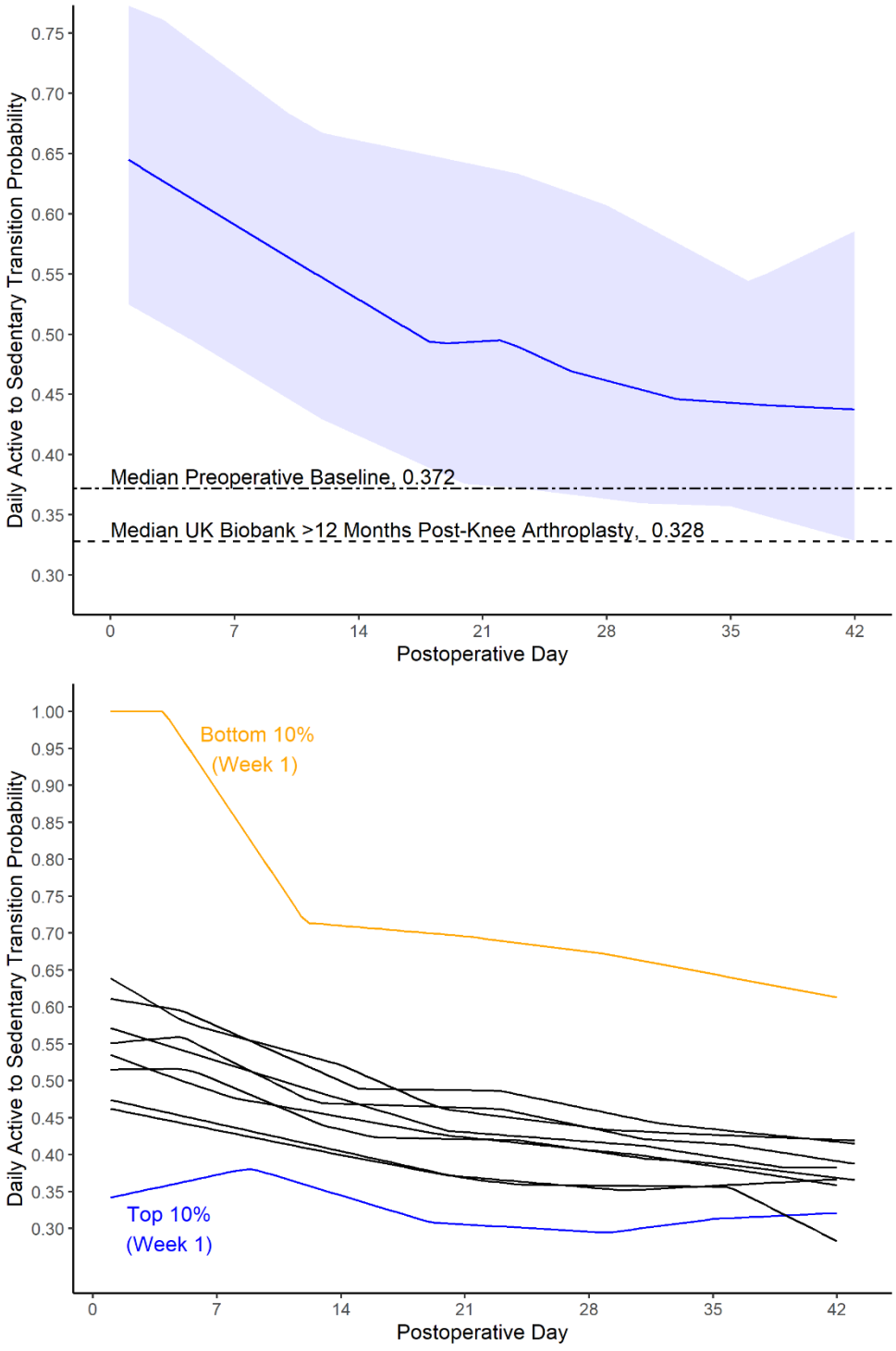


Figure 7-5: (Top) Quantile regression plot of decreasing activity fragmentation during the six weeks following knee arthroplasty, with solid line representing the median and shaded region representing the interquartile range. (Bottom) Median quantile regression plots of participant deciles based on median fragmentation during the first postoperative week. The most active and least active deciles are highlighted, while deciles making up the middle 80% of participants are shown in black.

7.3.3.4 Trajectory of Recovery Using Machine Learning Activity Classification

The change in postoperative time spent in classified light activity is presented in Figure 7-6. During the postoperative monitoring period, most non-sedentary or sleep time is spent in light physical activity. The pattern of increasing light activity closely matches the pattern of increasing overall acceleration, and activity levels in Week 1 similarly reflect subsequent Week 6 activity. The least active Week 1 decile ended with a Week 6 median 179 [108-269] minutes per day in light activity, compared to a Week 6 mean 334 [252-419] minutes per day of light activity in the most active decile at Week 1. Change in classified time spent in MVPA is presented in Figure 7-7. In contrast to other activity metrics, separation between participants is less clear when classifying the low levels of MVPA across most study participants. The most active decile of participants spent a median of only 2 [0-14] minutes per day in MVPA during Week 6.

Classified time spent in sedentary behaviour is presented in Figure 7-8. A steady decrease in inactive time was observed during the monitoring period. In this cohort, the most active decile spent a median 467 [377-503] minutes per day in sedentary behaviour during Week 6, compared to 666 [548-882] minutes per day in the least active decile. Classified sleep time is presented in Figure 7-9. Elevated sleep is observed in the first postoperative week, followed by a levelling off in the next five postoperative weeks. The decile with least classified sleep time spent a median 457 [281-509] minutes per day sleeping during Week 6, compared to 642 [607-687] minutes of sleep per day in the decile with most sleep.

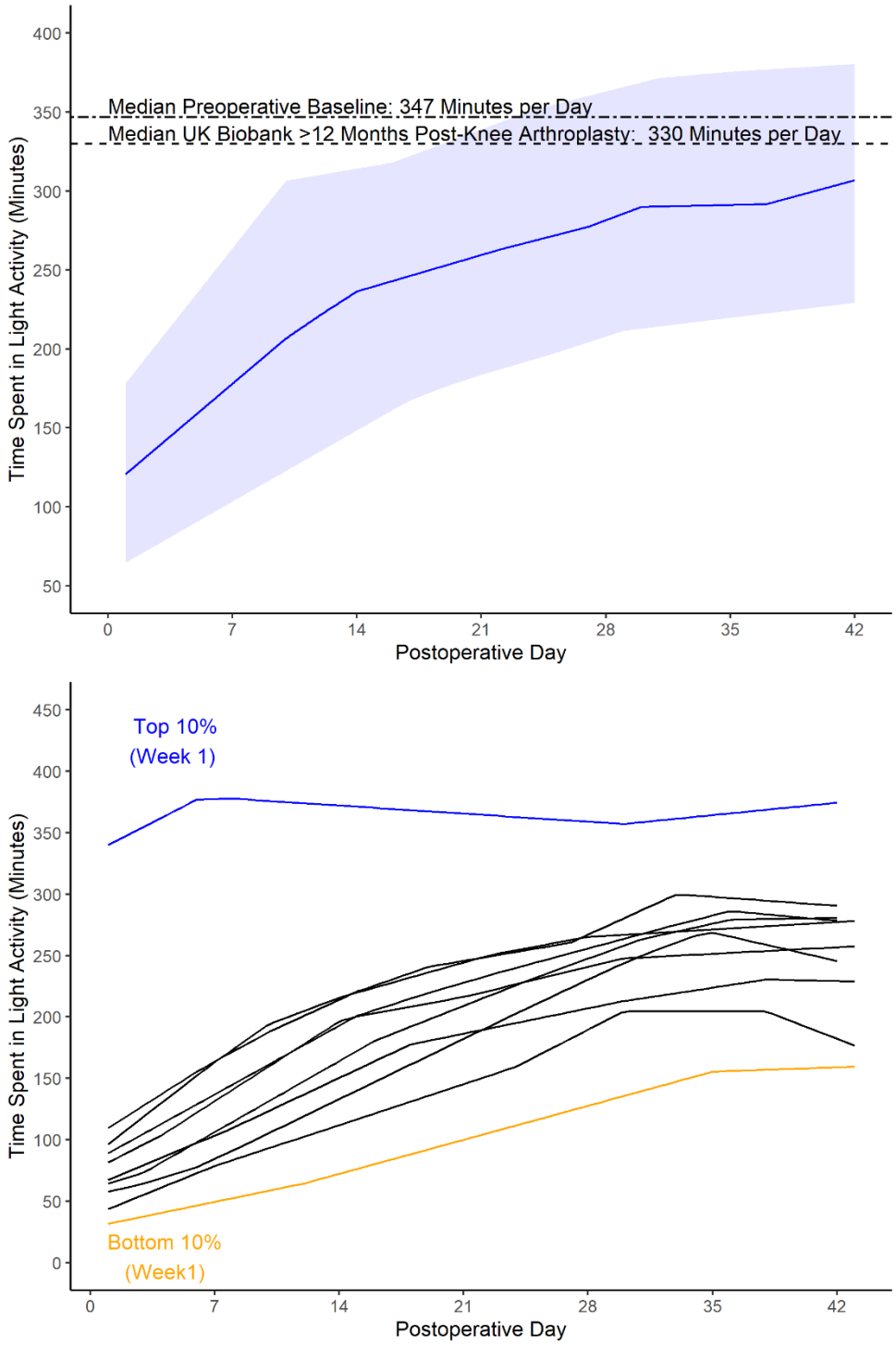


Figure 7-6: (Top) Quantile regression plot of increasing light activity during the six weeks following knee arthroplasty, with solid line representing the median and shaded region representing the interquartile range. (Bottom) Median quantile regression plots of participant deciles based on median light activity during the first postoperative week. The most active and least active deciles are highlighted, while deciles making up the middle 80% of participants are shown in black.

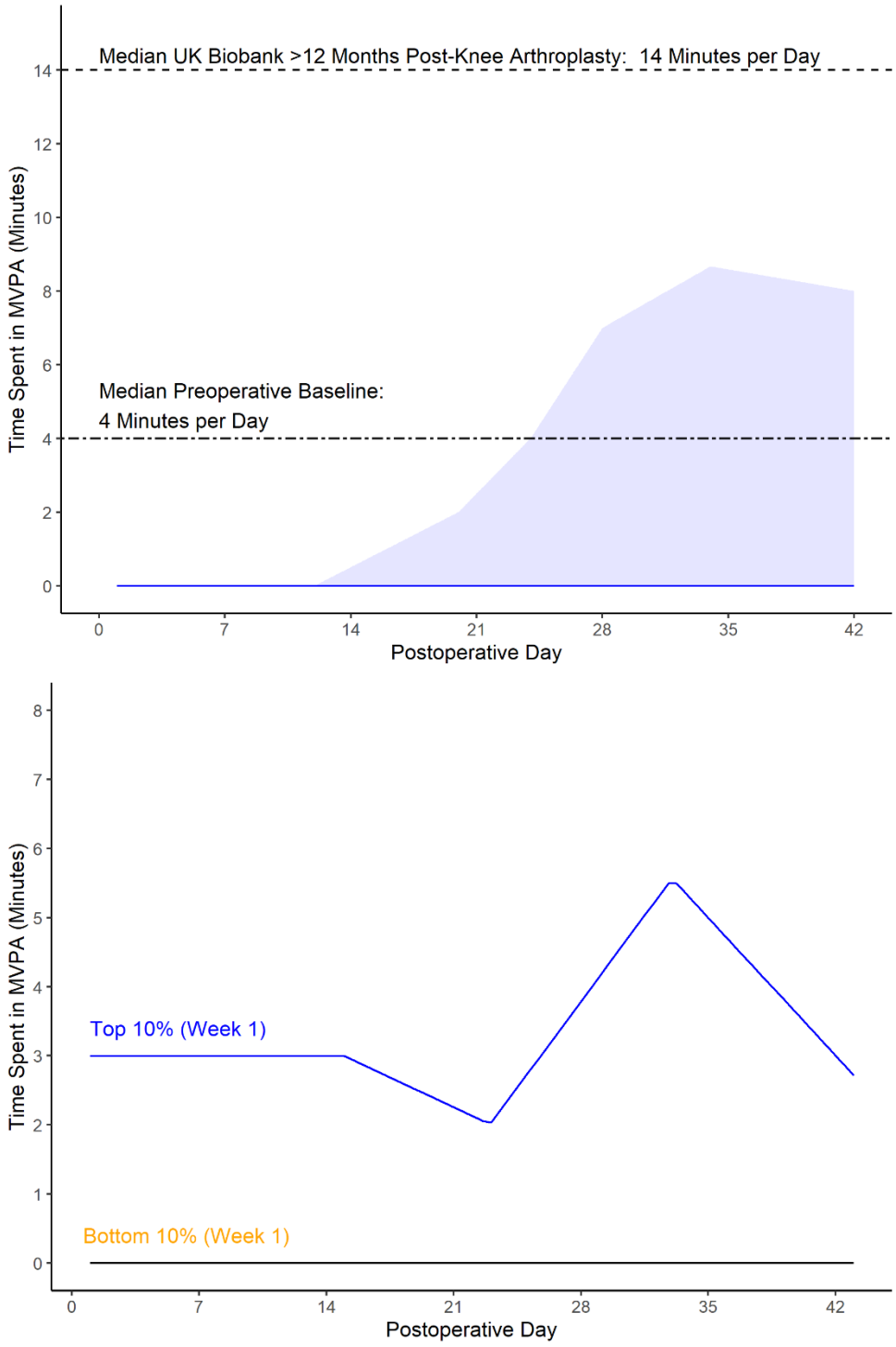


Figure 7-7: (Top) Quantile regression plot of increasing MVPA during the six weeks following knee arthroplasty, with solid line representing the median and shaded region representing the interquartile range. (Bottom) Median quantile regression plots of participant deciles based on median MVPA during the first postoperative week. The most active and least active deciles are highlighted, while deciles making up the middle 80% of participants are shown in black.

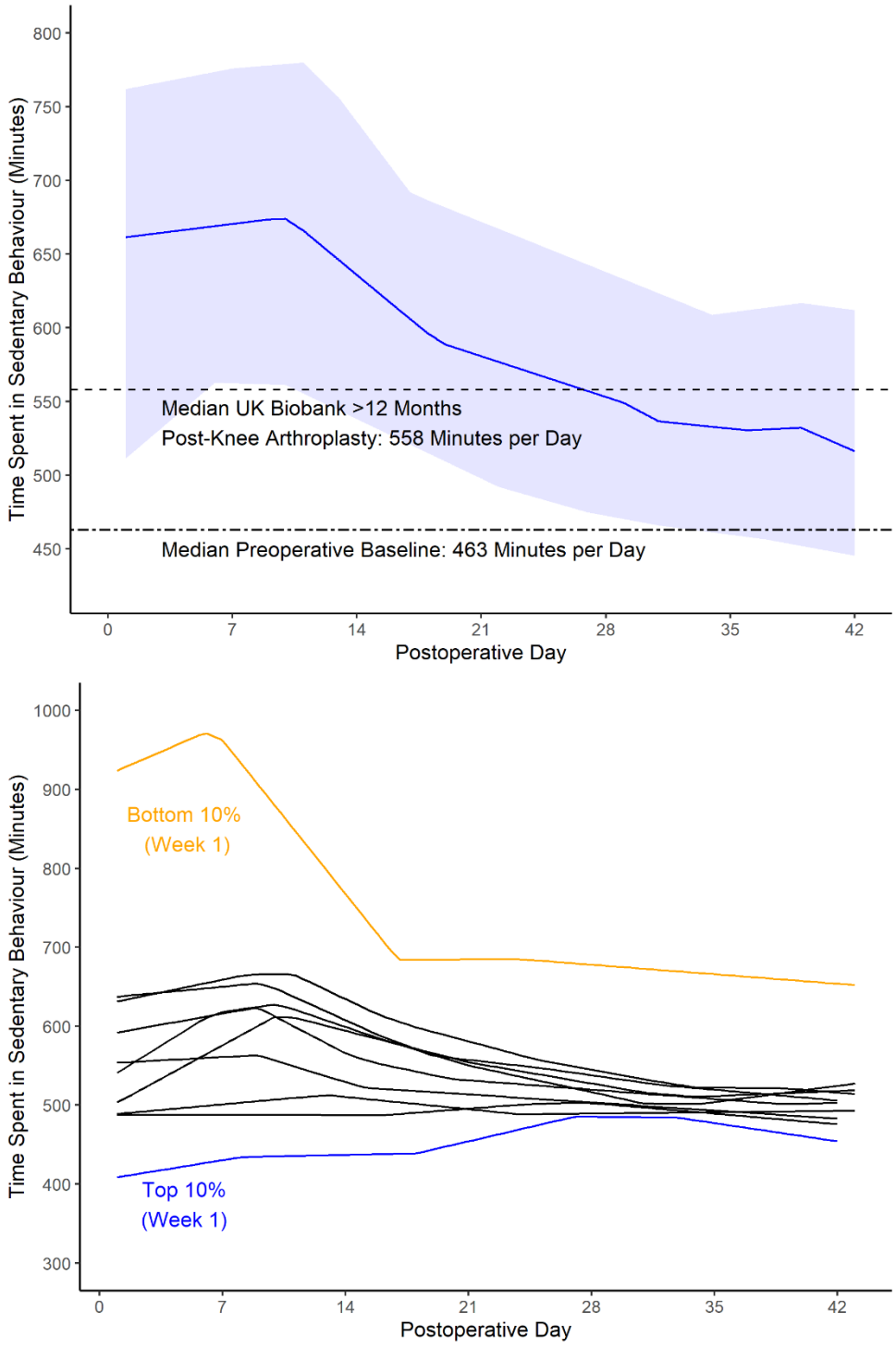


Figure 7-8: (Top) Quantile regression plot of decreasing sedentary activity during the six weeks following knee arthroplasty, with solid line representing the median and shaded region representing the interquartile range. (Bottom) Median quantile regression plots of participant deciles based on median sedentary during the first postoperative week. The most active and least active deciles are highlighted, while deciles making up the middle 80% of participants are shown in black.

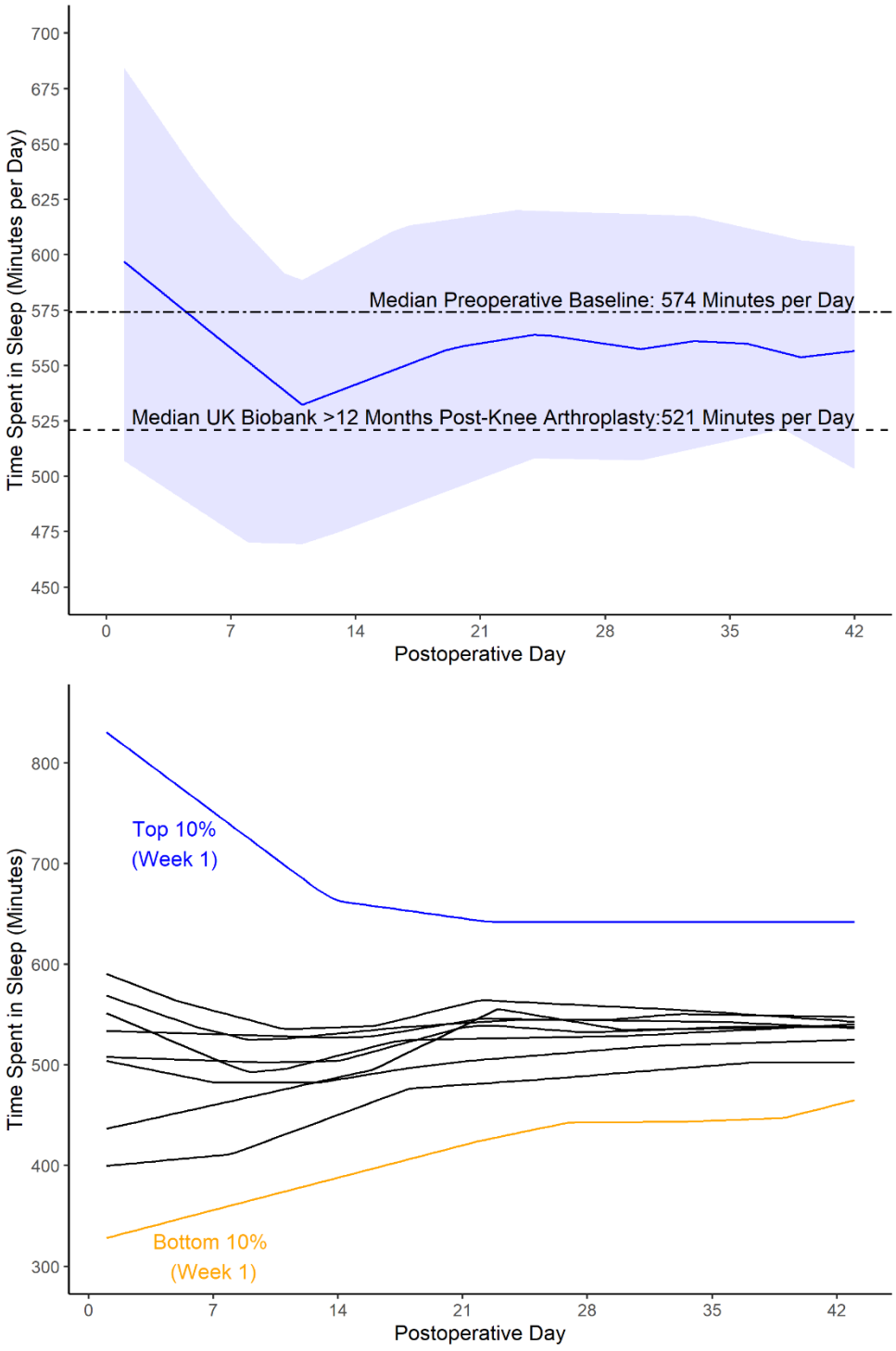


Figure 7-9: (Top) Quantile regression plot of sleep during the six weeks following knee arthroplasty, with solid line representing the median and shaded region representing the interquartile range. (Bottom) Median quantile regression plots of participant deciles based on median sleep during the first postoperative week. The most active and least active deciles are highlighted, while deciles making up the middle 80% of participants are shown in black.

7.3.3.5 Correlation Between Activity Metrics and Self-Reported Physical Activity and PROMs

Correlation coefficients for Week 6 physical activity measures and Week 6 self-reported activity questionnaires and PROMs are presented in Figure 7-10. Overall summary metrics of acceleration and step count were highly correlated with each other (0.76) and with associated acceleration or step-based metrics. Moderate correlation was observed between self-reported physical activity and PROMs. Week 6 pain scores were negatively correlated with PROMs and physical activity metrics, but positively correlated with increased ASTP and sedentary time.

Correlation was low between PROMs and objectively measured physical activity, with the OKS having the highest correlation with Week 6 acceleration (0.19) and step count (0.19). Self-reported activity questionnaires were only moderately correlated with objectively measured physical activity, with a 0.29 correlation between step count and UCLA score, and 0.26 correlation between UCLA score and overall acceleration.

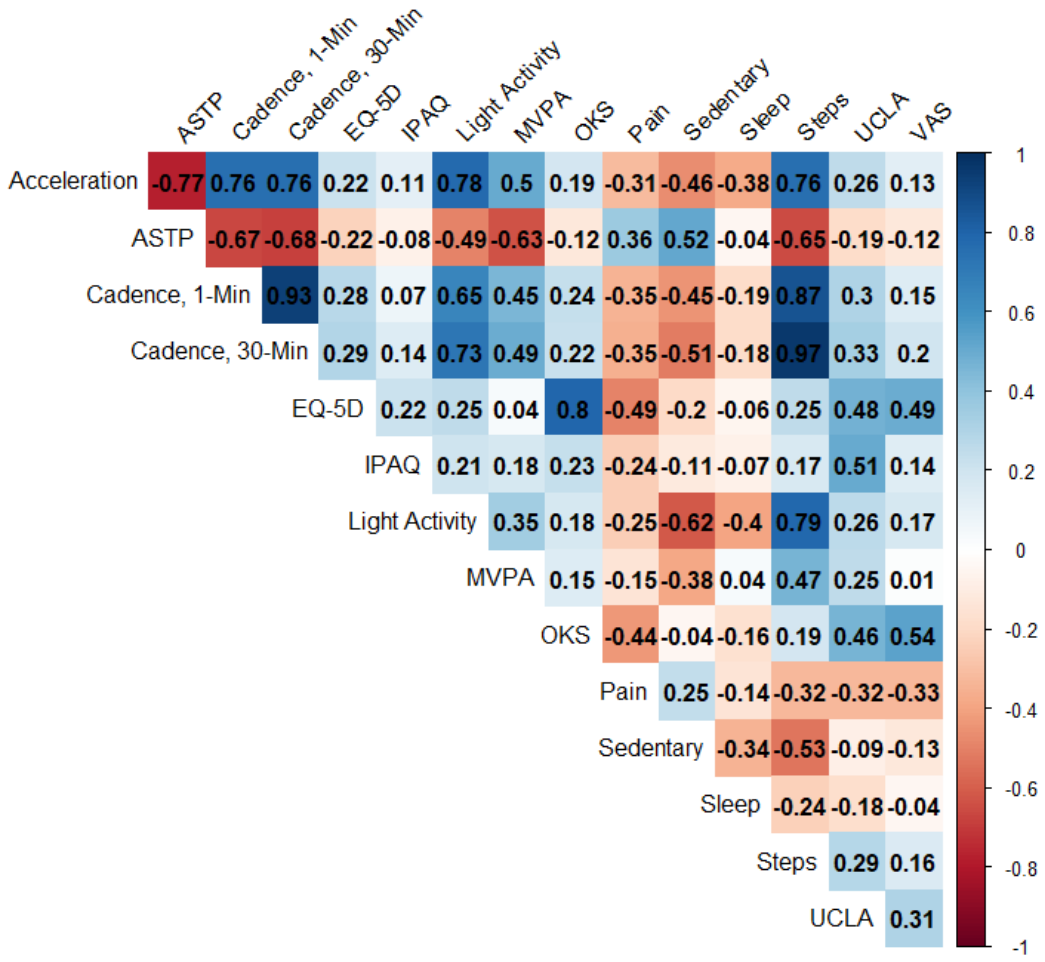


Figure 7-10: Postoperative correlation matrix using Spearman's Rank Test including objectively measured activity parameters, patient reported outcome measures, and self-reported activity questionnaires collected during postoperative Week 6.

Correlation coefficients of objectively measured physical activity between early and late weeks in the postoperative monitoring period are presented in Figure 7-11 and Figure 7-12. These plots parallel the patterns seen in decile plots in Sections 7.3.3.2 - 7.3.3.4 demonstrating high correlation between early and late physical activity. Week 1 acceleration was highly correlated with Week 6 acceleration, 0.61, as was step count (0.58). Similarly moderate to high correlation can also be observed across the derivative physical activity measures, with correlation

coefficients of 0.60, 0.58, 0.58, 0.49, and 0.61 between Week 1 and Week 6 ASTP, 1-minute cadence, light activity, sedentary behaviour, and sleep, respectively. Week 1 pain was moderately correlated with Week 6 pain (0.39) and negatively associated with both Week 1 acceleration (-0.31) and step count (-0.32).

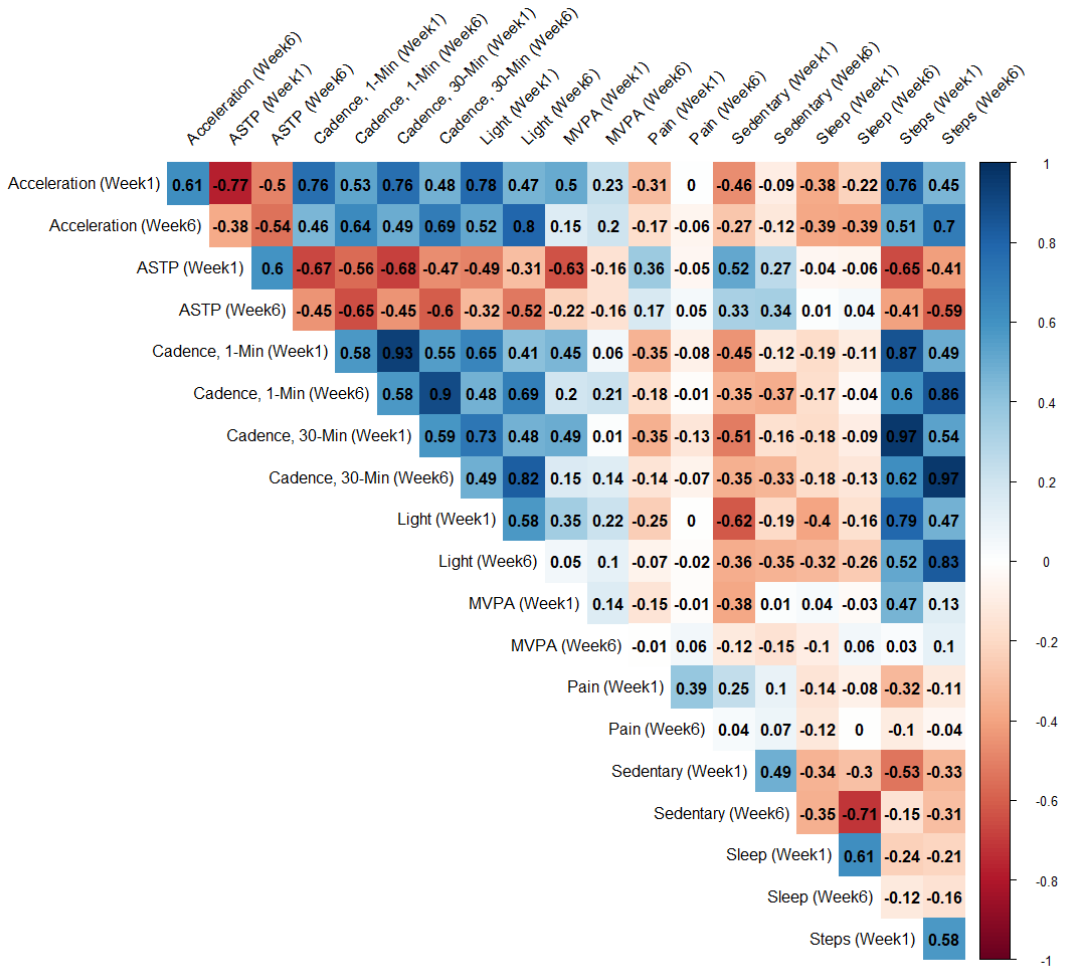


Figure 7-11: Correlation plots comparing Spearman's rank correlation coefficient of activity and pain measurements during postoperative Weeks 1 and 6.

Similar patterns, but generally stronger correlations were observed between Week 2 and Week 6 activity measures, as presented in Figure 7-12. Week 2 acceleration (0.77) and step count (0.76) were both highly correlated with the associated Week 6 measures. Compared to Week 1, Week 2 pain was much more strongly correlated with pain in Week 6, with a correlation coefficient of 0.61.

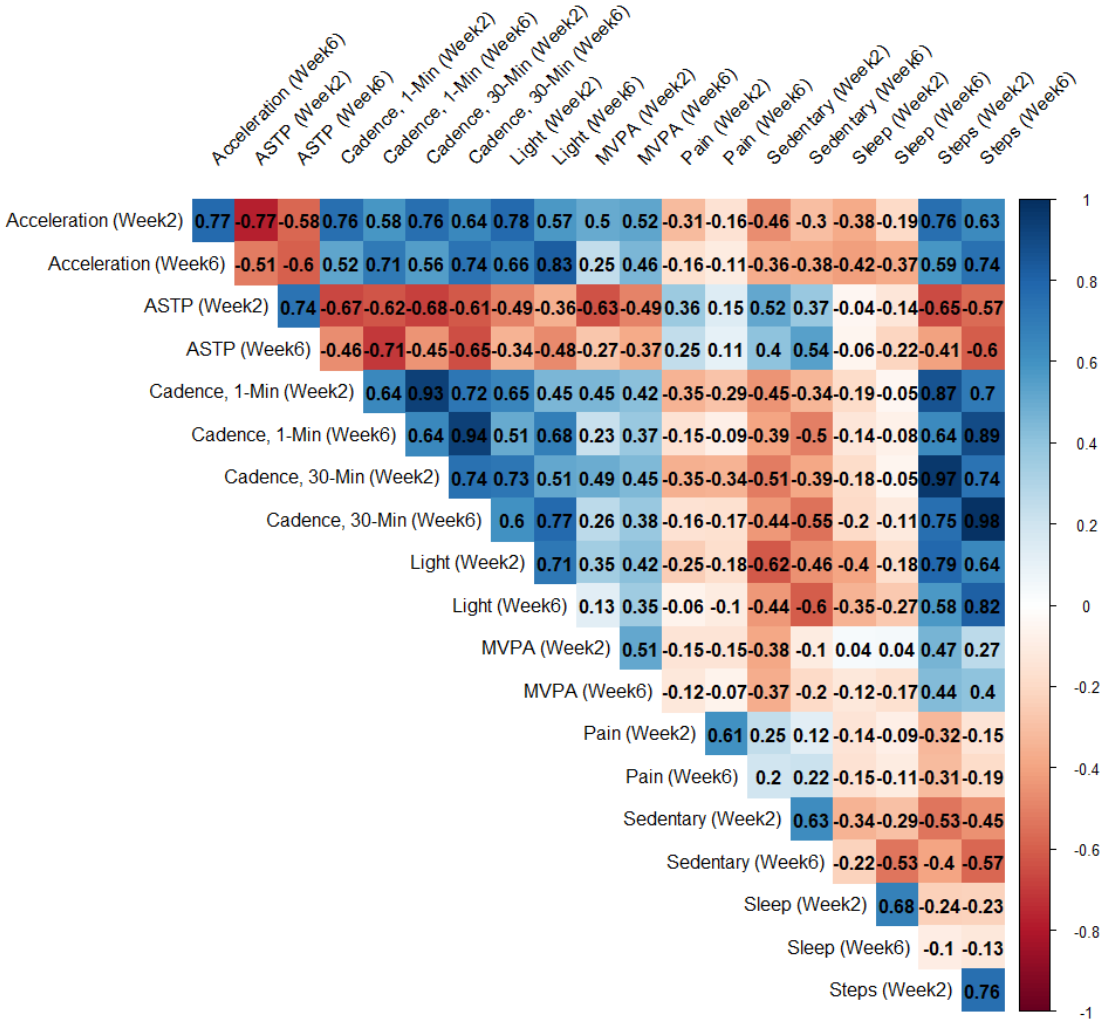


Figure 7-12: Correlation plots comparing Spearman's rank correlation coefficient of activity and pain measurements during postoperative Weeks 2 and 6.

7.3.3.6 Regression Analysis for Postoperative Recovery of Primary Physical Activity Metrics

Coefficients for the mixed-models regression of postoperative step count are presented in Table 7-6 and Table 7-7. The model presented in Table 7-6 includes covariates of routinely collected clinical data without including preoperative step count monitoring. In this regression, participant age, preoperative use of sticks as a walking aid, and postoperative day were the only predictors with statistically significant coefficient estimates. The baseline use of sticks or crutches was associated with a lower step count of 2,283 ($p=0.003$) and 2,965 ($p=0.027$) steps, respectively, while the use of a wheeled frame was also associated with a lower step count, but with a non-significant p-value. Step count significantly increased by 168 steps each postoperative day, with a polynomial correction of -1.5 steps each day^2 ($p<0.001$). The overall model has an R^2 value of 0.34.

Table 7-7 presents the same model with the addition of preoperative step count. In this model, postoperative day ($p<0.001$), and the baseline use of sticks ($p=0.025$) remain statistically significant covariates. Preoperative VAS is a significant covariate in this model with a decrease of 33 steps per VAS point ($p=0.039$). Each additional step in baseline preoperative measurement as associated with 0.22 more steps in the postoperative model ($p<0.001$). This overall model has an R^2 value of 0.40.

Table 7-6: Regression Analysis of Postoperative Step Count without Including Preoperative Baseline Step Count as a Predictor

Covariate (Reference)	Estimate	Confidence Interval	p
Preoperative Step Count	----	----	----
Age	-46	(-91, -3)	0.065
Body Mass Index	-47	(-133, 38)	0.333
Sex (Female)			
Male	-813	(-1,722, 95)	0.120
Ethnicity (White)			
Asian	-302	(-4,265, 3,661)	0.894
Black	-1,475	(-5,512, 2,562)	0.522
Other	-57	(-1,311, 1,197)	0.936
Baseline Walking Aid (Unaided)			
Crutch(s)	**-2,965**	**(-5,280, -651)**	**0.027**
Stick(s)	**-2,283**	**(-3,613, -954)**	**0.003**
Wheeled Frame	-3,167	(-5,977, -358)	0.051
Prior Hip Arthroplasty (None)			
Unilateral	240	(-1,440, 1,920)	0.802
Prior Hip Arthroplasty (None)			
Unilateral	240	(-1,440, 1,920)	0.802
Bilateral	-736	(-2,760, 1,287)	0.524
Self-Reported Disabled (No)			
Yes	-291	(-1,313, 730)	0.618
Preoperative EQ-5D	-1,228	(-5,023, 2,567)	0.571
Preoperative OKS	26	(-54, 107)	0.569
Preoperative UCLA	215	(-55, 485)	0.165
Preoperative VAS	-30	(-59,0)	0.083
Operation Performed (TKA)			
UKA	865	(-87, 1,817)	0.114
Length of Hospital Stay	29	(-247, 304)	0.855
Postoperative Day Number	**168**	**(151, 186)**	**<0.001**
Postoperative Day Number^2	**-1.5**	**(-2, -1)**	**<0.001**

Table 7-7: Regression Analysis of Postoperative Step Count Including Preoperative Baseline Step Count as a Predictor

Covariate (Reference)	Estimate	Confidence Interval	p
Preoperative Step Count	**0.22**	(0.14, 0.32)	**<0.001**
Age	-1	(-46, 44)	0.976
Body Mass Index	-3	(-84, 77)	0.944
Sex (Female)			
Male	-623	(-1,494, 168)	0.168
Ethnicity (White)			
Asian	-190	(-3,805, 3,424)	0.927
Black	-1,320	(-5,002, 2,362)	0.533
Other	180	(-969, 1,328)	0.785
Baseline Walking Aid (Unaided)			
Crutch(s)	-2,256	(-4,392, -123)	0.069
Stick(s)	**-1,633**	**(-2,881, -388)**	**0.025**
Wheeled Frame	-2,109	(-4,717, 494)	0.161
Prior Hip Arthroplasty (None)			
Unilateral	-597	(-2,173, 982)	0.510
Prior Hip Arthroplasty (None)			
Unilateral	-597	(-2,173, 982)	0.510
Bilateral	-413	(-2,265, 1,438)	0.698
Self-Reported Disabled (No)			
Yes	-60	(-997, 877)	0.912
Preoperative EQ-5D	-12	(-3,518, 3,489)	0.995
Preoperative OKS	5	(-69, 80)	0.900
Preoperative UCLA	142	(-106, 391)	0.321
Preoperative VAS	**-33**	**(-60, -5)**	**0.039**
Operation Performed (TKA)			
UKA	589	(-288, 1,466)	0.245
Length of Hospital Stay	16	(-236, 268)	0.912
Postoperative Day Number	**168**	**(151, 186)**	**<0.001**
Postoperative Day Number^2	**-1.5**	**(-2, -1)**	**<0.001**

Coefficients for the mixed-models regression of postoperative acceleration are presented in Table 7-8 and Table 7-9. The model presented in Table 7-8 includes covariates of routinely collected clinical data without including preoperative step count monitoring. In this regression, participant age, sex, preoperative use of sticks as a walking aid, and postoperative day were the only predictors with statistically significant coefficient estimates. The baseline use of sticks was associated with a 3.96 m*g* lower acceleration (p=0.014); while other walking aids were also associated with lower acceleration, but with non-significant p-values. Daily acceleration decreased by 0.13 m*g* for each additional year of participant age (p=0.015). Acceleration significantly increased by 0.24 m*g* each postoperative day, with a polynomial correction of -0.002 m*g* each day^2 (p<0.001). Unlike the step count model, male sex was associated with a lower acceleration (3.96 m*g*, p<0.001). The overall model has an R^2 value of 0.31.

Table 7-9 presents the same model with the addition of preoperative acceleration. Each m*g* of preoperative acceleration was associated with a 0.33 m*g* increase in postoperative acceleration (p<0.001). In this model, postoperative day (p<0.001) and postoperative day^2 (p<0.001) remain significant predictors of physical activity, while stick use is no longer significant predictor when preoperative acceleration is included in the model (p=0.124). Preoperative VAS is a significant predictor of postoperative acceleration in this model, with a decrease of 0.07 m*g* per point on the VAS scale (0.025), while male sex is associated with a 2.92 m*g* decrease in postoperative activity (p=0.002). The overall model has an R^2 value of 0.43.

Table 7-8: Regression Analysis of Postoperative Acceleration without Including Preoperative Baseline Acceleration as a Predictor

	Estimate (mg)	Confidence Interval	p
Preoperative Acceleration	----	----	----
Age	**-0.13**	**(-0.22, -0.04)**	**0.015**
Body Mass Index	**-0.23**	**(-0.40, -0.05)**	**0.025**
Sex (Female)			
Male	**-3.96**	**(-5.83, -2.10)**	**<0.001**
Ethnicity (White)			
Asian	-0.40	(-8.55, 7.75)	0.932
Black	-0.12	(-8.41, 8.18)	0.981
Other	0.81	(-1.77, 3.39)	0.582
Baseline Walking Aid (Unaided)			
Crutch(s)	-5.02	(-9.78, -0.26)	0.068
Stick(s)	**-3.96**	**(-6.69, -1.22)**	**0.013**
Wheeled Frame	-3.55	(-9.32, 2.23)	0.284
Prior Hip Arthroplasty (None)			
Unilateral	2.02	(-1.43, 5.48)	0.307
Bilateral	0.29	(-3.87, 4.45)	0.904
Prior Knee Arthroplasty (None)			
Contralateral	-1.23	(-3.15, 0.70)	0.265
Self-Reported Disabled (No)			
Yes	-0.57	(-2.67, 1.53)	0.635
Preoperative EQ-5D	-1.59	(-9.40, 6.21)	0.720
Preoperative OKS	0.01	(-0.16, 0.17)	0.931
Preoperative UCLA	0.54	(-0.02, 1.09)	0.093
Preoperative VAS	-0.07	(-0.13, -0.01)	0.054
Operation Performed (TKA)			
UKA	1.21	(-0.75, 3.17)	0.281
Length of Hospital Stay	-0.03	(-0.59, 0.54)	0.944
Postoperative Day	**0.24**	**(0.21, 0.27)**	**<0.001**
Postoperative Day^2	**-0.002**	**(-0.003, -0.002)**	**<0.001**

Table 7-9: Regression Analysis of Postoperative Acceleration Including Preoperative Baseline Acceleration as a Predictor

	Estimate (m*g*)	Confidence Interval	p
Preoperative Acceleration	**0.33**	**(0.23, 0.43)**	**<0.001**
Age	0.01	(-0.07, 0.10)	0.808
Body Mass Index	-0.07	(-0.22, 0.09)	0.441
Sex (Female)			
Male	**-2.92**	**(-4.51, -1.32)**	**0.002**
Ethnicity (White)			
Asian	-0.98	(-7.80, 5.84)	0.803
Black	1.13	(-5.81, 8.08)	0.776
Other	0.78	(-1.38, 2.93)	0.532
Baseline Walking Aid (Unaided)			
Crutch(s)	-3.09	(-7.11, 0.93)	0.183
Stick(s)	-2.10	(-4.45, 0.26)	0.124
Wheeled Frame	-2.16	(-7.00, 2.70)	0.440
Prior Hip Arthroplasty (None)			
Unilateral	0.45	(-2.48, 3.37)	0.791
Bilateral	-0.00	(-3.48, 3.48)	0.999
Prior Knee Arthroplasty (None)			
Contralateral	-1.67	(-3.29, -0.06)	0.074
Self-Reported Disabled (No)			
Yes	-0.41	(-2.17, 1.35)	0.685
Preoperative EQ-5D	2.11	(-4.52, 8.73)	0.580
Preoperative OKS	-0.05	(-0.19, 0.09)	0.565
Preoperative UCLA	0.31	(-0.16, 0.78)	0.247
Preoperative VAS	**-0.07**	**(-0.12, -0.02)**	**0.025**
Operation Performed (TKA)			
UKA	0.67	(-0.98, 2.31)	0.480
Length of Hospital Stay	-0.10	(-0.58, 0.37)	0.711
Postoperative Day	**0.24**	**(0.21, 0.27)**	**<0.001**
Postoperative Day^2	**-0.002**	**(0.003, 0.002)**	**<0.001**

7.3.4 Wearable Physical Activity Monitoring for the Comparison of Unicompartmental and Total Knee Arthroplasty Patients

For comparison of postoperative recovery between participants undergoing UKA and TKA, weekly median daily steps for both clinical cohorts are presented in Figure 7-13. Weekly postoperative acceleration is presented in Figure 7-14, while weekly comparisons of objectively measured physical activity across summary and derived physical activity metrics are presented in Table 7-10, Table 7-11, and Table 7-12. Box plots of derived physical activity metrics are presented in Appendix A.3.

Of the 95 participants contributing valid postoperative accelerometer data, a total of 42 participants underwent TKA while 53 participants underwent UKA. Prior to surgery, there were no statistically significant differences in median preoperative baseline objectively measured physical activity between the UKA and TKA cohorts as defined by any activity metric. This equivalence in physical activity between the two clinical cohorts remained during the first postoperative week; however, consistent differences in activity began to emerge within the second postoperative week, when the UKA cohort demonstrated 1,122 more daily steps (p=0.015) and 13.3 more steps per minute in peak 30-minute cadence (p=0.009). By Week 2, TKA patients were spending a median 67 more minutes in sedentary behaviour than UKA patients (p=0.014). In postoperative Week 3 the UKA cohort demonstrated a 19% lower ASTP (p=0.018). A statistically significant divergence in overall acceleration began in postoperative Week 5, with the UKA cohort generating 12.6% more daily acceleration (p=0.008). During the sixth postoperative week, the UKA cohort demonstrated better activity across all metrics, apart from time spent in MVPA. Both TKA and UKA cohorts retained

statistically significant lower physical activity in Week 6 when compared to all baseline preoperative activity metrics. No difference in sleep time was observed at any time point between UKA and TKA participants.

As presented in Table 7-12, statistically significant differences between week-to-week activity metrics were observed across nearly all metrics and instances from preoperative baseline to Week 5. However, in both UKA and TKA cohorts, levelling off of early recovery began in Week 6, with no statistically significant difference in acceleration-based metrics between Weeks 5 and 6 in the UKA cohort. Increases in daily step count and cadence metrics remained significant in the UKA cohort between weeks 5 and 6. When comparing Week 5 and Week 6 in the TKA cohort, a significant improvement in activity metrics was only observed in terms of daily steps and ASTP.

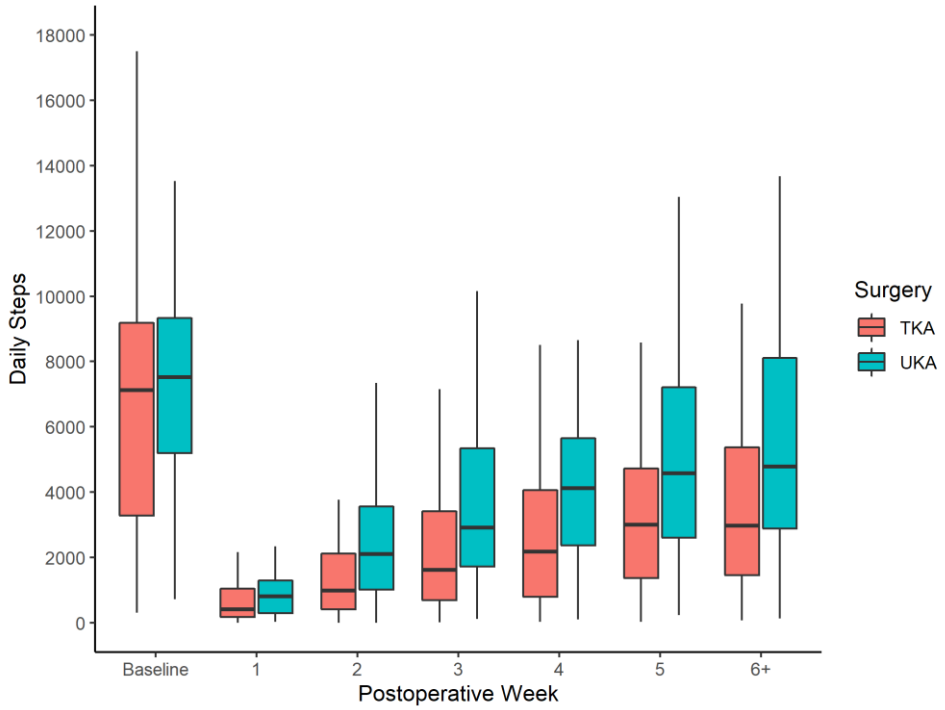


Figure 7-13: Boxplot of daily step count by operative intervention compared with preoperative baseline in participants contributing postoperative data.

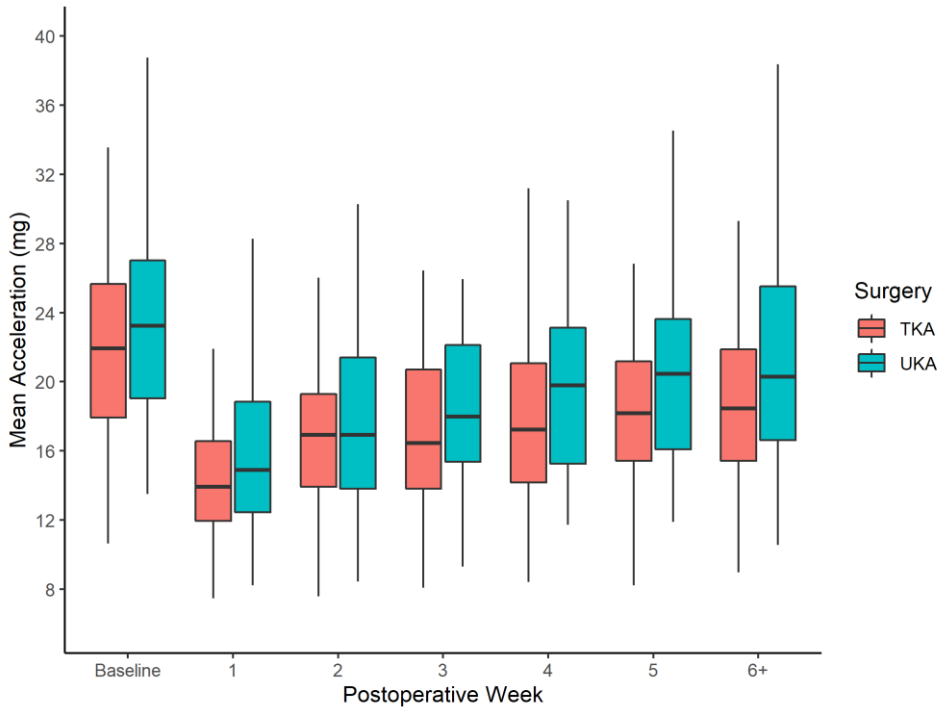


Figure 7-14: Boxplot of acceleration by operative intervention compared with preoperative baseline in all participants contributing postoperative data.

Table 7-10: Comparison of Physical Activity Between UKA and TKA Patients with Valid Wear During Postoperative Recovery (Baseline – Week 3)

Activity Metric	Baseline		Week 1		Week 2		Week 3	
	UKA	TKA	UKA	TKA	UKA	TKA	UKA	TKA
Daily Steps	7,522 [5,189-9,331]	7,122 [3,241-9,270]	806 [294-1,285]	412 [171-1,055]	2,101 [1,003-3,551]	979 [408-2,165]	2,906 [1,720-5,345]	1,613 [674-3,459]
	p=0.191		*p=0.091*		***p=0.015***		***p=0.004***	
Overall Acceleration (mg)	23.3 [19.0-27.0]	21.9 [17.8-25.9]	14.9 [12.4 -18.8]	13.9 [11.8-16.6]	16.9 [13.8-21.4]	16.9 [13.9-19.6]	18.0 [15.4-22.1]	16.4 [13.8-20.8]
	p=0.324		*p=0.098*		*p=0.301*		*p=0.149*	
MVPA (min/day)	3 [1-15]	6 [3-12]	0 [0-0]	0 [0-0]	0 [0-3]	0 [0-1]	0 [0-6]	0 [0-4]
	p=0.393		*p=0.415*		*p=0.319*		*p=0.257*	
Light (min/day)	342 [318-458]	357 [304-420]	164 [96-222]	131 [81-197]	220 [160-310]	203 [119-304]	259 [205-321]	238 [152-324]
	p=0.592		*p=0.231*		*p=0.306*		*p=0.196*	
Sedentary (min/day)	459 [425-528]	467 [392-509]	644 [522-746]	685 [600-771]	627 [529-720]	694 [643-796]	571 [498-664]	625 [559-705]
	p=0.961		*p=0.111*		***p=0.014***		***p=0.015***	
Sleep (min/day)	569 [513-604]	580 [538-631]	586 [525-635]	564 [483-638]	552 [509-588]	530 [442-573]	560 [508-621]	540 [488-603]
	p=0.221		*p=0.587*		*p=0.108*		*p=0.225*	
ASTP	0.342 [0.302-0.433]	0.390 [0.334-0.447]	0.600 [0.503-0.733]	0.629 [0.541-0.766]	0.520 [0.420-0.672]	0.611 [0.501-0.667]	0.463 [0.376-0.578]	0.573 [0.434-0.648]
	p=0.204		*p=0.231*		*p=0.068*		***p=0.018***	
1 min Cadence (steps/min)	105.0 [100.1-110.8]	105.0 [91.3-109.0]	73.0 [56.0-82.4]	60.4 [42.9-81.4]	80.1 [71.1-93.0]	72.1 [53.3-85.9]	92.3 [78.1-101.1]	83.8 [61.1-93.0]
	p=0.221		*p=0.095*		***p=0.014***		***p=0.006***	
30 min Cadence (steps/min)	81.2 [69.9-90.5]	80.4 [55.8-87.2]	27.1 [15.0-35.1]	14.3 [8.3-33.2]	40.5 [30.5-59.6]	27.2 [13.3-46.0]	54.4 [39.3-70.0]	40.2 [23.1-56.4]
	p=0.303		*p=0.067*		***p=0.009***		***p=0.003***	

All values presented as median [interquartile range]

n = 95, TKA:42, UKA: 53

Table 7-11: Comparison of Physical Activity Between UKA and TKA Patients with Valid Wear During Postoperative Recovery (Week 4 – Week 6)

Activity Metric	Week 4		Week 5		Week 6	
	UKA	TKA	UKA	TKA	UKA	TKA
Daily Steps	4,111 [2,364-5,644]	2,175 [770-4,083]	4,568 [2,595-7,212]	3,004 [1,351-4,921]	4,788 [2,883-8,104]	2,971 [1,403-5,492]
	p=0.004		***p=0.008***		***p=0.014***	
Overall Acceleration (mg)	19.8 [15.2-23.1]	17.2 [14.2-21.2]	20.5 [16.1-23.6]	18.2 [15.4-21.2]	20.3 [16.6-25.5]	18.5 [15.3-22.0]
	p=0.080		***p=0.029***		***p=0.028***	
MVPA (min/day)	2 [0-6]	1 [0-5]	3 [0-10]	2 [0-8]	4 [0-10]	3 [0-6]
	p=0.647		*p=0.341*		*p=0.931*	
Light (min/day)	286 [226-349]	244 [176-350]	311 [251-371]	273 [209-350]	328 [248-394]	272 [216-361]
	p=0.139		*p=0.088*		***p=0.039***	
Sedentary (min/day)	541 [471-598]	598 [528-678]	527 [469-583]	566 [507-628]	519 [448-560]	557 [483-637]
	p=0.023		***p=0.022***		***p=0.030***	
Sleep (min/day)	570 [524-619]	562 [521-600]	557 [518-613]	562 [529-600]	554 [525-607]	552 [533-600]
	p=0.429		*p=0.985*		*p=0.896*	
ASTP	0.459 [0.339-0.565]	0.525 [0.435-0.652]	0.430 [0.324-0.516]	0.492 [0.419-0.640]	0.402 [0.330-0.516]	0.474 [0.379-0.607]
	p=0.011		***p=0.009***		***p=0.032***	
1 min Cadence (steps/min)	97.5 [79.4-102.7]	87.8 [73.1-95.0]	97.6 [80.1-104.4]	92.2 [78.0-99.7]	99.3 [86.4-107.0]	92.0 [74.6-101.3]
	p=0.009		***p=0.023***		***p=0.005***	
30 min Cadence (steps/min)	66.2 [43.5-77.1]	46.5 [23.8-64.2]	65.5 [49.0-78.7]	52.8 [33.7-69.7]	65.1 [53.9-86.2]	55.8 [37.7-69.6]
	p=0.003		***p=0.012***		***p=0.011***	

All values presented as median [interquartile range]
n = 95, TKA:42, UKA: 53

Table 7-12: Wilcoxon Rank Test P-Values of Week-to-Week Change in Activity Metrics
UKA Postoperative Cohort

Activity Metric	Baseline to Week 1	Week 1 to Week 2	Week 2 to Week 3	Week 3 to Week 4	Week 4 to Week 5	Week 5 to Week 6	Week 1 to Week 6	Baseline to Week 6
Daily Steps	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.009**	**<0.001**	**<0.001**
Overall Acceleration	**<0.001**	**<0.001**	**<0.001**	**0.010**	**<0.001**	0.119	**<0.001**	**<0.001**
MVPA	**<0.001**	**0.003**	**0.042**	**0.022**	**0.007**	0.080	**<0.001**	**0.031**
Light	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	0.114	**<0.001**	**<0.001**
Sedentary	**<0.001**	0.182	**<0.001**	**0.004**	**0.004**	0.173	**<0.001**	**<0.001**
Sleep	0.098	**<0.001**	**0.007**	0.530	0.325	0.818	0.081	0.626
ASTP	**<0.001**	**<0.001**	**<0.001**	0.074	**0.003**	0.183	**<0.001**	**<0.001**
1 min Cadence	**<0.001**	**<0.001**	**<0.001**	**0.011**	0.146	**0.002**	**<0.001**	**<0.001**
30 min Cadence	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.010**	**0.003**	**<0.001**	**<0.001**

TKA Postoperative Cohort

Activity Metric	Baseline to Week 1	Week 1 to Week 2	Week 2 to Week 3	Week 3 to Week 4	Week 4 to Week 5	Week 5 to Week 6	Week 1 to Week 6	Baseline to Week 6
Daily Steps	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.021**	**<0.001**	**<0.001**
Overall Acceleration	**<0.001**	**<0.001**	**0.015**	0.464	**0.029**	0.078	**<0.001**	**<0.001**
MVPA	**<0.001**	**0.009**	0.093	**0.044**	**0.025**	0.222	**<0.001**	**0.009**
Light	**<0.001**	**<0.001**	**<0.001**	**0.029**	**0.013**	0.105	**<0.001**	**<0.001**
Sedentary	**<0.001**	**0.908**	**<0.001**	**0.001**	**0.012**	0.276	**<0.001**	**<0.001**
Sleep	0.214	**0.003**	**<0.001**	**0.039**	0.452	1.000	0.586	**0.002**
ASTP	**<0.001**	**0.005**	**<0.001**	0.153	**0.003**	**0.005**	**<0.001**	**<0.001**
1 min Cadence	**<0.001**	**0.001**	**<0.001**	**<0.001**	**0.014**	0.322	**<0.001**	**<0.001**
30 min Cadence	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	0.064	**<0.001**	**<0.001**

7.3.5 Case Studies

7.3.5.1 Case Study 1: Participant Requiring Manipulation Under Anesthesia

The first case study participant had a preoperative baseline activity level of 8,266 steps per day, 103 steps per minute peak cadence, 31.4 m*g* overall acceleration, and walked without aid. Upon completion of a TKA, the patient had 50 degrees of early knee flexion and was given two crutches to ambulate during early postoperative recovery. Daily step counts for this participant, as plotted against the clinical study mean recovery curve, are presented in Figure 7-15. While the participant's baseline preoperative activity is higher than average across most metrics, postoperative activity remained well below average for most of the postoperative window, though daily step counts increased aggressively in the sixth postoperative week.

Postoperative recovery of daily acceleration for this participant is presented in Figure 7-16. Counter to daily step count, this participant's daily acceleration more closely followed the mean overall acceleration recovery curve for the entire study population. At the Week 6 postoperative clinical evaluation, the participant presented to the clinical walking unaided with 70 degrees of knee flexion. In terms of clinical PROMs, the participant's OKS dropped by 4 points from baseline to 6 weeks postop, EQ-5D index score dropped by 0.194 points, and the VAS score dropped 20 points from baseline to Week 6. This participant was subsequently readmitted to the hospital and underwent a manipulation under anaesthesia (MUA) for treatment of limited joint range of motion within a month following the Week 6 clinical evaluation.

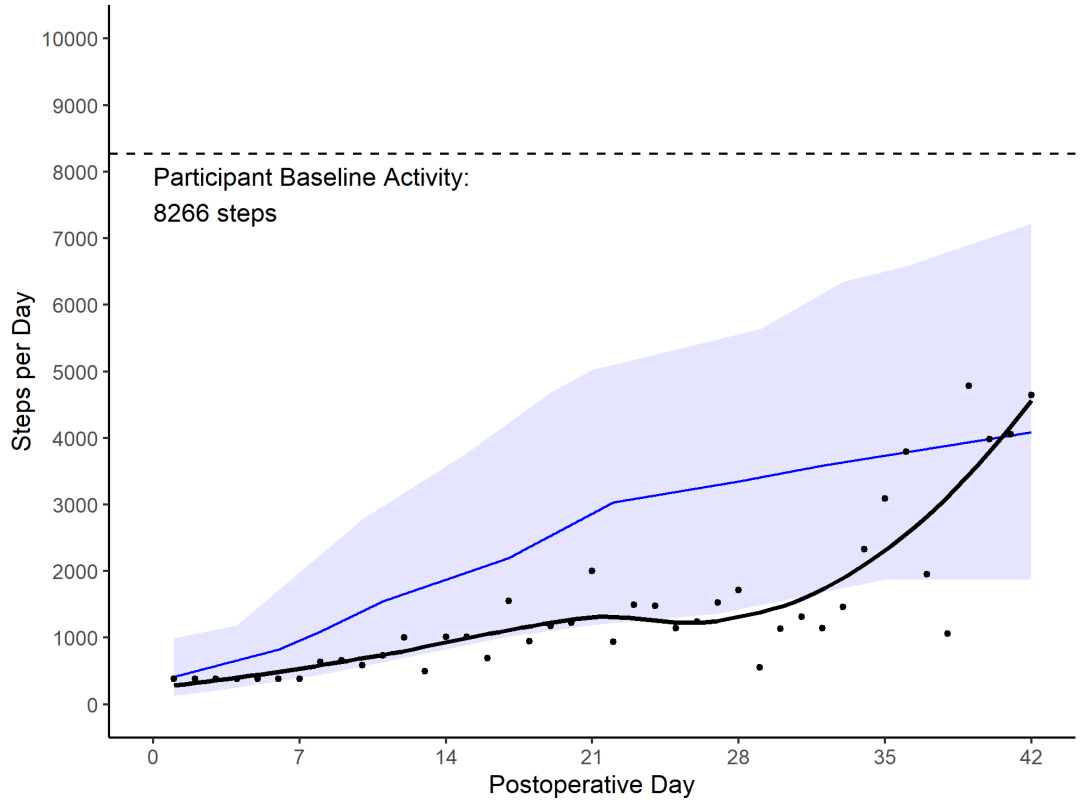


Figure 7-15: Case Study Participant 1: Individual daily step counts (dots) and LOESS curve (black) in a single participant compared to the study population quantile regression plot, with the solid blue line representing the population median with shaded population interquartile range. This patient was later admitted for a manipulation under anaesthesia due to poor knee range of motion.

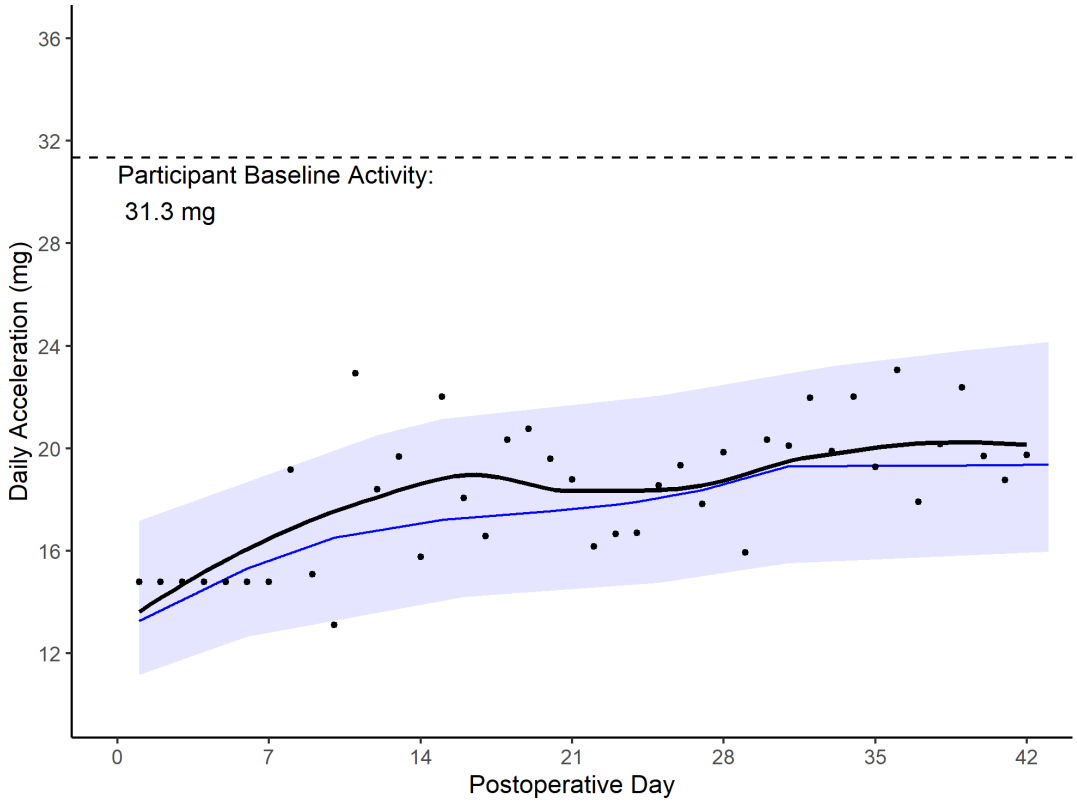


Figure 7-16: Case Study Participant 1: Individual daily acceleration (dots) and LOESS curve (black) in a single participant compared to the study population quantile regression plot, with the solid blue line representing the population median with shaded population interquartile range.

7.3.5.2 Case Study 2: Participant with an Early Downturn in Activity

The second case study participant had a preoperative baseline activity level of 9,270 steps per day, 102 steps per minute peak cadence, 32.4 m*g* overall acceleration, and walked unaided. Upon completion of a TKA, the patient had 80 degrees of early knee flexion and was given two crutches to ambulate during early recovery. Daily step counts for this participant, plotted against the clinical study mean recovery curve, are presented in Figure 7-17. This participant demonstrated higher than average postoperative step counts during the first four postoperative weeks, followed by an abrupt downturn in daily step count in the fifth and sixth postoperative weeks. Postoperative acceleration, not presented, closely follows the pattern shown in daily step count.

Postoperative ASTP for this participant is presented in Figure 7-18. Similar to daily step count, the participant's ASTP indicates above average mobility within the first postoperative month, followed by an abrupt increase in activity fragmentation. At the sixth week clinical evaluation, this patient reported having a setback in Week 4 associated with a painful "turn". Nevertheless, the participant presented well at the Week 6 evaluation, walking unaided with greater than 90 degrees of knee flexion, and was discharged normally. In terms of clinical PROMs, the participant's OKS increased by 12 points from baseline to 6 weeks postop, EQ-5D index score increased by 0.031 points, and the VAS score increased 33 points from baseline to Week 6.

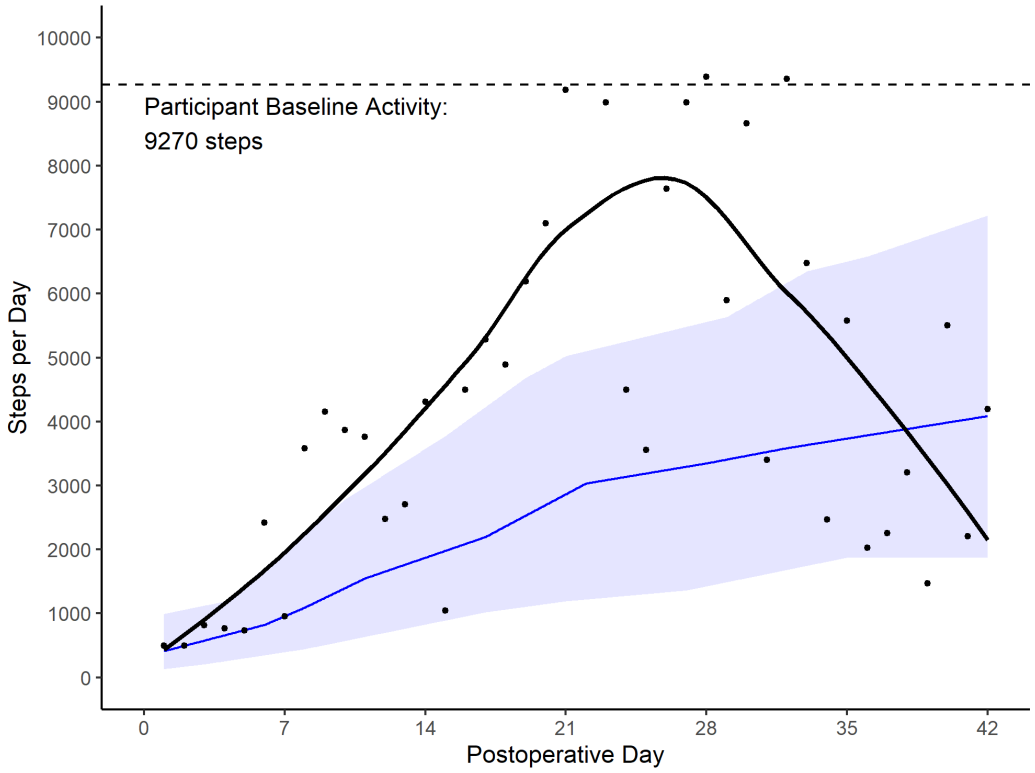


Figure 7-17: Case Study Participant 2: Individual daily step counts (dots) and LOESS curve (black) in a single participant compared to the study population quantile regression plot, with the solid blue line representing the population median with shaded population interquartile range. This patient reported an at-home injury on the operative knee during the fourth postoperative week.

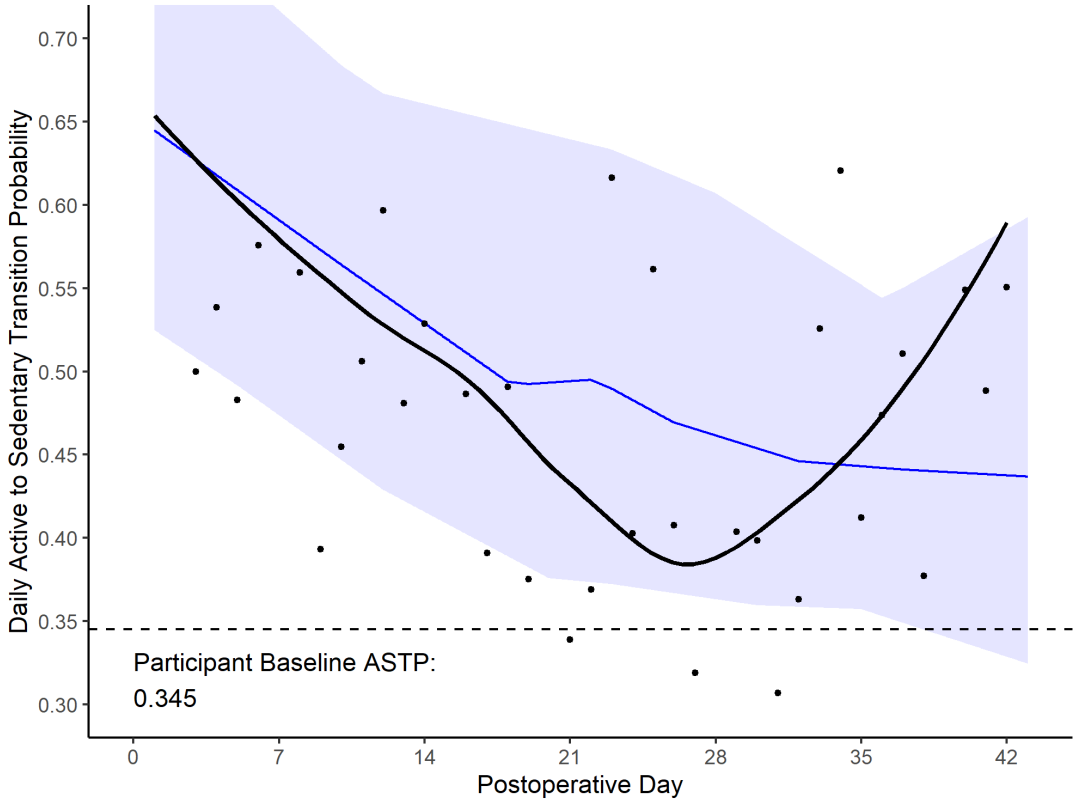


Figure 7-18: Case Study Participant 2: Individual daily activity fragmentation (dots) and LOESS curve (black) in a single participant compared to the study population quantile regression plot, with the solid blue line representing the population median with shaded population interquartile range. The data presents a dramatic shift in mobility post-injury during the early postoperative recovery period.

7.4 Discussion

The aim of this study was to describe the trajectory of early recovery following knee arthroplasty, while assessing the utility of remote patient monitoring in the early postoperative clinical setting. Through baseline and postoperative data collection, this study set out to identify trends in postoperative recovery while identifying metrics appropriate for the assessment of perioperative patient mobility. Two primary methods for integrating wrist-based accelerometry into perioperative monitor were assessed, the use of monitoring to differentiate recovery patterns in normal and poor recovers, and the use of monitoring as a clinical trial outcome measure. When evaluating postoperative recovery trajectory, a clear increase in overall physical activity, decrease in sedentary behaviour, and decrease in pain was observed in most patients during the first six weeks. When tested as a clinical trials outcome, earlier and greater improvements in physical activity were observed in patients undergoing UKA, a less invasive surgical intervention, when compared to TKA. This study demonstrates the considerable potential of wearable-based patient monitoring as a tool for improving patient care and evaluating surgical techniques and rehabilitation pathways.

Nonlinear patterns of recovery were observed in pain score and across all physical activity metrics during the six-week postoperative monitoring period, with fastest activity gains observed in the first three weeks, followed by a slightly slower, yet steady rate of recovery in postoperative weeks four through six. Study participants exhibited significant week-to-week increases in daily step count, acceleration, cadence, and light and moderate-to-vigorous activity, with associated decreases in sedentary behaviour and activity fragmentation. As demonstrated through exploratory decile grouping, trends of recovery were well in place within the first

one to two weeks of recovery, with the least and most active participants within these earliest weeks remaining so throughout the monitoring period. In further analysis, activity data from the first postoperative week was highly correlated with final observed activity levels in the sixth postoperative week, with even stronger correlation between Week 2 and Week 6 activity metrics. Due to COVID-19 related protocol adjustments, data collection for many participants did not begin until the third or fourth postoperative day, thus further work is needed to collect activity data beginning at Day 0 to confirm the predictive power of activity recovery in the first and second weeks. If confirmed, a short monitoring period beginning immediately following surgery and following up to the first physiotherapy evaluation during at 7-10 postoperative days could provide significant insight to inform treatment planning and identify areas of early concern.

The most convenient clinical deployment of activity monitoring would be to eliminate preoperative data collection and distribute sensors only on the day of operation. Regression of overall summary metrics was performed in this analysis both excluding and including preoperative activity measurement to assess the relative model quality without incorporating preoperative monitoring. Similar to the preoperative regression performed in Chapter 6, age and the use of sticks as a walking aid were both significantly associated with postoperative activity in terms of both step count and overall acceleration when incorporating only postoperative monitoring. As expected, postoperative day (days since surgery) was additionally a significant factor in postoperative activity. When including preoperative activity into the regression models, age was no longer a significant predictor; however, postoperative day number, use of sticks, and preoperative VAS scores had significant associations with postoperative activity. The inclusion of preoperative

monitoring in the regression of postoperative recovery seems to account for most age-related activity differences, while improving the model fit as demonstrated by an increase in R^2 from 0.34 to 0.40 in the step model and 0.31 to 0.43 in the acceleration model.

Shorter hospital stays, improved function, and reduced morbidity following UKA in comparison to TKA is well documented[247]; however, comparative differences in early postoperative, objectively measured physical activity following UKA and TKA have not been previously reported. The use of wearable sensors for perioperative monitoring allows for well-defined assessment of patient mobility and return to activity when comparing the two surgical techniques and associated rehabilitation pathways. Greater physical activity and faster recovery in a UKA patient cohort would be anticipated based on prior studies incorporating PROMs and self-reported measures of activity[248,249]. In this study, no statistically significant differences in preoperative physical activity were observed between the two surgical cohorts. Following surgery, UKA patients demonstrated a statistically significant improvement in physical activity versus TKA patients beginning in the second postoperative week. Significant differences in mobility were first identified through daily step count and the closely correlated peak 30-minute cadence in Week 2, followed by ASTP-measured activity fragmentation in Week 3, and finally by overall acceleration by Week 5. The varied sensitivity to differences in surgical cohorts by the included activity metrics is apparent here. There may be no one-size-fits-all activity metric, and the use of a combination of acceleration and step-based metrics may be most appropriate when evaluating activity in certain populations of interest.

Two case studies were selected to examine the potential role of physical activity monitoring at the individual participant level. In the first case study, the participant required an MUA within a month following the end of the physical activity monitoring period due to persistent stiffness of the operative knee and limited joint flexion. MUA is necessary in a small percentage of patients and is a common reason for readmission following knee arthroplasty[250]. This case is particularly interesting in that while there is a clear trend of below-average step count within the first four postoperative weeks, the participant's overall acceleration follows closely with the study mean trajectory throughout the monitoring period. Though step counting is susceptible to underestimation bias due to walking aids, independent walking is a key evaluation criterion at the Week 6 postoperative appointment. Therefore, as suggested in this case study, an underestimation of steps may point to prolonged use of a walking aid as an indication of reduced mobility, a trend identifiable through step counting but not overall acceleration. It is unclear why this participant had a sudden increase in physical activity in the last two monitoring weeks, however, for the first 4 weeks the participant remained well below the mean recovery curve in terms of step count. In a clinical implementation of perioperative activity monitoring, this patient would, ideally, trigger an alert for poor mobility, flagging an issue to the clinical care team.

The second case study involves a participant who seemed to be recovering well until a painful twisting incident that led to a downturn in physical activity. In this instance both step count and acceleration demonstrated similar patterns. ASTP was investigated and showed a sudden increase in activity fragmentation, associated with a decrease in participant mobility. For this participant, a monitoring system would be in place to encourage the participant during the time of good

recovery and initiate a system to check on the patient when the change in recovery trajectory was identified.

7.5 Conclusion

The results of the current study demonstrate a key area of opportunity in using physical activity monitoring to follow and assess patient recovery in the early postoperative period. Clear trends of increasing mobility and activity in the weeks following surgery, reproducibly using both step count and acceleration-based activity metrics are demonstrated in this patient population. The analysis presented demonstrates the feasibility of activity monitoring at both the individual and cohort level in the orthopaedic setting. Wearable-based monitoring is well-tolerated in the knee arthritis patient population and offers potential as a minimally burdensome tool for patient-centric monitoring or as an outcome measure for clinical trials. Summary activity metrics of acceleration and step count overlap in elucidating patient activity, but both may have distinct advantages in a patient population where ambulatory mobility is a targeted therapeutic goal. Quantifying both metrics may be helpful in distinguishing patient mobility from patient activity.

Chapter 8: Summary and Conclusions

The primary clinical conclusions from this thesis are as follows:

- Step count is an easily interpretable physical activity summary metric closely related to a patient's mobility. Within the UK Biobank and the clinical population, step count and step cadence were more sensitive to changes in participant mobility than acceleration-based metrics in this population of interest.
- Wearables can differentiate trajectories of recovery between individual patients in the early postoperative period. With further study clinical implementation of activity monitoring may have potential to identify participants at risk of a poor outcome, allowing the shifting of resources towards the patients who need them.
- In UKA and TKA patient cohorts of similar baseline activity levels, UKA results in faster recovery in terms of objectively measured early postoperative patient mobility and physical activity.
- True physical activity is an aspect of the patient recovery not well correlated with PROMs. Wearable monitoring is not a replacement for PROMs but offers an opportunity to gain insight into another facet of recovery that is equally important to the overall patient experience.

8.1 Summary of Findings

Following the introduction of background literature, this thesis began in Chapter 2 with a scoping review that identified a small yet growing trend in the use of wearable sensors for evaluation of knee arthroplasty patient outcomes. Three primary areas of research interest where wearable sensors are currently imbedded include functional analysis, physical activity monitoring, and the study of joint instability. Studies incorporating wearable sensors into functional analysis most frequently utilised two or more inertial measurement units (IMU), collecting accelerometer, magnetometer, and gyroscopic data over short periods of time through specified activities or movements. The use of wearables for the study of joint instability is a small research area, but one for which IMUs are well suited. Accelerometer-based physical activity monitoring is a growing research area with good potential for incorporation into routine clinical care through extended longitudinal patient monitoring. The review identified 26 studies monitoring physical activity in knee arthroplasty patients. Study protocols varied, negatively impacting potential for interstudy comparability. Most activity studies of knee arthroplasty patients incorporated monitoring windows of seven or fewer days, at six weeks or six months following surgery. A significant opportunity was identified for the integration of wrist-based accelerometry as a minimally-disruptive clinical tool for monitoring early postoperative recovery of mobility following knee arthroplasty. The results of this scoping review informed the subsequent chapters of this thesis, identifying a need for:

1) Sampling rate validation for enhanced interstudy comparability,
2) Improved validation of wrist-based step counting metrics,

3) The establishment of reference cohorts for end-stage knee arthritis and knee arthroplasty directly comparable to large scale population-based physical activity studies, and
4) Assessment of longitudinal physical activity monitoring in the clinical orthopaedic setting during the early postoperative period.

In Chapter 3, a previously undescribed measurement bias associated with accelerometer sampling rate was profiled in the popular Axivity AX3 device. In the collection of wearable sensor data during physical activity studies, sampling rate is an often-overlooked methodological decision that directly influences the allowable monitoring window. While wearable sensor technology is continually improving, with devices incorporating more storage and longer battery life, reduced sampling rate remains a key methodological feature for extending data collection opportunities in clinical populations with transient longitudinal activity profiles. Through 24-hour monitoring of 54 enrolled participants, data transformations for acceleration metrics and machine learning activity classifications were established between data collected at 25 Hz and 100 Hz. Applying these transformations, a reduced sampling rate can be used for valid interstudy comparison with more commonly employed 100 Hz protocols.

In Chapter 4, a ground truth validated, wrist-based step count algorithm integrating machine learning activity classification was developed and demonstrated superior accuracy when compared with traditional peak detection step counting algorithms. In physical activity epidemiology, step counting is often bypassed in preference to validated acceleration-based activity metrics. Step count is a popular and intuitive activity metric in personal activity tracking; however, true step counts are difficult to accurately quantify. Wrist-based accelerometry is associated with high levels of

participant compliance, and wrist-based activity tracking is the preferred body placement for consumer products. As such, wrist-based step counting is an important and desired physical activity measure for many patients. The work presented in Chapter 4 was conducted to generate an improved wrist-based step counting algorithm, while creating an open-access, ground-truth labelled dataset for future step-based algorithm validation. Using a cohort of 39 participants ranging in age from 19 to 81 years, a machine learning-based step counting algorithm was created. Using a random forest binary walking classification model in conjunction with heuristic peak detection, the developed algorithm predicted step count with a mean bias of -0.1% and a root mean square error (RMSE) of 13.1% in the free-living environment. In external validation, the model resulted in a 0.8% mean step count bias with a 5.1% MAPE in an independent dataset collected using a device from a different manufacturer.

Following the development of the machine learning step counting algorithm, wrist-based step count was calculated for 96,673 UK Biobank participants, including over 7,009 participants with a history of knee or hip arthroplasty. As a continuation of Chapter 4, Chapter 5 was designed to: 1) assess the face validity of step based physical activity metrics in comparison with acceleration-based activity in a well-established activity cohort, and 2) to establish reference datasets comparing physical activity of non-arthritic persons and those with either end-stage lower limb osteoarthritis or who have fully recovered from unilateral lower limb joint arthroplasty. Within the UK Biobank physical activity cohort, daily step count was highly correlated (0.74) with overall acceleration in all participants. Age and sex-related trends in acceleration-based activity were preserved in step count analyses, while step cadence closely followed trends in self-reported

walking pace. In a matched analysis with non-arthritic controls, 11% (p=0.019) and 9% (p=0.011) lower daily step counts were observed in participants with end stage knee, and hip osteoarthritis, respectfully. No difference in overall acceleration was observed, suggesting a unique sensitivity to mobility when using step-based metrics in this population. In participants greater than one year post-arthroplasty, no significant difference in daily step count or overall acceleration was observed between non-arthritic controls and either knee or hip arthroplasty patients.

The implementation of longitudinal physical activity monitoring of knee arthroplasty patients in the first six weeks of postoperative recovery are presented in Chapter 6 and Chapter 7. In these chapters, patient monitoring with a wrist-based wearable sensor was well tolerated by the patient population, with enrolment and valid baseline data collection from 68% of eligible patients approached during recruitment. Of those participants contributing baseline data, only 12% elected not to continue with any postoperative monitoring. A patient and participant involvement group corroborated high acceptability of perioperative wearable-based monitoring, indicating a desire for feedback regarding their recovery progress. Physical activity in this participant cohort was lower than UK Biobank baseline levels, taking a median 6855 steps per day with an overall acceleration of 21.5 m*g* in the weeks prior to surgery. Moderate correlation between preoperative self-reported and objectively measured physical activity was detected; however, no correlation was observed between patient reported outcome measures (PROMs) and objectively measured activity.

As reported in Chapter 7, a clear trend in overall activity was demonstrated in the first six postoperative weeks. Strong correlation was exhibited between Week 1 and Week 6 activity, with correlation coefficients of 0.61 and 0.58 between Week 1

and Week 6 acceleration and step counts, respectively. Patient age, use of walking aids, postoperative day number, preoperative VAS and preoperative activity levels were significantly associated with physical activity in mixed-models regressions. When used as an outcome measure, wearable-based monitoring was sensitive to different recovery trajectories for patients with total (TKA) and unicompartmental knee arthroplasty (UKA). While baseline activity levels were equivalent in the two surgical cohorts, patients undergoing UKA had a faster recovery and higher overall physical activity than the TKA cohort.

8.2 Study Limitations

There are limitations to the methods used in this work. First, non-clinical analyses in Chapter 3 and Chapter 4 were conducted with data collected from a healthy sample of convenience. As a result, the population demographics are significantly different than a typical clinical knee arthroplasty patient population. A convenience sample was selected due to the expedited ethical procedures and increased number of potential participants compared to a clinical recruitment pool. For both studies, care was taken to actively recruit a diversely aged study group to broaden variability in gait and activity patterns. While work investigating differences in activity metrics based on sampling rate should be similar regardless of participant demographics, walking classification and step detection are likely to be more sensitive to differences in gait dynamics in older populations. Further work is needed to confirm step detection accuracy in a clinical orthopaedic population. Similarly, step counts and associated cadence are known to be sensitive to walking aids, with underreporting of step count in the slowest and most aided gait patterns. However, as a result of some potential walking aid bias, step count metrics likely demonstrate changes in overall mobility that may not be as readily

apparent in acceleration-based metrics. An increase in postoperative step count, therefore, is likely to reflect both an increase in overall activity as well as a decreased reliance on walking aids. Machine learning activity classifications of light activity, sedentary activity, moderate-to-vigorous activity, and sleep were also calculated using a model trained on a healthy population. It is unclear how well these classification models generalise to an osteoarthritic population and should be appropriately retrained prior to implementation as the primary activity metrics within an epidemiological study of older, less active patient participants.

Additional limitations associated with the COVID-19 pandemic must be acknowledged with respect to the clinical study presented in Chapter 6 and Chapter 7. During this study recruitment, the United Kingdom entered several iterations of government regulations restricting normal movement and social interaction. At some timepoints, restrictions were highly variable throughout the country due to a tier system implemented in late 2020. As a result, patient activity and behaviour could be differentially affected during monitoring periods. Further, the addition of patient contact limitations in the clinical setting resulted in the need to send postoperative monitors by post instead of delivering directly following surgery. This led to high levels of missing data from Day 0 to Day 4 and may affect the interpretation of data in the first postoperative week.

8.3 Clinical Implementation and Future Directions

This thesis demonstrates the potential that wearable sensors offer for improved patient care following knee arthroplasty; however, further work is needed to fully establish the utility of a wearable-based perioperative program monitoring the recovery of physical activity and mobility. The following section outlines an overview of short-term and long-term next steps for continuing the current line of

research toward the integration of wearable-based physical activity monitoring into clinical practice.

8.3.1 Near-term Next Steps

The near-term next step in this study is to collect additional follow-up data from the enrolled study population to evaluate the association between early activity and long-term rehabilitation and recovery. The collection of PROMs alongside an additional week of objectively measured physical activity at 12 months following surgery is planned as a continuation of the current study. Because the scope of the clinical study reported in Chapter 6 and Chapter 7 was focused on short-term perioperative recovery, the implications of early physical activity recovery patterns on long-term outcomes remain broadly unknown. An extension of this study into the first postoperative year will add significant insight into the ability to anticipate 1-year postoperative outcomes based on the trajectory of physical activity recovery in the early postoperative period.

8.3.2 Clinical Implementation

The primary step towards clinical implementation of this work will be to initiate a multicentre study wherein a large and diverse group of patients is actively monitored and evaluated, with appropriate intervention delivered for individuals who are identified as at-risk for a poor clinical outcome. Two primary questions must be addressed in the proposed study to enable successful implementation of a perioperative wearable-based system for the recovery of physical activity: 1) What is the appropriate intervention associated with a wearable-based physical activity monitoring system, and 2) How long of a monitoring window is needed for efficient and cost-effective assessment of recovery?

Firstly, the specific intervention that should come through postoperative activity monitoring must be identified. Prior wearable-based passive coaching programs have primarily focused on to the distribution of app-based physiotherapy advice or workout routines[158,251,252]. Unfortunately, the largest of these coaching-only programs have demonstrated little success in terms of improving functional outcomes[158,251]. The lack of functional improvement from these wearable-based coaching programs point to the potential need for a more dynamic system wherein patients are actively monitored as a safety net, and an appropriate intervention is given when a need is identified. Following the work of this thesis, a key to the success of an intervention-based program will be the successful early detection of patients who are falling away from their expected trajectory of recovery, and who are struggling to build towards their preoperative levels of mobility.

Secondly, and perhaps most crucially, further data is needed to adequately assess the cost-benefit trade-off between postoperative monitoring limited to the first 1-2 postoperative weeks versus a full 6-week, or more, monitoring protocol. In a cost minimising healthcare system like the United Kingdom's National Health Service, patients may feel disconnected or alone in their recovery, particularly with a lack of routine contact with surgical and care team between weeks two and six following surgery. A wearable-based program in which their recovery is being monitored, even if only by an algorithm, could support patients during the period of their recovery with otherwise limited contact with their clinicians. An important aspect of this system would be to facilitate improved patient-clinician communication, while avoiding additional burden on the clinical care team. Establishing the combined role of coaching, resource signposting, activity monitoring, risk flagging, and improved communication between the patient and

clinical team will be critical to developing a robust and effective wearable-based monitoring system.

In contrast to continuous 6-week monitoring, there may be justification to integrate limited wearable-based monitoring into a shorter monitoring window during the immediate postoperative period. Currently at the Nuffield Orthopaedic Centre, patients are routinely discharged from the hospital on the same day as their surgery. Following surgical discharge, patients are not seen again by the clinical care team until an appointment with a physiotherapist 7 days later. A potential area of high value for the use of wearables would be to inform the physiotherapy care team of at-home activity levels undertaken by the patient in the first postoperative week. If the strong correlation between Week 1 and Week 6 physical activity (Section 7.3.3.5) can be confirmed in a broader study, the use of a wearable for all patients during the first days of recovery may help identify a subset of patients at risk for slow progression and in need of additional monitoring. With a clinical implementation focused on monitoring only in the first postoperative week, most patients would be identified as not needing close monitoring and patients would be expected to progress normally until their final routine clinical evaluation following the sixth postoperative week. Subsequently, monitoring and intervention resources would be focused only on those patients at highest identified risk for developing associated complications or having a poor long-term outcome.

Further to the question of the timing of patient monitoring, aims of this larger study should more fully investigate the need for preoperative activity data collection. As presented in Chapter 7, preoperative baseline activity was more highly correlated with postoperative activity than any other measured factor. While this week of

preoperative monitoring may be highly predictive of postoperative activity levels, several disadvantages exist when relying on the collection of preoperative data. As demonstrated in Chapter 6, a significant proportion of patients do not progress from preoperative clinical assessment to surgery. The collection of unused preoperative data is thus at high risk for unnecessary waste due to surgery cancellations or extended delays. The preoperative distribution of a wearable sensor introduces additional logistical issues in terms of data collection and device recovery. If the burden of preoperative monitoring can be eliminated without sacrificing postoperative effectiveness, the efficiency and ease of clinical activity monitoring is significantly increased.

8.3.3 Other Considerations

8.3.3.1 Data Logging

An important logistical aspect of integrating activity monitoring into the clinical care pathway is the comparison of device-logged data versus real-time monitored data. In a scenario where only data during the first postoperative week is collected, data logging on the wearable device without regular patient-to-clinic data transfer is ideal. Storage of data onboard the device removes the need for a linked smartphone or internet-enabled hub for remote data transfer and bypasses any associated security and privacy concerns. Logged data would be downloaded onsite and immediately available to the clinical care team for evaluation during the routine physiotherapy appointment on postoperative day seven. Conversely, to facilitate continuous monitoring and feedback, near real-time remote data transfer would be required. Cloud-based data transfer enables regular clinical monitoring and feedback, but is more expensive than onboard data storage, and comes with the additional considerations of security, device design and connectivity. Further

study into the optimal monitoring window and patient feedback system will dictate the final device parameters and data logging or transfer requirements.

8.3.3.2 Comparison to Novel PROMs

Additional investigation should be done to correlate early physical activity with the recently developed Oxford Early Recovery Score (OERS)[5,19]. This new PROM is specifically designed for the evaluation of patient recovery at postoperative Week 6, in contrast to other orthopaedic PROMs which are validated for use at 6+ months following surgery. Because PROMs and objectively measured activity tend to reflect different aspects of the patient experience, describing the correlation between physical activity recovery and the OERS will further improve understanding of short-term postoperative outcomes. When deployed collectively, these two measures may provide substantial insight across several domains of patient recovery and could be used in conjunction to evaluate and optimize perioperative care. To ensure clinical usefulness of these early postoperative measures, it will be important to associate both metrics with long-term clinical outcomes.

8.3.4 Broadened Clinical Application

In addition to the direct continuation of the current study into the clinical care pathway following knee arthroplasty, work can also be done to expand the utility of perioperative activity monitoring beyond adult joint reconstruction. A natural progression of this thesis would be the evaluation of physical activity recovery after hip arthroplasty and in sports medicine applications such as reconstruction of the anterior cruciate ligament and recovery following ankle fracture. Even more broadly, wearable-based physical activity monitoring is ideally suited for monitoring long-term conditions with transient patient activity profiles. A prime

example of potential non-orthopaedic implementation physical activity monitoring is in the management of myelodysplastic syndromes wherein patients require regular blood transfusions to sustain energy levels and overall quality of life[253]. Continuous activity monitoring of this clinical population could aid decision-making regarding the necessary timing for blood transfusions at the cohort or individual patient level. Recent studies conducted elsewhere monitor deterioration of physical activity associated with degenerative and terminal clinical conditions[77]. In these instances, wearable-based physical activity monitoring may help inform decisions for treatment and palliative care based on level of sustained decline in patient mobility and activity.

8.4 Conclusions

In summary, this thesis has established a framework for the basis of clinically-implemented physical activity monitoring as a tool for assessing postoperative recovery in an orthopaedic patient population. This work has expanded methodology for valid interstudy comparison of accelerometer-based measurements while demonstrating the utility of wrist-based step counting as a metric for quantifying mobility in the knee arthroplasty patient population. The clinical implementation of activity monitoring in this study reveals the potential for valid and impactful accelerometer-based assessment of postoperative recovery at the individual patient level, and additionally, at the cohort level for evaluation of medical devices and clinical pathways.

Appendix

A.1 Postoperative Activity Diary

Participant ID: _____
Device ID: _____

Special Activity and Wear-time Diary -- Recovery Week 1

Please complete the diary each day to record your activity and wrist monitor wear time.

Please enter the date for each day (Day/Month/Year)	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
At what times, if any, did you take an extended walk?							
At what times, if any, did you ride a bicycle?							
Was there any time in the past 24 hours that you removed your activity monitor?							
Using the pain scale provided (0-10), how would you rate your average knee pain over the past 24 hours?							

Physical Activity Levels During Recovery Following Knee Arthroplasty
Chief Investigator: Scott Small

Daily Diary v1.3 27/01/2020
IRAS: 260855 REC: 19/SW/0151
Page 1 of 7

Figure A-1: Example first page from the postoperative diary given to participants following surgery.

A.2 Exploratory Cluster Analysis

A.2.1 Methods

An exploratory cluster analysis was performed using time series data from postoperative step count, relative-to-baseline step count, acceleration, and relative-to-baseline acceleration. K-means clustering was performed, with preliminary clustering iterated over 2 to 9 clusters with within-cluster sum of squares and silhouette widths calculated. The number of clusters for each of the four cluster analyses was selected while balancing considerations for inflection points observed on scree plots, maximizing silhouette distance, and creating enough clusters to distinguish the expected relatively small groups likely to clinically present as unsatisfied with their surgical intervention (potentially up to 20%[238]). Following clustering, demographic data and pre/postoperative self-reported PROMs and activity were reported for all clusters.

A.2.2 Results

Quantile regression curves for median daily step count of clusters are presented in Figure A-2. Four clusters were selected in postoperative step count based on consideration of total within-cluster sum of squares and silhouette widths. Summary demographic data, self-reported activity, and PROMs for each cluster is presented in Table A-1. In this analysis the lowest activity cluster presents with the highest median age, highest proportion of total knee arthroplasty surgeries, and highest proportion of baseline walking aids needed. This cluster demonstrated a 6 point increase in OKS and 5 point increase in VAS following surgical intervention, compared to a 13 point increase in OKS and 22 point increase in VAS in the most active cluster.

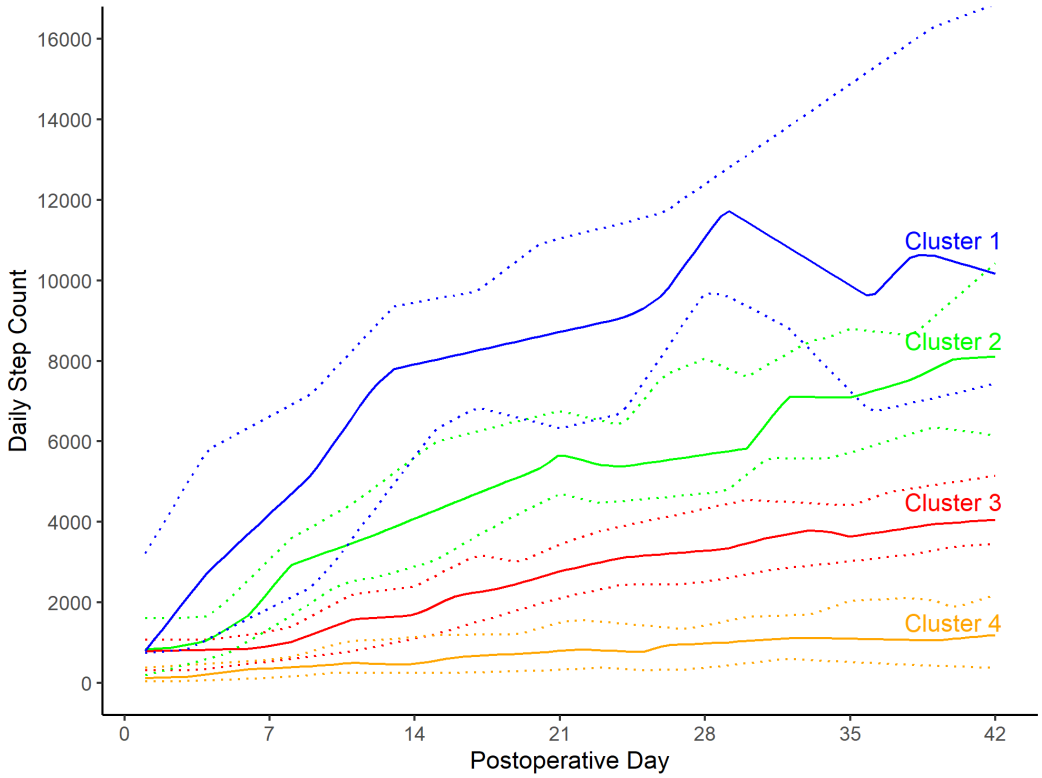


Figure A-2: Quantile regression plot of daily step count during the six weeks following knee arthroplasty for each cluster group, with solid line representing the median and dashed lines representing the lower and upper 25% quantiles.

Table A-1: Characteristics of Clusters Derived from Postoperative Step Count Time Series

	Cluster 1 (n=7)	Cluster 2 (n=25)	Cluster 3 (n = 31)	Cluster 4 (n=32)
Age	65.7 [62.2 - 72.5]	65.0 [58.7 - 70.4]	71.8 [61.3 - 74.9]	72.5 [64.5 - 80.3]
BMI	27.4 [22.1 - 31.9]	31.1 [26.7 - 33.6]	29.6 [26.9 - 32.8]	29.4 [26.2 - 34.2]
Sex				
Female	4	13	16	17
Male	3	12	15	15
Surgery				
TKA	1	9	13	19
UKA	6	16	18	13
Baseline Walking Aid				
Aided	0	0	8	13
Unaided	7	25	23	19
Baseline Median Daily Steps	12,805 [9,878 – 16,077]	8,582 [7,954 – 10,178]	8,160 [5,340 – 9,812]	3,522 [1,729 – 5,422]
OKS				
Preop	24 [21 - 30]	26 [19 - 28]	21 [13 - 28]	23 [16 - 28]
Postop	37 [34 - 42]	31 [28 - 34]	29 [25 - 33]	29 [23 - 35]
EQ-5D				
Preop	0.569 [0.566 - 0.602]	0.598 [0.388 - 0.659]	0.596 [0.294 - 0.659]	0.503 [0.349 - 0.659]
Postop	0.853 [0.762-1.000]	0.698 [0.659 - 0.761]	0.667 [0.596 - 0.730]	0.659 [0.503 - 0.761]
IPAQ MET-Min				
Preop	2,472 [1,332 – 3,083]	1,497 [1,026 – 2,571]	1,497 [687 – 3,765]	920 [264 – 2,070]
Postop	1,386 [891 – 2,064]	852 [693 - 1550]	1,173 [568 – 1,765]	957 [314 – 1,609]
UCLA				
Preop	5 [4 - 7]	5 [4 - 6]	5 [4 - 6]	4 [3 -5]
Postop	6 [4 - 6]	4 [4 - 6]	4 [4 - 6]	4 [3 -5]
VAS				
Preop	63 [55-69]	65 [50 - 70]	70 [61 - 81]	70 [60 - 70]
Postop	85 [75 - 88]	80 [65 - 88]	83 [70 - 87]	75 [60 - 81]

All values presented as median [interquartile range]

Time series cluster groups for postoperative step count relative to baseline levels are presented in Figure A-3, with summary demographics and self-reported activity and PROMs for each cluster presented in Table A-2. Five clusters were created when clustering step counts relative to baseline step counts, including one single-participant cluster. In this analysis, the cluster with lowest postoperative activity relative to baseline contained the highest proportion of total knee arthroplasty surgeries and gained a single point in postoperative OKS compared to gains of 8 to 10 points in the other clusters. This cluster had a median age of 65 years, a 7.5 year younger cluster than the least active cluster based on absolute step count, indicating this as a group of younger participants struggling to regain preoperative activity levels.

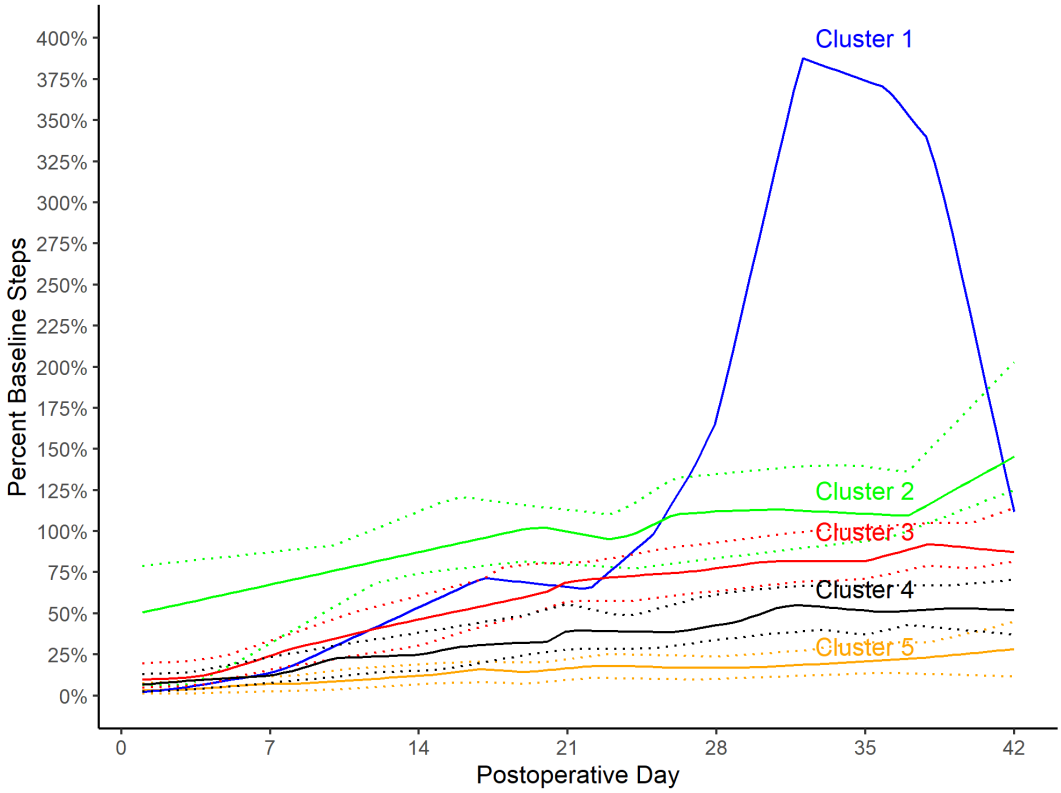


Figure A-3: Quantile regression plot of percent daily step count relative to baseline during the six weeks following knee arthroplasty for each cluster group, with solid line representing the median and dashed lines representing the lower and upper 25% quantiles.

Table A-2: Characteristics of Clusters Derived from Postoperative Relative-to-Baseline Step Count Time Series

	Cluster 1 (n=1)	Cluster 2 (n=8)	Cluster 3 (n=26)	Cluster 4 (n=33)	Cluster 5 (n=27)
Age	*80+	62.5 [57.9 - 72.7]	68.9 [65.5 - 73.6]	71.5 [60.1 - 77.3]	65.0 [60.8 - 75.2]
BMI	30	31.8 [31.1 - 36.4]	29.0 [26.6 - 33.6]	29.7 [26.7 - 33.6]	28.1 [26.0 - 32.6]
Sex					
Female	*	4	15	21	10
Male	*	4	11	12	17
Surgery					
TKA	0	4	5	16	17
UKA	1	4	21	17	10
Baseline Walking Aid					
Aided	1	1	4	7	8
Unaided	0	7	22	26	19
Baseline Median Daily Steps	709	7,086 [4,848 – 7,514]	6,998 [4,632 – 8,685]	7,868 [3,821 – 10,138]	7,233 [5,130 – 11,112]
OKS					
Preop	34	23 [17 - 28]	23 [12 - 30]	22 [14 - 28]	25 [20 - 29]
Postop	42	33 [28 - 38]	31 [26 - 34]	31 [27 - 36]	26 [21 - 34]
EQ-5D					
Preop	0.659	0.627 [0.385 - 0.674]	0.566 [0.307 - 0.659]	0.596 [0.294 - 0.659]	0.621 [0.372 - 0.659]
Postop	0.659	0.772 [0.710 - 1.000]	0.698 [0.596 - 0.761]	0.690 [0.659 - 0.761]	0.659 [0.534 - 0.746]
IPAQ MET-Min					
Preop	2236	1,139 [231 – 2,030]	1,293 [660 – 3,573]	1,413 [508 – 2,571]	1,044 [588 – 2,828]
Postop	2537	575 [396 – 1,808]	897 [462 – 2,038]	1,263 [696 – 1,638]	990 [588 – 1,760]
UCLA					
Preop	6	5 [4 - 6]	5 [4 - 6]	4 [4 - 6]	4 [3 - 6]
Postop	7	6 [5 - 6]	4 [4 - 6]	4 [4 - 5]	4 [3 - 6]
VAS					
Preop	90	66 [60 - 70]	61 [50 - 80]	70 [60 - 80]	70 [60 - 80]
Postop	92	80 [73 - 86]	80 [65 - 85]	80 [70 - 90]	75 [64 - 85]

** Redacted to preserve anonymity of single-participant cluster*
All values presented as median [interquartile range]

Quantile regression curves for five clusters based on daily acceleration time series clustering are presented in Figure A-4. Summary demographic data, self-reported activity, and PROMs for each cluster is presented in Table A-3. In this analysis the lowest two activity clusters present with the highest median age, with similarly high proportions of total knee arthroplasty and walking aid usage. The lowest acceleration cluster demonstrated a 4 point increase in OKS and a 13 point increase in VAS following surgical intervention, compared to an 8 point increase in OKS and 22 point increase in VAS in the most active cluster.

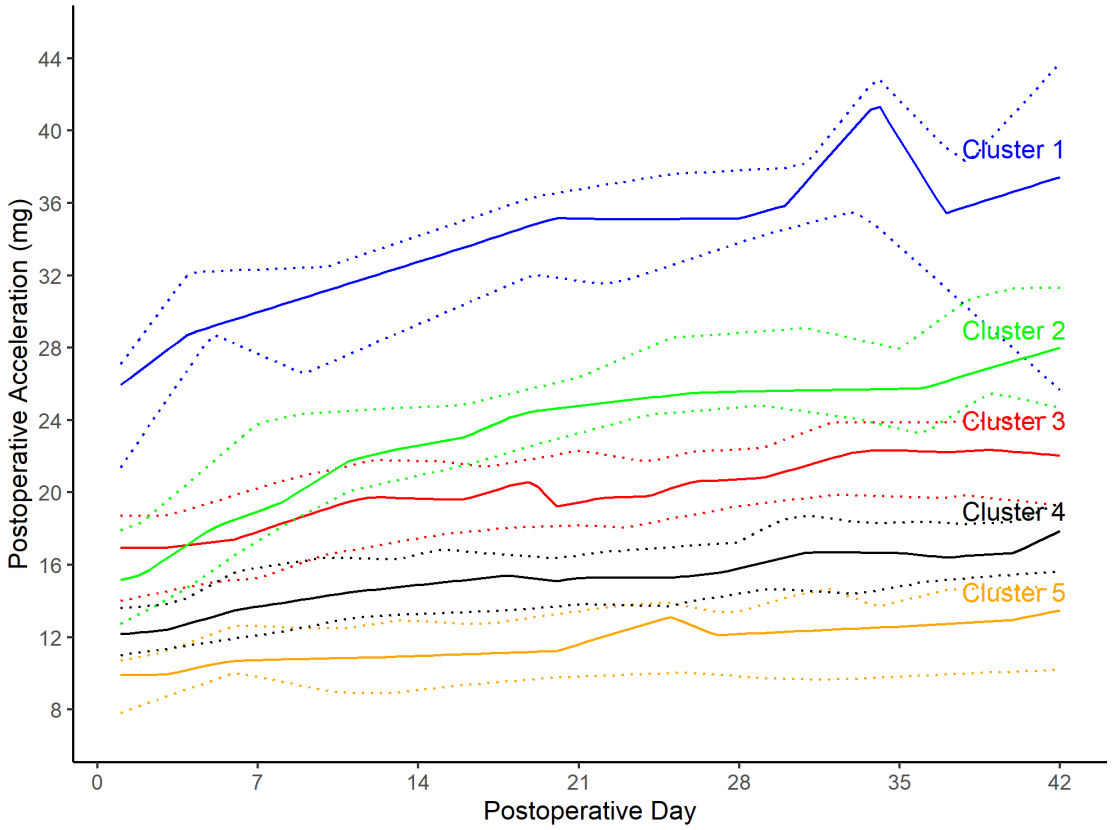


Figure A-4: Quantile regression plot of daily acceleration during the six weeks following knee arthroplasty for each cluster group, with solid line representing the median and dashed lines representing the lower and upper 25% quantiles.

Table A-3: Characteristics of Clusters Derived from Postoperative Acceleration Time Series

	Cluster 1 (n=3)	Cluster 2 (n=14)	Cluster 3 (n=30)	Cluster 4 (n=36)	Cluster 5 (n=12)
Age	65.7 [62.3 - 70.5]	63.5 [57.0 - 71.6]	65.9 [58.7 - 73.5]	73.7 [65.5 - 76.8]	73.5 [66.2 - 76.4]
BMI	21.4 [21.4 - 27.0]	29.4 [26.7 - 31.2]	30.1 [26.6 - 35.0]	29.6 [27.1 - 34.9]	30.1 [25.0 - 33.4]
Sex					
Female	3	10	15	17	5
Male	0	4	15	19	7
Surgery					
TKA	0	4	14	18	6
UKA	3	10	16	18	6
Baseline Walking Aid					
Aided	0	0	5	10	4
Unaided	3	14	25	24	8
Baseline Median Daily Acceleration (m*g*)	45.0 [41.5 – 48.3]	26.6 [25.7 – 33.5]	23.9 [21.8 – 28.2]	19.7 [17.2 – 23.0]	15.2 [13.8 – 18.2]
OKS					
Preop	29 [27 - 30]	24 [19 - 28]	25 [15 - 29]	22 [14 - 27]	24 [19 - 30]
Postop	37 [35 - 40]	31 [27 - 36]	31 [27 - 35]	29 [21 - 34]	28 [24 - 34]
EQ-5D					
Preop	0.566 [0.566 - 0.587]	0.627 [0.382 - 0.659]	0.597 [0.357 - 0.659]	0.522 [0.321 - 0.659]	0.581 [0.376 - 0.674]
Postop	0.853 [0.724 - .926]	0.730 [0.659 - 0.814]	0.710 [0.659 - 0.793]	0.659 [0.503 - 0.730]	0.659 [0.613 - 0.746]
IPAQ MET-Min					
Preop	2,860 [2,303 – 3,083]	1,442 [1,040 – 2,373]	1,386 [462 – 2,759]	1,629 [473 – 3,611]	627 [214 – 1,124]
Postop	1,386 [891 - 1542]	867 [462 – 1,732]	1,324 [693 – 1,748]	912 [337 – 2,058]	957 [378 – 1,638]
UCLA					
Preop	7 [6 - 8]	5 [4 - 6]	5 [4 - 6]	5 [3 - 6]	4 [3 - 5]
Postop	6 [5 - 6]	4 [4 - 5]	5 [4 - 6]	4 [3 - 5]	4 [3 - 5]
VAS					
Preop	63 [61 - 65]	60 [40 - 70]	70 [60 - 80]	74 [60 - 88]	64 [60 - 78]
Postop	85 [77 - 87]	73 [60 - 85]	82 [70 - 90]	78 [65 - 85]	77 [55 - 80]

All values presented as median [interquartile range]

Time series clusters for postoperative acceleration relative to baseline levels are presented in Figure A-5, with summary demographics and self-reported activity and PROMs for each cluster presented in Table A-4. Three clusters were created when clustering relative to baseline acceleration. In this analysis, the cluster with lowest postoperative activity relative to baseline contained only one female participant. This group had the highest median self-reported physical activity from the IPAQ questionnaire and the lowest proportion of baseline walking aid usage. This cluster presented a 5 point increase in median OKS and 12 point increase in VAS, compared to 9 and 6 point increase in OKS and 12 and 10 point increases in VAS for the clusters with the higher percentage of activity relative to baseline levels.

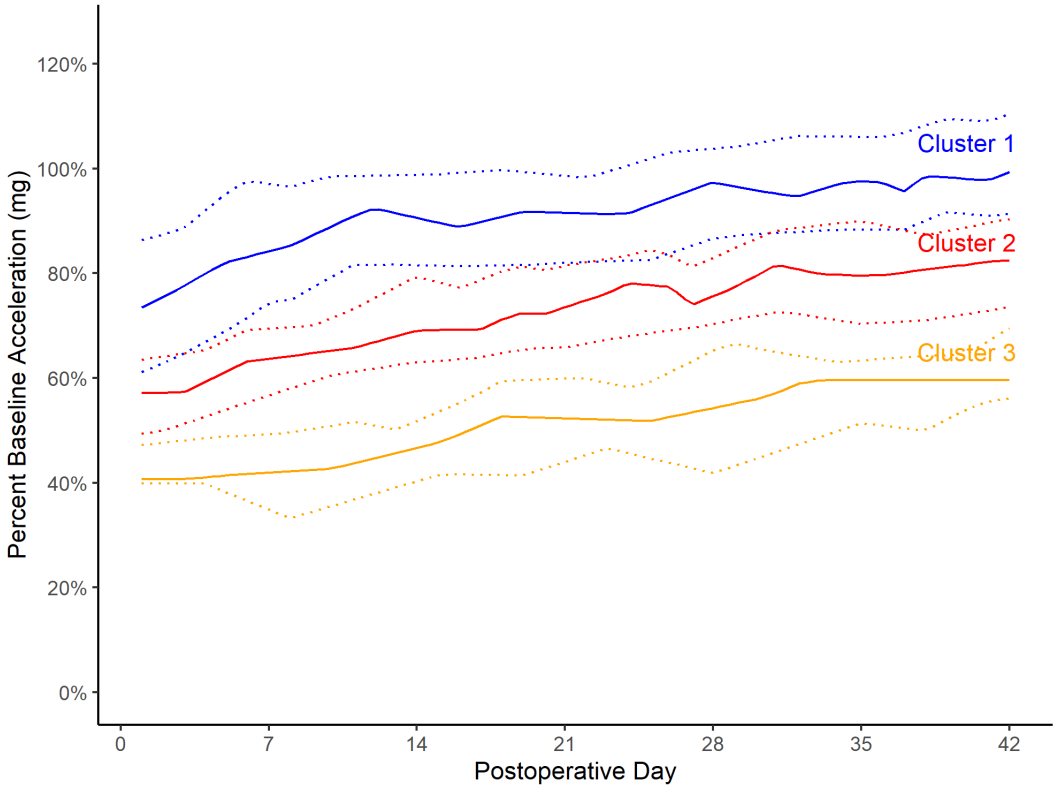


Figure A-5: Quantile regression plot of percent daily acceleration to baseline during the six weeks following knee arthroplasty for each cluster group, with solid line representing the median and dashed lines representing the lower and upper 25% quantiles.

Table A-4: Characteristics of Clusters Derived from Postoperative Relative-to-Baseline Acceleration Time Series

	Cluster 1 (n=44)	Cluster 2 (n=39)	Cluster 3 (n=12)
Age	72.8 [66.3 - 79.3]	65.0 [58.6 - 73.8]	64.8 [60.0 - 73.9]
BMI	29.6 [26.6 - 33.6]	30.2 [26.7 - 34.3]	27.7 [25.9 - 30.3]
Sex			
Female	27	22	1
Male	17	17	11
Surgery			
TKA	20	15	7
UKA	24	24	5
Baseline Walking Aid			
Aided	12	8	1
Unaided	32	31	11
Baseline Median Daily Acceleration (m*g*)	21.4 [17.1 – 23.8]	22.5 [19.2 – 29.1]	31.4 [26.0 – 33.5]
OKS			
Preop	22 [12 - 27]	25 [20 - 29]	23 [20 - 30]
Postop	31 [28 - 36]	31 [24 - 34]	28 [24 - 34]
EQ-5D			
Preop	0.568 [0.293 - 0.659]	0.608 [0.357 - .659]	0.621 [0.425 - 0.659]
Postop	0.694 [0.659 - 0.814]	0.659 [0.581 - 0.761]	0.659 [0.484 - 0.754]
IPAQ MET-Min			
Preop	1217 [485 - 2493]	1386 [685 - 2378]	3573 [640 - 5487]
Postop	846 [396 - 2348]	1272 [693 - 1716]	1077 [693 - 1542]
UCLA			
Preop	4 [3 - 5]	5 [4 - 7]	5 [4 - 7]
Postop	4 [4 - 6]	4 [3 - 5]	4 [4 - 6]
VAS			
Preop	68 [57 - 75]	70 [60 - 85]	68 [57 - 83]
Postop	80 [70 - 88]	80 [67 - 85]	80 [63 - 87]

All values presented as median [interquartile range]

A.2.3 Discussion

As demonstrated in the exploratory cluster analysis, substantially different clusters are created when investigating the least active patients in terms of absolute activity, and in those least active clusters when tracking activity relative to baseline levels. Wherein the least active clusters in terms of absolute activity tended to be older, less active preoperatively, and showed substantial improvement in PROMs, poorly recovering participants as clustered by relative-to-baseline activity were younger, previously more active, and demonstrated minimal change in PROMs. As PROMs and functional performance are well associated with overall patient satisfaction following knee arthroplasty, this current data strongly suggests that these patients are at high risk of an unsatisfactory outcome[254,255].

A.3 Additional Postoperative Plots

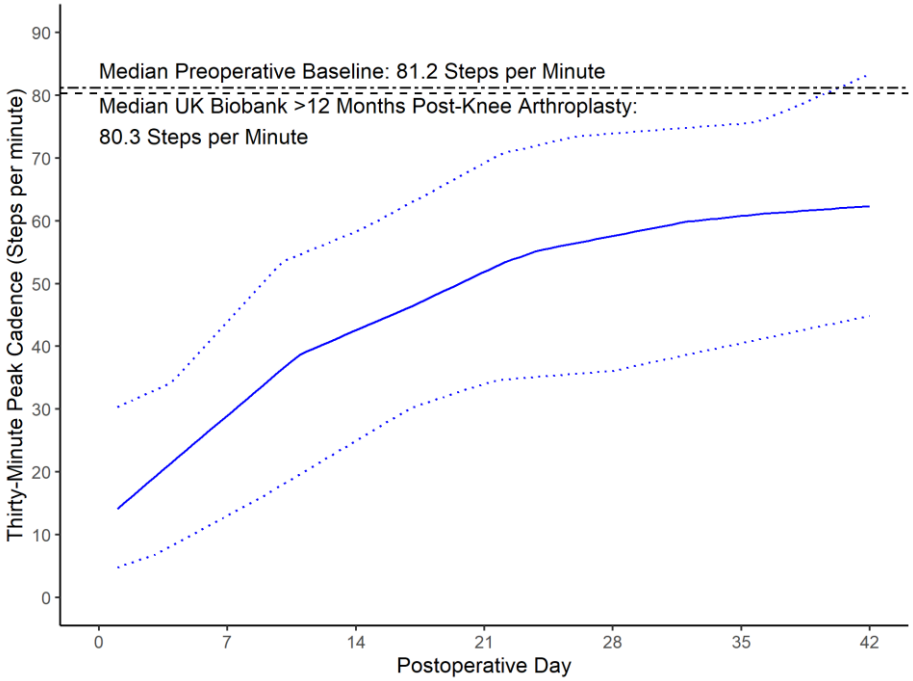


Figure A-6: Quantile regression plot of increasing peak 30-minute cadence during the six weeks following knee arthroplasty, with shaded solid line representing the study population median and the dashed lines representing the 25% and 75% quantiles.

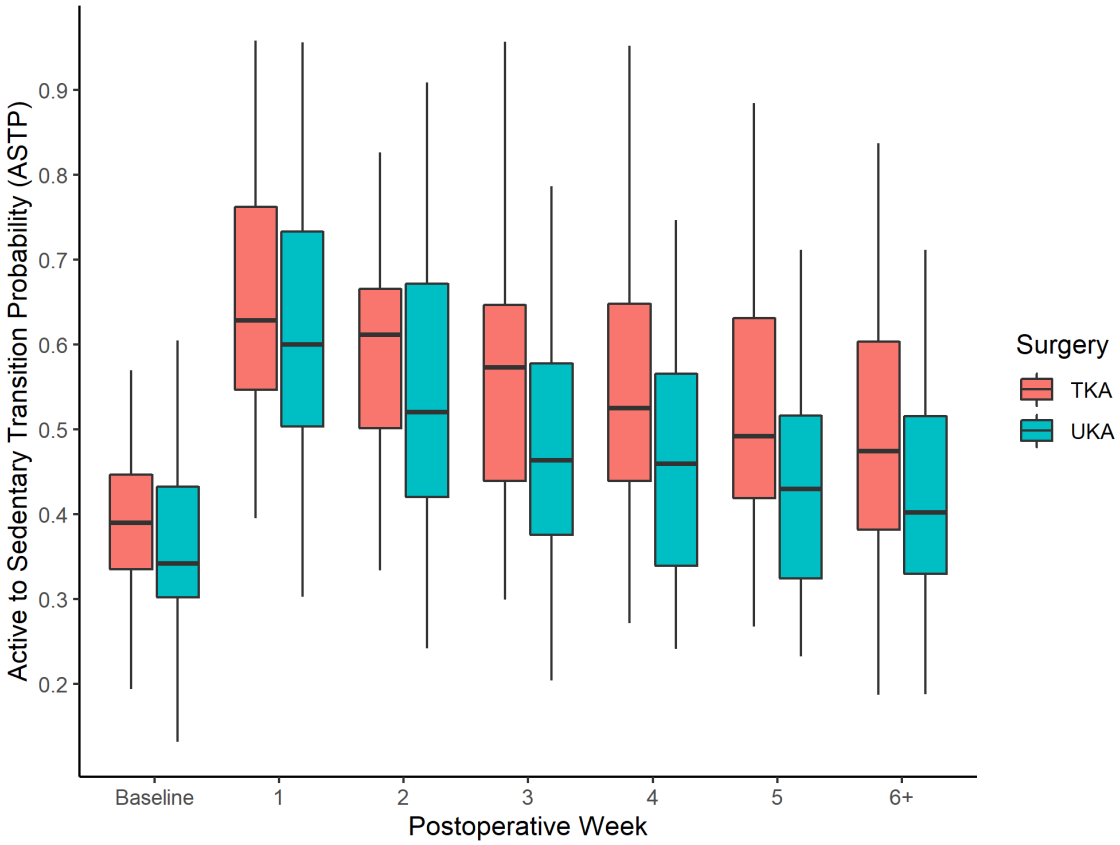


Figure A-2: Boxplot of active-to-sedentary probability (ASTP) by operative intervention compared with preoperative baseline in all participants contributing postoperative data.

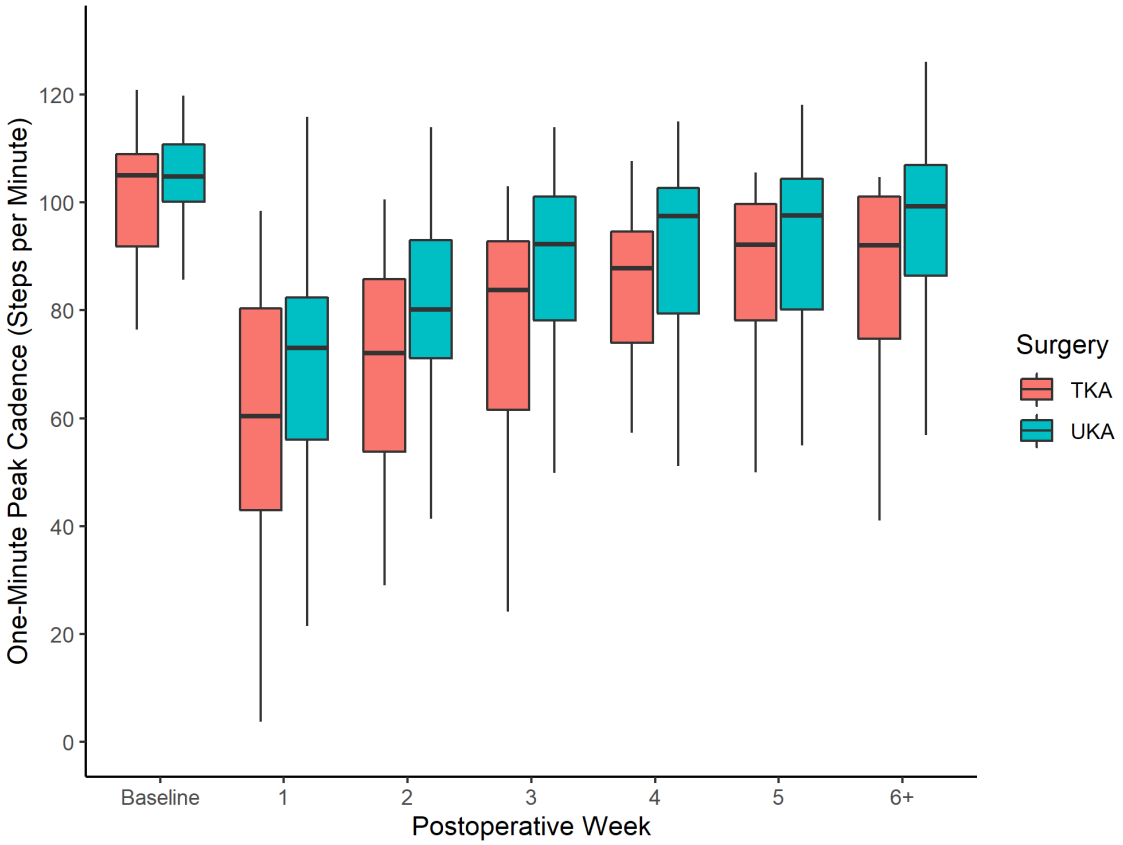


Figure A-3: Boxplot of peak 1-minute cadence by operative intervention compared with preoperative baseline in all participants contributing postoperative data.

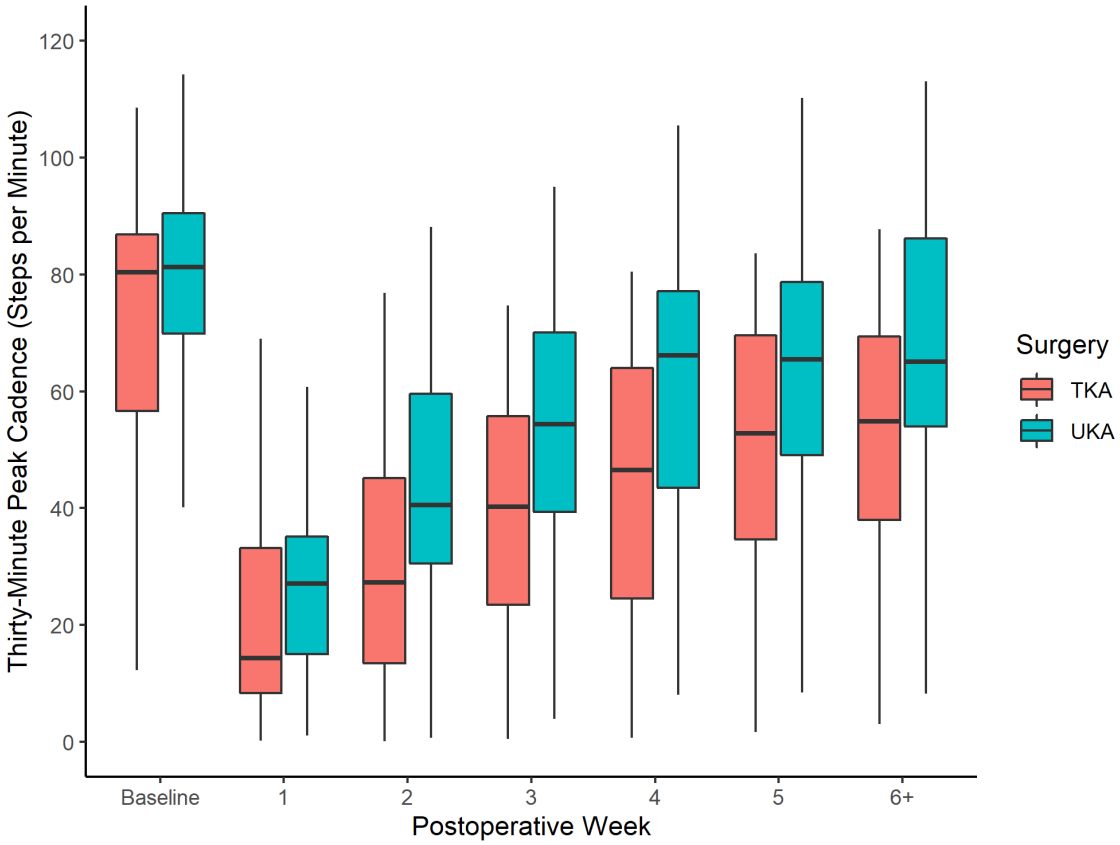


Figure A-4: Boxplot of peak 30-minute cadence by operative intervention compared with preoperative baseline in all participants contributing postoperative data.

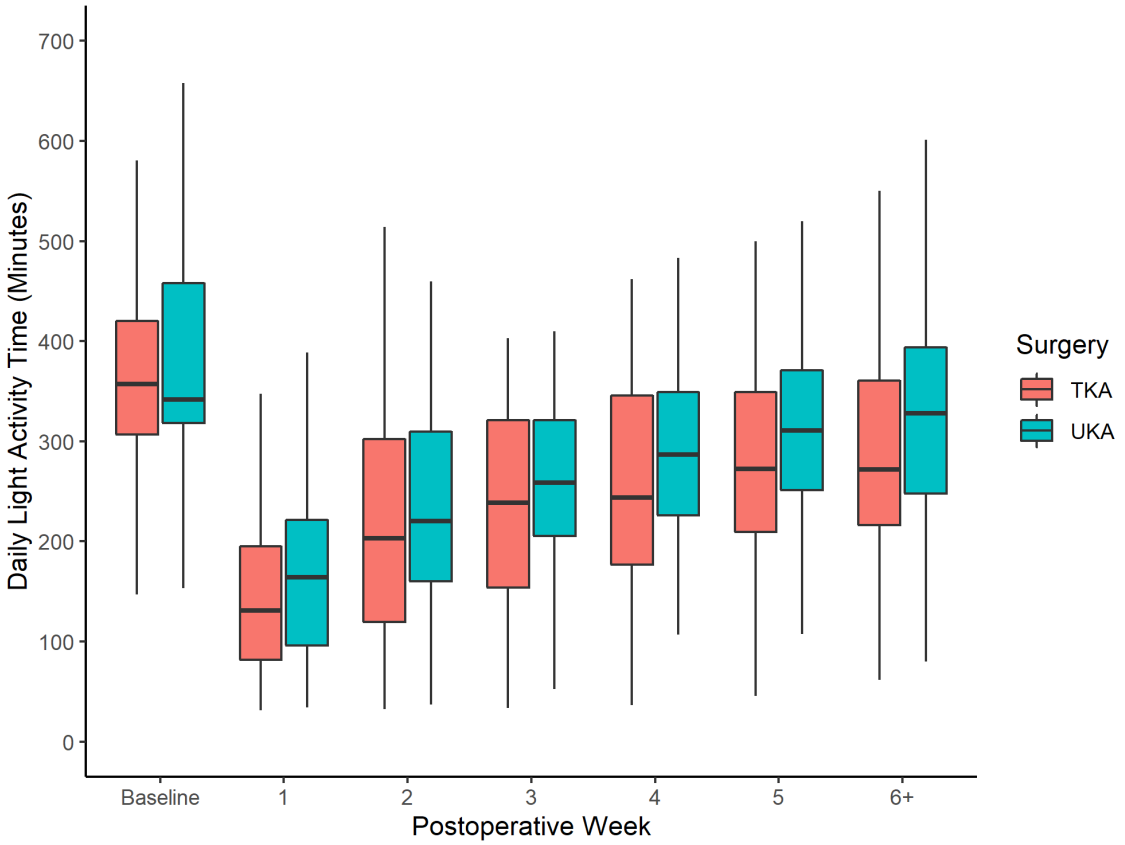


Figure A-5: Boxplot of daily time spent in light activity by operative intervention compared with preoperative baseline in all participants contributing postoperative data.

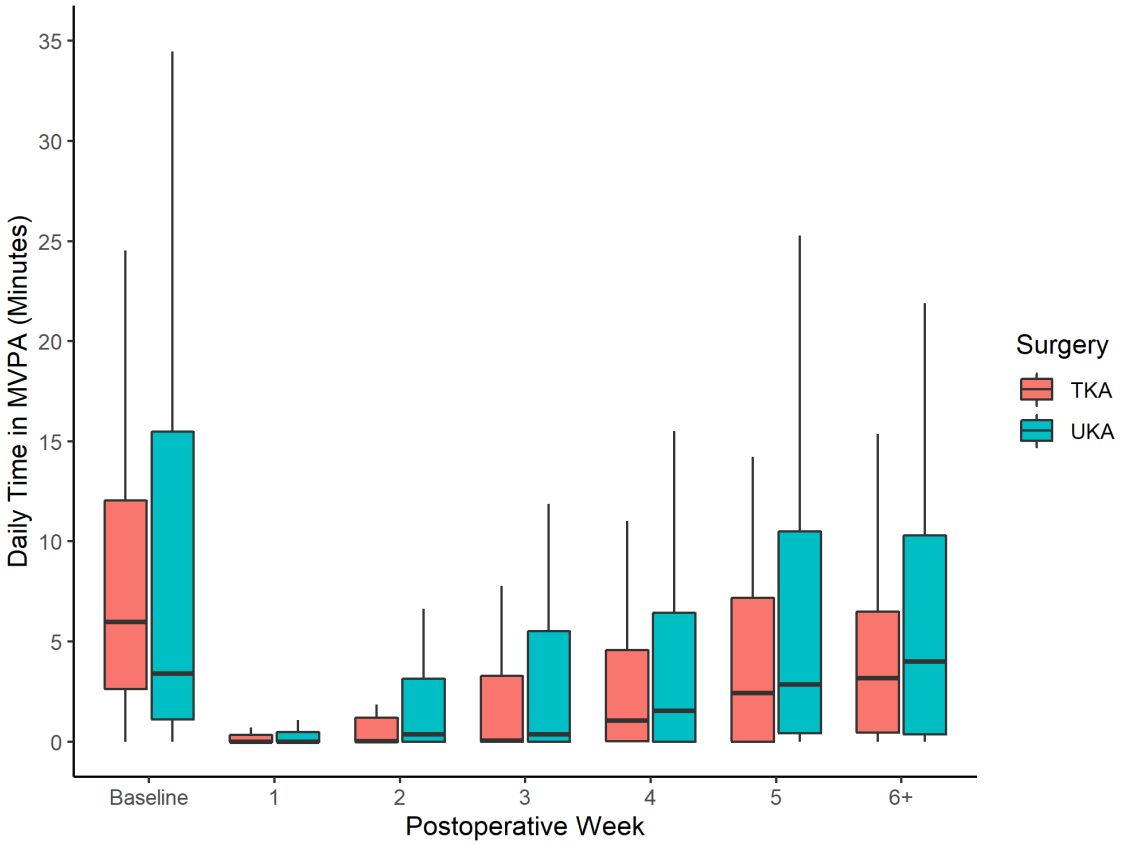


Figure A-6: Boxplot of daily time spent in moderate-to-vigorous activity (MVPA) by operative intervention compared with preoperative baseline in all participants contributing postoperative data.

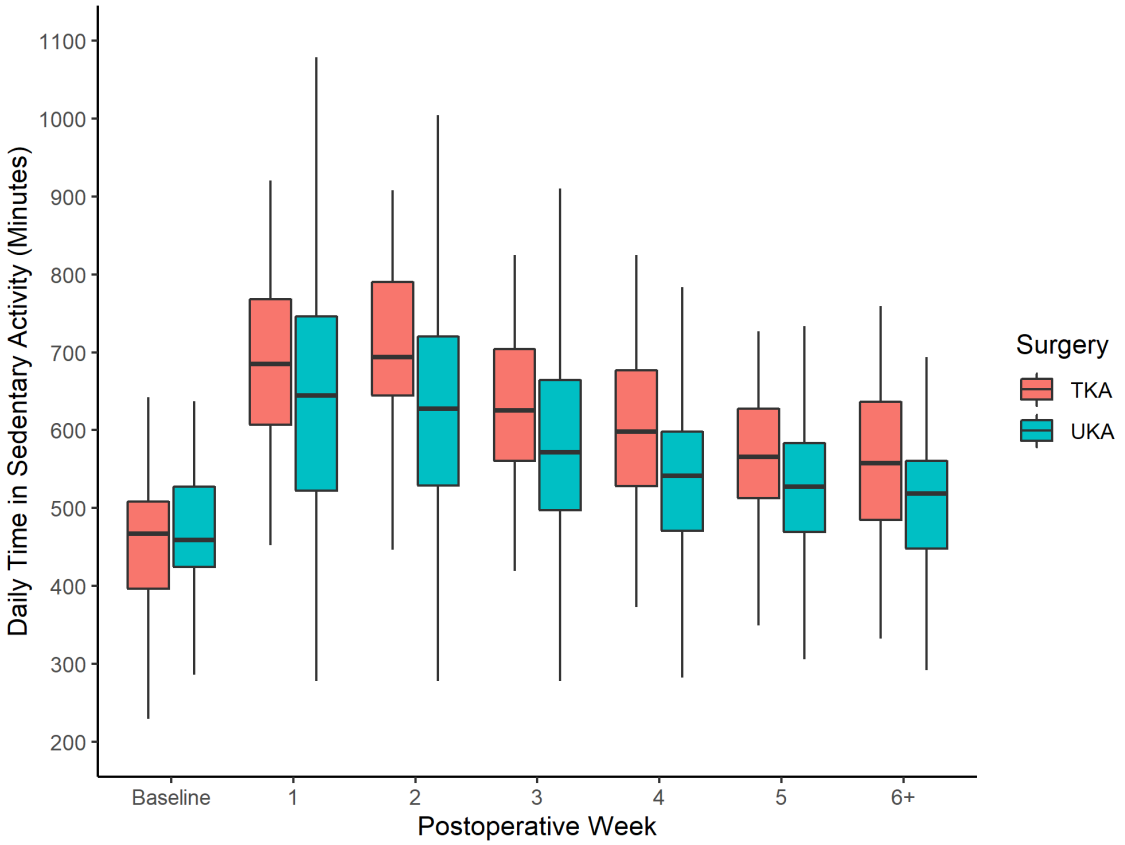


Figure A-7: Boxplot of daily time spent in sedentary activity by operative intervention compared with preoperative baseline in all participants contributing postoperative data.

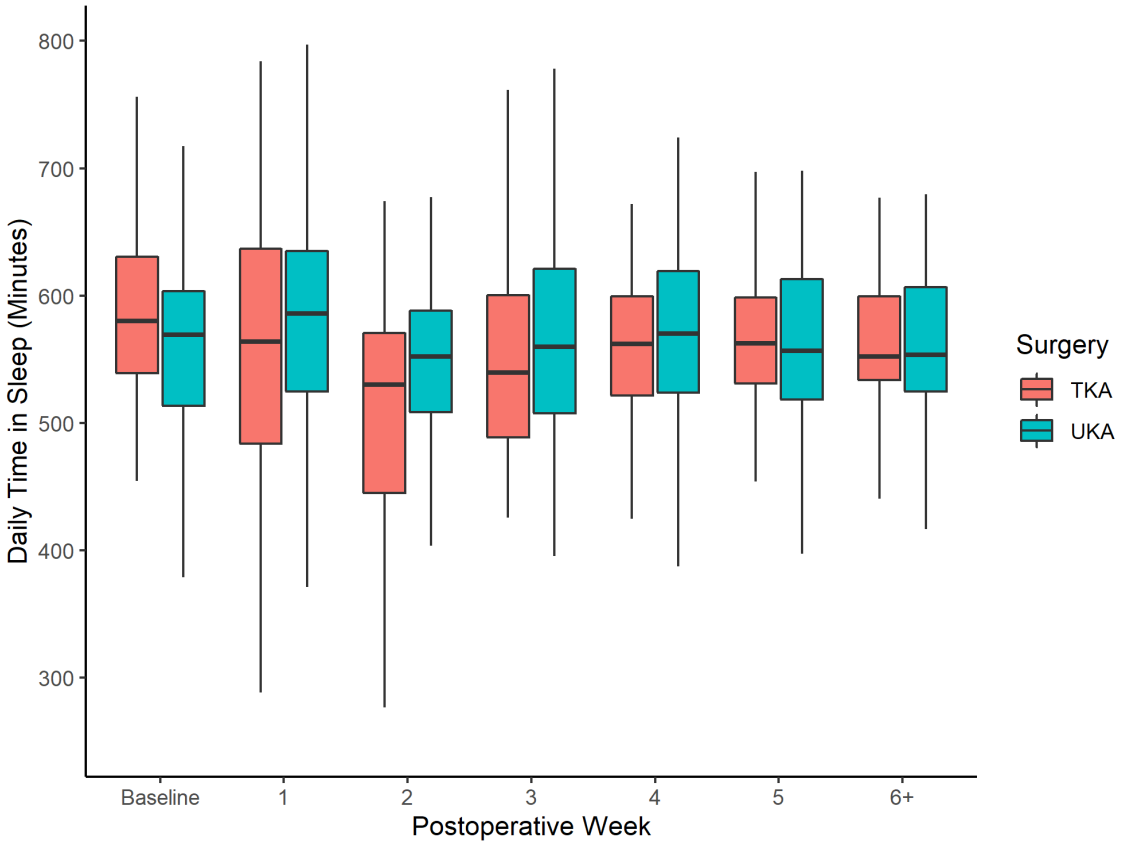


Figure A-13: Boxplot of daily time spent in sleep by operative intervention compared with preoperative baseline in all participants contributing postoperative data.

References

1. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. *Lancet*. 2019;393(10182):1745-1759. doi:10.1016/S0140-6736(19)30417-9
2. Silverwood V, Blagojevic-Bucknall M, Jinks C, Jordan JL, Protheroe J, Jordan KP. Current evidence on risk factors for knee osteoarthritis in older adults: a systematic review and meta-analysis. *Osteoarthr Cartil*. 2015;23(4):507-515. doi:10.1016/J.JOCA.2014.11.019
3. Carr AJ, Robertsson O, Graves S, et al. Knee replacement. In: *The Lancet*. Vol 379. Lancet; 2012:1331-1340. doi:10.1016/S0140-6736(11)60752-6
4. NJR Editorial Board. *18th Annual Report 2021*.; 2021. https://reports.njrcentre.org.uk/Portals/0/PDFdownloads/NJR 18th Annual Report 2021.pdf
5. Strickland LH, Rahman A, Jenkinson C, et al. Early Recovery Following Total and Unicompartmental Knee Arthroplasty Assessed Using Novel Patient-Reported Measures. Strickland L, ed. *J Arthroplasty*. 2021;36(10):3413-3420. doi:10.1016/j.arth.2021.05.025
6. Ranawat CS, Meftah M, Windsor EN, Ranawat AS. Cementless fixation in total knee arthroplasty. *J Bone Joint Surg Br*. 2012;94-B(11_Supple_A):82-84. doi:10.1302/0301-620X.94B11.30826
7. Ritter MA, Meneghini RM. Twenty-year survivorship of cementless anatomic graduated component total knee arthroplasty. *J Arthroplasty*. 2010;25(4):507-513. doi:10.1016/j.arth.2009.04.018
8. Murray DW, Parkinson RW. Usage of unicompartmental knee arthroplasty. *Bone Jt J*. 2018;100B(4):432-435. doi:10.1302/0301-620X.100B4.BJJ-2017-0716.R1
9. Rand JA, Ilstrup DM. Survivorship analysis of total knee arthroplasty. Cumulative rates of survival of 9200 total knee arthroplasties. *J Bone Jt Surg - Ser A*. 1991;73(3):397-409. doi:10.2106/00004623-199173030-00011
10. Halawi MJ, MJ H. Outcome measures in total joint arthroplasty: Current status, challenges, and future directions. *Orthopedics*. 2015;38(8):e685-e689. doi:10.3928/01477447-20150804-55
11. Gagnier JJ, Mullins M, Huang H, et al. A Systematic Review of Measurement Properties of Patient-Reported Outcome Measures Used in Patients Undergoing Total Knee Arthroplasty. *J Arthroplasty*. 2017;32(5):1688-1697.e7. doi:10.1016/j.arth.2016.12.052
12. Siljander MP, McQuivey KS, Fahs AM, Galasso LA, Serdahely KJ, Karadsheh MS. Current Trends in Patient-Reported Outcome Measures in Total Joint Arthroplasty: A Study of 4 Major Orthopaedic Journals. *J Arthroplasty*. 2018;33(11):3416-3421. doi:10.1016/j.arth.2018.06.034
13. Harris K, Dawson J, Gibbons E, et al. Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty. *Patient Relat Outcome Meas*. 2016;Volume 7:101-108. doi:10.2147/prom.s97774

14. Ramkumar PN, Harris JD, Noble PC. Patient-reported outcome measures after total knee arthroplasty: a systematic review. *Bone Jt Res*. 2015;4(7):120-127. doi:10.1302/2046-3758.47.2000380

15. Dowsey MM, Choong PFM. The utility of outcome measures in total knee replacement surgery. *Int J Rheumatol*. 2013;2013. doi:10.1155/2013/506518

16. Dawson J, Beard DJ, McKibbin H, et al. Development of a patient-reported outcome measure of activity and participation (the OKS- APQ) to supplement the Oxford knee score. *Bone Jt J*. 2014;96(3):332-340. doi:10.1302/0301-620X.96B3

17. Halawi MJ, Jongbloed W, Baron S, Savoy L, Cote MP, Lieberman JR. Patient-Reported Outcome Measures are not a Valid Proxy for Patient Satisfaction in Total Joint Arthroplasty. *J Arthroplasty*. 2020;35(2):335-339. doi:10.1016/j.arth.2019.09.033

18. Strickland LH, Hamilton TW, Jenkinson CC, Murray DW, Pandit HG. Patient-Reported Outcome Measure for Early Postoperative Recovery Following Lower Limb Arthroplasty: A Systematic Review. *J Arthroplasty*. 2016;31(12):2933-2940. doi:10.1016/j.arth.2016.06.026

19. Strickland LH, Jenkinson C, Murray DW, Pandit HG, Jenkinson C. Development of a patient-reported outcome measure (PROM) and change measure for use in early recovery following hip or knee replacement. Strickland L, ed. *J Patient-Reported Outcomes*. 2020;4(1). doi:10.1186/s41687-020-00262-1

20. Dobson F, Hinman RS, Roos EM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. *Osteoarthr Cartil*. 2013;21(8):1042-1052. doi:10.1016/j.joca.2013.05.002

21. Richardson S. The Timed “Up & Go”: A Test of Basic Functional Mobility for Frail Elderly Persons. *J Am Geriatr Soc*. 1991;39(2):142-148. doi:10.1111/j.1532-5415.1991.tb01616.x

22. Mizner RL, Petterson SC, Clements KE, et al. Measuring Functional Improvement After Total Knee Arthroplasty Requires Both Performance-Based and Patient-Report Assessments. A Longitudinal Analysis of Outcomes. *J Arthroplasty*. 2011;26(5):728-737. doi:10.1016/j.arth.2010.06.004

23. Jacobs CA, Christensen CP. Factors influencing patient satisfaction two to five years after primary total knee arthroplasty. *J Arthroplasty*. 2014;29(6):1189-1191. doi:10.1016/j.arth.2014.01.008

24. Kornuijt A, de Kort GJL, Das D, Lenssen AF, van der Weegen W. Recovery of knee range of motion after total knee arthroplasty in the first postoperative weeks: poor recovery can be detected early. *Musculoskelet Surg*. 2019;103(3):289-297. doi:10.1007/s12306-019-00588-0

25. Devers BN, Conditt MA, Jamieson ML, Driscoll MD, Noble PC, Parsley BS. Does Greater Knee Flexion Increase Patient Function and Satisfaction After Total Knee Arthroplasty? *J Arthroplasty*. 2011;26(2):178-186.

doi:10.1016/j.arth.2010.02.008

26. Urbanek JK, Zipunnikov V, Harris T, Crainiceanu C, Harezlak J, Glynn NW. Validation of Gait Characteristics Extracted from Raw Accelerometry during Walking Against Measures of Physical Function, Mobility, Fatigability, and Fitness. *Journals Gerontol - Ser A Biol Sci Med Sci*. 2018;73(5):676-681. doi:10.1093/gerona/glx174

27. Baker R. Gait analysis methods in rehabilitation. *J Neuroeng Rehabil*. 2006;3(1):1-10. doi:10.1186/1743-0003-3-4

28. Theologis T, Stebbins J. The Use of Gait Analysis in the Treatment of Pediatric Foot and Ankle Disorders. *Foot Ankle Clin*. 2010;15(2):365-382. doi:10.1016/j.fcl.2010.02.002

29. Edwards TA, Thompson N, Prescott RJ, Stebbins J, Wright JG, Theologis T. A comparison of conventional and minimally invasive multilevel surgery for children with diplegic cerebral palsy. *Bone Joint J*. 2021;103-B(1):192-197. doi:10.1302/0301-620X.103B1.BJJ-2020-0714.R1

30. Lopez-Nava IH, Angelica MM. Wearable Inertial Sensors for Human Motion Analysis: A review. *IEEE Sens J*. 2016;PP(99). doi:10.1109/JSEN.2016.2609392

31. Kobsar D, Masood Z, Khan H, et al. Wearable Inertial Sensors for Gait Analysis in Adults with Osteoarthritis—A Scoping Review. *Sensors (Switzerland)*. 2020;20(24):1-24. doi:10.3390/s20247143

32. Granat MH, Bolink SAAN. Evaluating physical function and activity in the elderly patient using wearable motion sensors. *EFORT Open Rev*. Published online 2016. doi:10.1302/2058-5241.1.160022

33. Follis S, Chen Z, Mishra S, Howe CL, Toosizadeh N, Dohm M. Comparison of wearable sensor to traditional methods in functional outcome measures: A systematic review. *J Orthop Res*. 2021;39(10):2093-2102. doi:10.1002/jor.24950

34. Troiano RP, Stamatakis E, Bull FC. How can global physical activity surveillance adapt to evolving physical activity guidelines? Needs, challenges and future directions. *Br J Sports Med*. 2020;54(24):1468-1473. doi:10.1136/bjsports-2020-102621

35. Bull FC, Al-Ansari SS, Biddle S, et al. *World Health Organization 2020 Guidelines on Physical Activity and Sedentary Behaviour*. Vol 54. BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine; 2020:1451-1462. doi:10.1136/bjsports-2020-102955

36. Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-Country reliability and validity. *Med Sci Sports Exerc*. 2003;35(8):1381-1395. doi:10.1249/01.MSS.0000078924.61453.FB

37. Zahiri CA, Schmalzried TP, Szuszczewicz ES, Amstutz HC. Assessing activity in joint replacement patients. *J Arthroplasty*. 1998;13(8):890-895. doi:10.1016/S0883-5403(98)90195-4

38. Smith WA, Zucker-Levin A, Mihalko WM, et al. A Randomized Study of Exercise and Fitness Trackers in Obese Patients After Total Knee

Arthroplasty. *Orthop Clin North Am*. 2019;50(1):35-45. doi:10.1016/j.ocl.2018.08.002

39. Paxton RJ, Melanson EL, Stevens-Lapsley JE, Christiansen CL, Christi-Ansen CL. Physical activity after total knee arthroplasty: A critical review. *World J Orthop*. 2015;6(8):614-622. doi:https://dx.doi.org/10.5312/wjo.v6.i8.614

40. Terwee CB, Bouwmeester W, van Elsland SLL, et al. Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Osteoarthr Cartil*. 2011;19(6):620-633. doi:https://dx.doi.org/10.1016/j.joca.2011.01.002

41. Naal FD, Impellizzeri FM. How active are patients undergoing total joint arthroplasty? A systematic review. *Clin Orthop Relat Res*. 2010;468(7):1891-1904. doi:10.1007/s11999-009-1135-9

42. Barber-Westin SD, Noyes FR, S.D. BW, Barber-Westin SD, Noyes FR, S.D. BW. Aerobic Physical Fitness and Recreational Sports Participation After Total Knee Arthroplasty. *Sport Heal A Multidiscip Approach*. 2016;8(6):553-560. doi:10.1177/1941738116670090

43. Mills K, Falchi B, Duckett C, et al. Minimal change in physical activity after lower limb joint arthroplasty, but the outcome measure may be contributing to the problem: a systematic review and meta-analysis. *Physiother (United Kingdom)*. 2019;105(1):35-45. doi:10.1016/j.physio.2018.04.003

44. Arnold JB, Walters JL, Ferrar KE, J.B. A, J.L. W. Does Physical Activity Increase After Total Hip or Knee Arthroplasty for Osteoarthritis? A Systematic Review. *J Orthop Sport Phys Ther*. 2016;46(6):431-442. doi:10.2519/jospt.2016.6449

45. S W, V G, PP K, RC van G, RW P, GM K. Return to Sports and Physical Activity After Total and Unicondylar Knee Arthroplasty: A Systematic Review and Meta-Analysis. *Sports Med*. 2016;46(2):269-292. doi:10.1007/S40279-015-0421-9

46. Naal FD, Impellizzeri FM, Leunig M. Which is the best activity rating scale for patients undergoing total joint arthroplasty? *Clin Orthop Relat Res*. 2009;467(4):958-965. doi:10.1007/s11999-008-0358-5

47. Stewart AL, Mills KM, King AC, Haskell WL, Gillis D, Ritter PL. CHAMPS Physical Activity Questionnaire for Older Adults: outcomes for interventions. *Med Sci Sport Exerc*. 2001;33(7):1126-1141. Accessed October 8, 2021. http://www.acsm-msse.org

48. Almeida GJ, Irrgang JJ, Fitzgerald GK, Jakicic JM, Piva SR. Reliability of Physical Activity Measures During Free-Living Activities in People After Total Knee Arthroplasty. *Phys Ther*. 2016;96(6):898-907. doi:10.2522/ptj.20150407

49. Jasper L. Comparison of Tools to Measure Physical Activity and Stationary Behavior in Older Adults Following Total Knee Arthroplasty. Published online 2020.

50. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. *Br J Sports Med*. 2003;37(3):197-206.

doi:10.1136/bjsm.37.3.197

51. Migueles JH, Aadland E, Andersen LB, et al. GRANADA consensus on analytical approaches to assess associations with accelerometer-determined physical behaviours (physical activity, sedentary behaviour and sleep) in epidemiological studies. *Br J Sports Med*. 2021;0:1-9. doi:10.1136/bjsports-2020-103604

52. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, Mcdowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc*. 2008;40(1):181-188. doi:10.1249/mss.0b013e31815a51b3

53. Doherty A, Jackson D, Hammerla N, et al. Large scale population assessment of physical activity using wrist worn accelerometers: The UK biobank study. *PLoS One*. 2017;12(2):1-14. doi:10.1371/journal.pone.0169649

54. Beg MS, Gupta A, Stewart T, Rethorst CD. Promise of wearable physical activity monitors in oncology practice. *J Oncol Pract*. 2017;13(2):82-89. doi:10.1200/JOP.2016.016857

55. Sitthipornvorakul E, Klinsophon T, Sihawong R, Janwantanakul P. The effects of walking intervention in patients with chronic low back pain: A meta-analysis of randomized controlled trials. *Musculoskelet Sci Pract*. 2018;34:38-46. doi:10.1016/J.MSKSP.2017.12.003

56. L A, RH K, K S, ED de B. Wearable systems for monitoring mobility-related activities in chronic disease: a systematic review. *Sensors (Basel)*. 2010;10(10):9026-9052. doi:10.3390/S101009026

57. Mills K, Hunt MA, Ferber R. Biomechanical deviations during level walking associated with knee osteoarthritis: A systematic review and meta-analysis. *Arthritis Care Res*. 2013;65(10):1643-1665. doi:10.1002/acr.22015

58. Matthews CE, Hagströmer M, Pober DM, Bowles HR. Best Practices for Using Physical Activity Monitors. *Med Sci Sport Exerc*. 2012;44(1):1-17. doi:10.1249/MSS.0b013e3182399e5b.BEST

59. Crowley P, Skotte J, Stamatakis E, et al. Comparison of physical behavior estimates from three different thigh-worn accelerometers brands: A proof-of-concept for the Prospective Physical Activity, Sitting, and Sleep consortium (ProPASS). *Int J Behav Nutr Phys Act*. 2019;16(1). doi:10.1186/s12966-019-0835-0

60. Skotte J, Korshøj M, Kristiansen J, Hanisch C, Holtermann A. Detection of physical activity types using triaxial accelerometers. *J Phys Act Heal*. 2014;11(1):76-84. doi:10.1123/jpah.2011-0347

61. Toth LP, Park S, Springer CM, FEYERABEND MD, Steeves JA, Bassett DR. Video-Recorded Validation of Wearable Step Counters under Free-living Conditions. *Med Sci Sports Exerc*. 2018;50(6):1315-1322. doi:10.1249/MSS.0000000000001569

62. Small SR, Bullock GS, Khalid S, Barker K, Trivella M, Price AJ. Current clinical utilisation of wearable motion sensors for the assessment of outcome following knee arthroplasty: a scoping review. *BMJ Open*. 2019;9(12):e033832. doi:10.1136/bmjopen-2019-033832

63. Shirer M, Llama R, Ubrani J. *Wearable Device Shipments Slow in Q1 2018 as Consumers Shift from Basic Wearables to Smarter Devices , According to IDC*.; 2018. Accessed April 9, 2019. https://www.idc.com/getdoc.jsp?containerId=prUS43900918

64. Troiano RP, Mcclain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. *J Sport Med*. 2015;48(13):1019-1023. doi:10.1136/bjsports-2014-093546.Evolution

65. Van Hees VT, Sabia S, Anderson KN, et al. A novel, open access method to assess sleep duration using a wrist-worn accelerometer. *PLoS One*. 2015;10(11). doi:10.1371/journal.pone.0142533

66. White T, Westgate K, Hollidge S, et al. Estimating energy expenditure from wrist and thigh accelerometry in free-living adults: a doubly labelled water study. *Int J Obes*. Published online July 17, 2019:370247. doi:10.1038/s41366-019-0352-x

67. Willetts M, Hollowell S, Aslett L, Holmes C, Doherty A. Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96 , 220 UK Biobank participants. *Sci Rep*. 2018;(May):1-10. doi:10.1038/s41598-018-26174-1

68. Migueles JH, Cadenas-Sanchez C, Tudor-Locke C, et al. Comparability of published cut-points for the assessment of physical activity: Implications for data harmonization. *Scand J Med Sci Sport*. 2019;29(4):566-574. doi:10.1111/sms.13356

69. St Fleur RG, St George SM, Leite R, Kobayashi M, Agosto Y, Jake-Schoffman DE. Use of fitbit devices in physical activity intervention studies across the life course: Narrative review. *JMIR mHealth uHealth*. 2021;9(5). doi:10.2196/23411

70. Axivity L. AX3 User Manual. Accessed January 9, 2020. https://axivity.com/userguides/ax3/settings/

71. Migueles JH, Cadenas-Sanchez C, Ekelund U, et al. Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. *Sport Med*. 2017;47(9):1821-1845. doi:10.1007/s40279-017-0716-0

72. Khan A, Hammerla N, Mellor S, Plötz T. Optimising sampling rates for accelerometer-based human activity recognition. *Pattern Recognit Lett*. 2016;73:33-40. doi:10.1016/j.patrec.2016.01.001

73. Hart TL, Swartz AM, Cashin SE, Strath SJ. How many days of monitoring predict physical activity and sedentary behaviour in older adults? *Int J Behav Nutr Phys Act*. 2011;8:62. doi:10.1186/1479-5868-8-62

74. Aguilar-Farias N, Martino-Fuentealba P, Salom-Diaz N, Brown WJ. How many days are enough for measuring weekly activity behaviours with the ActivPAL in adults? *J Sci Med Sport*. 2019;22(6):684-688. doi:10.1016/j.jsams.2018.12.004

75. Da Silva SG, Evenson KR, Ekelund U, et al. How many days are needed to estimate wrist-worn accelerometry-assessed physical activity during the second trimester in pregnancy? *PLoS One*. 2018;14(6).

doi:10.1371/journal.pone.0211442

76. Armitage LC, Chi Y, Santos M, et al. Monitoring activity of hip injury patients (MoHIP): A sub-study of the World Hip Trauma Evaluation observational cohort study. *Pilot Feasibility Stud*. 2020;6(1):3D.2. doi:10.1186/s40814-020-00612-2

77. Dadhania S, Pakzad-Shahabi L, Mistry S, et al. BrainWear: Longitudinal, objective assessment of physical activity in 42 HGG patients. *Neuro Oncol*. 2021;23(Supplement_2):ii13-ii14. doi:10.1093/neuonc/noab180.043

78. Migueles JH, Cadenas-Sanchez C, Rowlands A V, et al. Comparability of accelerometer signal aggregation metrics across placements and dominant wrist cut points for the assessment of physical activity in adults. *Sci Rep*. 2019;9(1):18235. doi:10.1038/s41598-019-54267-y

79. White T, Westgate K, Wareham NJ, Brage S. Estimation of physical activity energy expenditure during free-living from wrist accelerometry in UK adults. *PLoS One*. 2016;11(12). doi:10.1371/journal.pone.0167472

80. Migueles JH, Cadenas-Sanchez C, Alcantara JMA, et al. Calibration and Cross-Validation of Accelerometer Cut-Points to Classify Sedentary Time and Physical Activity from Hip and Non-Dominant and Dominant Wrists in Older Adults. Published online 2021. doi:10.3390/s21103326

81. Walmsley R, Chan S, Smith-Byrne K, et al. *Reallocation of Time between Device-Measured Movement Behaviours and Risk of Incident Cardiovascular Disease*. Vol 0. BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine; 2021:bjsports-2021-104050. doi:10.1136/bjsports-2021-104050

82. Sasaki JE, Hickey AM, Staudenmayer JW, John D, Kent JA, Freedson PS. Performance of activity classification algorithms in free-living older adults. *Med Sci Sports Exerc*. 2016;48(5):941-949. doi:10.1249/MSS.0000000000000844

83. Ellis K, Kerr J, Godbole S, Staudenmayer J, Lanckriet G. Hip and wrist accelerometer algorithms for free-living behavior classification. *Med Sci Sports Exerc*. 2016;48(5):933-940. doi:10.1249/MSS.0000000000000840

84. Narayanan A, Desai F, Stewart T, Duncan S, MacKay L. Application of raw accelerometer data and machine-learning techniques to characterize human movement behavior: A systematic scoping review. *J Phys Act Heal*. 2020;17(3):360-383. doi:10.1123/JPAH.2019-0088

85. Migueles JH, Rowlands A V, Huber F, Hees VT Van. GGIR : A Research Community – Driven Open Source R Package for Generating Physical Activity and Sleep Outcomes From Multi-Day Raw Accelerometer Data. *J Meas Phys Behav*. 2019;2(3):188-196.

86. Brønd JC, Arvidsson D. Sampling frequency affects the processing of Actigraph raw acceleration data to activity counts. *J Appl Physiol*. 2016;120(3):362-369. doi:10.1152/japplphysiol.00628.2015

87. Godfrey A, Hetherington V, Shum H, Bonato P, Lovell NH, Stuart S. From A to Z: Wearable technology explained. *Maturitas*. 2018;113(April):40-47. doi:10.1016/j.maturitas.2018.04.012

88. Clark CCT, Nobre GC, Fernandes JFT, et al. Physical activity characterization: does one site fit all? *Physiol Meas*. Published online 2018:aadad0. doi:10.1088/1361-6579/aadad0

89. *The Joanna Briggs Institute Reviewer's Manual 2015: Methodology for JBI Scoping Reviews*.; 2015. Accessed November 30, 2018. www.joannabriggs.org

90. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med*. 2018;169(7):467. doi:10.7326/M18-0850

91. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev*. 2016;5(1):210. doi:10.1186/s13643-016-0384-4

92. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med*. 2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097

93. National Joint Registry for England, Wales NI and the I of M. *15th Annual Report*.; 2018. doi:10.1038/nmat2505

94. Rahman J, Tang Q, Monda M, Miles J, Mccarthy I. Gait assessment as a functional outcome measure in total knee arthroplasty: a cross-sectional study. *BMC Musculoskelet Disord*. 2015;16(1):66. doi:10.1186/s12891-015-0525-2

95. Ali R, Atallah L, Lo B, Yang GZ. Detection and analysis of transitional activity in manifold space. *IEEE Trans Inf Technol Biomed*. 2012;16(1):119-128. doi:10.1109/TITB.2011.2165320

96. Christiansen CL, Bade MJ, Paxton RJ, Stevens-Lapsley JE. Measuring movement symmetry using tibial-mounted accelerometers for people recovering from total knee arthroplasty. *Clin Biomech*. 2015;30(7):732-737. doi:10.1016/j.clinbiomech.2015.04.013

97. Kwasnicki RM, Ali R, Jordan SJ, et al. A wearable mobility assessment device for total knee replacement: Alongitudinal feasibility study. *Int J Surg*. 2015;18:14-20. doi:10.1016/j.ijsu.2015.04.032

98. Rahman J, Tang Q, Monda M, Miles J, McCarthy I. Gait assessment as a functional outcome measure in total knee arthroplasty: A cross-sectional study. *BMC Musculoskelet Disord*. 2015;16:66. doi:10.1186/s12891-015-0525-2

99. Senden R, Grimm B, Meijer K, Savelberg H, Heyligers IC. The importance to including objective functional outcomes in the clinical follow up of total knee arthroplasty patients. *Knee*. 2011;18(5):306-311. doi:10.1016/j.knee.2010.07.008

100. Storey AST, Myrah AM, Bauck RA, Brinkman DM, Friess SN, Webber SC. Indoor and outdoor mobility following total knee arthroplasty. *Physiother Canada*. 2013;65(3):279-288. doi:10.3138/ptc.2012-36

101. Zhang H hua, Yan S hua, Fang C, Guo X yuan, Zhang K. Clinical Evaluation and Gait Characteristics before and after Total Knee Arthroplasty Based on

a Portable Gait Analyzer. *Orthop Surg*. 2016;8(3):360-366. doi:10.1111/os.12270

102. Chapman RM, Moschetti WE, Van Citters DW. Stance and swing phase knee flexion recover at different rates following total knee arthroplasty: An inertial measurement unit study. *J Biomech*. 2019;84:129-137. doi:10.1016/j.jbiomech.2018.12.027

103. Bolink SAAN, Grimm B, Heyligers IC. Patient-reported outcome measures versus inertial performance-based outcome measures: A prospective study in patients undergoing primary total knee arthroplasty. *Knee*. 2015;22(6):618-623. doi:10.1016/j.knee.2015.04.002

104. Chiang CY, Chen KH, Liu KC, Hsu SJP, Chan CT. Data collection and analysis using wearable sensors for monitoring knee range of motion after total knee arthroplasty. *Sensors (Switzerland)*. 2017;17(2):418. doi:10.3390/s17020418

105. Hiyama Y, Asai T, Wada O, et al. Gait variability before surgery and at discharge in patients who undergo total knee arthroplasty: A cohort study. Williams BO, ed. *PLoS One*. 2015;10(1):e0117683. doi:10.1371/journal.pone.0117683

106. Kluge F, Hannink J, Pasluosta C, et al. Pre-operative sensor-based gait parameters predict functional outcome after total knee arthroplasty. *Gait Posture*. 2018;66:194-200. doi:10.1016/j.gaitpost.2018.08.026

107. Ramkumar PN, Haeberle HS, Ramanathan D, et al. Remote Patient Monitoring Using Mobile Health for Total Knee Arthroplasty: Validation of a Wearable and Machine Learning–Based Surveillance Platform. *J Arthroplasty*. Published online 2019:1-7. doi:10.1016/j.arth.2019.05.021

108. Youn IH, Youn JH, Zeni JA, Knarr BA. Biomechanical gait variable estimation using wearable sensors after unilateral total knee arthroplasty. *Sensors (Switzerland)*. 2018;18(5):1577. doi:10.3390/s18051577

109. van Hemert WLW, Senden R, Grimm B, Kester ADM, Linde MJA, Heyligers IC. Patella retention versus replacement in total knee arthroplasty; functional and clinimetric aspects. *Arch Orthop Trauma Surg*. 2009;129(2):259-265. doi:10.1007/s00402-008-0640-8

110. van Hemert WLW, Meyers WGH, Kleijn LLA, Heyligers IC, Grimm B. Functional outcome of knee arthroplasty is dependent upon the evaluation method employed. *Eur J Orthop Surg Traumatol*. 2009;19(6):415-422. doi:10.1007/s00590-009-0450-x

111. van den Dikkenberg N, Meijer OG, van der Slikke RMA, et al. Measuring functional abilities of patients with knee problems: rationale and construction of the DynaPort knee test. *KNEE Surg Sport Traumatol Arthrosc*. 2002;10(4):204-212. doi:10.1007/s00167-002-0279-x

112. Daugaard R, Tjur M, Sliepen M, Lipperts M, Grimm B, Mechlenburg I. Are patients with knee osteoarthritis and patients with knee joint replacement as physically active as healthy persons? *J Orthop Transl*. 2018;14:8-15. doi:10.1016/j.jot.2018.03.001

113. Kahn TL, Schwarzkopf R. Does Total Knee Arthroplasty Affect Physical

Activity Levels? Data from the Osteoarthritis Initiative. *J Arthroplasty*. 2015;30(9):1521-1525. doi:10.1016/j.arth.2015.03.016

114. Kahn TL, Schwarzkopf R. Do Total Knee Arthroplasty Patients Have a Higher Activity Level Compared to Patients With Osteoarthritis? *Geriatr Orthop Surg Rehabil*. 2016;7(3):142-147. doi:10.1177/2151458516654518

115. Webber SC, Strachan SM, Pachu NS. Sedentary behavior, cadence, and physical activity outcomes after knee arthroplasty. *Med Sci Sports Exerc*. 2017;49(6):1057-1065. doi:10.1249/MSS.0000000000001207

116. Cooper NA, Rakel BA, Zimmerman B, et al. Predictors of multidimensional functional outcomes after total knee arthroplasty. *J Orthop Res*. 2017;35(12):2790-2798. doi:10.1002/jor.23596

117. Lützner C, Kirschner S, Lützner J. Patient Activity After TKA Depends on Patient-specific Parameters. *Clin Orthop Relat Res*. 2014;472(12):3933-3940. doi:10.1007/s11999-014-3813-5

118. De Groot IB, Bussmann HJ, Stam HJ, Verhaar JA. Small increase of actual physical activity 6 months after total hip or knee arthroplasty. *Clin Orthop Relat Res*. 2008;466(9):2201-2208. doi:10.1007/s11999-008-0315-3

119. Harding P, Holland AE, Delany C, Hinman RS. Do activity levels increase after total hip and knee arthroplasty? *Clin Orthop Relat Res*. 2014;472(5):1502-1511. doi:10.1007/s11999-013-3427-3

120. Hayashi K, Kako M, Suzuki K, et al. Impact of variation in physical activity after total joint replacement. *J Pain Res*. 2018;11:2399-2406. doi:10.2147/JPR.S178853

121. Vissers MM, Bussmann JB, De Groot IB, Verhaar JAN, Reijman M. Physical functioning four years after total hip and knee arthroplasty. *Gait Posture*. 2013;38(2):310-315. doi:10.1016/j.gaitpost.2012.12.007

122. Agarwal V, Smuck M, Shah NH. Quantifying the relative change in physical activity after Total Knee Arthroplasty using accelerometer based measurements. In: *AMIA Joint Summits on Translational Science Proceedings. AMIA Joint Summits on Translational Science*. Vol 2017. ; 2017:463-472. Accessed November 5, 2018. http://www.ncbi.nlm.nih.gov/pubmed/28815146%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5543365

123. Bolszak S, Casartelli NC, Impellizzeri FM, Maffiuletti NA. Validity and reproducibility of the Physical Activity Scale for the Elderly (PASE) questionnaire for the measurement of the physical activity level in patients after total knee arthroplasty. *BMC Musculoskelet Disord*. 2014;15(1):46. doi:10.1186/1471-2474-15-46

124. Christiansen MB, Thoma LM, Master H, et al. The feasibility and preliminary outcomes of a physical therapist-administered physical activity intervention after total knee replacement. *Arthritis Care Res (Hoboken)*. Published online 2019. doi:https://dx.doi.org/10.1002/acr.23882

125. Fenten MGE, Bakker SMK, Scheffer GJ, Wymenga AB, Stienstra R, Heesterbeek PJC. Femoral nerve catheter vs local infiltration for analgesia in fast track total knee arthroplasty: short-term and long-term outcomes. *Br*

J Anaesth. 2018;121(4):850-858. doi:10.1016/j.bja.2018.05.069

126. Frimpong E, McVeigh JA, van der Jagt D, et al. Light intensity physical activity increases and sedentary behavior decreases following total knee arthroplasty in patients with osteoarthritis. *Knee Surgery, Sport Traumatol Arthrosc*. 2018;In Press:1-10. doi:10.1007/s00167-018-4987-2

127. Losina E, Collins JE, Deshpande BR, et al. Financial Incentives and Health Coaching to Improve Physical Activity Following Total Knee Replacement: A Randomized Controlled Trial. *Arthritis Care Res*. 2018;70(5):732-740. doi:10.1002/acr.23324

128. Luna IE, Kehlet H, Wede HR, Hoevsgaard SJ, Aasvang EK. Objectively measured early physical activity after total hip or knee arthroplasty. *J Clin Monit Comput*. 2019;33(3):509-522. doi:10.1007/s10877-018-0185-5

129. Lützner C, Beyer F, Kirschner S, Lützner J. How Much Improvement in Patient Activity Can Be Expected After TKA? *Orthopedics*. 2016;39(3 Suppl):S18-S23. doi:10.3928/01477447-20160509-15

130. Pellegrini CA, Chang RW, Dunlop DD, et al. Comparison of a Patient-Centered Weight Loss Program starting before versus after knee replacement: A pilot study. *Obes Res Clin Pract*. 2018;12(5):472-478. doi:https://doi.org/10.1016/j.orcp.2018.06.009

131. Schotanus MGM, Bemelmans YFL, Grimm B, Heyligers IC, Kort NP. Physical activity after outpatient surgery and enhanced recovery for total knee arthroplasty. In: *Knee Surgery, Sports Traumatology, Arthroscopy*. Vol 25. Springer Berlin Heidelberg; 2017:3366-3371. doi:10.1007/s00167-016-4256-1

132. Taniguchi M, Sawano S, Kugo M, Maegawa S, Kawasaki T, Ichihashi N. Physical Activity Promotes Gait Improvement in Patients With Total Knee Arthroplasty. *J Arthroplasty*. 2016;31(5):984-988. doi:10.1016/j.arth.2015.11.012

133. Tsuji S, Tomita T, Fujii M, Laskin RS, Yoshikawa H, Sugamoto K. Is minimally invasive surgery-total knee arthroplasty truly less invasive than standard total knee arthroplasty? A quantitative evaluation. *J Arthroplasty*. 2010;25(6):970-976. doi:10.1016/j.arth.2009.06.016

134. Twiggs J, Salmon L, Kolos E, Bogue E, Miles B, Roe J. Measurement of physical activity in the pre- and early post-operative period after total knee arthroplasty for Osteoarthritis using a Fitbit Flex device. *Med Eng Phys*. 2018;51:31-40. doi:10.1016/j.medengphy.2017.10.007

135. Brandes M, Ringling M, Winter C, Hillmann A, Rosenbaum D. Changes in physical activity and health-related quality of life during the first year after total knee arthroplasty. *Arthritis Care Res*. 2011;63(3):328-334. doi:10.1002/acr.20384

136. Hayes DA, Watts MC, Anderson LJ, Walsh WR. Knee arthroplasty: A cross-sectional study assessing energy expenditure and activity. *ANZ J Surg*. 2011;81(5):371-374. doi:10.1111/j.1445-2197.2010.05570.x

137. Vissers MM, De Groot IB, Reijman M, Bussmann JB, Stam HJ, Verhaar JAN. Functional capacity and actual daily activity do not contribute to patient

satisfaction after total knee arthroplasty. *BMC Musculoskelet Disord*. 2010;11(1):121. doi:10.1186/1471-2474-11-121

138. Khan H, Walker PS, Zuckerman JD, et al. The Potential of Accelerometers in the Evaluation of Stability of Total Knee Arthroplasty. *J Arthroplasty*. 2013;28(3):459-462. doi:10.1016/j.arth.2012.07.025

139. Roberts D, Khan H, Kim JH, Slover J, Walker PS. Acceleration-based joint stability parameters for total knee arthroplasty that correspond with patient-reported instability. *Proc Inst Mech Eng Part H J Eng Med*. 2013;227(10):1104-1113. doi:10.1177/0954411913493724

140. Soeno T, Mochizuki T, Tanifuji O, et al. No differences in objective dynamic instability during acceleration of the knee with or without subjective instability post-Total knee arthroplasty. Williams JL, ed. *PLoS One*. 2018;13(3):e0194221. doi:10.1371/journal.pone.0194221

141. Papagiannis GI, Triantafyllou AI, Roumpelakis IM, Papagelopoulos PJ, Babis GC. Gait analysis methodology for the measurement of biomechanical parameters in total knee arthroplasties. A literature review. *J Orthop*. 2018;15(1):181-185. doi:10.1016/j.jor.2018.01.048

142. Davis RB, Õunpuu S, Tyburski D, Gage JR. A gait analysis data collection and reduction technique. *Hum Mov Sci*. 1991;10(5):575-587. doi:10.1016/0167-9457(91)90046-Z

143. Terwee CB, Mokkink LB, Steultjens MPM, Dekker J. Performance-based methods for measuring the physical function of patients with osteoarthritis of the hip or knee: A systematic review of measurement properties. *Rheumatology*. 2006;45(7):890-902. doi:10.1093/rheumatology/kei267

144. Ramkumar PN, Haeberle HS, Ramanathan D, et al. Remote Patient Monitoring Using Mobile Health for Total Knee Arthroplasty: Validation of a Wearable and Machine Learning–Based Surveillance Platform. *J Arthroplasty*. 2019;34(10):2253-2259. doi:10.1016/j.arth.2019.05.021

145. Wang C, Lu W, Narayanan MR, Redmond SJ, Lovell NH. Low-power technologies for wearable telecare and telehealth systems: A review. *Biomed Eng Lett*. 2015;5(1):1-9. doi:10.1007/s13534-015-0174-2

146. Walter RT, Thompson WR. Worldwide Survey of Fitness Trends for 2019. *Am Coll Sport Med Heal Fit J*. 2018;22(6):10-17. doi:10.1249/FIT.0000000000000252

147. Lim SER, Dodds R, Sayer DBAA, Roberts HC. Physical activity among hospitalised older people : insights from upper and lower limb accelerometry. *Aging Clin Exp Res*. 2018;30(11):1363-1369. doi:10.1007/s40520-018-0930-0

148. Trost SG, Zheng Y, Wong WK. Machine learning for activity recognition: hip versus wrist data. *Physiol Meas*. 2014;35(11):2183-2189. doi:10.1088/0967-3334/35/11/2183

149. Amiri A, Peltier N, Goldberg C, et al. WearSense: Detecting Autism Stereotypic Behaviors through Smartwatches. *Healthcare*. 2017;5(1):11. doi:10.3390/healthcare5010011

150. Zhang S, Murray P, Zillmer R, Eston RG, Catt M, Rowlands A V. Activity classification using the genea: Optimum sampling frequency and number of axes. *Med Sci Sports Exerc*. 2012;44(11):2228-2234. doi:10.1249/MSS.0b013e31825e19fd

151. Zhang S, Rowlands A V., Murray P, Hurst TL. Physical activity classification using the GENEA wrist-worn accelerometer. *Med Sci Sports Exerc*. 2012;44(4):742-748. doi:10.1249/MSS.0b013e31823bf95c

152. Cleland I, Kikhia B, Nugent C, et al. Optimal placement of accelerometers for the detection of everyday activities. *Sensors (Basel)*. 2013;13(7):9183-9200. doi:10.3390/s130709183

153. Hammett T, Simonian A, Austin M, et al. Changes in Physical Activity After Total Hip or Knee Arthroplasty: A Systematic Review and Meta-Analysis of Six- and Twelve-Month Outcomes. *Arthritis Care Res (Hoboken)*. 2018;70(6):892-901. doi:https://dx.doi.org/10.1002/acr.23415

154. Frimpong E, McVeigh JA, van der Jagt D, et al. Light intensity physical activity increases and sedentary behavior decreases following total knee arthroplasty in patients with osteoarthritis. *Knee Surgery, Sport Traumatol Arthrosc*. 2018;0:1-10. doi:10.1007/s00167-018-4987-2

155. Brandes M, Ringling M, Winter C, Hillmann A, Rosenbaum D. Changes in physical activity and health-related quality of life during the first year after total knee arthroplasty. *Arthritis Care Res*. 2011;63(3):328-334. doi:10.1002/acr.20384

156. Stevens-Lapsley JE, Schenkman ML, Dayton MR. Comparison of Self-Reported Knee Injury and Osteoarthritis Outcome Score to Performance Measures in Patients After Total Knee Arthroplasty. *PM R*. 2011;3(6):541-549. doi:10.1016/j.pmrj.2011.03.002

157. A Prospective Multicenter Longitudinal Cohort Study of the Mymobility Platform - Full Text View - ClinicalTrials.gov. Accessed February 1, 2019. https://clinicaltrials.gov/ct2/show/NCT03737149

158. Crawford DA, Duwelius PJ, Sneller MA, et al. 2021 Mark Coventry Award: Use of a smartphone-based care platform after primary partial and total knee arthroplasty: A prospective randomized controlled trial. *Bone Jt J*. 2021;103-B(7):3-12. doi:10.1302/0301-620X.103B6.BJJ-2020-2352.R1

159. Thompson D, Peacock O, Western M, Batterham AM. Multidimensional Physical Activity. *Exerc Sport Sci Rev*. 2015;43(2):67-74. doi:10.1249/jes.0000000000000039

160. Panda N, Solsky I, Huang EJ, et al. Using Smartphones to Capture Novel Recovery Metrics After Cancer Surgery. *JAMA Surg*. 2019;02114:1-7. doi:10.1001/jamasurg.2019.4702

161. Soeno T, Mochizuki T, Tanifuji O, et al. No differences in objective dynamic instability during acceleration of the knee with or without subjective instability post-Total knee arthroplasty. Williams JL, ed. *PLoS One*. 2018;13(3):e0194221. doi:10.1371/journal.pone.0194221

162. Ekelund U, Tarp J, Steene-Johannessen J, et al. Dose-response associations between accelerometry measured physical activity and

sedentary time and all cause mortality: Systematic review and harmonised meta-analysis. *BMJ*. 2019;366. doi:10.1136/bmj.l4570

163. Redfield MM, Anstrom KJ, Levine JA, et al. Isosorbide Mononitrate in Heart Failure with Preserved Ejection Fraction. *N Engl J Med*. 2015;373(24):2314-2324. doi:10.1056/NEJMoa1510774

164. Wolff-Hughes DL, McClain JJ, Dodd KW, Berrigan D, Troiano RP. Number of accelerometer monitoring days needed for stable group-level estimates of activity. *Physiol Meas*. 2016;37(9):1447-1455. doi:10.1088/0967-3334/37/9/1447

165. Ricardo LIC, Wendt A, Galliano LM, et al. Number of days required to estimate objectively measured physical activity constructs in different age groups. *bioRxiv*. Published online April 15, 2019:610030. doi:10.1101/610030

166. Taraldsen K, Vereijken B, Thingstad P, Sletvold O, Helbostad JL. Multiple days of monitoring are needed to obtain a reliable estimate of physical activity in hip-fracture patients. *J Aging Phys Act*. 2014;22(2):173-177. doi:10.1123/JAPA.2012-0130

167. Clevenger KA, Brønd JC, Arvidsson D, et al. Impact of ActiGraph Sampling Rate and Intermonitor Comparability on Measures of Physical Activity in Adults. *J Meas Phys Behav*. 2021;1(aop):1-11. doi:10.1123/JMPB.2021-0016

168. Clevenger KA, Pfeiffer KA, Mackintosh KA, et al. Effect of sampling rate on acceleration and counts of hip- and wrist-worn ActiGraph accelerometers in children. *Physiol Meas*. 2019;40(9):095008. doi:10.1088/1361-6579/ab444b

169. Karas M, Bai J, Strączkiewicz M, et al. Accelerometry data in health research: challenges and opportunities Review and examples. *Stat Biosci*. 2019;11(2):210-237. doi:10.1055/s-0034-1387696

170. van Hees VT, Fang Z, Langford J, et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. *J Appl Physiol*. 2014;117(7):738-744. doi:10.1152/japplphysiol.00421.2014

171. Hildebrand M, Van Hees VT, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist-and hip-worn monitors. *Med Sci Sports Exerc*. 2014;46(9):1816-1824. doi:10.1249/MSS.0000000000000289

172. Vähä-Ypyä H, Vasankari T, Husu P, Suni J, Sievänen H. A universal, accurate intensity-based classification of different physical activities using raw data of accelerometer. *Clin Physiol Funct Imaging*. 2015;35(1):64-70. doi:10.1111/cpf.12127

173. Mannini A, Intille SS, Rosenberger M, Sabatini AM, Haskell W. Activity recognition using a single accelerometer placed at the wrist or ankle. *Med Sci Sports Exerc*. 2013;45(11):2193-2203. doi:10.1249/MSS.0b013e31829736d6

174. de Almeida Mendes M, da Silva ICM, Ramires V V., Reichert FF, Martins

RC, Tomasi E. Calibration of raw accelerometer data to measure physical activity: A systematic review. *Gait Posture*. 2018;61:98-110. doi:10.1016/j.gaitpost.2017.12.028

175. Tudor-Locke C, Craig CL, Brown WJ, et al. How many steps/day are enough? for adults. *Int J Behav Nutr Phys Act*. 2011;8. doi:10.1186/1479-5868-8-79

176. ANSI/CTA-2056. *Physical Activity Monitoring for Fitness Wearables: Step Counting*. Vol ANSI/CTA-2. Consumer Technology Association; 2016. https://webstore.ansi.org/standards/ansi/cta20562016ansi

177. Feito Y, Bassett DR, Thompson DL, Y F, DR B, DL T. Evaluation of activity monitors in controlled and free-living environments. *Med Sci Sports Exerc*. 2012;44(4):733-741. doi:10.1249/MSS.0b013e3182351913

178. Alinia P, Cain C, Fallahzadeh R, Shahrokni A, Cook D, Ghasemzadeh H. How accurate is your activity tracker? A comparative study of step counts in low-intensity physical activities. *JMIR mHealth uHealth*. 2017;5(8):e106. doi:10.2196/mhealth.6321

179. Larkin L, Nordgren B, Purtill H, Brand C, Fraser A, Kennedy N. Criterion validity of the activPAL activity monitor for sedentary and physical activity patterns in people who have Rheumatoid arthritis. *Phys Ther*. 2016;96(7):1093-1101. doi:10.2522/ptj.20150281

180. Kooner P, Schubert T, Howard JL, Lanting BA, Teeter MG, Vasarhelyi EM. Evaluation of the effect of gait aids, such as canes, crutches, and walkers, on the accuracy of step counters in healthy individuals. *Orthop Res Rev*. 2021;13:1-8. doi:10.2147/ORR.S292255

181. De Ridder R, De Blaiser C, R. DR, et al. Activity trackers are not valid for step count registration when walking with crutches. *Gait Posture*. 2019;70(February):30-32. doi:10.1016/j.gaitpost.2019.02.009

182. Johnston W, Judice PB, García PM, et al. Recommendations for determining the validity of consumer wearable and smartphone step count: Expert statement and checklist of the INTERLIVE network. *Br J Sports Med*. 2020;0:1-14. doi:10.1136/bjsports-2020-103147

183. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act*. 2015;12(159):1-22. doi:10.1186/s12966-015-0314-1

184. Bagui S, Fang X, Bagui S, et al. An improved step counting algorithm using classification and double autocorrelation. *Int J Comput Appl*. Published online February 12, 2020:1-10. doi:10.1080/1206212X.2020.1726006

185. John D, Morton A, Arguello D, Lyden K, Bassett D. “What is a step?" Differences in how a step is detected among three popular activity monitors that have impacted physical activity research. *Sensors (Switzerland)*. 2018;18(4):1-15. doi:10.3390/s18041206

186. Karas M, Stra M, Fadel W, Harezlak J, Crainiceanu CM, Urbanek JK. Adaptive empirical pattern transformation (ADEPT) with application to walking stride segmentation. *Biostatistics*. Published online 2019:1-17. doi:10.1093/biostatistics/kxz033

187. Mattfeld R, Jesch E, Hoover A. Evaluating pedometer algorithms on semi-regular and unstructured gaits. *Sensors*. 2021;21(13):13-16. doi:10.3390/s21134260

188. Gershuny J, Harms T, Doherty A, et al. Testing Self-Report Time-Use Diaries against Objective Instruments in Real Time. *Sociol Methodol*. Published online December 30, 2019:008117501988459. doi:10.1177/0081175019884591

189. Cho Y, Cho H, Kyung CM. Design and Implementation of Practical Step Detection Algorithm for Wrist-Worn Devices. *IEEE Sens J*. 2016;16(21):7720-7730. doi:10.1109/JSEN.2016.2603163

190. Kang X, Huang B, Qi G. A novel walking detection and step counting algorithm using unconstrained smartphones. *Sensors (Switzerland)*. 2018;18(1). doi:10.3390/s18010297

191. Gu F, Khoshelham K, Shang J, Yu F, Wei Z. Robust and accurate smartphone-based step counting for indoor localization. *IEEE Sens J*. 2017;17(11):3453-3460. doi:10.1109/JSEN.2017.2685999

192. Hickey A, Del Din S, Rochester L, Godfrey A. Detecting free-living steps and walking bouts: Validating an algorithm for macro gait analysis. *Physiol Meas*. 2017;38(1):N1-N15. doi:10.1088/1361-6579/38/1/N1

193. Brajdic A, Harle R. Walk detection and step counting on unconstrained smartphones. In: *UbiComp 2013 - Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ; 2013:225-234. doi:10.1145/2493432.2493449

194. Hurt CP, Lein, Donald H J, Smith CR, et al. Assessing a novel way to measure three common rehabilitation outcome measures using a custom mobile phone application. *Gait Posture*. 2019;73:246-250. doi:10.1016/j.gaitpost.2019.07.251

195. Fortune E, Lugade V, Morrow M, Kaufman K. Validity of using tri-axial accelerometers to measure human movement - Part II: Step counts at a wide range of gait velocities. *Med Eng Phys*. 2014;36(6):659-669. doi:10.1016/j.medengphy.2014.02.006

196. Östlind E, Sant'Anna A, Eek F, Stigmar K, Ekvall Hansson E. Physical activity patterns, adherence to using a wearable activity tracker during a 12-week period and correlation between self-reported function and physical activity in working age individuals with hip and/or knee osteoarthritis. *BMC Musculoskelet Disord*. 2021;22(1). doi:10.1186/s12891-021-04338-x

197. Ooms A, Dutton SJ, Parsons S, et al. Statistical analysis plan for a pragmatic phase III randomised controlled trial examining behaviour change physiotherapy intervention to increase physical activity following hip and knee replacements: the PEP-TALK trial. *Trials 2021 221*. 2021;22(1):1-9. doi:10.1186/S13063-021-05362-X

198. Kline PW, Melanson EL, Sullivan WJ, et al. Improving Physical Activity Through Adjunct Telerehabilitation Following Total Knee Arthroplasty: Randomized Controlled Trial Protocol. *Phys Ther*. 2019;99(1):37-45. doi:10.1093/ptj/pzy119

199. Brandes M, Wirsik N, Niehoff H, et al. Impact of a tailored activity counselling intervention during inpatient rehabilitation after knee and hip arthroplasty - An explorative RCT. *BMC Musculoskelet Disord*. 2018;19(1):N.PAG-N.PAG. doi:10.1186/s12891-018-2130-7

200. Pellegrini CA, Lee J, DeVivo KE, Harpine CE, Del Gaizo DJ, Wilcox S. Reducing sedentary time using an innovative mHealth intervention among patients with total knee replacement: Rationale and study protocol. *Contemp Clin Trials Commun*. 2021;22:100810. doi:10.1016/j.conctc.2021.100810

201. Christiansen MB, Thoma LM, Master H, et al. Feasibility and Preliminary Outcomes of a Physical Therapist–Administered Physical Activity Intervention After Total Knee Replacement. *Arthritis Care Res*. 2020;72(5):661-668. doi:10.1002/acr.23882

202. Bolink SAANAN, Grimm B, Heyligers IC, Granat MH, Heyligers IC. Patient-reported outcome measures versus inertial performance-based outcome measures: A prospective study in patients undergoing primary total knee arthroplasty. *Knee*. 2015;22(6):618-623. doi:10.1016/J.KNEE.2015.04.002

203. Bassett DR, Toth LP, LaMunion SR, Crouter SE. Step Counting: A Review of Measurement Considerations and Health-Related Applications. *Sport Med*. 2017;47(7):1303-1315. doi:10.1007/s40279-016-0663-1

204. Saint-Maurice PF, Troiano RP, Bassett DR, et al. Association of Daily Step Count and Step Intensity with Mortality among US Adults. *JAMA - J Am Med Assoc*. 2020;323(12):1151-1160. doi:10.1001/jama.2020.1382

205. Paluch AE, Bajpai S, Bassett DR, et al. Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. *Lancet Public Heal*. 2022;7(3):e219-e228. doi:10.1016/s2468-2667(21)00302-9

206. Hall KS, Hyde ET, Bassett DR, et al. Systematic review of the prospective association of daily step counts with risk of mortality, cardiovascular disease, and dysglycemia. *Int J Behav Nutr Phys Act*. 2020;17(1). doi:10.1186/s12966-020-00978-9

207. Schrack JA, Kuo PL, Wanigatunga AA, et al. Active-to-Sedentary Behavior Transitions, Fatigability, and Physical Functioning in Older Adults. *Journals Gerontol - Ser A Biol Sci Med Sci*. 2019;74(4):560-567. doi:10.1093/gerona/gly243

208. Wanigatunga AA, Di J, Zipunnikov V, et al. Association of Total Daily Physical Activity and Fragmented Physical Activity with Mortality in Older Adults. *JAMA Netw Open*. 2019;2(10):e1912352-e1912352. doi:10.1001/jamanetworkopen.2019.12352

209. Fraysse F, Post D, Eston R, Kasai D, Rowlands A V, Parfitt G. Physical Activity Intensity Cut-Points for Wrist-Worn GENEActiv in Older Adults. *Front Sport Act Living*. 2021;2:579278. doi:10.3389/fspor.2020.579278

210. UKBiobank. UK Biobank Showcase: Data-Field 924. Accessed September 27, 2021. https://biobank.ndph.ox.ac.uk/showcase/field.cgi?tk=X93aHbDKXiAcPMuRRGAEZ1Jug8FQ0lQA185025&id=924

211. UK Biobank. *UK Biobank Hospital Inpatient Data*.; 2020. Accessed September 21, 2021. http://www.ukbiobank.ac.uk/

212. NHS Digital. *Indicator Specification: Summary Hospital-Level Mortality Indicator*.; 2014.

213. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic Dis*. 1987;40(5):373-383. doi:10.1016/0021-9681(87)90171-8

214. National Joint Registry for England Wales Northern Ireland and the Isle of Man. OPCS Codes relevant to procedures recorded on the NJR v5. 2017;(July):1-16. Accessed September 21, 2021. www.njrcentre.org.uk/njrcentre/Portals/0/Documents/England/Data collection forms/OPCS Procedure codes relevant to NJRv5.pdf?ver=2017-07-18-144930-523

215. Chudasama Y V., Khunti KK, Zaccardi F, et al. Physical activity, multimorbidity, and life expectancy: a UK Biobank longitudinal study. *BMC Med*. 2019;17(1):108. doi:10.1186/s12916-019-1339-0

216. Tudor-Locke C, Han H, Aguiar EJ, et al. How fast is fast enough? Walking cadence (steps/min) as a practical estimate of intensity in adults: a narrative review. *Br J Sports Med*. 2018;52(12):776-788. doi:10.1136/BJSPORTS-2017-097628

217. Katz JN. Total joint replacement in osteoarthritis. *Best Pract Res Clin Rheumatol*. 2006;20(1):145-153. doi:10.1016/j.berh.2005.09.003

218. Arnold JB, Walters JL, Ferrar KE. Does physical activity increase after total hip or knee arthroplasty for osteoarthritis? A systematic review. *J Orthop Sports Phys Ther*. 2016;46(6):431-442. doi:10.2519/jospt.2016.6449

219. Höll S, Blum A, Gosheger G, Dieckmann R, Winter C, Rosenbaum D. Clinical outcome and physical activity measured with StepWatch 3™ Activity Monitor after minimally invasive total hip arthroplasty. *J Orthop Surg Res*. 2018;13(1). doi:10.1186/S13018-018-0775-4

220. Cooper NA, Rakel BA, Zimmerman B, et al. Predictors of multidimensional functional outcomes after total knee arthroplasty. *J Orthop Res*. 2017;35(12):2790-2798. doi:10.1002/jor.23596

221. Daugaard R, Tjur M, Sliepen M, Lipperts M, Grimm B, Mechlenburg I. Are patients with knee osteoarthritis and patients with knee joint replacement as physically active as healthy persons? *J Orthop Transl*. 2018;14(9):8-15. doi:10.1016/j.jot.2018.03.001

222. Lützner C, Kirschner S, Lützner J. Patient Activity After TKA Depends on Patient-specific Parameters. *Clin Orthop Relat Res*. 2014;472(12):3933-3940. doi:10.1007/s11999-014-3813-5

223. Walker DJ, Heslop PS, Chandler C, Pinder IM. Measured ambulation and self-reported health status following total joint replacement for the osteoarthritic knee. *Rheumatology*. 2002;41(7):755-758. doi:10.1093/rheumatology/41.7.755

224. Vissers MM, Bussmann JBJJ, de Groot IB, et al. Walking and chair rising performed in the daily life situation before and after total hip arthroplasty. *Osteoarthr Cartil*. 2011;19(9):1102-1107. doi:10.1016/J.JOCA.2011.06.004

225. Kersten RFMR, Stevens M, van Raay JJAM, Bulstra SK, van den Akker-Scheek I. Habitual physical activity after total knee replacement. *Phys Ther*. 2012;92(9):1109-1116. doi:10.2522/ptj.20110273

226. Matsunaga-Myoji Y, Fujita K, Ide S, Tabuchi Y, Mawatari M, Makimoto K. Changes in actual daily physical activity and patient-reported outcomes up to 2 years after total knee arthroplasty with arthritis. *Geriatr Nurs (Minneap)*. 2020;41(6):949-955. doi:10.1016/j.gerinurse.2020.07.006

227. Modave F, Guo Y, Bian J, et al. Mobile device accuracy for step counting across age groups. *JMIR mHealth uHealth*. 2017;5(6). doi:10.2196/mhealth.7870

228. Simpson LA, Eng JJ, Klassen TD, et al. Capturing step counts at slow walking speeds in older adults: Comparison of ankle and waist placement of measuring device. *J Rehabil Med*. 2015;47(9):830-835. doi:10.2340/16501977-1993

229. David G Lewallen JAS. Dependence on Walking Aids and Patient-Reported Outcomes after Total Knee Arthroplasty. *J Arthritis*. 2015;04(02):2-6. doi:10.4172/2167-7921.1000149

230. Desmeules F, Dionne CE, Belzile É, Bourbonnais R, Frémont P. Waiting for total knee replacement surgery: Factors associated with pain, stiffness, function and quality of life. *BMC Musculoskelet Disord*. 2009;10(1). doi:10.1186/1471-2474-10-52

231. Pua YH, Seah FJT, Clark RA, Lian-Li Poon C, Tan JWM, Chong HC. Factors associated with gait speed recovery after total knee arthroplasty: A longitudinal study. *Semin Arthritis Rheum*. 2017;46(5):544-551. doi:10.1016/j.semarthrit.2016.10.012

232. Gabr A, Withers D, Pope J, Santini A. Functional outcome of staged bilateral knee replacements. *Ann R Coll Surg Engl*. 2011;93(7):537-541. doi:10.1308/147870811X13137608454803

233. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Jt Surg - Ser B*. 1998;80(1):63-69. doi:10.1302/0301-620X.80B1.7859

234. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy (New York)*. 1990;16(3):199-208. doi:10.1016/0168-8510(90)90421-9

235. Williams DP, Price AJ, Beard DJ, et al. The effects of age on patient-reported outcome measures in total knee replacements. *J Bone Jt Surg - Ser B*. 2013;95 B(1):38-44. doi:10.1302/0301-620X.95B1.28061

236. Amstutz HC, Thomas BJ, Jinnah R, Kim W, Grogan T, Yale C. Treatment of primary osteoarthritis of the hip. A comparison of total joint and surface replacement arthroplasty. *J Bone Jt Surg - Ser A*. 1984;66(2):228-241. doi:10.2106/00004623-198466020-00010

237. Sirard JR, Slater ME. Compliance with wearing physical activity accelerometers in high school students. *J Phys Act Heal*. 2009;6(SUPPL. 1):S148-55. doi:10.1123/jpah.6.s1.s148

238. Gunaratne R, Pratt DN, Banda J, Fick DP, Khan RJK, Robertson BW. Patient Dissatisfaction Following Total Knee Arthroplasty: A Systematic Review of the Literature. *J Arthroplasty*. 2017;32(12):3854-3860. doi:10.1016/j.arth.2017.07.021

239. Lützner C, Beyer F, Kirschner S, Lützner J. How Much Improvement in Patient Activity Can Be Expected After TKA? *Orthopedics*. 2016;39(3):S18-S23. doi:10.3928/01477447-20160509-15

240. Lyman S, Hidaka C, Fields K, Islam W, Mayman D. Monitoring Patient Recovery After THA or TKA Using Mobile Technology. *HSS J*. 2020;16(2):358-365. doi:10.1007/s11420-019-09746-3

241. Morcos MW, Teeter MG, Somerville LE, Lanting B. Correlation between hip osteoarthritis and the level of physical activity as measured by wearable technology and patient-reported questionnaires. *J Orthop*. 2020;20:236-239. doi:10.1016/j.jor.2019.11.049

242. Bradbury KE, Guo W, Cairns BJ, Armstrong MEG, Key TJ. Association between physical activity and body fat percentage, with adjustment for BMI: a large cross-sectional analysis of UK Biobank. *BMJ Open*. 2017;7(3):e011843. doi:10.1136/BMJOPEN-2016-011843

243. Bergmann JHM, McGregor AH. Body-worn sensor design: What do patients and clinicians want? *Ann Biomed Eng*. 2011;39(9):2299-2312. doi:10.1007/s10439-011-0339-9

244. Papi E, Belsi A, McGregor AH. A knee monitoring device and the preferences of patients living with osteoarthritis: A qualitative study. *BMJ Open*. 2015;5(9):1-8. doi:10.1136/bmjopen-2015-007980

245. Papi E, Murtagh GM, McGregor AH. Wearable technologies in osteoarthritis: A qualitative study of clinicians' preferences. *BMJ Open*. 2016;6(1):1-7. doi:10.1136/bmjopen-2015-009544

246. Stockwell S, Trott M, Tully M, et al. Changes in physical activity and sedentary behaviours from before to during the COVID-19 pandemic lockdown: a systematic review. *BMJ Open Sport Exerc Med*. 2021;7(1):e000960. doi:10.1136/BMJSEM-2020-000960

247. Wilson HA, Middleton R, Abram SGF, et al. Patient relevant outcomes of unicompartmental versus total knee replacement: systematic review and meta-analysis. *BMJ*. 2019;364. doi:10.1136/BMJ.L352

248. Witjes S, Gouttebarge V, Kuijer PPFM, van Geenen RCI, Poolman RW, Kerkhoffs GMMJ. Return to Sports and Physical Activity After Total and Unicondylar Knee Arthroplasty: A Systematic Review and Meta-Analysis. *Sport Med*. 2016;46(2):269-292. doi:10.1007/s40279-015-0421-9

249. Kleeblad LJ, Van Der List JP, Zuiderbaan HA, Pearle AD. Larger range of motion and increased return to activity, but higher revision rates following unicompartmental versus total knee arthroplasty in patients under 65: a systematic review. *Knee Surgery, Sport Traumatol Arthrosc*. 2018;26:1811-

1822. doi:10.1007/s00167-017-4817-y

250. Kornuijt A, Das · D, Sijbesma · T, De Vries · L, Van Der Weegen · W. Manipulation under anesthesia following total knee arthroplasty: a comprehensive review of literature. *Musculoskelet Surg*. 2018;102:223-230. doi:10.1007/s12306-018-0537-9

251. Crawford DA, Lombardi A V., Berend KR, et al. Early outcomes of primary total hip arthroplasty with use of a smartphone-based care platform: A prospective randomized controlled trial. *Bone Jt J*. 2021;103-B(7):91-97. doi:10.1302/0301-620X.103B7.BJJ-2020-2402.R1

252. Harmelink KEM, Zeegers AVCM, Tönis TM, et al. The effectiveness of the use of a digital activity coaching system in addition to a two-week home-based exercise program in patients after total knee arthroplasty: Study protocol for a randomized controlled trial. *BMC Musculoskelet Disord*. 2017;18(1):290. doi:http://dx.doi.org/10.1186/s12891-017-1647-5

253. McQuilten ZK, Polizzotto MN, Wood EM, Sundararajan V. Myelodysplastic syndrome incidence, transfusion dependence, health care use, and complications: an Australian population-based study 1998 to 2008. *Transfusion*. 2013;53(8):1714-1721. doi:10.1111/TRF.12054

254. Baryeh K, Maillot C, Gummaraju A, Rivière C. Disappointing Relationship between Functional Performance and Patient Satisfaction of UKA Patients: A Cross Sectional Study. *Orthop Traumatol Surg Res*. 2021;107(3). doi:10.1016/J.OTSR.2021.102865

255. Williams DP, O'Brien S, Doran E, et al. Early postoperative predictors of satisfaction following total knee arthroplasty. *Knee*. 2013;20(6):442-446. doi:10.1016/J.KNEE.2013.05.011