

# UC Davis

## UC Davis Electronic Theses and Dissertations

### Title

Measuring running outside the lab to better predict overuse injury development

### Permalink

<https://escholarship.org/uc/item/9cb9f84g>

### Author

Kiernan, Dovin

### Publication Date

2024

Peer reviewed|Thesis/dissertation

Measuring Running Outside the Lab to Better Predict Overuse Injury Development

By

DOVIN KIERNAN  
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

---

Blaine Christiansen, Chair

---

Susan Stover

---

David Fyhrie

Committee in Charge

2024

## Dedication

The opportunity to attend university and pursue a PhD has been an incredible privilege. I did not believe this path was for me and owe so much to Margaret Little for giving me the encouragement, confidence, and opportunity to follow it. I never would have started this journey without you. To my wife, Lindsay Bowman, I owe my ability to navigate this path to you. You taught me the skills to succeed and continue to be both my most critical editor and my most ardent supporter. I truly could not have done this without you. I have also been incredibly lucky to have many mentors who supported me throughout this journey. Tim Welsh introduced me to the world of research and through this introduction and his constant support kindled my passion for research and made me feel like I could achieve anything. Ross Miller and Jae Shim built on this introduction, shaping me into a biomechanics researcher and acting as my constant champions. And, finally, the people that supported me throughout my PhD, my advisors David Hawkins and Blaine Christiansen and my committee members Susan Stover and David Fyhrie, whose guidance and thoughtful discussion have shaped both this dissertation and me as a researcher.

## Abstract

Between 45-65 million Americans run, making running the most popular physical activity in the US (other than walking). Although running is associated with many health benefits, it also causes as many as 33 running-related injuries (RRIs) per 1000 hours of running. These RRIs cause pain and reduce performance, cause a reduction or cessation of physical activity, necessitate medical visits and expenditures, and cause absenteeism from work, creating a large health and financial burden. Thus, reducing RRI has been a target of much biomechanics research. Unfortunately, this research has not reduced RRI incidence, likely because it has been constrained to lab environments that may misrepresent the number and magnitude of the loads (ground reaction forces) thought to cause RRI. This dissertation overcomes this limitation by using small wearable devices to measure running in real-world settings. First, it presents a first-of-its-kind proof-of-concept study demonstrating that—when measured in the real world—the number and magnitude of loads relates to RRI development. Next, this dissertation presents a series of studies comparing methods to estimate the number and magnitude of loads using small wearable devices then validates these methods across a range of real-world conditions. Finally, applying these validated methods, this dissertation replicates the initial proof-of-concept study in a large heterogenous sample across a longer period of time. Collectively, these studies demonstrate a novel approach to studying running and RRI, provide the tools to execute that approach, and show support for the hypothesis that the magnitude and number of loads experienced by runners relates to RRI development.

## Table of Contents

1. Introduction .....	1
1.1. Physical activity and overuse injury .....	1
1.2. Running-related injury .....	2
1.2.1 Running-related as a useful model of overuse injury .....	2
1.2.2 Popularity of running .....	2
Figure 1.1: US running participation across time.....	3
1.2.3. Health benefits of running .....	3
1.2.4. High incidence of running-related injuries .....	3
Equation 1.1: Conservative estimate of the financial burden of RRI in the US. ....	4
1.3. Previous research on running-related injury .....	4
1.4. A novel approach to studying running-related overuse injuries .....	7
Equation 1.2:.....	7
Figure 1.2: Illustration of the difference between actual values and those predicted with accelerometer-based estimations .....	9
1.5. Potential impacts of this novel approach .....	10
2. Accelerometer-based prediction of running injury in National Collegiate Athletic Association track athletes .....	12
2.1. Introduction .....	12
Figure 2.1: Conceptual representation of changes in a musculoskeletal structure’s strength.....	13
2.2. Methods.....	15
2.2.1. Participants .....	15
Table 2.1: Participant characteristics.....	15
2.2.2. Training prescriptions .....	16
2.2.3. Training questionnaires .....	16
Table 2.2: Pain/fatigue scale.....	16
2.2.4. Accelerometer data .....	16
Figure 2.3: Illustration of the approach used to quantify hip acceleration .....	18
Figure 2.4: Comparison of known and imputed data .....	19
2.2.5. Multiple Imputation.....	19
2.2.6. Analyses .....	20
2.3. Results.....	20
2.3.1. Injury .....	20
2.3.2. Pain vs. injury .....	21

Figure 2.5: Outcome metrics for injured and uninjured runners .....	21
Figure 2.6: Relations between loading profiles, pain, and injury .....	22
2.3.3. Biomechanics vs. injury .....	22
2.4. Discussion.....	23
2.6. Acknowledgements.....	27
2.7. Funding .....	27
3. Unsupervised gait event identification with a single wearable accelerometer and/or gyroscope: A comparison of methods across running speeds, surfaces, and foot strike patterns .....	28
3.1. Introduction .....	28
Figure 3.1: Gait phases.....	29
Table 3.1: Summary of methods to identify gait events from running using data from a single wearable on the shank or sacrum/lower back. ....	32
3.2. Materials & Methods .....	33
3.2.1. IMU calibration .....	33
Equation 3.1 .....	33
3.2.2. Participants .....	33
Figure 3.2: Sample characteristics. ....	34
3.2.3. Protocol.....	34
Figure 3.3: IMU fixation and coordinate system. ....	35
3.2.4. Data processing.....	36
3.2.5. Analysis .....	37
Equation 3.2 .....	37
3.3. Results.....	38
3.3.1. Failure to identify gait events or step side.....	38
Figure 3.4: Percentage of trials per participant without any gait events recognized.....	38
Figure 3.5: Percentage of trials per participant where side was not correctly recognized	39
3.3.2. Initial Contact.....	39
Figure 3.6: IC estimation .....	40
Figure 3.7: IC estimation for methods with LOAs within $\pm 200$ ms .....	41
Table 3.2. Effects of surface, running speed, and foot strike angle on Initial Contact and Terminal Contact estimation .....	41
Figure 3.8: Mean absolute error in IC predicted by mixed effects models for each of the best performing .....	43
3.3.3. Terminal Contact.....	44
Figure 3.9: TC estimation .....	44
Figure 3.10: TC estimation for methods with LOAs within $\pm 200$ ms .....	45

Figure 3.11: Mean absolute error in TC predicted by mixed effects models for each of the best performing methods .....	46
3.3.4. Processing time .....	46
Figure 3.12. Mean time to process 30-min of steady state running data sampled at 1000 Hz .....	47
3.4. Discussion.....	47
3.5. Acknowledgements.....	53
3.6. Funding .....	53
3S: Supplement to chapter 3 .....	54
3S.1 Overview .....	54
3S.1.a Timing Constraints .....	54
3S.1.b. Common output.....	55
3S.2. Shank-mounted wearable methods.....	55
3S.2.1. Mizrahi method.....	55
Figure 3S.1: Mizrahi method.....	56
Figure 3S.2: Mizrahi results.....	56
3S.2.2. Mercer method .....	57
Figure 3S.3: Mercer method. ....	58
Figure 3S.4: Mercer results. ....	59
3S.2.3. Purcell method.....	59
Figure 3S.5: Purcell method.....	60
Figure 3S.6: Purcell results.....	61
3S.2.4. Aminian/O'Donovan method.....	61
Figure 3S.7: Aminian/O'Donovan method.....	63
Figure 3S.8: Aminian/O'Donovan results.....	65
3S.2.5. Aminian/O'Donovan modified method .....	65
Figure 3S.9: Aminian/O'Donovan modified method .....	66
3S.2.6. Greene/McGrath method .....	67
Equation 3S.1 .....	67
Equation 3S.2 .....	67
Equation 3S.3 .....	67
Equation 3S.4 .....	67
Equation 3S.5 .....	67
Equation 3S.6 .....	67
Figure 3S.11: Greene/McGrath method.....	68

Figure 3S.12: Greene/McGrath results.....	69
3S.2.7. Greene/McGrath modified method.....	69
Figure 3S.13: Greene/McGrath modified results.....	70
3S.2.8. Sinclair method .....	70
Figure 3S.14: Sinclair method .....	71
Figure 3S.15: Sinclair results .....	72
3S.2.9. Whelan method .....	72
Figure 3S.16: Whelan method .....	73
Figure 3S.17: Whelan results .....	73
3S.2.10. Norris method .....	74
Figure 3S.18: Norris method .....	74
Figure 3S.19: Norris results.....	75
3S.2.11. Schmidt method.....	75
Figure 3S.20: Schmidt method.....	76
Figure 3S.21: Schmidt results.....	77
3S.2.12. Aubol method. ....	78
Figure 3S.22: Aubol method .....	78
Figure 3S.23: Aubol results .....	79
3S.2.13. Fadillioglu method. ....	79
Figure 3S.24: Fadillioglu method .....	80
Figure 3S.25: Fadillioglu results .....	81
3S.2.14. Bach method. ....	81
Figure 3S.26: Bach method .....	82
3S.2.15. Bach modified method.....	83
Figure 3S.28: Bach results .....	84
3S.3. Lower back-mounted wearable methods.....	84
3S.3.1. Auvinet method. ....	84
Figure S29: Auvinet method. ....	85
Figure 3S.30: Auvinet results .....	86
3S.3.2. Lee method. ....	86
Figure 3S.31: Lee method .....	86
Figure 3S32: Lee results .....	87
3S.3.3. Wixted method. ....	87
Figure 3S.33: Wixted method .....	88
Figure 3S.34: Wixted results. ....	88



3S.3.4. Bergamini method. ....	89
Figure 3S.35: Bergamini method .....	89
Figure 3S.36: Bergamini results .....	90
3S.3.5. Benson method. ....	90
Figure 3S.37: Benson method .....	91
Figure 3S.38: Benson results .....	92
3S.3.6. Reenalda method. ....	92
Figure 3S.39: Reenalda method .....	93
Figure 3S.40: Reenalda results .....	93
3S.4. IMU data processing .....	94
3S.4.1. Calibration .....	94
Equation 3S.7 .....	95
Figure S3.41: IMU calibration .....	95
3S.4.2. Quiet period identification .....	96
Equation 3S.8 .....	96
Equation 3S.9 .....	96
Figure 3S.42: Quiet period identification .....	96
3S.4.3. Bias removal .....	97
Equation 3S.10 .....	97
Equation 3S.11 .....	97
Equation 3S.12 .....	97
Equation 3S.13 .....	97
Equation 3S.14 .....	98
Figure 3S.43: Bias removal .....	98
3S.4.4. Saturation correction .....	98
Figure 3S.44: Saturation correction .....	99
3S.4.5. Low pass filtering .....	99
Figure 3S.45: Low-pass filtering .....	100
3S.4.6. Drift correction .....	100
Figure 3S.46: Drift correction .....	101
3S.4.7. Orientation estimation and gravity subtraction .....	101
Figure 3S.47: Orientation estimation and gravity subtraction. ....	101
3S.4.8. Coordinate system definition .....	102
Figure 3S.48: Coordinate system definition .....	104

4. Acceleration-based estimation of vertical ground reaction forces during running: A comparison of methods across running speeds, surfaces, and foot strike patterns.....	105
4.1. Introduction .....	105
Table 4.1: Published ground reaction force estimation methods. ....	109
Table 4.2: Twenty-seven methods derived or adapted from 13 publications.....	110
4.2. Methods.....	110
4.2.1. IMU calibration .....	110
4.2.2. Participants .....	111
Figure 4.1: Participant details. ....	111
4.2.3. Protocol.....	112
4.2.4. IMU data processing .....	114
4.2.5. Force data processing .....	115
Figure 4.4: Vertical GRF and HiF reconstruction for an example stance .....	115
4.2.6. Analysis .....	116
Equation 4.1 .....	116
Equation 4.2 .....	117
Equation 4.3 .....	117
Equation 4.4 .....	117
Equation 4.5 .....	117
4.3. RESULTS .....	117
4.3.1. First peak.....	117
Table 4.3: Best performing first peak estimation methods .....	119
4.3.2. Loading rate .....	120
Table 4.4: Best performing loading rate estimation methods.....	122
4.3.3. Second peak.....	124
Figure 4.9: Second peak estimation.....	124
Table 4.5: Best performing second peak estimation methods.....	125
Figure 4.10: Bias in second peak estimates predicted by mixed effects models for each of the best-performing methods .....	127
4.3.4. Average force .....	127
Table 4.6: Best performing average force estimation methods .....	128
4.3.5. Time series .....	129
Table 4.7: Best performing force time series estimation methods .....	130
4.4. Discussion.....	132
4.5. Conclusions & practical applications .....	137

4.6. Funding .....	137
4S: Supplement to Chapter 4.....	138
4S.1. Description of force estimation methods and their implementation.....	138
4S.1.1. Neugebauer method .....	139
Equation 4S.1 .....	139
Equation 4S.2 .....	139
Equation 4S.3 .....	139
Equation 4S.4 .....	139
Figure 4S.1: Neugebauer method.....	140
4S.1.2. Charry method .....	140
Equation 4S.5 .....	140
Equation 4S.6 .....	141
Figure 4S.2: Charry method .....	141
4S.1.3. Wundersitz method .....	142
Equation 4S.7 .....	143
Figure 4S.3: Wundersitz method .....	143
4S.1.4. Meyer method .....	143
Equation 4S.8 .....	144
Equation 4S.9 .....	144
4S.1.5. Gurchiek method .....	144
Equation 4S.10 .....	144
4S.1.6. Thiel method .....	145
Equation 4S.11 .....	145
Equation 4S.12 .....	145
Equation 4S.13 .....	145
Figure 4S.4: Thiel method.....	146
4S.1.7. Kiernan method .....	146
Equation 4S.14 .....	147
Equation 4S.15 .....	147
Figure 4S.5: Kiernan method .....	147
4S.1.8. Kim method.....	148
Figure 4S.6: Kim method.....	149
4S.1.9. Pogson method .....	149
Figure 4S.7: Pogson method .....	151
4S.1.10. Pogson xynorm method.....	151

4S.1.11. Day method.....	152
Equation 4S.16 .....	152
Figure 4S.8: Day method.....	153
4S.1.12. Higgins method .....	153
Equation 4S.17 .....	153
Equation 4S.18 .....	154
Equation 4S.19 .....	154
Equation 4S.20 .....	154
Figure 4S.9: Higgins method. ....	155
4S.1.13. Veras method.....	155
Equation 4S.21 .....	156
Equation 4S.22 .....	156
Figure 4S.10: Veras method.....	156
4S.2. Results across wearable, segment, and tilt-corrected coordinate systems .....	157
5. A 0.05 m change in inertial measurement unit placement alters time and frequency domain metrics during running.....	159
5.1. Introduction .....	159
5.2. Methods.....	161
5.2.1. Participants .....	161
5.2.2. IMU placement .....	162
Figure 5.2: Set up and protocol.....	163
5.2.3. Protocol.....	163
Figure 5.3 .....	164
5.2.4. Processing .....	164
5.2.5. Analysis .....	165
5.3. Results.....	166
5.3.1. Acceleration .....	166
Table 5.1: Acceleration differences observed between simultaneously recorded reference and misplaced IMUs. ....	166
Figure 5.4: Root mean square error (RMSE) between $a$ measured simultaneously by the reference IMU and the misplaced .....	167
Figure 5.5: (A) absolute and (B) normalized differences between peak $a$ magnitudes measured simultaneously by the reference IMU and the misplaced IMU .....	168
Figure 5.6: (A) absolute and (B) normalized differences between peak $a$ timings measured simultaneously by the reference IMU and the misplaced IMU .....	169
5.3.2. Angular velocity .....	170

Table 5.2: Angular velocity differences observed between simultaneously recorded reference and misplaced IMUs. ....	170
Figure 5.7: Root mean square error (RMSE) between $\omega$ measured simultaneously by the reference IMU and the misplaced IMU .....	170
Figure 5.8: (A) absolute and (B) normalized differences between peak $\omega$ magnitudes measured simultaneously by the reference IMU and the misplaced IMU .....	171
Figure 5.9: (A) absolute and (B) normalized differences between peak $\omega$ timings measured simultaneously by the reference IMU and the misplaced IMU .....	172
5.3.3. Estimated outcome variables .....	173
Table 5.3: Differences between contact times estimated from simultaneously recorded reference and misplaced IMUs. ....	173
Figure 5.10: Differences in (A) initial contact and (B) terminal contact gait event timings estimated using data from the reference IMU and the misplaced IMU.....	173
Table 5.4: Differences between ground reaction forces estimated from simultaneously recorded reference and misplaced IMUs. ....	174
Figure 5.11: Differences in (A) vertical ground reaction force second peak magnitude, (B) average vertical ground reaction force during stance, and (C) vertical ground reaction force time series during stance .....	174
5.4. Discussion.....	175
5.5. Conclusions .....	179
5.6. Funding .....	180
5S. Supplement to Chapter 5 .....	181
5S.1. Time series plots .....	181
5S.1.1. Shank acceleration .....	181
Figure 5S.1: Shank x-axis accelerations.....	182
Figure 5S.2: Shank y-axis accelerations .....	183
Figure 5S.3: Shank z-axis accelerations.....	184
5S.1.2. Shank angular velocity .....	185
Figure 5S.4: Shank x-axis angular velocities.....	185
Figure 5S.5: Shank y-axis angular velocities.....	186
5S.1.3. Pelvis acceleration.....	188
Figure 5S.7: Pelvis x-axis accelerations .....	188
5S.1.4. Pelvis angular velocity .....	191
5S.1.5. Sacrum acceleration.....	194
5S.1.6. Sacrum angular velocity .....	197
5.S.2. Frequency domain analyses.....	200
5S.2.1. Shank acceleration .....	201

5S.2.2. Shank angular velocity .....	204
5S.2.3. Pelvis acceleration.....	207
5S.2.4. Pelvis angular velocity.....	210
5S.2.5. Sacrum acceleration.....	213
5S.2.6. Sacrum angular velocity.....	216
5.3 Wearable coordinate system results .....	218
Figure 5S.37: Wearable coordinate system root mean square error in acceleration. ....	219
Figure 5S.38: Wearable coordinate system differences between peak <i>a</i> magnitudes... 220	
6. Cumulative loading does not relate to injury development over a 12-week prospective period .....	223
.....	
6.1. Introduction .....	223
6.2. Methods.....	226
6.2.1. Sample size calculations and participants .....	226
Figure 6.1: Participant details .....	227
6.2.2. Baseline data collection .....	227
Figure 6.2: IMU placement and belt design.....	228
6.2.3. Prospective survey data.....	229
6.2.4. Prospective IMU data.....	230
6.2.5. Attrition management .....	231
6.2.6. IMU data processing .....	231
6.2.7. IMU outcome metrics .....	232
Equation 6.1: Allometric scaling. ....	232
Equation 6.2: Cumulative weighted load.....	232
6.2.8. Analyses .....	233
Table 6.1: ‘Risk models’ proposed to predict overuse injury. ....	234
6.3. Results.....	234
6.3.1. Compliance and attrition .....	234
Figure 6.3: Attrition.....	235
6.3.2. Accuracy of self-reported mass .....	235
Figure 6.4: Self-reported vs. true mass at baseline .....	236
6.3.3. Injury .....	236
Figure 6.5: Injury development across time and definitions .....	237
6.3.4. Self-reported outcome variables across groups .....	237
Table 6.2: Self-reported outcome variables across groups. ....	237
Figure 6.6: Self-reported outcome variables across injured and uninjured runners .....	238

6.3.5. IMU-derived outcome variables across groups .....	238
Figure 6.7: IMU-estimated outcome variables across injured and uninjured runners....	239
Figure 6.8: Injury status as a function of load magnitudes and numbers .....	240
6.3.6. Risk models across groups .....	240
Figure 6.9: Risk model scores across injured and uninjured runners .....	240
Figure 6.10: Acute (6 day):Chronic (21 day) cumulative weighted loading over time ....	241
Figure 6.11: Training strain over time.....	241
Figure 6.12: 10-day exponentially weighted moving average over time .....	242
Figure 6.13: Week-to-week change in cumulative loading over time .....	242
6.4. Discussion.....	243
6.5. Conclusions .....	248
6.6. Funding .....	248
6S: Supplement to Chapter 6.....	249
6S.1: Baseline interview.....	249
6S.2: Weekly survey .....	250
6S.3: Weekly running log .....	261
7. Conclusions .....	262
8. References .....	264

# 1. Introduction

## 1.1. Physical activity and overuse injury

Physical activity offers many health benefits that improve quality of life and reduce morbidity and mortality. However, engaging in physical activity can also lead to *musculoskeletal overuse injury* and negative health consequences. Overuse injuries are common across sport, military, and occupational settings. When these injuries occur, they can decrease quality of life, degrade occupational performance, impose medical costs, and initiate a cycle of injury recurrence that ultimately decreases physical activity.

Overuse injuries result from bouts of cyclic submaximal loading of the musculoskeletal system that cause microtrauma accumulation over time. When microtrauma from a given bout of loading is limited and biological structures are allowed adequate time for repair, they can positively remodel, becoming stronger and less susceptible to injury [1, 2]. In contrast, when successive bouts of cyclic loading occur before microtrauma can be repaired, microtrauma accumulation overwhelms repair processes, elicits negative remodeling, and increases injury risk [3, 4]. In single bouts of continuous cyclic loading, the relationship between load magnitude and the number of loading cycles to structure failure can be described by an inverse exponential relationship for soft tissue [5, 6] and an inverse power relationship for bone [7]. Due to this etiology, overuse injuries typically have a gradual onset without a single identifiable precipitating event [8].



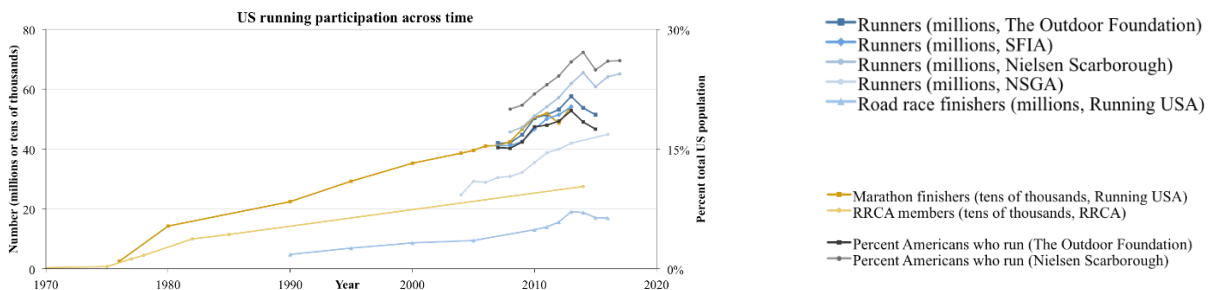
## 1.2. Running-related injury

### 1.2.1 Running-related as a useful model of overuse injury

One type of overuse injury that provides a particularly useful model for study is running related injury (RRI). Running provides a good model for study because it is (1) popular, (2) has high injury incidence, and (3) occurs across discrete bouts with relatively stereotyped repeated movements (cf. for ex. military settings where many different movement patterns could contribute to a single overuse injury).

### 1.2.2 Popularity of running

Recreational and competitive running participation has increased rapidly since the 1970s with 45-65 million Americans now running a total of ~4.7 billion runs a year (Figure 1.1) [9, 10, 11, 12]. These participation rates make running the most popular physical activity in the US (other than walking) in terms of both the number of participants and the number of yearly outings [9, 10]. Participation in running also extends beyond recreational involvement; for example, indoor and outdoor track and field and cross-country have the highest participation rates of any sport at the high school and junior high school levels [13, 14]. This popularity extends across sexes and ethnic/racial groups, and running is relatively accessible to people across a range of socio-economic backgrounds [9]. This broad appeal has resulted in a multi-billion-dollar industry providing runners with shoes, apparel, and – more recently – wearable activity monitors [11, 15, 14].



**Figure 1.1: US running participation across time.** Sources presented here [9, 10, 11, 12, 15, 16, 17] used relatively consistent methodology to sample at multiple time points. Units vary between sources and are presented in millions (blue) and tens of thousands (gold) on the left axis or percent of the total US population on the right axis (dark gray).

### 1.2.3. Health benefits of running

The popularity of running is due, in part, to running being a relatively accessible way to achieve healthy levels of physical activity and cardiovascular health benefits that reduce morbidity and mortality [18]. Indeed, running's popularity can be seen as a boon for public health. Each year physical inactivity kills 3.2 million people and accounts for 36 million years lost to ill-health, disability, and early death world-wide [19]. Countering the effects of inactivity, weight bearing physical activities such as running have demonstrated health benefits across > 25 diseases including major killers such as cardiovascular disease, obesity, diabetes, and some cancers [18, 20, 21, 22, 23, 24].

### 1.2.4. High incidence of running-related injuries

Unfortunately, in spite of the general health benefits of running, it also presents risks: Depending on the sample studied and the definition of "running" and "injury,"<sup>1</sup> between 2.5 and 33 RRIs occur per 1000 hours of running [24, 25]. These RRIs may cause pain and reduce performance, cause a reduction or cessation of physical activity (~50% of runners lose 1-5 weeks of training per year due to RRI), can necessitate medical visits and expenditures (~30% of runners see a doctor for RRI a mean 4 times per year), and can cause absenteeism from work (5% of runners lose time from work each year) [26, 27, 28, 29, 15, 30, 31, 32]. With millions of runners in the US, even the most conservative estimates of RRI incidence indicate there are millions of RRIs per year in the US. Unfortunately, each of these

---

<sup>1</sup> Typically, three different injury definitions are used: (1) any physical complaint/pain, (2) time-loss, and (3) medical attention [119, 26, 334]. These three definitions may capture different information. For example, pain and time-loss definitions appear to represent separate constructs. Time lost is more conservative [26, 27, 115, 114], does not appear related to pain across various RRIs [116], and does not evince the same statistical relationships with other variables [117, 118]. Due to these discrepancies, reported incidences may vary widely based on the injury definition.

injuries comes with an expected monetary cost in the hundreds of dollars [31], creating a large financial burden:

**Equation 1.1: Conservative estimate of the financial burden of RRI in the US.** The average runner reports running > 4 h per week [15], the lowest reported incidence of RRI is 2.5/1000 h [25], costs per RRI are estimated at 210 USD in a nation with socialized health care [31], and at least 45 million Americans run [11].

$$\begin{aligned}4 \text{ h/wk} * 52 \text{ wks/yr} &= 208 \text{ h/yr} \\208 \text{ h/yr} * 2.5 \text{ RRI/1000 h} &= 0.52 \text{ RRI/yr} \\0.52 \text{ RRI/yr} * 210 \text{ USD/RRI} &= 109.2 \text{ USD/yr} \\109.2 \text{ USD/yr} * 45\text{e}^6 \text{ US runners} &= 4.91\text{e}^9 \text{ USD/yr}\end{aligned}$$

To compound this problem, running injuries tend to recur (anywhere from ~20-70% of RRI recur) [33, 34] meaning that these costs are also likely to recur and – once caught in this cycle of injury recurrence – runners may curtail their participation, adding to the already large number of sedentary individuals in the US [30, 35]. Thus, these RRIs create health and economic burdens, and prevent people from maintaining healthy levels of physical activity.

### 1.3. Previous research on running-related injury

Given these figures, preventing RRI is of considerable interest both for its immediate public health and economic implications and for its utility as a model of overuse injury more generally. To that end, much excellent research has been conducted on the causes and prevention of RRI. Unfortunately, despite ~50 years of research examining potential RRI risk factors, causal relationships have been difficult to establish. A host of biomechanical, anatomical, and life history variables have been investigated (see [33, 36, 37] for review and summary), however, no significant reduction in RRI incidence has been achieved [38] (e.g., in a 1992 review, 24-65% annual incidence was reported [37] cf.

[39, 25]). The risk factors of RRI remain difficult to identify [40, 41, 42, 43] and the incidence of RRI remains high suggesting that new strategies are needed [44, 39].

In contrast to the breadth of variables that have been investigated, at the most basic level, animal [45, 46, 47, 48, 49, 50, 51], *in silico* [52, 53, 7, 54], and *ex vivo* [5, 6, 55, 56, 57, 58, 59, 60] evidence suggests that the mechanisms for RRI are bouts of loading with combinations of high load magnitudes and numbers that cause microtrauma accumulation and negative remodeling without sufficient time for repair between bouts [46, 59]. Indeed, while other risk factors may modify the loads a runner generates or their sensitivity to that loading, loading *must* occur to cause injury: A given risk factor can only cause RRI when loading reaches a threshold where it interacts with that factor [61].

Unfortunately, difficulty in measuring the *in vivo* loading of, and trauma to, injury-prone structures in humans [62, 63, 64, 65, 66, 67] has largely prevented researchers from directly investigating links between the internal loading of specific musculoskeletal structures and injury to those structures. To overcome this difficulty, biomechanists often assume that external loads imposed on the body are highly correlated with structure loading and are therefore associated with injury risk. Most notably, ground reaction forces (GRF; external forces imposed on the foot by the ground with each step) are often used as a surrogate measure for internal structure loading because they are the primary external force acting on runners. Unfortunately, evidence relating GRF and RRI is inconsistent and knowing a runner's in-lab GRF appears insufficient to predict and prevent RRIs [68, 41].

One potential reason for this inconsistency is that GRF may not actually correlate well with structure loading: Although GRF contributes to structure loading, there are limitations to the assumption that external and internal loading are directly related. Indeed, relations between external and internal loading are complex and non-intuitive due to inter-individual variability in musculoskeletal forces and morphology [69]. Despite this variability, the loading profile imposed on a musculoskeletal

structure during running depends on the interaction between their structural properties, muscle force profiles, and – critically – the external GRF profile. Thus, it remains reasonable to assume that increased external loading translates into increased internal structure loading and that tracking GRF provides information essential to RRI prevention. Indeed, recent work has demonstrated that GRFs explain 94% of the variance in structure loading at the ankle during level-ground running [70]. Thus, during level-ground running, GRFs may provide a reasonable surrogate of structure loading and injury risk.

Another reason for the inconsistent relation between GRF and RRI may be that biomechanics research has largely been restricted to lab and clinical settings. Thus, inconsistencies in observed relations with RRI may be due, in part, to previous research largely estimating ‘representative’ biomechanics from the mean of a few stances or strides observed in lab settings [71, 72]. Such research is constrained in several ways: (1) biomechanics obtained in labs may not accurately represent biomechanics in the real world [73]; (2) biomechanics may change throughout a repetitive task; for example, fatigue can alter biomechanics across long, exhausting runs [74, 75]; and (3) metrics based on a single representative stride tell an incomplete story that ignores the number of loads actually experienced by runners (a variable theorized to play a critical role in RRI causation [76, 61]) and are thus likely insufficient to predict RRI [77, 72, 78, 79, 80]. Therefore, extrapolation of lab-based, single-stride results may not realistically represent the number or magnitude of loads actually experienced by runners in the real world. Thus, lab-based observations may not be generalizable to real world behavior, undermining the ecological validity of extant findings and potentially explaining the failure to identify the causal factors of RRI [73].

Finally, with few exceptions, the bulk of biomechanics research to-date has been cross-sectional or retrospective. These types of investigations are limited and cannot establish causal relations between biomechanics and injury. Prevention hinges on understanding causal relationships between risk factors

and injury [81] which requires accurately quantifying loading profiles in prospective and ecologically valid designs.

#### 1.4. A novel approach to studying running-related overuse injuries

The difficulty in identifying RRI etiology has led to a realization that new methods are required. Although RRIs are caused by complex multifactorial interactions, recent perspectives advocate that RRIs are always precipitated by an instance in which the combination of load magnitudes and numbers exceed a musculoskeletal structure's capacity [61]. This perspective leads to two conclusions: 1) both load magnitudes and numbers must be quantified as they most directly cause RRI, and 2) it is critical to control for loading even when investigating other risk factors—a given risk factor only causes injury when loading is increased to the point where it interacts with that factor. Thus, researchers interested in RRI must quantify runners' *loading profiles* (load magnitude-number combinations).

To this end, several approaches have been developed to expand on single-stride loading metrics. Notably, '*per-unit-distance*' and '*cumulative loading*' metrics integrate waveform magnitudes within a stance to calculate load, then sum integrals across loading cycles [71, 72, 82, 83, 84, 85].

**Equation 1.2:** Where  $l_c$  is the cumulative load,  $l_i$  is the loading integral at stance  $i$ ,  $x$  is a weighting factor based on human musculoskeletal structure properties, and  $n$  is the total number of stances.

$$l_c = \sum_{i=1}^{i=n} l_i^x$$

These newer *cumulative loading* metrics promise to quantify runners' entire loading profiles, representing much more data than is captured in a single, representative stride. In contrast to single-

stride measures, cumulative loading represents loading profiles across large time periods by integrating waveform magnitudes across a stance to calculate load, then summing loads across multiple stances.

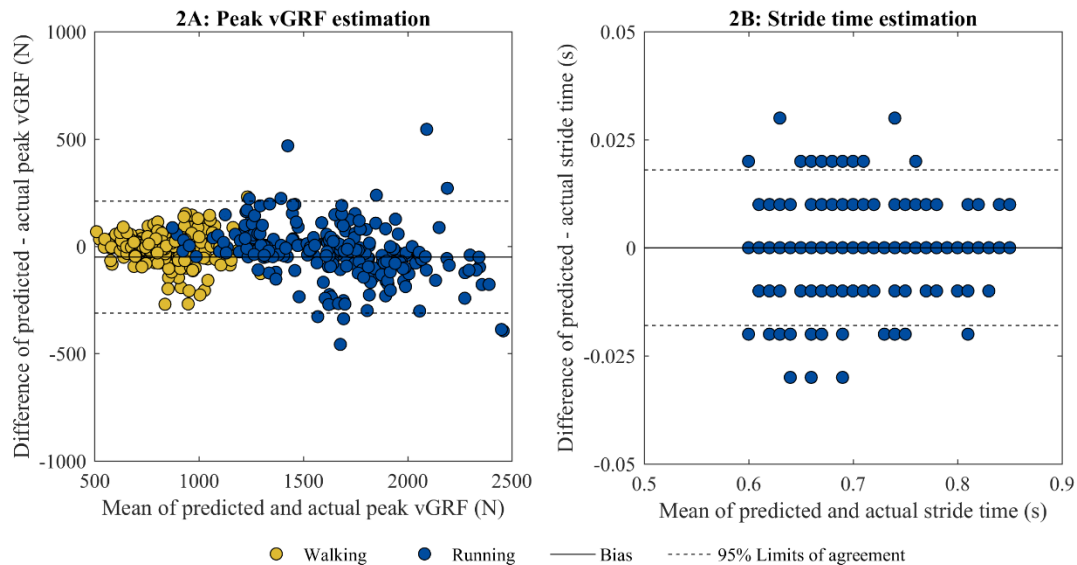
Findings from these studies have shown potentially non-intuitive results not predicted by single-stride metrics. For example, although reducing stride length decreased ankle joint loading during a single stance, across a 5 km run it *increased* cumulative loading [72]. Thus, in agreement with the conceptual model presented here, these novel approaches highlight the importance of broadening the focus from single-stride metrics to include repetitive loading in RRI research. To date, however, this approach has been limited to labs or simulations as there are currently no methods to collect cumulative loading in the field.

In this dissertation, I report a series of studies that advance this novel cumulative loading approach beyond lab and *in silico* settings and into ecologically valid real-world settings. These studies allow quantification of loading profiles over long time periods, in ecologically valid settings overcoming both the limitations of current lab-based single-stride GRF estimation approaches and current limitations of cumulative loading approaches. To do so, these dissertation studies use wearable activity monitors to estimate stance-by-stance external loading, allowing the entire loading profile of runners to be captured in the *real-world*. The assessment of runners' entire loading profiles, calculated from activity in the natural environment, offers ecologically valid data that can be used to quantify RRI risk, and may have tremendous implications for reducing RRI.

The advent and adoption of wearable activity monitors offers a promising way forward. Although commercially available products are not capable of measuring external loading, previous work from the UC Davis Human Performance Lab has demonstrated the potential for wearable activity monitors to collect novel and meaningful loading metrics. These studies have pioneered methods that overcome the limitations of the lab, estimating external loading from wearable activity monitors [86,

87]. In these studies, participants wore hip-mounted activity monitors and ran over-ground, while acceleration (from the activity monitor) and force (from force plates) were collected simultaneously. Vertical ground reaction force (vGRF) and acceleration peaks for the stance ipsilateral to the activity monitor were used to develop models that estimate peak ipsilateral vGRF from peak vertical acceleration. Using separate subsets of participants, the accelerometer-based vGRF estimation models were validated and showed impressive agreement with the ‘gold-standard’ force-plate-observed vGRF peaks: A model for youth estimated peak vGRF within 5.2% of observed vGRF values [86] while a model for adults estimated peak vGRF within 8.3% of observed values [87]. The validity of these models suggests external loading can be accurately estimated from wearable activity monitors, allowing calculation of loading profiles over long time periods in real-world settings.

Other work has shown that the number of loading cycles can be estimated from wearables. For example, using a sacrum-mounted activity monitor, Lee et al. [88] measured stride, step, and stance times during running with biases  $\leq 1$  ms.



**Figure 1.2: Illustration of the difference between actual values and those predicted with accelerometer-based estimations. (A)** Peak vGRF was estimated using multiple linear regression with factors of peak vertical acceleration, mass, and type of locomotion (walk/run). The regression equation estimated peak vGRFs within  $8.3 \pm 3.7\%$  of force plate-measured values (mean  $\pm$  SD) [87]. **(B)** Sacrum accelerations were used to estimate stride times with biases  $< 1$  ms when compared to motion capture (adapted with permission, Lee et al. [88], adapted with permission).



In this dissertation, I expand on these approaches in a series of five studies: First, using currently available methods to estimate stance and force, I conduct a proof-of-concept in a small sample of elite runners showing that cumulative loading relates to injury. Second, I validate and compare extant methods for estimating the number of loads (stances or strides) from a wearable accelerometer across a broad range of conditions and a large heterogeneous sample. Third, I validate and compare extant methods for estimating the external loads from a wearable accelerometer across a broad range of conditions and a large heterogeneous sample. Fourth, I quantify the potential impact that misplacement of a wearable sensor can have on estimated numbers and magnitudes of loads. Fifth, and finally, I build on the initial proof-of-concept by applying the improved methods discovered in studies two and three to a larger, more heterogeneous sample over a longer period of time.

By combining these approaches to estimate number and magnitude of loads experienced by runners in the real world, the current dissertation overcomes critical barriers of lab-based research. This novel approach allows us to test the hypothesis that GRF magnitudes and numbers measured in the real world can be used to calculate cumulative loads that predict RRI.

### **1.5. Potential impacts of this novel approach**

In comparison to capturing single representative stances in-lab, these approaches will better capture the repetitive loading that leads to overuse injury and the true range and variance of GRF magnitudes and numbers both within- and between-runners. If no differences are observed between injured and uninjured runners, it would suggest that GRF measurements are insufficient to predict injury even when accounting for the repetitive loading that causes overuse injury and collecting data in real-world settings. This finding would support the contention that GRFs are not well correlated with the

loading of injury prone structures and suggest that new methods to measure—or new surrogates of—structure loading are needed. If, however, significant differences between injured and uninjured runners are observed, it would suggest that GRFs—when measured appropriately—can be used to predict RRI. Such findings would allow us to establish injury risk thresholds and explore the relative importance of GRF numbers and magnitudes to injury development. Ultimately, positive findings could be used to monitor runners' loading profiles and provide feedback to prevent RRI by avoiding 'high risk' running behaviors.

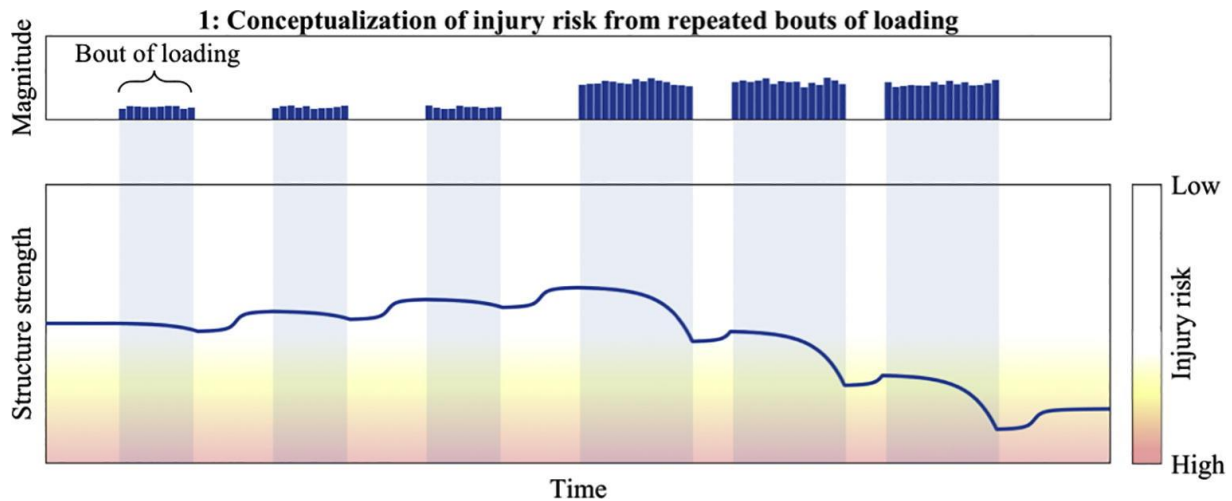
## 2. Accelerometer-based prediction of running injury in National Collegiate Athletic Association track athletes

### 2.1. Introduction

Long distance runners experience high rates of running-related injury (RRI), musculoskeletal overuse injuries causing a restriction of running speed, distance, duration, or frequency [77]. Depending on the population studied and the methods used to diagnose injury, incidences range from 6.9 to 92.4% per 1000 hours of running [89, 39, 44]. In addition to negatively affecting performance, these RRIs lead to both direct (e.g., health care), and indirect (e.g., time lost) costs [31]. With approximately 51.5 million Americans running [9] these injuries constitute a large health and economic burden. Thus, developing means to predict and prevent RRI can have meaningful health and economic impacts.

Overuse injuries such as RRI result from bouts of cyclic loading that cause microtrauma accumulation over time. When microtrauma from a given bout of loading is limited and biological structures are allowed adequate time for repair, they can positively remodel, becoming stronger and less susceptible to injury (e.g., [1, 2]). In contrast, when successive bouts of cyclic loading occur before microtrauma can be repaired, microtrauma accumulation overwhelms repair processes, elicits negative remodeling, and increases injury risk (Fig. 2.1) [3, 4, 52, 53]. In single bouts of continuous cyclic loading, the relationship between load magnitude and the number of loading cycles to structure failure can be described by an inverse exponential relationship for soft tissue [5, 6] and an inverse power relationship for bone [7]. In running, vertical ground reaction forces (vGRFs) are often used as surrogate measures for structure loading. As discussed by Miller and Hamill [69], however, relationships between external and internal loading are complex and non-intuitive due to inter-individual variability in muscle forces and structure morphology. Despite this variability, vGRF is the primary external force acting on runners

and it is reasonable to assume that increased external loading translates into increased structure loading. Thus, the mechanisms for RRI can be conceptualized as a combination of vGRFs (load magnitudes) and strides (load cycles) that cause microtrauma to a musculoskeletal structure, with insufficient time for recovery between runs (bouts of loading) [90, 91, 92, 77, 76].



**Figure 2.1: Conceptual representation of changes in a musculoskeletal structure's strength** elicited by the interaction between the number of cycles in a bout of activity, the magnitude of loading in the bout, and the time between bouts. During the first three bouts of loading, the number and magnitude of loading cycles are low and adequate time is allowed for positive remodelling; thus, the structure's strength increases and the likelihood of injury decreases. In contrast, during the last three bouts, the number and magnitude of loading cycles increases and insufficient time between bouts is allowed for positive remodelling; thus, the structure's strength decreases and injury risk increases.

In spite of this theoretical rationale and much high-quality research, findings relating vGRF and RRI are inconsistent (e.g., [68, 41]). This inconsistency may be due in part to previous research largely focusing on relationships between RRI and load magnitudes determined from single representative stances or strides observed in laboratory settings [71, 72]. Such research is constrained in several ways: (1) loading data obtained in laboratories may not accurately represent loading in the field [73]; (2) loading may change throughout a repetitive task; for example, fatigue can alter biomechanics across a long, exhausting run [74, 75]; and (3) given the importance of repetitive loading in overuse injury, metrics based on a single representative stride tell an incomplete story and are likely insufficient to

predict RRI [78, 79, 80, 77, 72]. Therefore, extrapolation of lab-based, single-stride results may not realistically represent the number or magnitude of loads actually experienced by runners.

Several approaches have been developed to expand on single-stride metrics. Edwards et al. [53] proposed a 'stressed-life' method in which the probability of positive and negative remodeling, and consequent tibial stress fracture risk, were estimated as a function of bone strain, number of strides, and structure adaptation. Although extremely useful in elucidating the role that repeated loading plays in RRI, modeling demands and assumptions limit this approach. More easily applied 'per-unit distance' and 'cumulative loading' metrics integrate waveform magnitudes within a stance to calculate load, then sum integrals across loading cycles [71, 82, 83, 72, 84, 85]. Findings from these studies have shown potentially non-intuitive results not predicted by single-stride metrics. For example, although reducing stride length decreased ankle joint loading during a single stance, across a 5 km run it increased cumulative loading [72]. Thus, in agreement with theory, these novel approaches highlight the importance of broadening the focus from single-stride metrics to include repetitive loading in RRI research.

Wearable activity monitors provide the opportunity to broaden the focus from lab-based single-stride metrics. These devices may be capable of non-invasively capturing loading profiles (load magnitude-number combinations) throughout entire runs in the field. For instance, to obtain ecologically valid estimates of peak vGRFs outside the lab, Neugebauer et al. [86, 87] simultaneously collected accelerations from hip-mounted activity monitors and GRFs from force plates. With a hold-back procedure, one group of participants was used to develop a multiple linear regression relating peak vertical acceleration, mass, and type of locomotion (walk/run) to log-transformed peak vGRF. A second group of participants was used to validate that this equation estimated peak vGRFs within  $8.3 \pm 3.7\%$  of force plate-measured values (mean  $\pm$  SD) (Fig. 1.2A). Other research groups have successfully used activity monitors to measure temporal parameters during running (e.g., [93, 94, 95, 96, 97]). For

example, using a sacrum-mounted activity monitor, Lee et al. [88] measured stride, step, and stance times during running with biases  $\leq 1$  ms (Fig. 1.2B).

Here, we combined accelerometer-based methods developed to estimate stride parameters [88] and peak vGRF [87] and quantified loading profiles throughout entire runs of elite track athletes. We quantified both traditional metrics (number of strides, peak vGRF magnitudes) and a cumulative loading metric and evaluated their relationships to RRI. We hypothesized that runners exhibiting large cumulative loading based on combinations of high peak vGRF and/or high numbers of strides were more likely to develop RRI.

## 2.2. Methods

### 2.2.1. Participants

Ten National Collegiate Athletic Association Division I runners competing in distance events were recruited from the University of California Davis Men's Track Team. One participant was excluded from analyses due to non-compliance, resulting in a final sample of nine (Table 2.1). Participants were excluded if they suffered a major injury within the past 12 months (physician diagnosed injury preventing training for  $\geq 2$  months) or a minor injury with a return-to-training less than six weeks before the study (trainer or self-diagnosed injury preventing training for  $\geq 1$  month). The University of California Davis Institutional Review Board approved all procedures and participants provided written informed consent.

**Table 2.1: Participant characteristics** (mean  $\pm$  SD). Injured participants were significantly older than uninjured participants (\*p = 0.02) but were otherwise statistically similar (ps > 0.05).

	n	Age (years)*	Weight (N)	Height (cm)
All participants	9	18.7 $\pm$ 1.0	629.37 $\pm$ 71.35	178.4 $\pm$ 4.6
Uninjured participants	6	18.2 $\pm$ 0.4	599.23 $\pm$ 64.0	178.3 $\pm$ 4.1
Injured participants	3	19.7 $\pm$ 1.2	689.64 $\pm$ 44.15	178.6 $\pm$ 6.4

### 2.2.2. Training prescriptions

Data were collected during training runs over a 60-day period during the Track and Field season. Coaches provided schedules prescribing daily training for individual participants. Running prescriptions for each training session were coded based on prescribed training time and optional training time (scale variables), whether the training was middle- or long-distance, self-supervised, contained high intensity portions (e.g., sprints), included a pre-race routine, and/or included any otherwise undescribed training (categorical variables).

### 2.2.3. Training questionnaires

After each training session, participants rated their overall pain/fatigue (Table 2.2), reported whether they completed the prescribed training session, and, if they did not, described the reason. If failure to complete training was due to RRI, participants provided details on injury location and severity.

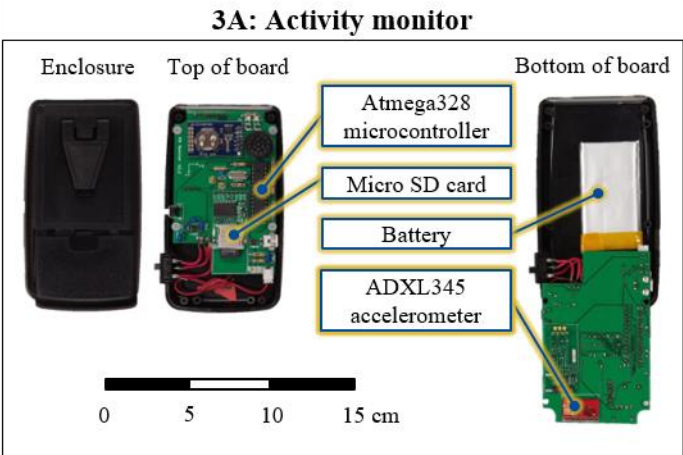
**Table 2.2: Pain/fatigue scale.** After each training session, participants rated overall musculoskeletal pain/fatigue on a scale of integers from 1 to 9.

Quantitative Rating	Qualitative Description
1	No pain or fatigue. <i>Your muscles/bones feel as though they are at optimal training levels.</i>
3	Minimal pain or fatigue. <i>You feel slightly less than optimal, but you still feel as though you can complete a rigorous workout.</i>
5	Moderate pain or fatigue. <i>You have noticeable pain/fatigue in your muscles/bones, but feel as though you could complete an average workout.</i>
7	High pain or fatigue. <i>You have a significant level of pain/fatigue and feel as though you would have difficulty completing an average workout.</i>
9	Extreme pain or fatigue. <i>You have extreme pain/fatigue in your muscles/bones and do not feel as though you could complete a workout.</i>

### 2.2.4. Accelerometer data

Each participant was assigned an activity monitor with a tri-axial linear accelerometer (Fig. 2.3A) (ADXL345, Analog Devices, Norwood, MA;  $\pm 8$  g; 48.4-52.4 Hz, twice the frequency observed in vGRFs during running; [98]). Participants were instructed on activity monitor use and wear. Before each

training session, participants fixed the monitor to their lateral right iliac crest with a neoprene belt and turned it on (Fig. 2.3B and 2.3C). Participants wore the monitor throughout the entire training session, turning it off and automatically generating a time-stamped data file after the session was complete. Each week, researchers collected data and ensured monitor function.

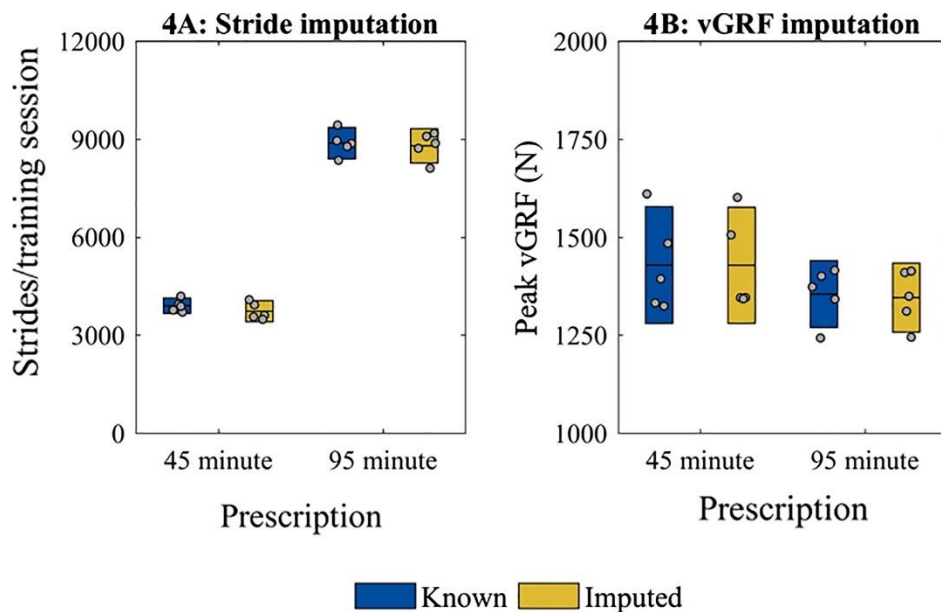




**Figure 2.3: Illustration of the approach used to quantify hip acceleration.** The activity monitor including accelerometer, battery, microcontroller, and memory card (A) was placed in a neoprene belt (B) and secured to the runner's lateral right iliac crest (C).

Custom MATLAB scripts (R2016a, The MathWorks, Natick, MA) were used to extract unfiltered accelerometer data. The DC component was calculated during a 30 s static period and subtracted from the signal, and monitor-specific calibrations were applied, yielding output in “gs” relative to a reference frame aligned with the activity monitor housing and participant [99]. The anterior-posterior axis was used to identify right and left foot strikes [88]. Left stances (contralateral to the monitor) were discarded. Peak vertical accelerations during right stances were extracted and entered into a regression equation to estimate right stance peak vGRF [87]. Data were collected throughout an entire training session and could include warm up drills, stretching, breaks in running to wait at traffic lights, etc. Thus, published stride times and peak vGRF magnitudes were used to eliminate data  $\pm 3$  SD outside the ranges expected for running [100, 101, 102, 103, 104, 105, 106, 107, 75]. Periods evincing vGRF magnitudes and stride times within expected ranges but with  $< 10$  consecutive strides were also eliminated to ensure participants had achieved steady running speeds and that pattern recognition algorithms were not biased by aberrant waveforms. Periods of running with  $\geq 10$  strides were concatenated for analysis.

Mean peak vGRF and total number of strides were calculated for each training session. Data were grouped by training prescription and evaluated for outliers exceeding  $\pm 2$  SD of prescription mean number of strides (e.g., Fig. 2.4A). High outliers were considered real and complete data but misrepresentative of the prescription (i.e., the participant violated the coach's instructions and ran longer than prescribed); thus, prescription data were removed to avoid biasing imputation (see below; 1.67% of data). Low outliers were considered potentially incomplete data collections (i.e., the accelerometer turned off during data collection); thus, peak vGRF and number of strides were deleted and imputed (see below; 2.15% of data).



**Figure 2.4: Comparison of known and imputed data** for number of strides and peak vGRF. Mean (black line) and 95% CI (colored bar) of (A) number of strides, and (B) estimated peak vGRF for known and imputed data from two representative training prescriptions: a 95-min training session and a 45-min training session with high intensity portions. Note that consistent with expectations there are (1) no significant differences between known and imputed data, (2) significantly more strides for the longer prescribed training session, and (3) a trend to higher vGRF for the shorter prescribed training session with high intensity portions. Gray dots represent individual data points (training sessions).

### 2.2.5. Multiple Imputation

A total of 419 training sessions were prescribed, however, item non-response (e.g., the participant forgot to turn on the accelerometer or fill out the questionnaire) or outlier deletion caused the loss of 22.4% of accelerometer and 43.4% of pain/fatigue data. Consistent with recommendations for the use of accelerometers in measuring physical activity [108, 109], missing data were multiply imputed using prescribed training, participant anthropometrics, and accelerometer variables, and assuming that data were missing at random. Multiple imputation uses associations with observed data to generate multiple plausible values for each missing data point. Each of these multiple plausible data sets is then separately analyzed and analyses are pooled. This procedure minimizes bias and results in valid statistical inferences that reflect the uncertainty due to missing values. Here, 50 imputed data sets were generated with SPSS (v24.0, IBM Corp., Armonk, NY) and pooled for analysis using Rubin's rules

[110]. To evaluate imputation accuracy, a second imputation was conducted where known data were deleted from two representative prescriptions. Imputed values for deleted data were not significantly different than original data (evaluated with uncorrected paired t-tests with significance set at  $p < 0.05$ ) (Fig. 2.4).

#### 2.2.6. Analyses

After imputation, mean estimated peak vGRF and number of strides per training session were calculated for each participant. Based on the S-N curve of tendon, an 'effective' load for soft-tissue injury was calculated by weighting the peak vGRF to the 9th power [84]. Weighted peak vGRFs were summed across each training session and a mean weighted cumulative load per training session was calculated. Participants were separated into injured (missed training due to self-reported RRI) and uninjured (no missed training due to RRI) groups. Injured and uninjured mean pain/fatigue, mean estimated peak vGRF, mean strides per training session, and mean weighted cumulative loading per training session were entered into independent samples t-tests. Significance was set at  $p < 0.05$  and corrected with a False Discovery Rate procedure [111].

### 2.3. Results

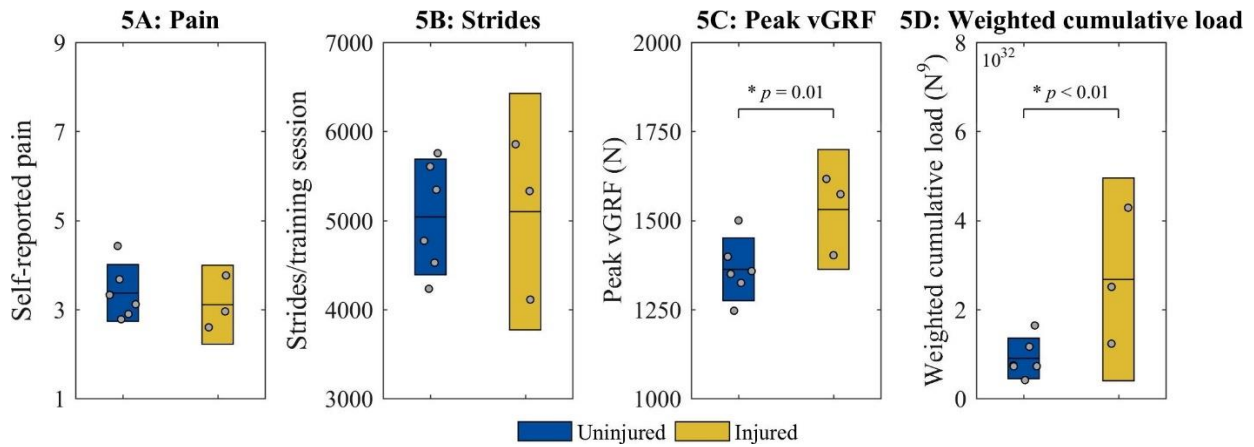
#### 2.3.1. Injury

Across the 60-day study, three participants (33%) lost training time to self-reported foot, adductor, and hamstring soft-tissue RRIs with 7, 10, and 33 training days lost respectively. Injured participants were significantly older than uninjured participants ( $p = 0.02$ ) but did not significantly differ in height or weight (evaluated with uncorrected independent samples t-tests with significance set at  $p < 0.05$ ) (Table 2.1).

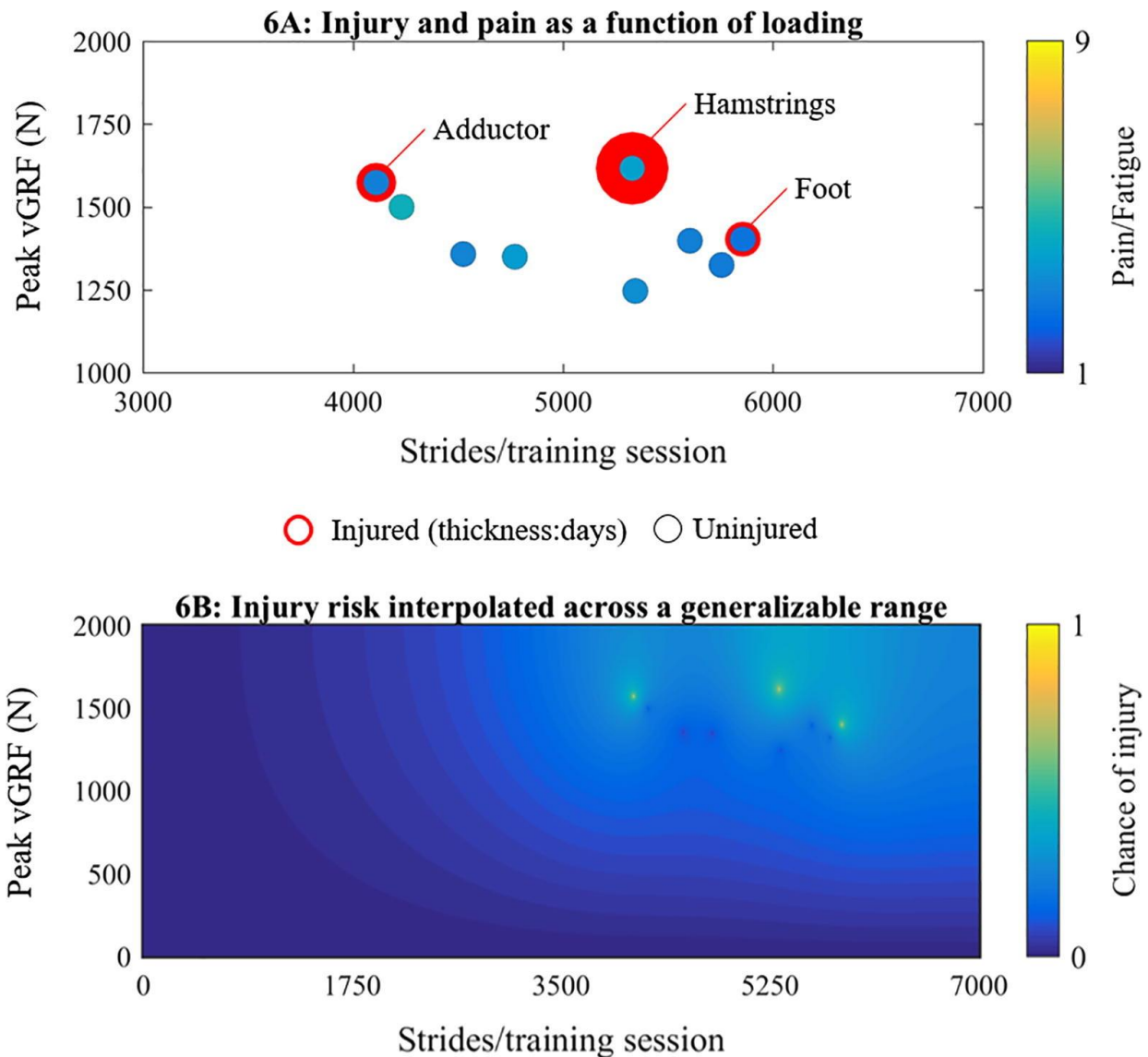
### 2.3.2. Pain vs. injury

Injured and uninjured participants did not report significantly different pain/fatigue ( $p = 0.56$ ).

This result suggests that high subjective pain/fatigue may not predict impending RRI and highlights the need for additional objective metrics (Fig. 2.5A; Fig. 2.6A).



**Figure 2.5: Outcome metrics for injured and uninjured runners.** Mean (black line) and 95% CI (colored bar) of (A) pain, (B) mean number of strides per training session, (C) mean estimated peak vGRF, and (D) mean weighted cumulative load (sum of estimated peak vGRF weighted to the ninth power) per training session. Gray dots represent individual data points (participant means).



**Figure 2.6: Relations between loading profiles, pain, and injury.** (A) Each participant's mean estimated peak vGRF vs. strides per training session. Color represents mean reported pain/fatigue. Thickness of encircling red line represents time lost due to injury (if any). (B) Empirical data coded as 0 for uninjured and 1 for injured and interpolated to conceptualize injury risk across a generalizable range of peak vGRF and strides/training session. Color represents chance of injury over two-months on a scale from 0 (not predicted to incur injury) to 1 (predicted to incur injury).

### 2.3.3. Biomechanics vs. injury

Mean number of strides per training session did not differ between injured and uninjured participants ( $p = 0.91$ ) (Fig. 2.5B). Injured participants did, however, have significantly greater mean estimated peak vGRF ( $p = 0.01$ ) (Fig. 2.5C) and mean weighted cumulative loading per training session ( $p < 0.01$ ) (Fig. 2.5D). Injured runners also appeared as outliers when plotted on a mean estimated peak

vGRF vs. mean strides per training session graph, suggesting greater injury risk at combinations of high loads and magnitudes (Fig. 2.6A). Participant's chance of injury across two months was coded as 1 for injured and 0 for uninjured then interpolated across a generalizable range of peak vGRF and stride number combinations. This process resulted in a contour pattern similar to the curves of Davis'/Wolfe's Laws [112, 113] (Fig. 2.6B), illustrating that the risk of injury increases when high load magnitudes are repeatedly applied.

## **2.4. Discussion**

In this study we used wearable activity monitors to prospectively measure runners' loading profiles in the field. We found that injured runners had higher estimated peak vGRFs and weighted cumulative loads than uninjured runners. To our knowledge this is the first study to (1) apply accelerometer-based estimates of peak vGRF and stride number to RRI prediction in a prospective sample, and (2) empirically validate cumulative loading metrics in RRI prediction. These novel results suggest that the accelerometer-based models used here are capable of capturing inter-participant differences in loading profiles that may be predictive of RRI. These findings support both popular and emerging theories in RRI research: namely, that high vGRF magnitudes may be a contributing factor in RRI as speculated by Cavanagh & LaFortune [100] and others, and that cumulative loading metrics may be a valid index of RRI risk (e.g., [83]). Further, the methods used here show promise in identifying safe limits of loading that could allow coaches and athletes to develop and adapt training prescriptions based on individual athlete loading profiles.

The development of objective metrics similar to those presented here appears critical given the high incidence of RRI [89, 39, 44] and apparent disconnect between subjective pain/fatigue and RRI. Although pain has been used to define RRI elsewhere, when RRI is instead defined as training time lost,

these variables appear to represent separate constructs. Indeed, time lost appears to be a much more conservative definition of injury [26, 114, 27, 115], does not appear related to pain across various RRIs [116], and does not evince the same statistical relationships with other variables as pain [117, 118]. Similarly, the present finding that injured runners did not report higher pain/fatigue, suggests that runners may be insensitive to impending injury. Indeed, it should be noted that injured runners actually tended to report lower pain than uninjured runners (mean 3.4 uninjured vs. 3.1 injured). Given the limited power in the present analysis and the risk of Type II error, however, we cannot rule out that sensitivity to pain may be protective, allowing runners to adapt motor patterns and avoid injury. In any case, subjective pain alone appears insufficient to predict time-loss injuries, highlighting both the important role that providing biomechanics-based feedback about RRI risk could play and the importance of discretely reporting injury as operationalized by medical, time-loss, and pain definitions to facilitate comparison across studies [119].

The activity monitor-derived objective metrics used here address several constraints in current biomechanics research. In the past, biomechanics research has largely been restricted to lab and clinical settings. This restriction has undermined the ecological validity of findings given lab-based observations may not be generalizable to real world behavior [73]. Further, the constraints of the lab have prevented accurate measurement of the number of loads actually experienced by runners – a variable theorized to play a critical role in RRI causation [76, 61]. The current results join an emerging body of literature (e.g., [120, 120, 121, 122, 123, 124]) that demonstrates the role activity monitor-based methods may play in overcoming previous constraints and collecting large ecologically valid data sets.

The results presented here were largely consistent with our hypotheses; however, the number of strides per training session did not significantly differ between injured and uninjured participants. Participants in the present study were elite athletes with training regimes highly constrained by coach's prescriptions. Thus, all participants ran for similar amounts of time and there was relatively little inter-

participant variability in the number of strides. It seems likely that in an unconstrained population, such as recreational runners, the number of strides completed in each training session would vary greatly; in which case, the number of strides might play a larger role in RRI. Further research with more heterogeneous samples is required to more fully investigate the role of stride number and cumulative loading in RRI.

The homogeneity of the present sample was, however, beneficial in constraining the time between training sessions. As outlined in the introduction, repeated bouts of loading play a key role in RRI causation, since adequate time is required between bouts for positive remodeling of injury-prone structures (Fig. 2.1). In spite of this key role, the present analyses focus on metrics across a mean training session, largely ignoring the role of repeated bouts of loading. We feel this decision is justified given the highly constrained training prescriptions in our sample: injured and uninjured participants trained 85.8% and 85.6% of potential days respectively, and times between training sessions were extremely similar. Thus, variability in the number of, and time between, bouts likely has little impact on the present results, and taking means across training sessions captures critical differences. In a more variable sample, however, it seems likely that the time between bouts could play a large role in RRI. This role could be captured either by identifying critical time periods over which load accumulates to cause RRI (e.g., the ratio of workload across one week relative to workload across four weeks predicts non-contact injury [125]; though this method is controversial; cf. [126]), by calculating a daily probability of injury in a given structure based on the probability of positive and negative remodeling as a function of estimated structure loading, number of strides, and structure adaptation (e.g., [53]), or by applying Miner's Rule to calculate cumulative damage (e.g., [85]). Enacting these potential methods requires further research to determine mathematical associations between external loading conditions, internal structure loading, microtrauma accumulation, the temporal healing response, and RRI.



The present investigation represents a preliminary proof-of-concept application of novel methods. Although we observed promising results, there are a number of methodological refinements that could improve the techniques used here. For example, some evidence suggests the magnitude and/or rate of loading associated with the first peak of the vGRF waveform is more predictive of RRI relative to the magnitude of the second peak (e.g., [68, 41] cf. [127, 128, 129]). The method used here is unable to discriminate between the two peaks; rather, absolute peak acceleration is used to estimate absolute peak vGRF. Further, previous studies quantifying cumulative loading have summed waveforms integrated across stance [71, 82, 83, 72, 84, 85]. Methodological limitations prevented the use of an integral in the current study; instead, we summed peak vGRFs. Though this method captures information about the number and maximum magnitude of loads, potentially important information about loading throughout the entire gait cycle and/or the load duration is lost. Finally, given the preponderance of unilateral RRIs [39] and the potential role bilateral asymmetries play in RRI (e.g., [130]), methods to calculate loading and RRI separately for each limb should be developed. These, and other, methodological improvements should be considered as the field of activity monitor-based biomechanics develops.

In sum, we observed significantly different loading profiles between injured and uninjured track runners across a 60-day prospective period. These promising preliminary results provide evidence that further work in this area is warranted. Follow up studies should build on the current methods and collect data from larger, more heterogeneous, samples. Calculating critical time periods over which load accumulates to cause injury and/or calculating rolling probabilities of injury based on loading may lead to the development of thresholds for RRI based on athlete-specific loading histories. Such results would help refine injury prediction models and provide the evidence necessary to develop adaptive feedback and training prescriptions that account for the mechanics and loading profiles of individual runners.

## **2.6. Acknowledgements**

Thanks to Coach Drew Wartenburg and the UC Davis Men's NCAA Track and Field team for their participation, and to Henry Luk, Amar Dholakia, and Shonit Sharma for contributions to earlier work.

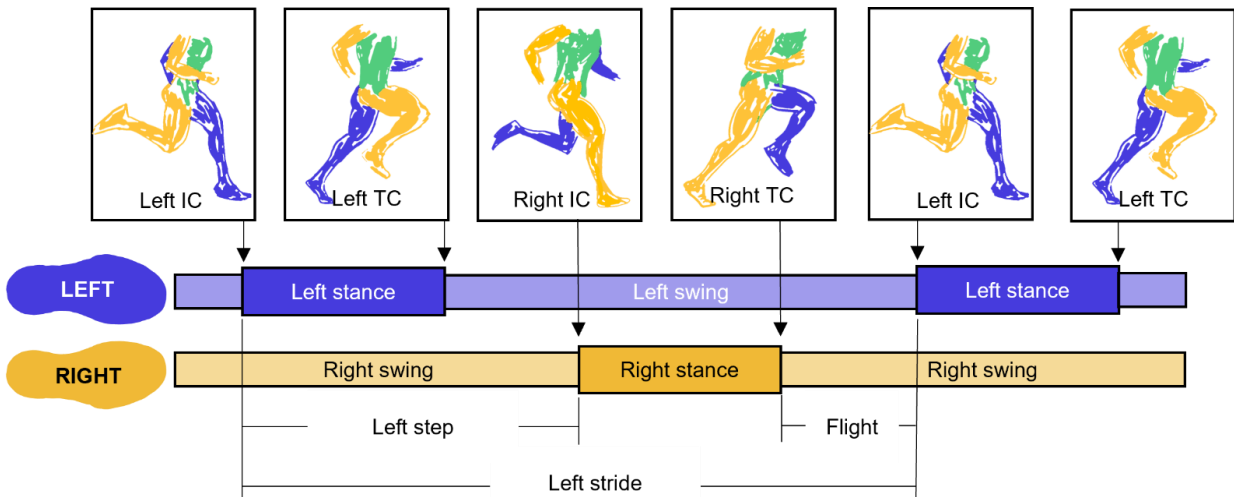
## **2.7. Funding**

DK was supported by a scholarship from the Natural Sciences and Engineering Research Council of Canada (CGSD3-502358-2017). CLC was supported by an Achievement Rewards for College Scientists (ARCS) scholarship. MACM and MM were supported by the UC Davis College of Biological Sciences Dean's Mentorship Award. Supporting agencies had no involvement in the design, collection, or interpretation of this study.

### **3. Unsupervised gait event identification with a single wearable accelerometer and/or gyroscope: A comparison of methods across running speeds, surfaces, and foot strike patterns**

#### **3.1. Introduction**

In this paper we explore the accuracy and precision of different methods to identify gait events during human running using a single wearable device (accelerometer, gyroscope, or inertial measurement unit). Identifying when the foot contacts and leaves the ground—the initial contact (IC) and terminal contact (TC) gait events, respectively—allows biomechanists to break the cyclic movements of running into discrete phases. Identifying IC and TC events allows the analysis of basic but highly useful temporal variables like stride (right IC to right IC or left IC to left IC), step (right IC to left IC or left IC to right IC), stance/contact (right IC to right TC or left IC to left TC), swing (right TC to right IC or left TC to left IC), and flight/float (right TC to left IC or left TC to right IC) times and frequencies (1/time) (Figure 3.1). These basic temporal variables are critical in the analysis and prediction of running speed and performance [131, 132, 133, 134]. Segmenting running into these discrete phases also forms the starting point for many more advanced biomechanical analyses that quantify and compare kinematic or kinetic features' means and variances. Thus, accurately identifying IC and TC events is a critical first step in the analysis and evaluation of running.



**Figure 3.1: Gait phases** defined by initial contact (IC) and terminal contact (TC) gait events. Left arm and leg represented in blue; right arm and leg represented in orange; trunk represented in green.

IC and TC events can be accurately identified with relative ease and reliability in labs using ground truth force plates or instrumented treadmills. Further, several methods have been developed to identify these gait events from lab-based motion capture [135, 136, 137, 138]. However, lab settings impose constraints on data collection: Force plates, instrumented treadmills, and motion capture systems require participants to visit a biomechanics lab, are expensive, and require training to operate, process, and analyze. These ‘captive’ systems may also cause participants to alter their gait (e.g., treadmill or short running track) and may limit the volume of data collection to just a few gait cycles [139, 140, 141, 142]. Thus, ‘captive’ systems can only provide ‘snapshots’ of data that may decrease ecological validity, limiting biomechanists’ ability to monitor changes over time or investigate phenomena that occur across many gait cycles such as fatigue, movement variability, or overuse injury [143]. Advances in technology such as instrumented insoles have allowed biomechanists to overcome some of these constraints but suffer from their own issues with durability and comfort [144]. Consequently, researchers remain largely reliant on ‘captive’ technology, decreasing the volume and ecological validity of running data available for analysis.

The emergence of small, low-cost accelerometers, gyroscopes, and inertial measurement units (IMUs) (collectively called ‘wearables’ for the remainder of this paper) may offer a new way forward [145]. These portable sensors can be used by researchers, clinicians, coaches, and athletes to continuously record data during prolonged runs in natural environments. Wearables have high rates of adoption in runners [15] and have already demonstrated accuracy in measuring or estimating several biomechanical parameters including gait events [146, 145, 147]. Based on these promising findings, our objective was to identify and evaluate wearables-based methods of gait event identification that met two criteria: (1) The method needed to minimize system complexity by relying on data from a single wearable located on the shank or sacrum/low-back (in contrast to sets of multiple wearables), and (2) the method needed to be automated, using an unsupervised processing code to prevent bottlenecks caused by manually processing thousands or millions of gait events.

In line with this objective, we identified 18 candidate methods (Table 3.1; see Supplement 3 for further description of each method). Each of these methods represents a promising approach for in-field gait event identification, however, comprehensive comparisons of these methods have not been conducted, leaving users with little guidance on selecting the best method for their application. Further, most methods have at least one of the following limitations: Most methods were developed and validated (1) using a small sample ( $11 \pm 13$  mean  $\pm$  SD; 1-54 range), (2) using a narrow range of running speeds, (3) using a single running surface, (4) without accounting for foot-strike kinematics, and (5) without adequate synchronization between wearable data and a ground truth reference. These limitations decrease the confidence with which we can apply these wearables-based methods broadly across conditions and participants.

To address these limitations and help users make an informed decision about which of the many available methods to use for their application, we attempted to replicate each method using code provided by the authors, by directly contacting the authors to acquire code, or by following descriptions

in the papers to produce new code (code for each method available at [https://github.com/DovinKiernan/REID\\_IMU\\_Running\\_Event\\_ID](https://github.com/DovinKiernan/REID_IMU_Running_Event_ID)). We then evaluated each method to determine how well it could identify IC and TC events from a large sample that was collected across a range of running speeds, surfaces, and foot strike angles. To do so, we quantified: (1) how often each method failed to identify gait events, (2) the method's accuracy (bias), (3) precision (error variance), (4) the effects of speed, surface, and foot strike angle on accuracy, and (5) the time required to execute each method.

**Table 3.1: Summary of methods to identify gait events from running using data from a single wearable on the shank or sacrum/lower back.**

Sensor location	Method	Sample	Foot-strike	Speed	Surface	Placement	Signals	Sampling frequency	Events	Ground truth	Sync
Shank	Mizrahi [148]	<i>n</i> = 14 (14 M) Healthy	NR	3.5 m/s	Treadmill	Tibial tuberosity	$a_{WCS,Y}$	1667 Hz	IC	None	N/A
	Mercer [149]	<i>n</i> = 10 (10 M) Healthy	NR	3.1 - 3.8 m/s	Treadmill	Anteromedial distal tibia	$a_{WCS,Y}$	1000 Hz	IC, TC	None	N/A
	Purcell [150]	<i>n</i> = 6 Healthy	NR	SS jog, run, & sprint	Overground	Anteromedial distal tibia	$a_{WCS,X,Y,&Z}$	250 Hz	IC, TC	Forceplate 1000 Hz	TTL pulse
	Aminian/O’Donovan [151, 152]	<i>n</i> = 1 (1 M) Healthy	NR	SS jog	Overground	Shank	$\omega_{SCS,Z}$	102.4 Hz	IC, TC	MoCap 200 Hz	TTL pulse
	Greene/McGrath [153, 154]	<i>n</i> = 5 (4 M; 1 F) Healthy	RF	0.6 - 3.3 m/s	Treadmill	Anterior aspect of mid shank	$\omega_{WCS,Z}$	102.4 Hz	IC, TC	MoCap 200 Hz	TTL pulse
	Sinclair [155]	<i>n</i> = 16 (11 M; 5 F)	RF	4.0 ± 0.2 m/s	Overground	Anteromedial distal tibia	$a_{WCS,Y}$	1000 Hz	IC, TC	Forceplate 1000 Hz	Synchronous recording
	Whelan [156]	<i>n</i> = 7 (3 M; 4 F) National & International sprinters	NR	≤ 50% max effort	Overground	Anteromedial mid-tibia	$a_{WCS,X}$	148.2 Hz	IC	Forceplate 1000 Hz	TTL pulse
	Norris [157]	<i>n</i> = 6 (1 M; 5 F) Recreational	NR	SS half-marathon training	Overground	Anteromedial distal tibia	$a_{WCS,Z}$	204.8 Hz	IC	None	N/A
	Schmidt [158]	<i>n</i> = 12 (10 M; 2 F) Track and Field athletes	NR	SS sprint	Overground	Lateral distal tibia	$a_{WCS,Y}$ $\omega_{WCS,Z}$	1000 Hz	IC, TC	Photocell	NR
	Aubol [159]	<i>n</i> = 19 (9 M; 10 F) ≥16.1 km/wk Injury free	RF	3.0 ± 0.2 m/s	Overground	Anteromedial distal tibia	$a_{WCS,X,Y,&Z}$	1000 Hz	IC	Forceplate 1000 Hz	Synchronous recording
	Fadillioglu [160]	<i>n</i> = 13 (13 M) Injury free	NR	SS walking and running	Overground	Leg	$\omega_{WCS,Z}$	1500 Hz	IC, TC	Forceplate 1000 Hz	TTL pulse
Bach [161]	<i>n</i> = 5 (3 M; 2 F) Healthy	NR	SS walking and running	Treadmill	Anteromedial proximal tibia	$a_{WCS,X,Y,&Z}$	142.9 Hz	IC, TC	Forceplate 1000 Hz	NR	
Sacrum/Lower back	Auvinet [162]	<i>n</i> = 7 (7 M) “top-level”	RF	5.2 ± 0.1 m/s	Overground	Lumbar spine	$a_{WCS,X,Y,&Z}$	100 Hz	IC, TC, RL	MoCap 200 Hz	Photoflash
	Lee [163]	<i>n</i> = 10 (6 M; 4 F) National standard runners	NR	2.8 - 5.3 m/s	Treadmill	Sacrum (S1)	$a_{WCS,X&Z}$	100 Hz	IC, TC, RL	MoCap 100 Hz	Vertical movement
	Wixted [95]	<i>n</i> = 2 Nationally ranked	NR	5.9 - 6.2 m/s	Overground	Lumbar spine (L3-L4)	$a_{WCS,X&Y}$	500 Hz	IC, TC	Insoles 500 Hz	Synchronous collection
	Bergamini [164]	<i>n</i> = 11 (7 M; 4 F) Amateur and National Track & Field team	NR	5.7 - 10.8 m/s	Overground	Lumbar spine (L1)	$\omega_{WCS,X,Y,&Z}$	200 Hz	IC,TC	Forceplate/Mocap 200Hz/300Hz	Hammer tap/ none
	Benson [165]	<i>n</i> = 54 (29 M; 25 F) Recreational	FF & RF	2.7 - 3.6 m/s	Treadmill & Overground	Lower back	$a_{WCS,X,Y,&Z}$	201.03 Hz	IC, TC, RL	Forceplate 1000 Hz	Vertical jump
	Reenalda [166]	<i>n</i> = 20 (15 M; 5 F) ≥15 km/week; no injuries	FF & RF	3.1 – 4.2 m/s	Treadmill	Sacrum	$a_{GCS,Y}$	240 Hz interpolated to 1000 Hz	IC	Forceplate 1000 Hz	x-correlated MoCap

M – Male; F – Female; NR – Not reported; FF – forefoot; RF – rearfoot; SS – Self-selected;  
 $\alpha$  – acceleration;  $\omega$  – angular velocity; SCS – segment coordinate system; WCS – wearable coordinate system; GCS – global coordinate system (coordinate conventions defined in Fig. 3)  
 IC – initial contact; TC – terminal contact; RL – right/left

## 3.2. Materials & Methods

### 3.2.1. IMU calibration

Tri-axial IMUs (ProMove MINI, Inertia Technology;  $\pm 16$  g primary,  $\pm 100$  g secondary,  $\pm 34.91$  rad/s, 1000 Hz) were secured to a centrifuge (ClearPath MCVC, Teknic) with custom 3-D printed jigs (SOLIDWORKS 2019, Dassault Systèmes) and calibrated in 6 orientations at 16 known accelerations (from 0 to 41.42 g where 1 g = 9.8 m/s<sup>2</sup> [167, 168]) and angular velocities (from 0 to 78.54 rad/s). Adapting methods from Coolbaugh et al. [99], known data (K) from the centrifuge and measured data (M) from the IMU were used to calculate 3 x 7 calibration matrices (C; 3 signed magnitude terms, 3 absolute magnitude terms, and one bias term per axis) and quantify sensor accuracy after subtracting out biases observed during a quiet period (B) (Equation 3.1). After calibration, IMU primary accelerometer errors were  $\leq 0.01 \pm 0.04$  g, secondary accelerometer errors were  $\leq 0.05 \pm 0.07$  g, and gyroscope errors were  $\leq 0.01 \pm 0.01$  rad/s.

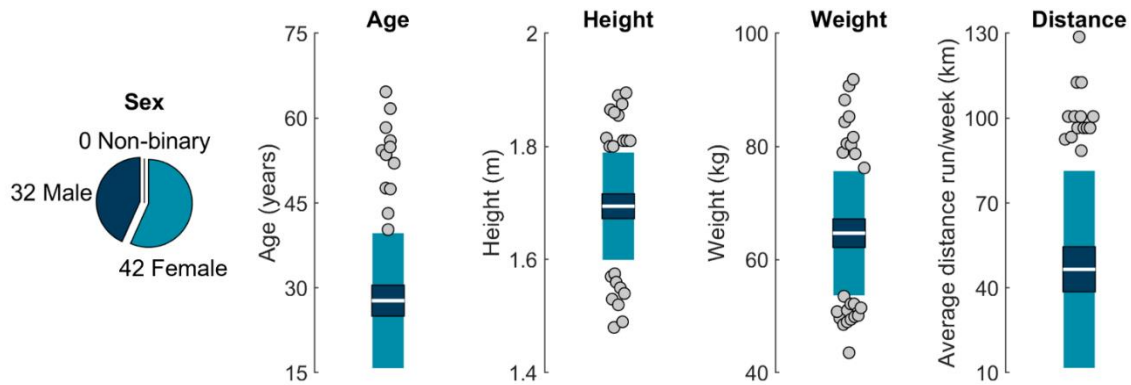
#### Equation 3.1

$$C*(M + B) = K$$

### 3.2.2. Participants

Seventy-seven runners were recruited and tested. Three participants were excluded from analysis due to movement of an IMU ( $n = 2$ ) or inability to complete the protocol as instructed ( $n = 1$ ) leaving a final sample of 74 (32 male; 42 female; age  $28 \pm 12$  years; Figure 3.2). All participants provided written informed consent and procedures were approved by the University of California, Davis Institutional Review Board.

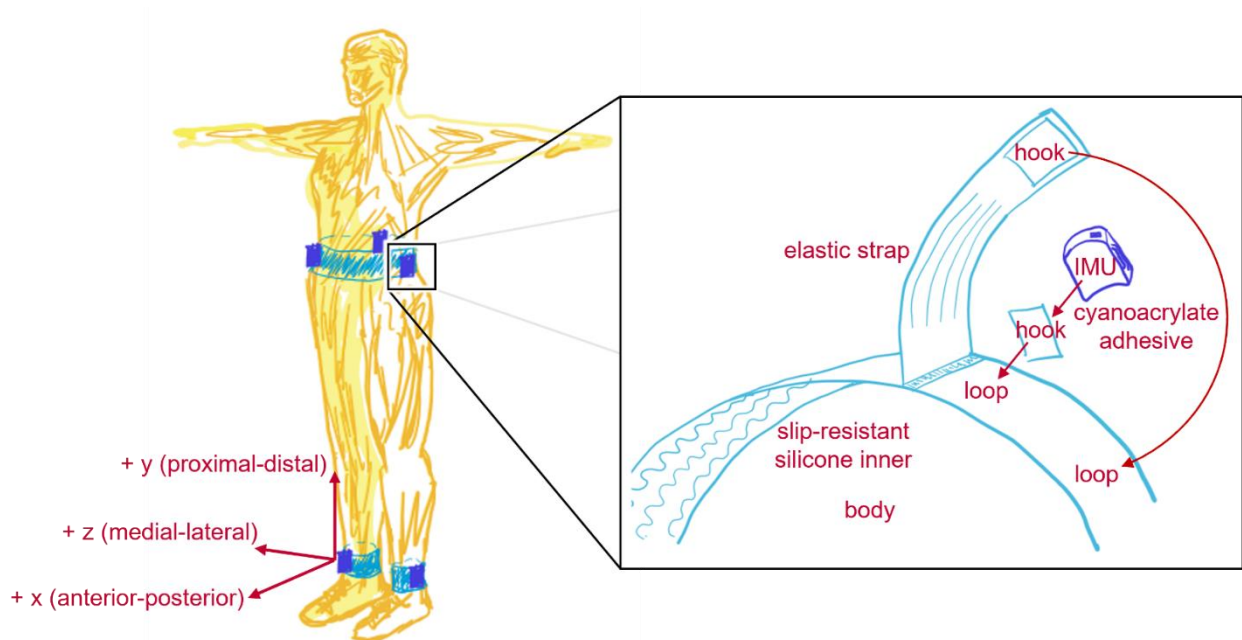




**Figure 3.2: Sample characteristics.** From left to right: Sample sex, age, height, weight, and self-reported average distance run per week. White horizontal line represents the mean, dark blue represents  $\pm 95\%$  confidence interval ( $\pm 1.96$  SEM) around the mean, light blue represents  $\pm 1$  SD around the mean. Gray dots represent participants outside  $\pm 1$  SD.

### 3.2.3. Protocol

Using adhesive-bonded hook-and-loop fasteners, IMUs were attached to neoprene belts with anti-slip silicone inners then wrapped with elastic straps as tightly as possible, within the limit of participant comfort. IMUs were placed anterior and superior to the lateral malleoli, the superior aspect of the iliac crest in-line with the greater trochanter, and the superior aspect of the sacrum in line with the spine (Figure 3.3). Only data from tibial and sacral IMUs are analyzed here.



**Figure 3.3: IMU fixation and coordinate system.** On the left, IMU placement and coordinate conventions. For consistency, different conventions used across methods have been standardized to ISB conventions [169]: Segment coordinate systems (SCS) were defined as anterior (+x), proximal (+y), and medial-lateral (with right defined as +z); and wearable-coordinate systems (WCS) were defined by the axes of the sensors within the housing of the wearable which was roughly aligned with the direction of progression (+x), longitudinal axis (+y), and right (+z). On the right, belt design and IMU fixation.

Participants wore their own shoes and ran a 25 m track with embedded force plate (Kistler 9281, Kistler Group; 1000 Hz). Running speed was recorded using two custom-built laser speed gates 2.5 m on each side of force plate center. Participants warmed up and practiced striking the force plate three times per side at *slowest* (“the slowest pace you would use on a run”), *typical* (“the pace you use for the majority of your running”), and *fastest* (“the fastest pace you would use on a run”) self-selected speeds. During this warmup, markers on the lateral calcaneus and base of the fifth metatarsal were recorded using a conventional camera (Exilim EX-FH25, Casio; 120 Hz). Foot strike angle was calculated by subtracting a neutral standing foot angle (Kinovea 0.9.5) with positive values indicating a more dorsiflexed foot at IC and values  $> 0.14$  rad corresponding to rear-foot strike,  $-0.03$  to  $0.14$  rad to mid-foot strike, and  $< -0.03$  rad to forefoot strike [170]. After warm-up, five stances per side were collected at each speed for two surface conditions: (1) With a track surface covering the running path and force plate (*track*) and (2) with no covering on the hardwood floor of a basketball court (*floor*). Participants always progressed from their slowest to fastest speeds, but the order of foot and surface were pseudo-randomized.

IMU data were synchronized within 100 ns of each other with a wireless network hub (Advanced Inertia Gateway, Inertia Technology; 1000 Hz). This hub periodically sent voltage pulse trains that were synchronously recorded by IMUs and a custom MATLAB script that simultaneously recorded the speed and force data (R2018b, the Mathworks). Pulse trains were cross correlated to synchronize signals.

#### 3.2.4. Data processing

Quiet periods were identified and used to remove biases from IMU data. Saturated frames from the primary accelerometer ( $|a| > 15.5$  g) were replaced with corresponding frames from the secondary accelerometer. Data were filtered with a 4<sup>th</sup> order 50-Hz low-pass Butterworth filter. Angular velocity was drift-corrected using a Madgwick filter [171] (shown to perform as well as, or better than, Kalman and Mahony filters [172, 173] and here validated to provide the best orientation estimates under experimental conditions). Starting at each quiet period, accelerations were used to estimate IMU position in the inertial reference frame, then angular velocities were used to estimate frame-by-frame changes in IMU orientation and remove the gravity component from accelerations [174]. Data were then transformed to a segment coordinate system based on the Principal Component that explained the most variance in angular velocity during running (the medial-lateral axis), the gravity vector during quiet standing, and their cross-product [121, 175]. Force plate data were filtered with a 4<sup>th</sup> order 50-Hz low-pass Butterworth filter. A vertical force threshold of 10 N was used to define the start and end of stance (IC and TC respectively).

During data processing we observed small discrepancies caused by the initialization of discrete MATLAB data acquisitions and small variances in the sampling rates of the IMU and MATLAB systems. Although extremely small, these discrepancies could accumulate over the course of the ~60-minute data collection leading to timing discrepancies between the first and last synch events of a data collection (10s of ms). Given this study's focus on event timing, a conservative approach was used and only trials that contained a synch event were analyzed (642 of 4440 trials). All other trials were discarded to ensure millisecond-level accuracy was maintained.

### 3.2.5. Analysis

To quantify how well each method agreed with ground truth gait events (IC and TC) measured by the force plate, errors were quantified for each method by subtracting method-estimated gait-event timings from force plate gait-event timings. To account for non-independence of data (642 trials from 74 participants) and ensure proper estimation of variance, IC and TC errors were entered into two separate linear mixed-effects models in R (v4.2.2; R Foundation for Statistical Computing) as described in Carstensen et al.'s approach to linked replicates (Equation 3.2) [176]:

#### Equation 3.2

$$y_{method,participant,trial} = a_{method} + b_{participant} + c_{participant,trial} + d_{method,participant} + e_{method,participant,trial}$$

This approach allowed us to quantify variance components and estimate: (1) Method biases (accuracy), (2) Limits Of Agreement (LOA) within which 95% of future errors for a given method are expected to fall, and (3) within-method standard deviations that quantify method repeatability (precision) (see also: [177, 178, 179]). Model assumptions of independence, normality, and homoscedasticity were validated by plotting within-participant variances against within-participant means, histograms of residuals, residuals for each level of random effect, and residuals as a function of fitted value.

To evaluate if any potential explanatory variables affected method error, a second set of linear mixed effect models was developed for each method. These models added surface, speed, and foot strike angle as fixed effects. Significance of  $p \leq 0.05$  was used to evaluate whether a fixed effect explained a significant amount of a given method's error.

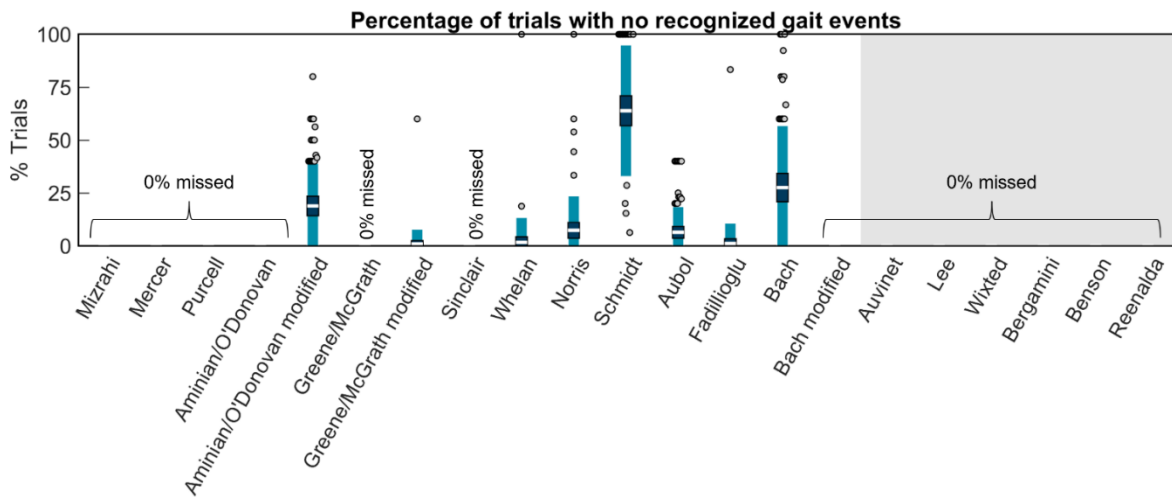
Finally, to quantify the processing time required to execute each method, the same computer (Intel Core i9-13900HX; Kingston Fury 2x32 GB DDR5 5600 MT/s; Gammix S70 Blade 7400 MB/s SSD) was used to process two 30-minute, 1.8 million frame (1000 Hz), steady state running trials obtained from a

separate study (but using the same IMUs, placement, and pre-processing) [180]. Each trial was processed 100 times per method and processing times were recorded.

### 3.3. Results

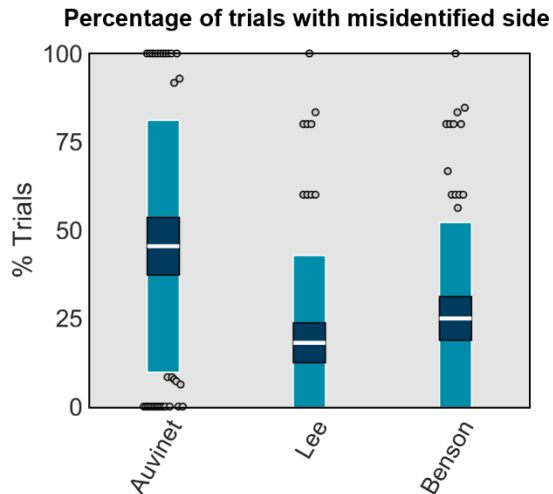
#### 3.3.1. Failure to identify gait events or step side

Some methods were unable to recognize any gait events within the 5 m of running that was segmented for analysis in each trial. The percentage of trials per participant where no gait events were identified is plotted in Figure 3.4.



**Figure 3.4: Percentage of trials per participant without any gait events recognized.** White horizontal line represents the mean, dark blue represents  $\pm 95\%$  confidence interval ( $\pm 1.96$  SEM) around the mean, light blue represents  $\pm 1$  SD around the mean. Gray dots represent participants outside  $\pm 1$  SD. No bars indicates that gait events were identified in every trial for every participant. Methods on the white background are for wearables on the shank. Methods on the gray background are for wearables on the sacrum/low-back.

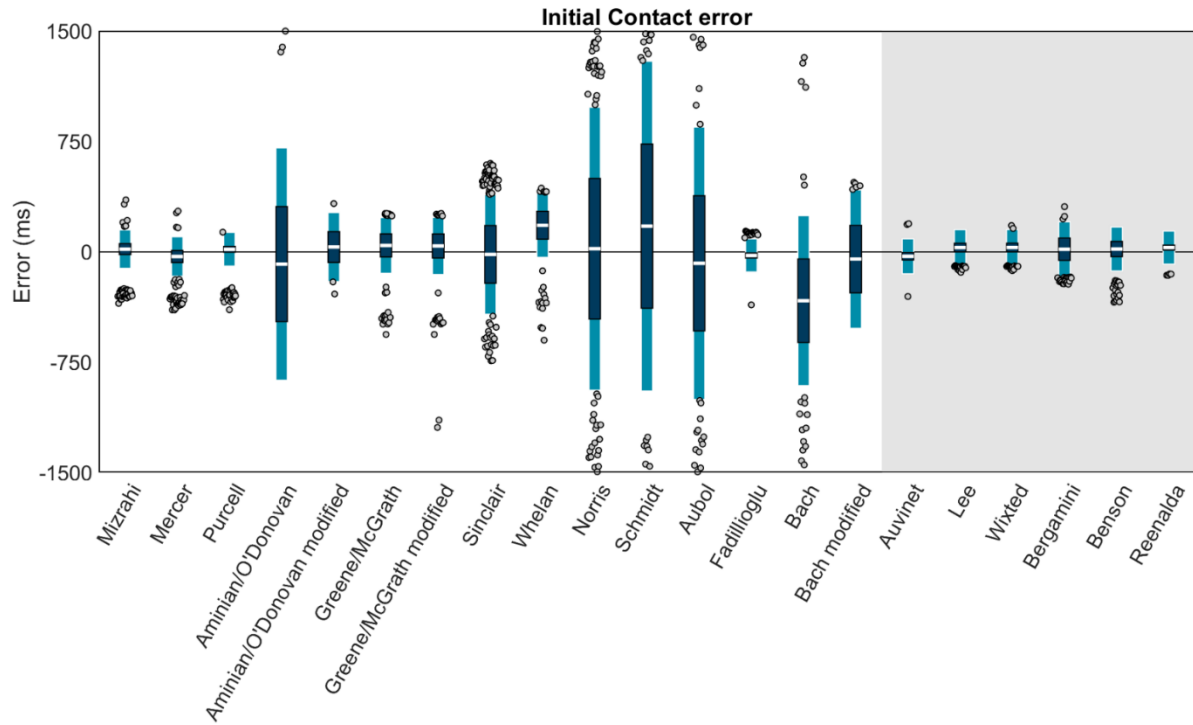
Methods using wearables placed on the shank only provided information ipsilateral to placement and thus side was always known. In contrast, methods using wearables placed on the sacrum/low-back provided information on bilateral gait events. Despite this, only three of the six sacrum/low-back placed methods were capable of identifying the side (i.e., right or left) contacting the ground. The accuracy of these methods is plotted in Figure 3.5.



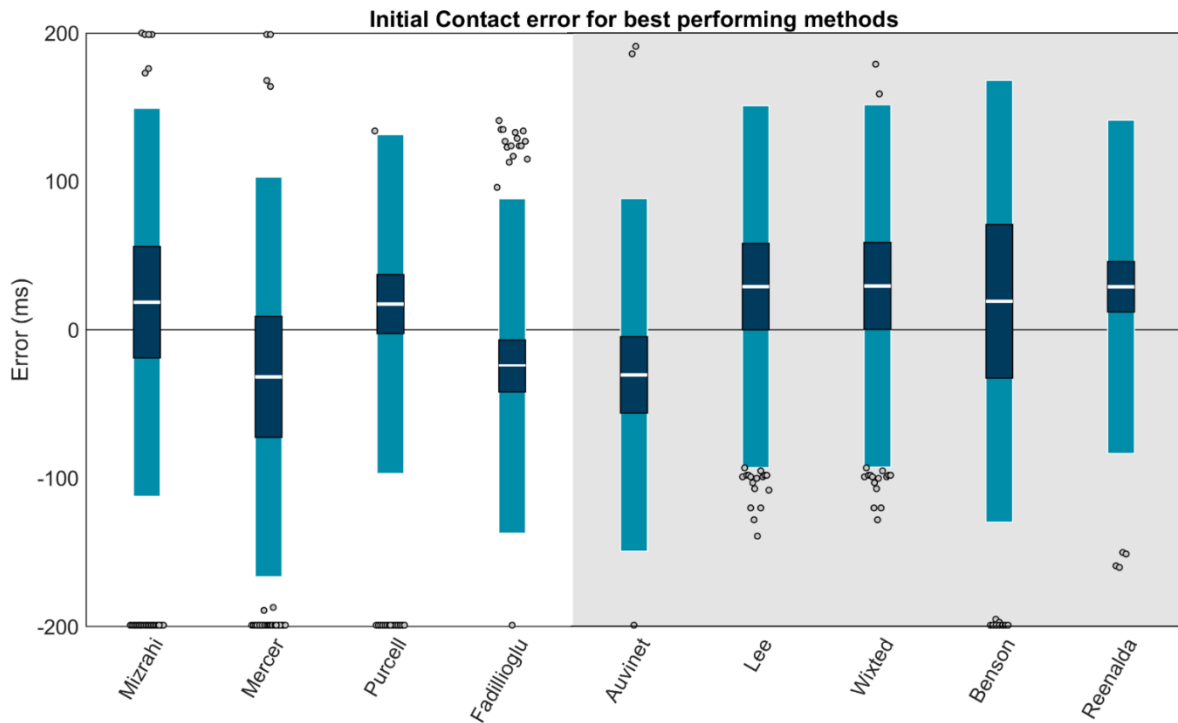
**Figure 3.5: Percentage of trials per participant where side was not correctly recognized.** White horizontal line represents the mean, dark blue represents  $\pm 95\%$  confidence interval ( $\pm 1.96$  SEM) around the mean, light blue represents  $\pm 1$  SD around the mean. Gray dots represent participants outside  $\pm 1$  SD.

### 3.3.2. Initial Contact

Using Carstensen's method for linked replicates [176], biases, within-method SDs, and 95% LOAs (1.96 SD of errors) for the IC estimation of each method were modeled (Figure 3.6). Results revealed high biases and/or LOAs for several methods; thus, a second figure displaying only the best performing methods (LOAs within  $\pm 200$  ms) is also provided (Figure 3.7). The best performing methods for shank mounted wearables were Mizrahi (0% unrecognized events, +18.58 ms bias, -112.16 to +149.32 ms LOA), Mercer (0% unrecognized events, -31.70 ms bias, -166.34 to +102.93 ms LOA), Purcell (0% unrecognized events, +17.39 ms bias, -96.81 to +131.58 ms LOA), and Fadillioglu (1.13% unrecognized events, -24.33 ms bias, -137.02 to +88.35 ms LOA). The best performing methods for shank/low-back mounted wearables were Auvinet (0% unrecognized events, -30.36 ms bias, -149.19 to +88.47 ms LOA), Lee (0% unrecognized events, +29.10 ms bias, -92.85 to +151.04 ms LOA), Wixted (0% unrecognized events, +29.56 ms bias, -92.53 to +151.65 ms LOA), Benson (0% unrecognized events, +19.25 ms bias, -129.64 to +168.13 ms LOA), and Reenalda (0% unrecognized events, +29.00 ms bias, -83.34 to +141.34 ms LOA).



**Figure 3.6: IC estimation.** Means (white bar),  $\pm 1$  within-method SD (dark blue), and  $\pm 95\%$  LOA (1.96 SD of errors; light blue) for each method. Gray dots represent trials falling outside the 95% LOA. A value of 0 indicates perfect agreement with the ground truth. Positive values indicate the IC was estimated later than the ground truth (after the force plate IC). Negative values indicate the IC was estimated earlier than the ground truth (before the force plate IC). Methods with a white background are for wearables on the shank. Methods on the gray background are for wearables on the sacrum/low-back.



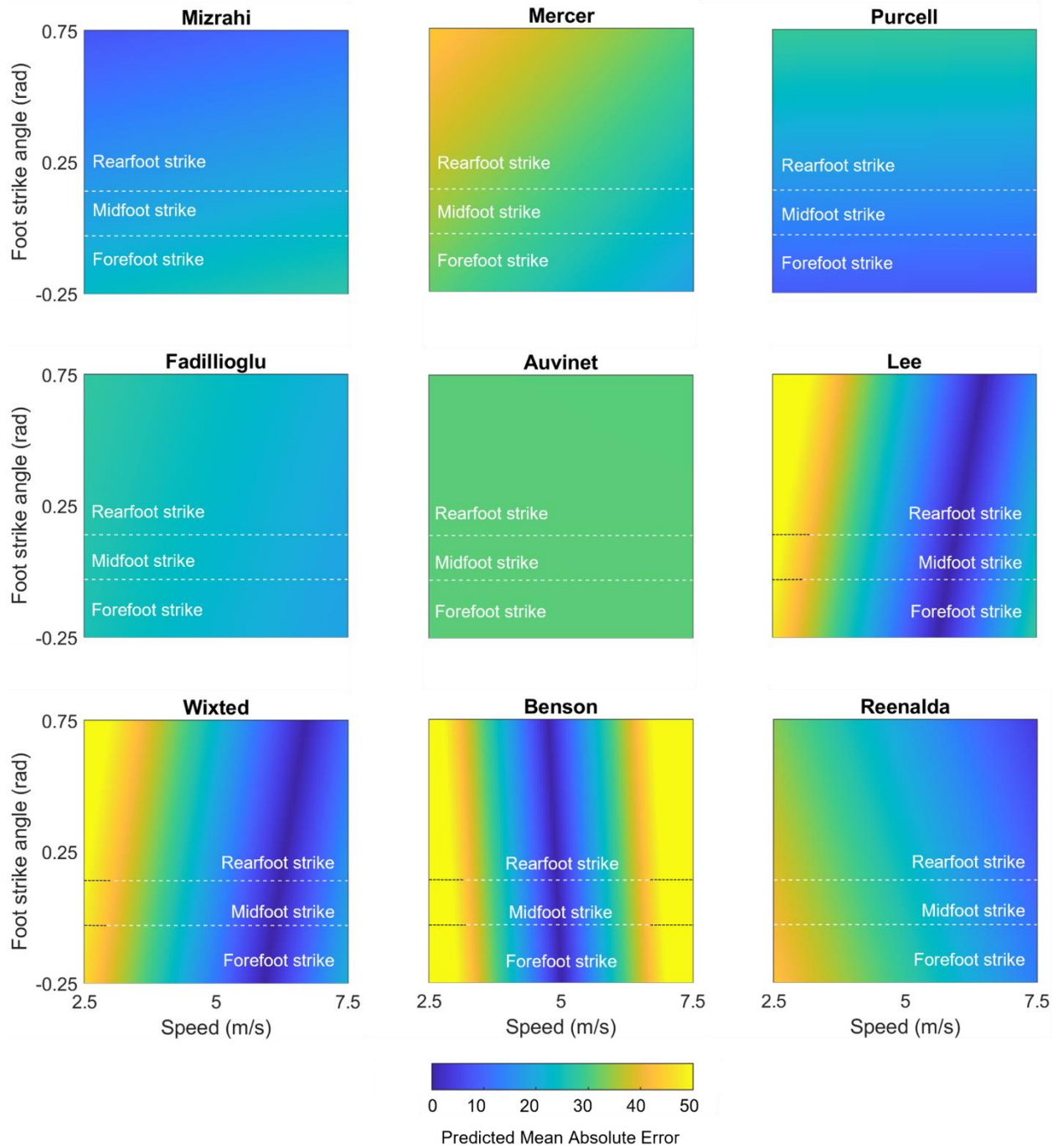
**Figure 3.7: IC estimation for methods with LOAs within  $\pm 200$  ms.** Means (white bar),  $\pm 1$  within-method SD (dark blue), and  $\pm 95\%$  LOA (1.96 SD of errors; light blue) for each method. Gray dots represent trials falling outside the 95% LOA. Outliers falling outside the  $\pm 200$  ms range are plotted at  $\pm 200$  ms. A value of 0 indicates perfect agreement with the ground truth. Positive values indicate the IC was estimated later than the ground truth (after the force plate IC). Negative values indicate the IC was estimated earlier than the ground truth (before the force plate IC). Methods with a white background are for wearables on the shank. Methods on the gray background are for wearables on the sacrum/low-back.

We performed a second set of linear mixed effect models on each method to examine the role of potential explanatory variables. These models revealed that none of the methods' IC estimations were significantly affected by surface (all  $p$ s  $> 0.05$ ). In contrast, speed explained a significant amount ( $p \leq 0.05$ ) of the error in 3 of 15 shank-placed methods and 5 of 6 sacrum/low-back placed methods, and foot strike angle explained a significant amount of the error in 1 of 15 shank- and 0 of 6 sacrum/low-back -placed methods (Table 3.2). To better illustrate these effects, model-predicted mean absolute errors were plotted as a function of speed and foot strike angle (Figure 3.8 for best performing methods; Chapter 3S for others).

**Table 3.2. Effects of surface, running speed, and foot strike angle on Initial Contact and Terminal Contact estimation** across methods. Numbers represent coefficients for the intercept of surface (added to model estimates for the floor condition but not the track condition) and the slopes of running speed and foot strike angle. \*significant ( $p \leq 0.05$ ) effects highlighted in blue.



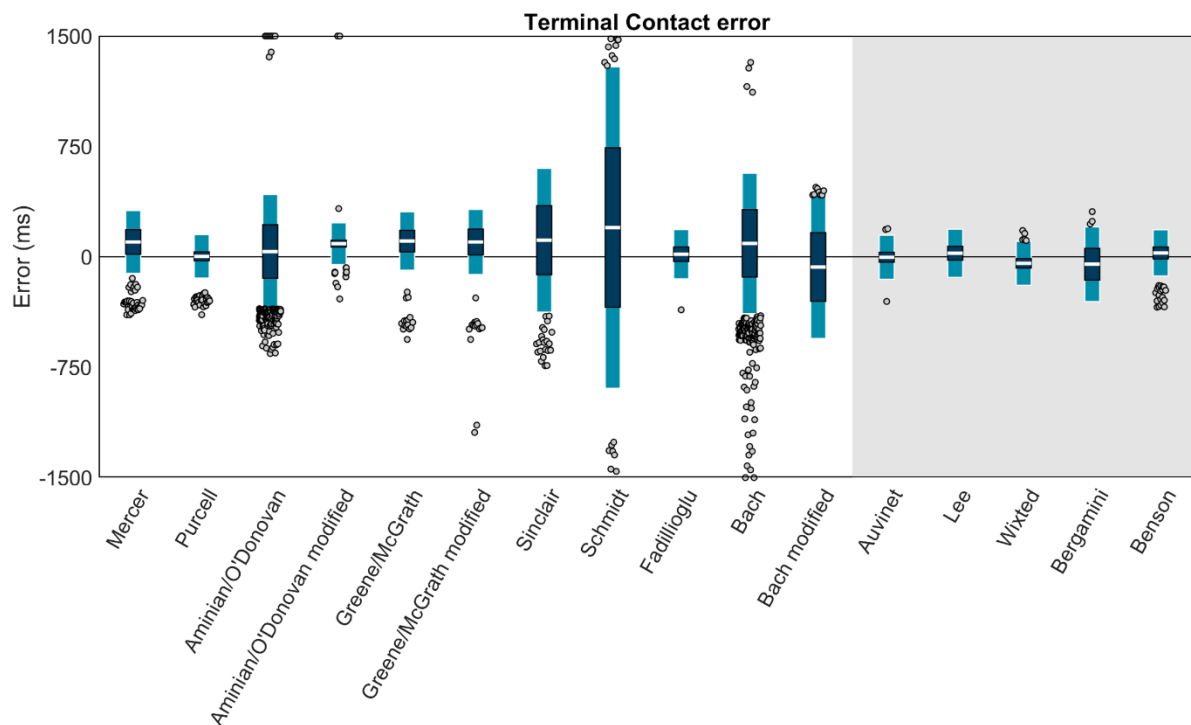
Method	Initial Contact (IC)			Terminal Contact (TC)		
	Surface	Speed	Foot strike	Surface	Speed	Foot strike
Mizrahi	-5.02	0.71	-15.91	n/a		
Mercer	-5.07	2.59	-11.99	-0.07	34.17*	-47.56
Purcell	0.74	-0.14	19.66	2.46	4.00*	4.02
Aminian/O'Donovan	2.81	-31.55	19.26	-15.10	11.29	-1.83
Aminian/O'Donovan modified	11.40	-5.58	-0.12	-1.26	-5.95*	10.76
Greene/McGrath	2.22	-5.53	23.35	1.12	1.18	-16.43
Greene/McGrath modified	-1.91	-2.70	9.89	3.79	0.69	-25.87
Sinclair	-7.24	0.18	-44.17	-9.19	-15.65	43.08
Whelan	-14.22	-13.89*	74.18*	n/a		
Norris	-23.89	-36.41	148.44	n/a		
Schmidt	-11.17	209.61*	49.38	-2.46	227.50*	18.75
Aubol	-26.90	-3.94	34.33	n/a		
Fadilliglu	0.20	1.44	-2.13	1.27	-4.44	37.21
Bach	19.63	9.71	32.00	-7.44	-3.37	-44.60
Bach modified	-1.07	36.60*	68.71	1.75	54.45*	75.07
Auvinet	-3.08	0.01	0.13	0.15	6.86*	-19.55
Lee	-1.03	-15.11*	13.29	0.74	-12.46*	4.05
Wixted	-2.99	-13.39*	10.53	-1.52	2.37	0.20
Bergamini	-9.77	-19.31*	12.63	-9.95	2.87	-34.54
Benson	0.11	-25.14*	-6.34	0.46	4.60*	-7.39
Reenalda	-0.13	-5.93*	-9.39	n/a		



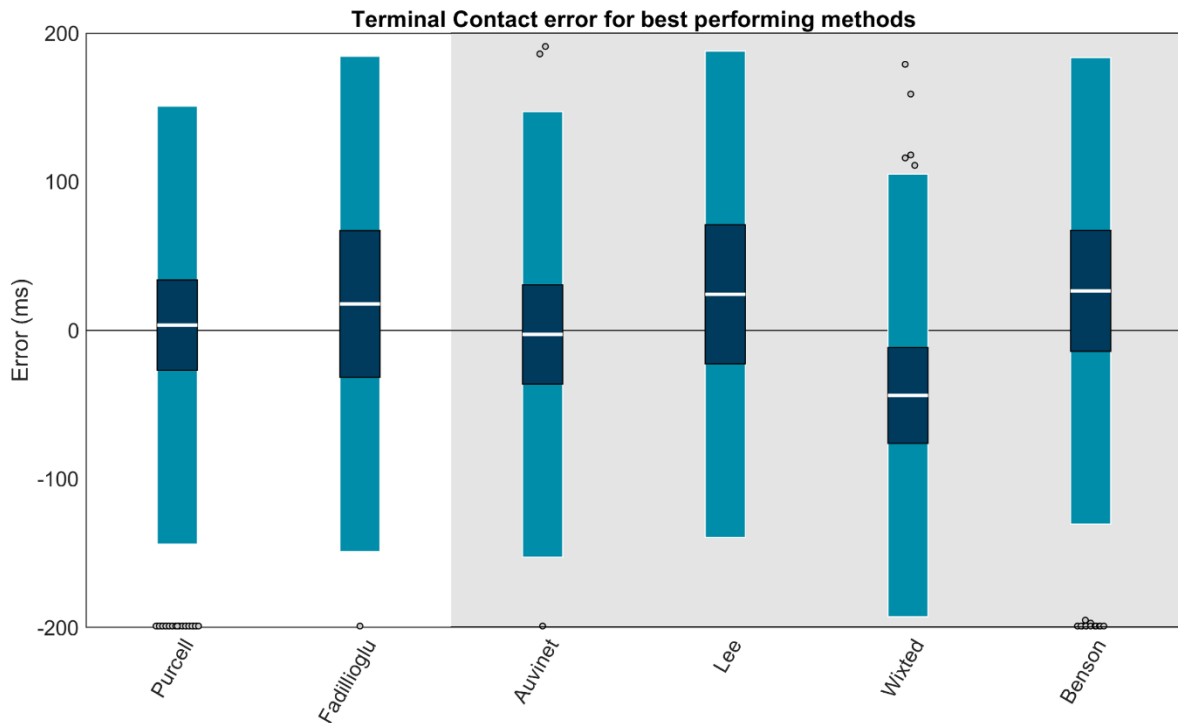
**Figure 3.8: Mean absolute error in IC predicted by mixed effects models for each of the best performing methods (LOAs within  $\pm 200$  ms). Plotted as a function of speed and foot strike angle. Darker blue values represent lower predicted mean absolute errors while brighter yellow values represent higher predicted mean absolute errors. To illustrate, Auvinet has a mean absolute error of  $\sim 30$  ms that is  $\sim$ constant across foot strike angles and speeds; in contrast, Wixted has lower mean absolute error at speeds above  $\sim 4$  m/s but higher mean absolute errors at speeds below  $\sim 4$  m/s.**

### 3.3.3. Terminal Contact

Using Carstensen’s method for linked replicates [176], biases, within-method SDs, and 95% LOAs for each method’s TC estimation were calculated (Figure 3.9). Results revealed high biases and/or LOAs for several methods; thus, a second figure displaying only the best performing methods (LOAs within  $\pm 200$  ms) is also provided (Figure 3.10). The best performing methods for shank mounted wearables were Purcell (0% unrecognized events, +3.51 ms bias, -143.86 to +150.87 ms LOA) and Fadillioğlu (1.13% unrecognized events, +17.75 ms bias, -148.90 to +184.40 ms LOA). The best performing methods for shank/low-back mounted wearables were Auvinet (0% unrecognized events, -2.75 ms bias, -152.67 to +147.17 ms LOA), Lee (0% unrecognized events, +24.26 ms bias, -139.40 to +187.91 ms LOA), Wixted (0% unrecognized events, -43.84 ms bias, -192.74 to +105.07 ms LOA), and Benson (0% unrecognized events, +26.55 ms bias, -130.41 to +183.51 ms LOA).

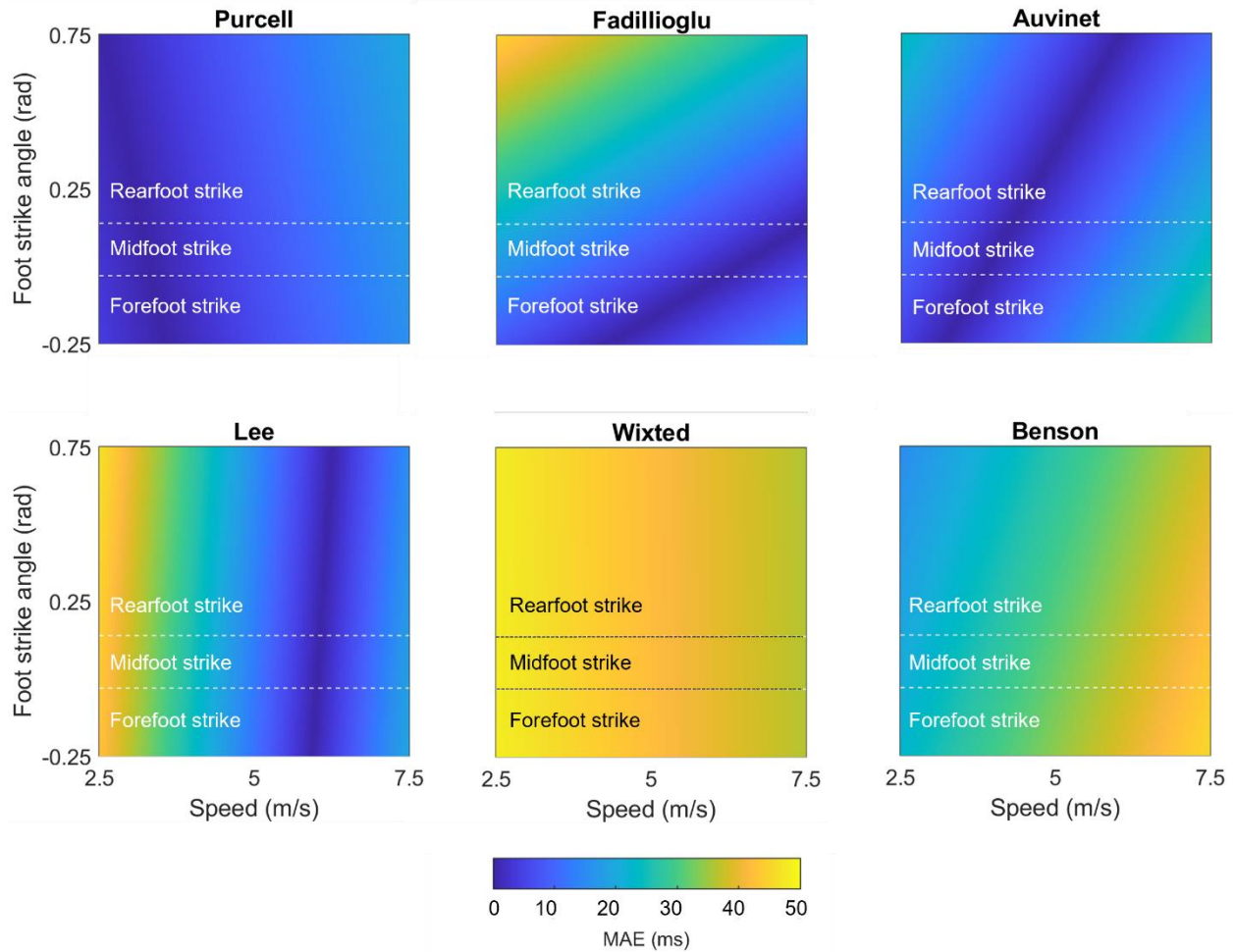


**Figure 3.9: TC estimation.** Means (white bar),  $\pm 1$  within-method SD (dark blue), and  $\pm 95\%$  LOA (1.96 SD of errors; light blue) for each method. Gray dots represent trials falling outside the 95% LOA. A value of 0 indicates perfect agreement with the ground truth. Positive values indicate the TC was estimated later than the ground truth (after the force plate TC). Negative values indicate the TC was estimated earlier than the ground truth (before the force plate TC). Methods with a white background are for wearables on the shank. Methods on the gray background are for wearables on the sacrum/low-back.



**Figure 3.10: TC estimation for methods with LOAs within  $\pm 200$  ms.** Means (white bar),  $\pm 1$  within-method SD (dark blue), and  $\pm 95\%$  LOA (1.96 SD of errors; light blue) for each method. Gray dots represent trials falling outside the 95% LOA. Outliers falling outside the  $\pm 200$  ms range are plotted at  $\pm 200$  ms. A value of 0 indicates perfect agreement with the ground truth. Positive values indicate the TC was estimated later than the ground truth (after the force plate TC). Negative values indicate the TC was estimated earlier than the ground truth (before the force plate TC). Methods with a white background are for wearables on the shank. Methods on the gray background are for wearables on the sacrum/low-back.

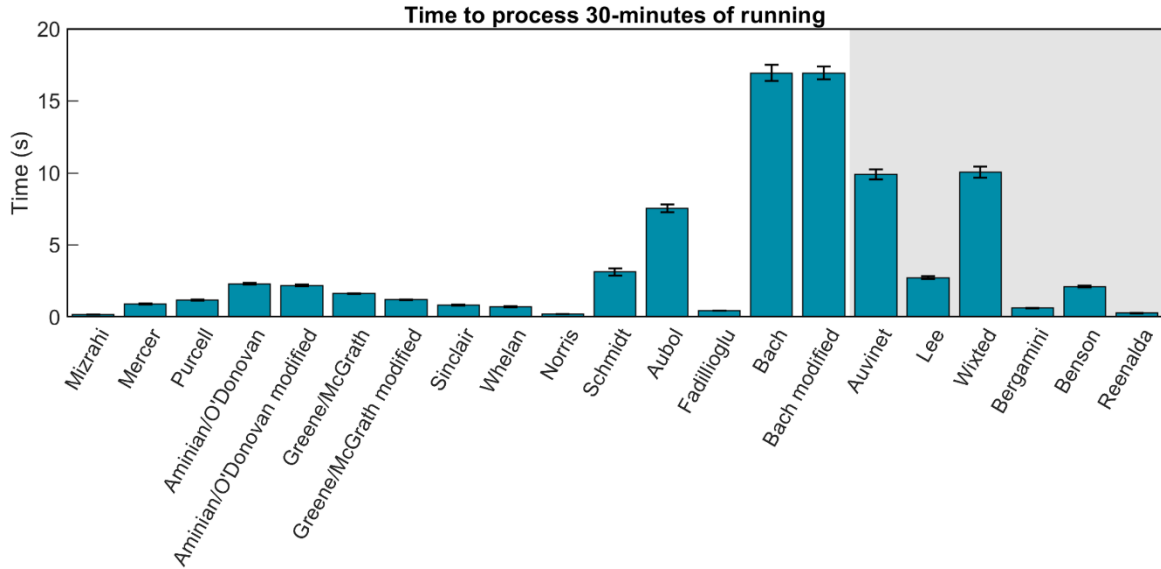
We performed a second set of linear mixed effect models on each method to examine the role of potential explanatory variables. These models revealed that none of the methods' TC estimations were significantly affected by surface or foot strike angle (all  $p$ s  $> 0.05$ ), while speed explained a significant amount ( $p \leq 0.05$ ) of the error in 5 of 11 shank-placed methods and 3 of 5 sacrum/low-back placed methods (Table 3.2). To better illustrate these effects, model-predicted mean absolute errors were plotted as a function of speed and foot strike angle (Figure 3.11 for best performing methods; Supplement 3 for others).



**Figure 3.11: Mean absolute error in TC predicted by mixed effects models for each of the best performing methods (LOAs within  $\pm 200$  ms).** Plotted as a function of speed and foot strike angle. Darker blue values represent lower predicted mean absolute errors while brighter yellow values represent higher predicted mean absolute errors. To illustrate, Wixted has a mean absolute error of  $\sim 44$  ms that is  $\sim$ constant across foot strike angles and speeds; in contrast, Lee has lower mean absolute error at speeds above  $\sim 3$  m/s but higher mean absolute errors at speeds below  $\sim 3$  m/s.

### 3.3.4. Processing time

Finally, to quantify the time required to execute each method, two 30-min, 1.8 million frame, steady-state running trials were processed 100 times with each method. Processing times were recorded and plotted in Figure 3.12.



**Figure 3.12.** Mean time to process 30-min of steady state running data sampled at 1000 Hz. Error bars represent  $\pm 1$  SD about the mean. All processing conducted on the same computer (Intel Core i9-13900HX; Kingston Fury 2x32 GB DDR5 5600 MT/s; Gammix S70 Blade 7400 MB/s SSD).

### 3.4. Discussion

We identified 18 separate methods to estimate initial contact (IC) and terminal contact (TC) gait events from running using a single wearable sensor on the shank or sacrum/low-back. We modified three of these original methods in an attempt to improve performance, resulting in a total of 21 methods. For each method we either adapted (4 of 18) or created (14 of 18) code to automatically process data (code available at [https://github.com/DovinKiernan/Running\\_Gait\\_Event](https://github.com/DovinKiernan/Running_Gait_Event)). We then used each automated method to estimate gait events from 74 runners across two different surfaces (wood floor and running track), three self-selected speeds (slowest, typical, and fastest), and foot strike angles (ranging from forefoot to rearfoot strike patterns). To quantify error, these estimated gait events were compared to ground truth events from a time-synchronized force plate.

Overall, we found error to be higher than reported in the original studies (with two exceptions: Benson et al. [165], where we found errors below the 50 ms for IC and 30 ms for TC originally reported, and Reenalda et al. [166], where we found errors on par with their reported 21-27 ms). This is likely a

function of several factors: First, in contrast to some studies, we were able to exactly time synchronize our gold-standard force plates with our IMUs using voltage pulses. Second, to our knowledge, the current study represents the largest sample used to validate and compare running gait event identification methods. Most studies used samples of  $11 \pm 13$  participants (mean  $\pm$  SD) to both develop and validate their method (Benson was an outlier here, using 54 participants). Third, the current study used a range of conditions including different surfaces, speeds, and foot strike angles. In contrast, the methods studied were developed and validated on a narrow set of conditions (again, with the exception of Benson who validated across several conditions). Thus, the error values reported for most studies are likely a function of both developing and validating the method on the same small sample of participants and narrow set of conditions. When adopting a gait event identification method, users should consider the conditions and participants with which the method was developed and validated and not assume that the method will work under other conditions or for other participants.

For example, O'Donovan et al. [152] and McGrath et al. [153] adapted approaches developed for walking [151, 154] without specifying any changes to address the differences between walking and running. These methods had high error, often mislabeling gait events from pre-/proceeding steps as the event of interest due to their large time windows. When modified with timings based on running, however, the performance of the Aminian/O'Donovan method did improve. On the other side of the spectrum, although it fit the criteria for inclusion in this study, Schmidt et al. [158] developed their method specifically for sprinting and never advocated for its application to running more broadly. This likely explains the high error observed across the range of running speeds we used here. As a final example, Bach et al. [161] trained an Echo State Network to estimate gait events using data from a narrow set of conditions. To replicate their model, we used the full running data set that they generously published with their original paper; however, we still observed high error when using it to estimate gait events from our data. This error likely stems from the model being applied to a broad set

of conditions on which it was not trained. It is possible that further training across these conditions could yield improved results.

These observations should also be applied to the current study: We only quantified error for over-ground running on two level surfaces. It should not be assumed that our results would hold for incline or decline running, treadmill running, or running on other surfaces (e.g., grass/turf, asphalt, concrete). That said, the fact that no method was significantly affected by differences in the two surfaces used here, suggests that they would also work on similar surfaces (like concrete and asphalt) though further research should confirm they work on softer surfaces (like grass and sand). Finally, all methods studied here have been developed and validated on a relatively homogenous group of runners. Thus, if studying participants drawn from different populations, our results may not be representative. For example, individuals with lower extremity amputation could have higher frequency components in their data that may disrupt pattern recognition in some methods [98]. Similarly, data collected from individuals with more subcutaneous body fat could contain more noise from soft tissue artefact.

With those caveats in mind, for IC estimation with a shank-placed wearable, we recommend either the Purcell [150] or Fadillioglu [160] method. Both methods were able to identify gait events in almost all trials (100% and 98.87%) and stood out with low biases (+17.39 and -24.33 ms) and LOAs (-96.81 to +131.58 and -137.02 to +88.35 ms). Further, neither method was significantly affected by running surface, speed, or foot strike angle, suggesting that both methods could be applied broadly across conditions. Compared to one another, Purcell had slightly lower bias whereas Fadillioglu had slightly lower LOAs and slopes for speed and foot strike angle (though, again, neither was significantly affected by these potential explanatory variables).

For IC estimation with a sacrum or low-back placed wearable, we recommend the Auvinet [162] or Reenalda [166] method. Both methods were able to identify gait events in 100% of trials and stood



out with low biases (-30.36 and +29.00 ms) and LOAs (-149.19 to +88.47 and -83.34 to +141.34 ms). The Auvinet method was not significantly affected by running surface, speed, or foot strike angle. In contrast, the Reenalda method was significantly affected by running speed with errors increasing at low running speeds as well as a trend toward errors increasing at lower foot strike angles. Thus, the Reenalda method may be preferable when speeds are known to exceed ~4 m/s while the Auvinet method may be preferable for speeds below ~4 m/s or when conditions are unknown. The Lee, Wixted, and Benson methods [163, 95, 165] may also present viable options when speed is known to fall within certain ranges (see Figure 3.7).

For TC estimation with a shank-placed wearable, we recommend the Purcell method [150]. This method identified gait events in 100% of trials with low bias (+3.51 ms) and LOAs (-143.86 to +150.87 ms). Error was not significantly affected by surface or foot strike angle. And, although it was significantly affected by speed, the slope was quite low, indicating error only changed a small amount across speeds (4.00 ms per 1 m/s).

For TC estimation with a sacrum/low-back placed wearable, we recommend the Auvinet method [162]. This method identified gait events in 100% of trials with low bias (-2.75 ms) and LOAs (-152.67 to +147.17 ms). Error for TC estimation with the Auvinet method was not significantly affected by surface or foot strike angle but was significantly affected by speed. At speeds below ~5 m/s, the Lee method [163] could also be used. Although the Lee method's bias (+24.26 ms) and LOAs (-139.40 to 187.91 ms) were higher overall, this appears to be largely driven by the significant effect of speed, with lower speeds causing larger errors.

To identify side with a sacrum/low-back placed wearable, we recommend the Lee method [163] which correctly identified side in 81.95% of gait events. The Benson method performed almost as well, correctly identifying side for 75.05% of gait events. Given these high success rates and the predictable

left-right-left-right pattern of running, the probability of misidentifying gait events should decrease exponentially as a function of the number of steps, quickly reaching negligible values. To illustrate, based on the number of steps observed in the 5 m trials used here, we observed a probability of misidentification of 0.18. With another 5 m of steps, we would again have a 0.18 probability of misidentification. Therefore, we can roughly estimate the probability of misidentifying all steps in 10 m should be  $0.18^2$  or 0.03. Over 1 km this would fall to roughly  $0.18^{(1000 \text{ m}/5 \text{ m})}$  or  $1 \times 10^{-149}$ . Thus, over longer data collections, the chances of misidentifying the side a gait event is occurring on becomes extremely small.

As seen in the recommendations above, users should balance reported accuracy of a method against the potential negative effects that running speed and foot strike angle could have on it. One approach to address these concerns and improve accuracy may be to include speed and foot strike angle as model inputs. For example, Patoz et al. [181] provided their contact and flight time estimation model with running speed data and reported impressively low error. Unfortunately, this method estimated the *mean* contact and flight times across a data set rather than the timing of individual gait events. Despite this limitation, Patoz et al. illustrate the potential benefits of including speed as a model input. Indeed, speed is relatively easy to calculate from GPS and several methods have already been proposed to estimate speed and foot strike angle from accelerations or angular velocities [182, 183, 184]. Thus, the inclusion of speed, foot strike angle, and other potential explanatory variables may be a relatively easy way to improve the performance of future models and should be a target of additional research.

Future methods should also compare their results against the existing methods presented here. As seen by the 18 methods examined here, the field is saturated with options for running gait event identification. Further, almost all methods were developed and evaluated in isolation and there have been few direct comparisons of methods. This saturation and lack of comparison may lead to confusion about which method is best for a given application.

With the continued development of this field, and of data collection and processing methods that can be used both in-lab and in-field, biomechanics will become increasingly accessible. Although increasing accessibility was a key goal of this study (through minimizing system complexity and data processing supervision), this increase in accessibility also brings challenges. For example, when deploying wearables in-field, they may be placed on participants by individuals with little training. This could lead to misalignment and degrade accuracy. To mitigate this potential issue, we used a segment coordinate system. In contrast, the majority of methods we studied here (16 of 18) originally used wearable coordinate systems defined by the axes of the sensors within the wearable. Using a wearable coordinate system is less reliable due to (1) differences in wearable manufacturing (sensor axes may not be perfectly aligned with their housing or each other), (2) participant geometry (e.g., tibial or sacral morphology may differ across participants), or (3) wearable placement (e.g., wearable may be placed upside down or at an ‘improper’ angle). Thus, we recommend (1) calibrating the wearable to ensure the output is accurately expressed in a consistent coordinate system [99] and (2) creating a segment coordinate system based off (a) the gravity vector during quiet standing and (b) the first Principal Component calculated from calibration motions such as rotation about the medial-lateral axis (e.g., legs swings or inverted pendulum about the ankles) [121, 175], and/or (b) the gravity vector while lying prone/supine. Using this approach, even if a wearable is ‘misaligned’ the segment coordinate system should be unaffected. These or similar methods will help mitigate the potential pitfalls of deploying wearables, particularly when deploying in-field. Although we believe this is the best practice approach, for the sake of comparison, we also executed our analyses on our data expressed in both wearable and global coordinate systems (see the Supplement 3). In line with our *a priori* expectation, these supplemental analyses show that—even though the same trained experimenter attempted to consistently fix and align every wearable in this study—the segment coordinate system had lower error than the wearable coordinate system across most methods. Indeed, the mean absolute error in the

methods recommended above was, on average, 59.35 ms lower when calculated in the SCS than the WCS.

Finally, in line with our goal to identify methods capable of easily processing large volumes of running data, we constrained our study to automated/unsupervised methods. Schmidt et al. [158] originally recommended that their method be used with thresholds defined by user input. Thus, we have included an option to run their method with user supervision in our code. However, given the goals of our paper, we did not include its results for comparison here. It is likely that with supervision, the errors observed for the Schmidt method would decrease.

### **3.5. Acknowledgements**

Thanks to Gerome Alejandro Manson for reviewing our code and Mijke Rhemtulla for statistical consultation. Thanks to Marcus Schmidt and Cagla Fadillioglu for their help implementing their original code and to Benson et al. [165] and Bach et al. [161] for providing their original code online.

### **3.6. Funding**

This research was funded in part by the American College of Sports Medicine Foundation Doctoral Student Research Grant (18-00671) and the Sigma Xi Scientific Research Honor Society Grant-in-Aid of Research (G2018031595024439). The APC was partially funded by the University of California Davis Library's Open Access Fund. DK was supported by a scholarship from the Natural Sciences and Engineering Research Council of Canada (CGSD3-502358-2017), the Maury L Hull Endowed Fellowship for Musculoskeletal Biomechanics Research, and a University of California Davis Professors for the Future Fellowship.

## 3S: Supplement to chapter 3

### 3S.1 Overview

Here we provide details for both our own IMU data processing and for each of the gait event identification methods included in our main paper. For details beyond what we provide here please refer to the original sources. Figures have been included showing the process each method uses to identify initial contact (IC) and terminal contact (TC) gait events. All data for these figures was taken from the same randomly selected trial (a 1.64 m tall 62.37 kg female running 3.73 m/s on the ‘floor’ surface and contacting the force plate with a forefoot strike angle of -0.12 rad). Wherever possible, figures originally published with each method were digitized (<https://apps.automeris.io/wpd/>) and run through the code provided with this paper to ensure we could faithfully reproduce results (not included here due to copyright). For clarity, we have standardized the original coordinate conventions used in each method to be consistent with the segment- and wearable-coordinate systems (SCS and WCS) described briefly in our main paper (Figure 3.3), described more fully in our IMU processing Supplement (below), and with ISB recommendations [126]. Finally, each method’s performance is individually reported here. Please note that figure axes and descriptions are consistent across all methods; consequently, the results for some methods may not be fully plotted if they exceed the axis range. Also note, the results figures are only fully described the first time they appear (Figure 3S.2).

#### *3S.1.a Timing Constraints*

For several methods, novel timing constraints were added to facilitate pattern recognition and exclude observations falling outside the values reported for steady-state running [100, 101, 102, 103,

104, 133, 106, 107, 75]. Based on these sources, we adopted a maximum step frequency of 4.75 Hz, stance times between 95 and 270 ms, and swing times between 200 and 600 ms.

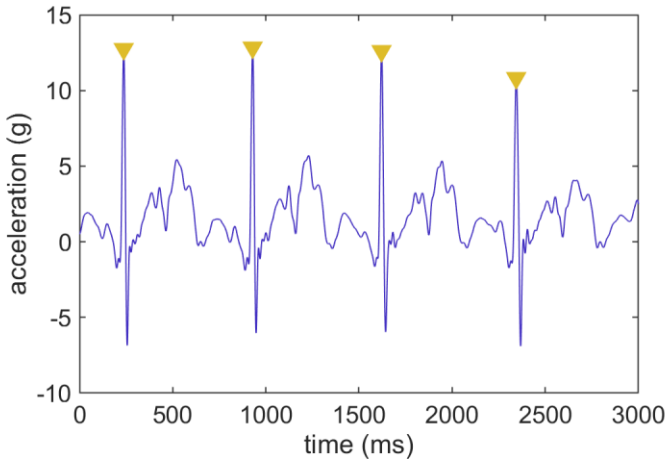
### *3S.1.b. Common output*

Finally, for all methods, to prevent errors (e.g., empty IC and TC outputs, impossible patterns of results such as IC-IC-TC-IC, etc.) that were common to many methods (all except: Mizrahi, Benson, Norris, Reenalda, and Whelan) and to ensure a common output, we implemented a function that took the raw IC, TC, and side output from each function and checked it to ensure: (1) gait events were identified (if a method was supposed to deliver both ICs and TCs, only paired gait events were accepted; i.e., no IC without TC), (2) gait event timings were unique and in temporal order, and (3) gait events followed a logical IC-TC-IC-TC pattern.

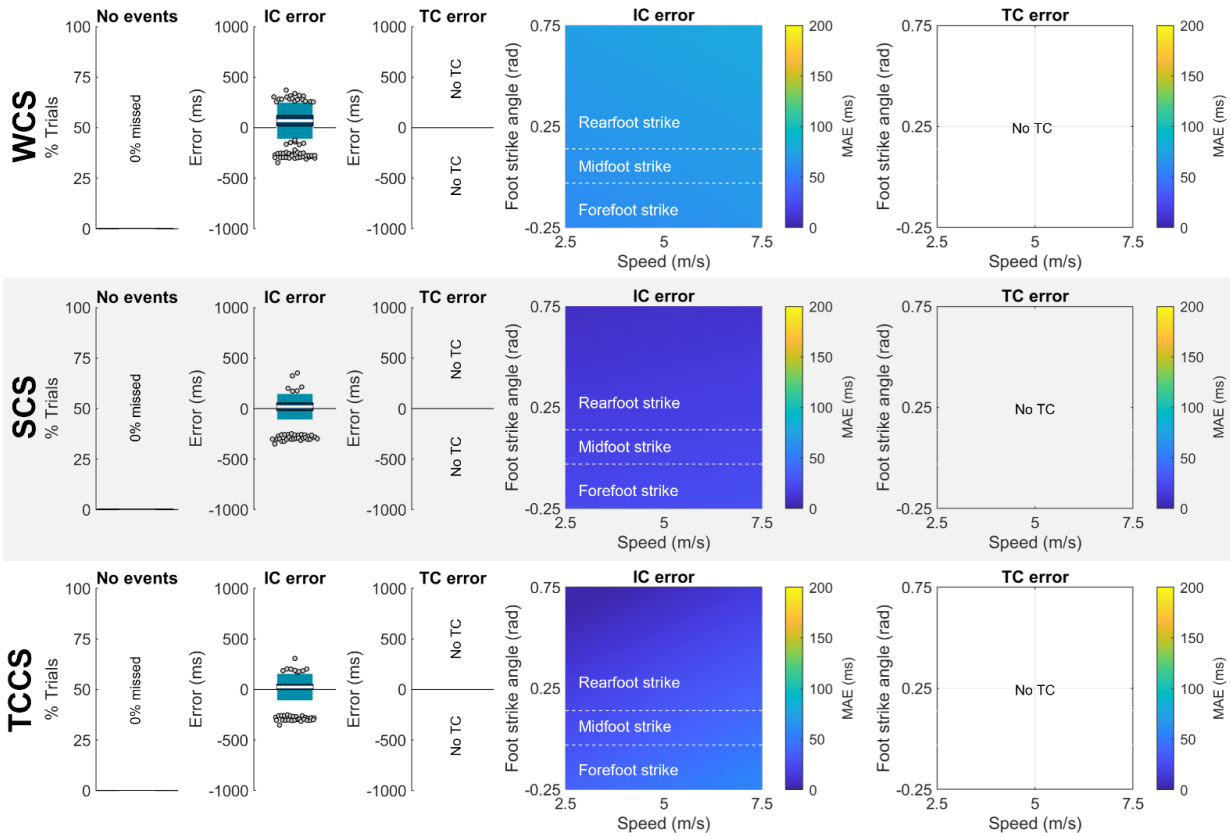
## **3S.2. Shank-mounted wearable methods**

### *3S.2.1. Mizrahi method*

Developing and validating a gait-event identification method was not the primary focus of Mizrahi et al. [148]. The authors did, however, report identifying IC from the peak acceleration of a uniaxial accelerometer on the tibial tuberosity aligned with the longitudinal axis of the tibia ( $a_{WCS,y}$ ). The authors reported using an automated code that we attempted to replicate based on their paper. Our code finds maxima in y-axis acceleration (proximal-distal in the SCS or ~longitudinal in the WCS) with a minimum separation of ~422 ms (based on the maximum step frequencies reported across previous studies) (Figure 3S.1). For results see Figure 3S.2.



**Figure 3S.1: Mizrahi method.** Y-axis acceleration (proximal-distal in the SCS; ~longitudinal in the WCS; dark blue line) maxima are found with a minimum separation of ~422 ms and selected as IC gait events (gold triangles). No TC identification.



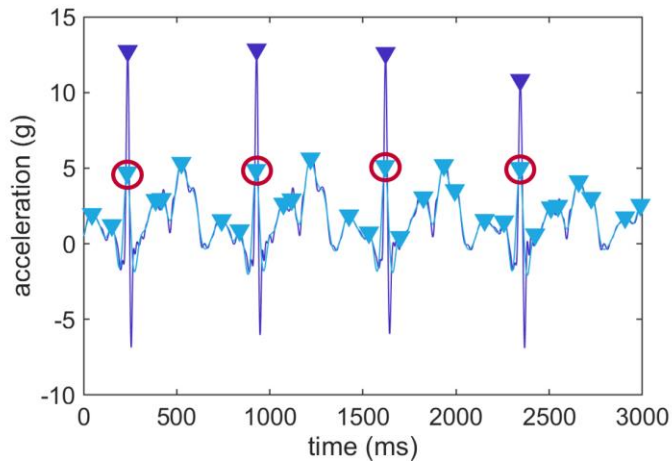
**Figure 3S.2: Mizrahi results.** Results were calculated using the methods described in each coordinate system (calculated as described in part 8 of our IMU processing breakdown at the end of this Supplement). Results in the top row are for the wearable coordinate system (WCS), results in the middle row are for the segment coordinate system (SCS), and results in the bottom row are for a pseudo-global coordinate system, the “tilt-corrected” coordinate system (TCCS). Reading left to right, in the first column the white horizontal line represents the mean percentage of trials per participant without any gait events recognized, dark blue represents  $\pm 95\%$  confidence interval ( $\pm 1.96$  SEM) around the mean, light blue represents  $\pm 1$  SD around the mean. Gray dots represent participants outside  $\pm 1$  SD. No bars indicates that gait events were identified in every trial for every participant. In the second column IC error means (white bar),  $\pm 1$  within-method SD (dark blue), and  $\pm 95\%$  LOA (1.96 SD of

errors; light blue) are plotted. Gray dots represent trials falling outside the 95% LOA. A value of 0 indicates perfect agreement with the ground truth. Positive values indicate the IC was estimated later than the ground truth (after the force plate IC). Negative values indicate the IC was estimated earlier than the ground truth (before the force plate IC). In the third column TC error means (white bar),  $\pm 1$  within-method SD (dark blue), and  $\pm 95\%$  LOA (1.96 SD of errors; light blue) are plotted. Gray dots represent trials falling outside the 95% LOA. A value of 0 indicates perfect agreement with the ground truth. Positive values indicate the TC was estimated later than the ground truth (after the force plate TC). Negative values indicate the TC was estimated earlier than the ground truth (before the force plate TC). In the fourth column mean absolute error in IC predicted by a mixed effects model. Plotted as a function of speed and foot strike angle. Darker blue values represent lower predicted mean absolute errors while brighter yellow values represent higher predicted mean absolute errors. In the fifth column mean absolute error in TC predicted by a mixed effects model. Plotted as a function of speed and foot strike angle. Darker blue values represent lower predicted mean absolute errors while brighter yellow values represent higher predicted mean absolute errors.

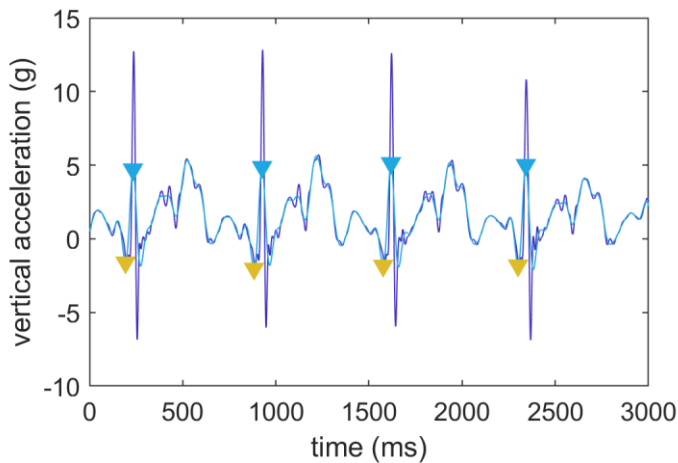
### 3S.2.2. Mercer method

Developing and validating a gait-event identification method was not the primary focus of Mercer et al. [149], however, they provided the first published details on a gait-event identification method previously used by Shorten and Winslow [185], Hamill et al. [186], and Derrick et al. [187, 188]. We attempted to translate these details into an automated code. To identify IC, the authors reported placing an accelerometer on the anteromedial distal tibia and finding a local minimum immediately preceding a global maximum in  $a_{WCS,y}$  (Figure 3S.3A-C). To identify TC, they reported finding the “minimum after a second local maximum.” We found that small local maxima could generate results inconsistent with the figures presented in Mercer’s paper when looking for local maxima to identify TCs (indeed, their figure depicts the TC after three small local maxima). Thus, to exclude smaller local maxima, we added an additional constraint that only the four largest maxima in the window of interest would be considered. Results are shown in Figure 3S.4.

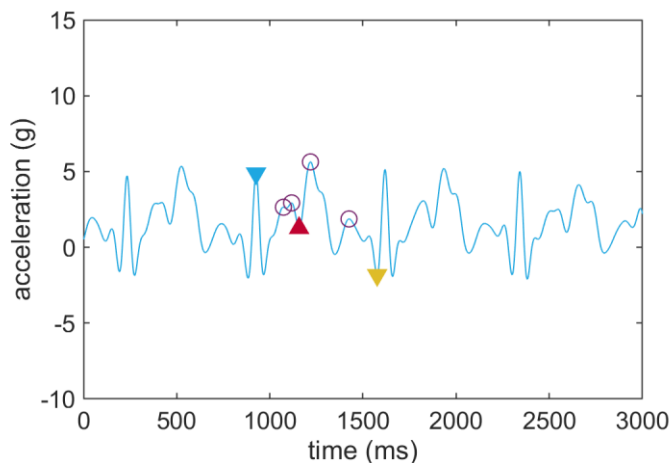




**Figure 3S.3: Mercer method. (A)** Y-axis acceleration (proximal-distal in the SCS or ~longitudinal in the WCS; dark blue line) maxima (dark blue triangles) are found with a minimum separation of ~422 ms. The signal is then low-pass filtered at 15 Hz (light blue line) and maxima are found in the filtered signal with no minimum separation criterion (light blue triangles). The peak in the filtered signal closest to the peak in the original signal is selected as a peak of interest (red circle).



**(B)** Walking back from each peak of interest (light blue triangles) the first point when the signal magnitude begins to increase is selected as an IC gait event (gold triangles)



**(C)** The four largest maxima (red circles) are found between each peak of interest (light blue triangle) and its preceding IC (gold triangle). Stepping forward from the second maxima, the first point when the signal magnitude begins to increase is selected as a TC gait event (red triangle).

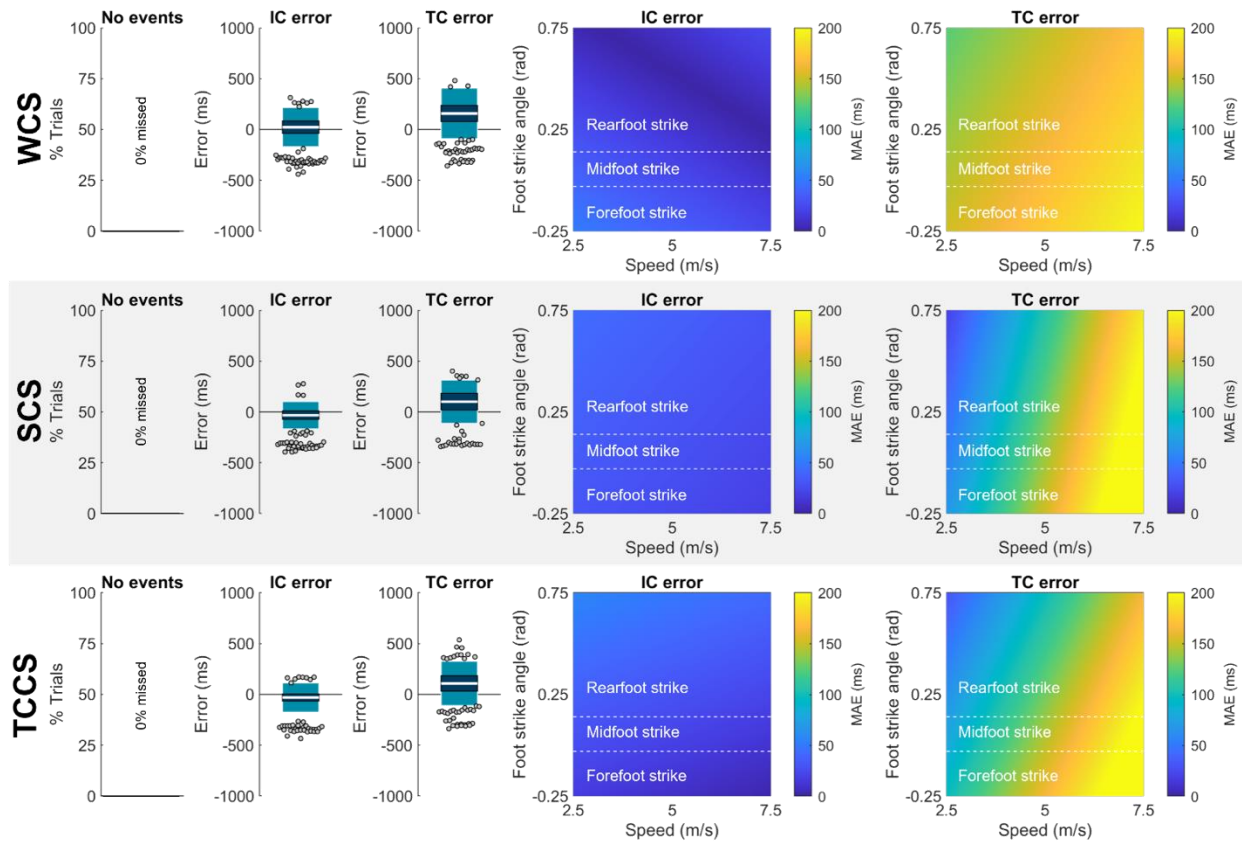
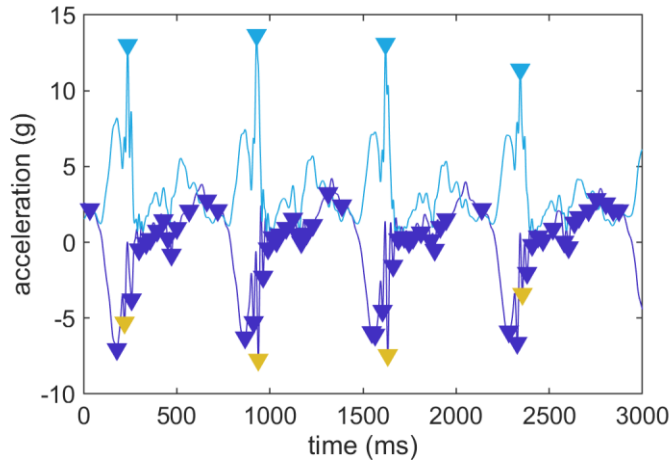


Figure 3S.4: Mercer results. See Figure 3S.2 for description.

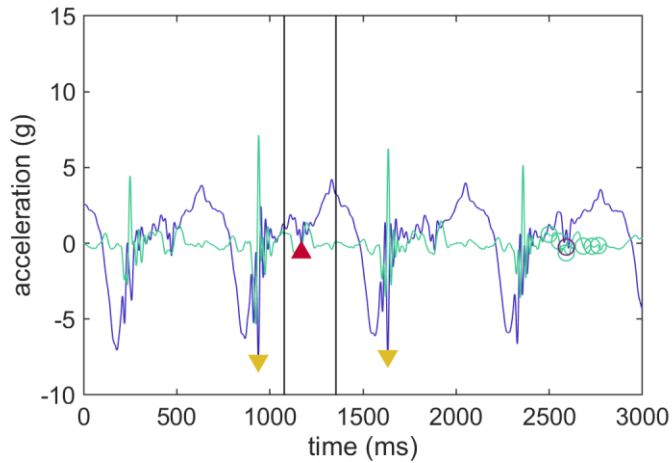
### 3S.2.3. Purcell method

Purcell et al. [150] placed a tri-axial accelerometer on the anteromedial tibia. To define IC, the authors report using minima in  $a_{WCS,x}$  (as defined by the coordinate convention they report standardized to ours) that corresponds in time with a maxima in  $a_{WCS,res}$ . However, the data they present (and the method which we were successfully able to reproduce) suggest that they misreported their coordinates and actually used minima in  $a_{WCS,x}$  (rather than maxima). To define TC, they take the average time stamp of local minima in  $a_{WCS,x}$  and  $a_{WCS,z}$  (after correcting their coordinate system) (Figure 3S.5A-B). The  $a_{WCS,z}$  minima were not very pronounced; thus, to aid performance, we imposed a

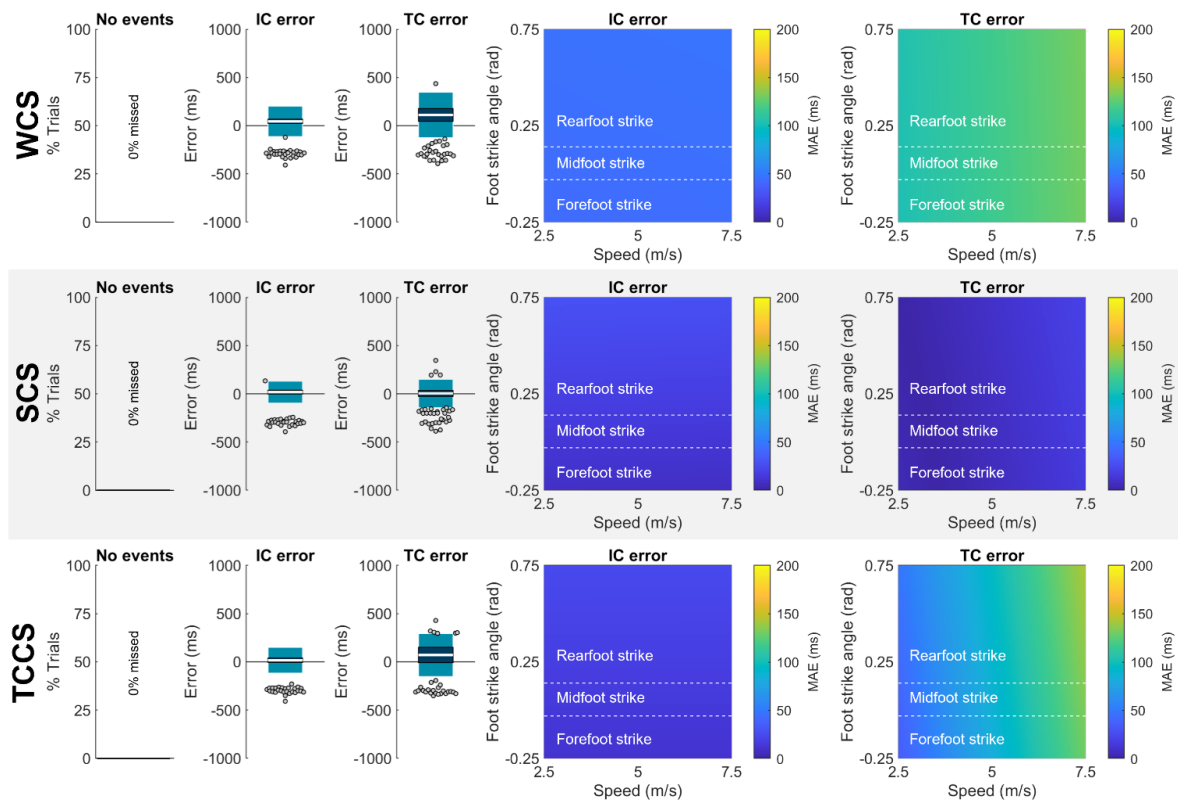
further constraint here that the window to search for these local minima should be between 20-60% of one IC to the next. Results are shown in Figure 3S.6.



**Figure 3S.5: Purcell method. (A)** X-axis acceleration (anterior-posterior in SCS, ~direction of progression in WCS; dark blue line) minima (dark blue triangles) are identified. Then, the resultant acceleration is calculated (light blue line) and peaks separated by at least ~422 ms are identified (light blue triangles). IC is defined as the x-axis minima closest to resultant maxima (gold triangle).



**(B)** A window of interest (vertical black lines) is created from 20-60% of one IC to the preceding IC (gold triangles). The greatest magnitude x-axis minima in that window (blue circle) is found. All z-axis (medial-lateral in SCS, ~right-left in WCS) minima within that window (green circles) are found. The z-axis minimum closest to the large x-axis minimum (purple circle) is found. TC is defined as the midpoint between those two points (red triangle).

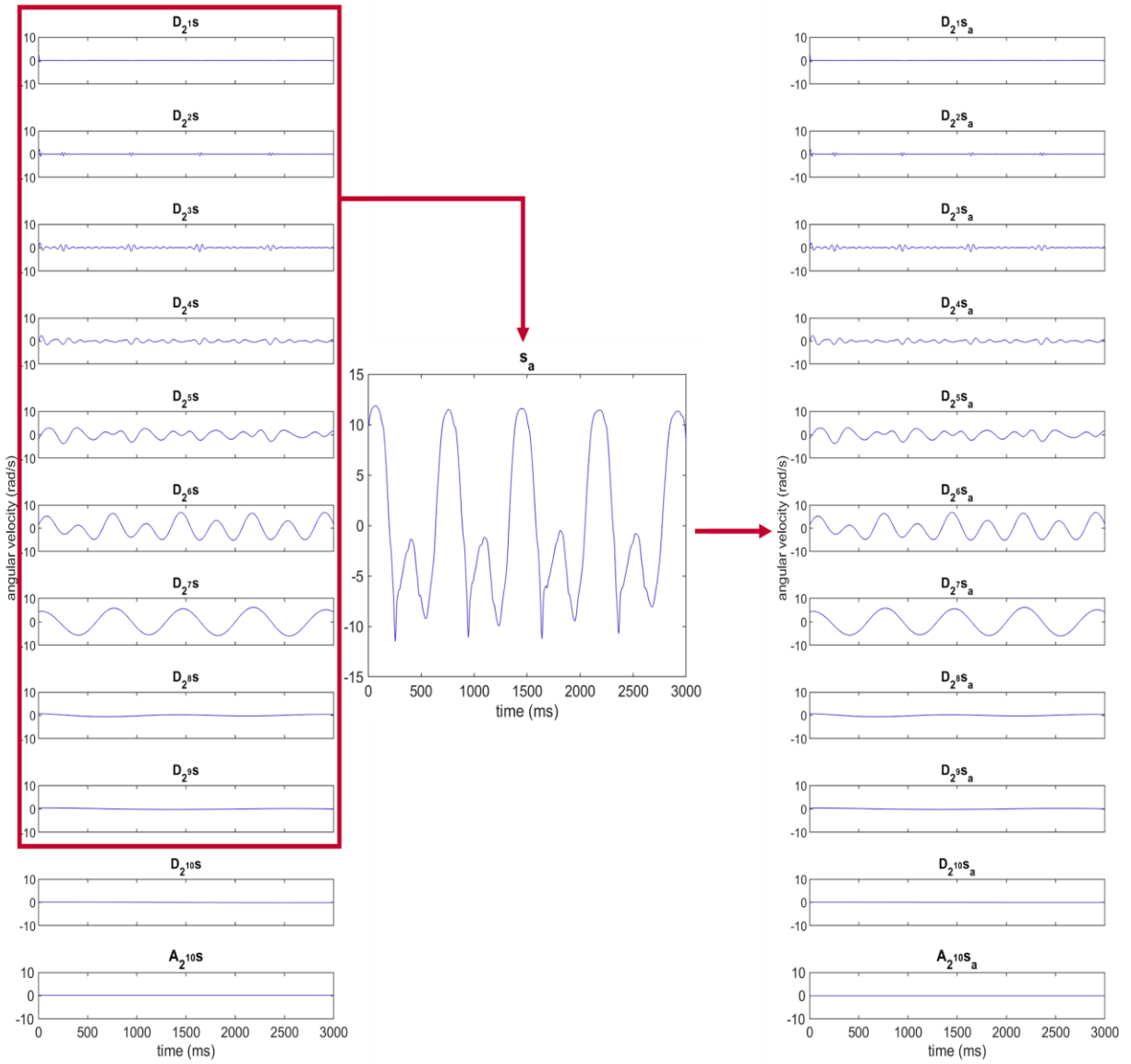


**Figure 3S.6: Purcell results.** See Figure C3S.2 for description.

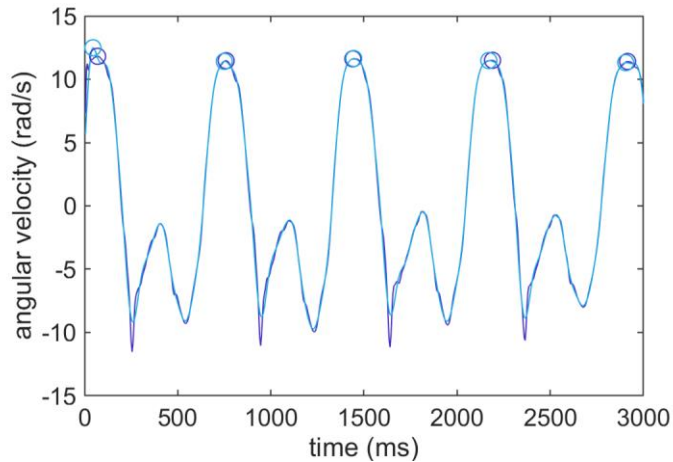
### 3S.2.4. Aminian/O'Donovan method

Aminian et al. [151] developed a method to identify gait events during walking that was later applied to running by O'Donovan et al. [152]. This method decomposes  $\omega_{SCS,z}$  using a series of wavelet multi-resolution analyses (MRA) that split the signal into high-scale ‘approximation’ (low-frequency) and low-scale ‘detail’ (high-frequency) components. Aminian et al. report using two iterative 10-level 5th order Coiflet wavelet analyses. Given each level in MRA represents a frequency band that is a function of the sampling frequency used during data collection, we first resample data to match Aminian et al.’s reported 200 Hz sampling frequency (thus, yielding frequency bands that should be identical to theirs for each level; note, however, that Aminian et al. state their approximation “only considers” up to 36 Hz—consistent with a 72 Hz sampling frequency and not the reported 200 Hz sampling frequency—it is

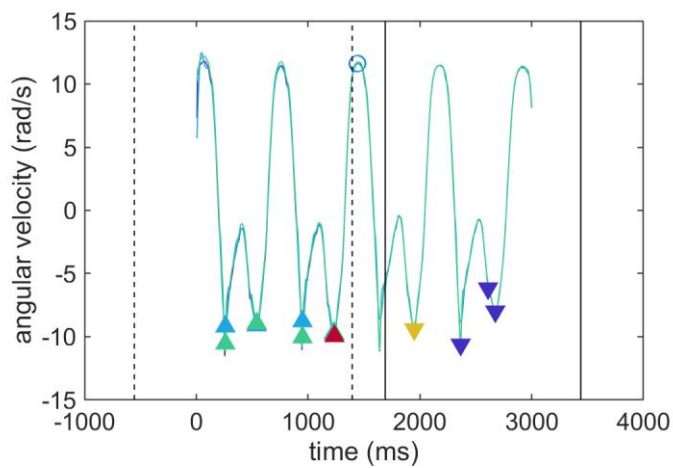
unclear where this discrepancy originates from but for the purposes of this paper we adhered to their stated sampling frequency and equations). After resampling, MRA is used to obtain a new signal approximation (“ $s_a$ ”) by summing details 1:9. Then, a second MRA is conducted on  $s_a$  and two new signal approximations are created: (1)  $A_{2^1}s_a - A_{2^9}s_a$  is designed to enhance the IC component by subtracting approximation 9 from approximation 1 and (2)  $A_{2^3}s_a - A_{2^9}s_a$  is designed to enhance the TC component by subtracting approximation 9 from approximation 3. These new signal approximations are then searched for maxima/minima within specific time constraints. Results shown in Figure 3S.8.



**Figure 3S.7: Aminian/O'Donovan method.** Angular velocity about the z-axis (medial-lateral in the SCS, ~right in the WCS) is resampled to 200 Hz then decomposed with a 10-level 5<sup>th</sup> order Coiflet wavelet multi-resolution analysis (MRA) **(A)**. Details 1 through 9 are summed to yield signal approximation  $s_a$  **(B)**.  $s_a$  is then entered into a second 10-level 5<sup>th</sup> order Coiflet wavelet MRA **(C)**.



**(D)** Two new approximations are created:  $A_{2^1}s_a - A_{2^0}s_a$  (dark blue line) and  $A_{2^3}s_a - A_{2^0}s_a$  (light blue line). Maxima separated by at least  $\sim 422$  ms are found in each signal approximation (circles). These maxima approximately correspond to mid-swing.



**(E)** A window of interest preceding each maxima in  $A_{2^1}s_a - A_{2^0}s_a$  (+250 to +2000 ms; solid vertical lines) is used to look for local minima corresponding to potential ICs (dark blue triangles). A window of interest preceding each maxima in  $A_{2^3}s_a - A_{2^0}s_a$  (-2000 to -50 ms; dashed vertical lines) is used to look for local minima (light blue triangles) then the minimum values in the original signal  $s$  (green line) within 0 to +75 ms are selected as potential TCs (green triangles). Starting with the IC closest to the mid-swing maxima in time iterate backwards in time through potential TCs. The first potential IC-potential TC pair that satisfies the condition  $100 \text{ ms} < (\text{IC} - \text{TC}) < 2500 \text{ ms}$  is accepted. If no pairing meets this condition iterate forward in time to the next potential IC. Note, the windows of interest stipulated by Arminian et al. for walking analysis were so large that they required the x-scale to be adjusted and extend beyond the data selected for analysis.

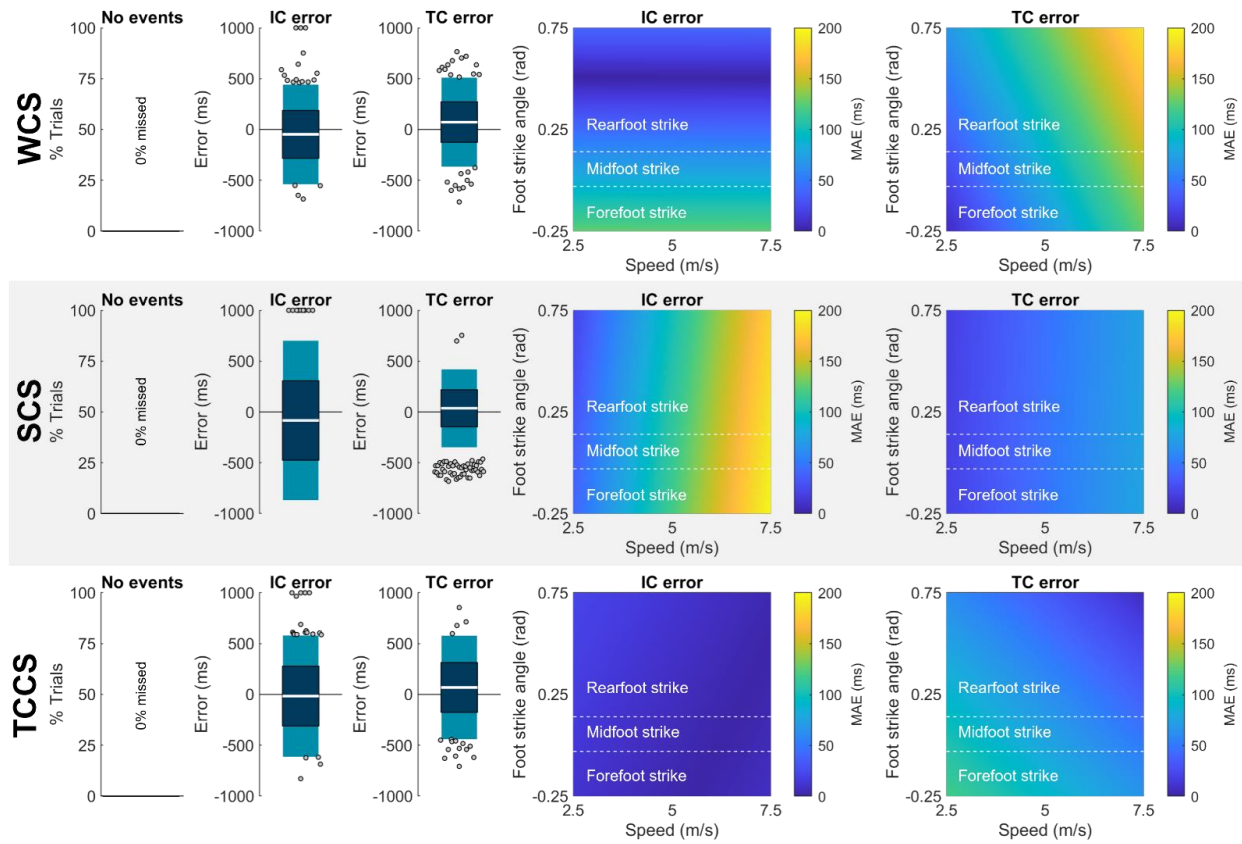
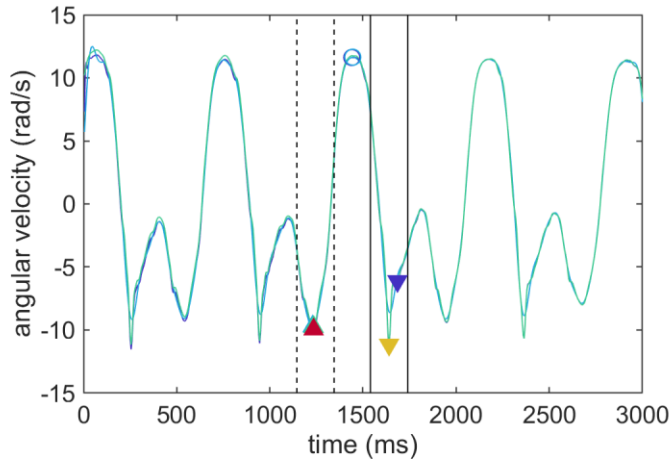


Figure 3S.8: Aminian/O'Donovan results. See Figure 3S.2 for description.

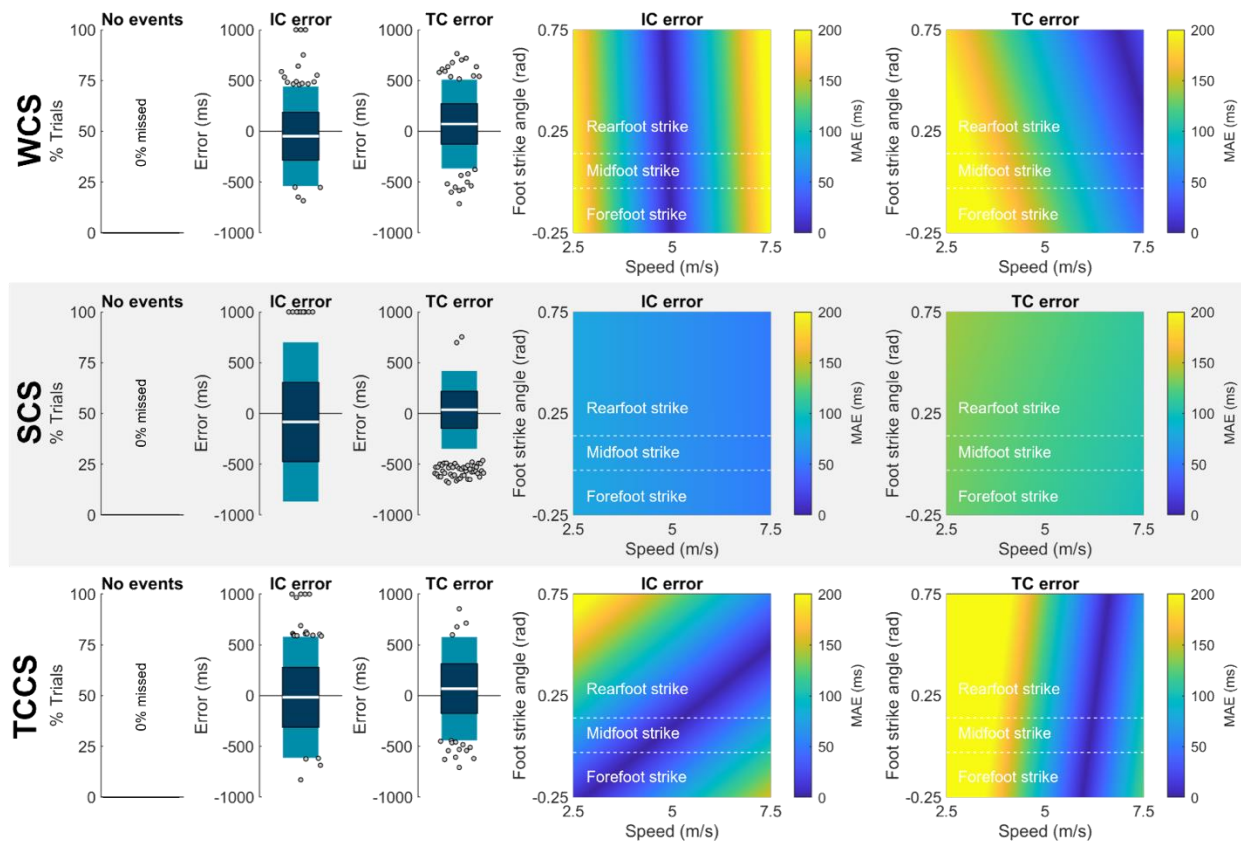
### 3S.2.5. Aminian/O'Donovan modified method

As stated in the previous section, Aminian et al. [151] developed their method to identify gait events in walking. O'Donovan et al. [152] later applied this method to running but did not report any adjustments to the windows of interest based on the temporal differences between walking and running. Here we adjust the windows of interest to better correspond to swing times presented in the running literature. The approach is otherwise as described above. Results shown in Figure 3S.10.





**Figure 35.9: Aminian/O'Donovan modified method.** New running-based windows of interest are defined for the Aminian/O'Donovan approach. A window of interest preceding each maxima in  $A_{2^1}s_a - A_{2^9}s_a$  (+100 to +300 ms; solid vertical lines) is used to look for local minima corresponding to potential ICs (dark blue triangles). A window of interest preceding each maxima in  $A_{2^3}s_a - A_{2^9}s_a$  (-300 to -100 ms; dashed vertical lines) is used to look for local minima (light blue triangles) then the minimum values in the original signal  $s$  (green line) within 0 to +75 ms are selected as potential TCs (green triangles). Starting with the IC closest to the mid-swing maxima in time iterate backwards in time through potential TCs. The first potential IC-potential TC pair that satisfies the condition  $200 \text{ ms} < (\text{IC} - \text{TC}) < 600 \text{ ms}$  is accepted. If no pairing meets this condition iterate forward in time to the next potential IC.



**Figure 35.10: Aminian/O'Donovan modified results.** See Figure 35.2 for description.

### 3S.2.6. Greene/McGrath method

Greene et al. [154] developed a method for gait event identification during walking that was later adapted by McGrath et al. [153] to identify running gait events. McGrath et al. placed IMUs mid-shank to capture  $\omega_{WCS,z}$ . Data were low-pass filtered with a 5 Hz 5<sup>th</sup> order Butterworth filter then thresholds calculated based on the data's properties were used to determine IC and TC. Finally, the sequence and temporal spacing of gait events was checked to ensure they were logical (i.e., must follow IC-TC-IC-TC pattern). Results are shown in Figure 3S.12.

**Equation 3S.1**

$$th_1 = 0.6 * \max(\omega_{WCS,z})$$

**Equation 3S.2**

$$th_2 = 0.8 * \frac{1}{N} \sum_{i=1}^N (\omega_{WCS,z,i} > \underline{\omega}_{WCS,z})$$

**Equation 3S.3**

$$th_3 = 0.8 * \left| \frac{1}{N} \sum_{i=1}^N (\omega_{WCS,z,i} < \underline{\omega}_{WCS,z}) \right|$$

**Equation 3S.4**

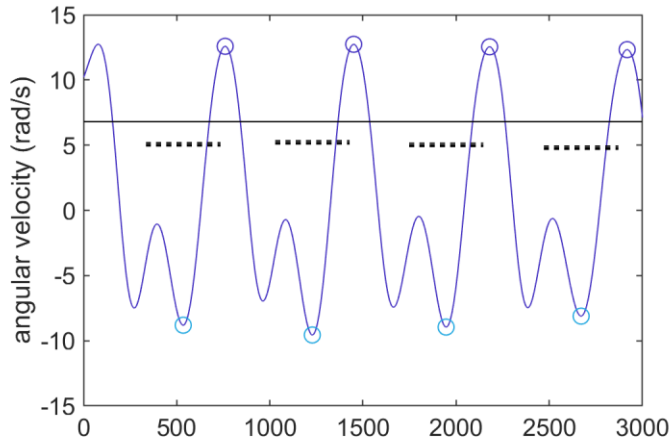
$$th_4 = 0.8 * \frac{1}{N} \sum_{i=1}^N (\omega_{WCS,z,i} < \underline{\omega}_{WCS,z})$$

**Equation 3S.5**

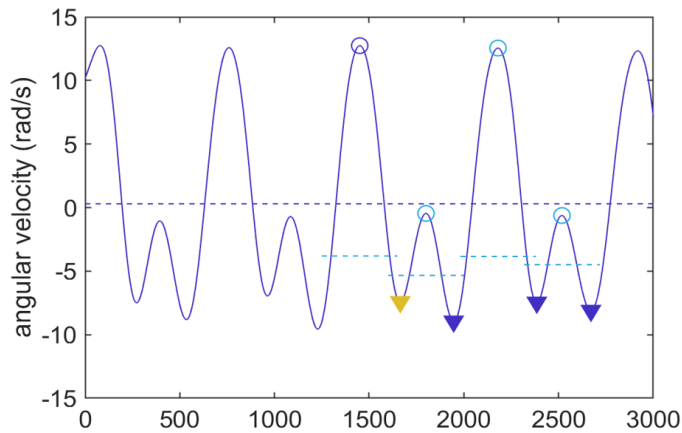
$$th_5 = \bar{\omega}_{WCS,z}$$

**Equation 3S.6**

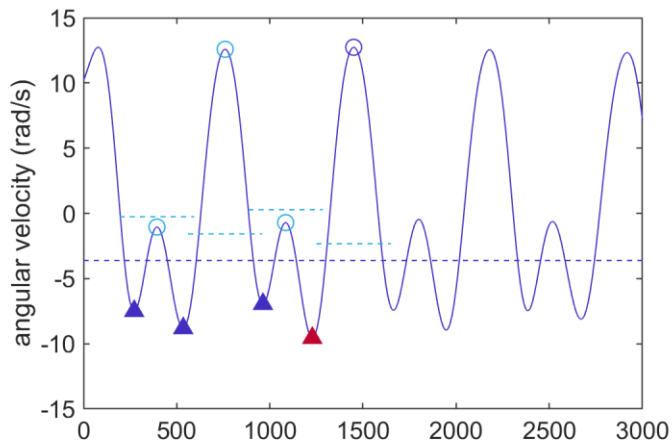
$$th_6 = 2 * th_3$$



**Figure 3S.11: Greene/McGrath method. (A)** Angular velocity about the z-axis (medial-lateral in the SCS, ~right in the WCS; dark blue line) is filtered with a 5-Hz 5<sup>th</sup> order low-pass Butterworth filter. Maxima separated by at least  $t_1$  (500 ms) are identified (dark blue circles). Any identified maximum without a preceding minimum (light blue circles) at least  $th_1$  below the maximum (dashed lines) or with a value below  $th_2$  (solid line) is discarded.



**(B)** Iterating backwards through the remaining maxima (dark blue circle) preceding minima are identified as potential ICs (dark blue triangle). Potential ICs are rejected if their preceding maxima (light blue circle) is not at least  $th_3$  greater than their magnitude (dashed light blue line) or if their magnitude is  $th_5$  or greater (dashed dark blue line). The potential IC satisfying these conditions and closest in time to the maxima is labeled as the IC.



**(C)** Similarly, minima preceding the maxima are identified as potential TCs (dark blue triangles). Potential TCs are rejected if their magnitudes are  $th_4$  or greater (dashed dark blue line) or their preceding maxima (light blue circle) is not  $th_6$  greater than the potential TC (dashed light blue line; note: Greene et al. stipulate this threshold applies to the preceding maximum; however, we found that maxima preceding TCs rarely meet this condition—including on Greene et al.'s own data digitized via <https://apps.automeris.io/wpd/>—thus, we believe this was a typo and that this condition is intended to be applied to proceeding maxima). The timing and sequence of events is then checked to ensure events always alternate between IC and TC, ICs and their preceding TCs occur within 2500 ms of each other, and TCs and their preceding ICs occur within 7500 ms of each other.

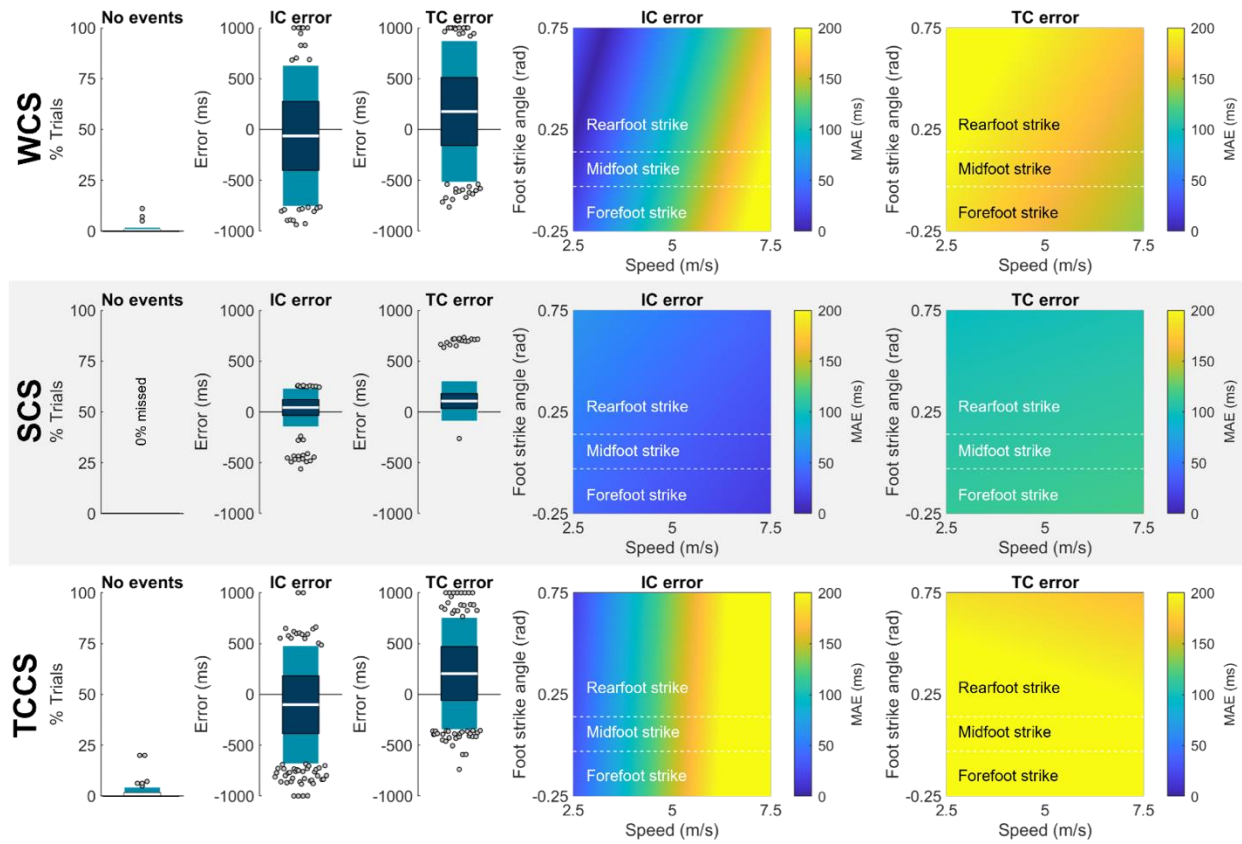


Figure 3S.12: Greene/McGrath results. See Figure 3S.2 for description.

### 3S.2.7. Greene/McGrath modified method

As stated in the previous section, Greene et al. [154] developed their method to identify gait events in walking. McGrath et al. [153] later applied this method to running but did not report any adjustments to the algorithm based on the temporal differences between walking and running. Here we adjust the algorithm to better correspond to the timing between gait events presented in the running literature. The approach is otherwise as described above. Results are shown in Figure 3S.13.

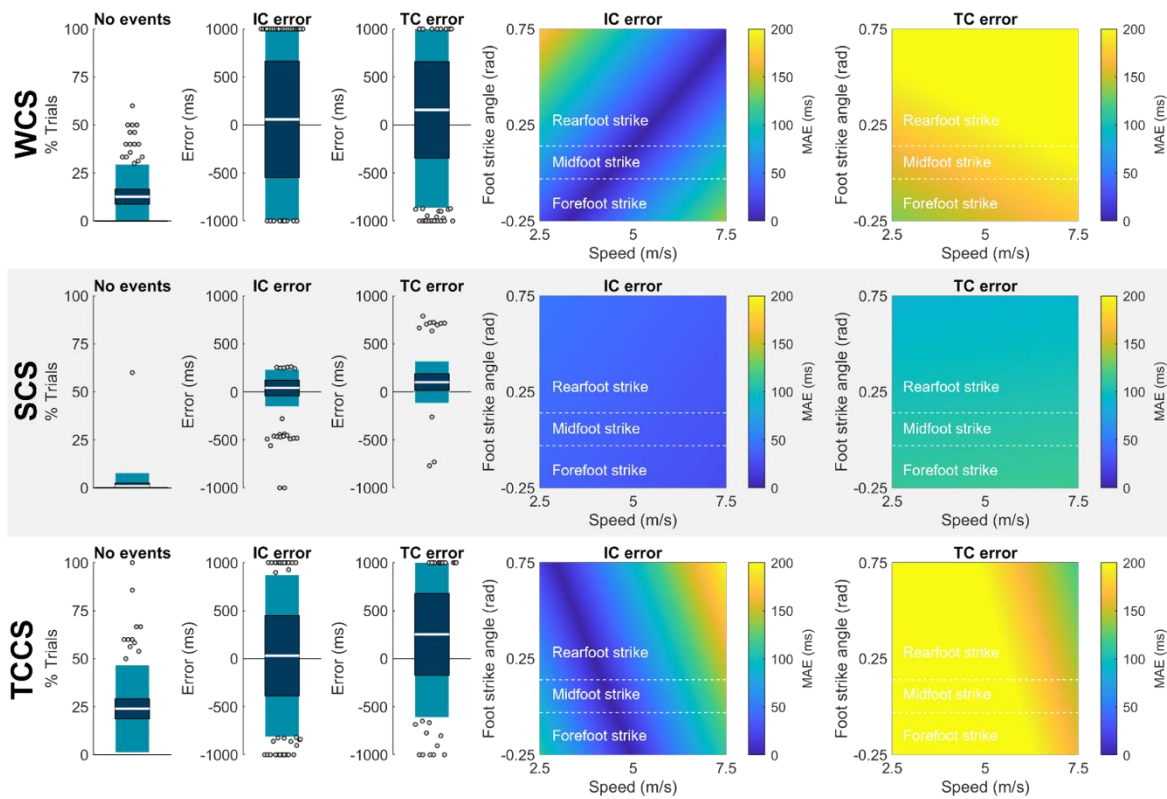
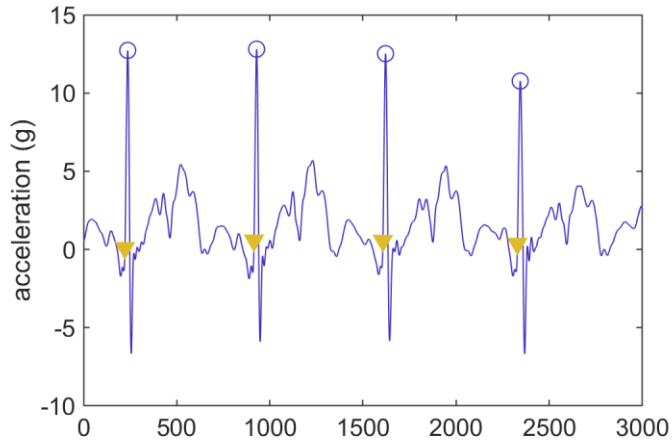


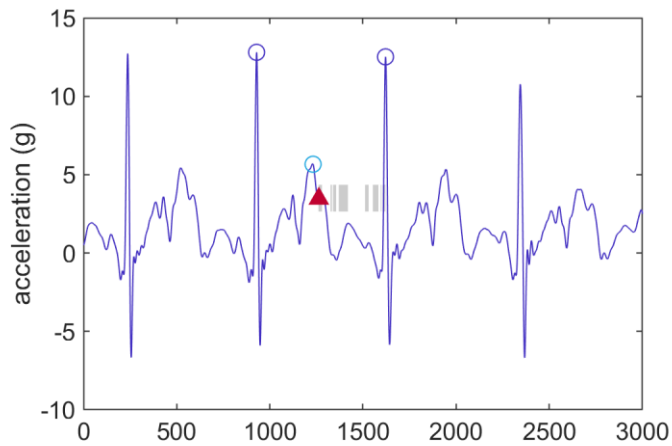
Figure 3S.13: Greene/McGrath modified results. See Figure 3S.2 for description.

### 3S.2.8. Sinclair method

Sinclair et al. [155] placed a tri-accelerometer on the anteromedial tibia and used  $a_{WCS,Y}$  to identify IC and TC gait events. Data were low-pass filtered then maxima in the  $\sim$ longitudinal acceleration were identified. IC was defined as the zero-crossing point preceding each maxima while TC was defined as a plateau following the largest local maximum between longitudinal maxima. Results are shown in Figure 3S.15.



**Figure 3S.14: Sinclair method. (A)** Y-axis acceleration (proximal-distal in the SCS; ~longitudinal in the WCS; blue line) is filtered with a 4<sup>th</sup>-order 60-Hz low-pass Butterworth. Maxima separated by at least ~422 ms are identified (blue circles). Walking back from each maxima, the first zero-crossing point followed by 20 ms of positive data is defined as the IC.



**(B)** To identify TC, a local maximum (light blue circle) is found between pairs of maxima. Data between this local maximum and the preceding y-axis maximum is differentiated and portions with where the magnitude is decreasing by less than 2% of the previous frame's magnitude per second are selected (gray shading). TC is defined as the first frame of the selection occurring earliest in time with a duration of at least ~5 ms.

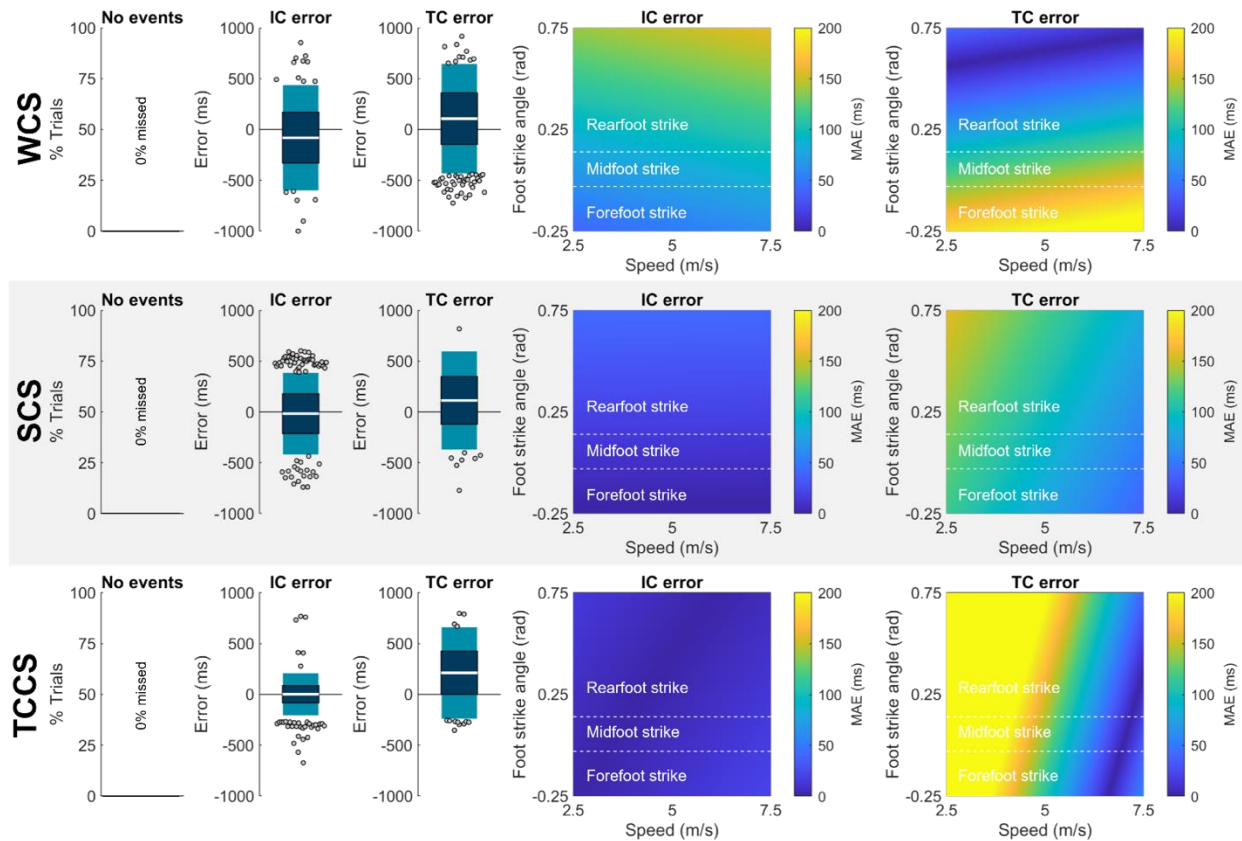
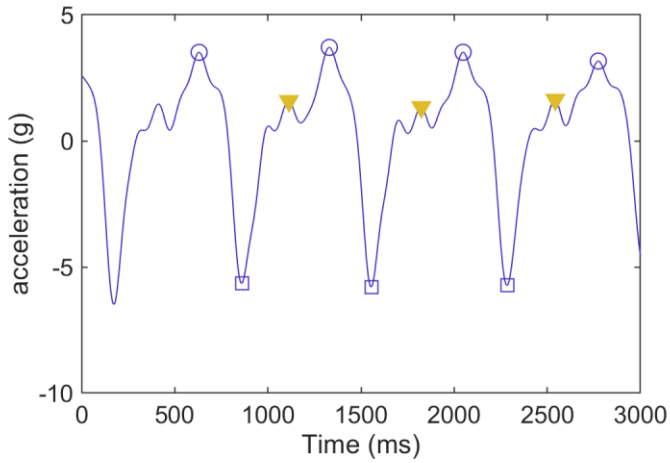


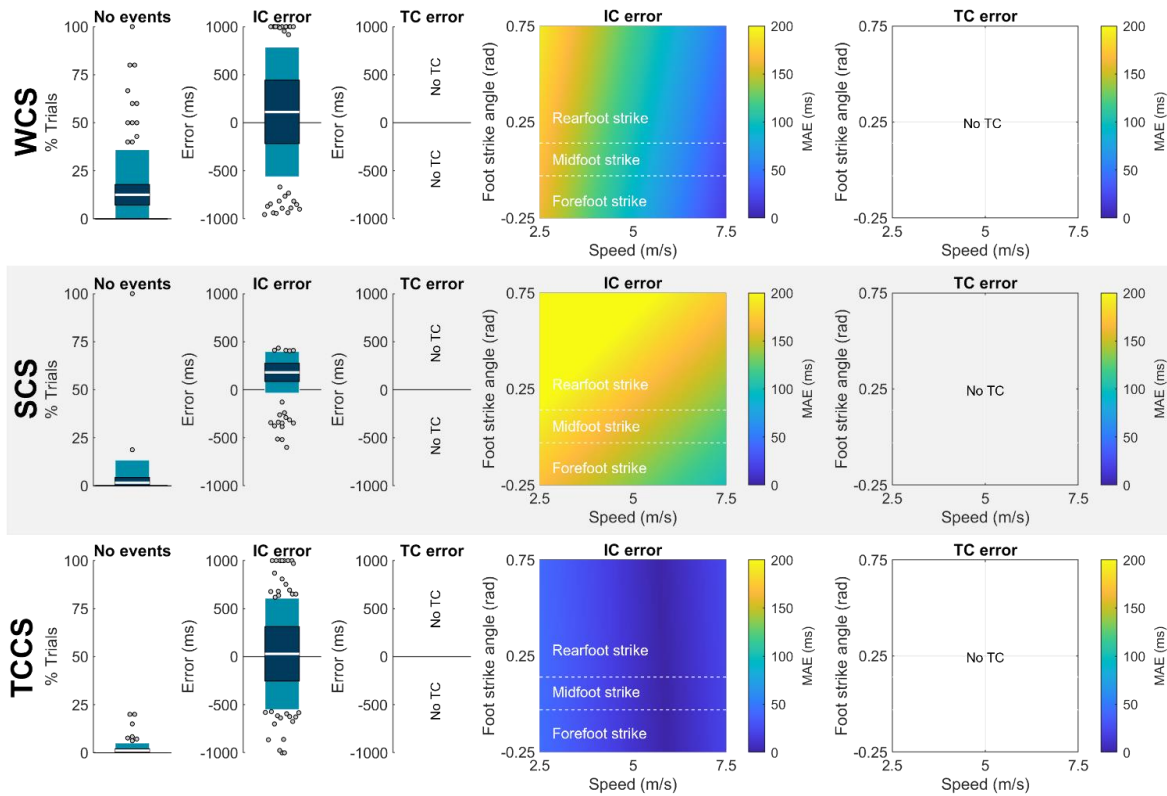
Figure 3S.15: Sinclair results. See Figure 3S.2 for description.

### 3S.2.9. Whelan method

Whelan et al. [156] used  $a_{WCS,X}$  from a triaxial accelerometer placed on the tibialis anterior (anteromedial on the midshaft of the tibia) to identify IC events in sprinters running at up to 50% of their maximum effort. Data were first filtered using a 10 Hz low-pass Butterworth filter (order not reported; we assumed 4<sup>th</sup>). Then they reported using “peak acceleration” to define IC. Their figures, however, suggest that they defined IC using a local maximum immediately preceding a larger maximum. We developed automated code based on their figure. Results are shown in Figure 3S.17.



**Figure 3S.16: Whelan method.** X-axis acceleration (anterior-posterior in the LCS;  $\sim$ direction of progression in the WCS) was filtered using a 4<sup>th</sup> order 10 Hz low-pass Butterworth filter (dark blue line). Maxima were then found in the signal (we added a constraint that they must be separated by at least  $\sim$ 422 ms) (dark blue circles). The minimum value between successive maxima was then found (dark blue squares). Finally, a local maximum between the minimum and it's proceeding maximum was found and labelled as the IC (yellow triangles). TC was not identified using this method.

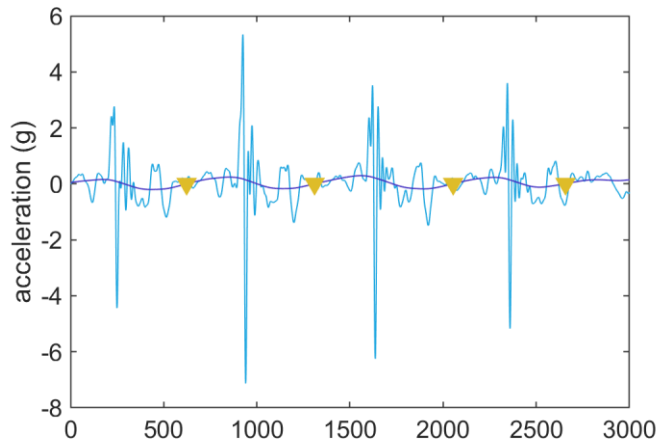


**Figure 3S.17: Whelan results.** See Figure 3S.2 for description.



### 3S.2.10. Norris method.

Norris et al. [157] proposed a running gait event identification method where  $a_{WCS,Z}$  from a tri-axial accelerometer placed on the anteromedial distal tibia were filtered with a 2 Hz 2<sup>nd</sup> order low-pass Butterworth filter and a zero-crossing was used to identify IC. Results are shown in Figure 3S.19.



**Figure 3S.18: Norris method.** Z-axis acceleration (medial-lateral in the SCS; ~right in the WCS; light blue line) is filtered with a 2<sup>nd</sup>-order 2 Hz Butterworth filter (dark blue line). IC is defined as any positive zero-crossing point. No TC definition is provided.

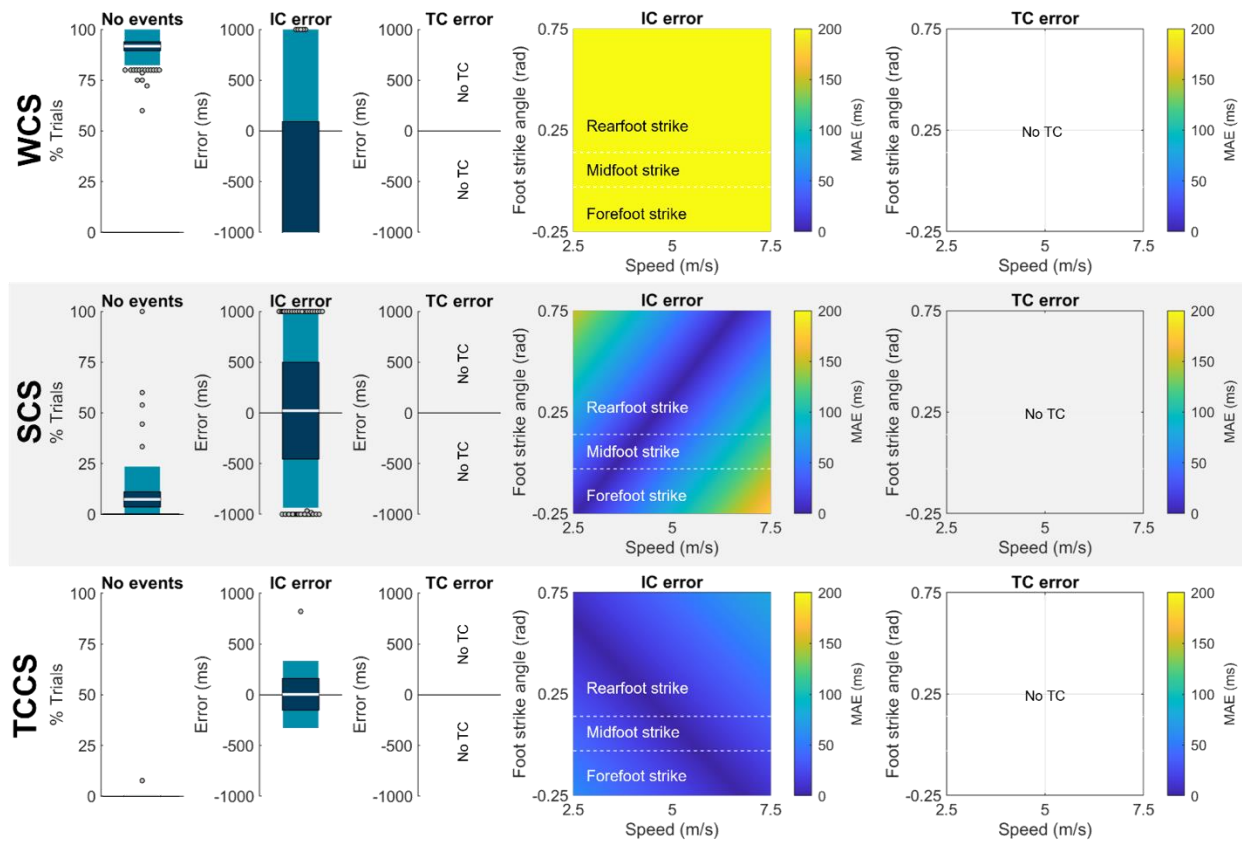
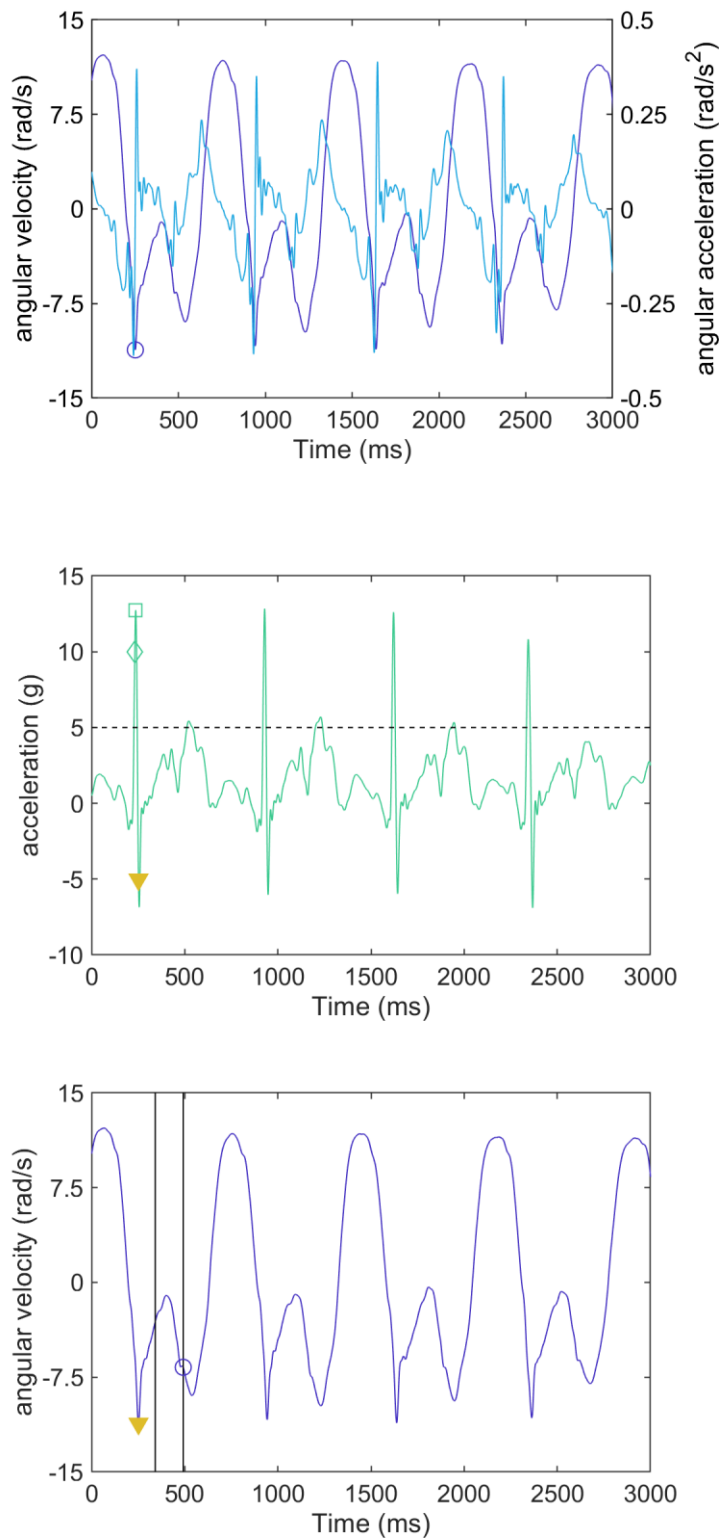


Figure 3S.19: Norris results. See Figure 3S.2 for description.

### 3S.2.11. Schmidt method.

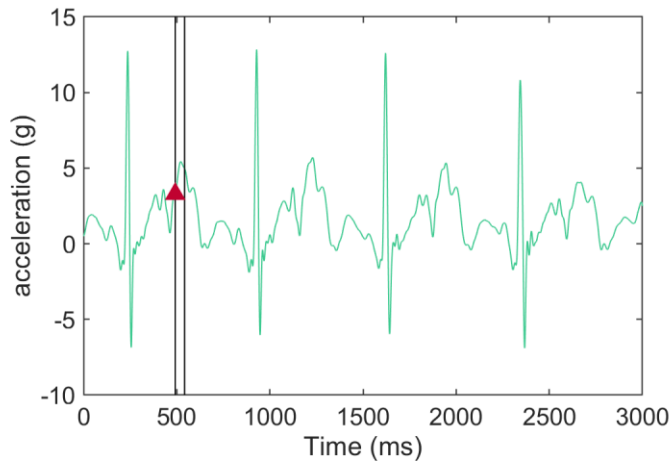
Schmidt et al. [158] developed a gait event identification method for sprinting. It should be noted that this method was not developed with, or intended for, use across a broad range of running speeds. To identify IC and TC events, Schmidt et al. placed a triaxial IMU on the distal lateral tibia and identified points in  $a_{WCS,Y}$  and  $\omega_{WCS,Z}$  that met user-defined thresholds. Given our focus on unsupervised methods, however, we adapted this code to use default thresholds. We report only the unsupervised results here, it is likely that results would improve with supervision. Results are shown in Figure 3S.21.



**Figure 3S.20: Schmidt method. (A)** Find the global minimum in angular velocity about the z-axis (medial-lateral in SCS; ~right in WCS; dark blue line). Then, walk through each frame of data and find the local minimum in angular velocity across the next 100 ms and the differential in angular velocity (angular acceleration; light blue line) for the next 50 ms. If the local minimum is equal to the global minimum or all angular accelerations are negative and the linear acceleration in the y-axis (proximal-distal in the SCS; ~longitudinal in the WCS; green line; **B**) exceeds a critical threshold (5 g or user-defined; dashed black line) then flag the current frame as a point of interest (In the example data, a single frame meets these criteria—the global minimum in angular velocity about the y-axis; dark blue circle).

**(B)** For each point of interest, find the maximum in y-axis acceleration occurring between 20 ms before to 100 ms after the point of interest (green square). Then, find the y-axis acceleration minimum occurring between 20 ms before the point of interest to the y-axis acceleration maximum (green diamond). If the minimum occurs after the point of interest label it as the IC; otherwise, label the point of interest as the IC (In this case, the point of interest was labeled as the IC; dark blue triangle).

**(C)** Next, look through all ICs. Within a 150 ms (or user defined) time window starting 90 ms (or user defined) after the IC (vertical black lines) find the minimum angular velocity (dark blue circle).



**(D)** Starting from this minimum, within a ~50 ms window (black vertical lines) and find the minimum y-axis acceleration (in this case, the same frame as the minimum angular velocity from the previous step). Label this as the TC (red triangle). If multiple estimated timings are obtained for a single gait event take the mean value.

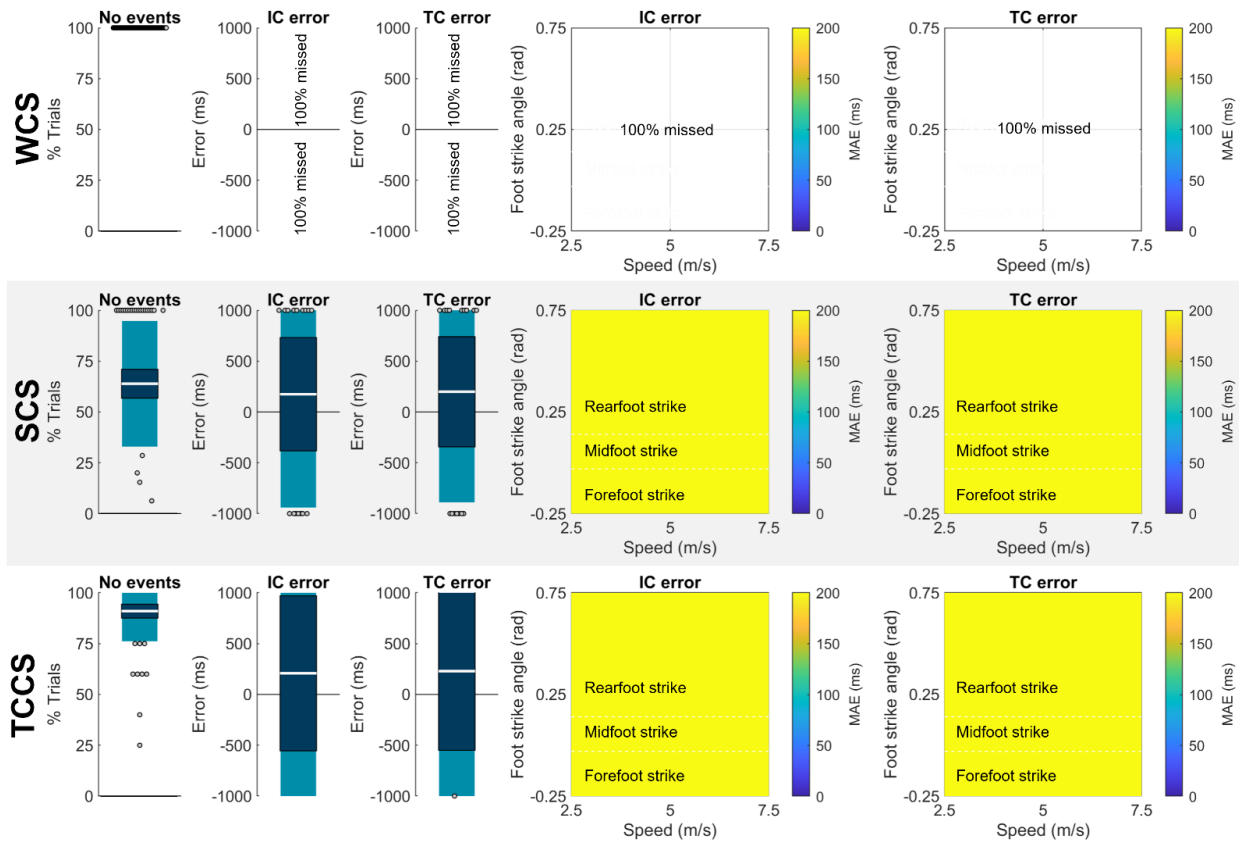
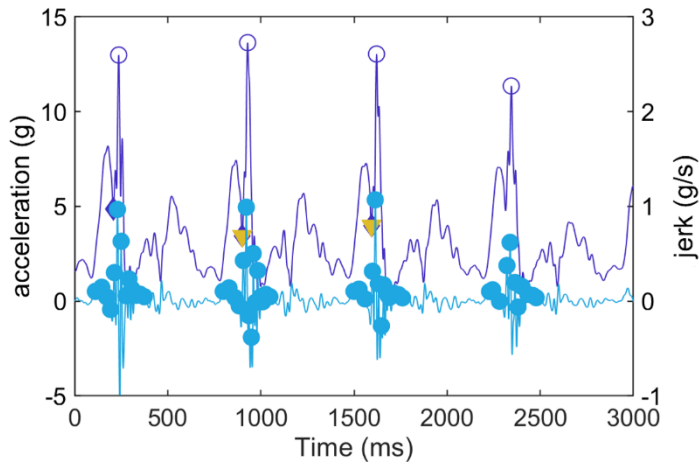


Figure 3S.21: Schmidt results. See Figure S2 for description.

### 3S.2.12. Aubol method.

Aubol et al. [159] used  $a_{WCS,resultant}$  from a triaxial accelerometer mounted on the anteromedial distal tibia to estimate IC. First, each axis' acceleration was filtered with a 70-Hz 4<sup>th</sup>-order low-pass Butterworth filter. Then features of resultant acceleration and jerk were used to identify IC. Results are shown in Figure 3S.23.



**Figure 3S.22: Aubol method.** Tri-axial accelerations were filtered using a 70-Hz 4<sup>th</sup> order low-pass Butterworth filter and used to calculate resultant acceleration (dark blue line) and jerk (light blue line). Maxima in the resultant acceleration separated by at least ~422 ms were identified (dark blue circles; the timing constraint was a novel constraint we introduced). Next, maxima in the resultant jerk were identified, however, only maxima within ~150 ms of a resultant acceleration maximum were retained (light blue circles). Acceleration minima with a prominence of at least 0.2\*the magnitude of the third largest resultant acceleration peak were found (dark blue diamonds). Any minimum with a resultant acceleration peak occurring in the preceding ~25 ms is eliminated. Looking between pairs of subsequent resultant acceleration maxima (dark blue circles), the earliest occurring minimum is labeled as the IC.

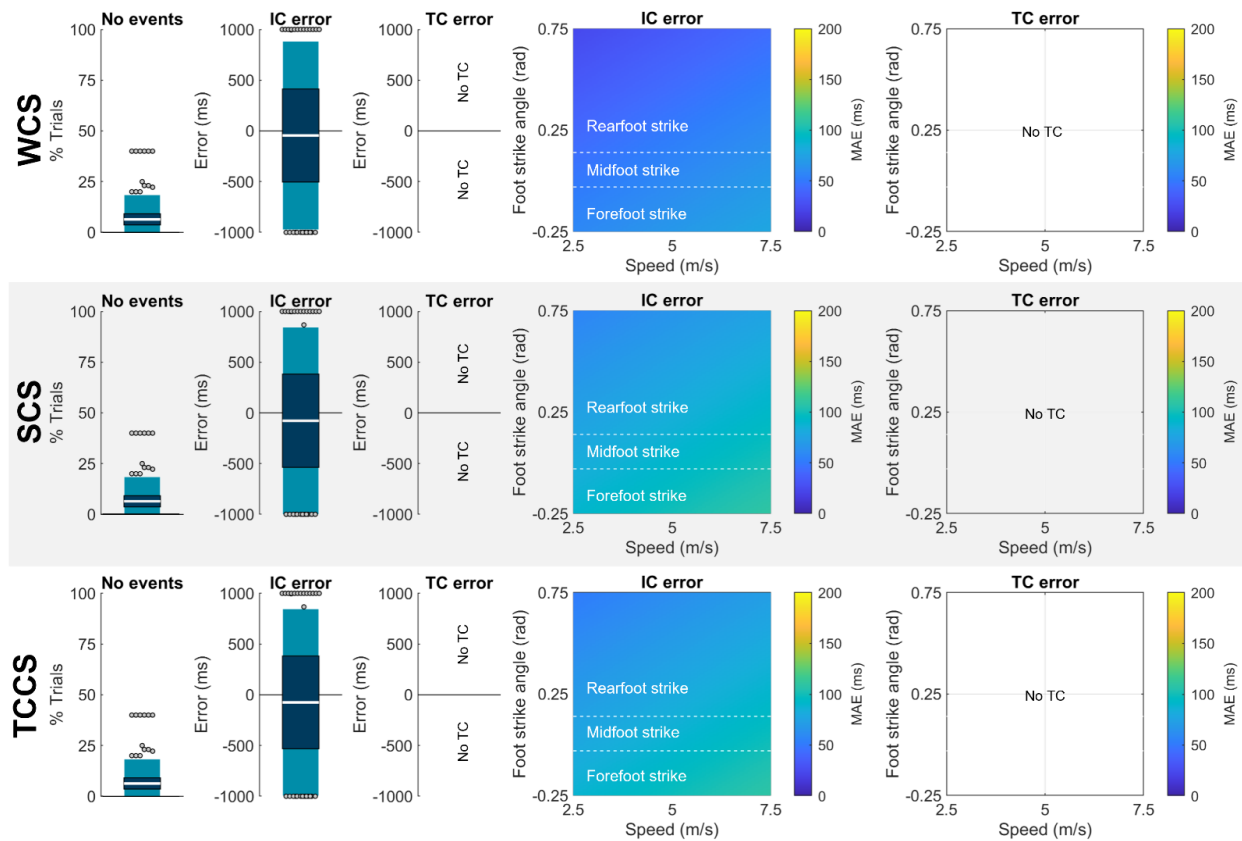
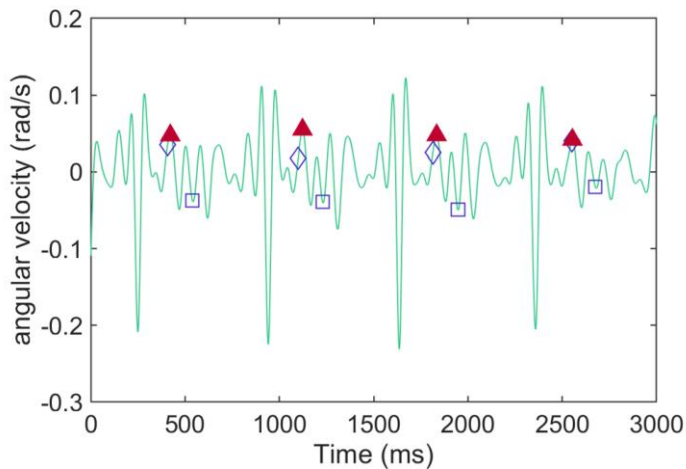
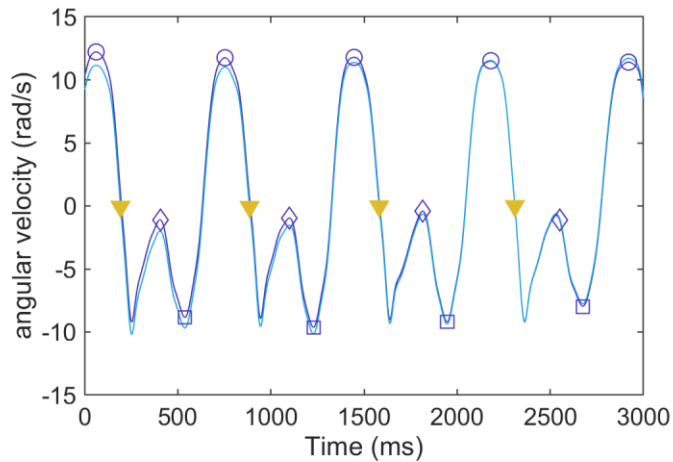


Figure 3S.23: Aubol results. See Figure 3S.2 for description.

### 3S.2.13. Fadillioglu method.

Fadillioglu et al. [160] aimed to create a method capable of identifying gait events during both walking and running. To implement this method, Fadillioglu et al. placed a uniaxial gyroscope on the lateral distal tibia to capture  $\omega_{WCS,Z}$ . The signal was low-pass filtered then midswing peaks were identified. A complementary signal was created by subtracting the filtered signal from the original signal and then features of this complementary signal were used to identify IC and TC. Results are shown in Figure 3S.25.



**Figure 3S.24: Fadillioglu method. (A)** Angular velocity about the x-axis (medial-lateral in the SCS and ~right in the WCS; dark blue line) is filtered with a 15 Hz 4<sup>th</sup> order low-pass Butterworth filter. Maxima with a magnitude of at least 4.63 rad/s and with a minimum separation of at least ~333 ms are identified as midswing peaks (dark blue circles). The signal is then detrended (light blue line) and for each midswing peak, the first zero-crossing following the peak is found and labelled as the IC (yellow triangles). Next, for each sequential pair of midswing peaks, the minimum value in the non-detrended signal (dark blue square) is found within a window from halfway between the peaks (dark blue diamond) to the second peak + 0.1\*the time between the peaks.

**(B)** Next, a complementary signal is created by subtracting the non-detrended filtered angular velocity from the unfiltered angular velocity. The complementary signal is low-pass filtered with a 2<sup>nd</sup> order 10-Hz Butterworth filter. A window of interest is defined based on the time between sequential midswing maxima: For times less than ~1000 ms (as in the example where the time is ~690 ms), TC (red triangle) is defined as the maximum complementary signal within a search window that starts halfway between midswing maxima (dark blue diamond) and ends at the negative peak (dark blue square). For times greater than or equal to ~1000 ms, TC is defined as the minimum complementary signal within a window starting halfway between midswing peaks and ending at the negative peak + 0.1\*the time between the peaks.

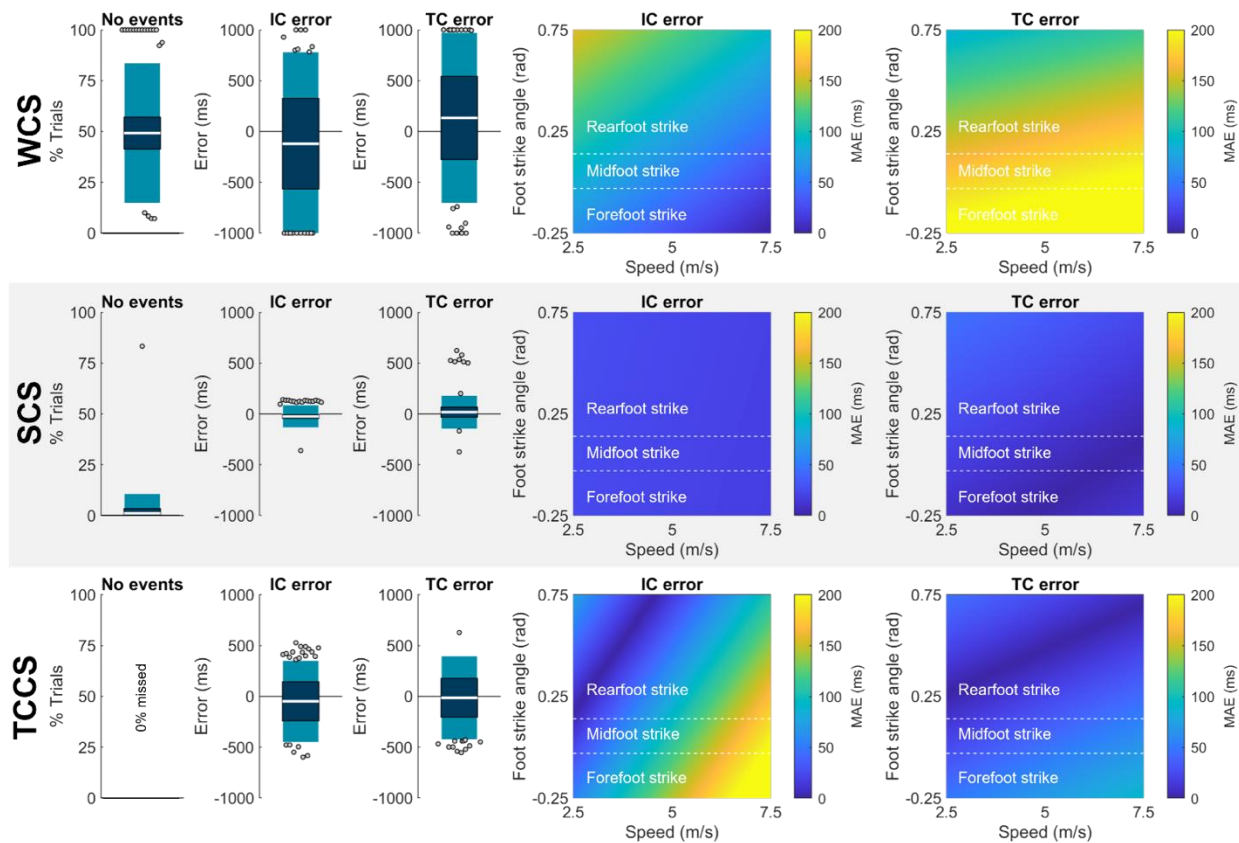
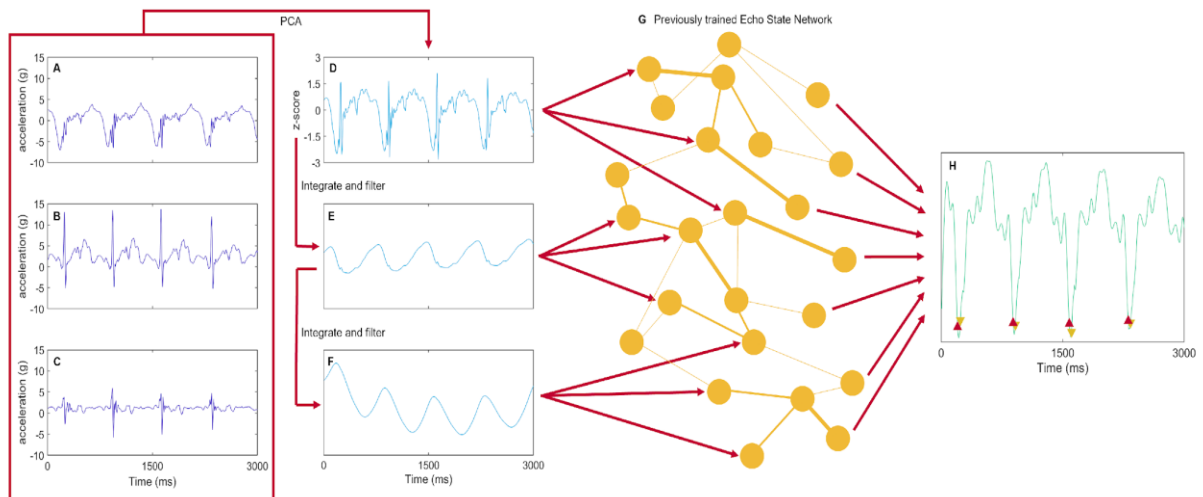


Figure 3S.25: Fadillioglu results. See Figure 3S.2 for description.

### 3S.2.14. Bach method.

Bach et al. [161] placed a tri-axial accelerometer on the anteromedial proximal tibia. They entered acceleration data into a principal component model and retained the principal component explaining the most variance for further analysis. The PC was normalized to its standard deviation and filtered and integrated to obtain velocity and position. Each signal was then entered into a machine learning algorithm and trained to estimate ground reaction forces. Estimated ground reaction forces were then used to identify IC and TC events. Results are shown in Figure 3S.27.





**Figure 3S.26: Bach method.** First, tri-axial acceleration data are resampled to match Bach et al.'s original sampling frequency (A-C) (the Echo State Network was trained on data collected at ~142.9 Hz and data collected at higher frequencies may have signal features that are novel to the network and degrade performance). Next, accelerations are entered into a Principal Component Analysis and the first component (explaining the most variance in the data) is retained (D). The principal component "acceleration" is then normalized to its standard deviation and integrated to find principal component "velocity." The principal component velocity is high-pass filtered with a 2nd order 1-Hz Butterworth filter (E) then integrated to obtain principal component "position." The principal component position is then filtered using the same parameters (F). Principal component acceleration, velocity, and position signals are entered into an Echo State Network that has been previously trained using Bach et al.'s published data (G) yielding estimated ground reaction forces (H) that are used to define IC and TC events (yellow and red triangles). (In this example, it appears the Echo State Network has almost captured the shape and location of four distinct stances in the ground reaction force signal, however, these distinct stances are separated by noise and large troughs that interfere with the method's ability to accurately identify IC and TC events).

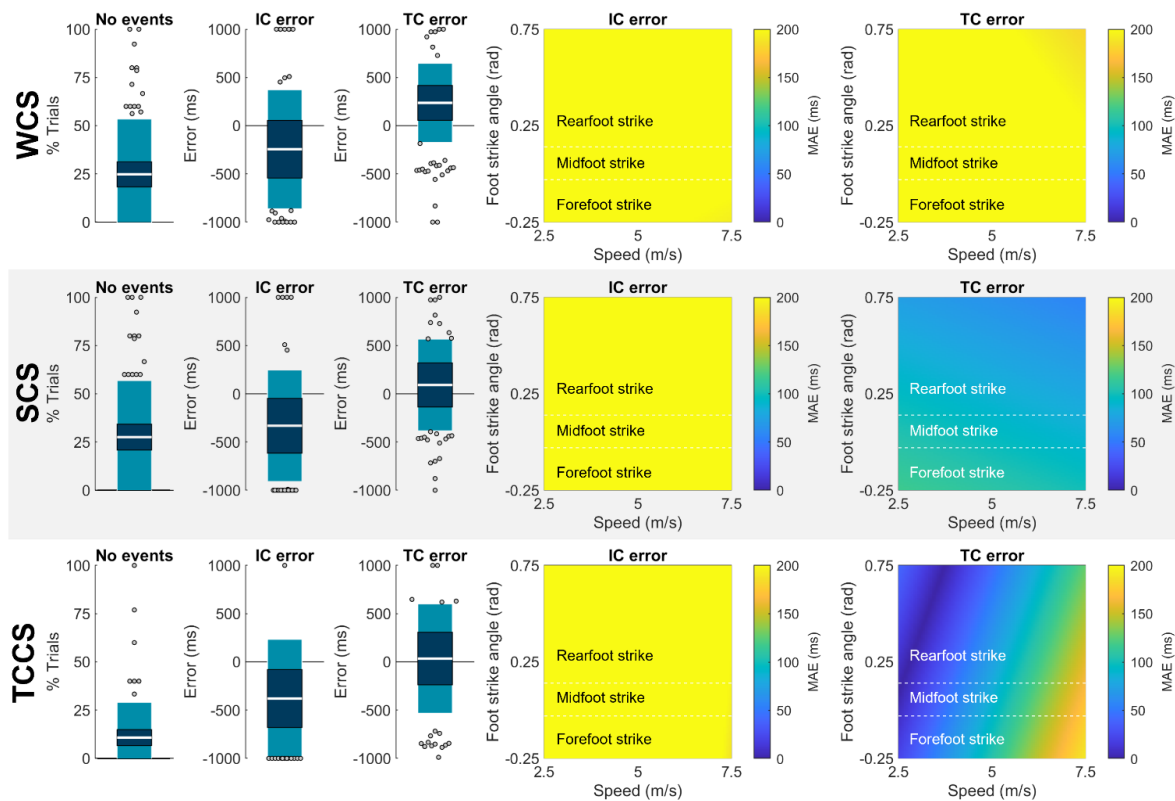


Figure 3S.27: Bach results. See Figure 3S.2 for description.

### 3S.2.15. Bach modified method

We also provide a modified version of the Bach et al. method [161]. This version addresses the large amount of negative noise we observed between the vGRF waveforms by zeroing out any vGRF values below zero then multiplying the vGRF by negative one before feeding it into Bach et al.'s gait event estimation algorithm (to cancel out an unexplained negative one multiplication that method performs). Results are shown in Figure 3S.28.

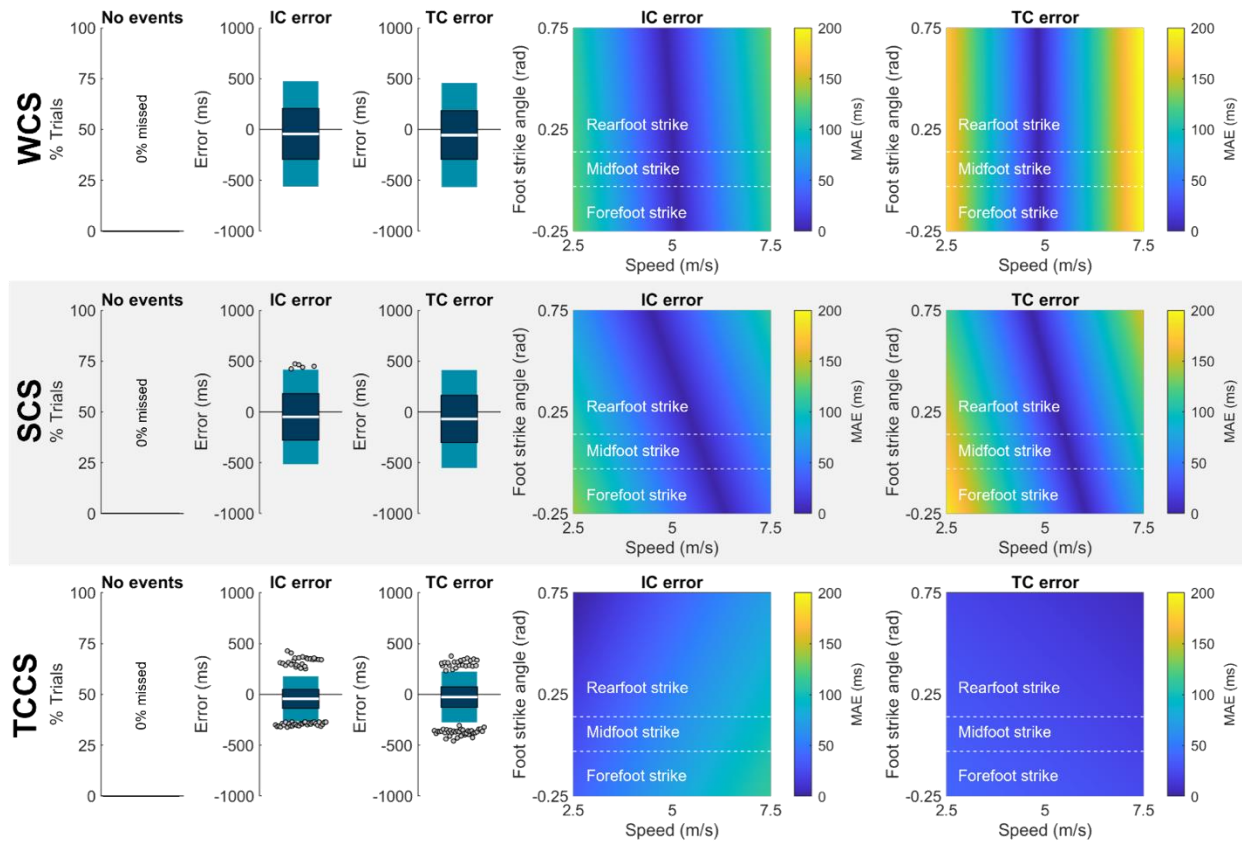
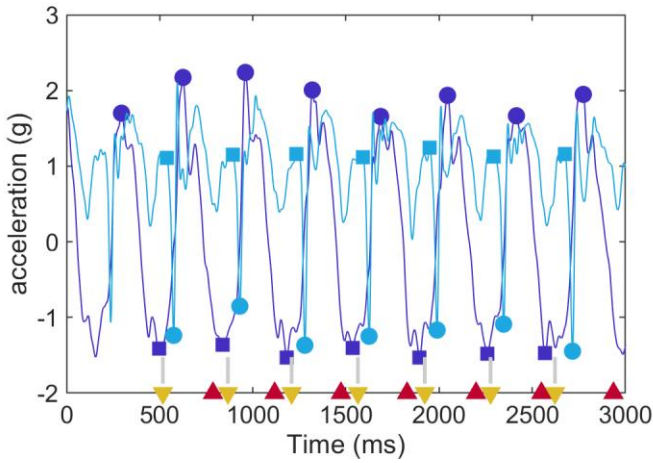


Figure 3S.28: Bach results. See Figure 3S.2 for description.

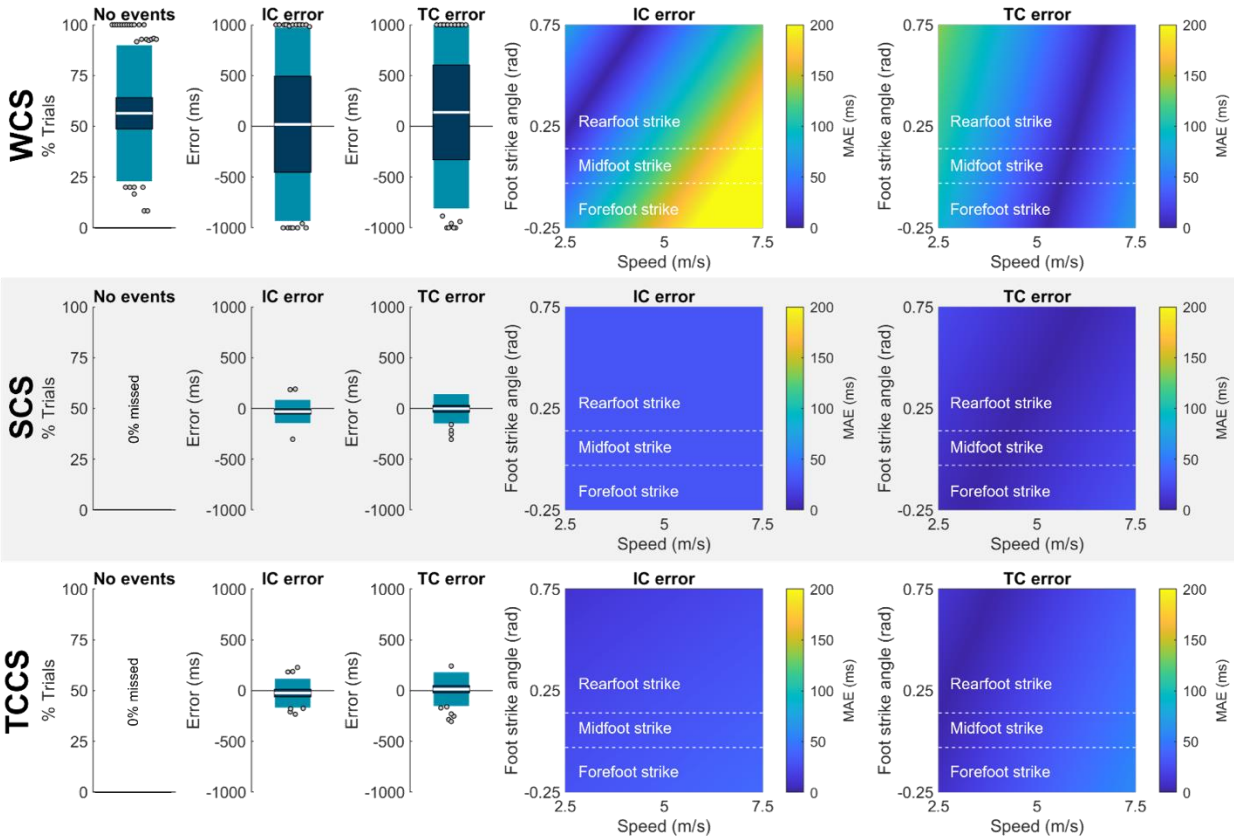
### 3S.3. Lower back-mounted wearable methods

#### 3S.3.1. Auvinet method.

Auvinet et al. [94] identified specific features associated with gait events in the signal of a tri-axial accelerometer placed on the mid-sagittal lumbar spine. Based on their simultaneous recording of acceleration and video data they identified IC as the start of a large peak in  $a_{WCS,Y}$  co-occurring with the start of a deep minima in the  $a_{WCS,X}$ . TC was identified as the end of the peak in  $a_{WCS,Y}$ . Right and left side were also identifiable based on  $a_{WCS,Z}$ . We developed a novel automated code to identify gait events based on these features. Results are shown in Figure 3S.30.



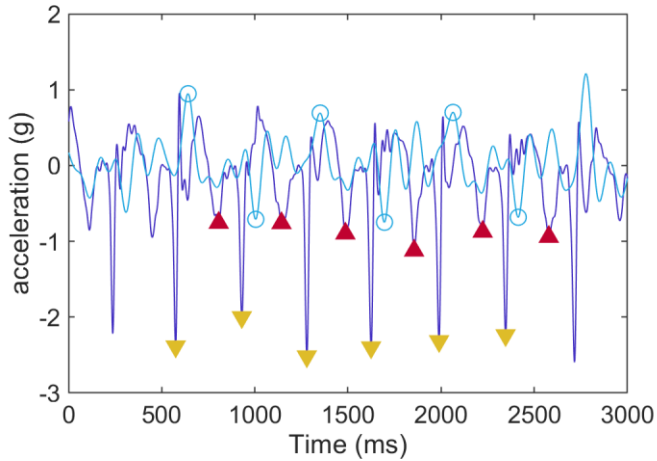
**Figure S29: Auvinet method.** Identify y-axis (proximal-distal in the SCS, ~longitudinal in the WCS; dark blue line) maxima separated by at least ~422 ms (dark blue circles). Then look for the minimum value between successive maxima (dark blue squares). Next, in the x-axis acceleration (anterior-posterior in the SCS, ~direction of progression in the WCS; light blue line) find minima (light blue circles) occurring between the y-axis minimum and its preceding maximum. Then, walk back until values start decreasing (light blue squares). Define IC (yellow triangles) as the average frame between the points where y-axis minimum and where x-axis acceleration started decreasing. To identify TC (red triangles), look for the first frame following each y-axis maximum where y-axis acceleration magnitude falls below -1 g. Finally, to determine side, take the mean value of the z-axis acceleration (medial-lateral in ~SCS, ~right in WCS) in a window extending ~10 ms to each side of the IC (gray bars). If the magnitude is less than 0, call it a left stance. If the magnitude is more than 0, call it a right stance.



**Figure 3S.30: Auvinet results.** See Figure 3S.2 for description.

### 3S.3.2. Lee method.

Lee et al. [88] placed an IMU with tri-axial accelerometer on the sacrum. They identified IC and TC as maxima in  $a_{WCS,X}$  and used  $a_{WCS,Z}$  to identify side. Results are shown in Figure 3S.32.



**Figure 3S.31: Lee method.** To identify IC, find minima in the x-axis acceleration (anterior-posterior in SCS, ~direction of progression in WCS) separated by at least ~211 ms. Label this point the IC (gold triangles). Look from one IC plus ~85 ms to the preceding IC and find a secondary x-axis minima. Label this point the TC (red triangles). To identify stance side, low-pass filter z-axis accelerations (medial-lateral in SCS, ~right in WCS) with a 10-Hz 4<sup>th</sup> order Butterworth filter. Then, looking between successive IC and TC events, find positive and negative peaks. If the absolute value of the negative peak is greater label the stance as left. If the positive peak is greater label the stance as a right.

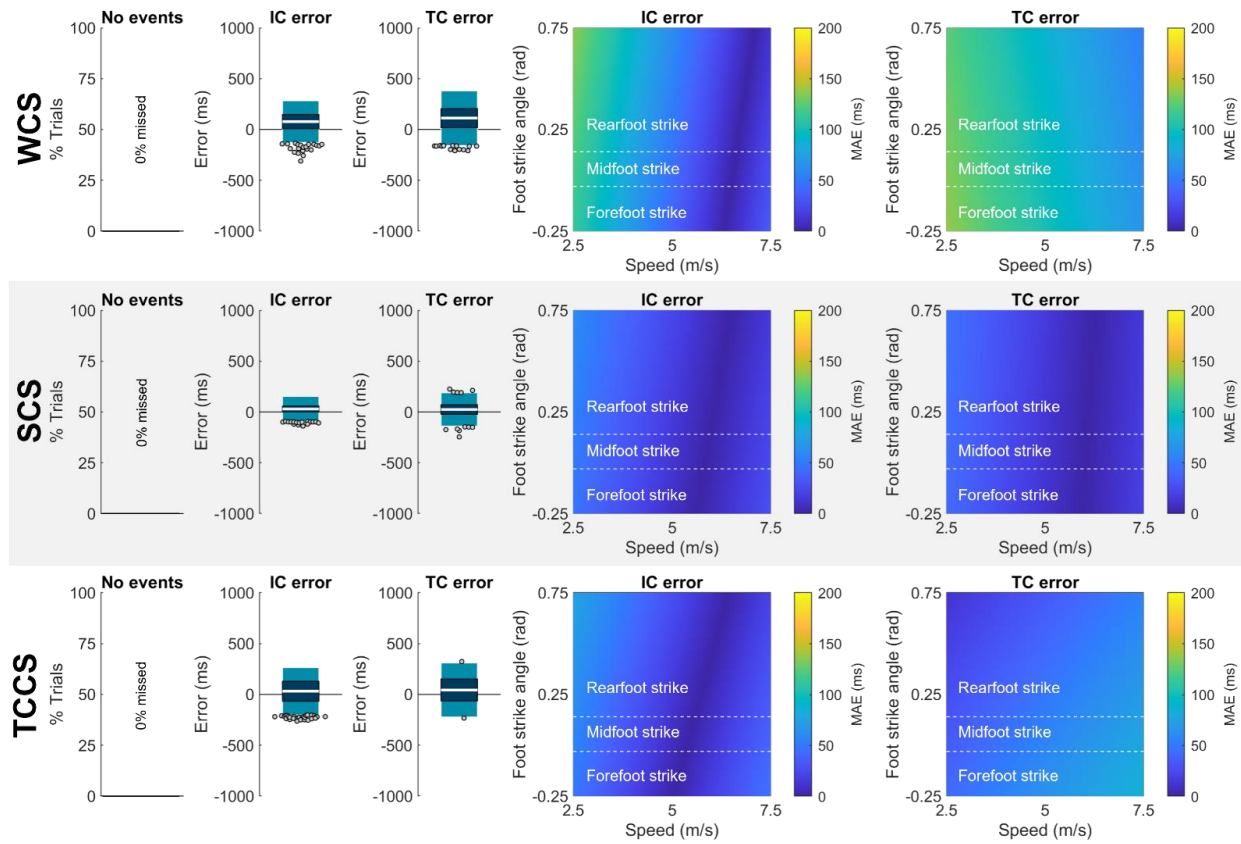
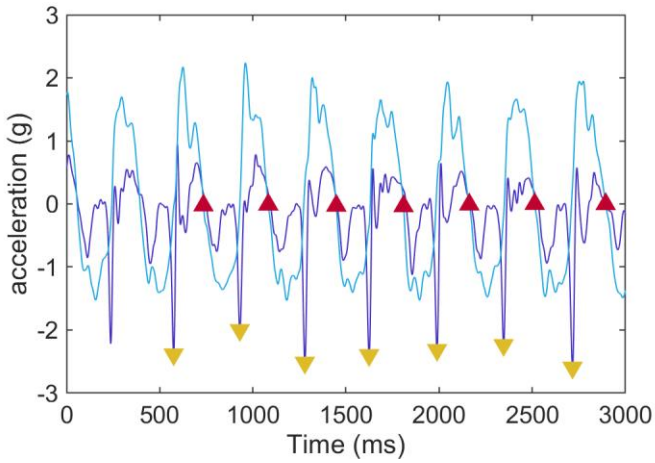


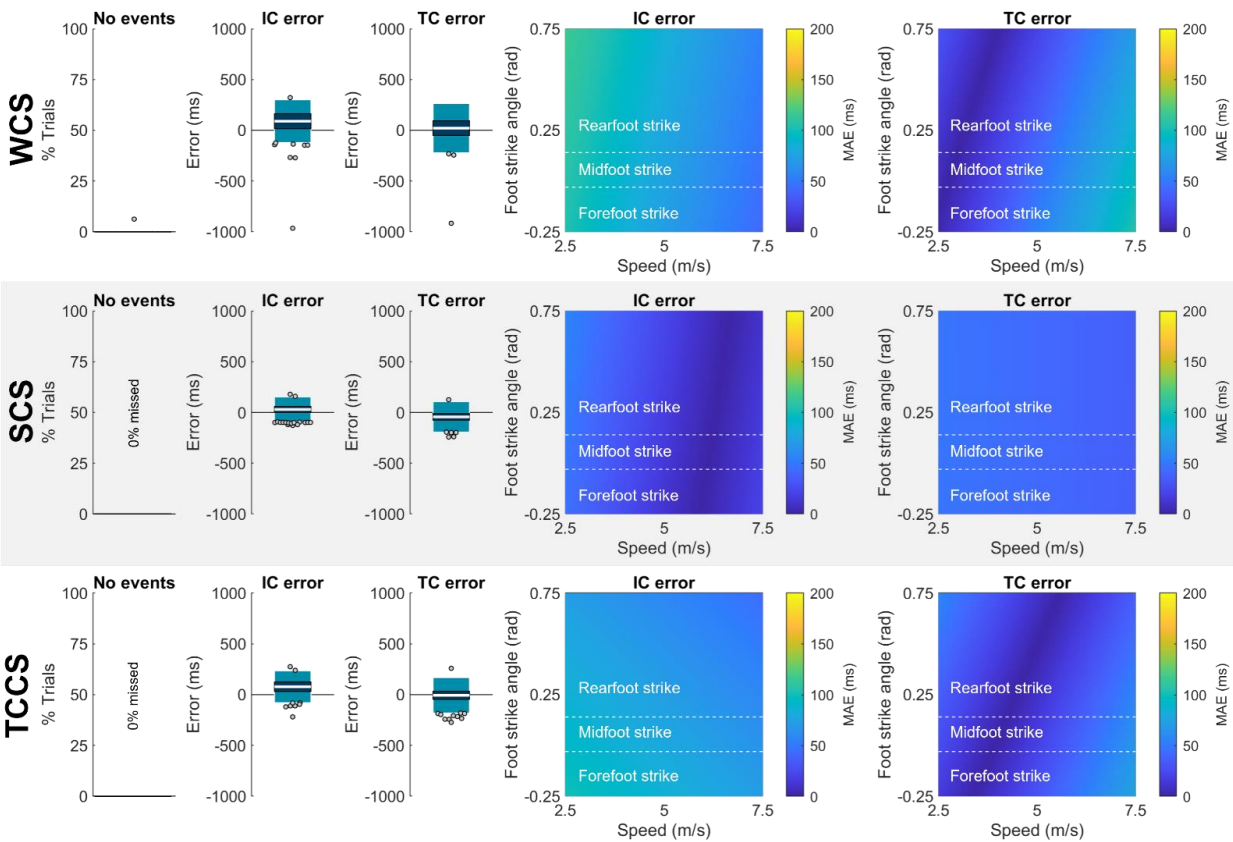
Figure 3S32: Lee results. See Figure 3S.2 for description.

### 3S.3.3. Wixted method.

To characterize acceleration signals related to gait events, Wixted et al. [95] simultaneously recorded accelerations from L3-L4 and pressures from in-shoe insoles. They observed that IC occurred around the time of a negative peak in  $a_{WCS,X}$  (similar to Lee et al. [88]) while TC occurred around the time of a negative zero-crossing in  $a_{WCS,Y}$  (similar to Auvinet et al. [94]). Results are shown in Figure 3S.34.



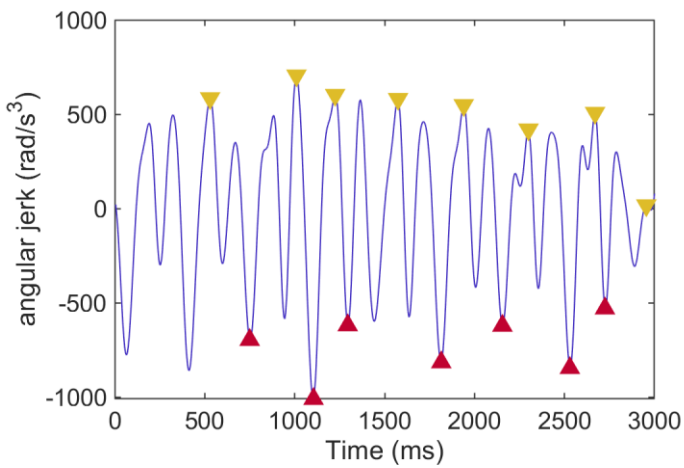
**Figure 3S.33: Wixted method.** To identify IC, find negative peaks in the x-axis acceleration (anterior-posterior in SCS, ~direction of progression in WCS; dark blue line) separated by at least ~211 ms. Label these points the ICs (gold triangles). Starting ~85 ms after each IC, find the first y-axis acceleration value (proximal-distal in the SCS, ~vertical in the WCS; light blue line) that goes below 0. Label this point the TC (red triangles).



**Figure 3S.34: Wixted results.** See Figure 3S.2 for description.

### 3S.3.4. Bergamini method.

Bergamini et al. [164] placed a tri-axial IMU on the lumbar spine and used Luo et al.'s wavelet-mediated differentiation technique [189] to double differentiate  $\omega_{WCS,resultant}$ . Positive peaks in the double differentiated signal are used to define IC while minima are used to define TC. Results are shown in Figure 3S.36.



**Figure 3S.35: Bergamini method.** Angular velocity is resampled to 200 Hz, matching Bergamini et al.'s original paper. Then resultant angular velocity is calculated and double differentiated using 4-level quadratic spline discrete wavelet differentiation as described by Luo et al. [32] (dark blue line). Positive peaks separated by at least  $\sim 211$  ms are found in this new angular jerk signal and are labelled as IC (yellow triangles). Minima between each successive pair of positive peaks are found and labelled as TCs (red triangles).



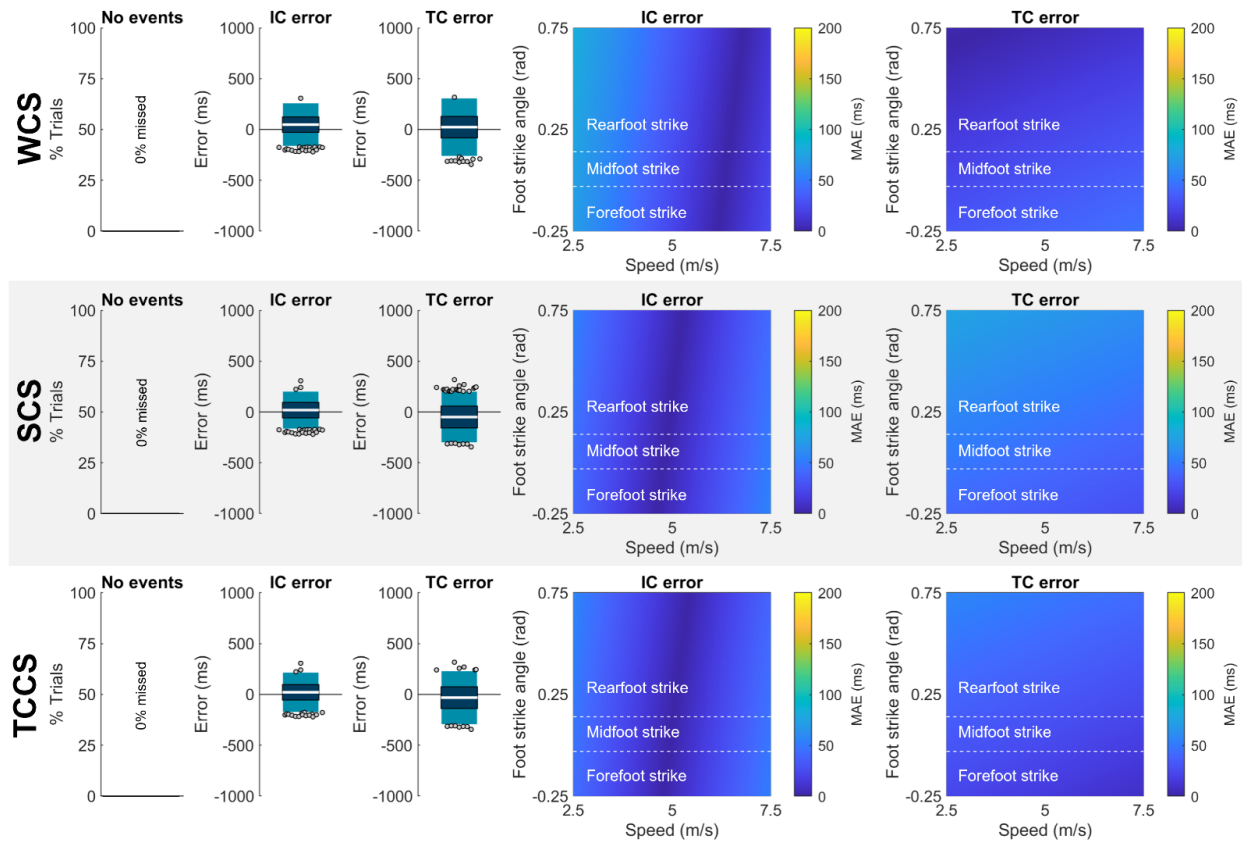
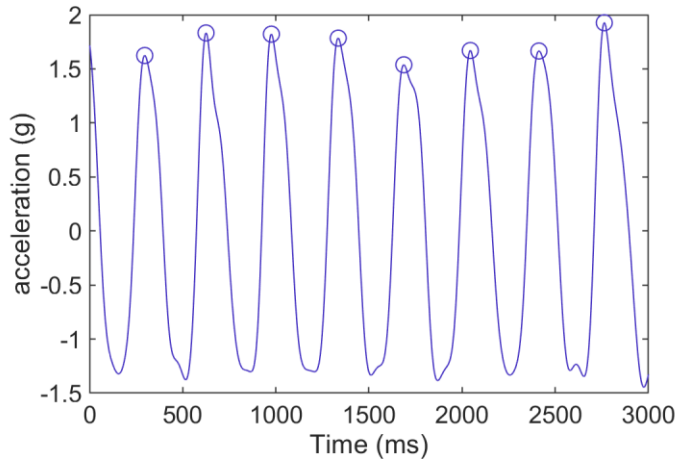


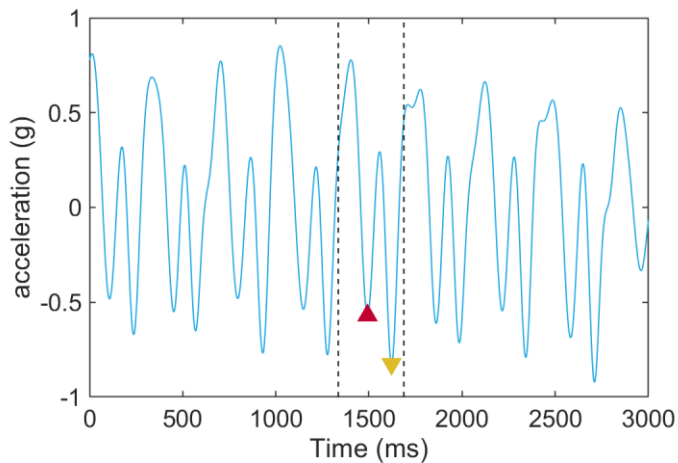
Figure 3S.36: Bergamini results. See Figure 3S.2 for description.

### 3S.3.5. Benson method.

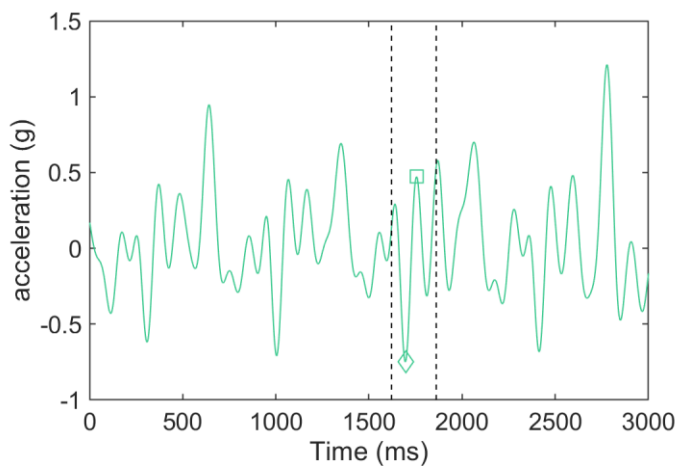
Benson et al. [165] provide several well-documented functions to identify gait events from acceleration profiles of wearables mounted on the foot or sacrum. Here, we've minimally adapted their sacrum function to work with the same inputs and provide same outputs as other methods in this package. Results are shown in Figure 3S.38.



**Figure 3S.37: Benson method. (A)** First, accelerations are low-pass filtered with a 10-Hz 4<sup>th</sup> order Butterworth filter. Then positive peaks separated by at least ~250 ms are found in the y-axis acceleration (proximal-distal in the SCS, ~vertical in the WCS; dark blue line).



**(B)** Next, find negative x-axis (anterior-posterior in the SCS, ~direction of progression in the WCS; light blue line) acceleration peaks between each pair of successive y-axis peaks (vertical black lines). If there are multiple negative x-axis peaks, then rank them 1...n by both magnitude (largest to smallest) and by timing (latest to earliest). IC (yellow triangle) is defined as the peak with the lowest mean magnitude and timing ranking. If there is a tie in mean rankings, the later peak is accepted as IC. Then, look from the preceding y-axis peak to IC - ~100 ms and accept the latest occurring negative x-axis peak as the TC (red triangle). Note, if no negative x-axis peaks can be identified, find the x-axis minimum and label it as the IC. Then find the x-axis maximum between the preceding y-axis peak and the IC. Find the peak x-axis jerk between the x-axis maximum and the IC.



**(C)** To identify stance side, find the largest positive (green square) and negative (green diamond) peaks in the z-axis acceleration (medial-lateral in the SCS, ~right in the WCS; green line) during each stance (vertical black lines). If there is no positive (rightward) peak, then set as a left stance (and vice versa). If both positive and negative peaks exist, if the positive peak is closer to TC than to the negative peak, set as a left step; otherwise, if the negative peak occurs within ~15 ms of the IC, find a new negative peak between the current negative peak and the TC; otherwise, set as a left stance.

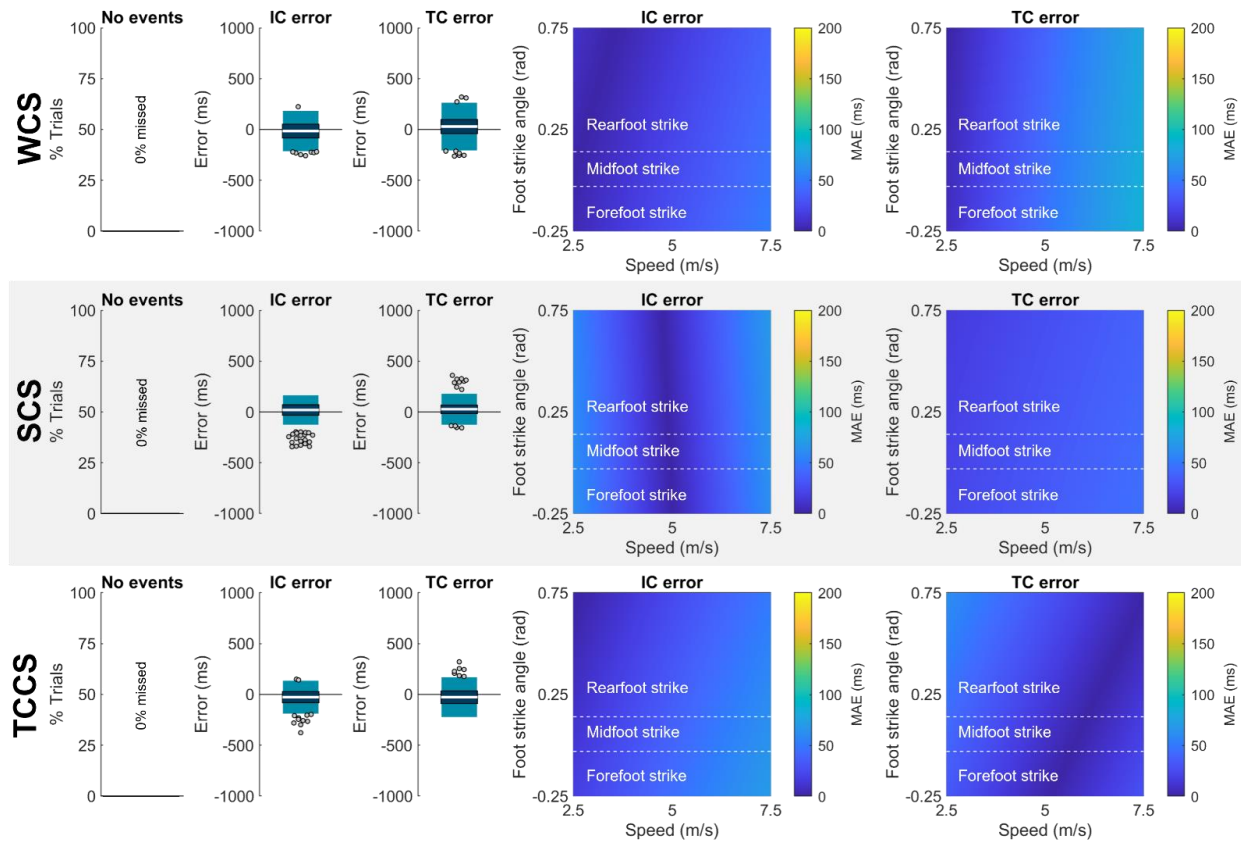
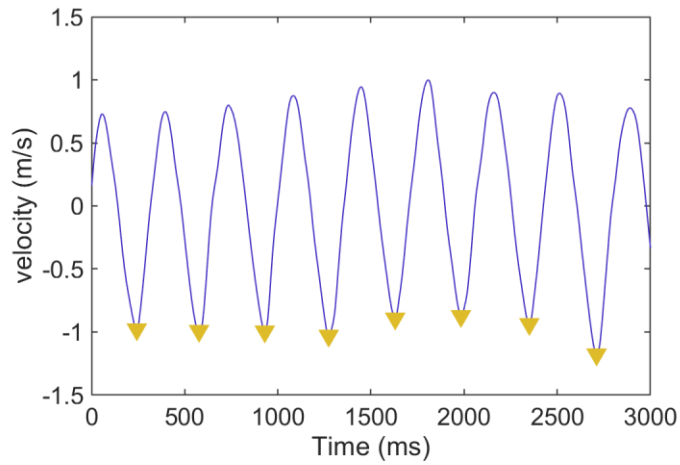


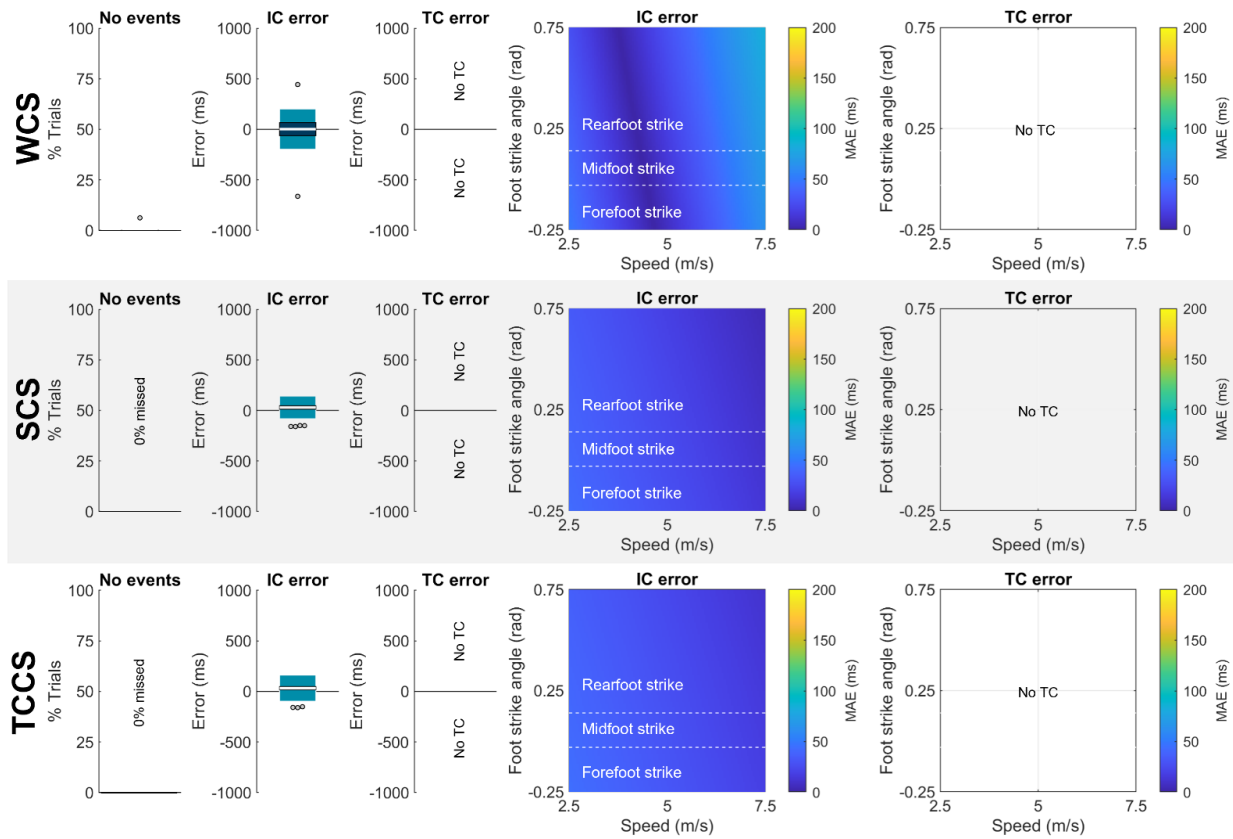
Figure 3S.38: Benson results. See Figure 3S.2 for description.

### 3S.3.6. Reenalda method.

Reenalda et al. [166] placed an IMU on runners while they ran a treadmill using different foot strike patterns. They defined IC as the peak downward velocity of the pelvis by integrating  $a_{GCS,Y}$  (obtained using proprietary algorithms provided by the IMU manufacturer). Results are shown in Figure 3S.40.



**Figure 3S.39: Reenalda method.** *y*-axis acceleration (vertical in the GCS, proximal-distal in the SCS, and ~vertical in the WCS) is integrated to obtain *y*-axis velocity. Negative peaks are found and labelled as IC. We added an additional constraint that these peaks must be separated by at least ~211 ms. No method is described to determine TC or stance side.



**Figure 3S.40: Reenalda results.** See Figure 3S.2 for description.

### 3S.4. IMU data processing

An overview of our IMU data processing is provided in the main text. Here, we expand on those processing details. After collecting raw data from each IMU, data were downloaded and processed offline using the following steps: (1) Calibration, (2) Quiet period identification, (3) Bias removal, (4) Saturation correction, (5) Low-pass filtering, (6) Drift correction, (7) Orientation estimation and gravity subtraction, and (8) Coordinate system transformation. After these processing steps each of the 642 trials we analyzed was segmented using the speed gate signals (to include data between the times when the participant crossed the gate 2.5 m before force plate center and the gate 2.5 m after force plate center) and entered into each of the 21 individual methods as an  $n \times 7$  matrix (with  $n$  rows representing the number of frames and the 7 columns representing time stamps, 3 acceleration axes, and 3 angular velocity axes). This resulted in a total of  $642 \text{ trials} \times 21 \text{ methods} = 13482$  possible events for comparison to the corresponding 642 ground truth force plate events for each IC, TC, and stance side.

#### 3S.4.1. Calibration

All IMU data collected during this experiment were corrected with IMU-specific calibration matrices. These matrices were calculated by conducting a calibration procedure that ensured each IMU accurately expressed accelerations and angular velocities in an orthogonal coordinate system oriented square to the IMU housing.

Each IMU was secured to a centrifuge (ClearPath MCVC, Teknic, Victor, USA) with custom 3-D printed jigs (SOLIDWORKS 2019, Dassault Systèmes, Vélizy-Villacoublay, France) and calibrated in 6 orientations at 16 known accelerations (from 0 - 41.42 g where  $1 \text{ g} = 9.8 \text{ m/s}^2$ ) and angular velocities (from 0 - 78.54 rad/s). Adapting methods from Coolbaugh et al. [99], known data (K) from the centrifuge and measured data (M) from the IMU were used to calculate  $3 \times 7$  calibration matrices for each IMU (C;

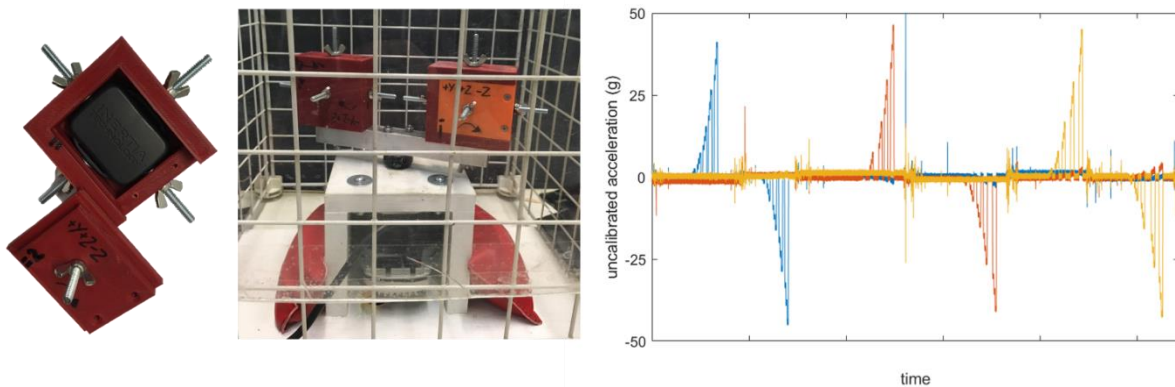
3 signed magnitude terms, 3 absolute magnitude terms, and one bias term per axis) and quantify sensor accuracy with a hold-back procedure after subtracting out biases observed during a quiet period (B).

Equation 3S.7

$$C*(M + B) = K$$

One potential limitation of this procedure is that it treats each triaxial sensor independently (primary accelerometer, secondary accelerometer, and gyroscope) and assumes their values do not affect each other. This assumption was tested while piloting this calibration procedure by quantifying inter-sensor dependencies between the primary accelerometer and gyroscope and between the secondary accelerometer and gyroscope. Observed dependencies were negligible and independent sensor calibration matrices yielded the best results; thus, we felt confident using this approach (which avoids the indeterminacy of the primary and secondary accelerometer having the same K values).

After calibration, IMU primary accelerometer errors were  $\leq 0.01 \pm 0.04$  g, secondary accelerometer errors were  $\leq 0.05 \pm 0.07$  g, and gyroscope errors were  $\leq 0.01 \pm 0.01$  rad/s.



**Figure S3.41: IMU calibration.** Left: An IMU in its 3-D printed housing. Computer-aided design software was used to ensure IMUs were friction fit square to their housing. Middle: Two IMUs in their 3-D printed housings mounted on the centrifuge. IMUs were checked for square with an engineer's square and level with a bullseye level. Right: Example of measured triaxial accelerations for the secondary accelerometer (M). The 16 accelerations being applied to the IMU in each of 6 orientations correspond to known (K) values from the centrifuge. Accelerations between each orientation correspond to the IMU being repositioned on the centrifuge and checked for square and level.

### 3S.4.2. Quiet period identification

Quiet periods were identified throughout data collection (e.g., participant resting, participant preparing at the start of the runway, participant standing while receiving instruction) and used to periodically check for changes in bias (as bias can vary with battery life and temperature) and reset orientation algorithms (as orientation estimates are prone to drift over prolonged periods; discussed further below). These quiet periods were defined as any period where...

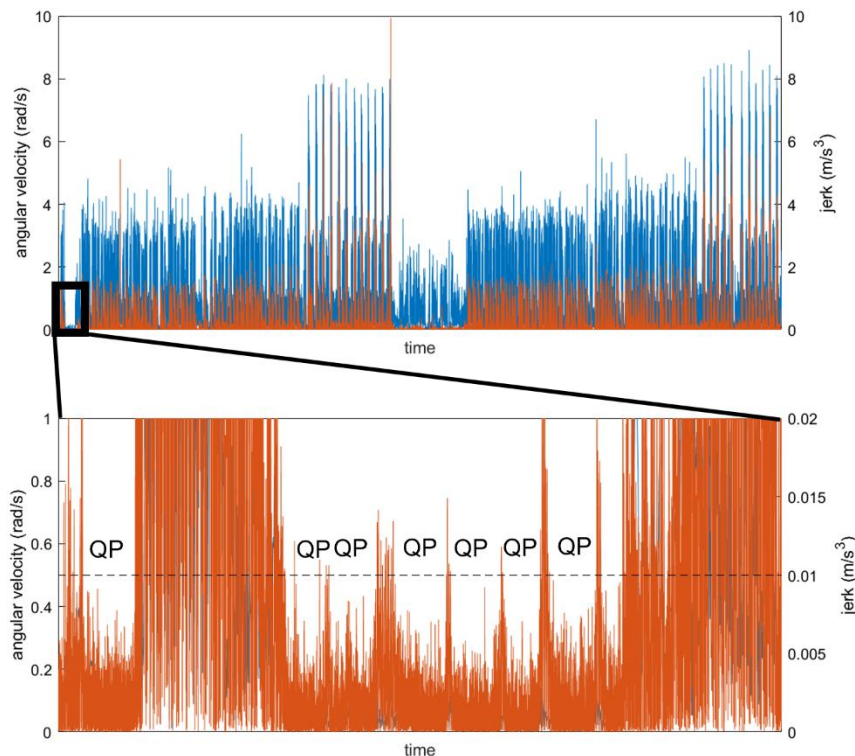
Equation 3S.8

$$\omega_{resultant} < 0.5 \text{ rad/s and}$$

Equation 3S.9

$$j_{resultant} < 0.01 \text{ m/s}^3$$

...for at least 100 ms.



**Figure 3S.42: Quiet period identification.** Top: Resultant angular velocity (blue) and jerk (orange) at the sacrum across an entire data collection for the randomly selected example participant. Bottom: Zoomed in to show quiet periods (QP) where resultant angular velocities are  $< 0.5 \text{ rad/s}$  and resultant jerks are  $< 0.01 \text{ m/s}^3$  for at least 100 ms. Thresholds noted with the dashed horizontal line.

### 3S.4.3. Bias removal

When the IMU is quiet, we know that it is not accelerating or rotating and thus, the only thing loading the axes should be the gravity vector. Based on this knowledge we can create a temporary inertial coordinate system based on gravity:

**Equation 3S.10**

$$Y = \frac{\sum_{\text{first quiet frame}}^{\text{last quiet frame}} a_{\text{calibrated}}}{n_{\text{quiet frames}}}$$

We can express Y as a unit vector, then make X and Z orthogonal unit vectors (with arbitrary sense). Using these vectors, we can create a temporary rotation matrix that will align our data with gravity:

**Equation 3S.11**

$$R_{\text{temp}} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

We can then express our data in this temporary inertial coordinate system (and given our calibration we know that the axes of the accelerometer and gyroscope are exactly aligned so the same rotation matrix can be used for both):

**Equation 3S.12**

$$a_{\text{temp}} = R_{\text{temp}} * a_{\text{calibrated}}$$

Given the IMU is quiet, in every frame  $a_{\text{temp}}$  and  $\omega_{\text{temp}}$  should now equal  $[0 \ 1 \ 0]$  g and  $[0 \ 0 \ 0]$  rad/s, respectively. Thus, we can calculate bias (B) in acceleration and angular velocity as the average deviation from those values across the quiet period:

**Equation 3S.13**

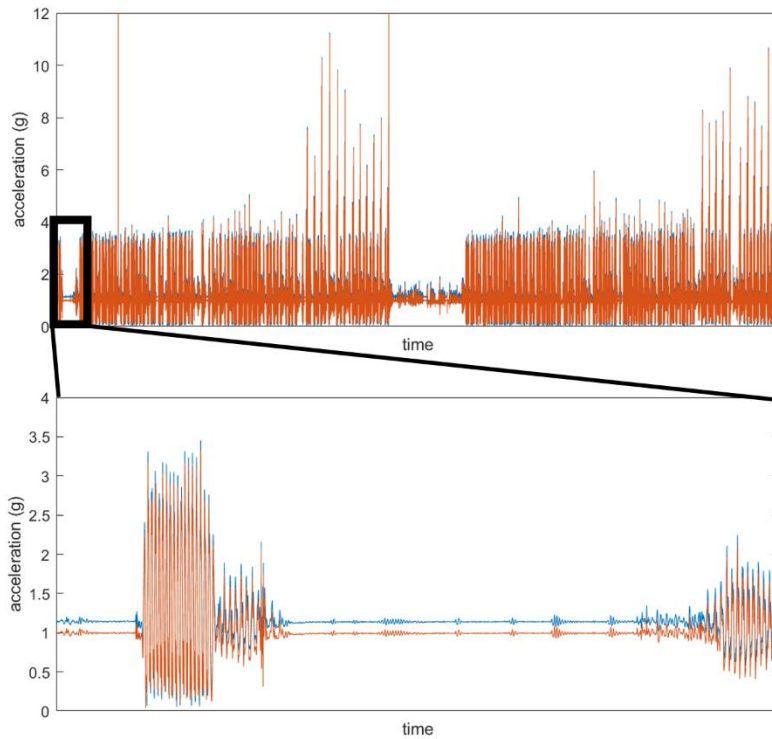
$$B_a = \frac{\sum_{\text{first quiet frame}}^{\text{last quiet frame}} (a_{\text{temp}} - [0 \ 1 \ 0])}{n_{\text{quiet frames}}}$$

We can then remove bias and re-express our data in its original coordinates:



Equation 3S.14

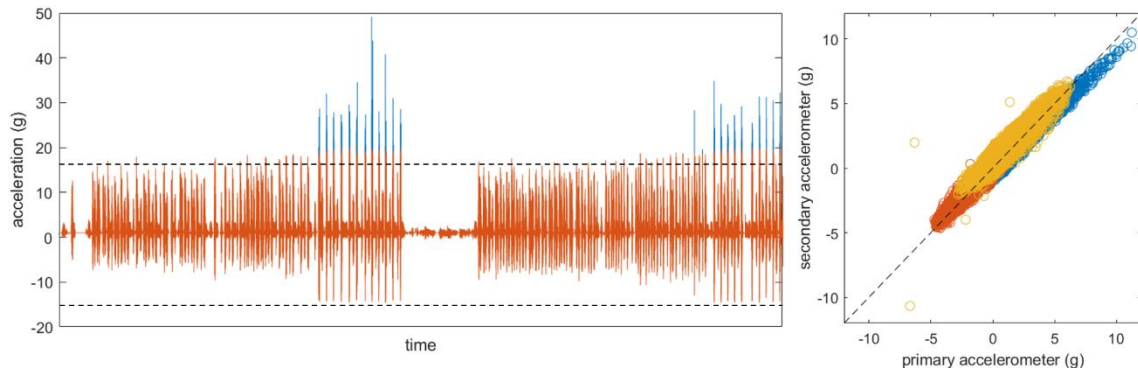
$$a_{debias} = a_{calibrated} - R_{temp}^{-1}B_a$$



**Figure 3S.43: Bias removal.** Top: Uncorrected resultant sacral acceleration (blue) and de-biased resultant sacral acceleration (orange) across the entire data collection for the randomly selected participant. Bottom: Zoomed in to show that uncorrected resultant acceleration does not equal 1 g during quiet periods while de-biased acceleration equals exactly 1 g.

#### 3S.4.4. Saturation correction

Our IMU contained two tri-axial accelerometers with different ranges. The primary accelerometer had a range of 16 g while the secondary accelerometer had a range of 100 g. Although 16 g is a large enough range to capture the majority of accelerations at the tibia and sacrum during running, we wanted to ensure that saturation did not occur, particularly at the tibia [190]. Thus, we used a threshold of 15.5 g and replaced any value above this threshold in our primary accelerometer with the corresponding frame from our secondary accelerometer (these values were highly correlated across the  $\pm 16$  g range they could both measure). Secondary accelerometer data were then discarded.

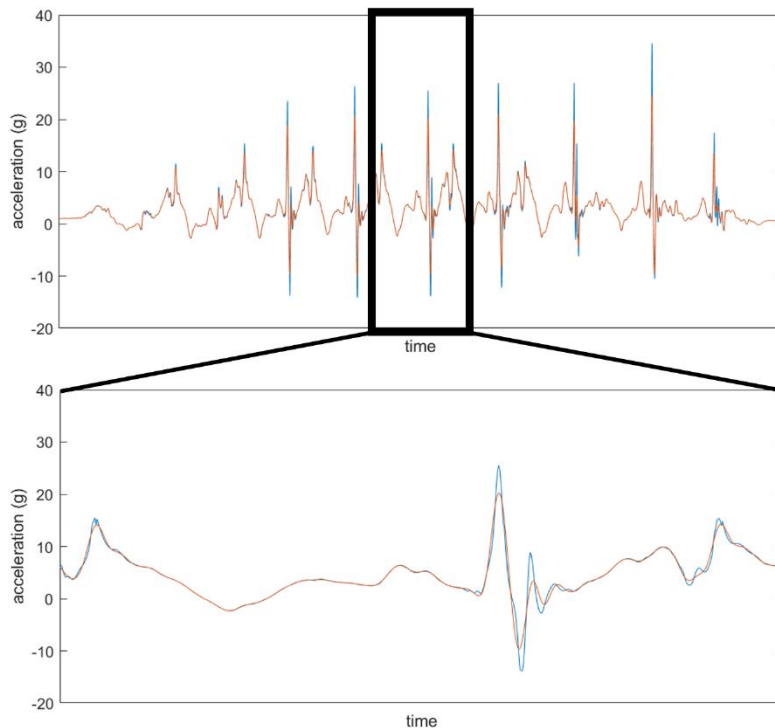


**Figure 3S.44: Saturation correction.** Left: Tibia longitudinal axis accelerations from primary accelerometer (orange) and secondary accelerometer (blue) across the entire data collection for the example participant. The horizontal black line indicates data outside the primary accelerometer's range (defined as  $|a| > 15.5$  g). Right: Secondary accelerometer measurements plotted against primary accelerometer measurements for each axis (different colors). Black dashed diagonal line indicates perfect agreement. Here correlations between each axis ranged from  $r = 0.82$  to  $r = 0.96$ . In general, correlations across the ranges shared between primary and secondary accelerometers were  $\geq 0.90$ .

### 3S.4.5. Low pass filtering

Next, accelerations and angular velocities were filtered with a 4th order 50 Hz low-pass

Butterworth filter.



**Figure 3S.45: Low-pass filtering.** Top: Worst-case example of unfiltered (blue) and filtered (orange) longitudinal tibial acceleration (that experienced saturation and is now composed of data from the primary and secondary accelerometers). Data is taken from a single trial from the example participant. Bottom: Zoomed in on one step to better visualize differences between filtered and unfiltered signal. Filter parameters were chosen to qualitatively balance the preservation of major signal features (particularly peak magnitudes and locations) with the removal of high-frequency noise.

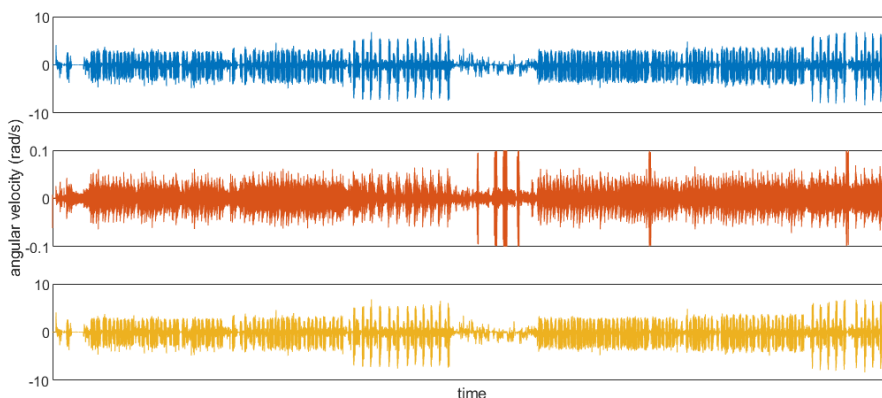
### 3S.4.6. Drift correction

Angular velocity measured by IMUs is prone to drift. This drift makes it difficult to integrate angular velocities and calculate the orientation of an IMU in space. Several sensor fusion algorithms have been developed to correct this drift including Kalman filters [191], Mahoney filters [192], and Madgwick filters [193]. We explored the use of each of these filters and found that converting our data to quaternion representation and entering it into a Madgwick filter (with beta set to 0.05 and no magnetometer fusion due to the amount of magnetic interference in our lab) was the most successful in eliminating drift in a “worst case” recreation of our experimental conditions (an 80 minute data capture with extreme angular rotations and accelerations and no quiet period corrections yielded 1.66 rad rotation error). The code we used to execute the Madgwick filter is available from x-io at:

<https://x-io.co.uk/open-source-imu-and-ahrs-algorithms/>

And from MATLAB at:

<https://www.mathworks.com/products/sensor-fusion-and-tracking.html>

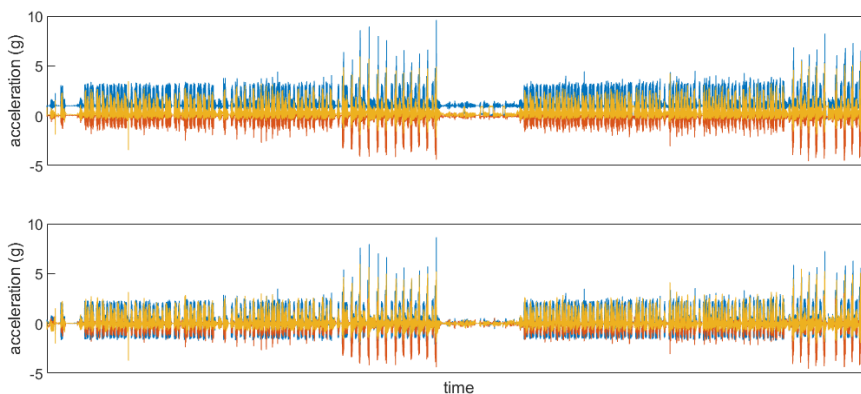


**Figure 3S.46: Drift correction.** Top: Uncorrected angular velocity about the longitudinal axis of the sacrum (blue) for the entire data collection of the example participant. Middle: Difference between uncorrected and Madgwick filter-estimated angular velocities (orange). Bottom: Madgwick filter-estimated angular velocity about the longitudinal axis of the sacrum (yellow).

### 3S.4.7. Orientation estimation and gravity subtraction

After drift-correcting angular velocity with the Madgwick algorithm, we create a rotation matrix based on the loading of gravity during quiet periods (see 3S.4.3. above) then used it to create a “tilt-corrected” coordinate system (see 3S.4.8. below). Then, between each quiet period, we used angular velocity to calculate changes in orientation based on Equations 2 and 3 in McGinnis & Perkins [174]. This provided a rotation matrix from the wearable coordinate system to the “tilt-corrected” coordinate system for each time step.

Using these time-varying rotation matrices, acceleration data for each frame was expressed in the “tilt-corrected” coordinate system then 1 g was subtracted from the y-axis (in line with gravity). This procedure removed the gravity component from the accelerometer data. To create the wearable and segment coordinate systems (in the next step), data were then re-expressed in their original coordinate system using the inverse of the time-varying rotation matrices.



**Figure 3S.47: Orientation estimation and gravity subtraction.** Top: Uncorrected acceleration of the sacrum for the entire data collection of the example participant (colors represent different axes). Bottom: Acceleration of the sacrum after subtracting 1 from the y-axis in the “tilt-corrected” coordinate system and then re-expressing in the wearable coordinate system (colors represent different axes).

### 3S.4.8. Coordinate system definition

Finally, data were expressed in three different coordinate systems for analysis. First, data were expressed in the Wearable Coordinate System (WCS). This is not the raw coordinate system of the IMU. Rather, all data were corrected with the calibration matrices described in 3S.4.1. above. These calibration matrices ensured that data were expressed in orthogonal axes aligned with the IMU housing. The IMU housing was positioned such that, during quiet standing, the WCS axes were oriented roughly in the direction of progression (+x), the longitudinal axis (+y), and to the right (+z).

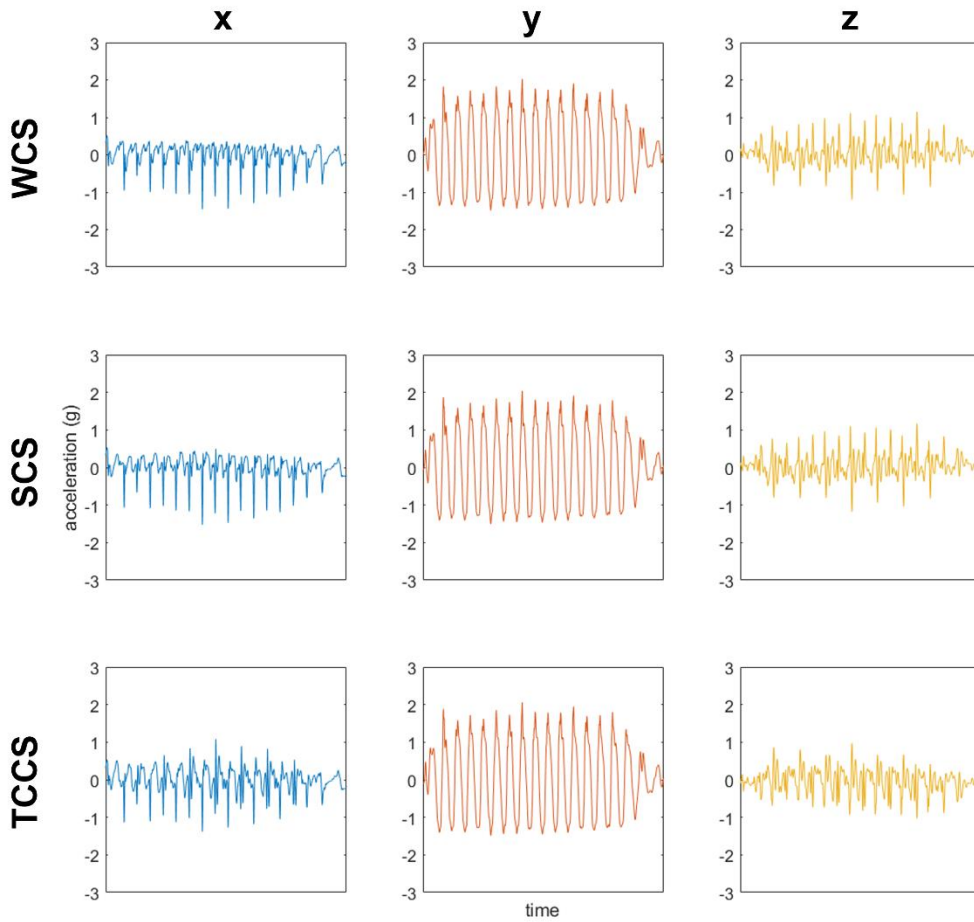
Data were also expressed in a Segment Coordinate System (SCS). This coordinate system was defined using an approach described in the Supplemental Material of Cain et al. [121] which can be found at:

<http://dx.doi.org/10.1016/j.gaitpost.2015.10.022>

In brief, accelerations during a quiet standing trial were used to define a gravity vector (similar to 3S.4.3. above) assuming that the segment was aligned with gravity during the standing trial. This gravity-based vector was defined as the proximal-distal axis (+y proximal). Then a period of steady-state running was manually selected from the data set. Angular velocities from this period were entered into a Principal Component Analysis and the Principal Component accounting for the most variability in angular velocity was selected to represent the average axis of rotation. During running the average axis of rotation was assumed to correspond to the medial-lateral axis. We defined this as the z axis (+z right). The anterior-posterior axis was then defined as the cross-product of y and z (+x anterior). Finally, the z axis was recalculated as the cross-product of x and y to ensure orthogonality. These three unit vectors were then used to create a rotation matrix that transformed data from the WCS to the SCS.

Finally, data were expressed in a pseudo-global system similar to Cain et al.'s "tilt-corrected" coordinate system (TCCS) [121]. First, we created a rotation matrix based on the loading of gravity

during quiet periods (see 3S.4.3. above). This rotation matrix expressed data with the y-axis aligned with gravity during quiet standing (+y vertical). Next, the acceleration of each axis was double integrated to obtain displacement then entered into a Principal Component Analysis. The Principal Component accounting for the most variation in displacement was taken as the projection of the direction of progression onto the horizontal plane (+x direction of progression). Then the projection of the medial-lateral axis onto the horizontal plane was defined as the cross product of x and z (+z right). The x-axis was then recalculated to ensure orthogonality. These three unit vectors were then used to create a rotation matrix and multiplied by the time-varying rotation matrices described in 3S.3.7. to express data in the TCCS. Thus, TCCS data is always expressed with y aligned with gravity but with x and z free to rotate about y as the participant moves.



**Figure 35.48: Coordinate system definition.** All plots show sacral accelerations from the same trial from the example participant. The top row shows data in the Wearable Coordinate System (WCS). the middle row shows data in the Segment Coordinate System. The bottom row shows data in the Tilt-Corrected Coordinate System (TCCS). The first column shows x-axis data (blue). The second column shows y-axis data (orange). The third column shows z-axis data (yellow). Due to the similarities between all three coordinate systems at the sacrum, discrepancies are minor.

## **4. Acceleration-based estimation of vertical ground reaction forces during running: A comparison of methods across running speeds, surfaces, and foot strike patterns**

### **4.1. Introduction**

Ground reaction forces (GRFs) are external reaction forces created with equal magnitude and opposite sense to the force that the foot applies to the ground with each step. Quantifying GRFs is fundamental to running biomechanics research because: (1) per Newton's second law, GRFs dictate the center of mass' (COM) acceleration and can therefore be used to study whole-body motion, (2) this whole-body motion both causes, and is caused by, muscle activity and thus GRF provides insight into that activity [194], (3) in combination with this muscle activity, GRF contributes to the internal loads experienced by structures within the body (e.g., bone, ligament, tendon, cartilage) leading to its frequent (though much contested) investigation as a risk-factor for injury [195, 196, 41, 68, 197, 198, 199, 200, 201]; (4) the magnitude and sense of GRF is used to assess performance [202, 203, 204, 205, 206]; and finally, (5) because GRFs are critical to inverse dynamics calculations, allowing estimation of joint forces and moments and more advanced analysis of behavior. Thus, accurate quantification of GRF during running is an important goal.

Accurate quantification of GRF is relatively easy with a force plate or instrumented treadmill, however, this equipment is generally 'captive' to lab environments that may not represent the actual conditions runners experience. To increase ecological validity, previous research has attempted to replicate field conditions within the lab [207, 208, 209, 210, 211]. Even with such attempts, however, spatial constraints may cause participants to alter their gait on a treadmill or short running track [212] and temporal constraints still limit the duration and volume of data collection. Thus, lab measured GRFs



may not accurately represent the millions of GRFs that occur over many long bouts of running in the field, limiting our understanding of GRF and its relation to other variables.

Some of these constraints have been overcome through the advancement of force sensing insoles [213, 214, 215, 216, 217, 218], wearable load cells [219, 220, 221], and instrumented shoes [222, 223, 224]. However, this equipment still suffers from issues with durability, comfort, changing the mechanical properties of a shoe (i.e., making it more rigid), and interfering with the foot-ground (or foot-shoe) interface. Consequently, biomechanists remain largely reliant on 'captive' technology to measure GRF, decreasing the ecological validity and volume of available data.

Accelerometers offer a promising alternative to overcome this reliance on 'captive' technology. These small, low-cost wearable devices may allow the capture of greater volumes of more ecologically valid data than traditional 'captive' equipment. Sets of multiple wearables can estimate GRFs during walking [225, 226, 227, 228] and running [229, 230, 231] and several methods have been proposed that truly capitalize on the advantages of accelerometers (minimizing system complexity, preparation time, participant discomfort, and costs) by estimating GRFs with a single accelerometer. These single-accelerometer methods often place the accelerometer on the shank or locations intended to approximate whole body COM such as the hip, lower back, or upper back (other locations, such as the wrist, have also been investigated but show poor correlations with GRFs [232, 233]).

Researchers attempting to estimate GRF from shank accelerations build on observations that shank acceleration and GRF signals are closely related [234, 235, 236] and therefore should allow estimation of one from the other [237, 238, 239]. Further, some argue that forces applied at the ground are damped as they travel up the body and thus, measuring accelerations at the shank better reflects GRFs than more proximal mounting locations [240]. Researchers also point to similarities in the timing of

peaks in the tibial acceleration and GRF signals [241] (cf. [242]) and argue for a mechanical coupling of these peaks [243].

Relations between GRF and accelerometers mounted on the approximate COM have also been explored [244, 232]. Use of these COM locations is based on Newton's second law which states that whole body acceleration is inversely proportional to the mass of a body and proportional to the net forces acting on that body [245, 246, 247, 248]. Given a constant mass, if forces other than GRF are relatively small (e.g., air resistance), then the COM acceleration is a function of gravitational force and GRF. A limitation of this 'COM' approach is the assumption that a single accelerometer with a static position can capture whole body COM acceleration even though the COM location can move during running (due to limb movements and changes in posture). Despite this movement, previous research supports this assumption and demonstrates that a sacrum mounted accelerometer captures whole body COM acceleration during running fairly well [249, 250, 251, 252, 253].

Based on these arguments there have been many attempts to estimate GRF from acceleration at these two mounting locations [254], however, only one study has conducted a head-to-head comparison of methods and that study only compared two of the many methods available [255]. Thus, there are no comprehensive recommendations to guide users on which method to use for a given application. To overcome this gap in the literature, methods to estimate GRF from a single accelerometer were replicated and compared. Methods were required to be non-participant or -trial specific (cf. [161]), non-proprietary (cf. [256, 257, 242]), report promising results (cf. [258]), and be capable of providing stance-by-stance estimates of at least one feature of the vertical GRF (first peak, loading rate, second peak, average, or time series) (cf. [248]) using only easy-to-measure anthropometrics and/or features of an acceleration input signal (cf. [248, 259, 208, 260]) from a single sensor on the shank or COM (cf. [229, 230, 231, 261, 262, 263]). In total, 27 methods derived from 13 publications met these criteria. The 13 original publications are described in Table 4.1 while the

methods derived from those publications and the vertical GRF feature they are capable of estimating are described in Table 4.2.

To evaluate which of these 27 methods provides the most accurate, reliable, and precise estimate of vertical GRF first peak, loading rate, second peak, average, and/or time series, errors were calculated relative to a gold-standard force plate. For each method, errors were compared across a range of speeds, foot strike angles, and running surfaces to explore whether method performance varied across these conditions. Results demonstrate the best method to estimate vertical GRF parameters from a single accelerometer under given surface, foot strike, and running speed conditions. Code to automatically execute each of the methods on stance-segmented accelerometer data is provided at [https://github.com/DovinKiernan/MTFBWY\\_running\\_vGRF\\_from\\_a](https://github.com/DovinKiernan/MTFBWY_running_vGRF_from_a).

**Table 4.1: Published ground reaction force estimation methods.** Thirteen publications met our inclusion criteria and presented one or more methods to estimate at least one feature of the vertical GRF (first peak, loading rate, second peak, average, or time series) from a single accelerometer on the shank or approximate COM (hip, sacrum, lower back, or upper back).

Publication	Sample	Foot-strike	Speed	Surface	Placement	Signals	Range & Frequency	Targets	Ground truth	Sync
Neugebauer 2012, 2014 [86, 87]	$n = 35$ (20 F 15 M) children [86] $n = 39$ (20 F 19 M) injury free adults [87]	NR	2.2 - 4.1 m/s	90 [86] and 15 m [87] overground	Right iliac crest	$\alpha_{WCSres}$ [86] $\alpha_{WCSx,y}$ [87]	NR 40 Hz [86] $\pm 6$ g 100 Hz [87]	$F_{y,max}$	Force plate 1000 Hz	Average 30 [86] or 10 s [87]
Charry 2013 [264]	$n = 3$	NR	1.7 - 7.2 m/s	Overground	Medial mid-tibia	$\alpha_{WCSy}$	$\pm 24$ g 100 Hz	$F_{y,second}$	Force plate 300 Hz	Video
Wundersitz 2013 [246]	$n = 17$ (5 F 12 M) uninjured team sport	NR	2.5 - 7.4 m/s	10 m Overground	2 <sup>nd</sup> Thoracic vertebra	$\alpha_{WCSy,res}$	$\pm 8$ g 100 Hz	$F_{y,max}$	Force plate 100 Hz	Video
Meyer 2015 [265]	$n = 13$ (3 F 10 M) moderately active children	NR	1.7 - 2.8 m/s	10 m overground	Hip	$\alpha_{WCSy}$	$\pm 8$ g 100 Hz	$F_{y,max}$	Force plate 2400 Hz	Average 8-15 steps
Gurchiek 2017 [247]	$n = 15$ (3 F 12 M)	NR	Sprinting and cutting	Overground	Sacrum	$\alpha_{GCSx,y,z}$	$\pm 24$ g 450 Hz	$F_{y,t}$ $F_{y,average}$	Force plate 1000 Hz	Counter-movement jumps
Thiel 2018 [240]	$n = 3$ Elite sprinters	NR	Sprint	Overground	Above medial malleolus	$\alpha_{WCSx,y,z}$	$\pm 16$ g 250 Hz	$F_{y,max}$	Force plates 1000 Hz	LED flash
Kiernan 2020 [266]	$n = 40$ (NR)	NR	NR	25 m Overground	Sacrum; iliac crest	$\alpha_{SCSy}$	$\pm 100$ g 1000 Hz	$F_{y,first}$ $F_{y,second}$	Force plate 1000 Hz	TTL pulses
Kim 2020 [267]	$n = 7$ (0 F 7 M)	NR	2.9 m/s	Treadmill	Sacrum	$\alpha_{WCSx,y,z}$	200 Hz	$F_{y,t}$	Force plate 400 Hz	NR
Pogson 2020 [268]	$n = 15$ (5 F 10 M) team sport players	NR	2.0 - 8.0 m/s	Overground	Back of upper torso	$\alpha_{WCSx,y,z}$	$\pm 16$ g 100 Hz	$F_{y,t}$	Force plate 3000 Hz	Synchronous recording
Day 2021 [269]	$n = 30$ (21 F 9 M) NCAA Div 1 xcountry	NR	3.8 - 5.4 m/s	Treadmill	Posterior waistband	$\alpha_{WCSx,y,z}$	500 Hz	$F_{y,t}$ $F_{y,max}$	Instrumented treadmill 500 Hz	Average 10 s (jump)
Higgins 2021 [232]	$n = 30$ (15 F 15 M) healthy	NR	$\sim 1.8 - 5.0$ m/s	23 m overground	Superior to lateral malleolus; hip	$\alpha_{WCSy,res}$	$\pm 8$ g 100 Hz	$F_{y,first}$ $\frac{dy}{dx} F_{y,first}$	Force plate 1000 Hz	Vertical jumps
Veras 2022 [255]	$n = 131$	NR	1.9 – 3.9 m/s	Treadmill	Superior to lateral malleolus; iliac crest; sacrum	$\alpha_{WCSy\&res}$	$\pm 16$ g 100 Hz	$\frac{dy}{dx} F_{y,first}$ $F_{y,max}$	Instrumented treadmill 1000 Hz	Manual correction and cross-correlation

M – Male; F – Female; NR – Not reported; FF – forefoot; RF – rearfoot; SS – Self-selected;

$\alpha$  – acceleration; SCS – segment coordinate system; WCS – wearable coordinate system; GCS – global coordinate system (coordinate conventions defined in Fig. 2)

$F_{y,first}$  – first (or ‘impact’) peak;  $\frac{dy}{dx} F_{y,first}$  – loading rate;  $F_{y,second}$  – second (or ‘active’) peak;  $F_{y,max}$  – maximum, assumed to correspond to second (or ‘active’) peak;  $F_{y,t}$  – time series

**Table 4.2: Twenty-seven methods derived or adapted from 13 publications** in Table 3.1. Methods are sorted by accelerometer placement location. Methods that originally placed accelerometers on the lumbar or thoracic spine have been adapted to the sacrum. Not all methods could estimate all potential vertical GRF features. If a method was originally designed to estimate a feature it is specified as ‘designed’ and highlighted blue. If a method was designed to estimate a time series, all discrete vertical GRF features were then derived from that time series. These are marked ‘derived’ and highlighted pink.

Sensor location	Method	Estimated force variable				
		First peak	Loading rate	Second peak	Average	Time series
Shank	Cherry			designed		
	Thiel			designed		
	Veras shank res		designed	designed		
	Veras shank y		designed	designed		
	Higgins shank	designed	designed			
Hip	Neugebauer			designed		
	Meyer			designed		
	Kiernan hip	designed		designed		
	Veras hip res		designed	designed		
	Veras hip y		designed	designed		
	Higgins hip	designed	designed			
Sacrum	Gurchiek	derived	derived	derived	designed	designed
	Kim acceleration	derived	derived	derived	derived	designed
	Kim displacement	derived	derived	derived	derived	designed
	Kiernan sacrum	designed		designed		
	Veras sacrum res		designed	designed		
	Veras sacrum y		designed	designed		
	Day 5 Hz	derived	derived	designed	derived	designed
	Day 10 Hz	derived	derived	designed	derived	designed
	Day 30 Hz	derived	derived	designed	derived	designed
	Wundersitz 10 Hz			designed		
	Wundersitz 15 Hz			designed		
	Wundersitz 20 Hz			designed		
	Wundersitz 25 Hz			designed		
	Wundersitz raw			designed		
	Pogson	derived	derived	derived	derived	designed
	Pogson xynormed	derived	derived	derived	derived	designed

## 4.2. Methods

Data from this study were first reported in a separate analysis in Chapters 3 and 3S but methods are briefly repeated here for convenience.

### 4.2.1. IMU calibration

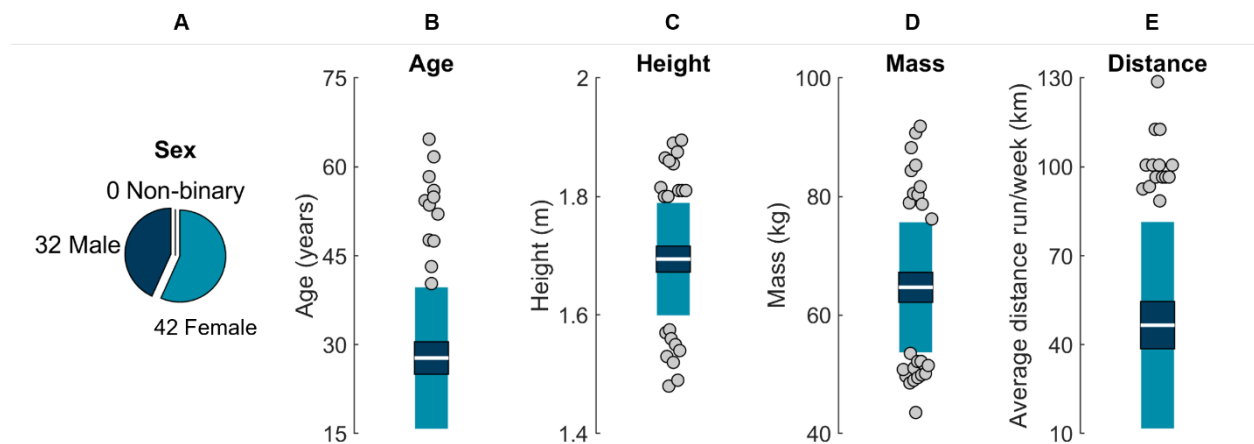
Adapting methods from Coolbaugh et al. [99], tri-axial IMUs (ProMove MINI, Inertia Technology, Enschede, The Netherlands;  $\pm 16$  g primary,  $\pm 100$  g secondary,  $\pm 34.91$  rad/s, 1000 Hz; see

<https://inertia-technology.com/wp-content/uploads/2022/02/ProMove-mini-datasheet.pdf> for further

details on device specifications and operation) were calibrated with a centrifuge (ClearPath MCVC, Teknic, Victor, NY, USA) and custom 3-D printed jigs (SOLIDWORKS 2019, Dassault Systèmes, Vélizy-Villacoublay, France). After calibration, IMU primary accelerometer errors were  $\leq 0.01 \pm 0.04$  g, secondary accelerometer errors were  $\leq 0.05 \pm 0.07$  g, and gyroscope errors were  $\leq 0.01 \pm 0.01$  rad/s.

#### 4.2.2. Participants

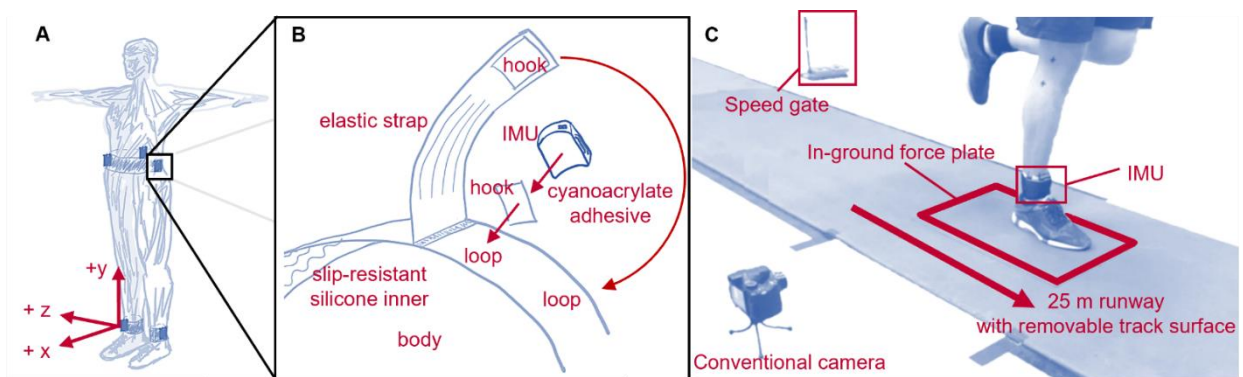
Seventy-seven participants were recruited from the University of California Davis, local running clubs, and the community at large. Participants were  $\geq 18$  years old and reported running  $\geq 16.09$  km per week for  $\geq 6$  months. Three participants were excluded from analysis due to movement of an IMU ( $n = 2$ ) or inability to complete the protocol as instructed ( $n = 1$ ), leaving a final sample of 74 (32 males; 42 females; 0 non-binary; age  $28 \pm 12$  years; Figure 4.1). All participants provided written informed consent, and procedures were approved by the University of California Davis Institutional Review Board.



**Figure 4.1: Participant details.** (A) Sex, (B) age, (C) height, (D) mass, and (E) self-reported average distance run per week. The white horizontal line represents the mean; dark blue represents  $\pm 95\%$  confidence interval ( $\pm 1.96$  SEM) around the mean; and light blue represents  $\pm 1$  SD around the mean. Gray dots represent participants outside  $\pm 1$  SD.

#### 4.2.3. Protocol

Participant mass, height, and distance of the left and right lateral malleolus, fibular head, lateral epicondyle, and superior aspect of greater trochanter from the ground were measured. Using adhesive-bonded hook-and-loop fasteners, IMUs were attached to neoprene belts with anti-slip silicone inners, then wrapped with elastic straps as tightly as possible, within the limit of participant comfort. IMUs were placed anterior and superior to the lateral malleoli (*shank*), on the superior aspect of the iliac crests in line with the greater trochanter (*hip*), and on the superior aspect of the sacrum in line with the spine (*sacrum*) (Figure 4.2).



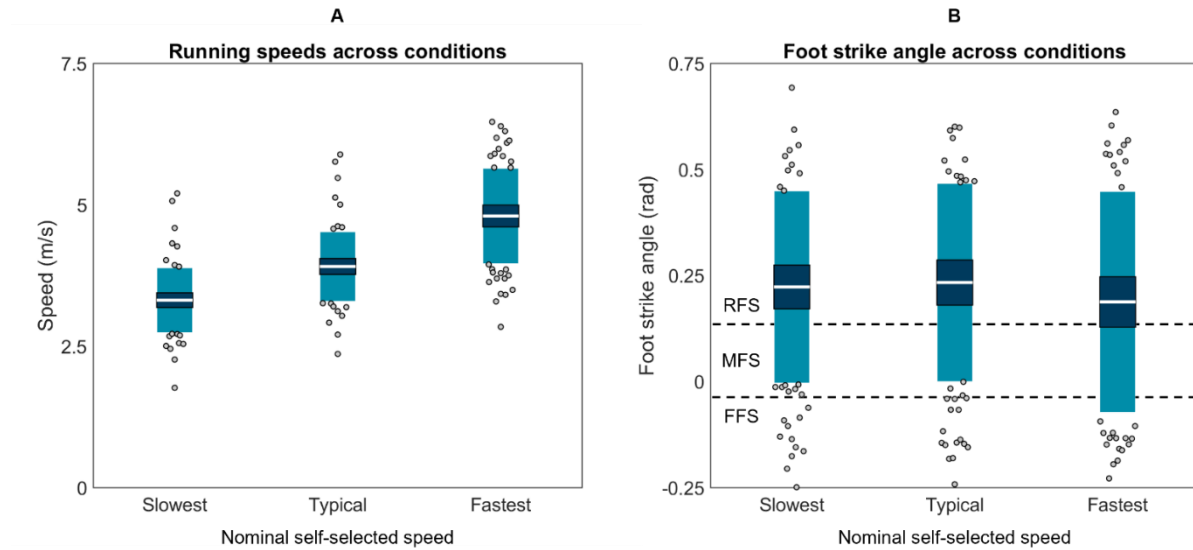
**Figure 4.2: Experimental set up. (A)** IMU placement and coordinate conventions. For consistency, different conventions used across methods have been standardized to ISB conventions [169]: Segment coordinate systems (SCS) were defined as anterior (+x), proximal (+y), and medial-lateral (with right defined as +z); wearable coordinate systems (WCS) were defined square to the IMU housing, which was roughly aligned with the direction of progression (+x), longitudinal axis (+y proximal), and right (+z); Tilt-corrected coordinate systems (TCCS) were defined as vertical (+y) and the projections of direction of progression (+x) and the medial-lateral axis (+z right) onto the horizontal plane. **(B)** Belt design and IMU fixation. **(C)** Experimental setup.

Participants wore their own shoes and ran a 25 m runway with an embedded force plate (Kistler 9281, Kistler Group, Winterthur, Switzerland; 1000 Hz). Running speed was recorded using two custom-built laser speed gates, placed 2.5 m on each side of force plate center. Participants warmed up and practiced striking the force plate three times per side at their *slowest* (“the slowest pace you would use on a run”), *typical* (“the pace you use for the majority of your running”), and *fastest* (“the fastest pace you would use on a run”) self-selected speeds (Figure 4.3). During this warmup, markers on the lateral calcaneus and base of the fifth metatarsal were recorded using a conventional video camera (Exilim EX-

FH25, Casio; 120 Hz). Foot strike angle was calculated by subtracting a neutral standing foot angle from the foot angle at initial contact (Kinovea 0.9.5). Positive values indicate a more dorsiflexed foot at initial contact with values  $> 0.14$  rad corresponding to rear-foot strike,  $-0.03$  to  $0.14$  rad to mid-foot strike, and  $< -0.03$  rad to forefoot strike [270]. After warm-up, five stances per side were collected at each speed for two surface conditions: (1) with a *track* surface covering the runway and force plate, and (2) with no covering on the hardwood *floor* of a basketball court. Participants always progressed from their slowest to fastest speeds, but the order of foot and surface was pseudo-randomized.

IMU data were synchronized within 100 ns of each other with a wireless network hub (Advanced Inertia Gateway, Inertia Technology, Enschede, The Netherlands). This hub sent voltage pulses that were synchronously recorded by IMU software and a custom MATLAB script that simultaneously recorded the speed and force data (R2018b, MathWorks, Natick, USA). Pulse trains were cross correlated to synchronize signals. During data processing, we observed small timing discrepancies caused by the initialization of discrete MATLAB data acquisitions and small variances between the sampling rates of the IMU and MATLAB systems. Although extremely small, these discrepancies could accumulate over the course of the  $\sim 60$  min data collection, leading to timing differences between the first and last synch events of a data collection (on the order of 10s of ms). To ensure the input acceleration data were perfectly matched with the target force data, a conservative approach was used and only trials containing a synch event were analyzed (642 of 4440 trials). All other trials were discarded to ensure millisecond-level accuracy.





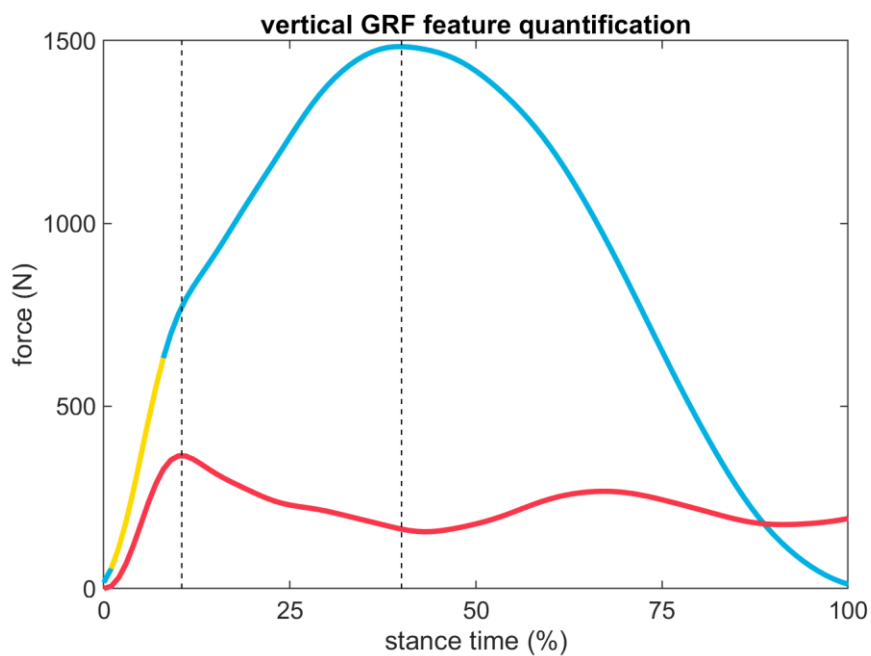
**Figure 4.3: Observed speeds and foot strike angles. (A)** Mean speeds and **(B)** foot strike angles used by each participant across the slowest, typical, and fastest conditions (RFS is rear foot strike, MFS is mid foot strike, and FFS is fore foot strike). The white horizontal line represents the mean; dark blue represents  $\pm 95\%$  confidence interval ( $\pm 1.96$  SEM) around the mean; and light blue represents  $\pm 1$  SD around the mean. Gray dots represent participants outside  $\pm 1$  SD.

#### 4.2.4. IMU data processing

Calibration matrices were applied to IMU data. Quiet periods were identified (angular velocity  $< 0.5$  rad/s and jerk  $< 0.01$  m/s<sup>3</sup> for at least 100 ms) and used to remove biases. Saturated frames from the primary accelerometer ( $|a| > 15.5$  g) were replaced with corresponding frames from the secondary accelerometer. Data were filtered with a 4th-order 50-Hz low-pass Butterworth filter. Angular velocity was drift-corrected using a Madgwick filter [171, 193, 173, 172]. Starting at each quiet period, accelerations were used to estimate the IMU's position in the inertial reference frame, then angular velocities were used to estimate frame-by-frame changes in IMU orientation and remove the gravity component from accelerations [174]. Data were then expressed in a segment coordinate system based on the Principal Component that explained the most variance in angular velocity during running (the medial-lateral axis) and the gravity vector during quiet standing [121, 175]. IMU data during stance were extracted based on the instant the time-synchronized vertical force crossed a 10 N threshold. For more detailed IMU processing see Chapter 3S.4.

#### 4.2.5. Force data processing

Force data were filtered with a 4th-order 50-Hz low-pass Butterworth filter. A vertical force threshold of 10 N was used to define the start and end of stance. The first (or 'impact') peak was identified by performing a Fourier Transform on the vertical GRF then reconstructing a time domain signal from the  $\geq 10$  Hz high frequency ('HiF') components with an Inverse Fourier Transform [271]. First peak magnitude was defined as the magnitude of the original vertical GRF signal at the time when the HiF signal achieved its earliest peak occurring after 5% of stance duration. The loading rate was calculated from 20 - 80% of stance onset to first peak [272]. The second (or 'active') peak was defined as the maximum magnitude of the vertical GRF (or the magnitude of the second peak if two peaks were present). Average was also calculated across stance. These methods are depicted for a single stance in Figure 4.4.



**Figure 4.4: Vertical GRF and HiF reconstruction for an example stance** (blue and red, respectively). Vertical dashed lines indicate the timing of first and second vertical GRF peaks. Note, although the first peak is difficult to visually identify in the

original signal (blue), a consistent point in the HiF signal can still be identified. The yellow highlighted region from 20-80% of the first peak was used to calculate loading rate.

#### 4.2.6. Analysis

First peak, loading rate, second peak, average force, and force time series were estimated with each capable method (Table 4.2). For methods that required a model to be built (see Chapter 4S), these features were estimated using a leave one out analysis where 74 models were iteratively trained with data from 73 participants then used to estimate features for the single participant the model was not trained on. Errors were calculated by subtracting the ground truth force plate value from the estimated value (for first peak, loading rate, second peak, or average) or by calculating the RMSE (for time series). One sample t-tests were used to compare each method's error to the gold standard (0 error). Significance was set at  $p \leq 0.05$  with a False Discovery Rate (FDR) procedure to correct for multiple comparisons.

To account for the non-independence of the data (642 trials from 74 participants) and ensure proper estimation of variance, an adaptation of the Bland-Altman method was used [179]. Errors were entered into linear mixed-effects models in R (v4.2.2; R Foundation for Statistical Computing, Indianapolis, USA) as described in Carstensen et al.'s approach to linked replicates [176, 273]:

##### Equation 4.1

$$y_{method,participant,trial} = \alpha_{method} + b_{participant} + C_{method,participant} + D_{participant,trial} + \varepsilon_{method,participant,trial}$$

Where  $y$  corresponds to the model estimated error, lower case terms correspond to fixed effects, upper case terms correspond to random effects, and  $\varepsilon$  corresponds to error. Model assumptions of independence, normality, and homoscedasticity were validated by plotting within-participant variances against within-participant means, histograms of residuals, residuals for each level of random effect, and residuals as a function of fitted value. Method-specific variance components were extracted using the 'MethComp' package for R [274]:

**Equation 4.2**

$$C_{method,participant} \sim N(0, \tau_{method}^2)$$

**Equation 4.3**

$$\varepsilon_{method,participant,trial} \sim N(0, \sigma_{method}^2)$$

Where the values of  $C$  and  $\varepsilon$  are normally distributed about zero with variances of  $\tau^2$  and  $\sigma^2$  for each method. This allowed estimation of: (1) method biases that quantify accuracy (mean error); (2) repeatability coefficients (RC) that quantify the largest absolute difference predicted between two measurements on the same participant under identical circumstances; and (3) limits of agreement (LOA) that quantify precision (limits within which 95% of future errors for a given method are expected to fall), using the equations:

**Equation 4.4**

$$RC_{method} = \pm 2.83\sigma_{method}$$

**Equation 4.5**

$$LOA_{method} = \pm 1.96\sqrt{\tau_0^2 + \tau_{method}^2 + \sigma_0^2 + \sigma_{method}^2}$$

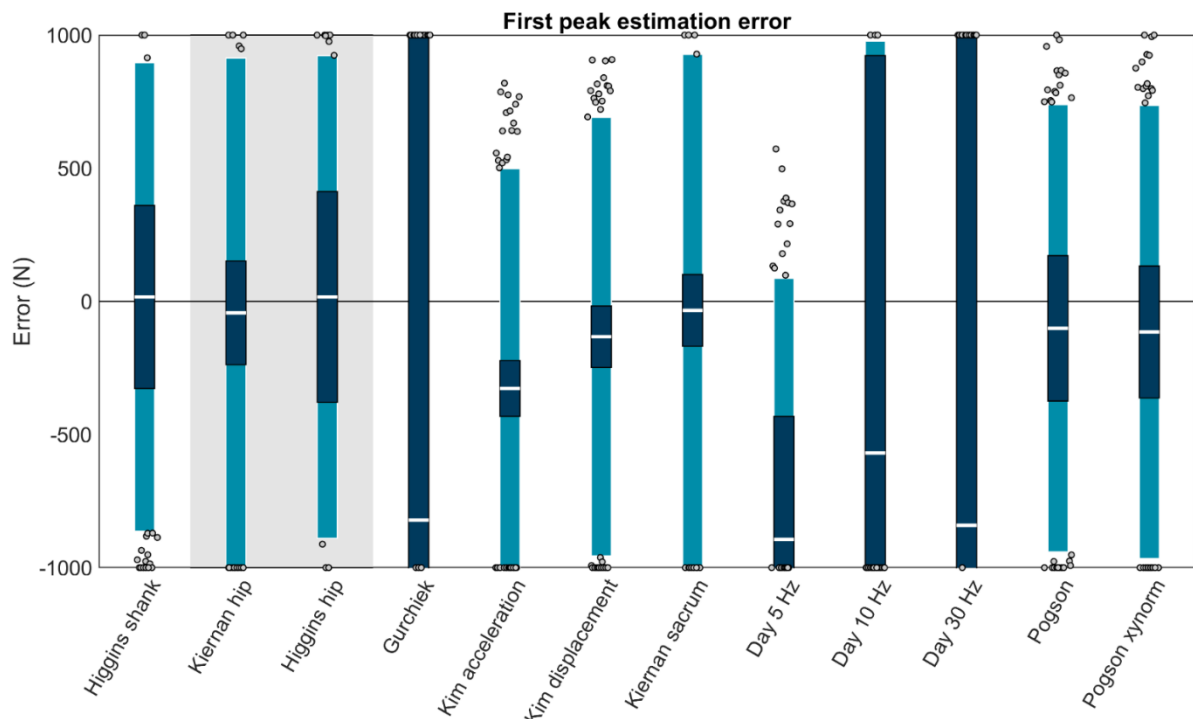
Where  $\tau_0^2$  and  $\sigma_0^2$  correspond to variances for the gold standard.

To evaluate if any potential explanatory variables affected method error, a second set of linear mixed effects models was developed for each method. These models added surface, speed, and foot strike angle as fixed effects. A  $p \leq 0.05$  for any fixed effect was interpreted as that fixed effect accounting for a significant amount of a method's error (i.e., model-estimated force was significantly affected by running surface, speed, and/or foot strike angle).

**4.3. RESULTS****4.3.1. First peak**

Using Carstensen's method for linked replicates [176], biases, RCs, and LOAs were calculated for each method capable of first peak estimation (Figure 4.5; Table 4.3). First peak magnitude was

estimated at the shank by one method, which was among the best performing (biases < 200 N): ‘Higgins shank’ (+16.46 ± 879.51 N or +1.35 ± 72.23%; bias ± LOA). At the hip by ‘Kiernan hip’ (-43.18 ± 957.18 N or -3.55 ± 78.61%) and ‘Higgins hip’ (+16.46 ± 906.02 N or +1.35 ± 74.41%). At the sacrum, by nine methods, with four among the best performing: ‘Kim displacement’ (-132.74 ± 823.12 N or -10.90 ± 67.60%), ‘Kiernan sacrum’ (-33.81 ± 961.22 N or -2.78 ± 78.94%), ‘Pogson’ (-101.25 ± 839.57 N or -8.32 ± 68.95%), and ‘Pogson xynorm’ (-115.06 ± 850.78 N or -9.45 ± 69.87%).




**Figure 4.5: First peak estimation** bias (white bar), ± RC (dark blue), and ± LOA (light blue) for each capable method. Gray dots represent trials falling outside the LOA. Values outside ± 1000 N are plotted at the axis limits. A value of 0 represents perfect agreement with the force plate. Positive values indicate the method overestimated the first peak. Negative values indicate the method underestimated the first peak. The method with a white background on the left is for accelerometers on the shank, methods with a gray background are for the hip, and methods with a white background on the right are for the sacrum.

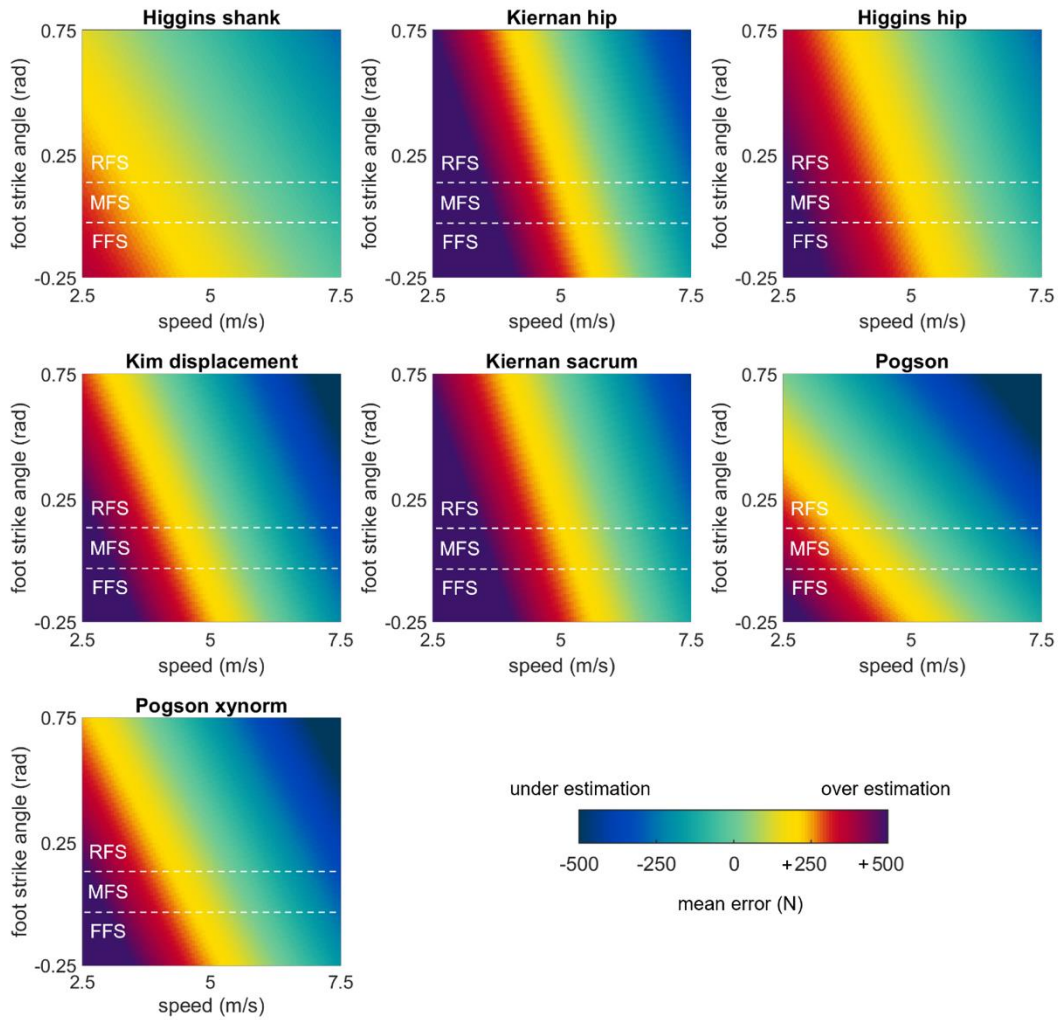
We performed a second set of linear mixed effects models on each of the best performing methods (biases < 200 N) to examine the role of running speed, surface, and foot strike angle as potential explanatory variables. These models revealed that error in each of the best performing methods was significantly explained by running speed and foot strike angle ( $p \leq 0.05$ ) but not by

running surface ( $p_s > 0.05$ ) (Table 4.3). To illustrate these effects, model-predicted biases were plotted as a function of speed and foot strike angle (Figure 4.6). Yellow, red, and purple colors in these plots correspond to over estimations (positive errors where the estimate has a greater magnitude than the gold standard), while darker blue colors correspond to under estimations (negative errors where the estimate has a smaller magnitude than the gold standard). All methods showed the same general pattern with overestimates at low speeds and foot strike angles, and underestimates at high speeds and foot strike angles.

**Table 4.3: Best performing first peak estimation methods.** Overall performance shown as biases (accuracy), RCs (repeatability), and LOAs (precision) color-coded from the absolute minimum (green) to absolute maximum (purple) values observed within-column. Biases that significantly differed from 0 ( $p \leq 0.05$  with FDR correction) marked \*. Performance across conditions shown as coefficients for the intercept of surface (added to model estimated error for the track condition but not the floor condition) and slopes for running speed (in m/s) and foot strike angle (in rad). If surface, speed, or foot strike explains a significant ( $p \leq 0.05$ ) amount of error it is highlighted pink and marked \*.

Method	Overall performance			Performance across conditions		
	Bias (N)	RC (N)	LOAs (N)	Speed	Surface	Foot strike
Higgins shank	+16.46	343.88	879.51	-83.70*	11.89	-243.59*
Kiernan hip	-43.18*	194.73	957.18	-191.48*	-5.69	-300.01*
Higgins hip	+16.46	395.53	906.02	-131.03*	-1.19	-231.61*
Kim displacement	-132.74*	115.26	823.12	-181.87*	-12.86	-387.83*
Kiernan sacrum	-33.81*	134.60	961.22	-183.19*	-11.26	-312.66*
Pogson	-101.25*	273.04	839.57	-135.50*	-14.55	-535.86*
Pogson xynormed	-115.06*	247.59	850.78	-170.00*	2.42	-427.28*

column |min|  column |max|

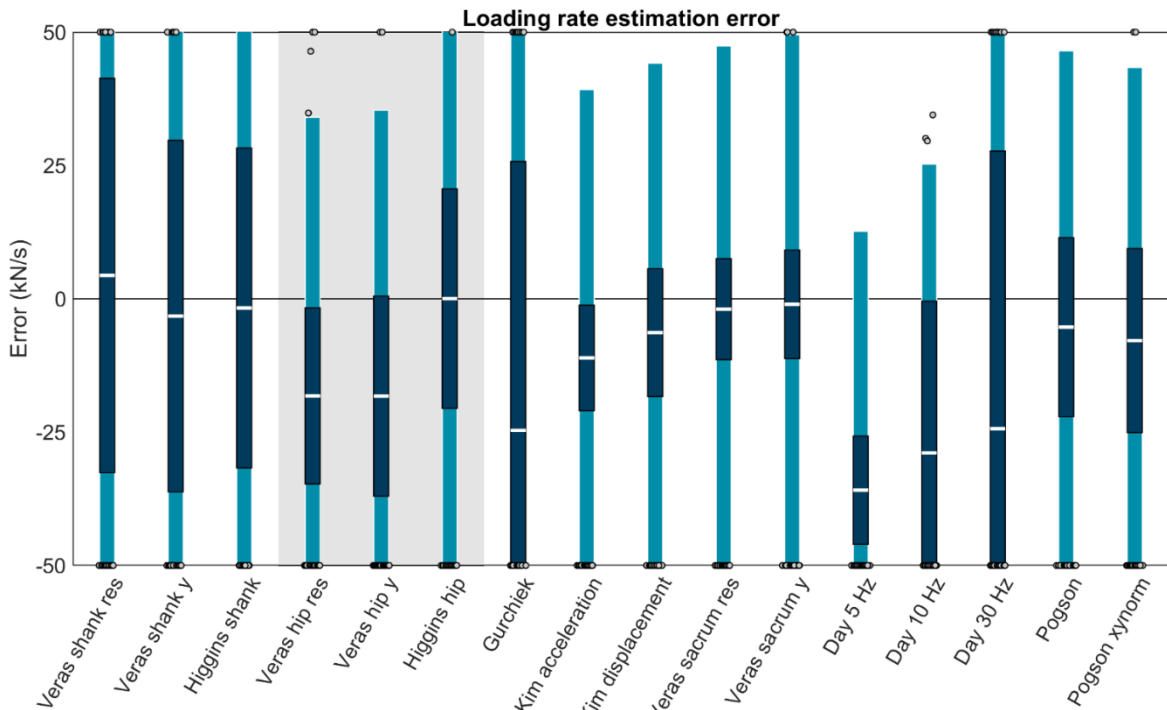


**Figure 4.6: Bias in first peak estimates predicted by mixed effects models for each of the best-performing methods**, plotted as a function of speed and foot strike angle. Green values represent perfect agreement with the gold standard; yellow, red, and purple values represent positive biases (overestimates); darker blue values represent negative biases (underestimates). Foot strike angles corresponding to rear-, mid-, and fore-foot strike patterns have been labelled (RFS, MFS, and FFS) and divided with dashed white lines.

#### 4.3.2. Loading rate

Three methods were capable of estimating loading rate from shank accelerations and were among the best performing methods (biases < 10 kN/s): ‘Veras shank res’ ( $+4.37 \pm 62.56$  kN/s or  $+8.26 \pm 118.16\%$ ; bias  $\pm$  LOA), ‘Veras shank y’ ( $-3.25 \pm 59.96$  kN/s or  $-6.13 \pm 113.25\%$ ), and ‘Higgins shank’ ( $-1.71 \pm 56.69$  kN/s or  $-3.23 \pm 107.08\%$ ). Three were capable of estimating loading rate for hip acceleration but

only ‘Higgins hip’ was among the best performing ( $+0.02 \pm 52.36$  kN/s or  $+0.04 \pm 98.90\%$ ). Ten methods were capable of estimating loading rate from accelerations at the sacrum, of these the best performing were: ‘Kim displacement’ ( $-6.34 \pm 50.52$  kN/s or  $-11.98 \pm 95.43\%$ ), ‘Veras sacrum res’ ( $-1.95 \pm 49.34$  kN/s or  $-3.68 \pm 93.19\%$ ), ‘Veras sacrum y’ ( $-1.04 \pm 50.52$  kN/s or  $-1.96 \pm 95.42\%$ ), ‘Pogson’ ( $-5.32 \pm 51.82$  kN/s or  $-10.04 \pm 97.87\%$ ), and ‘Pogson xynorm’ ( $-7.83 \pm 51.21$  kN/s or  $-14.80 \pm 96.73\%$ ) (Figure 4.7).




**Figure 4.7:** Loading rate bias (white bar),  $\pm$  RC (dark blue), and  $\pm$  LOA (light blue) for each capable method. Gray dots represent trials falling outside the LOA. Values outside  $\pm 50$  kN/s are plotted at the axis limits. A value of 0 represents perfect agreement with the force plate. Positive values indicate the method overestimated the loading rate. Negative values indicate the method underestimated the loading rate. Methods with a white background on the left are for accelerometers on the shank, methods with a gray background are for the hip, and methods with a white background on the right are for the sacrum.

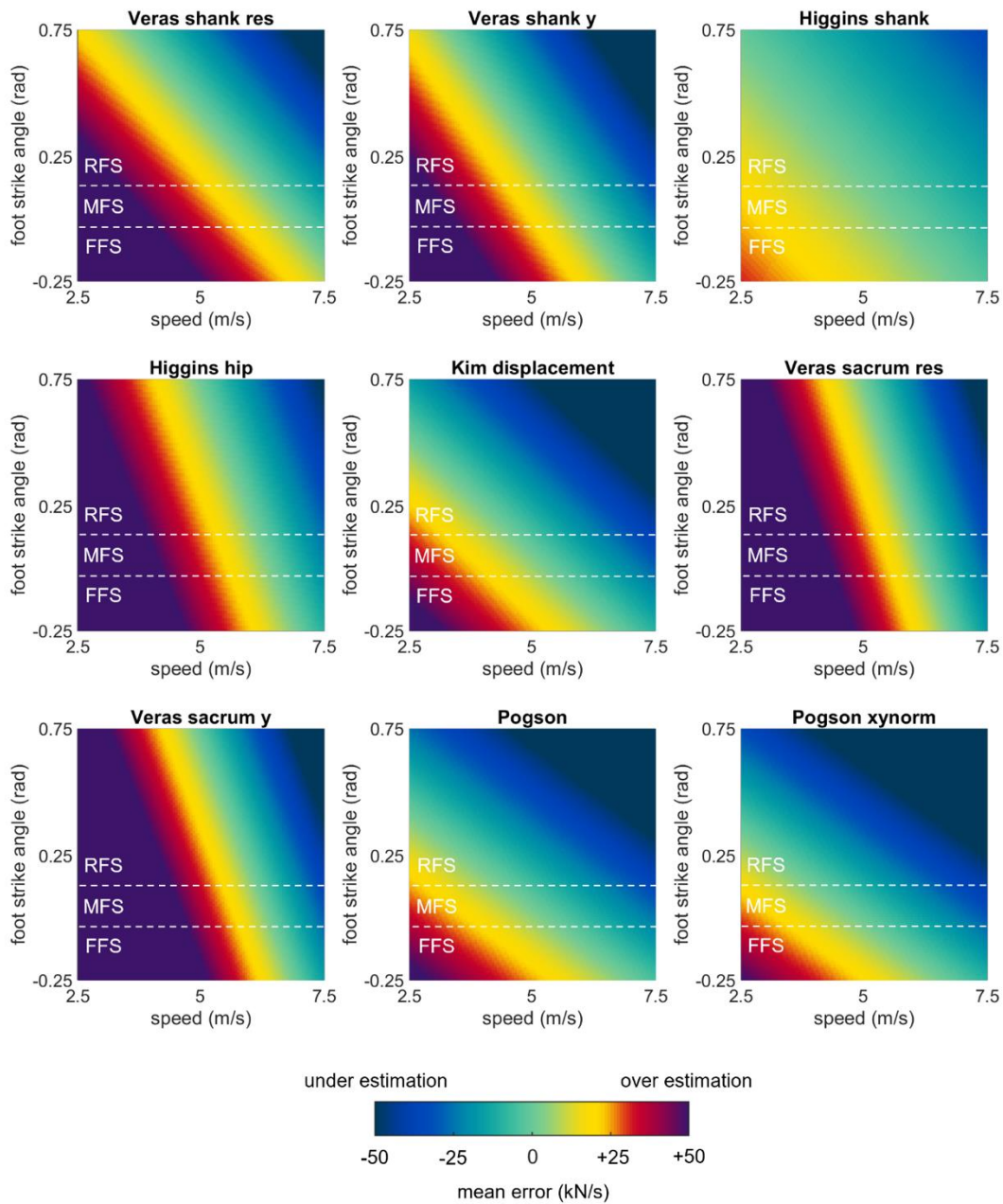
The linear mixed effects models examining the role of running speed, surface, and foot strike angle on the best performing methods (biases  $< 10$  kN/s) revealed that error in all methods was significantly explained by running speed and foot strike angle ( $p \leq 0.05$ ) but not by running surface ( $p > 0.05$ ) (Table 4.4). All methods showed a similar pattern of overestimating loading rate at low speeds and foot strike angles and underestimating at high speeds and foot strike angles (Figure 4.8).



**Table 4.4: Best performing loading rate estimation methods.** Overall performance shown as biases (accuracy), RCs (repeatability), and LOAs (precision) color-coded from the absolute minimum (green) to absolute maximum (purple) values observed within-column. Biases that significantly differed from 0 ( $p \leq 0.05$  with FDR correction) marked \*. Performance across conditions shown as coefficients for the intercept of surface (added to model estimated error for the track condition but not the floor condition) and slopes for running speed (in m/s) and foot strike angle (in rad). If surface, speed, or foot strike explains a significant ( $p \leq 0.05$ ) amount of error it is highlighted pink and marked \*.

Method	Overall performance			Performance across conditions		
	Bias (kN/s)	RC (kN/s)	LOAs (kN/s)	Speed	Surface	Foot strike
Veras shank res	+4.37*	36.96	62.56	-8.28*	-0.29	-36.00*
Veras shank y	-3.25*	32.96	59.96	-9.14*	-0.52	-33.12*
Higgins shank	-1.71	29.99	56.69	-3.47*	0.92	-16.17*
Higgins hip	+0.02	20.58	52.36	-10.66*	-0.61	-21.41*
Kim displacement	-6.34*	11.98	50.52	-6.90*	-1.45	-37.74*
Veras sacrum res	-1.95	9.47	49.34	-12.36*	-0.96	-22.08*
Veras sacrum y	-1.04	10.18	50.52	-13.83*	-0.96	-27.55*
Pogson	-5.32*	16.78	51.82	-5.98*	-1.04	-40.57*
Pogson xynorm	-7.83*	17.25	51.21	-5.99*	-1.03	-42.16*

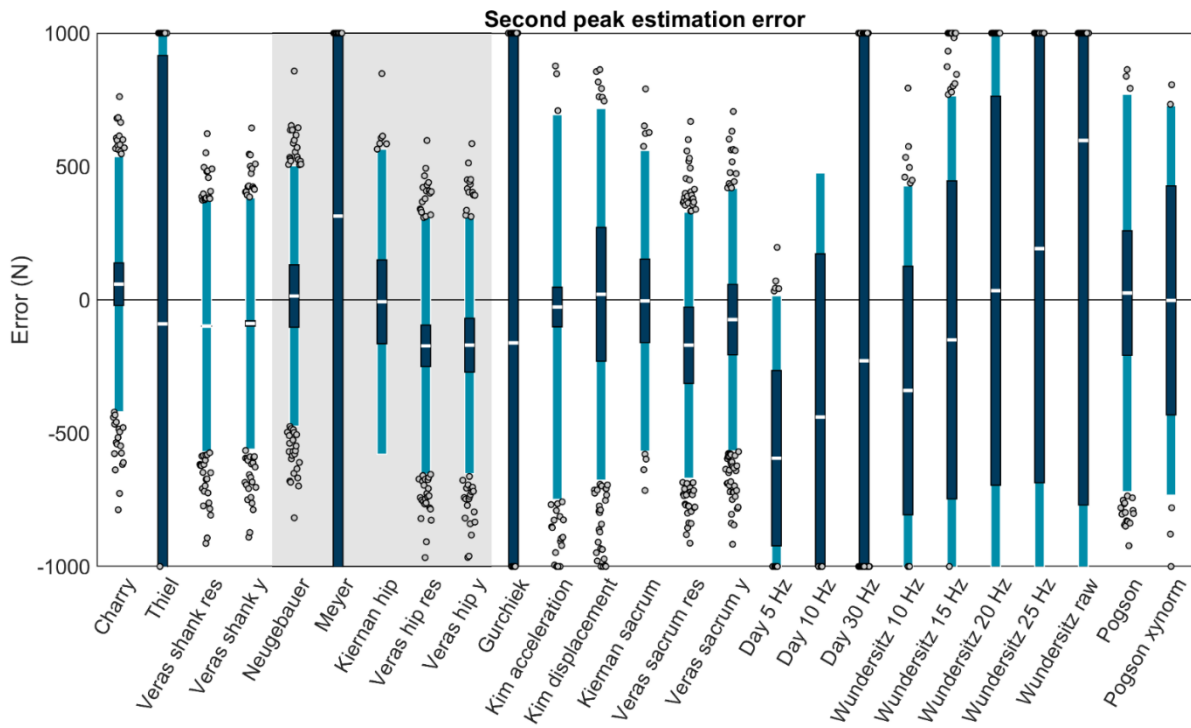
column |min|  column |max|



**Figure 4.8: Bias in loading rate estimates predicted by mixed effects models for each of the best-performing methods, plotted as a function of speed and foot strike angle.** Green values represent perfect agreement with the gold standard; yellow, red, and purple values represent positive biases (overestimates); darker blue values represent negative biases (underestimates). Foot strike angles corresponding to rear-, mid-, and fore-foot strike patterns have been labelled (RFS, MFS, and FFS) and divided with dashed white lines.

### 4.3.3. Second peak

Four methods could estimate second peak from shank accelerations, all were among the best performing methods (biases < 100 N): ‘Charry’ ( $+58.48 \pm 478.42$  N or  $+3.53 \pm 28.85\%$ ; bias  $\pm$  LOA), ‘Thiel’ ( $-90.32 \pm 1162.97$  N or  $-5.45 \pm 70.12\%$ ), ‘Veras shank res’ ( $-98.67 \pm 470.96$  N or  $-5.95 \pm 28.40\%$ ), and ‘Veras shank y’ ( $-88.75 \pm 471.45$  N or  $-5.35 \pm 28.43\%$ ). Five methods could estimate second peak from hip accelerations, with two among the best performing: ‘Neugebauer’ ( $+14.04 \pm 488.09$  N or  $+0.85 \pm 29.43\%$ ) and ‘Kiernan hip’ ( $-7.56 \pm 572.04$  N or  $-0.46 \pm 34.49\%$ ). At the sacrum, 16 methods could estimate second peak, with seven among the best performing: ‘Kim acceleration’ ( $-27.21 \pm 721.47$  N or  $-1.64 \pm 43.50\%$ ), ‘Kim displacement’ ( $+20.68 \pm 696.86$  N or  $+1.25 \pm 42.02\%$ ), ‘Kiernan sacrum’ ( $-4.20 \pm 563.80$  N or  $-0.25 \pm 33.99\%$ ), ‘Veras sacrum y’ ( $-74.18 \pm 492.08$  N or  $-4.47 \pm 29.67\%$ ), ‘Wundersitz 20 Hz’ ( $+34.19 \pm 1089.85$  N or  $+2.06 \pm 65.71\%$ ), ‘Pogson’ ( $+25.36 \pm 745.01$  N or  $+1.53 \pm 44.92\%$ ), and ‘Pogson xynorm’ ( $-2.39 \pm 730.18$  N or  $-0.14 \pm 44.02\%$ ).




**Figure 4.9:** Second peak estimation bias (white bar),  $\pm$  RC (dark blue), and  $\pm$  LOA (light blue) for each capable method. Gray dots represent trials falling outside the LOA. Values outside  $\pm 1000$  N are plotted at the axis limits. A value of 0 represents

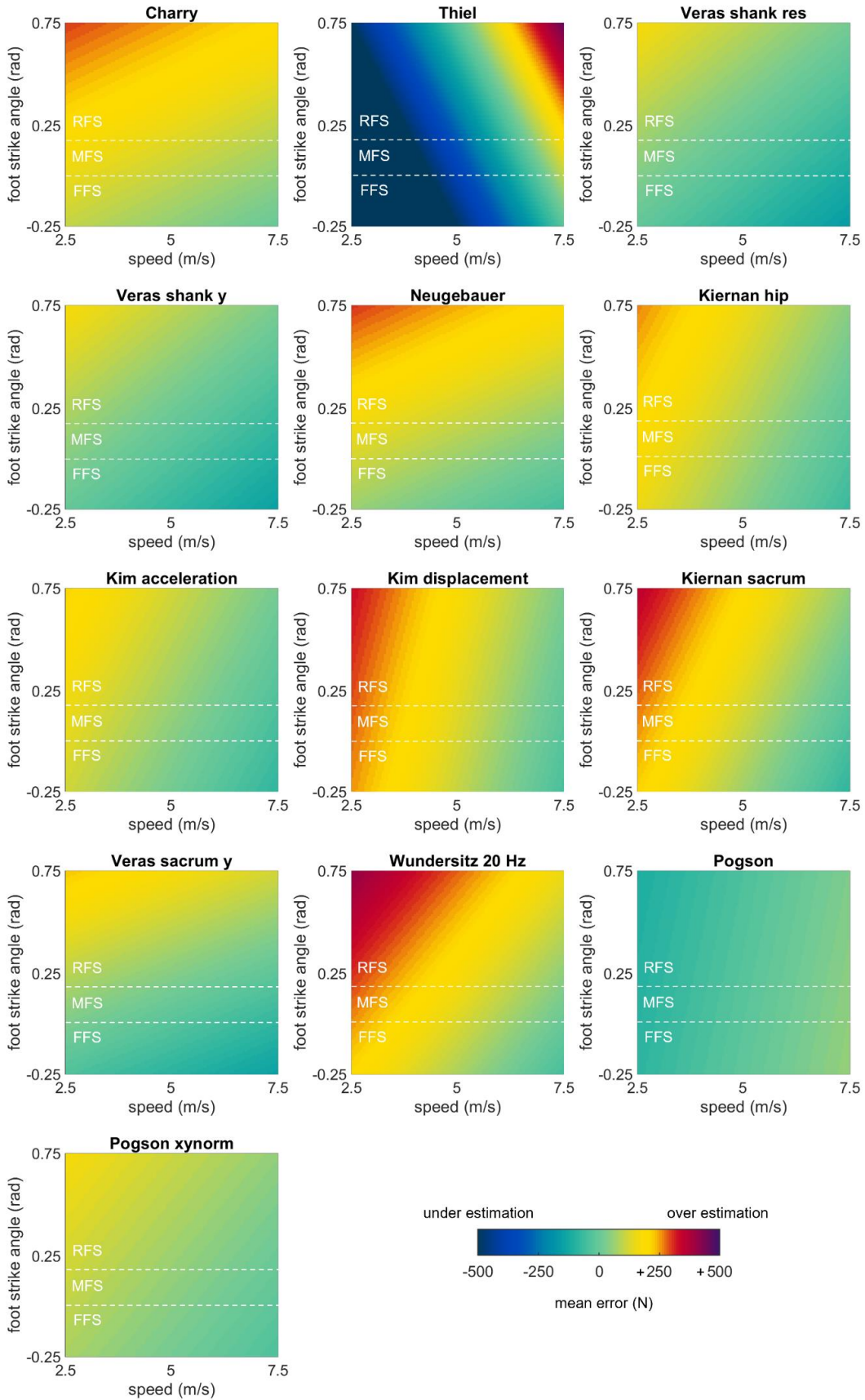
perfect agreement with the force plate. Positive values indicate the method overestimated the second peak. Negative values indicate the method underestimated the second peak. Methods with a white background on the left are for accelerometers on the shank, methods with a gray background are for the hip, and methods with a white background on the right are for the sacrum.

The linear mixed effects models examining the role of running speed, surface, and foot strike angle on the best performing methods revealed that error in all methods was significantly explained by running speed ( $p \leq 0.05$ ). Foot strike angle also significantly explained error in eight of 13 methods ( $p \leq 0.05$ ) but not in ‘Kim acceleration,’ ‘Kim displacement,’ ‘Wundersitz 20 Hz,’ ‘Pogson,’ and ‘Pogson xynorm’ ( $p > 0.05$ ). Surface explained significant variation in only one method: ‘Wundersitz 20 Hz’ (all other  $p > 0.05$ ) (Table 4.5). Most methods showed the same general pattern with overestimates at low speeds and high foot strike angles. There were two exceptions: ‘Thiel’ had underestimates at low speeds and foot strike angles and overestimates at high speeds and foot strike angles, while ‘Pogson’ showed generally stable performance with a small (but significant) increase in error at faster speeds (Figure 4.10).

**Table 4.5: Best performing second peak estimation methods** (bias < 100 N). Overall performance shown as biases (accuracy), RCs (repeatability), and LOAs (precision) color-coded from the absolute minimum (green) to absolute maximum (purple) values observed within-column. Biases that significantly differed from 0 ( $p \leq 0.05$  with FDR correction) marked \*. Performance across conditions shown as coefficients for the intercept of surface (added to model estimated error for the track condition but not the floor condition) and slopes for running speed (in m/s) and foot strike angle (in rad). If surface, speed, or foot strike explains a significant ( $p \leq 0.05$ ) amount of error it is highlighted pink and marked \*.

Method	Overall performance			Performance across conditions		
	Bias (N)	RC (N)	LOAs (N)	Speed	Surface	Foot strike
Charry	58.48*	79.70	478.42	-20.20*	4.46	200.08*
Thiel	-90.32*	1006.17	1162.97	192.97*	17.93	484.50*
Veras shank res	-98.67*	0.53	470.96	-30.82*	3.04	182.12*
Veras shank y	-88.75*	10.07	471.45	-29.67*	3.88	178.48*
Neugebauer	14.04	116.89	488.09	-23.53*	1.69	244.81*
Kiernan hip	-7.56	157.16	572.04	-44.59*	9.34	113.21*
Kim acceleration	-27.21*	74.26	721.47	-39.70*	0.25	102.31
Kim displacement	20.68	250.96	696.86	-61.10*	1.81	75.90
Kiernan sacrum	-4.20	156.73	563.80	-60.27*	8.40	138.82*
Veras sacrum y	-74.18*	132.12	492.08	-22.25*	4.19	268.09*
Wundersitz 20 Hz	34.19	729.30	1089.85	-51.84*	63.98*	205.90
Pogson	25.36	233.50	745.01	28.92*	2.74	-27.38
Pogson xynorm	-2.39	429.45	730.18	-23.62*	18.74	93.89

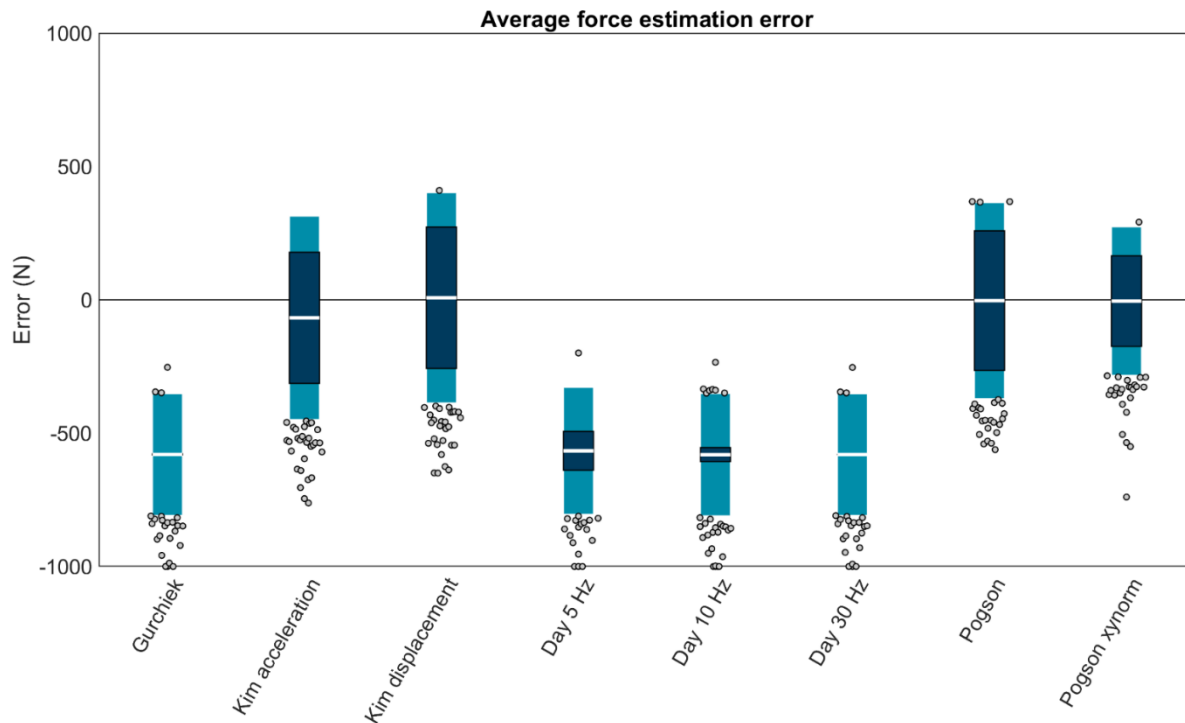
column |min|  column |max|



**Figure 4.10: Bias in second peak estimates predicted by mixed effects models for each of the best-performing methods,** plotted as a function of speed and foot strike angle. Green values represent perfect agreement with the gold standard; yellow, red, and purple values represent positive biases (overestimates); darker blue values represent negative biases (underestimates). Foot strike angles corresponding to rear-, mid-, and fore-foot strike patterns have been labelled (RFS, MFS, and FFS) and divided with dashed white lines.

#### 4.3.4. Average force

Eight sacrum methods could calculate average force from an estimated time series, the four best performing (biases < 100 N) were: ‘Kim acceleration’ ( $-67.76 \pm 381.52$  N or  $-6.87 \pm 38.68\%$ ; bias  $\pm$  LOA), ‘Kim displacement’ ( $+7.66 \pm 394.37$  N or  $+0.78 \pm 39.98\%$ ), ‘Pogson’ ( $-3.18 \pm 367.81$  N or  $-0.32 \pm 37.29\%$ ), and ‘Pogson xynorm’ ( $-4.87 \pm 278.80$  N or  $-0.49 \pm 28.26\%$ ).




**Figure 4.11: Average force estimation** bias (white bar),  $\pm$  RC (dark blue), and  $\pm$  LOA (light blue) for each capable method. Gray dots represent trials falling outside the LOA. Values outside  $\pm 1000$  N are plotted at the axis limits. A value of 0 represents perfect agreement with the force plate. Positive values indicate the method overestimated the average force. Negative values indicate the method underestimated the average force. All methods were for the sacrum.

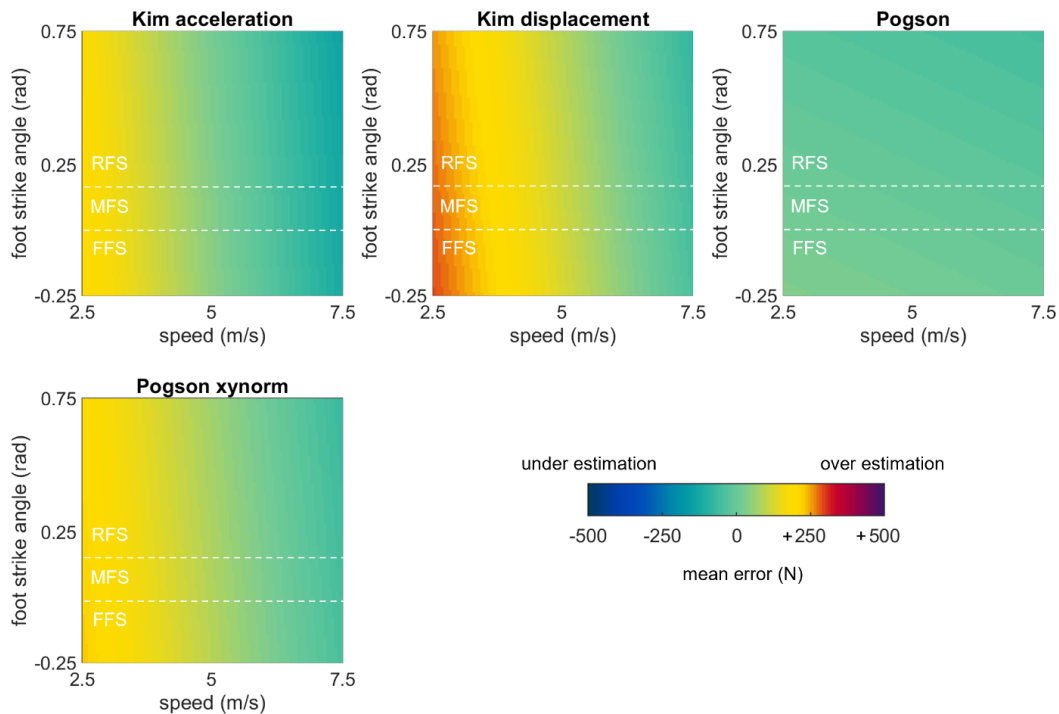
The linear mixed effects models examining the role of running speed, surface, and foot strike angle on each of the best performing methods revealed that error in all methods except ‘Pogson’ was significantly explained by running speed ( $p \leq 0.05$ ). Foot strike angle significantly explained error in only

'Pogson' (all other  $ps > 0.05$ ). Surface significantly explained error in only 'Pogson xynorm' (all other  $ps > 0.05$ ) (Table 4.6). Three methods showed the same general pattern with overestimates at low speeds. 'Pogson' was an exception showing generally stable performance with a small (but significant) increase in error at lower foot strike angles (Figure 4.12).

**Table 4.6: Best performing average force estimation methods** (biases < 100 N). Overall performance shown as biases (accuracy), RCs (repeatability), and LOAs (precision) color-coded from the absolute minimum (green) to absolute maximum (purple) values observed within-column. Biases that significantly differed from 0 ( $p \leq 0.05$  with FDR correction) marked \*. Performance across conditions shown as coefficients for the intercept of surface (added to model estimated error for the track condition but not the floor condition) and slopes for running speed (in m/s) and foot strike angle (in rad). If surface, speed, or foot strike explains a significant ( $p \leq 0.05$ ) amount of error it is highlighted pink and marked \*.

Method	Overall performance			Performance across conditions		
	Bias (N)	RC (N)	LOAs (N)	Speed	Surface	Foot strike
Kim acceleration	-67.76*	245.71	381.52	-66.45*	3.96	-24.06
Kim displacement	7.66	265.11	394.37	-67.73*	5.20	-36.48
Pogson	-3.18	261.96	367.81	-5.48	3.68	-59.42*
Pogson xynorm	-4.87	169.92	278.80	-54.79*	14.12*	-24.00

column |min|  column |max|

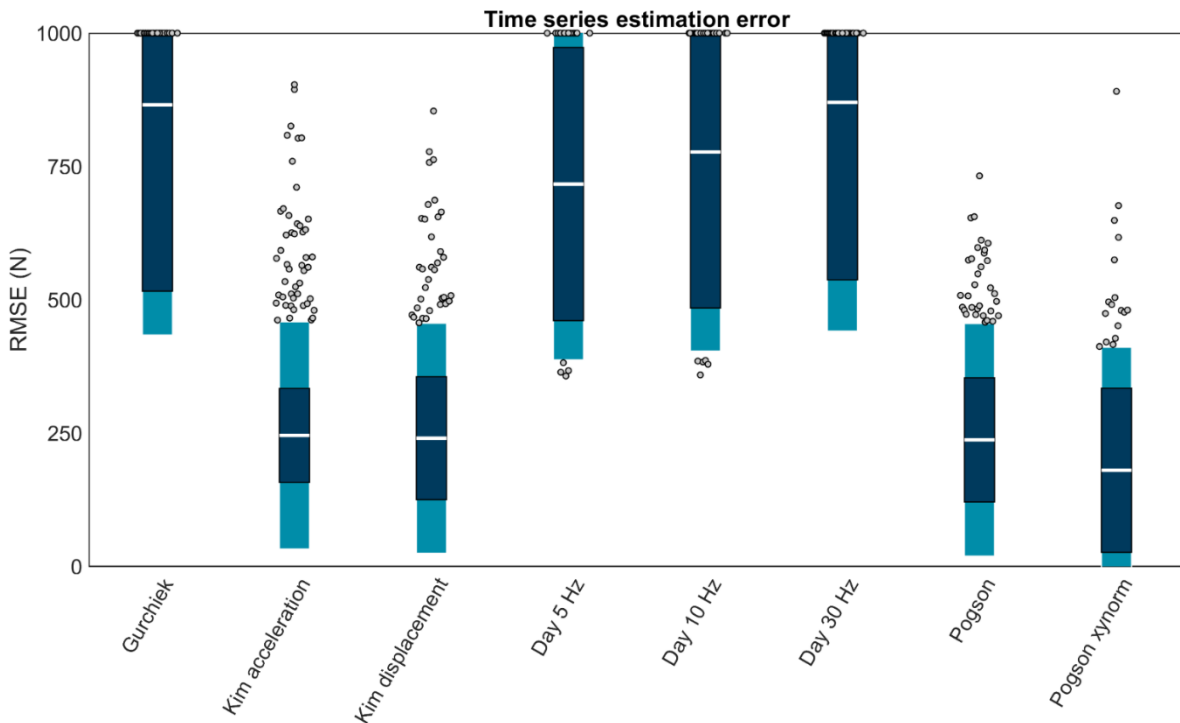


**Figure 4.12: Bias in average force estimates predicted by mixed effects models for each of the best-performing methods,** plotted as a function of speed and foot strike angle. Green values represent perfect agreement with the gold standard; yellow, red, and purple values represent positive biases (overestimates); darker blue values represent negative biases

(underestimates). Foot strike angles corresponding to rear-, mid-, and fore-foot strike patterns have been labelled (RFS, MFS, and FFS) and divided with dashed white lines.

#### 4.3.5. Time series

Eight sacrum methods could estimate force time series, of these the four best performing (RMSEs < 250 N) were: ‘Kim acceleration’ ( $245.69 \pm 212.85$  N or  $24.91 \pm 21.58\%$ ; mean  $\pm$  LOA), ‘Kim displacement’ ( $240.43 \pm 215.57$  N or  $24.37 \pm 21.85\%$ ), ‘Pogson’ ( $237.33 \pm 218.09$  N or  $+24.06 \pm 22.11\%$ ), and ‘Pogson xynorm’ ( $180.32 \pm 230.62$  N or  $18.28 \pm 23.38\%$ ).



**Figure 4.13: Time series estimation** mean RMSE (white bar),  $\pm$  RC (dark blue), and  $\pm$  LOA (light blue) for each capable method. Gray dots represent trials falling outside the LOA. Values above 1000 N are plotted at the axis limit. A value of 0 represents perfect agreement with the force plate. Positive values indicate larger errors. All methods were for the sacrum.

The linear mixed effects models examining the role of running speed, surface, and foot strike angle on each of the best performing methods (RMSEs < 250 N) revealed that error in all methods was significantly explained by running speed ( $p_s \leq 0.05$ ) but not foot strike angle or running surface ( $p_s > 0.05$ ) (Table 4.7). All methods showed the same general pattern with higher predicted RMSEs at higher

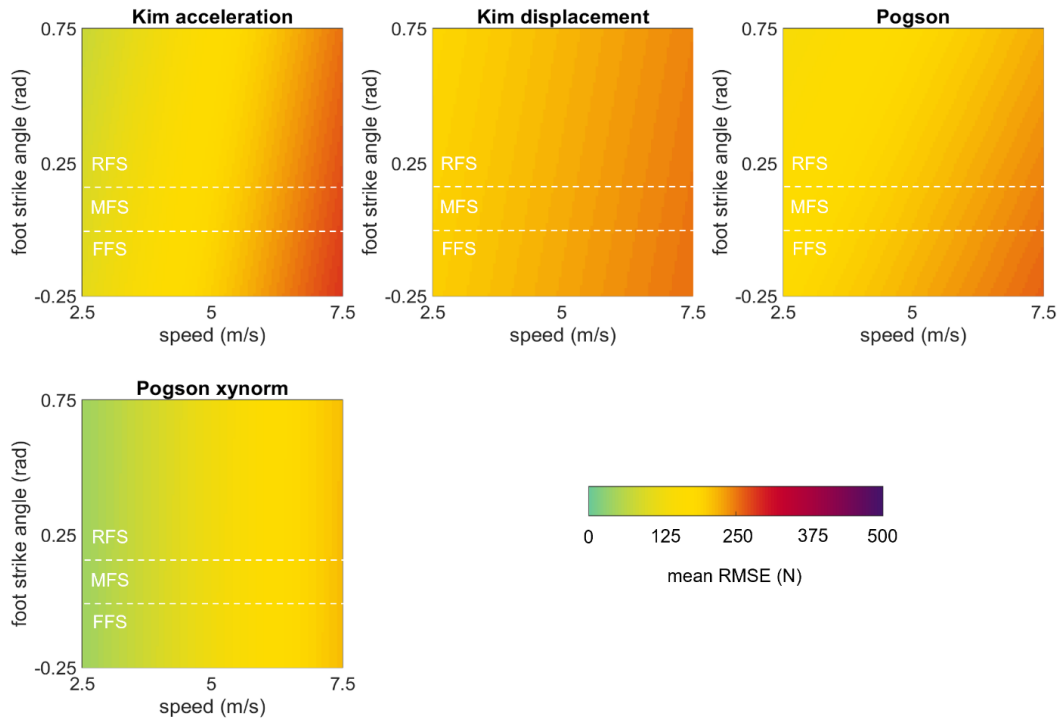


speeds but ‘Pogson xynorm’ had lower predicted mean RMSEs than any other method across its entire range (Figure 4.14).

**Table 4.7: Best performing force time series estimation methods** (RMSE < 250 N). Overall performance shown as biases (accuracy), RCs (repeatability), and LOAs (precision) color-coded from the absolute minimum (green) to absolute maximum (purple) values observed within-column. RMSEs that significantly differed from 0 ( $p \leq 0.05$  with FDR correction) marked \*. Performance across conditions shown as coefficients for the intercept of surface (added to model estimated error for the track condition but not the floor condition) and slopes for running speed (in m/s) and foot strike angle (in rad). If surface, speed, or foot strike explains a significant ( $p \leq 0.05$ ) amount of error it is highlighted pink and marked \*.

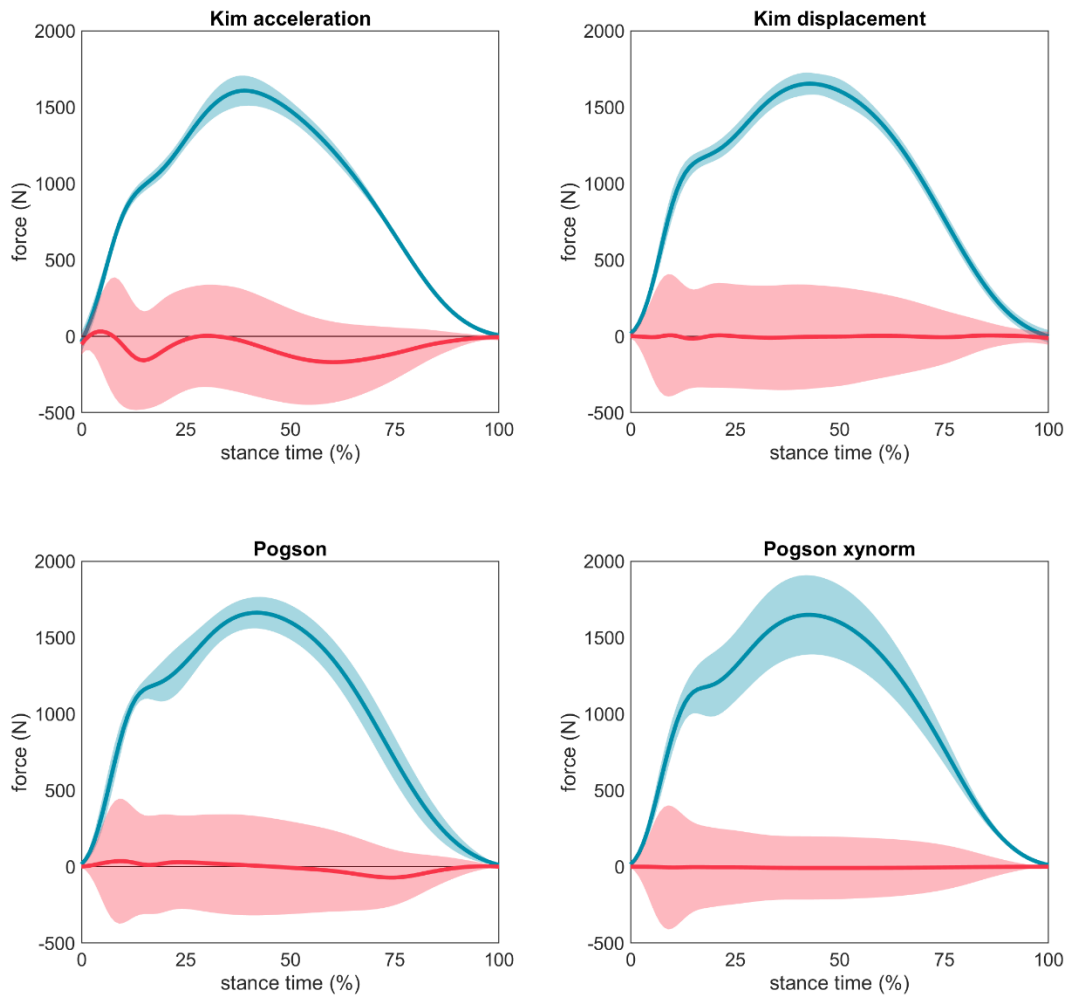
Method	Overall performance			Performance across conditions		
	RMSE (N)	RC (N)	LOAs (N)	Speed	Surface	Foot strike
Kim acceleration	245.69*	88.18	212.85	36.30*	-3.45	-32.21
Kim displacement	240.43*	115.29	215.57	11.64*	4.84	-10.52
Pogson	237.33*	116.38	218.09	16.84*	-4.09	-40.05
Pogson xynorm	180.32*	153.83	230.62	34.94*	0.31	-0.85

column |min|  column |max|



**Figure 4.14: Mean RMSE in estimated force time series predicted by mixed effects models for each of the best-performing methods**, plotted as a function of speed and foot strike angle. Green values represent perfect agreement with the gold standard; yellow, red, and purple values represent higher predicted mean RMSEs. Foot strike angles corresponding to rear-, mid-, and fore-foot strike patterns have been labelled (RFS, MFS, and FFS) and divided with dashed white lines.

The 'Kim acceleration' and 'Kim displacement' methods produced highly stereotyped time series with little variation (i.e., no matter the input acceleration the output force estimate was similar) (Figure 4.15). This caused relatively large errors, particularly at the locations of the first and second peak. In contrast, 'Pogson xynorm' had much more variation in its estimated values and better fit the data (lower errors), however, errors remained high around the first peak.



**Figure 4.15: Mean time series estimated by each method (dark blue line)  $\pm 1$  SD (light blue shading) and the error between the gold standard time series (dark red line)  $\pm 1$  SD (light red shading). A red line at 0 indicates perfect agreement with the force plate, values above 0 indicate an overestimation of force at that time point, values below 0 indicate an underestimation of force at that time point.**

#### 4.4. Discussion

Performance was evaluated for 27 methods of estimating vertical GRF during running from a single wearable accelerometer on the shank or approximate COM. For each method, forces were estimated from 74 runners across two different surfaces (wood floor, running track), three self-selected speeds (slowest, typical, fastest), and a range of foot strike angles (including fore-, mid-, and rear-foot strike patterns). Errors were quantified as the difference between estimated and ground truth forces.

Based on the observed errors, we recommend the 'Pogson' or 'Pogson xynorm' methods for several reasons: First, these methods use a single accelerometer on the sacrum to estimate bilateral forces. This is advantageous over shank and hip methods that either (a) do not allow estimation of bilateral forces, (b) require an assumption of bilateral symmetry, or (c) require two accelerometers. Second, in contrast to most other methods, these methods can estimate every feature of the vertical GRF investigated here (first peak, loading rate, second peak, average, and time series). Third, these methods had relatively stable performance across speeds, foot strike angles, and running surfaces. Fourth, these methods were consistently high performers (had low biases, RCs, and LOAs) for second peak, average, and time series estimation. Potential users should, however, balance their own design needs when choosing a method. For example, if estimating second peak under known speed and foot strike conditions, other methods may have similar accuracy but better reliability and precision (e.g., the 'Neugebauer,' 'Kim,' or 'Kiernan' methods).

For first peak and loading rate estimation, the 'Pogson' methods were outperformed by other methods (e.g., 'Higgins,' 'Kiernan,' or 'Veras'). However, this does not affect our overall recommendation due to the high LOAs observed for first peak and loading rate estimation: Even the best performing methods had LOAs exceeding  $\pm 67.60$  and  $\pm 93.19\%$  of first peak and loading rate target magnitudes. These LOAs are likely larger than any potential between-group effects or within-participant

changes, suggesting that first peak and loading rate cannot be estimated with sufficient precision with any of the methods investigated here. Given these results, we do not currently recommend the estimation of first peak or loading rate from accelerometers.

Despite the poor results for first peak and loading rate estimation, observed biases for all estimated vertical GRF features were at or below those originally reported for nine of the 13 publications from which we derived the methods [246, 265, 86, 87, 264, 240, 255, 266, 268]. Some of the error we did observe was attributable to speed, surface, and foot strike angle. Thus, one approach to decrease error may be to include these explanatory variables as model inputs. For example, Alcantara et al. [208] added speed, slope of the surface, and foot strike pattern to their force prediction model and reported RMSEs of 106.78 N (or 6.4%) for time series estimation. These results are an improvement over the methods recommended here (although direct comparison is difficult as they included prediction of zero forces during swing in their RMSE calculations, which may have reduced their errors relative to our analysis of stance phase only, when forces are non-zero) (see also: [259, 248, 266]). Thus, including speed, foot strike angle, and surface may improve the performance of future methods provided these variables can be quantified precisely and accurately in the field (e.g., [182, 183, 184]).

The inclusion of other explanatory variables may also improve performance. Thirteen of the 27 methods investigated here estimated force using linear regressions that included anthropometrics (mass, height, and/or leg length), sex, and/or age as explanatory variables [86, 87, 264, 240, 266, 232, 255]. Although these methods did not receive our final recommendation, many had very reasonable results. These results are particularly salient in contrast to the poor performance of methods that estimated force by simply multiplying acceleration by mass without additional explanatory variables [246, 265, 247, 269]. This performance difference suggests that including explanatory variables improved performance and may be warranted in future models.

Reductions in error could also be achieved by fitting participant-specific models. For example, Kiernan et al. [266] reported that including a random effect of participant reduced error in their model by ~40% (see also [255, 232, 86]). Thus, participant-specific models likely offer more accuracy, reliability, and precision than participant-general models and should be used when time and resources allow. Participant-specific models were not examined here to ensure that the methods studied were broadly applicable and accessible. This design choice maximizes the in-field utility of the models studied and allows them to be applied to novel participants without the necessity of taking gold-standard measurements in-lab and using computational resources to develop a model for each participant.

Conversely, the choice to use acceleration data that were time synchronized to target force data with ms accuracy stands in contrast to the goal of applying these methods in-field. This choice to made to ensure that errors in accelerometer-based stance identification did not affect results and lead to erroneous conclusions. When taking measures outside the lab, however, the use of time synchronized force plate data is not available. Currently, even the best methods to identify stance have errors of  $-30.4 \pm 118.8$  and  $-2.8 \pm 149.9$  ms (bias  $\pm$  LOA) for initial and terminal contact, respectively [207]. These errors may interfere with the accurate segmentation of acceleration data during stance and thus, until more exact acceleration-based stance identification is possible, acceleration signals will vary in duration relative to their target force signals. It is likely that the regression methods that used peak acceleration values to estimate discrete force variables are more robust to these discrepancies in duration [86, 87, 264, 240, 266, 232, 255]. The peaks used by these methods tend to be greater than acceleration values immediately pre- or proceeding initial and terminal contact and tend to be ~mid-stance so are unlikely to be removed accidentally (see Chapter 4S.1). In contrast, methods attempting to estimate continuous force time series are likely vulnerable to variability in the input data duration.

To explore this, errors for the recommended 'Pogson xynorm' method were recalculated using acceleration input data that were stance-segmented based on acceleration signal features instead of a

force plate threshold (using the 'Auvinet' stance identification method [162] as implemented by Kiernan et al. [207]). In contrast to the expectation that errors would increase, they were comparable to those observed when segmenting based on the time synchronized force signal (first peak:  $-104.75 \pm 876.71$  N; loading rate:  $-4.71 \pm 53.31$  kN/s; second peak:  $-5.03 \pm 683.64$  N; average:  $-5.83 \pm 341.02$  N; time series:  $178.25 \pm 207.58$  N;). Model performance may have been maintained due to providing a more stereotyped input (i.e., acceleration data segmented based on acceleration features) and/or due to providing a larger set of training data (since ms accuracy was not required all 4440 trials were used vs. the 642 exactly synchronized trials available for the other methods). In any case, the 'Pogson\_Auvinet' method's performance demonstrates the promise of current methods to estimate vertical GRF second peak, average, and time series in the field.

Before applying these force estimation methods in the field, potential users should consider the conditions and participants used to develop and validate the method. It should not be assumed that these methods will work under other conditions or for other participants. For example, only over-ground running on two level surfaces was quantified here, so it should not be assumed that results would hold for incline/decline running, treadmill running, or running on other surfaces (e.g., sand, grass/turf, asphalt, concrete). That said, consistent with previous findings that changes in surface do not affect vertical GRF [275], results demonstrated that surface rarely explained error, suggesting that estimated forces are robust to changes in surface. Potential users should also consider that the current sample represents a relatively homogenous group of runners (Figure 4.1). Thus, if studying participants drawn from different populations, these results may not be representative, and the included code/models may not produce estimations with the accuracy, reliability, and precision reported here.

Careful consideration should also be given to any differences in acceleration processing and/or coordinate conventions as any differences in acceleration inputs could affect force outputs. For example, the 'Wundersitz,' 'Meyer,' 'Gurchiek,' and 'Day' methods [246, 265, 247, 269] all used the

same general approach of multiplying acceleration by mass to estimate force. Despite this common approach, results differed across these methods due to differences in acceleration processing. Expressing accelerations in different coordinate systems will also change the input and affect the output. All but two of the publications the methods were derived from used a WCS (Table 4.1). These coordinate systems assume alignment with segments or the inertial vector and may be particularly prone to altering acceleration inputs. For example, an accelerometer placed on the sacrum could deviate from assumed inertial alignment due to lumbosacral curvature and adiposity or a participant leaning forward during running [276, 277]. Any discrepancies in the placement of an accelerometer could also change the data. More consistent data may be obtained with SCS and TCCS. Thus, all analyses in this paper are presented in the SCS. Although not included here, analyses were repeated in the WCS and TCCS. In contrast to the *a priori* expectation that SCS and TCCS would outperform the WCS, a systematic effect of coordinate system was only found in methods that multiplied acceleration by mass with no other coefficients [246, 265, 247, 269]. When unique model coefficients, weights, and/or biases were calculated for each coordinate system there were no systematic differences between them (Chapter 4S.2). This result should, however, be interpreted with caution: Each of the IMUs in this study was placed by the same experimenter, their positions were monitored throughout data collection, and any movement of the IMU led to elimination of a participant ( $n = 2$ ). Thus, the WCS in this study is likely more consistent than under field conditions where wearables may be placed across many repeated data collections by individuals with little training, leading to misalignment and inconsistency. Thus, we still caution against the use of the WCS.

#### 4.5. Conclusions & practical applications

We recommend the ‘Pogson’ methods to estimate second peak, average, and time series. We do not currently recommend the estimation of first peak or loading rate due to the large observed limits of agreement for these variables. For each method, code to automatically process stance-segmented accelerometer data is available at [https://github.com/DovinKiernan/MTFBWY\\_running\\_vGRF\\_from\\_a](https://github.com/DovinKiernan/MTFBWY_running_vGRF_from_a). This code should be applied with careful consideration of the sample it was developed and validated on and the data processing that was used on the acceleration inputs used to train the models. Future research should investigate whether the inclusion of anthropometrics, sex, age, field measures of speed and foot strike angle, or other explanatory variables can improve model performance. The results reported here should be used as a benchmark for the performance of future models and details on accuracy, reliability, and precision should be reported.

#### 4.6. Funding

This research was funded in part by the American College of Sports Medicine Foundation Doctoral Student Research Grant (18-00671) and the Sigma Xi Scientific Research Honor Society Grant-in-Aid of Research (G2018031595024439). The APC was funded by the University of California Davis Library’s Open Access Fund and by *Sensors*. DK was supported by a scholarship from the Natural Sciences and Engineering Research Council of Canada (CGSD3-502358-2017), the Maury L Hull Endowed Fellowship for Musculoskeletal Biomechanics Research, and a University of California Davis Professors for the Future Fellowship.



## 4S: Supplement to Chapter 4

### 4S.1. Description of force estimation methods and their implementation

Here we provide details on each of the force estimation methods included in our study. Brief descriptions, equations, and figures have been included to show the force estimation process for each method. For additional detail please refer to the original publications.

All data for the figures in this section was taken from the same randomly selected trial (a 75.3 kg male running 3.20 m/s on the ‘floor’ surface with a rear foot strike angle of 0.44 rad). For clarity, the different coordinate systems used across the original publications have been standardized to the conventions described briefly in the main text of Chapter 4 and in detail in Section S3. All example data are shown in the SCS regardless of the original coordinate system used by the method, but Section 4.2. demonstrates the effects of different coordinate systems on the data.

Finally, our implementation of each of these methods is freely available at:

[https://github.com/DovinKiernan/MTFBWY\\_running\\_vGRF\\_from\\_a](https://github.com/DovinKiernan/MTFBWY_running_vGRF_from_a)

To use these implementations simply feed acceleration data during a single stance into the provided function as an  $f \times 4$  matrix where  $f$  is frame and the columns correspond to time (in ms) and accelerations in the  $x$ ,  $y$ , and  $z$  axes (in  $g$ ).

#### 4S.1.1. Neugebauer method

Neugebauer and colleagues published two papers using linear regression to estimate maximum vertical GRF from accelerometers worn on the hips of children and adults [86, 87]. To do so, they used the generalized (non-participant-specific) equation:

**Equation 4S.1**

$$\ln(F_{y,max}) = \alpha_0 + \alpha_1 a_{WCS,y,max} + \alpha_2 m + \alpha_3 L + \alpha_4 L a_{WCS,y,max}$$

Where  $m$  corresponds to participant mass and  $L$  corresponds to the type of locomotion (with walking equal to 0 and running equal to 1). Thus, for running, their equation simplifies to:

**Equation 4S.2**

$$\ln(F_{y,max}) = (\alpha_0 + \alpha_3) + (\alpha_1 + \alpha_4) a_{WCS,y,max} + \alpha_2 m$$

**Equation 4S.3**

$$\ln(F_{y,max}) = c_1 + c_2 a_{WCS,y,max} + c_3 m$$

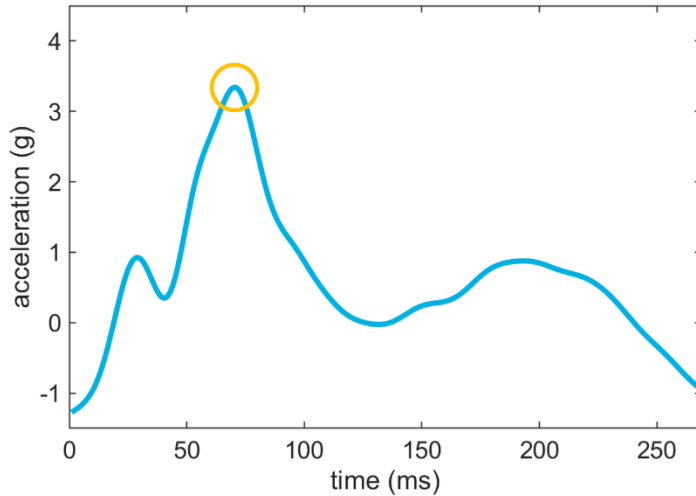
**Equation 4S.4**

$$F_{y,max} = e^{c_1 + c_2 a_{WCS,y,max} + c_3 m}$$

Where  $c_1$ ,  $c_2$ , and  $c_3$  are constants replacing  $(\alpha_0 + \alpha_3)$ ,  $(\alpha_1 + \alpha_4)$ , and  $\alpha_2$ , respectively. Note, that Neugebauer and colleagues published a third paper estimating maximum vertical GRF from the accelerations of soldiers carrying a load during walking that can also be reduced to Equation 4S.4 if carrying a load of 0 [278].

Here, we made two assumptions: (1) that Neugebauer defined their WCS with ~vertical as positive, and (2) that  $F_{y,max}$  is analogous to  $F_{y,second}$ . We then used a leave-one-out cross-validation to iteratively calculate the error for one participant using  $c_1$ ,  $c_2$ , and  $c_3$  values calculated from the other 73 participants while the final  $c_1$ ,  $c_2$ , and  $c_3$  values were calculated using data from all 74 of our participants. Also note, that in their 2012 and 2014 papers, Neugebauer and colleagues were unable to synchronize their accelerometers and force plates and thus used  $a_{WCS,y,max}$  averaged across 30 s and

10 s, respectively, for those papers. In contrast, we use  $a_{WCS,y,max}$  during the single stance for which  $F_{y,second}$  is being estimated.



**Figure 4S.1: Neugebauer method.** Hip y-axis acceleration (proximal-distal in the SCS; ~ longitudinal in the WCS; blue line) maxima (yellow circle) and participant mass are entered into Equation S12 to estimate vertical GRF second peak magnitude.

#### 4S.1.2. Charry method

Charry et al. [279] placed a  $\pm 24$  g accelerometer on the medial midshaft of three participants tibiae and measured  $a_{WCS,y}$  at 100 Hz while the participants ran 1.7 to 7.2 m/s overground. Charry et al. evaluated the efficacy of four potential acceleration variables to predict vertical GRF second peak magnitudes: (1) heel-strike, (2) initial peak acceleration, (3) minimum peak (specified as maximum in their paper but minimum here based on differences in coordinate conventions), and (4) peak-to-peak (Fig. S10A). They found that the minimum peak acceleration was the best predictor of vertical GRF second peak magnitude and thus discarded the other three predictor variables. Charry et al. evaluated both a linear and logarithmic prediction equation and found that the logarithmic equation better predicted vertical GRF second peak magnitude; thus, they discarded the linear fit and used:

**Equation 4S.5**

$$F_{y,second} = \log_2(-a_{WCS,y,min} + 1)$$

Where 1 was added to the acceleration values to prevent taking the log of a negative value. Charry et al. then added additional terms to their prediction equation to create slopes and intercepts that were a function of participant mass, using the form:

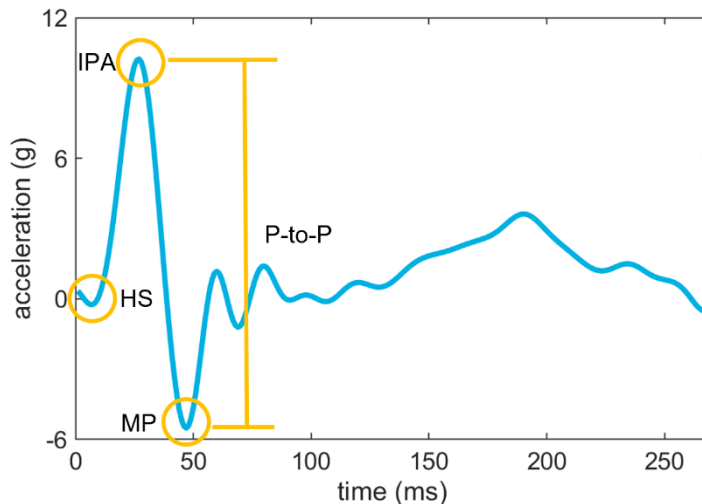
**Equation 4S.6**

$$F_{y,second} = s(m) * \log_2(-a_{WCS,y,min} + 1) + i(m)$$

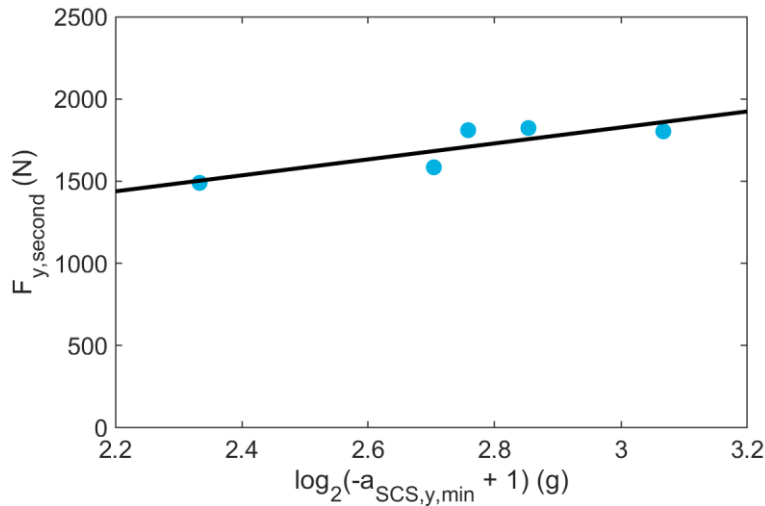
Where  $s(m)$  and  $i(m)$  were found by taking the participant- and leg-specific slopes and intercepts (respectively) from Equation 4S.5 (Figure 4S.2B) and plotting them against participant mass (Figure 4S.2C), then defining  $s(m)$  and  $i(m)$  as the equation for that line. Thus, Equation 4S.6 can be expanded to the form:

$$F_{y,second} = (s_1m + s_2) * \log_2(-a_{WCS,y,min} + 1) + (i_1m + i_2)$$

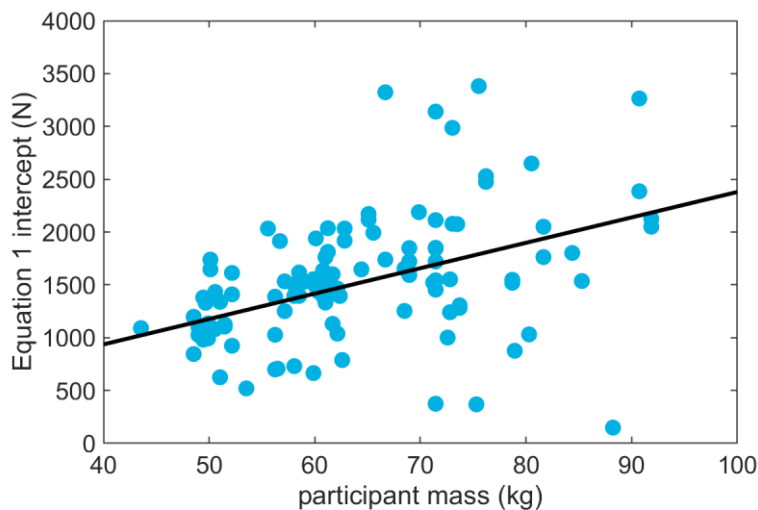
The original  $s(m)$  and  $i(m)$  were found on the three participants in Charry et al.'s study. Here, we used a leave-one-out cross-validation to iteratively calculate the error for one participant using  $s(m)$  and  $i(m)$  terms calculated from the other 73 participants while the final  $s(m)$  and  $i(m)$  terms were calculated using data from all 74 of our participants.



**Figure 4S.2: Charry method. (A)** Tibia y-axis acceleration (proximal-distal in the SCS; ~longitudinal in the WCS; blue line) features Charry et al. used to predict vertical GRF second peak magnitude. HS = Heel Strike, IPA = Initial Peak Acceleration, MP = minimum peak (originally specified as a maximum but here as a minimum due to differences in coordinate conventions), and P-to-P = peak-to-peak. 'Minimum peak' was found to be the best predictor and all other variables were discarded.



**(B)** Fit for Charry et al.'s Equation 1 for a single leg from the example participant. Each blue point represents one stance. To account for differences in coordinate conventions, the negative value of the minimum peak was used. Although 1 was added to all accelerations, 0.2% of Equation 1 results when using the SCS were still imaginary numbers. These results were discarded from analysis.



**(C)** Participant- and leg-specific intercepts from the line of best fit in the preceding figure plotted against participant mass for each participant and leg (blue points). The equation describing the line of best fit for the black line was used to define the  $i(m)$  term in Charry et al.'s Equations S14 and S15. A similar procedure was conducted using the slopes from the previous step (Figure S9B) to define the  $s(m)$  term.

#### 4S.1.3. Wundersitz method

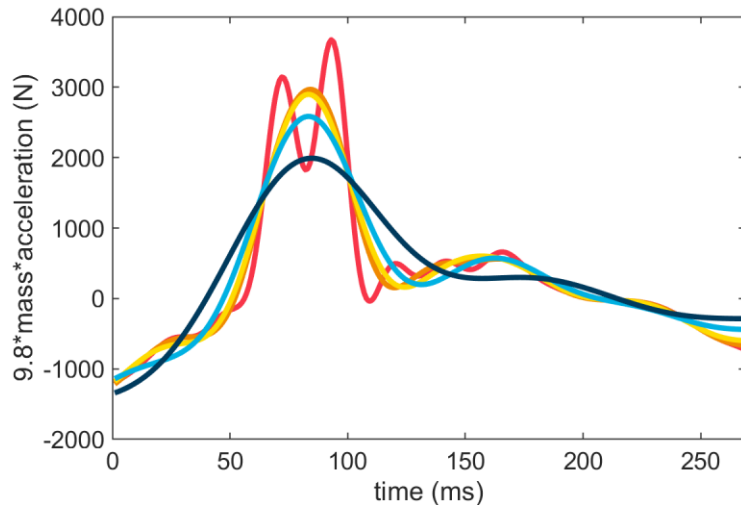
Wundersitz et al. [280] measured vertical GRF and second thoracic vertebra acceleration from 17 competitive team sport athletes while they ran either in a straight line or with a change of direction. To estimate force, they multiplied acceleration by mass. They then filtered the estimated force with a 4<sup>th</sup> order low-pass Butterworth filter of either 10, 15, 20, or 25 Hz and extracted the peak value. We assumed that the peak force they were estimating corresponds to the vertical GRF second peak such that:

**Equation 4S.7**

$$F_{y,second} = 9.8ma_{WCS,y,max}$$

Where  $m$  is mass in kg and  $a_{WCS,y,t}$  is the acceleration in g (with  $1\text{ g} = 9.8\text{ m/s}^2$ ).

Here, we adapted this upper back approach to our sacrum data.



**Figure 4S.3: Wundersitz method.** Sacrum y-axis acceleration (proximal-distal in the SCS; ~longitudinal in the WCS) was multiplied by participant mass then low-pass filtered at 10, 15, 20, 25-Hz, and no filter (dark blue, light blue, yellow, orange, and red lines, respectively). Maxima were then identified and used to estimate vertical GRF second peak.

**4S.1.4. Meyer method**

Meyer et al. [281] had 13 moderately active children perform a range of tasks including jogging and running while they wore triaxial accelerometers on their right hip. Meyer et al. did not explicitly define their coordinate convention but report extracting the “minimum acceleration of the vertical axis.” We assumed that this corresponds to a maximum in the ~longitudinal axis as defined in our WCS (see Figure 4S.1 above). Acceleration and force signals were not time synchronized; thus, Meyer et al. extracted and averaged their acceleration minima across 8-15 steps per trial. The correlation between these average accelerations and “peak impact forces in the vertical plane” were then explored (which we interpreted as maximum vertical GRF and *not* vertical GRF first peak magnitude based on their reported magnitudes and on the fact that their other tasks would not have an “impact peak” as the term

is commonly used in the running literature). Meyer et al. also explored sex, age, weight, height, and leg length as potential explanatory variables but found they were not significant predictors. Thus, they ultimately used a correlation between the mean  $a_{WCS,y,max}$  across 8-15 steps and the  $F_{y,max}$  of a single stance on the force plate (expressed in body weights), observing an  $r^2$  of 0.81. This can be expressed as:

**Equation 4S.8**

$$\frac{F_{y,max}}{9.8m} = a_{WCS,y,max}$$

And rearranged as:

**Equation 4S.9**

$$F_{y,second} = 9.8ma_{WCS,y,max}$$

Where  $m$  is mass in kg and  $a_{WCS,y,max}$  is the acceleration in g (with  $1\text{ g} = 9.8\text{ m/s}^2$ ) and with the assumption that the  $F_{y,max}$  they attempted to estimate corresponded to  $F_{y,second}$  at the running speeds they studied. This equation resembles the one used by Wundersitz et al. [280] (above) but with different data entered into the equation.

We used Equation 4S.9 to estimate  $F_{y,second}$  across our 74 participants using the  $a_{WCS,y,max}$  from the same stance the force was taken from.

#### 4S.1.5. Gurchiek method

Gurchiek et al. [247] had 15 participants perform sprint start and change of direction tasks while wearing an IMU on their sacrum. Data were low pass filtered at 30 Hz then re-expressed in a GCS. The force time series and average across stance were estimated by scaling acceleration by mass:

**Equation 4S.10**

$$F_{y,t} = 9.8ma_{GCS,y,t}$$

Where  $m$  is mass in kg and  $a_{GCS,y,t}$  is the acceleration in g (with  $1\text{ g} = 9.8\text{ m/s}^2$ ). This equation resembles the one used by Wundersitz et al. [280] and Meyer et al. [281] (above) but with different data entered into the equation.

#### 4S.1.6. Thiel method

Thiel et al. [240] placed IMUs above the medial malleoli of three elite sprinters and had them sprint 50 m on a track with embedded force plates. They calculated maximum vertical GRF as:

**Equation 4S.11**

$$F_{y,max} = c_1(n)a_{WCS,x} + c_2(n)a_{WCS,y} + c_3(n)a_{WCS,z}$$

Where  $c_1$ ,  $c_2$ , and  $c_3$  were coefficients that varied linearly as a function of stride number  $n$  and where we assumed that  $F_{y,max}$  was the vertical GRF second peak and that each “acceleration component” was the maximum value observed during stance, such that:

**Equation 4S.12**

$$F_{y,second} = (c_{1a}n + c_{1b})a_{WCS,x,max} + (c_{2a}n + c_{2b})a_{WCS,y,max} + (c_{3a}n + c_{3b})a_{WCS,z,max}$$

Thiel et al. noted that this stride-varying approach was suitable until the maintenance (steady-state) phase where an approximately constant relation between acceleration and force is expected. Thus, during steady state running, the coefficients could be assumed constant, and the equation simplified:

**Equation 4S.13**

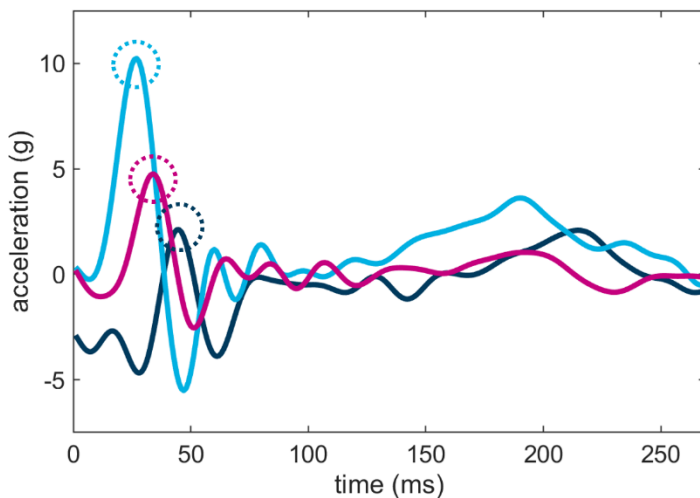
$$F_{y,second} = c_1a_{WCS,x,max} + c_2a_{WCS,y,max} + c_3a_{WCS,z,max}$$

Here, we assumed that (1) participants are running at a steady-state and therefore used the constant expression Equation 4S.13, and that (2) Thiel et al. used a WCS with ~vertical, ~anterior, and ~medial defined as positive (although they did not fully describe their coordinate convention, they did provide time-series acceleration figures that suggest this was their convention. Note, that their text



suggests that the right and left medial-lateral axes were defined with positive in opposite directions; however, we inferred that the minima was taken from the right and the maxima from the left, effectively making the method the maxima from the ~medial direction).

To execute this method, we used a leave-one-out cross-validation to iteratively calculate the error for one participant using  $c_1$ ,  $c_2$ , and  $c_3$  coefficients calculated using the other 73 participants while the final  $c_1$ ,  $c_2$ , and  $c_3$  coefficients were calculated using data from all 74 of our participants.



**Figure 4S.4: Thiel method.** Tibial accelerations from the x- (anterior-posterior in the SCS; ~direction of progression in the WCS; dark blue line), y- (proximal-distal in the SCS; ~longitudinal in the WCS; light blue line), and z-axes (medial-lateral in the SCS; ~right in the WCS; pink line) maxima were found (dashed circles). Note, we inferred that the largest medial value was taken from the z-axis; thus, for right stances (such as the one shown here) the z-axis was multiplied by -1. These values were multiplied by the coefficients  $c_1$ ,  $c_2$ , and  $c_3$  to estimate vertical GRF second peak magnitudes.

#### 4S.1.7. Kiernan method

Building on work by Neugebauer et al. [86, 87, 278], Kiernan et al. [282] had 40 participants run overground at slow, typical, and fast speeds while recording accelerations from their iliac crests and sacra. In an effort to estimate both the first and second peak of the vertical GRF, the  $a_{SCS,y}$  signal from each location was divided into signals composed of 0-8 Hz ‘LoF’ frequency content and  $\geq 10$  Hz ‘HiF’ frequency content [271]. For each of the two sensor locations, maxima from the LoF and HiF signals were found and entered into a linear regression along with sex, height, mass, and leg length to estimate

log transformed vertical GRF first and second peaks. Kiernan et al. also explored models including speed and participant as fixed and random effects. They found this could improve model performance; however, we use the generalizable form of their model:

**Equation 4S.14**

$$\ln(F_{y,second}) = c_1 + c_2 a_{SCS,y,LoF,max} + c_3 a_{SCS,y,HiF,max} + c_4 s + c_5 m + c_6 h + c_7 g$$

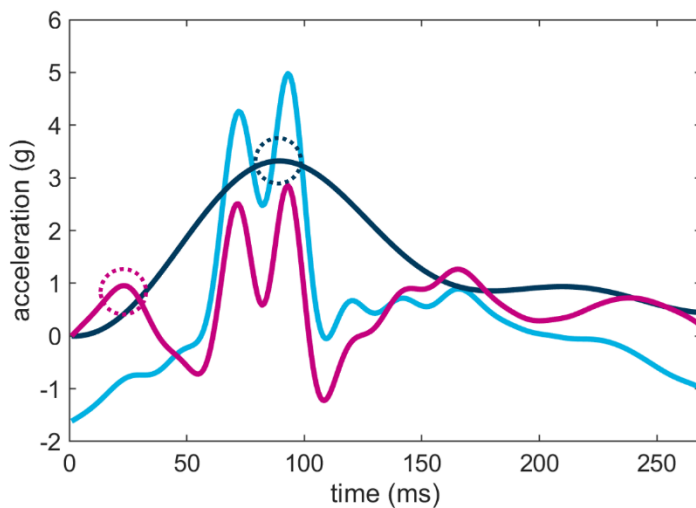
Which can be rearranged as:

**Equation 4S.15**

$$F_{y,second} = e^{c_1 + c_2 a_{SCS,y,LoF,max} + c_3 a_{SCS,y,HiF,max} + c_4 s + c_5 m + c_6 h + c_7 g}$$

Where  $c_{1:7}$  are constants,  $a_{SCS,y,LoF,max}$  and  $a_{SCS,y,HiF,max}$  are low and high frequency acceleration maxima,  $s$  is the self-reported participant sex (with female = 0, male = 1, and no non-binary reported),  $m$  is the participant mass,  $h$  is the participant height, and  $g$  is the height of the greater trochanter.

To replicate this method, we used a leave-one-out cross-validation to iteratively calculate the error for one participant using  $c_{1:7}$  values found using the other 73 participants while the final  $c_{1:7}$  values were found using data from all 74 of our participants. Note that the original model published by Kiernan et al. was developed with a subset of the current sample (40 of our 74 participants).



**Figure 4S.5: Kiernan method.** Sacrum y-axis acceleration (proximal-distal in the SCS; ~longitudinal in the WCS; light blue line) and its low frequency (dark blue line) and high frequency components (pink line). The peak in the low frequency was identified (dark blue circle) then the earliest occurring peak in the high frequency between the start of the stance and the low frequency

peak was identified (pink circle). These peaks were entered into a linear regression to estimate vertical GRF second peak magnitude.

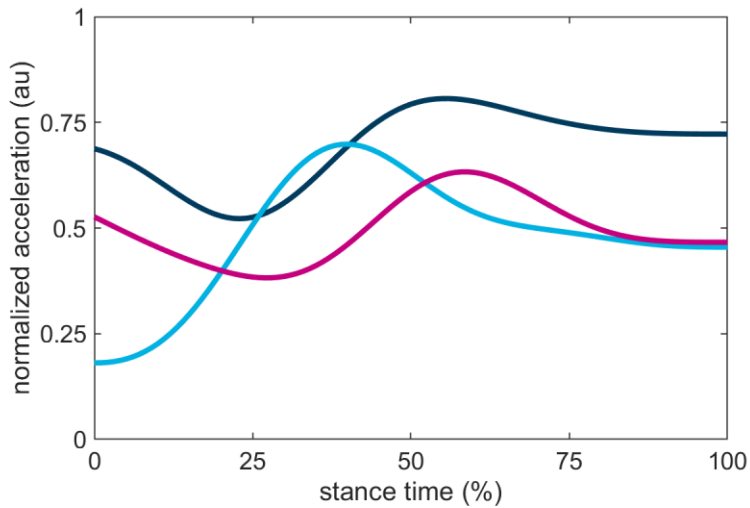
#### 4S.1.8. Kim method

To estimate GRF from acceleration during running, Kim et al. [267] used a feed-forward neural network (FFNN). They proposed two models: a “SLIP” model (spring loaded inverse pendulum) that estimated GRF from sacrum displacement (double integrated acceleration) and a “rigid dynamics” model that estimated GRF from sacrum acceleration.

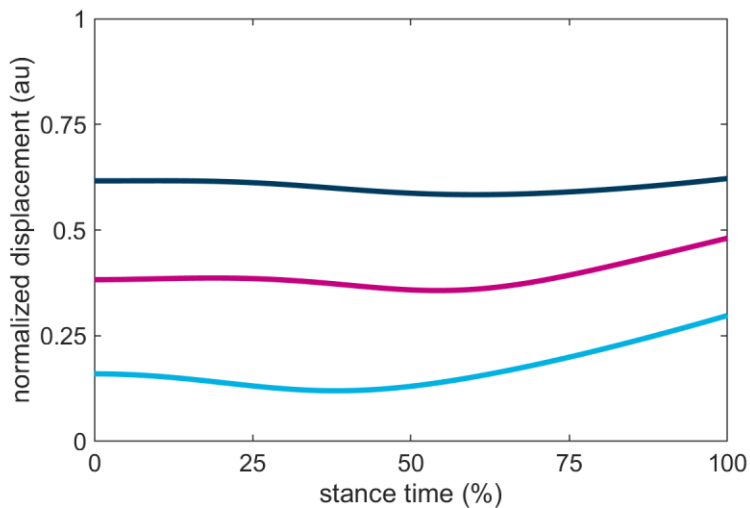
Kim et al. tested these two models on seven participants who ran on a treadmill while sacral acceleration was recorded via motion capture. Triaxial accelerations were low-pass filtered at 10 Hz with a 5<sup>th</sup> order Butterworth filter. Accelerations (or displacements) at a single time point, and the time point itself, were then entered into a FFNN with a single 10-node hidden layer to estimate triaxial forces normalized to body weight. Here, however, we followed the approach laid out in Equation 4S.7 (above) and instead multiplied the acceleration input by mass to estimate non-normalized force.

Kim et al. did not report a normalization procedure for the acceleration inputs to the FFNN but based on their previous work [226] we assumed that data were normalized by maximum acceleration such that all values ranged from 0 to 1. Similarly, we inferred from their figures that time was expressed as percent stance from 0 to 100. Few details were provided regarding model parameters; thus, we assumed that log-sigmoid and pure linear activation functions were used in the hidden and output layers and that a 2000 epoch Levenberg-Marquardt function was used for training.

To evaluate their model, we used a leave one out approach and iteratively trained a network to minimize mean square errors on 73 participants using an 80-20 training-validation split. We then calculated errors between estimated and actual  $F_{y,t}$  on the remaining participant. Weights and biases for the final model were calculated by training the FFNN on all 74 participants.



**Figure 4S.6: Kim method. (A)** Sacral accelerations normalized from 0 to 1 based off the minimum and maximum observed values in the x- (anterior-posterior in the SCS;  $\sim$ direction of progression in the WCS; dark blue line), y- (proximal-distal in the SCS;  $\sim$ longitudinal in the WCS; light blue line), and z-axes (medial-lateral in the SCS;  $\sim$ right in the WCS; pink line). Tri-axial values from each time point were used to estimate corresponding vertical GRF values.



**(B)** Sacral displacements normalized from 0 to 1 based off the minimum and maximum observed values in the x- (anterior-posterior in the SCS;  $\sim$ direction of progression in the WCS; dark blue line), y- (proximal-distal in the SCS;  $\sim$ longitudinal in the WCS; light blue line), and z-axes (medial-lateral in the SCS;  $\sim$ right in the WCS; pink line). Tri-axial values from each time point were used to estimate corresponding vertical GRF values.

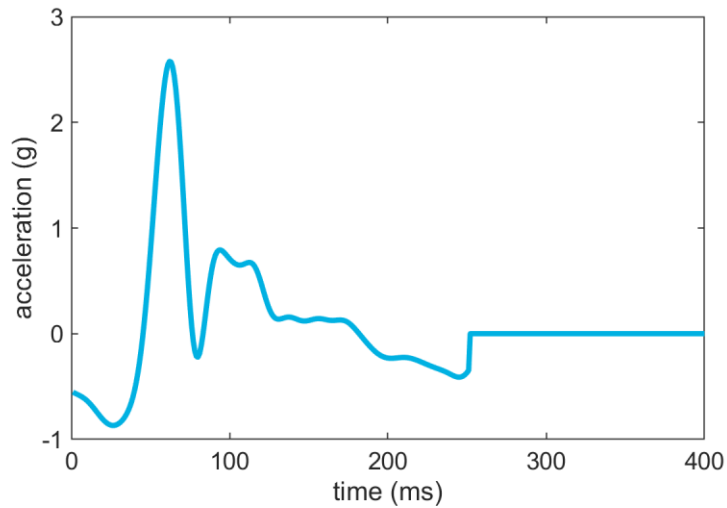
#### 4S.1.9. Pogson method

Pogson et al. [283] measured resultant forces and upper back accelerations while 15 team sport athletes ran overground. Given our goal of estimating vertical GRFs we instead used y-axis force and acceleration. Accelerations were segmented by stance then zero-padded to the duration of the longest stance (we zero-padded to 0.4 s to ensure the method could accommodate all future data). Acceleration and force data were then entered into a Principal Component Analysis. The acceleration principal components (PCs) and signal duration were then entered into a multilayer perceptron (MLP) which was

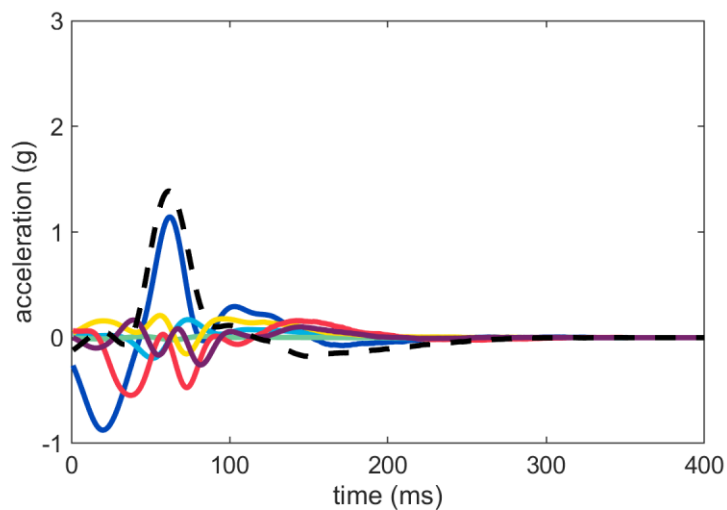
trained to estimate the force PCs. Pogson et al. used a stochastic particle swarm optimization to determine the number of inputs (acceleration PCs), hidden layers, number of nodes in each layer, and outputs (force PCs) that minimized force PC estimation error. They reported that all optimizations returned similar values but found slightly better results with an MLP using 6 acceleration PCs, with 5 hidden layers containing 45, 36, 45, 82, and 40 nodes, and trained to estimate 8 force PCs. The force PCs were then used to reconstruct the force signal. We assumed that input data were normalized from 0 to 1 and that sigmoid activation functions were used between each of the layers except for the output layer where a linear activation function was used.

The reconstructed force signal could vary in duration from the input and target signals. To address this issue, Pogson et al. reported appending the signal duration to the output layer (as a quantity to be estimated) then trimming the force signal to the estimated duration. During development, however, we observed that this approach often led to input and output signals of different durations, increasing error (consistent with Pogson et al.'s Figure 4). Given that the estimated force should always have the same duration as the acceleration input (because both are segmented based on stance onset and offset) we trimmed any values following the point where estimated force fell below 10 N then interpolated the force signal to match the duration of the input signal. This approach greatly reduced error.

To evaluate Pogson's method, we tuned the hyperparameters of the MLP solver using a grid search then used a leave one out approach to iteratively train a network that minimized mean square errors on 73 participants using an 80-20 training-validation split. We then calculated errors between estimated and actual  $F_{y,t}$  on the remaining participant. Weights and biases for the final model were calculated by training the MLP on all 74 participants.



**Figure 4S.7: Pogson method. (A)** Sacral acceleration in the y-axis (proximal-distal in the SCS; ~longitudinal in the WCS; light blue line) zero-padded to 400 ms.



**(B)** The first six Principal Component reconstructions for the example trial (solid lines; PC scores multiplied by coefficients) and mu (dashed line; the estimated means of each time point across the sample). The sum of these seven lines reconstructs the original signal. The scores used to calculate the solid lines were entered into the machine learning model to estimate vertical GRF PC scores.

#### 4S.1.10. Pogson xynorm method

We observed that the Pogson et al. [283] method was promising but that the majority of error originated from differences between the duration of the original force signal and the estimated force signal. To deal with this source of error we modified their approach: Instead of zero-padding the data we standardized all signals to 101 time points and followed the approach laid out in Equation 4S.7 (above), multiplying the acceleration input by mass before applying the Principal Component Analysis. After estimating the PC scores and reconstructing the signal it was rescaled to its original duration.

To evaluate this model, we tuned the hyperparameters of the MLP solver using a grid search then used a leave one out approach to iteratively train a network that minimized mean square errors on 73 participants using an 80-20 training-validation split. We then calculated errors between estimated and actual  $F_{y,t}$  on the remaining participant. Weights and biases for the final model were calculated by training the MLP on all 74 participants.

#### 4S.1.11. Day method

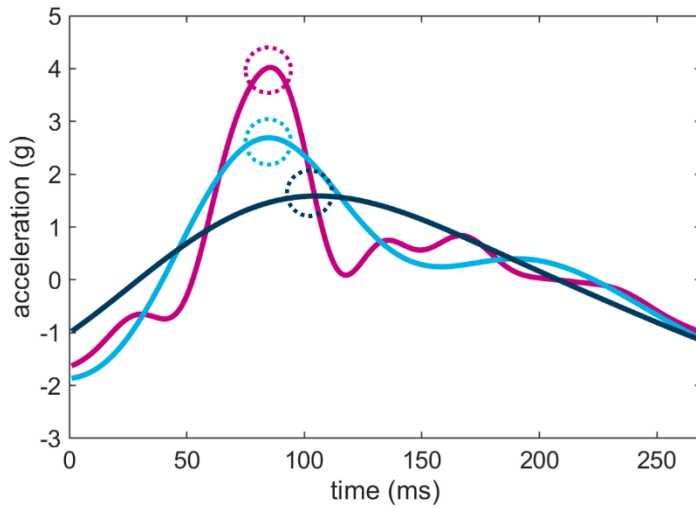
Day et al. [284] had 30 NCAA Division 1 cross country runners run on an instrumented treadmill while wearing an IMU clipped to their posterior waistband. They filtered  $a_{WCS,y}$  at 5, 10, and 30 Hz. Other filter parameters were not reported thus we assumed they used the same 8<sup>th</sup> order low-pass Butterworth filter that was used to filter their force data. Then, using an approach similar to Wundersitz et al. [280], Meyer et al. [281], and Gurchiek et al. [247], they multiplied acceleration by mass to estimate force:

#### Equation 4S.16

$$F_{y,t} = 9.8ma_{WCS,y,t}$$

Where  $m$  is mass in kg and  $a_{WCS,y,t}$  is the acceleration in g (with  $1\text{ g} = 9.8\text{ m/s}^2$ ) at time  $t$ . This equation resembles the one used by Wundersitz et al. [280], Meyer et al. [281], and Gurchiek et al. [247] (above) but with different data entered into the equation (namely, data had different filtering, a different coordinate system, and came from a different sensor location).

For this method, we found that the aggressive filtering (particularly the 5 Hz) led to signal distortion when using stance-segmented data and thus, in contrast to the other methods, here we filtered *before* segmenting by stance. If using this method, please be aware that your data must be filtered before passing your stance-segmented data into our provided code.



**Figure 4S.8: Day method.** Sacrum y-axis acceleration (proximal-distal in the SCS; ~ longitudinal in the WCS) low-pass filtered at 5, 10, and 30-Hz (dark blue, light blue, and pink lines, respectively) maxima (dashed circles) were multiplied by participant mass to estimate vertical GRF second peak magnitude.

#### 4S.1.12. Higgins method

Higgins et al. [285] had 30 participants perform a range of tasks including jogging and running down a 23 m pathway with an embedded force plate while wearing an accelerometer on their right hip and ankle. Higgins et al. took the maximum accelerations in  $a_{WCS,y}$  during each stance and entered them into linear mixed models (note that their coordinate convention was not explicitly defined outside of stating that the vertical vector is “typically in line with the majority of gravity related loading;” based on this statement, we inferred that they used a WCS with the ~vertical direction defined as positive). In addition to  $a_{WCS,y,max}$ , Higgins et al. explored several variables including sex, age, and “activity code” as potential predictors of either average vertical loading rate or vertical GRF first peak magnitude. Potential predictors that were not significant were iteratively removed leading to the development of the following equations:

#### Equation 4S.17

$$F_{y,first} = \alpha + \beta_1 a_{WCS,y,max} + \beta_2 A$$



**Equation 4S.18**

$$\frac{dy}{dx} F_{y,first} = \alpha + \beta_1 a_{WCS,y,max} + \beta_2 A + \beta_3 y$$

Where  $A$  is activity code (jogging = 1, running = 2) and  $y$  is age in years. Here, we assume a single type of activity (running) resulting in a constant activity code  $A$ . Further, although Higgins et al. originally specified  $\alpha$  as a participant-specific intercept, our goal here is to produce a generalizable method thus we simplify their approach to:

**Equation 4S.19**

$$F_{y,first} = c_1 + c_2 a_{WCS,y,max}$$

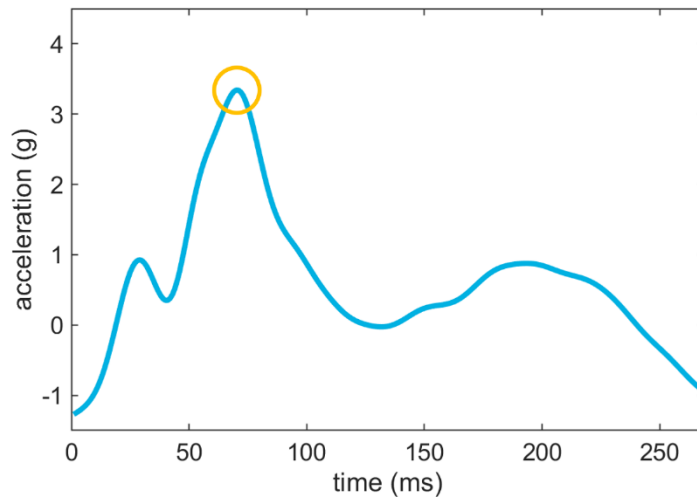
**Equation 4S.20**

$$\frac{dy}{dx} F_{y,first} = c_1 + c_2 a_{WCS,y,max} + c_3 y$$

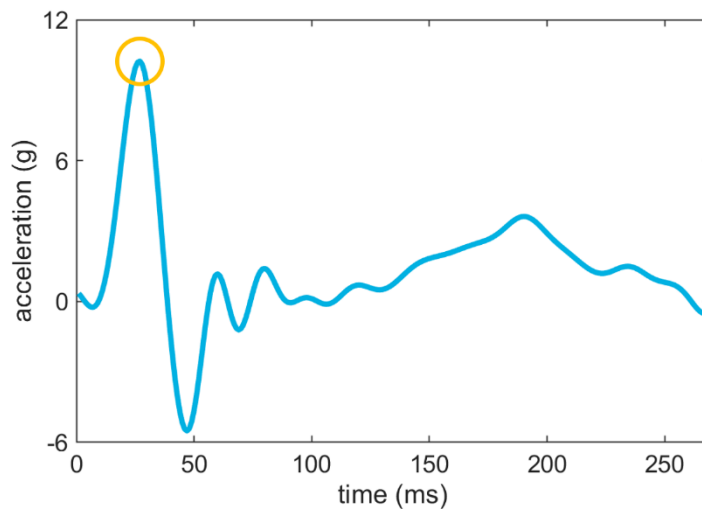
Where  $c_1$  corresponds to  $(\alpha + \beta_2 A)$  with  $A$  being constant and where  $c_2$  and  $c_3$  correspond to  $\beta_1$  and  $\beta_3$ .

Note that Higgins et al. also attempted to estimate average vertical loading rate and vertical GRF first peak magnitude using the same approach but with shank accelerations. They found that hip accelerations provided better estimates across the activities they studied and thus did not report full details on their shank models (despite reporting that  $a_{WCS,y,max}$  measured at the shank had the highest observed correlation with vertical loading rate during running). Thus, to replicate their shank method we assumed that it was identical to their hip method.

We used a leave-one-out cross-validation to iteratively calculate the error for one participant using  $c_1$ ,  $c_2$ , and  $c_3$  values found using the other 73 participants while the final  $c_1$ ,  $c_2$ , and  $c_3$  values were found using data from all 74 of our participants.



**Figure 4S.9: Higgins method. (A)** Hip y-axis acceleration (proximal-distal in the SCS; ~longitudinal in the WCS; blue line) maxima (yellow circle).



**(B)** Shank y-axis acceleration (proximal-distal in the SCS; ~longitudinal in the WCS; blue line) maxima (yellow circle).

#### 4S.1.13. Veras method

Veras et al. [255] had 131 participants walk and run on an instrumented treadmill while wearing accelerometers on their tibiae, hips, and sacra. Peak accelerations and jerks (first derivative of acceleration) were extracted from  $a_{WCS,y}$  and  $a_{WCS,res}$  signals low pass filtered with a 4<sup>th</sup> order 20 Hz Butterworth filter. These values were entered into linear regressions along with participant mass to predict  $F_{y,max}$  and  $\frac{dy}{dx}F_{y,first}$ . Body mass index and body mass index category were also explored as potential predictor variables but did not improve model performance. On the other hand, models were

improved by the inclusion of random effects for speed and participant. To maintain generalizability, however, we omit those variables resulting in the equations:

**Equation 4S.21**

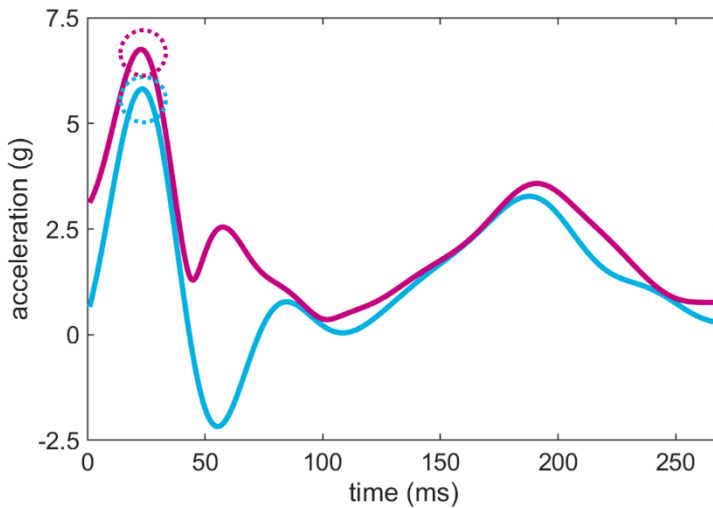
$$F_{y,max} = c_1 + c_2 a_{WCS,y,max} + c_3 m + c_4 m a_{WCS,y,max}$$

**Equation 4S.22**

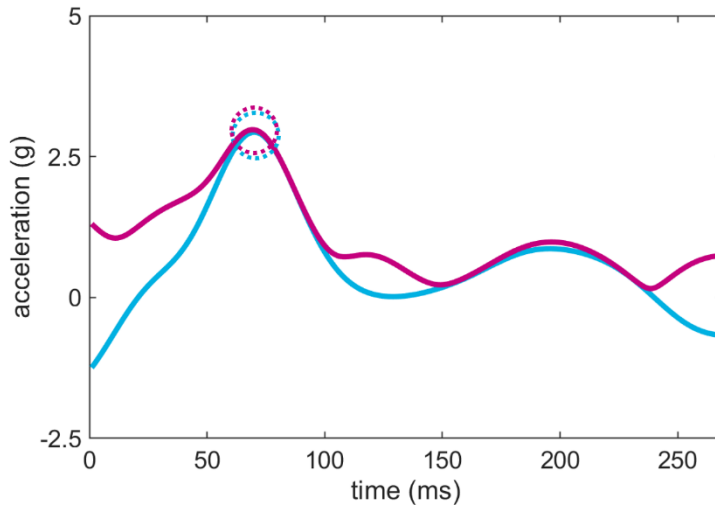
$$\frac{dy}{dx} F_{y,first} = c_1 + c_2 \frac{dy}{dx} a_{WCS,y,max} + c_3 m + c_4 m \frac{dy}{dx} a_{WCS,y,max}$$

Where  $c_1$ ,  $c_2$ ,  $c_3$ , and  $c_4$  are constants and  $m$  corresponds to participant mass.

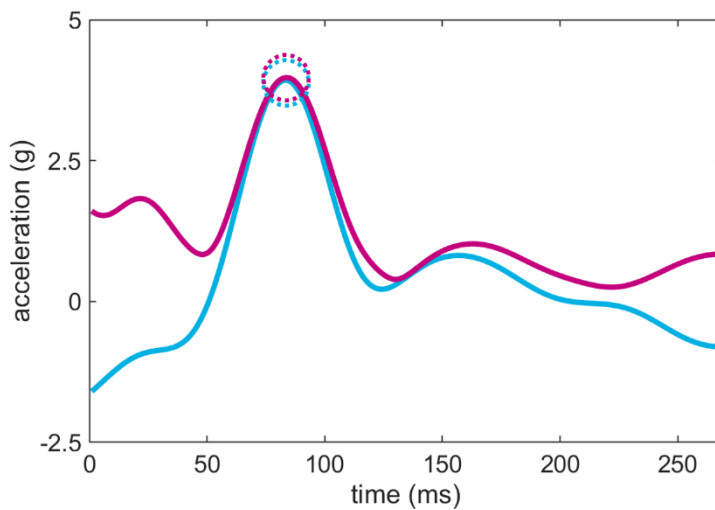
Here, we assumed that  $F_{y,max}$  is analogous to  $F_{y,second}$ . We then used a leave-one-out cross-validation to iteratively calculate the error for one participant using  $c_1$ ,  $c_2$ ,  $c_3$ , and  $c_4$  values found using the other 73 participants while the final  $c_1$ ,  $c_2$ ,  $c_3$ , and  $c_4$  values were found using data from all 74 of our participants.



**Figure 4S.10: Veras method. (A)** Shank y-axis (proximal-distal in the SCS; ~longitudinal in the WCS; blue line) and resultant (pink line) accelerations and maxima (dashed circles).



**(B)** Hip y-axis acceleration (proximal-distal in the SCS; ~longitudinal in the WCS; blue line) maxima (yellow circle).

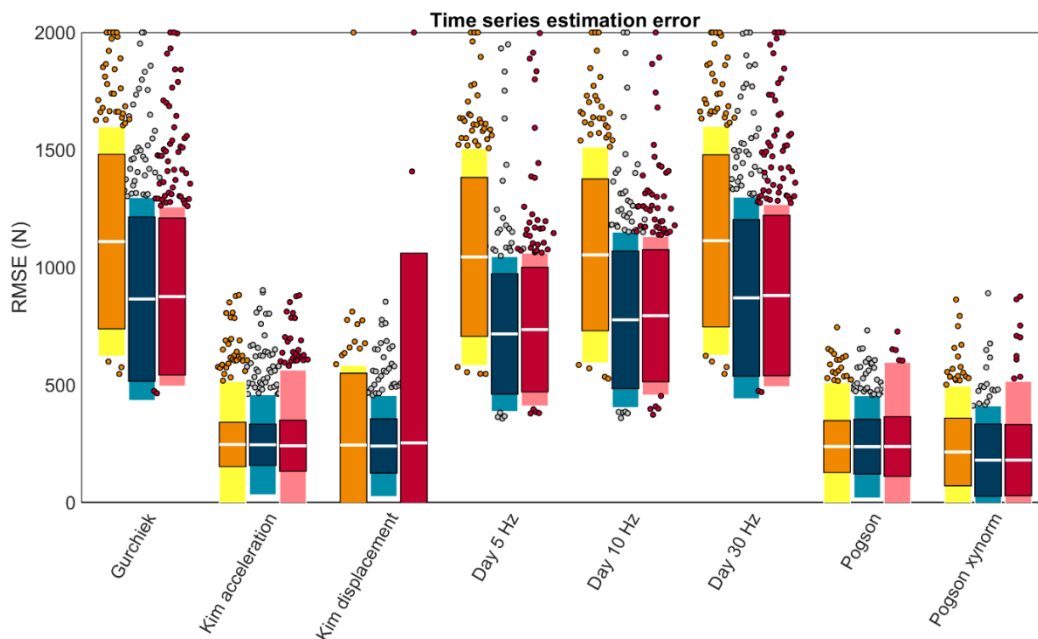


**(C)** Sacrum y-axis acceleration (proximal-distal in the SCS; ~longitudinal in the WCS; blue line) maxima (yellow circle).

#### 4S.2. Results across wearable, segment, and tilt-corrected coordinate systems

Despite the fact that the majority of the publications described in Section 4S.1. expressed their acceleration data in the WCS (11 of 13), we presented our results in the main paper based on acceleration data expressed in the SCS. We made this choice based on an *a priori* expectation that the WCS would be more prone to error due to IMU misplacement/misalignment, differences in participant morphology, and/or differences in participant posture/kinematics. We did, however, also execute the same analyses presented in the main paper but with WCS and TCCS acceleration data as inputs.

Overall, we found that for all linear regression and machine learning methods that had coefficients, weights, or biases fit to the specific input data (i.e., had different values for WCS, SCS, and TCCS), there were no systematic effects. In contrast, for methods that simply multiplied acceleration by mass, the WCS error was systematically higher. This result was consistent across all estimated features (first peak, loading rate, second peak, average, and time series); thus, we present only the time series results here to illustrate (Figure 4S.11).



**Figure 4S.11: Results across coordinate systems.** Mean RMSE for each method capable of estimating vertical GRF time series based on acceleration inputs. Orange and yellow represent estimations made with models trained on WCS input data, blue on SCS input data (identical to the main paper), and red on TCCS input data. The ‘Gurchiek’ and ‘Day’ methods did not have coefficients, weights, or biases fit to individual coordinate systems and showed systematically higher errors in the WCS.

## **5. A 0.05 m change in inertial measurement unit placement alters time and frequency domain metrics during running.**

### **5.1. Introduction**

Inertial measurement units are small, low-cost, light-weight devices that measure acceleration, angular velocity, and ferromagnetic fields. These wearable devices offer several key advantages over systems that are ‘captive’ to lab environments [286]. Captive systems (like force plates and video motion capture) are relatively expensive, require dedicated facilities, and are time-consuming to set up and operate [240, 287]. These factors limit the general population’s access to captive systems and the biomechanical analyses they can provide [287]. Further, even when accessible, captive systems may cause participants to alter their gait (e.g., the Hawthorne effect, running on a treadmill or short track, targeting force plates during over-ground running in lab), can limit the volume of data collected to a few ‘representative’ gait cycles [288], cannot be used to provide real-time feedback in the field, and cannot capture biomechanics that may only occur under certain real-world conditions (e.g., weather, running surfaces, races and training) [289, 225, 290, 291]. These limitations have led to most biomechanics studies capturing relatively brief ‘snapshots’ of running that may not accurately represent the millions of gait cycles that occur over many long bouts in the field [292], creating a gap in our understanding of running behavior [41].

IMUs overcome these limitations and can be used in-field, facilitating the collection of large volumes of running biomechanics data under real-world conditions [145]. These devices are much more accessible to the general population than captive systems, with >90% of runners already reporting regularly wearing a tracking device or watch (similar in size and cost to an IMU) to improve their training outcomes or avoid injury [15, 293, 294, 295]. These advantages have led to the use of IMUs to collect

data in ways that captive technology cannot. For example, IMUs can estimate gait events, external loading, running speed, and kinematics for entire runs in the field [207, 296, 297, 298], can be used to quantify biomechanical changes over long-duration runs in the field [299, 300, 301, 302, 303], can continuously monitor biomechanics that may lead to injury over many bouts of running [198, 304, 305, 306], and can be used to provide instantaneous feedback in the field [307, 308, 309, 310, 311]. Thus, IMUs have the potential to greatly expand the volume and ecological validity of data available to runners, coaches, researchers, and clinicians.

Despite this potential, there are challenges to deploying IMUs to collect in-field data across many runs with long durations. When collecting data on long runs, an IMU may change position over the duration of the run, potentially altering the data obtained. When collecting data over many runs, it is likely that the initial placement of the IMU will vary slightly, particularly when end users (e.g., coaches, runners) are not in a constrained lab or clinical environment and do not have the same training palpating anatomical landmarks as researchers and clinicians. Thus, inconsistencies in IMU placement between and within users, as well as IMU movement during data collection, may ultimately decrease the repeatability of measurements and reliability of results. In turn, errors in IMU-derived quantities could result from differences in placement rather than any difference between participants or conditions, leading to misleading findings.

Unfortunately, the critical effects of IMU misplacement and movement on running data are little explored. Previous research suggests that small variations in IMU location can affect estimated ground reaction forces, knee joint angles, and inter-session reliability in walking [312, 313, 314], shank and foot accelerations in running [315, 238, 316], and lumbar accelerations in cadavers [317]. However, systematic quantification of the effects of placement variation on acceleration and angular velocity time and frequency domain metrics *in vivo* is still lacking. To address this gap, this paper quantifies the effects that a 0.05 m difference in IMU placement has on the time and frequency domains during running. IMUs

were placed at three common locations (shank, pelvis, and sacrum) [318], then, to represent a worst-case misplacement/movement scenario, a second IMU was ‘misplaced’ 0.05 m away and data were recorded simultaneously. Potential differences between IMUs were quantified as (1) the root mean square error (RMSE) between time domain signals, (2) the magnitude and timing of peaks, (3) the differences in outcome variables commonly estimated with IMUs, including temporal (initial contact and terminal contact) and kinetic metrics (vertical ground reaction force second peak magnitude, average, and RMSE), (4) the magnitude-squared coherence between signals, and (5) the proportion of signal power contained in different frequency bins. In sum, these descriptive analyses provide a wholistic understanding of the potential effects that IMU misplacement or movement can have on acceleration and angular velocity time and frequency domain metrics and derived outcome variables.

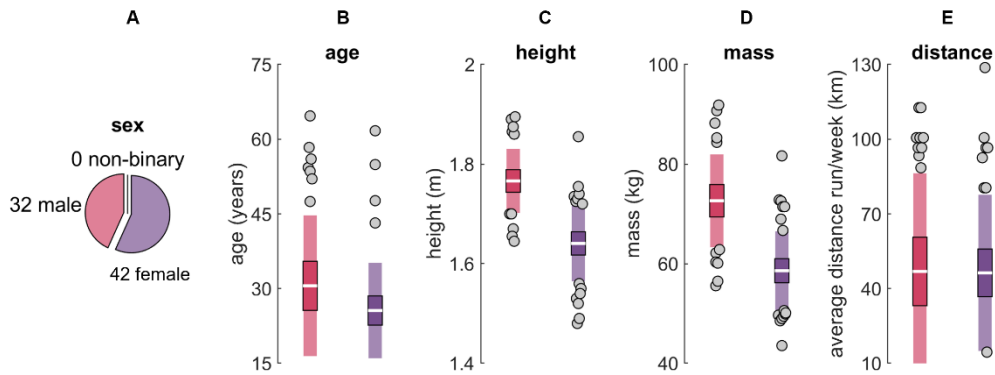
## **5.2. Methods**

Data collection for this study was first reported in separate analyses in Chapters 3 and 4 [207, 296] but is briefly repeated here for convenience.

### *5.2.1. Participants*

Seventy-seven participants were recruited from UC Davis, local running clubs, and the community at large. Participants were  $\geq 18$  years old and reported running  $\geq 16.09$  km per week for  $\geq 6$  months. Three participants were excluded from analysis due to an inability to complete the protocol as instructed ( $n = 1$ ) or an IMU moving from its original placement location across the duration of data collection (e.g., IMU belt rotated about the long axis of the shank or ‘rode up’ moving the IMU proximal;  $n = 2$ ), leaving a final sample of 74 (32 males; 42 females; 0 non-binary; age  $28 \pm 12$  years; Figure 5.1). All participants provided written informed consent, and procedures were approved by the UC Davis Institutional Review Board.



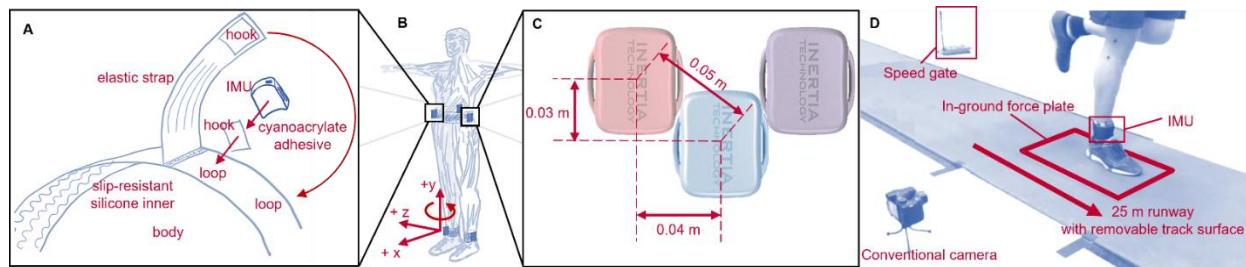


**Figure 5.1: Participant characteristics.** (A) sex, (B) age, (C) height, (D) mass, and (E) self-reported average distance run per week for males (red) and females (purple). The white horizontal line represents the mean; dark colors represent  $\pm 95\%$  confidence interval ( $\pm 1.96$  SEM) around the mean; and light colors represent  $\pm 1$  SD around the mean. Gray dots represent participants outside  $\pm 1$  SD.

### 5.2.2. IMU placement

Using adhesive-bonded hook-and-loop fasteners, IMUs, each with two tri-axial accelerometers and one tri-axial gyroscope (ProMove MINI, Inertia Technology, Enschede, The Netherlands;  $\pm 16$  g primary accelerometer with 0.0005 g resolution,  $\pm 100$  g secondary accelerometer with 0.05 g resolution,  $\pm 34.91$  rad/s gyroscope with 0.001 rad/s resolution, 1000 Hz; see <https://inertia-technology.com/wp-content/uploads/2022/08/ProMoveMiniAdvGwUserManual3.8.10.pdf>; accessed on 24 December 2023), were attached to neoprene belts with anti-slip silicone inners, then wrapped with elastic straps as tightly as possible, within the limit of participant comfort (Figure 5.2A). IMUs were ‘correctly’ placed at three locations commonly used for IMU-based research: (1) anterosuperior to the lateral malleoli (*shank*), (2) on the superior aspect of the iliac crests in line with the greater trochanter (*pelvis*), and (3) on the superior aspect of the sacrum in line with the spine (*sacrum*) (Figure 5.2B). The correctly placed IMU on the right shank ( $n = 26$ ), right pelvis ( $n = 24$ ), or sacrum ( $n = 24$ ) was then pseudo-randomly selected as the ‘reference’ IMU and another IMU was ‘misplaced’ 0.05 m on-center from the correctly placed reference IMU. The misplaced IMU was always 0.03 m more proximal than the reference IMU. Fifty percent of the time it was placed 0.04 m anterior/ventral and 50% of the time it

was placed 0.04 m posterior/dorsal (for shank and pelvis locations) or 0.04 m left and right (for the sacrum) (Figure 5.2C). Given the physical size of the IMUs, these were the smallest misplacements possible that still allowed the misplaced IMU to be secured to the participant in a manner identical to the reference IMU. This 0.05 m change in placement likely represents a ‘worst-case’ misplacement/movement scenario.

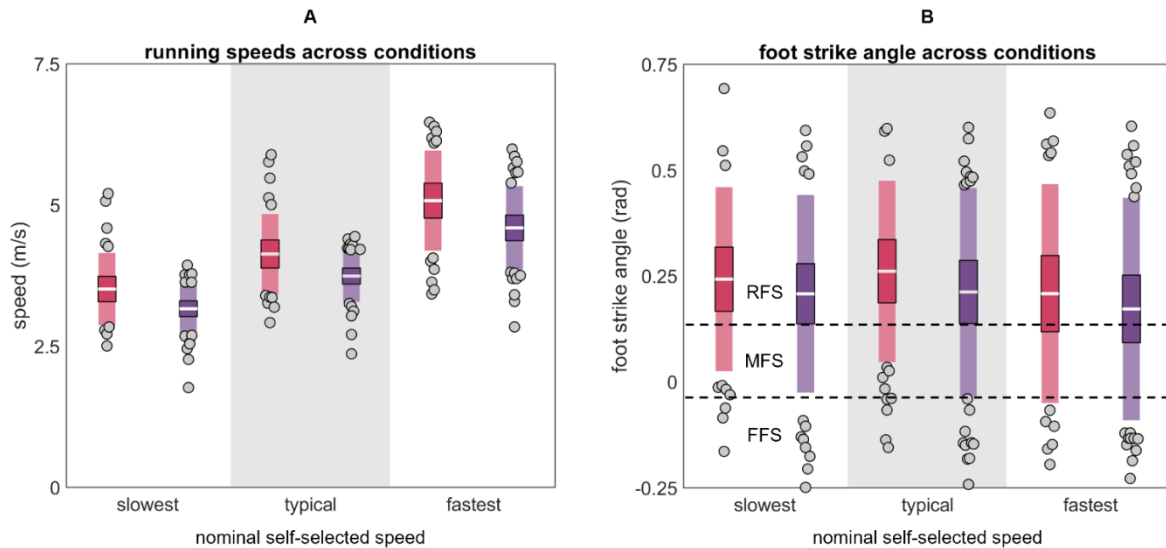


**Figure 5.2: Set up and protocol.** (A) Belt design and IMU fixation. (B) IMU placement and coordinate conventions. A segment coordinate system was defined as anterior (+x), proximal (+y), and medial-lateral (with right defined as +z) and adduction-abduction/right downward-upward tilt/right-left lateral bending, internal-external rotation/left-right axial rotation, and flexion-extension/anterior-posterior tilt were defined about the x, y, and z axes with the right hand rule [169]. (C) The reference IMU (blue location) was ‘correctly’ placed anterosuperior to the right lateral malleolus (*shank*), on the superior aspect of the right iliac crest in line with the greater trochanter (*pelvis*), or on the superior aspect of the sacrum in line with the spine (*sacrum*). A single ‘misplaced’ IMU was then positioned 0.03 m proximal and either 0.04 m to the left or right of the reference IMU (in the red and purple locations). (D) Experimental setup.

### 5.2.3. Protocol

Participants wore their own shoes and ran a 25 m runway with an embedded force plate (Kistler 9281, Kistler Group, Winterthur, Switzerland; 1000 Hz). Running speed was recorded using two custom-built laser speed gates, placed 2.5 m on each side of force plate center. Participants warmed up and practiced striking the force plate three times per side at their slowest (“the slowest pace you would use on a run”), typical (“the pace you use for the majority of your running”), and fastest (“the fastest pace you would use on a run”) self-selected speeds (Figure 5.3). After warm-up, five stances per side were collected at each speed for two surface conditions: (1) with a track surface covering the runway and force plate, and (2) with no covering on the hardwood floor of a basketball court. Participants always progressed from their slowest to fastest speeds, but the order of foot and surface was pseudo-randomized. IMU data were synchronized within 100 ns of each other with a wireless network hub

(Advanced Inertia Gateway, Inertia Technology, Enschede, The Netherlands). Data were rejected (5.83% of all trials) if visual inspection revealed atypical kinematics or kinetics suggesting that the participant was targeting the force plate, positively or negatively accelerating, or otherwise not exhibiting a steady state running pattern, resulting in a total of 4181 trials for analysis.



**Figure 5.3: Running speeds and foot strike angles across conditions.** (A) Mean speeds and (B) foot strike angles calculated from markers on the lateral calcaneus and base of the fifth metatarsal for males (red) and females (purple) across the slowest, typical, and fastest conditions (RFS is rear foot strike, MFS is mid foot strike, and FFS is fore foot strike [270]). The white horizontal line represents the mean; dark colors represent  $\pm 95\%$  confidence interval ( $\pm 1.96$  SEM) around the mean; and light colors represent  $\pm 1$  SD around the mean. Gray dots represent participants outside  $\pm 1$  SD.

#### 5.2.4. Processing

For full IMU processing details see Chapter 3S. In brief, calibration matrices were applied to IMU data. Quiet periods were identified (angular velocity  $< 0.5$  rad/s and jerk  $< 0.01$  m/s<sup>3</sup> for at least 100 ms) and used to remove biases. Saturated frames from the primary accelerometer ( $|a| > 15.5$  g) were replaced with corresponding frames from the secondary accelerometer. Data were filtered with a 4th-order 50-Hz lowpass Butterworth filter. Angular velocity was drift-corrected using a Madgwick filter [193, 171]. Starting at each quiet period, accelerations were used to estimate IMU position in the inertial reference frame, then angular velocities were used to estimate frame-by-frame changes in IMU orientation and remove the gravity component from accelerations [174] Data were then expressed in a

segment coordinate system based on the Principal Component that explained the most variance in angular velocity during running (the medial–lateral axis) and the gravity vector during quiet standing [121, 175]. This system was defined as anterior (+x), proximal (+y), and medial–lateral (with right defined as +z), and adduction–abduction/right downward–upward tilt/right–left lateral bending, internal–external rotation/left–right axial rotation, and flexion–extension/anterior–posterior tilt were defined about the x, y, and z axes with the right hand rule [169] (Figure 5.2B).

#### 5.2.5. Analysis

The Purcell method [150] (as implemented by Kiernan et al. [207]) was used to identify initial contact events from acceleration of the reference shank IMU. The stride (right foot initial contact to right foot initial contact) containing or immediately following force plate contact was identified and segmented for further analysis. Means and standard deviations were calculated and plotted for each axis of the reference and misplaced acceleration and angular velocity signals (Chapter 5S.1). Root mean squared error (RMSE) between these signals was calculated. The stride was then concatenated with itself, 50 ms was removed from the start and end, and peak absolute acceleration and angular velocity were found for each axis. A 101 ms search window centered on the reference peak was then used to find the peak absolute acceleration and angular velocity in the time-synchronized misplaced IMU signal. Differences in the magnitude and timing of reference and misplaced peaks were then calculated along with limits of agreement (LOAs;  $\pm 1.96$  SD) within which 95% of future differences are expected to fall.

To compare the potential consequences of misplacement on outcome metrics, gait events and vertical ground reaction forces were estimated from both reference and misplaced IMUs at the shank and sacrum. For the shank, gait events were estimated using the Purcell method [150] while vertical ground reaction force second peak magnitude was estimate using the Charry method [264] (as

implemented by [207] and [296], respectively). For the sacrum, gait events were estimated using the Auvinet method [162] while vertical ground reaction force second peak magnitude, stance averages, and time series were estimated using the Pogson-Auvinet method [268] (as implemented by [207] and [296], respectively). Differences and LOAs were then calculated.

To compare the frequency domains of reference and misplaced IMU signals, a Fourier transform was used to calculate power spectral density at frequencies from 0 to 50 Hz (the low-pass filter cut-off frequency) in 1 Hz bins. Magnitude squared coherence was calculated between reference and misplaced IMUs via the Welch method. The proportion of signal power in three equally sized bins from 0 to 50 Hz (0 to 16 Hz, 17 to 33 Hz, and 34 to 50 Hz) was then calculated [98]. Results from these frequency analyses are presented in Chapter 5S.

### 5.3. Results

Plots of reference and misplaced acceleration and angular velocity time series data for each axis and placement condition are included in Chapter 5S. Differences in those time series are summarized here as RMSEs.

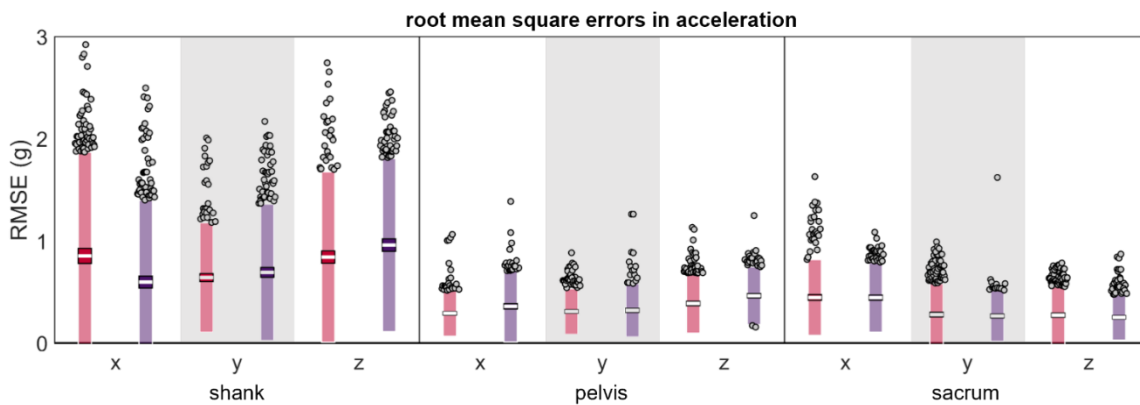
#### 5.3.1. Acceleration

Mean RMSEs for acceleration were less than 1 g across all conditions (Table 5.1; Figure 5.4). RMSEs had higher magnitudes and greater LOAs at the shank compared to the pelvis or sacrum.

**Table 5.1: Acceleration differences observed between simultaneously recorded reference and misplaced IMUs.**

location	axis	misplacement	RMSE (g)		$\Delta$  magnitude  (g)		$\Delta$  magnitude  (% reference)		$\Delta$ timing (ms)		$\Delta$ timing (% stride)	
			mean	LOA	mean	LOA	mean	LOA	mean	LOA	mean	LOA
shank	x	anterior-proximal	0.86	1.02	-0.11	3.61	0.46	37.69	-1.36	21.21	-0.19	3.12
		posterior-proximal	0.60	0.80	-0.01	3.69	0.49	41.47	0.07	15.43	0.02	2.28

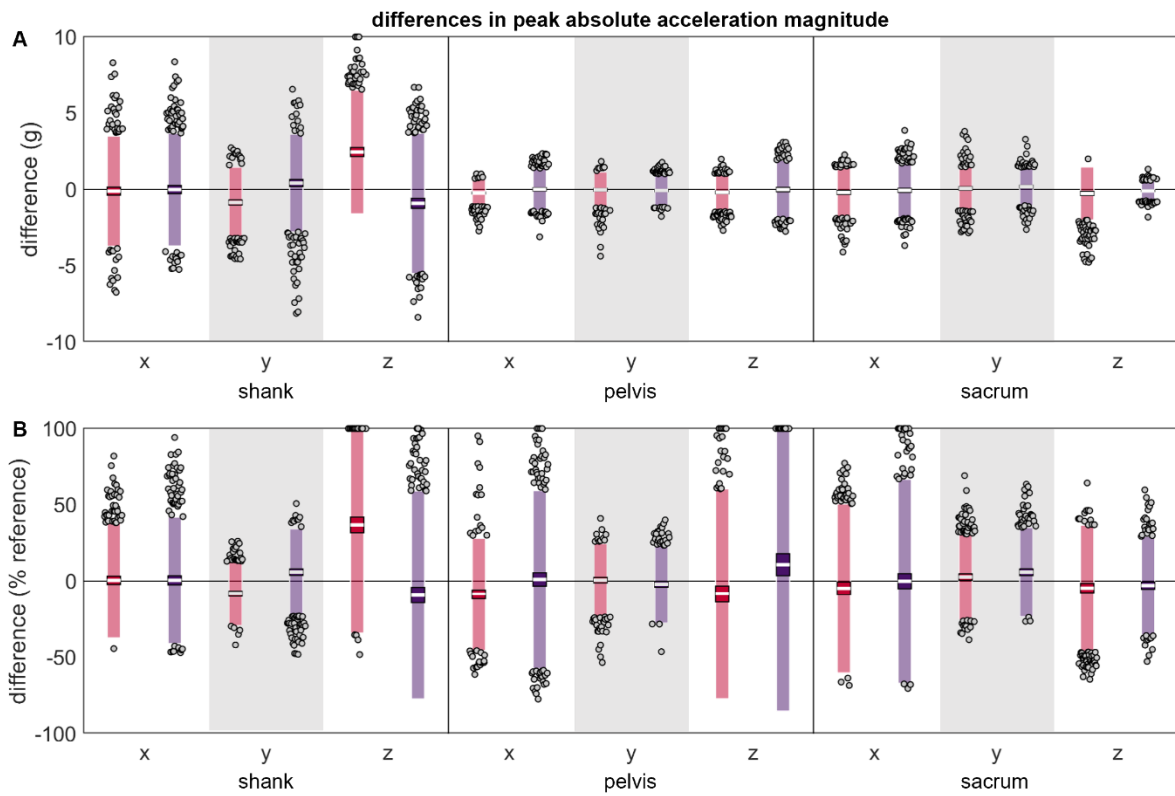
	y	anterior-proximal	0.64	0.54	-0.85	2.32	-8.22	20.91	0.76	8.52	0.11	1.20
		posterior-proximal	0.70	0.67	0.42	3.21	5.87	28.46	-0.14	13.50	-0.01	1.92
	z	anterior-proximal	0.84	0.83	2.45	4.05	36.82	70.88	1.16	20.14	0.15	2.85
		posterior-proximal	0.96	0.85	-0.91	4.60	-9.09	68.21	2.07	31.93	0.32	4.54
pelvis	x	anterior-proximal	0.29	0.22	-0.23	0.88	-8.71	36.82	3.47	32.00	0.50	4.50
		posterior-proximal	0.36	0.35	0.01	1.39	1.05	58.35	-2.33	30.34	-0.33	4.41
	y	anterior-proximal	0.31	0.22	-0.02	1.17	0.60	23.98	2.41	13.01	0.33	1.81
		posterior-proximal	0.32	0.26	-0.07	1.09	-2.09	25.48	-1.67	12.51	-0.25	1.86
	z	anterior-proximal	0.39	0.29	-0.19	1.21	-8.41	68.79	6.89	40.43	0.90	5.64
		posterior-proximal	0.47	0.28	-0.01	1.91	10.73	96.00	4.71	35.80	0.72	5.31
sacrum	x	left-proximal	0.45	0.37	-0.20	1.65	-4.83	55.43	6.57	24.52	0.97	3.74
		right-proximal	0.45	0.34	-0.06	1.77	-0.14	66.92	5.40	23.65	0.78	3.67
	y	left-proximal	0.28	0.30	0.06	1.41	2.67	28.12	0.63	15.29	0.10	2.22
		right-proximal	0.27	0.25	0.18	1.24	5.84	29.26	0.78	13.18	0.12	1.94
	z	left-proximal	0.28	0.28	-0.26	1.74	-4.69	41.44	0.94	26.11	0.11	3.83
		right-proximal	0.25	0.22	-0.09	0.65	-3.01	32.23	2.95	29.38	0.41	4.19



**Figure 5.4: Root mean square error (RMSE) between  $a$  measured simultaneously by the reference IMU and the misplaced IMU (red represents anterior- or left-proximal misplacement; purple represents posterior- or right-proximal misplacement). The white line represents the RMSE mean across all trials, the dark-colored box represents the confidence interval about the mean ( $\pm 1.96$  SEM), the light-colored box represents the limits of agreement ( $\pm 1.96$  SD), and the grey dots represent trials falling outside the limits of agreement (with trials falling outside the axis limit plotted at the limit).**

The differences between the reference and misplaced absolute acceleration magnitudes were generally zero-centered (Table 5.1 and Figure 5.5). The shank y and z axes were exceptions and exhibited systematic changes in the direction of error based on the misplacement location, with positive differences indicating the misplaced acceleration peak had a larger magnitude than the reference and negative errors indicating the misplaced acceleration peak had a smaller magnitude than the reference.

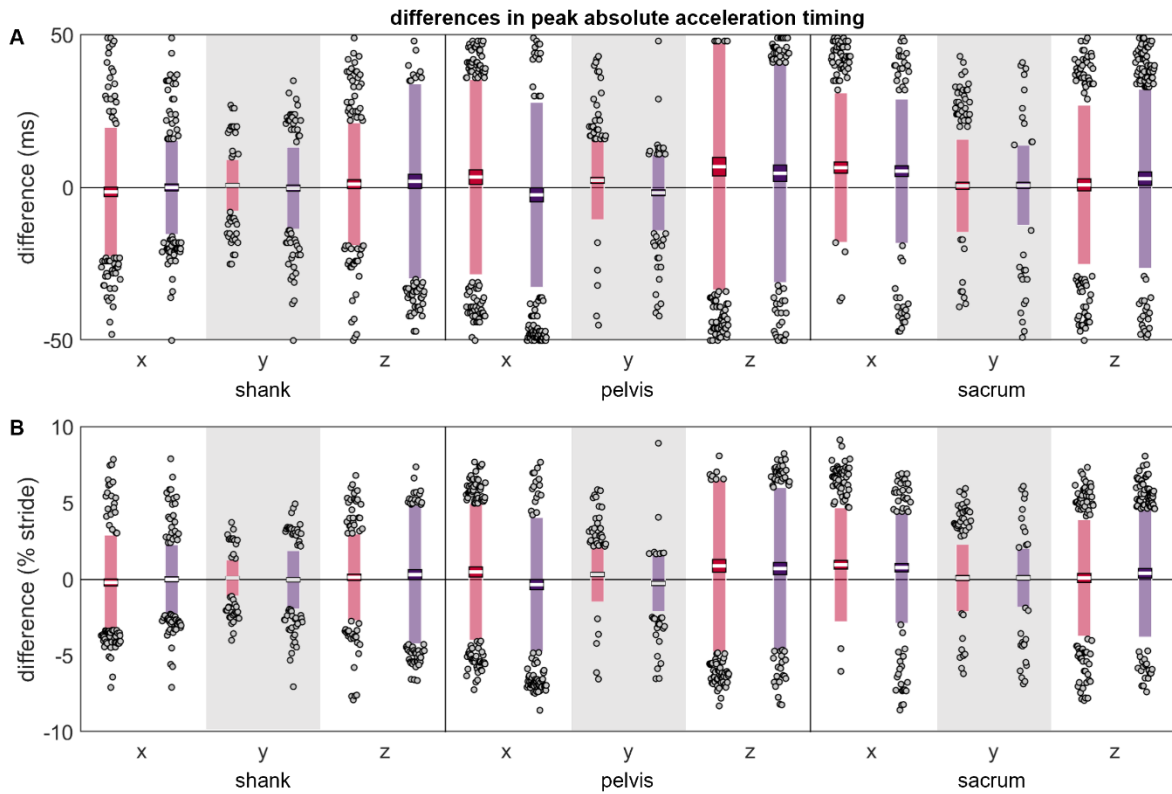
Pelvis x and z axes also tended to show systematic changes based on misplacement location; however, this difference did not appear large unless magnitudes were normalized by the reference magnitude. The LOAs for the differences between the reference and misplaced absolute acceleration magnitudes were greater for the shank compared to the pelvis or sacrum. Due to the larger magnitude of peaks observed at the shank, however, when differences were normalized to the magnitude of the reference peak, LOAs were more comparable between the shank, pelvis, and sacrum.



**Figure 5.5: (A) absolute and (B) normalized differences between peak  $|a|$  magnitudes measured simultaneously by the reference IMU and the misplaced IMU (red represents anterior- or left-proximal misplacement; purple represents posterior- or right-proximal misplacement). The white line represents the mean observed difference across trials (bias), the dark-colored box represents the confidence interval about the mean ( $\pm 1.96$  SEM), the light-colored box represents the limits of agreement ( $\pm 1.96$  SD), and the grey dots represent trials falling outside the limits of agreement (with trials falling outside the axis limits plotted at the limit). Positive differences indicate the misplaced  $|a|$  magnitude was greater than the reference  $|a|$  magnitude while negative differences indicate the misplaced  $|a|$  magnitude was less than the reference  $|a|$  magnitude.**

The differences between the reference and misplaced absolute acceleration timings were generally zero-centered (Table 5.1 and Figure 5.6). The pelvis x and y axes were exceptions and

exhibited systematic changes in the direction of error based on the misplacement location, with positive differences indicating that the misplaced IMU peak occurred after the reference peak, and negative differences indicating the misplaced IMU peak occurred before the reference peak. The pelvis z axis and sacrum x axis' misplaced IMU tended to have systematically later peaks than the reference IMU. The magnitudes and LOAs of the differences were generally comparable across the shank, pelvis, and sacrum, both in absolute terms and when normalized by stride duration.



**Figure 5.6: (A) absolute and (B) normalized differences between peak  $|a|$  timings measured simultaneously by the reference IMU and the misplaced IMU (red represents anterior- or left-proximal misplacement; purple represents posterior- or right-proximal misplacement). The white line represents the mean observed difference across trials (bias), the dark-colored box represents the confidence interval about the mean ( $\pm 1.96$  SEM), the light-colored box represents the limits of agreement ( $\pm 1.96$  SD), and the grey dots represent trials falling outside the limits of agreement (with trials falling outside the axis limits plotted at the limit). Positive differences indicate the misplaced  $|a|$  peak occurred after the reference peak while negative differences indicate the misplaced  $|a|$  peak occurred before the reference peak.**

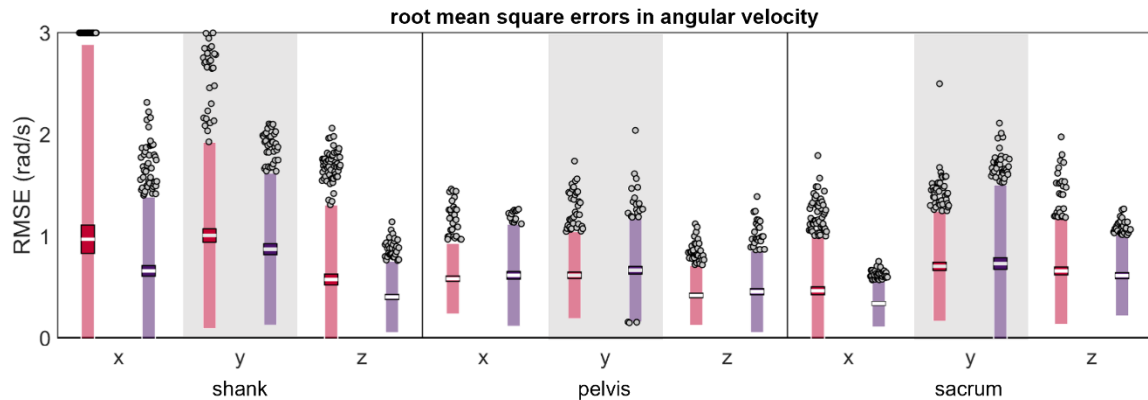


### 5.3.2. Angular velocity

Mean RMSEs were less than 1 rad/s across all conditions (except the shank anterior-proximal y axis which was 1.01 rad/s) (Table 5.2 and Figure 5.7); however, RMSEs tended to have higher magnitudes and greater LOAs at the shank than the pelvis or sacrum.

**Table 5.2: Angular velocity differences observed between simultaneously recorded reference and misplaced IMUs.**

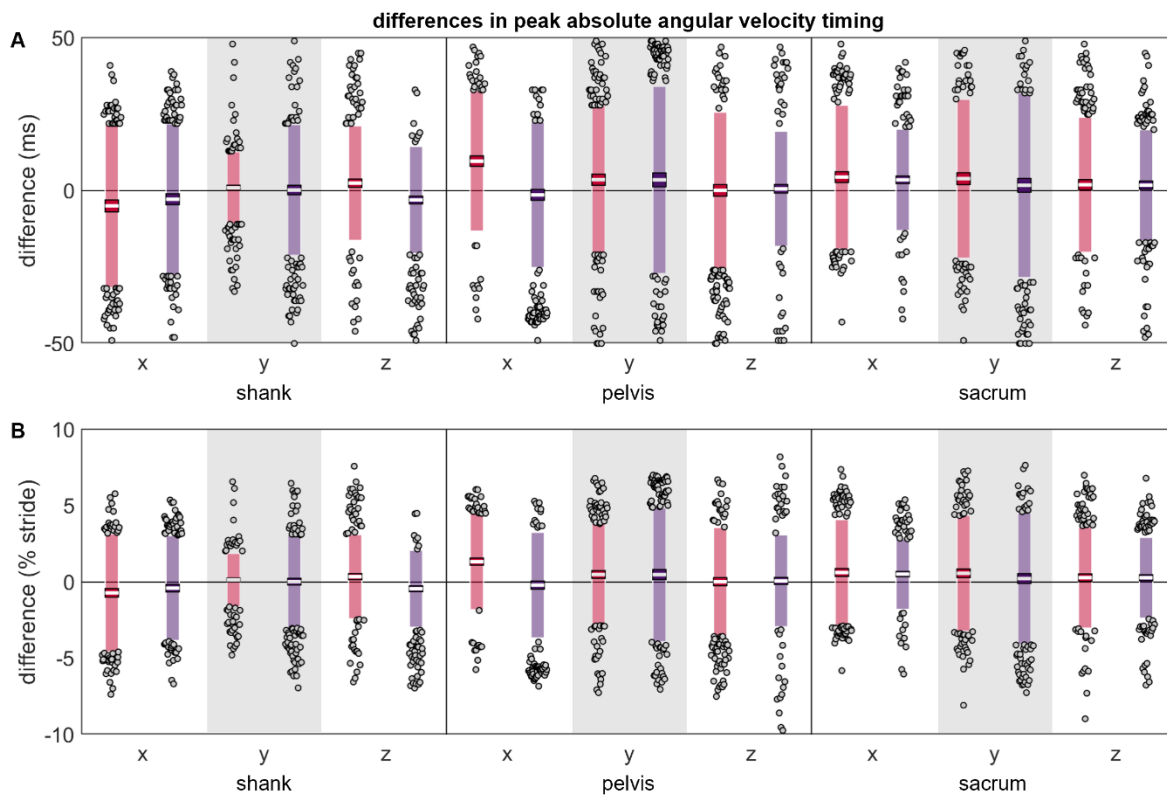
location	axis	misplacement	RMSE (rad/s)		$\Delta$  magnitude  (rad/s)		$\Delta$  magnitude  (% reference)		$\Delta$ timing (ms)		$\Delta$ timing (% stride)	
			mean	LOA	mean	LOA	mean	LOA	mean	LOA	mean	LOA
shank	x	anterior-proximal	0.97	1.92	0.20	4.29	0.21	54.21	-4.98	26.75	-0.71	3.86
		posterior-proximal	0.66	0.73	-0.16	4.53	0.56	49.71	-2.80	24.72	-0.40	3.44
	y	anterior-proximal	1.01	0.92	2.48	6.10	22.07	45.47	0.96	11.81	0.14	1.74
		posterior-proximal	0.87	0.75	0.10	4.86	-0.88	51.02	0.23	21.44	0.03	3.05
	z	anterior-proximal	0.58	0.73	-0.25	1.73	-4.57	30.68	2.48	18.75	0.36	2.76
		posterior-proximal	0.40	0.35	-0.11	0.85	-1.27	11.71	-3.04	17.54	-0.44	2.53
pelvis	x	anterior-proximal	0.58	0.35	-1.16	1.80	-56.34	89.81	9.68	22.94	1.34	3.16
		posterior-proximal	0.62	0.50	0.00	2.88	-0.69	130.63	-1.36	23.81	-0.20	3.47
	y	anterior-proximal	0.62	0.43	0.24	1.88	5.59	45.30	3.60	24.33	0.50	3.33
		posterior-proximal	0.67	0.51	-1.15	2.05	-27.89	51.56	3.53	30.64	0.49	4.39
	z	anterior-proximal	0.42	0.30	0.14	1.89	6.07	107.43	0.10	25.62	0.03	3.55
		posterior-proximal	0.46	0.40	0.35	1.78	30.00	125.17	0.62	18.87	0.08	3.01
sacrum	x	left-proximal	0.47	0.52	-0.06	1.83	-3.95	66.85	4.45	23.56	0.63	3.47
		right-proximal	0.34	0.23	0.05	0.97	4.96	49.84	3.56	16.61	0.52	2.31
	y	left-proximal	0.71	0.54	-0.02	1.68	0.46	36.49	3.93	25.99	0.57	3.79
		right-proximal	0.73	0.77	-0.79	2.32	-20.48	60.98	1.76	30.20	0.23	4.32
	z	left-proximal	0.66	0.53	-0.12	3.17	3.26	113.30	1.93	22.12	0.29	3.30
		right-proximal	0.61	0.40	0.03	3.17	-0.87	153.98	1.81	18.15	0.28	2.65



**Figure 5.7: Root mean square error (RMSE) between  $\omega$  measured simultaneously by the reference IMU and the misplaced IMU (red represents anterior- or left-proximal misplacement; purple represents posterior- or right-proximal misplacement). The**

white line represents the RMSE mean across all trials, the dark-colored box represents the confidence interval about the mean ( $\pm 1.96$  SEM), the light-colored box represents the limits of agreement ( $\pm 1.96$  SD), and the grey dots represent trials falling outside the limits of agreement (with trials falling outside the axis limit plotted at the limit).

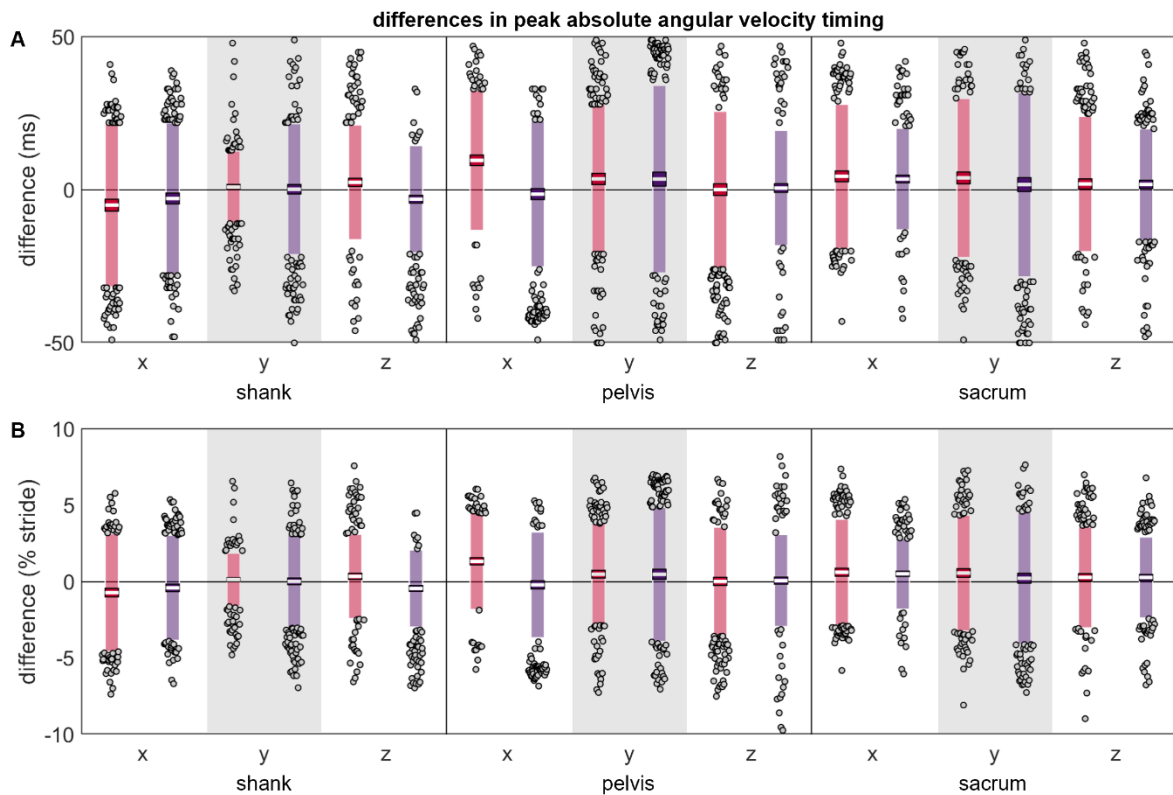
Although most absolute angular velocity peak magnitude differences were zero-centered (Table 5.2 and Figure 5.8), the shank y axis, all pelvis axes, and the sacrum y axis exhibited systematic changes in the direction/magnitude of differences based on the misplacement location, with positive differences indicating that the misplaced angular velocity peak had a larger magnitude than the reference and negative errors indicating that the misplaced angular velocity peak had a smaller magnitude than the reference. Again, before normalization, differences appeared larger at the shank than the pelvis or sacrum, but after normalizing to the reference, the peak magnitude differences appeared more similar across the shank, pelvis, and sacrum.



**Figure 5.8: (A) absolute and (B) normalized differences between peak  $|\omega|$  magnitudes measured simultaneously by the reference IMU and the misplaced IMU (red represents anterior- or left-proximal misplacement; purple represents posterior- or right-proximal misplacement). The white line represents the mean observed difference across trials (bias), the dark-colored box represents the confidence interval about the mean ( $\pm 1.96$  SEM), the light-colored box represents the limits of agreement ( $\pm$**

1.96 SD), and the grey dots represent trials falling outside the limits of agreement (with trials falling outside the axis limits plotted at the limit). Positive differences indicate the misplaced  $|\omega|$  magnitude was greater than the reference  $|\omega|$  magnitude while negative differences indicate the misplaced  $|\omega|$  magnitude was less than the reference  $|\omega|$  magnitude.

Most differences in the timing of the angular velocity peaks were not zero-centered (Table 5.2 and Figure 5.9). The timing differences at the pelvis and sacrum suggested that angular velocity peaks tended to occur later in time for the misplaced IMU than the reference IMU (positive differences). In contrast, three of six axis-placement conditions at the shank suggested that the shank angular velocity peaks tended to occur earlier in time for the misplaced IMU. The magnitudes and LOAs were generally comparable across the shank, pelvis, and sacrum, both in absolute and relative terms.



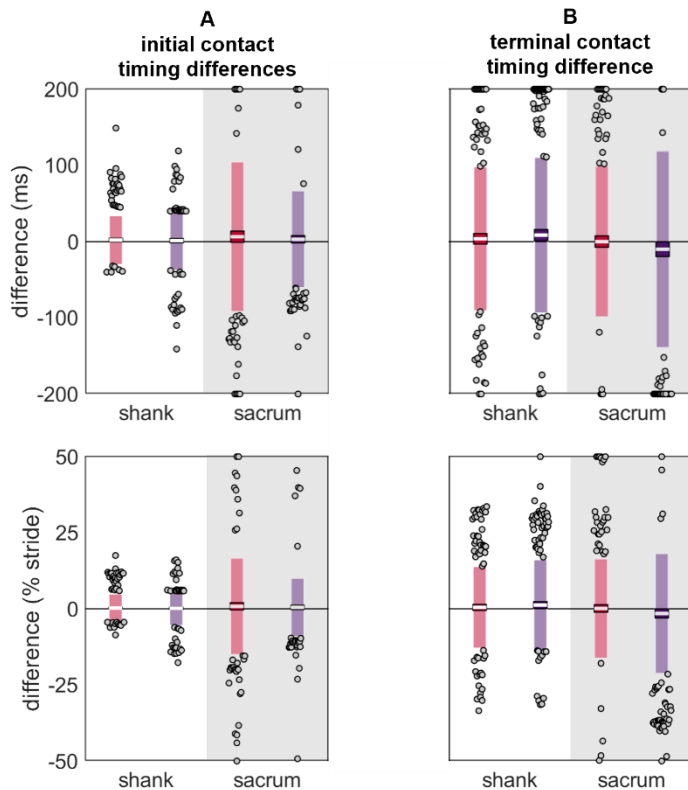
**Figure 5.9: (A) absolute and (B) normalized differences between peak  $|\omega|$  timings measured simultaneously by the reference IMU and the misplaced IMU (red represents anterior- or left-proximal misplacement; purple represents posterior- or right-proximal misplacement). The white line represents the mean observed difference across trials (bias), the dark-colored box represents the confidence interval about the mean ( $\pm 1.96$  SEM), the light-colored box represents the limits of agreement ( $\pm 1.96$  SD), and the grey dots represent trials falling outside the limits of agreement (with trials falling outside the axis limits plotted at the limit). Positive differences indicate the misplaced  $|\omega|$  peak occurred after the reference peak while negative differences indicate the misplaced  $|\omega|$  peak occurred before the reference peak.**

### 5.3.3. Estimated outcome variables

Despite the shank exhibiting larger RMSEs and peak differences, the initial contact differences were lower at the shank than the sacrum (Table 5.3 and Figure 5.10). In contrast, the terminal contact differences were comparable between the shank and sacrum.

**Table 5.3: Differences between contact times estimated from simultaneously recorded reference and misplaced IMUs.**

location	misplacement	$\Delta$ initial contact (ms)		$\Delta$ initial contact (% stride)		$\Delta$ terminal contact (ms)		$\Delta$ terminal contact (% stride)	
		mean	LOA	mean	LOA	mean	LOA	mean	LOA
shank	anterior-proximal	2.02	32.21	0.30	4.67	3.68	94.61	0.49	13.47
	posterior-proximal	1.28	38.62	0.18	5.68	8.43	102.29	1.25	14.97
sacrum	left-proximal	6.42	98.50	0.81	15.88	0.03	98.96	0.15	16.35
	right-proximal	3.10	63.70	0.49	9.61	-10.08	129.21	-1.51	19.71



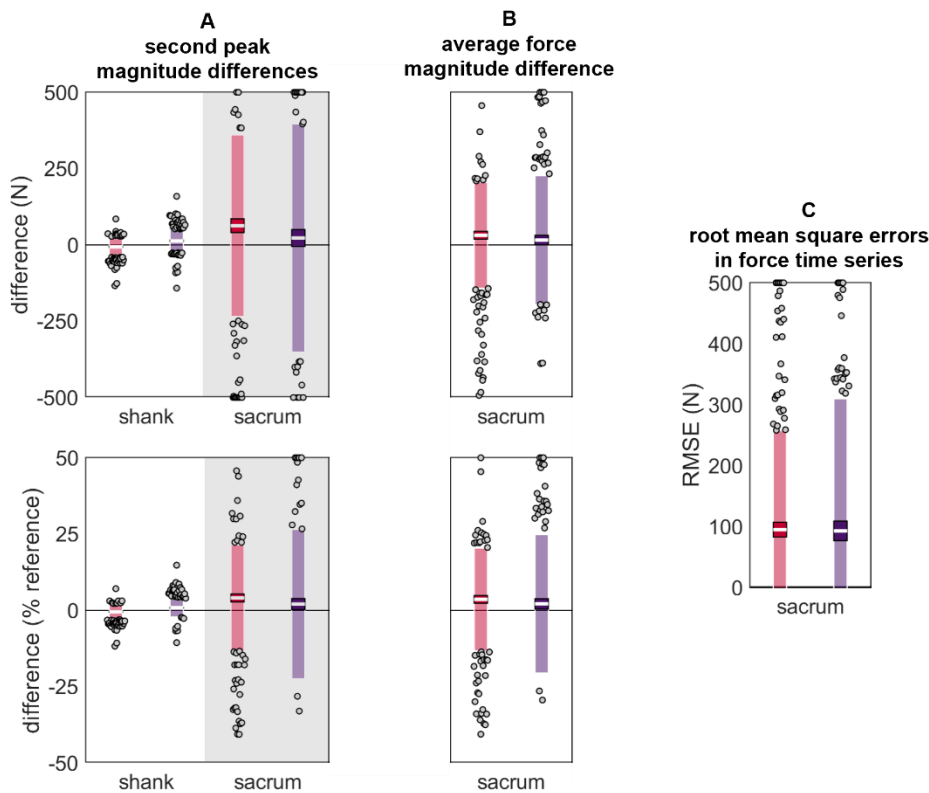
**Figure 5.10: Differences in (A) initial contact and (B) terminal contact gait event timings estimated using data from the reference IMU and the misplaced IMU (red represents anterior- or left-proximal misplacement; purple represents posterior- or right-proximal misplacement). The white line represents the mean observed difference across trials (bias), the dark-colored box represents the confidence interval about the mean ( $\pm 1.96$  SEM), the light-colored box represents the limits of agreement ( $\pm 1.96$  SD), and the grey dots represent trials falling outside the limits of agreement (with trials falling outside the axis limits plotted at the limit). Positive differences indicate the gait event estimated with the misplaced IMU occurred after the reference**

gait event while negative differences indicate the gait event estimated with the misplaced IMU occurred before the reference gait event.

A similar trend was observed in the estimated vertical ground reaction forces, with lower differences in estimated second peak at the shank than the sacrum (Table 5.4; Figure 5.11).

**Table 5.4: Differences between ground reaction forces estimated from simultaneously recorded reference and misplaced IMUs.**

location	misplacement	$\Delta$ second peak (N)		$\Delta$ second peak (% reference)		$\Delta$ average force (N)		$\Delta$ average force (% reference)		time series RMSE (N)	
		mean	LOA	mean	LOA	LOA	LOA	mean	LOA	mean	LOA
shank	anterior-proximal	-6.22	33.73	-0.45	2.60						
	posterior-proximal	13.23	38.18	0.85	3.03						
sacrum	left-proximal	62.99	298.04	4.07	17.37	31.48	173.58	3.59	16.84	95.06	162.42
	right-proximal	22.45	373.92	2.05	24.48	15.95	211.19	2.15	22.71	93.12	216.66



**Figure 5.11: Differences in (A) vertical ground reaction force second peak magnitude, (B) average vertical ground reaction force during stance, and (C) vertical ground reaction force time series during stance.** Estimated using data from the reference IMU and the misplaced IMU (red represents anterior- or left-proximal misplacement; purple represents posterior- or right-proximal misplacement). The white line represents the mean observed difference across trials (bias), the dark-colored box represents the confidence interval about the mean ( $\pm 1.96$  SEM), the light-colored box represents the limits of agreement ( $\pm 1.96$  SD), and the grey dots represent trials falling outside the limits of agreement (with trials falling outside the axis limits

plotted at the limit). For (A) and (B) positive differences indicate the misplaced IMU estimated a higher magnitude than the reference while negative values indicate the misplaced IMU estimated a lower magnitude than the reference.

#### 5.4. Discussion

To characterize the extent of changes that occur when an IMU is misplaced or moved, the current paper compared signals from a reference IMU ‘correctly’ placed on the shank, pelvis, or sacrum and an IMU ‘misplaced’ 0.05 m away (simulating a ‘worst-case’ misplacement/movement scenario). Overall, the time domain signals of the reference and misplaced IMUs exhibited the same general patterns (Chapter 5S), as evidenced by their low root mean square errors (RMSEs) ( $\leq 1$  g and  $\sim 1$  rad/s). Examining another commonly investigated feature of IMU signals—the peak magnitudes and timings—revealed that differences were generally small and zero-centered but could reach up to  $2.45 \pm 4.05$  g ( $36.82 \pm 70.88\%$  reference; mean  $\pm$  limits of agreement; LOA),  $2.48 \pm 6.10$  rad/s ( $22.07 \pm 45.47\%$  reference), and  $9.68 \pm 22.94$  ms ( $1.34 \pm 3.16\%$  stride duration) depending on the axis and direction of misplacement. Altogether, these data show that IMU users must be cautious about IMU misplacement and movement when collecting and interpreting data.

Acceleration and angular velocity magnitude errors were larger at the shank than the pelvis or sacrum. When normalized by the reference magnitude, however, these errors were mitigated. Thus, the relatively large shank errors observed before normalization likely reflect the larger magnitude accelerations and angular velocities observed at the shank during running (cf. the pelvis or sacrum; see time series in Chapter 5S).

Although we are not aware of any previous investigations of IMU misplacement at the pelvis or sacrum during running, Sara et al. [315] and Zhang et al. [238] have previously reported the effects of IMU misplacement on proximal-distal (y-axis) acceleration magnitudes at the shank. Sara et al. placed a reference IMU on the medial malleolus and ‘misplaced’ another IMU 0.02 m proximal. They found that

peak proximal–distal accelerations during fast (but not typical or slow) running were  $\sim 1.23$  g (or  $\sim 13.00\%$ ) higher for the misplaced IMU than the reference IMU (estimated from their Figure 1A, using <https://apps.automeris.io/wpd/>; accessed on 7 November 2023). Using a similar approach, Zhang et al. positioned a reference IMU at the lateral malleolus and compared it to an IMU on the anteromedial distal tibia. They found that peak proximal accelerations were  $0.70$  g (or  $8.65\%$ ) greater for the anteromedial distal tibia than the lateral malleolus (calculated from their Table 1). Thus, both Sara et al. and Zhang et al. reported that a small proximal shift (coupled with a change from lateral to medial in Zhang et al.) increased observed proximal–distal acceleration peaks. These differences are inconsistent with other research demonstrating that large proximal shifts in location decrease acceleration peaks [319, 320] but are partially consistent with the current results: we found that a  $0.05$  m anterior-proximal misplacement caused a  $-0.85$  g ( $-8.22\%$ ) difference while a  $0.05$  m posterior-proximal misplacement caused a  $0.42$  g ( $5.87\%$ ) difference. The increase observed with a posterior-proximal placement may be due to greater movement of the IMU relative to the anatomical segment that it is measuring (soft tissue artefact, as the IMU sits more on the muscle). Conversely, the decrease observed with an anterior-proximal placement may be due to less soft tissue artefact (as the IMU sits more on the anterior aspect of the tibia) and is more consistent with the previous literature [319, 320] and with the pattern of results Sara et al. originally predicted.

These observed differences may have downstream effects when using IMU signals to estimate other outcome metrics (e.g., spatiotemporal events, ground reaction forces, running speed, segment and joint kinematics, etc.). To investigate this, we used the reference and misplaced IMU signals to estimate initial contact, terminal contact, vertical ground reaction force second peak magnitude, average vertical ground reaction force during stance, and vertical ground reaction force time series, and then quantified the differences between these estimates. This investigation revealed that, although the overall differences between signal magnitudes may be small (as evidenced by the RMSEs), even these

small differences can cause large downstream effects: At the sacrum, the mean difference in estimated terminal contact reached  $-10.08 \pm 129.21$  ms ( $-1.51 \pm 19.71\%$  stride duration), while the mean difference in estimated vertical ground reaction force second peak magnitude reached  $62.99 \pm 298.04$  N (or  $4.07 \pm 17.37\%$  reference). At the shank, differences in terminal contact times had similar magnitude errors ( $8.43 \pm 102.29$  ms or  $1.25 \pm 14.97\%$  stride), but initial contact times ( $2.02 \pm 32.21$  ms or  $0.30 \pm 4.67\%$  stride) and vertical forces  $13.23 \pm 38.18$  N ( $0.85 \pm 3.03\%$  reference) were less affected by misplacement, suggesting that shank-based estimates may be more robust to misplacement. To our knowledge Tan et al. [312] provide the only comparable results. They investigated misplacements of up to  $\pm 0.10$  m at one or more of eight simulated IMU placement locations (feet, shanks, thighs, sacrum, and trunk) and then used the accelerations and angular velocities of the eight simulated IMUs to estimate vertical ground reaction forces during walking. They found that the misplacement of a single IMU resulted in estimated force differences of up to 2.0%, comparable to the differences observed here ( $-0.45, 0.85, 2.05, \text{ and } 4.07\%$ ). It seems likely that even with their very large 0.10 m misplacement, the use of multiple IMUs to estimate force stabilized the estimate (i.e., seven of their eight simulated IMU signals were still unaffected). When Tan et al. misplaced all eight IMUs, they observed mean differences up to 6.0%, higher than even the largest differences observed here. This is likely a function of the large misplacement they used.

Overall, the differences observed both here and in previous work [315, 238, 312] underscore the importance of placing IMUs correctly and preventing their movement throughout data collection. Ruder et al. [313] previously demonstrated that IMUs placed by untrained participants have lower validity and inter-session reliability than IMUs placed by trained experimenters. Thus, care should be taken when deploying IMUs in-field and proper training should be provided to end users. Attempts should also be made to reduce IMU movement. Before executing the current study, we piloted our IMU fixation system by quantifying IMU movement across multiple 5.63 km runs. These runs were designed to elicit the



greatest possible movement (e.g., included dynamic warm-ups, sprinting, 'strides', 'Fartlek', moving through extreme ranges of motion, etc.). Even under these 'worst case' conditions, maximum displacements at the shank were only 0.0049 m proximal and 0.0027 m posterior with 0.07 rad of rotation, while maximum displacements at the sacrum were only 0.0068 m proximal and 0.0004 m left with 0 rad of rotation. These observed displacements are far lower than those studied here and are likely associated with smaller signal differences; however, it is unclear from the present results whether/how signal differences scale with the size of misplacement (i.e., error magnitudes may be non-linear).

The current results may not represent other fixation systems or populations. Previous work by Johnson et al. [321] demonstrates that fixation method can systematically alter IMU signals, with a looser fixation resulting in higher shank accelerations. The current misplacement results, and the IMU movements reported for the fixation system we used, were collected from a relatively lean sample. In a sample with greater adiposity there may be greater soft tissue artefact that alters both the IMU signal and how it is affected by misplacement [322]. Anecdotally, participants with greater adiposity may also induce greater IMU movement across data collection. The two participants that were eliminated from the current study due to the movement of an IMU were both outliers in terms of mass and body mass index. To ensure IMUs can be used to collect high-quality data from all participants, future research should characterize potential differences across participant subgroups and develop comfortable fixation systems that prevent IMU movement for all participants.

Finally, the current study did not investigate the effects of IMU rotations on IMU signals. Tan et al. [312] previously demonstrated that changes in orientation had a larger effect on IMU-derived estimates of vertical ground reaction forces during walking than even very large 0.10 m linear translations. Errors in orientation originating from IMU misplacement can be mitigated using a coordinate transformation from the 'wearable' coordinate system (WCS) to the 'segment' coordinate

system (SCS) (see Supplementary Materials Section SA). Thus, all the results presented in the current paper were derived from data expressed in an SCS. For the sake of comparison, we did, however, repeat the entire set of analyses on data in the WCS (Chapter 5S). These analyses show much higher errors for the same linear misplacements when data are processed and analyzed in the WCS rather than the SCS. Thus, we recommend using an SCS when possible. Unfortunately, using an SCS does not negate the possibility that an IMU may rotate/move across the duration of a data collection, which would still introduce error between the axis alignment at the start and end of the data collection. This type of movement was not studied here as we established coordinate systems at the start of the data collection and used a hook-and-loop attachment system that minimized movement and did not allow rotation of the IMU. To prevent IMU rotations, we recommend the use of a similar hook-and-loop fixation system or the use of double-sided tape to secure IMUs directly to a participant's skin, then tightly wrapping elastic straps over top.

## **5.5. Conclusions**

This paper provides a descriptive analysis of the effects that a 0.05 m IMU misplacement can have on acceleration and angular velocity signals during running. IMU users should characterize the magnitude of IMU misplacements and movements they expect for a specific use scenario with a specific fixation system and target population. The present results can then serve as a guide to estimate the signal differences that could be expected due to misplacement or movement. Those expected differences can then be compared to expected effect sizes to determine if IMUs will be sufficiently reliable for a given scenario.

The results from this paper suggest that signal coherence is high and differences in the frequency domain are minimal for most axes, while in the time domain, most differences are

approximately zero-centered with low bias. The limits of agreement may, however, be quite high, indicating a high degree of variability. Absolute differences are larger at the shank than at the pelvis or sacrum but are comparable when normalized by the reference magnitude. Further, the differences at the shank appear to have less of an effect on outcome variables like initial contact and ground reaction force when estimated from the shank versus the sacrum. Thus, when IMU movement or misplacement is likely, using a shank IMU may be preferable to using a sacrum IMU.

Future research should investigate smaller, incremental misplacements (more in line with those observed using the type of fixation system here) and changes in IMU orientation. To provide confidence for scenarios where participants place their own IMUs, research should also be conducted to characterize the size of misplacements across repeated placements by participants, and how much training is required to minimize those misplacements.

## **5.6. Funding**

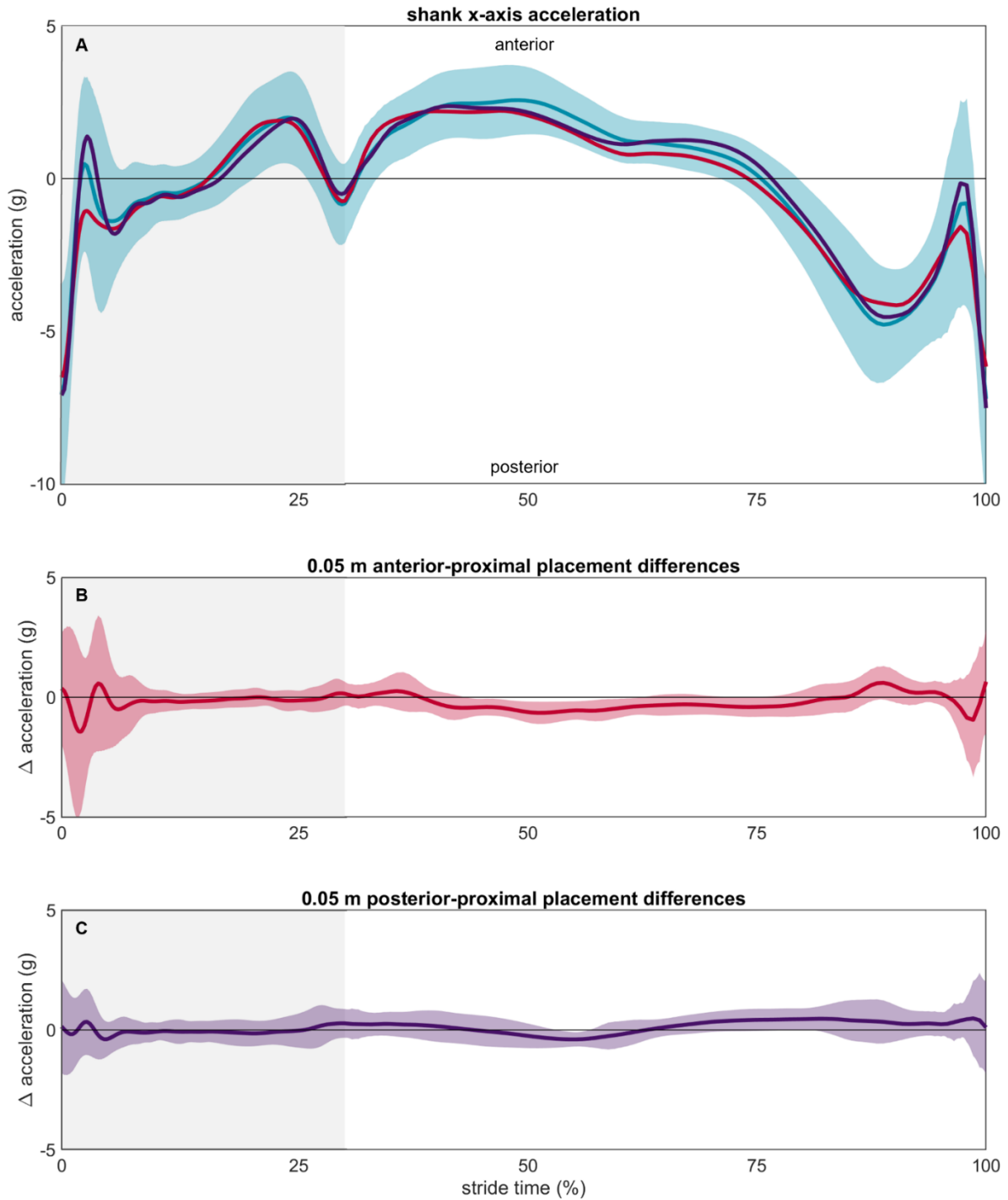
This research was funded in part by the American College of Sports Medicine Foundation Doctoral Student Research Grant (18-00671) and the Sigma Xi Scientific Research Honor Society Grant-in-Aid of Research (G2018031595024439). DK was supported by a scholarship from the Natural Sciences and Engineering Research Council of Canada (CGSD3-502358-2017), the Maury L Hull Endowed Fellowship for Musculoskeletal Biomechanics Research, and a University of California Davis Professors for the Future Fellowship.

## 5S. Supplement to Chapter 5

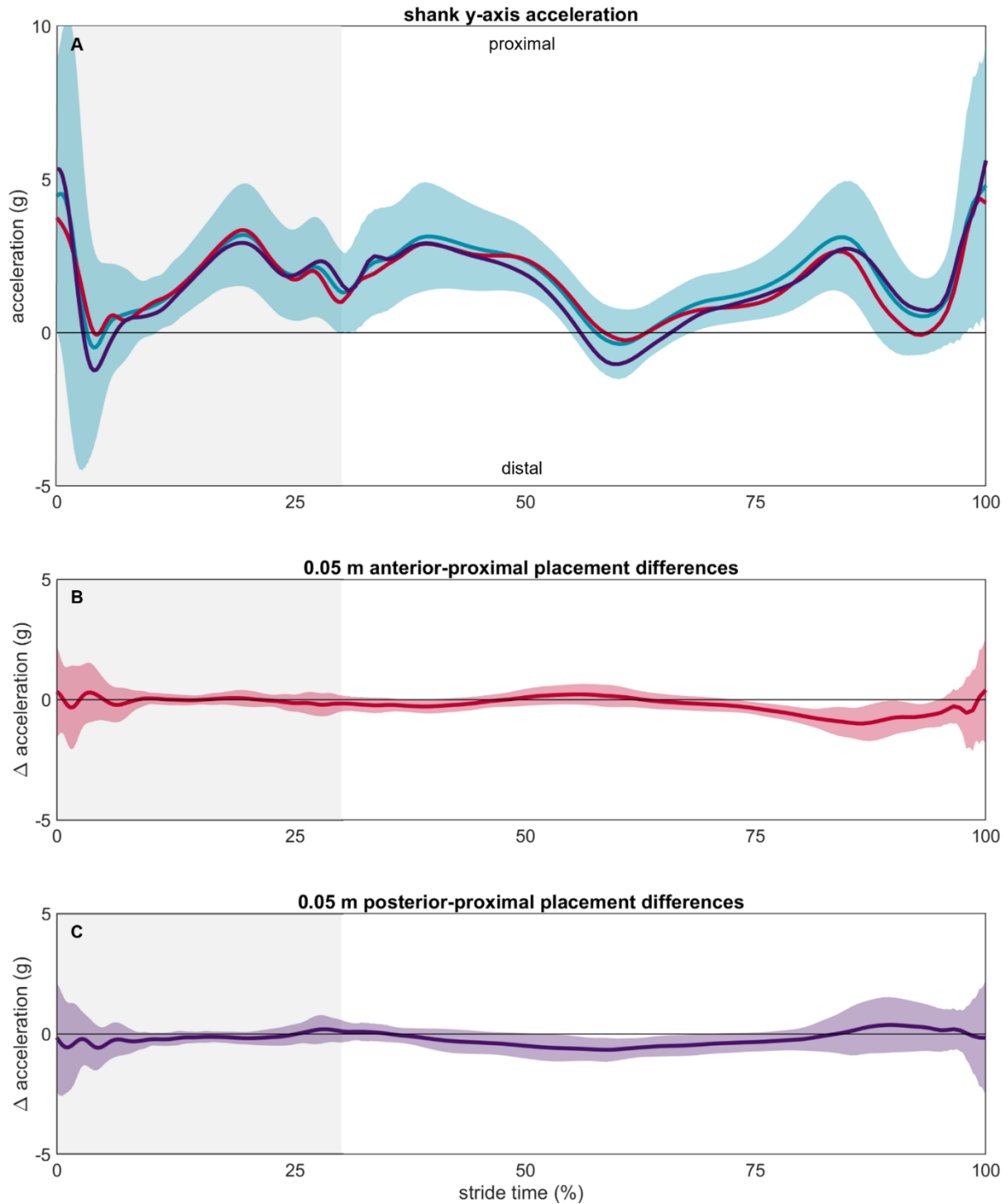
### 5S.1. Time series plots

Means and standard deviations were calculated and plotted for each axis of the reference and misplaced acceleration and angular velocity signals. Data are time-normalized such that right stance goes from 0-30% of stride and right swing (right terminal contact to right initial contact) goes from 31-100% of stride (stance was  $29.96 \pm 5.82\%$  of stride in our data).

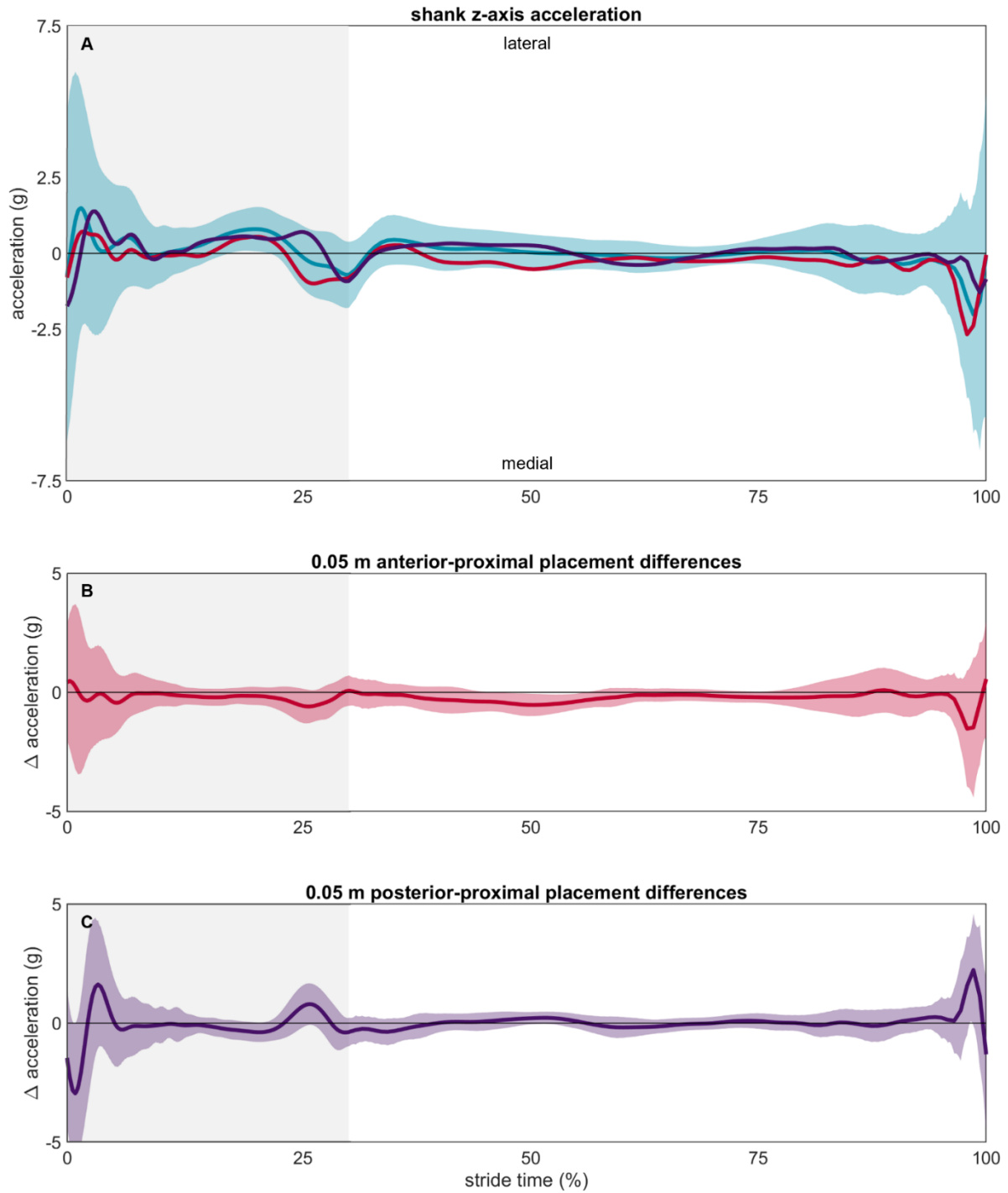
#### 5S.1.1. Shank acceleration



**Figure 55.1: Shank x-axis accelerations.** (A) measured at the reference (blue), anterior-proximal misplaced (red), and posterior-proximal (purple) IMUs. The light blue area represents  $\pm 1$  SD about the reference mean. The light grey background represents stance while the white background represents swing. Differences between the reference and (B) anterior-proximal and (C) posterior-proximal IMUs are shown in the panels below.

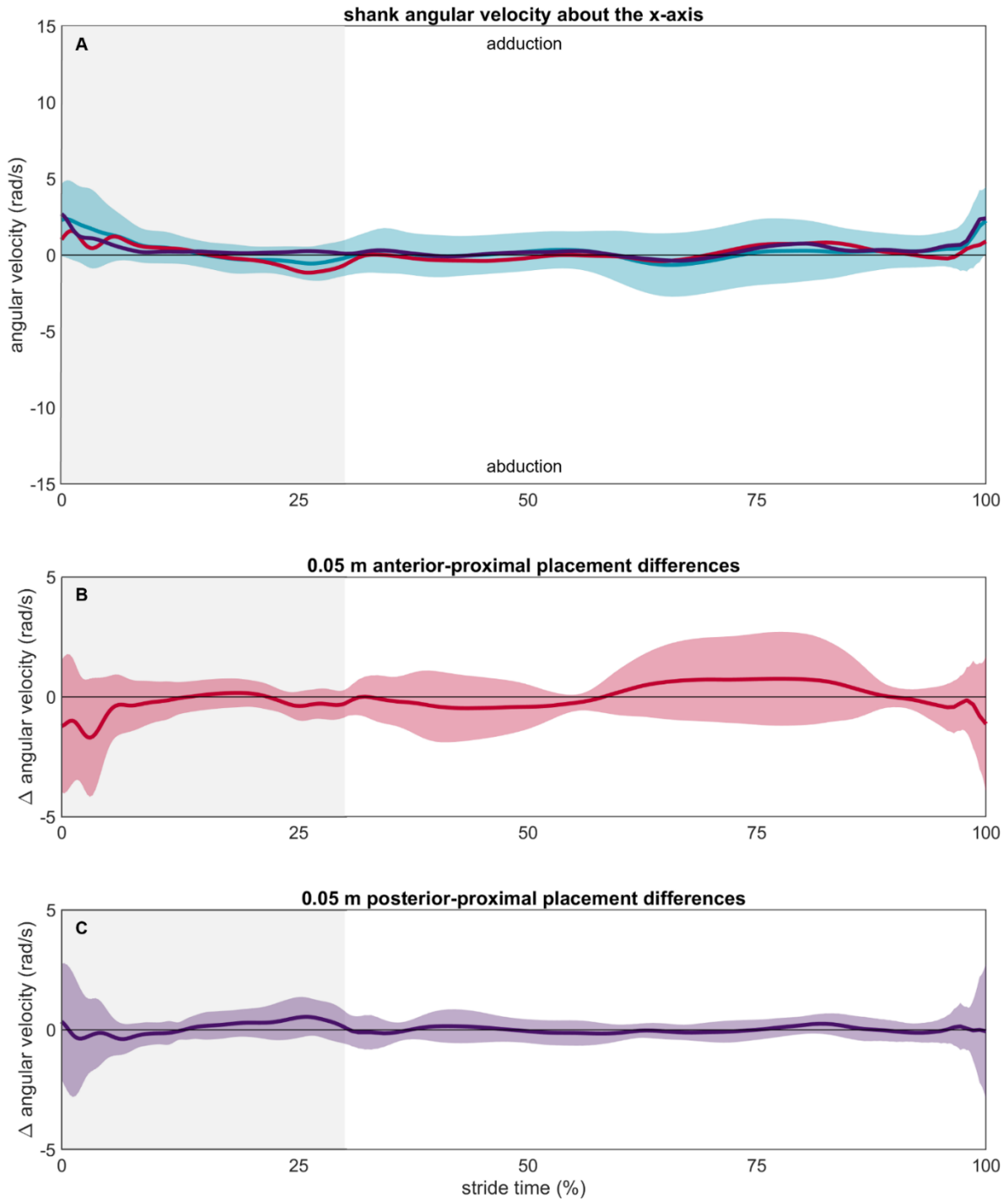


**Figure 5S.2: Shank y-axis accelerations (A)** measured at the reference (blue), anterior-proximal misplaced (red), and posterior-proximal (purple) IMUs. The light blue area represents  $\pm 1$  SD about the reference mean. The light grey background represents stance while the white background represents swing. Differences between the reference and **(B)** anterior-proximal and **(C)** posterior-proximal IMUs are shown in the panels below.



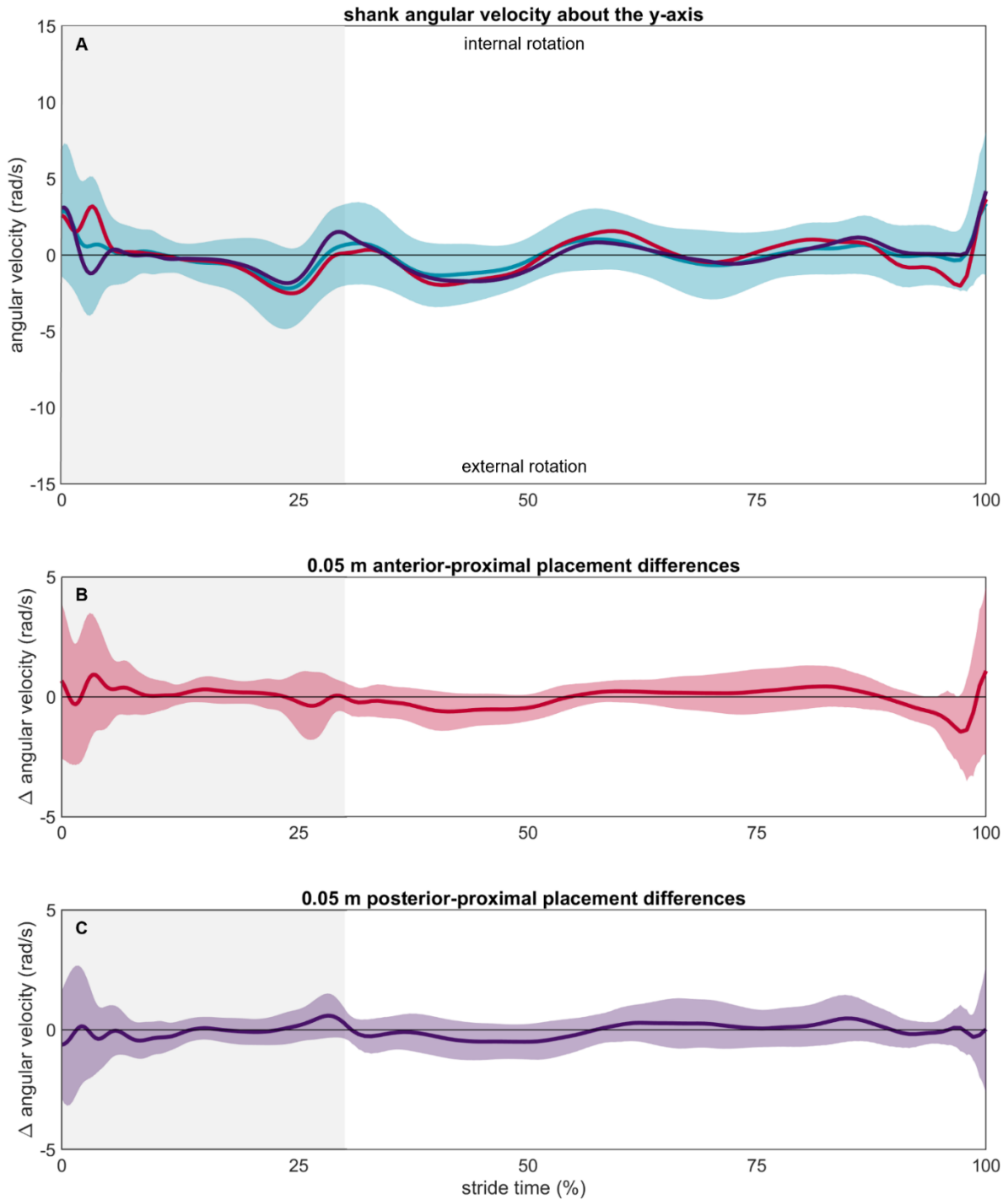
**Figure 5S.3: Shank z-axis accelerations (A)** measured at the reference (blue), anterior-proximal misplaced (red), and posterior-proximal (purple) IMUs. The light blue area represents  $\pm 1$  SD about the reference mean. The light grey background represents stance while the white background represents swing. Differences between the reference and **(B)** anterior-proximal and **(C)** posterior-proximal IMUs are shown in the panels below.

5S.1.2. Shank angular velocity

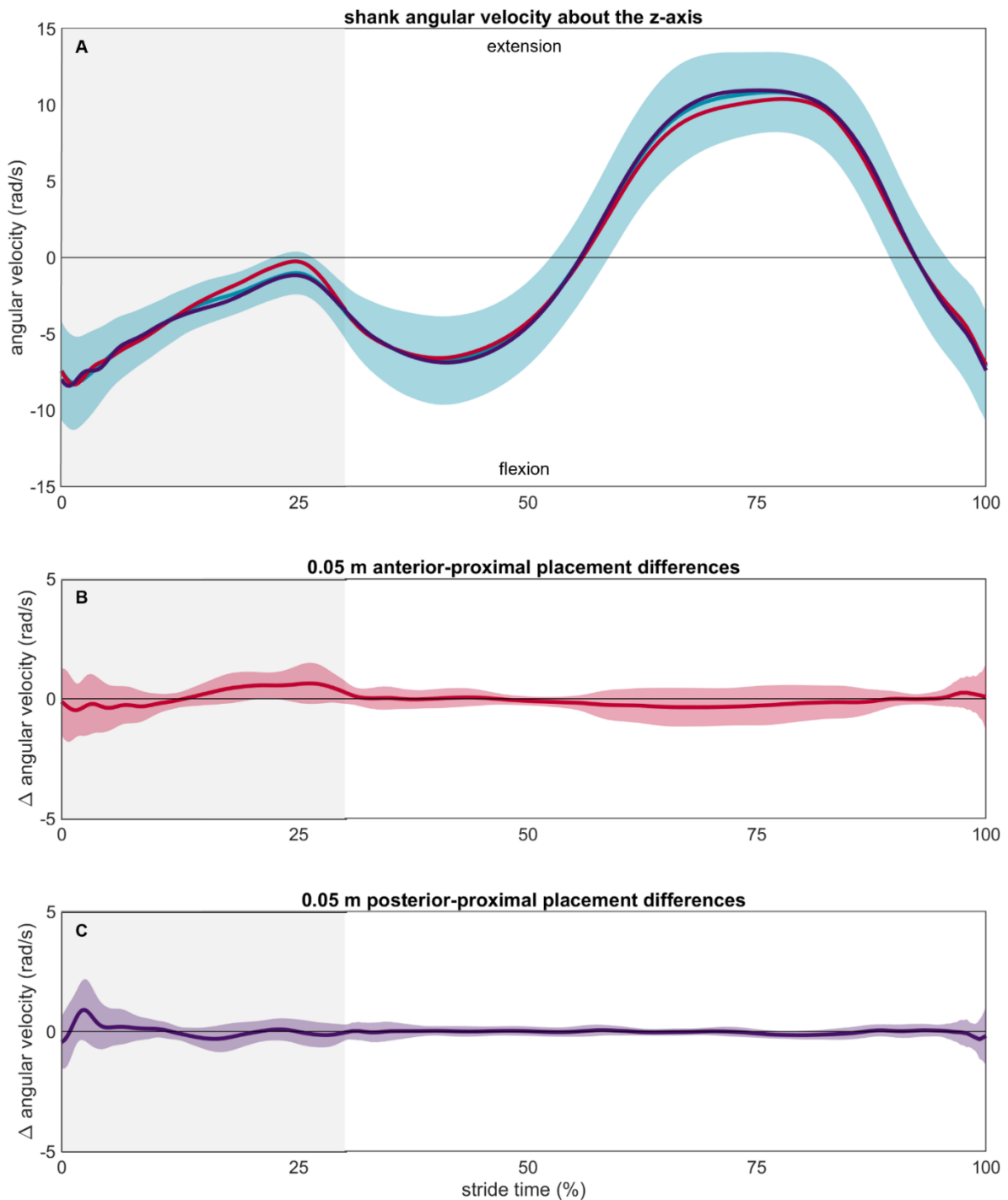


**Figure 5S.4: Shank x-axis angular velocities (A)** measured at the reference (blue), anterior-proximal misplaced (red), and posterior-proximal (purple) IMUs. The light blue area represents  $\pm 1$  SD about the reference mean. The light grey background represents stance while the white background represents swing. Differences between the reference and **(B)** anterior-proximal and **(C)** posterior-proximal IMUs are shown in the panels below.



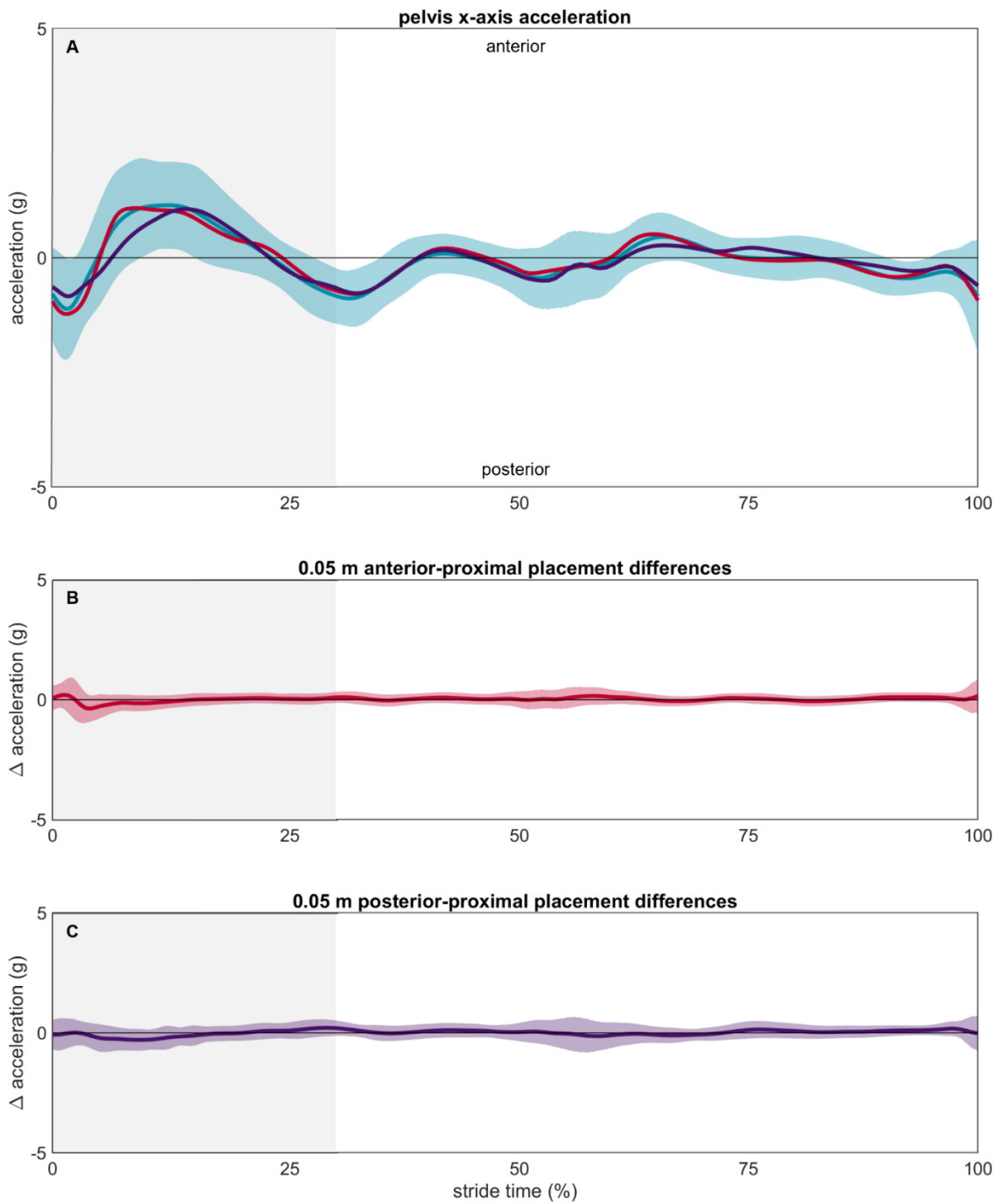


**Figure 5S.5: Shank y-axis angular velocities (A)** measured at the reference (blue), anterior-proximal misplaced (red), and posterior-proximal (purple) IMUs. The light blue area represents  $\pm 1$  SD about the reference mean. The light grey background represents stance while the white background represents swing. Differences between the reference and **(B)** anterior-proximal and **(C)** posterior-proximal IMUs are shown in the panels below.



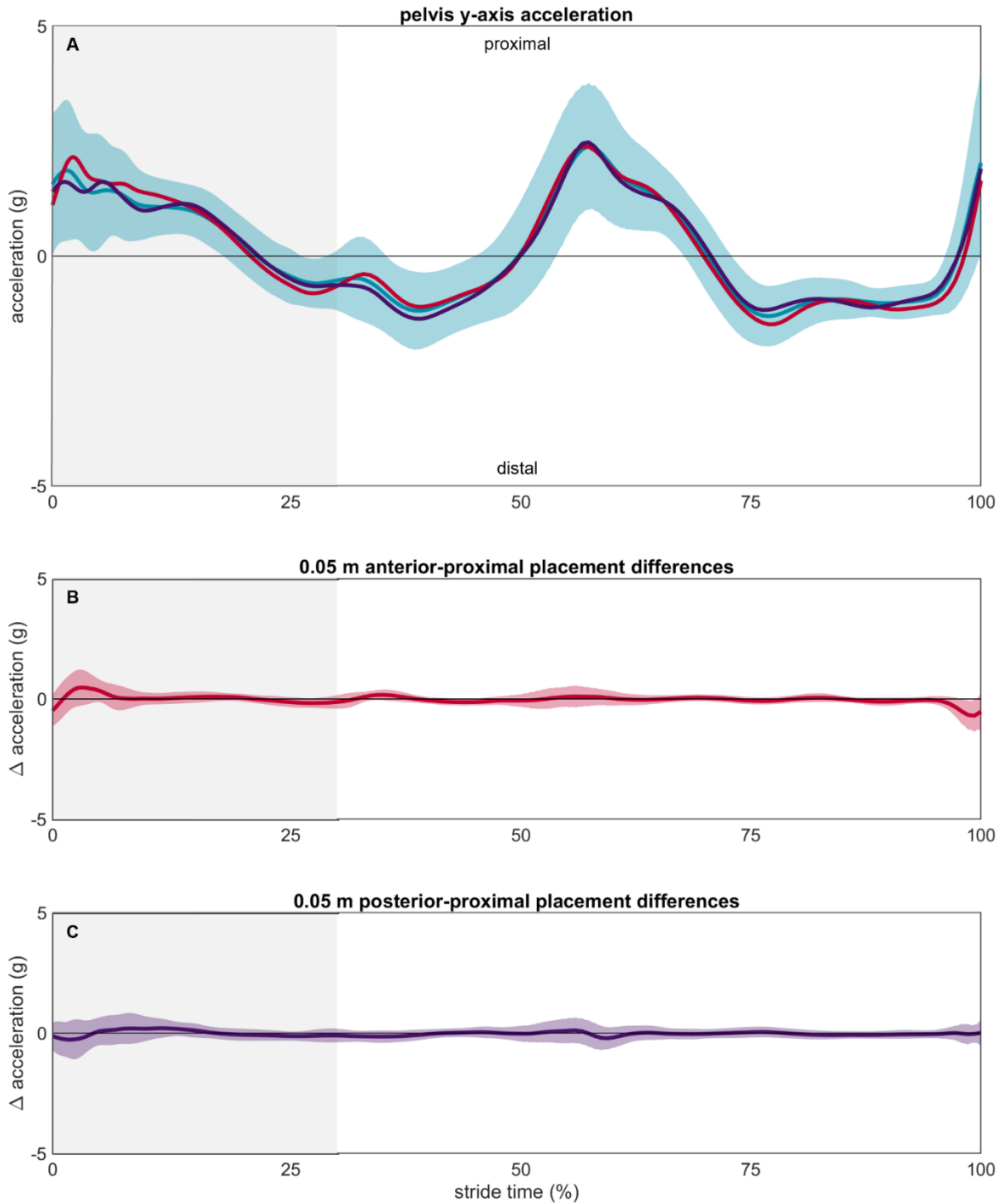
**Figure 5S.6:** Shank z-axis angular velocities (**A**) measured at the reference (blue), anterior-proximal misplaced (red), and posterior-proximal (purple) IMUs. The light blue area represents  $\pm 1$  SD about the reference mean. The light grey background represents stance while the white background represents swing. Differences between the reference and (**B**) anterior-proximal and (**C**) posterior-proximal IMUs are shown in the panels below.

### 5S.1.3. Pelvis acceleration

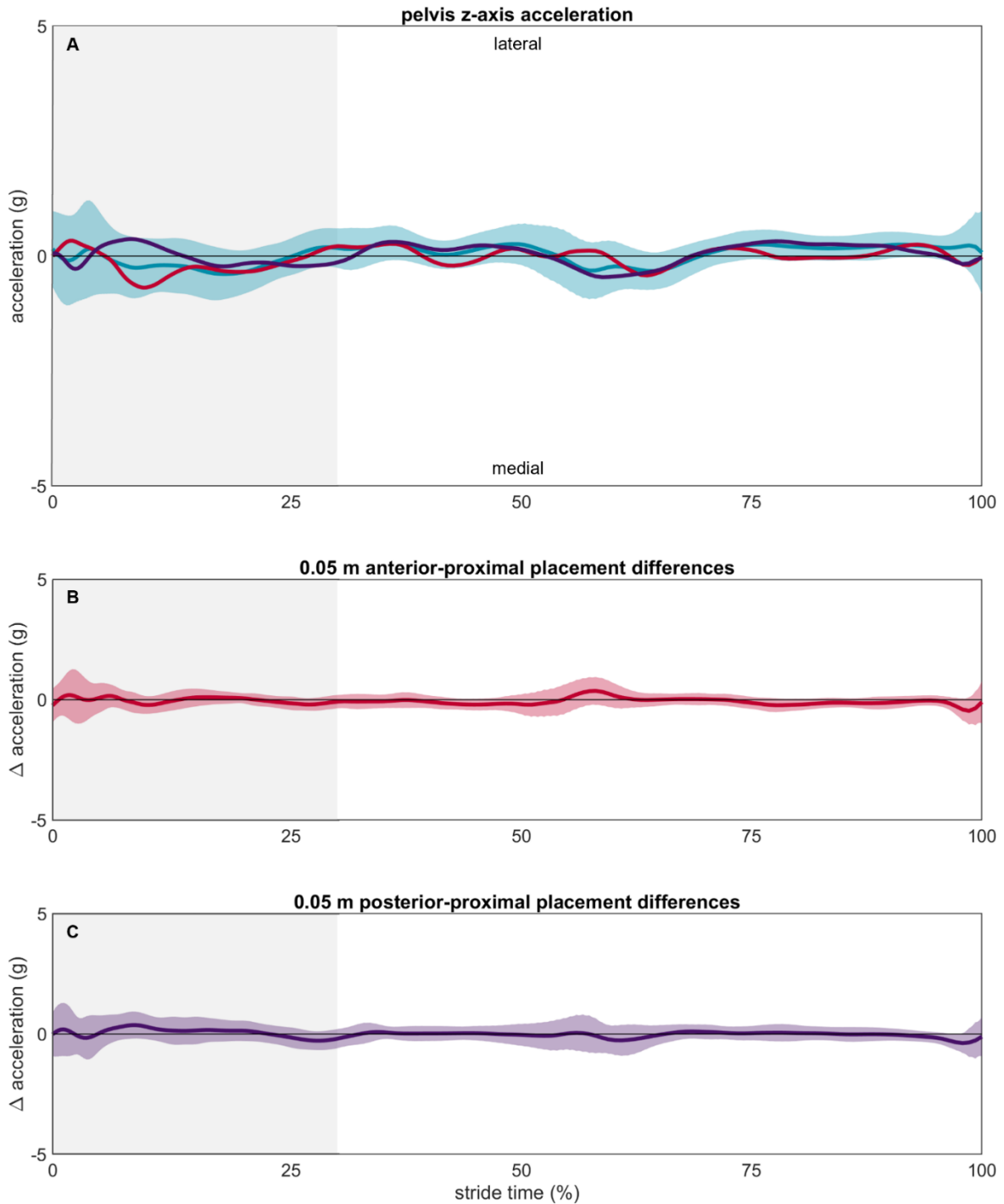


**Figure 5S.7: Pelvis x-axis accelerations (A)** measured at the reference (blue), anterior-proximal misplaced (red), and posterior-proximal (purple) IMUs. The light blue area represents  $\pm 1$  SD about the reference mean. The light grey background represents

stance while the white background represents swing. Differences between the reference and **(B)** anterior-proximal and **(C)** posterior-proximal IMUs are shown in the panels below.

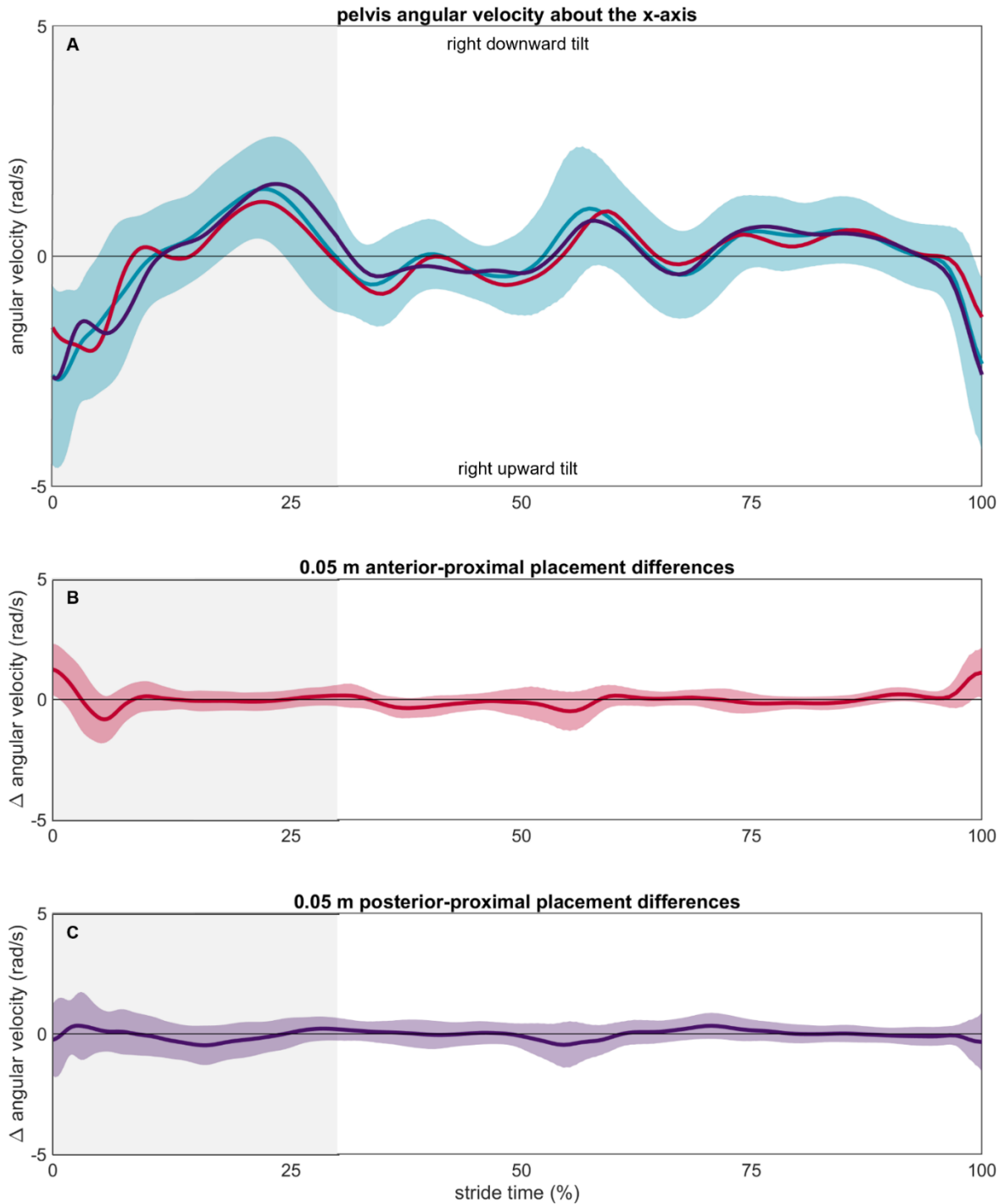


**Figure 5S.8: Pelvis y-axis accelerations (A)** measured at the reference (blue), anterior-proximal misplaced (red), and posterior-proximal (purple) IMUs. The light blue area represents  $\pm 1$  SD about the reference mean. The light grey background represents stance while the white background represents swing. Differences between the reference and **(B)** anterior-proximal and **(C)** posterior-proximal IMUs are shown in the panels below.

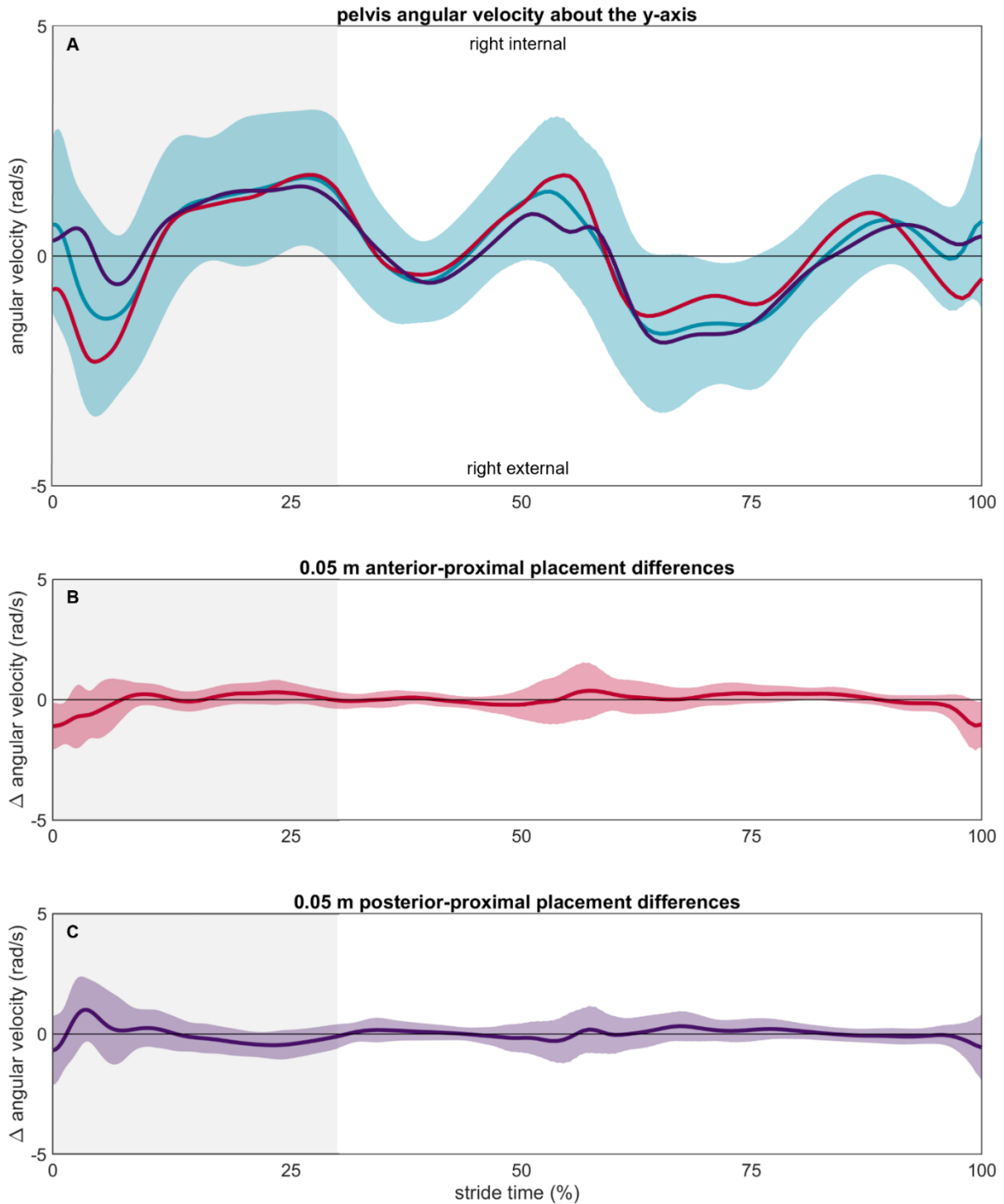


**Figure 5S.9: Pelvis z-axis accelerations (A)** measured at the reference (blue), anterior-proximal misplaced (red), and posterior-proximal (purple) IMUs. The light blue area represents  $\pm 1$  SD about the reference mean. The light grey background represents stance while the white background represents swing. Differences between the reference and **(B)** anterior-proximal and **(C)** posterior-proximal IMUs are shown in the panels below.

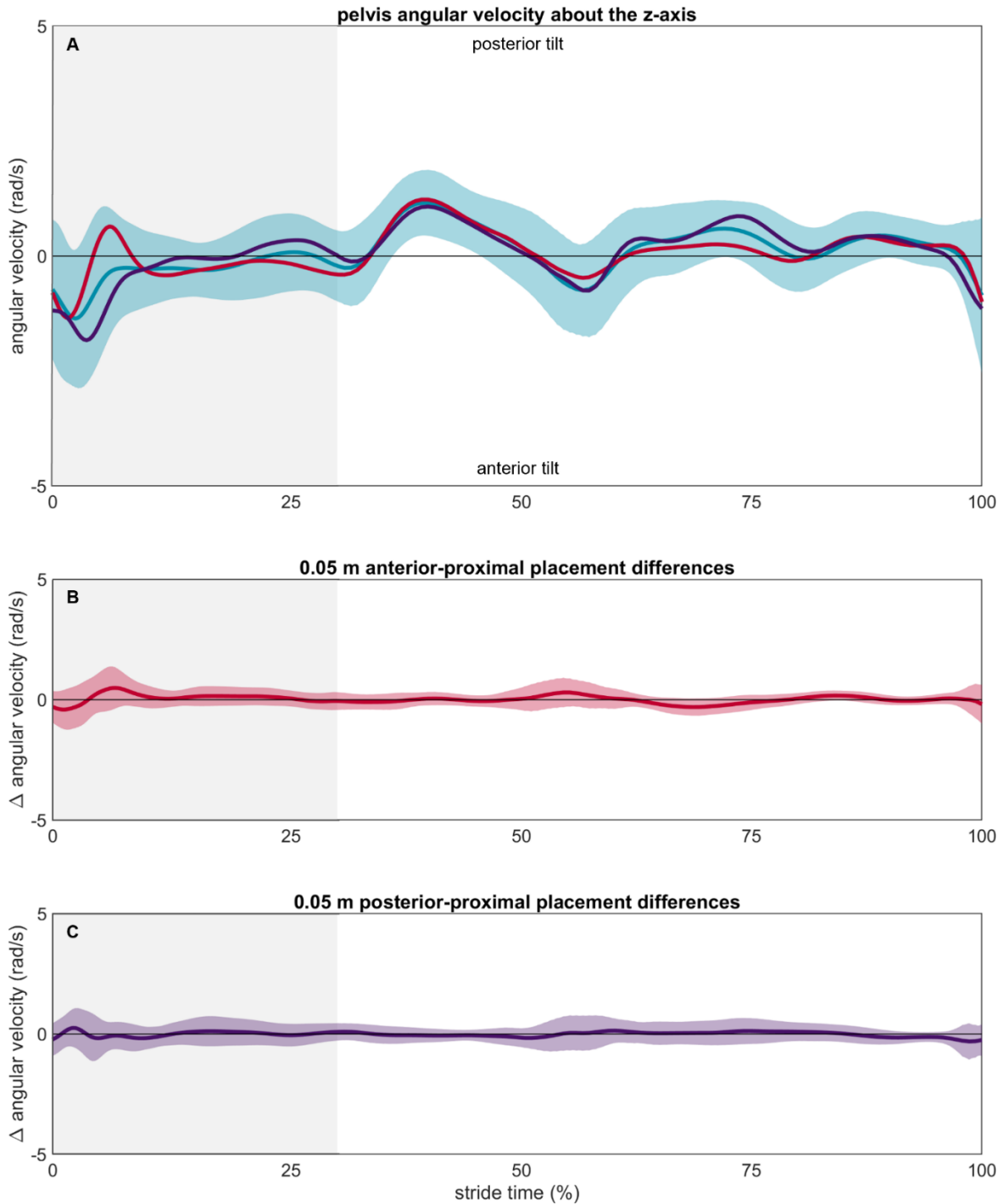
5S.1.4. Pelvis angular velocity



**Figure 5S.10: Pelvis x-axis angular velocities (A)** measured at the reference (blue), anterior-proximal misplaced (red), and posterior-proximal (purple) IMUs. The light blue area represents  $\pm 1$  SD about the reference mean. The light grey background represents stance while the white background represents swing. Differences between the reference and **(B)** anterior-proximal and **(C)** posterior-proximal IMUs are shown in the panels below.



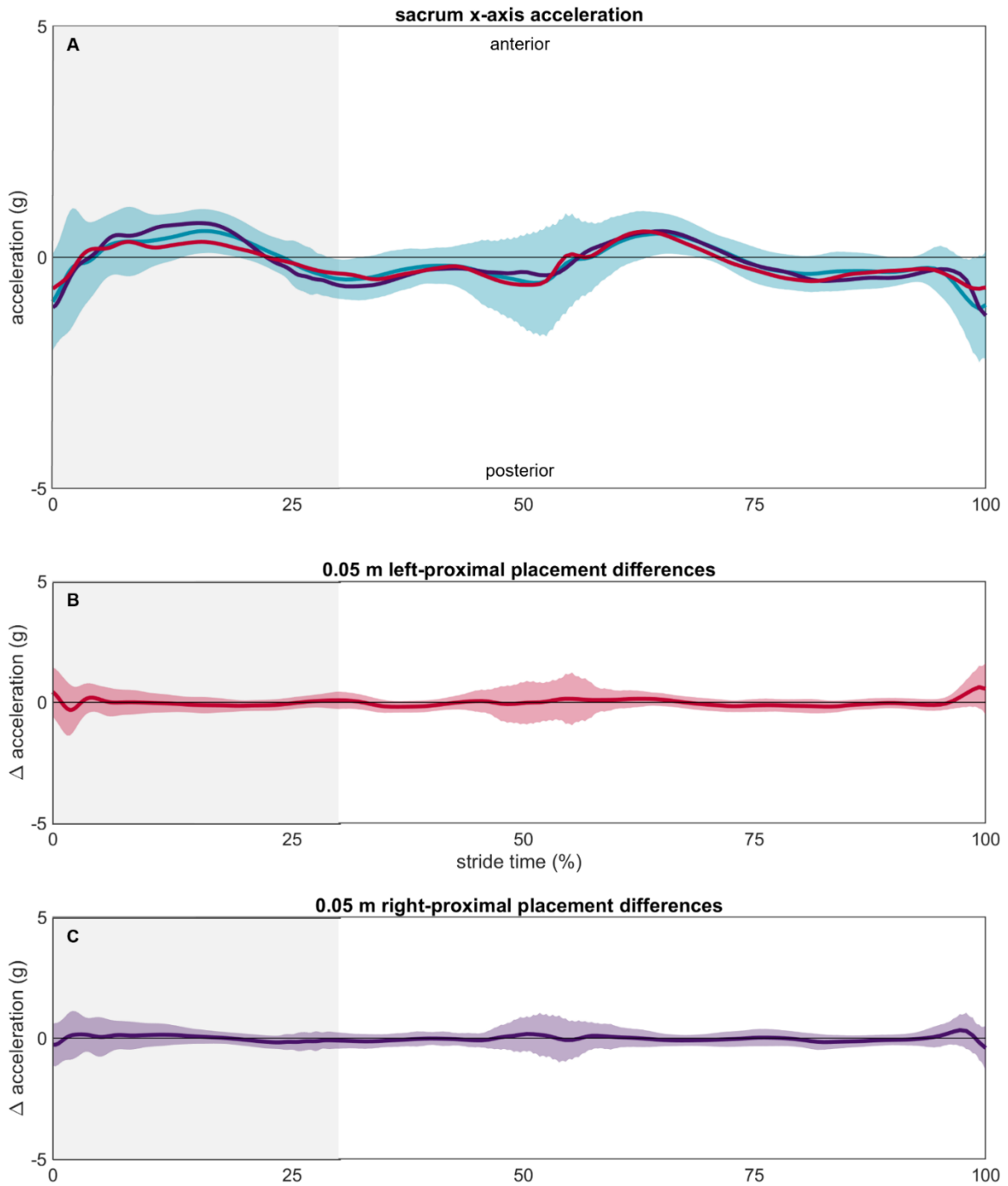
**Figure 5S.11: Pelvis y-axis angular velocities (A)** measured at the reference (blue), anterior-proximal misplaced (red), and posterior-proximal (purple) IMUs. The light blue area represents  $\pm 1$  SD about the reference mean. The light grey background represents stance while the white background represents swing. Differences between the reference and **(B)** anterior-proximal and **(C)** posterior-proximal IMUs are shown in the panels below.



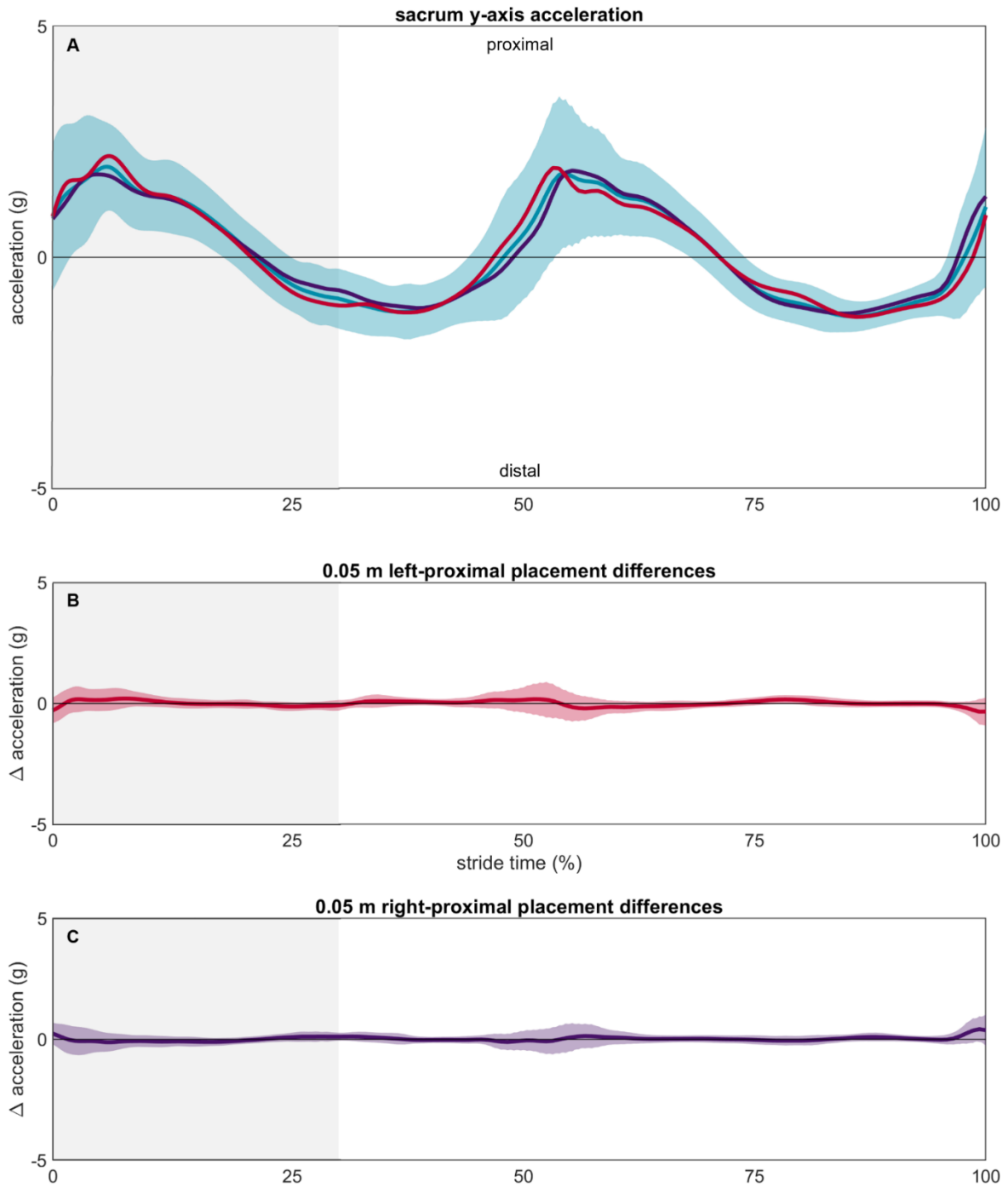
**Figure 5S.12: Pelvis z-axis angular velocities (A)** measured at the reference (blue), anterior-proximal misplaced (red), and posterior-proximal (purple) IMUs. The light blue area represents  $\pm 1$  SD about the reference mean. The light grey background represents stance while the white background represents swing. Differences between the reference and **(B)** anterior-proximal and **(C)** posterior-proximal IMUs are shown in the panels below.



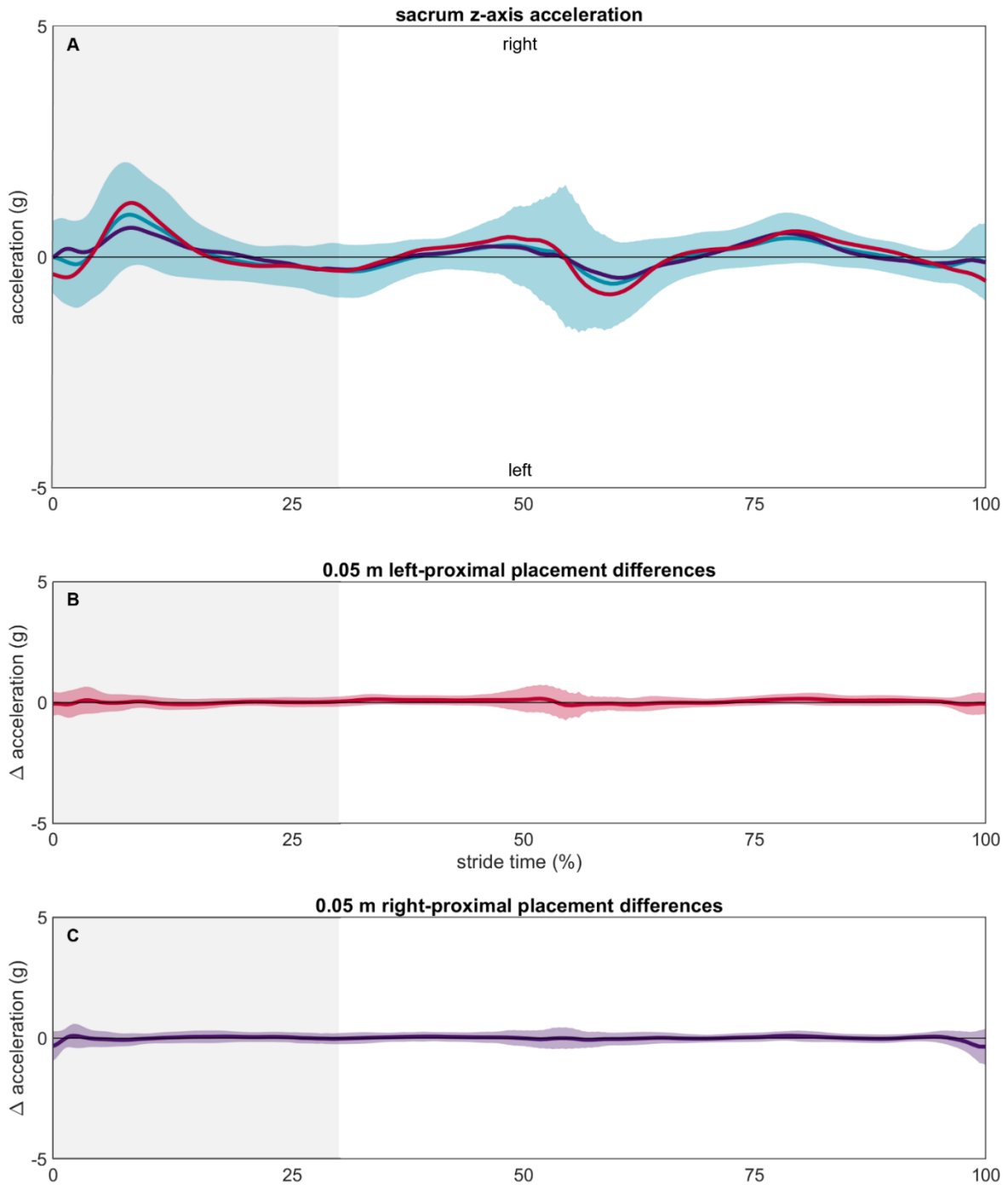
### 5S.1.5. Sacrum acceleration



**Figure 5S.13: Sacrum x-axis accelerations (A)** measured at the reference (blue), left-proximal misplaced (red), and right-proximal (purple) IMUs. The light blue area represents  $\pm 1$  SD about the reference mean. The light grey background represents stance while the white background represents swing. Differences between the reference and **(B)** left-proximal and **(C)** right-proximal IMUs are shown in the panels below.

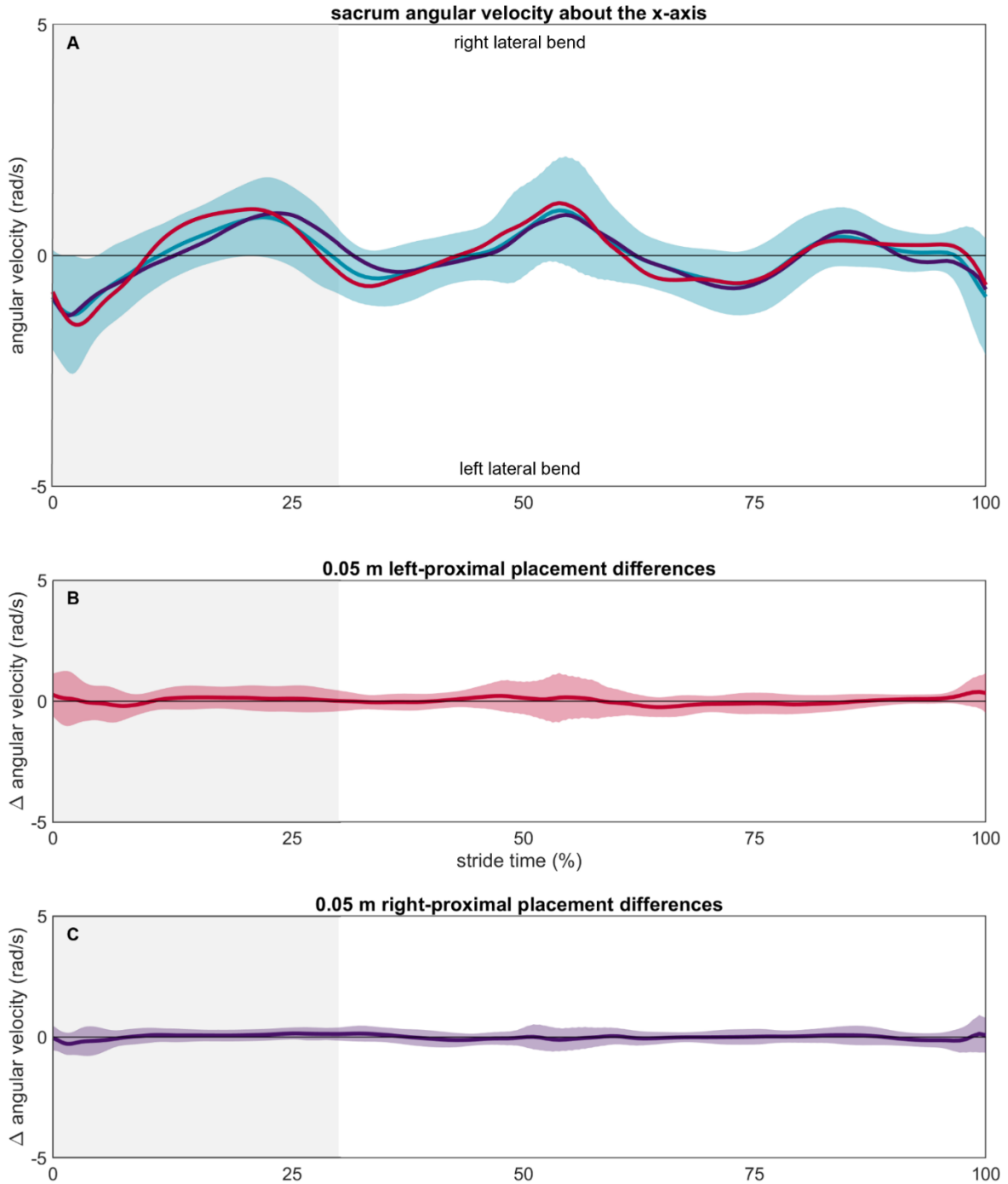


**Figure 5S.14: Sacrum y-axis accelerations (A)** measured at the reference (blue), left-proximal misplaced (red), and right-proximal (purple) IMUs. The light blue area represents  $\pm 1$  SD about the reference mean. The light grey background represents stance while the white background represents swing. Differences between the reference and **(B)** left-proximal and **(C)** right-proximal IMUs are shown in the panels below.

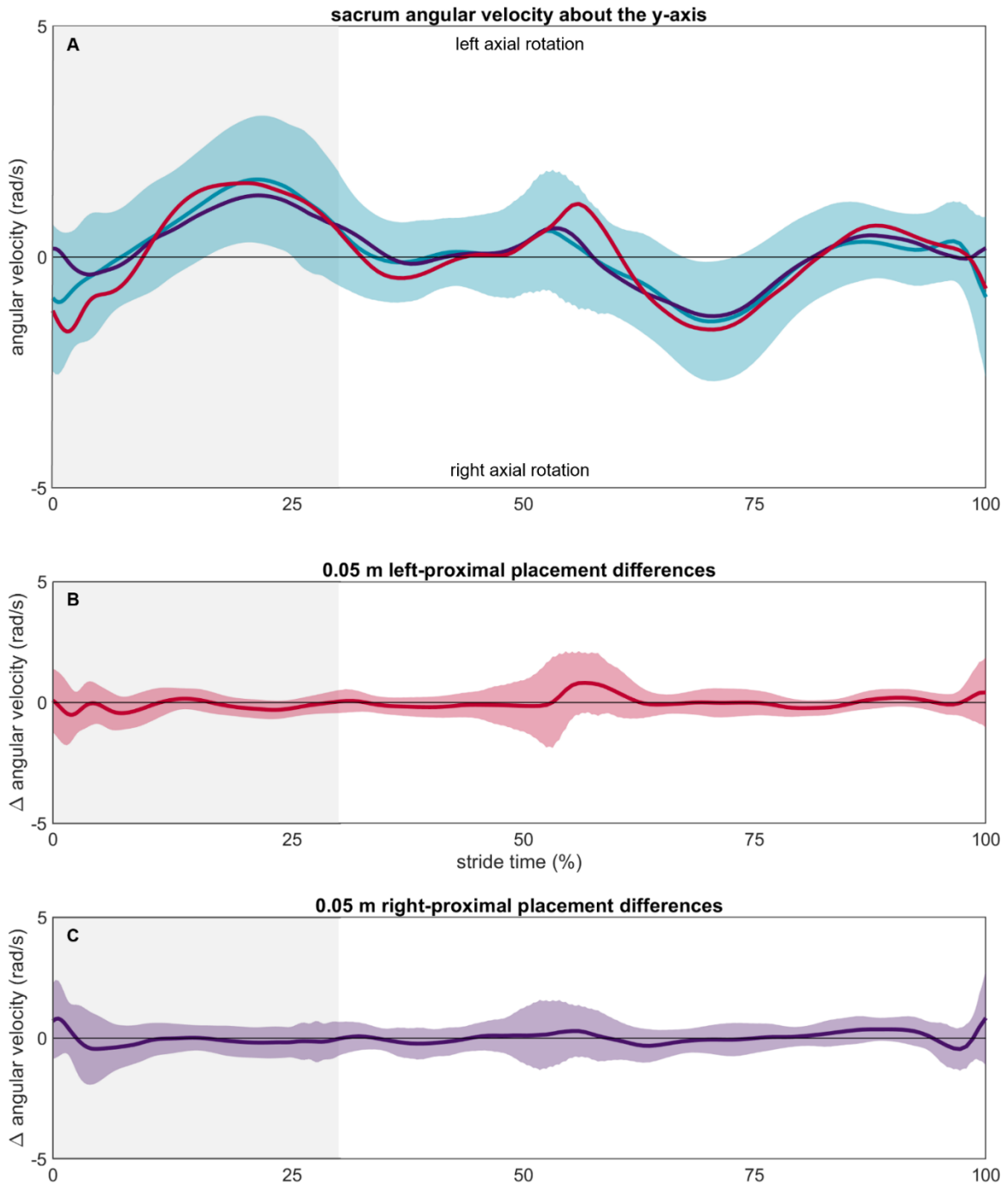


**Figure 5S.15: Sacrum z-axis accelerations (A)** measured at the reference (blue), left-proximal misplaced (red), and right-proximal (purple) IMUs. The light blue area represents  $\pm 1$  SD about the reference mean. The light grey background represents stance while the white background represents swing. Differences between the reference and **(B)** left-proximal and **(C)** right-proximal IMUs are shown in the panels below.

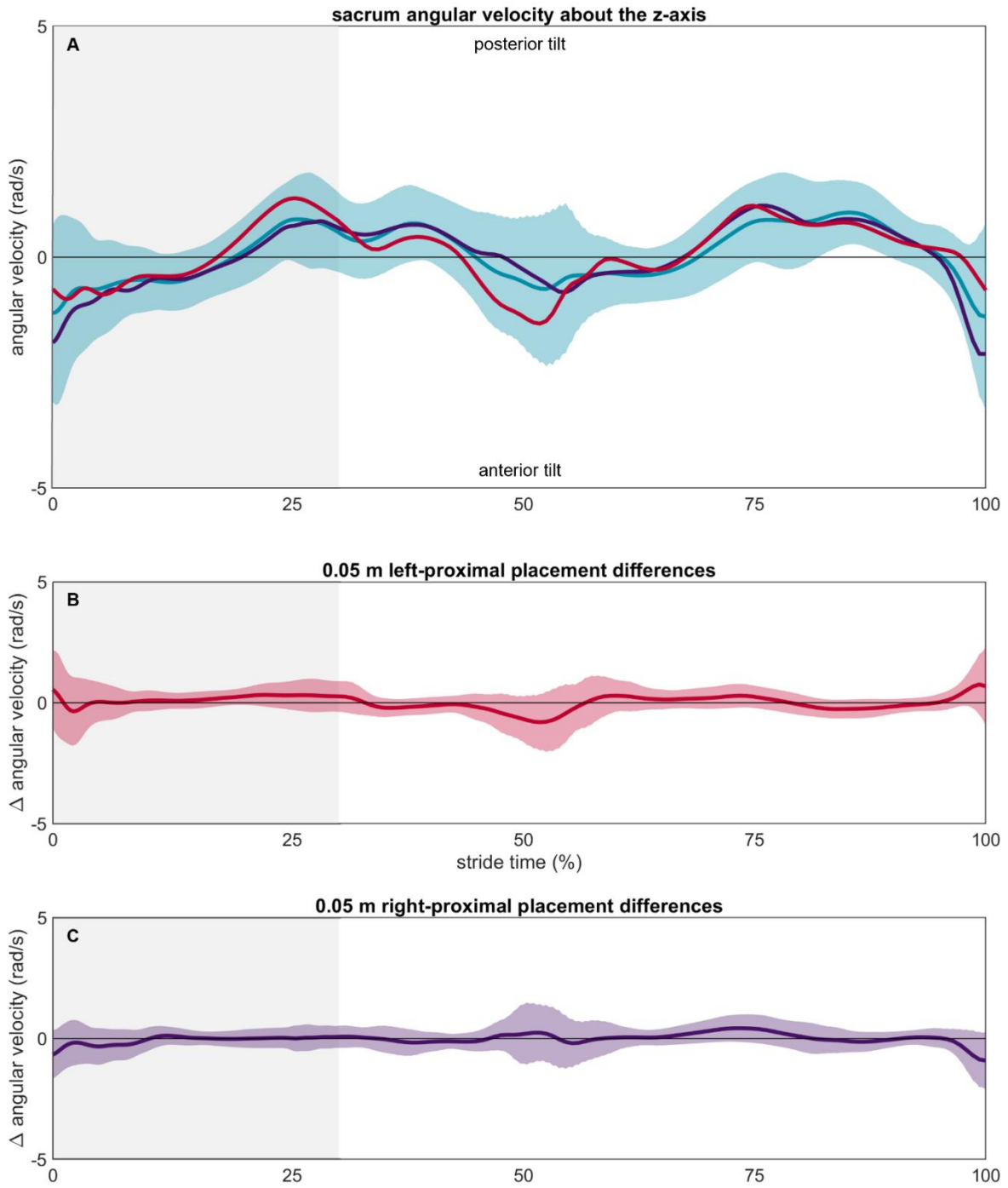
5S.1.6. Sacrum angular velocity



**Figure 5S.16: Sacrum x-axis angular velocity (A)** measured at the reference (blue), left-proximal misplaced (red), and right-proximal (purple) IMUs. The light blue area represents  $\pm 1$  SD about the reference mean. The light grey background represents stance while the white background represents swing. Differences between the reference and **(B)** left-proximal and **(C)** right-proximal IMUs are shown in the panels below.



**Figure 5S.17: Sacrum y-axis angular velocity (A)** measured at the reference (blue), left-proximal misplaced (red), and right-proximal (purple) IMUs. The light blue area represents  $\pm 1$  SD about the reference mean. The light grey background represents stance while the white background represents swing. Differences between the reference and **(B)** left-proximal and **(C)** right-proximal IMUs are shown in the panels below.

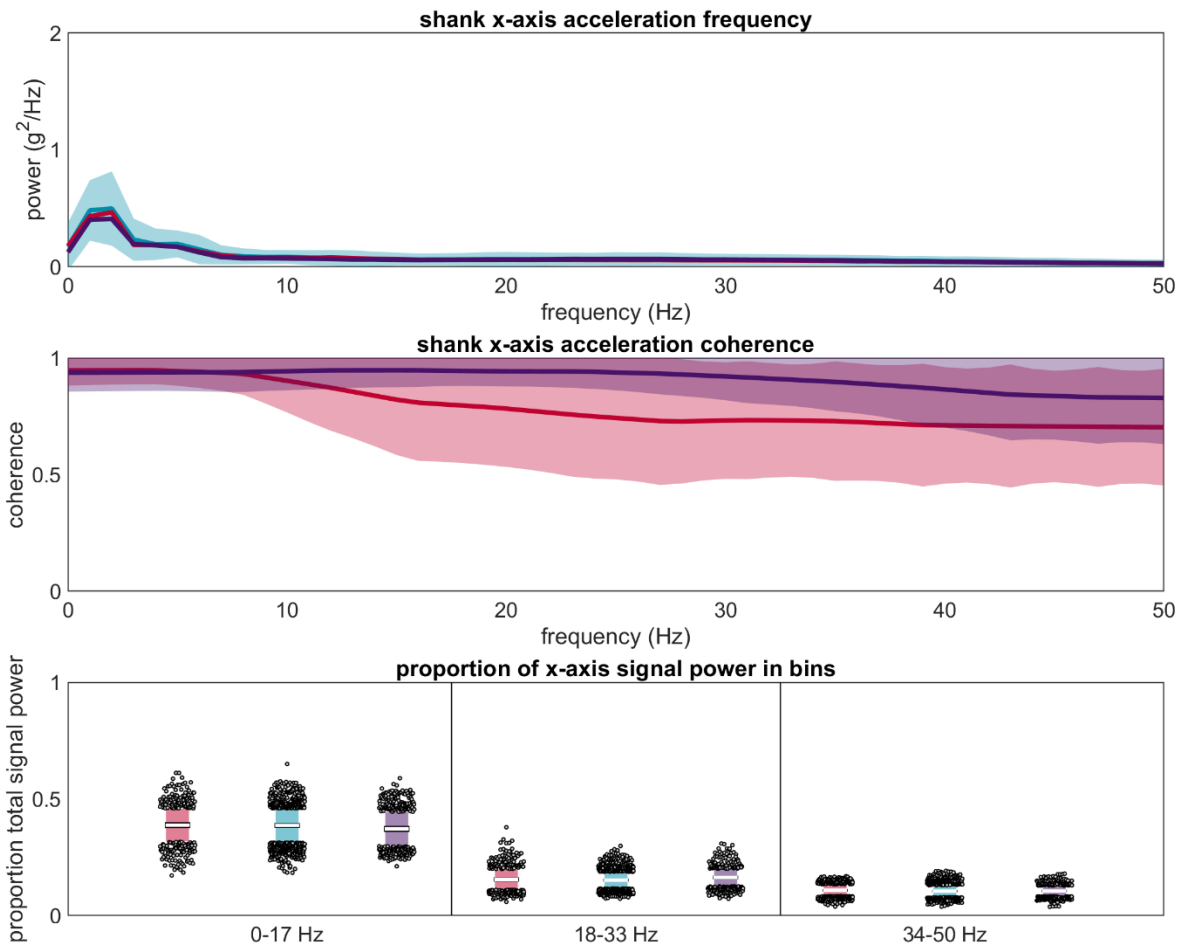


**Figure 5S.18: Sacrum y-axis angular velocity (A)** measured at the reference (blue), left-proximal misplaced (red), and right-proximal (purple) IMUs. The light blue area represents  $\pm 1$  SD about the reference mean. The light grey background represents stance while the white background represents swing. Differences between the reference and **(B)** left-proximal and **(C)** right-proximal IMUs are shown in the panels below.

### 5.S.2. Frequency domain analyses

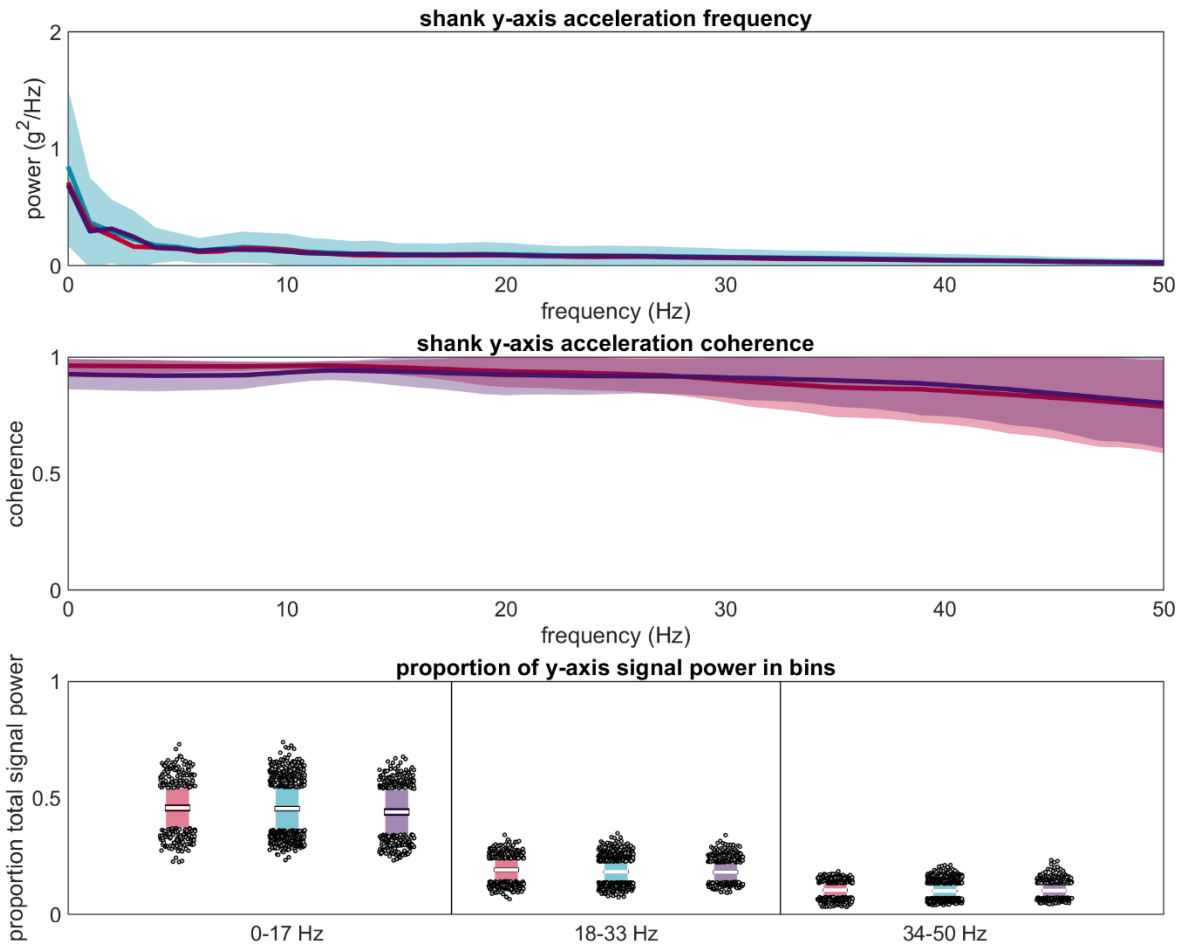
Each stride previously segmented in the time domain was zero-padded to the nearest power of two that exceeded the stride duration. The power spectral density from 0 to the Nyquist frequency (500 Hz) was then calculated using a Fast Fourier Transform. Powers and frequencies were interpolated to 1 Hz bins, then power was normalized so that the sum of powers from 0 to 500 Hz was equal to the mean squared amplitude of the time domain. Thus, final units were  $g^2/\text{Hz}$  or  $(\text{rad/s})^2/\text{Hz}$  [186, 187, 323]. Magnitude-squared coherences between reference and misplaced IMU signals were then calculated via the Welch method. This yields a number between 0 and 1 at each frequency with 0 indicating that the misplaced signal cannot be predicted from the reference signal using a linear model and 1 indicating that the two signals are perfectly related through a linear model. Finally, the proportion of signal power in three equally sized bins from 0 to 50 Hz (0 to 16 Hz, 17 to 33 Hz, and 34 to 50 Hz) was calculated [98].

5S.2.1. Shank acceleration

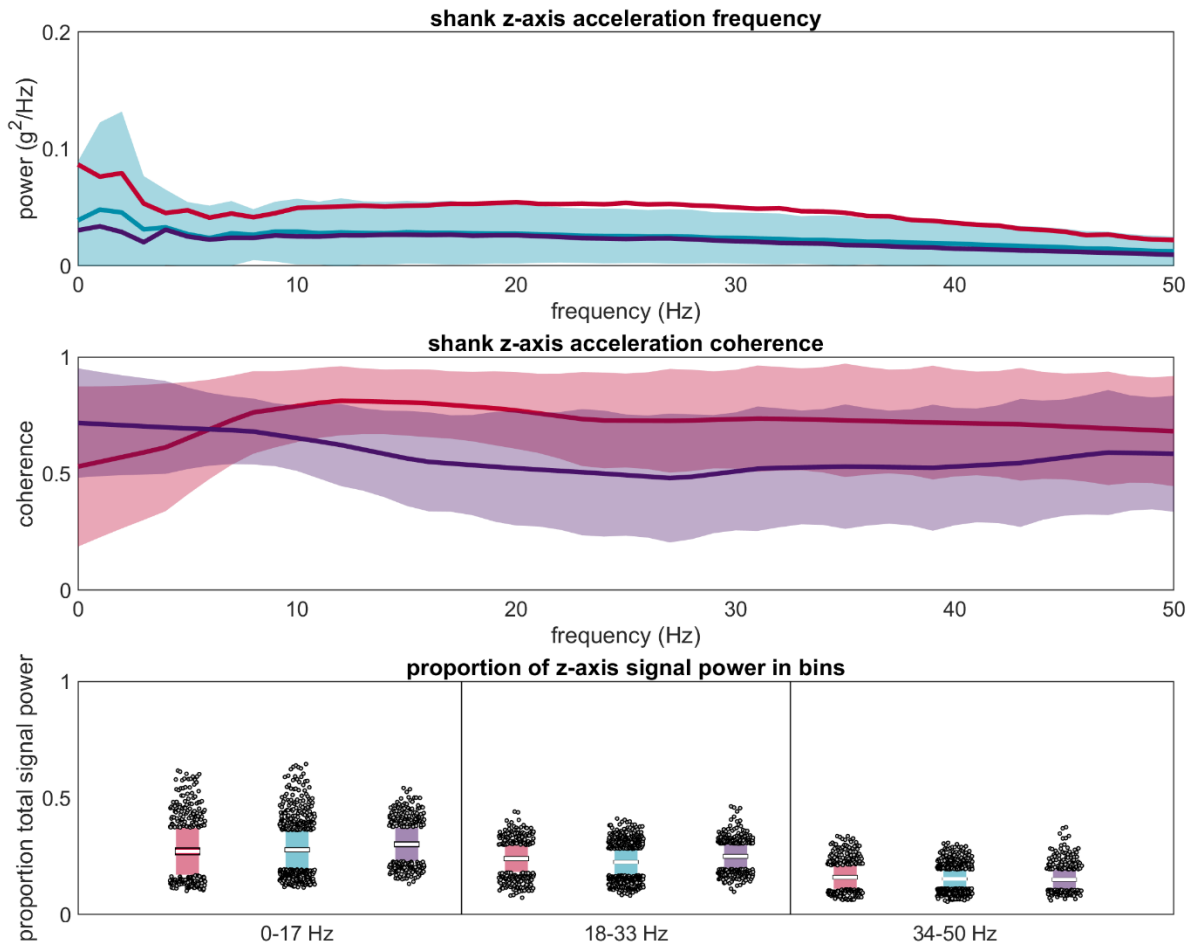


**Figure 5S.19: Frequency analysis of shank x-axis acceleration. (Top)** Power spectral density from 0 to 50 Hz. The blue line represents the mean for the reference IMU, the blue shaded area represents  $\pm 1$  SD, and the red and purple lines represent the anterior-proximal and posterior-proximal misplaced IMUs. **(Middle)** Magnitude squared coherence between each misplaced IMU and the reference IMU. Plotted as mean  $\pm 1$  SD. **(Bottom)** Proportion of total signal power observed in the 0-17 Hz, 18-33 Hz, and 34-50 Hz bins. Central white line represents the mean across trials, dark shaded area represents  $\pm 1.96$  SEM (confidence interval; difficult to see due to small size), light shaded area represents  $\pm 1$  SD, and grey dots represent trials falling outside  $\pm 1$  SD.



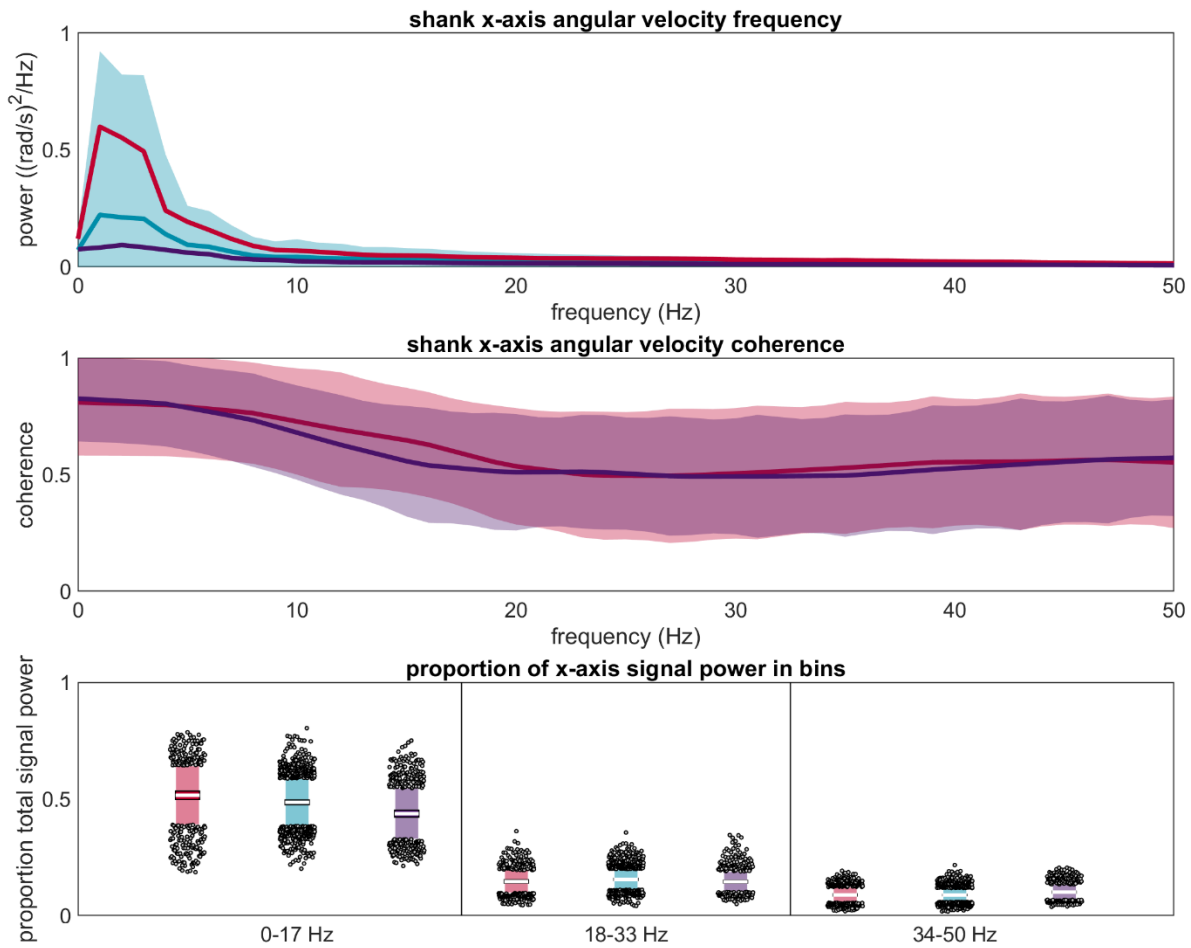


**Figure 5S.20: Frequency analysis of shank y-axis acceleration. (Top)** Power spectral density from 0 to 50 Hz. The blue line represents the mean for the reference IMU, the blue shaded area represents  $\pm 1$  SD, and the red and purple lines represent the anterior-proximal and posterior-proximal misplaced IMUs. **(Middle)** Magnitude squared coherence between each misplaced IMU and the reference IMU. Plotted as mean  $\pm 1$  SD. **(Bottom)** Proportion of total signal power observed in the 0-17 Hz, 18-33 Hz, and 34-50 Hz bins. Central white line represents the mean across trials, dark shaded area represents  $\pm 1.96$  SEM (confidence interval; difficult to see due to small size), light shaded area represents  $\pm 1$  SD, and grey dots represent trials falling outside  $\pm 1$  SD.

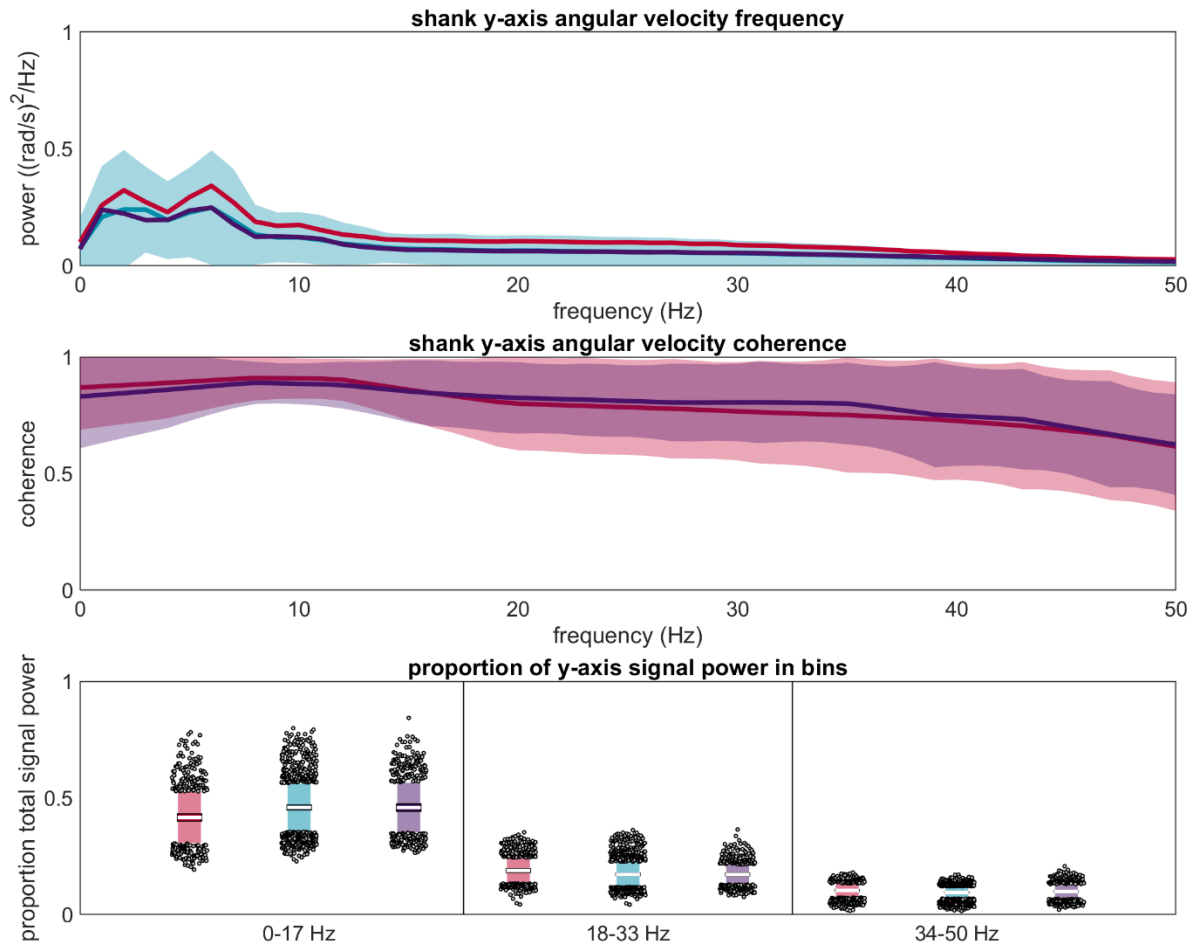


**Figure 5S.21: Frequency analysis of shank z-axis acceleration. (Top)** Power spectral density from 0 to 50 Hz. The blue line represents the mean for the reference IMU, the blue shaded area represents  $\pm 1$  SD, and the red and purple lines represent the anterior-proximal and posterior-proximal misplaced IMUs. **(Middle)** Magnitude squared coherence between each misplaced IMU and the reference IMU. Plotted as mean  $\pm 1$  SD. **(Bottom)** Proportion of total signal power observed in the 0-17 Hz, 18-33 Hz, and 34-50 Hz bins. Central white line represents the mean across trials, dark shaded area represents  $\pm 1.96$  SEM (confidence interval; difficult to see due to small size), light shaded area represents  $\pm 1$  SD, and grey dots represent trials falling outside  $\pm 1$  SD.

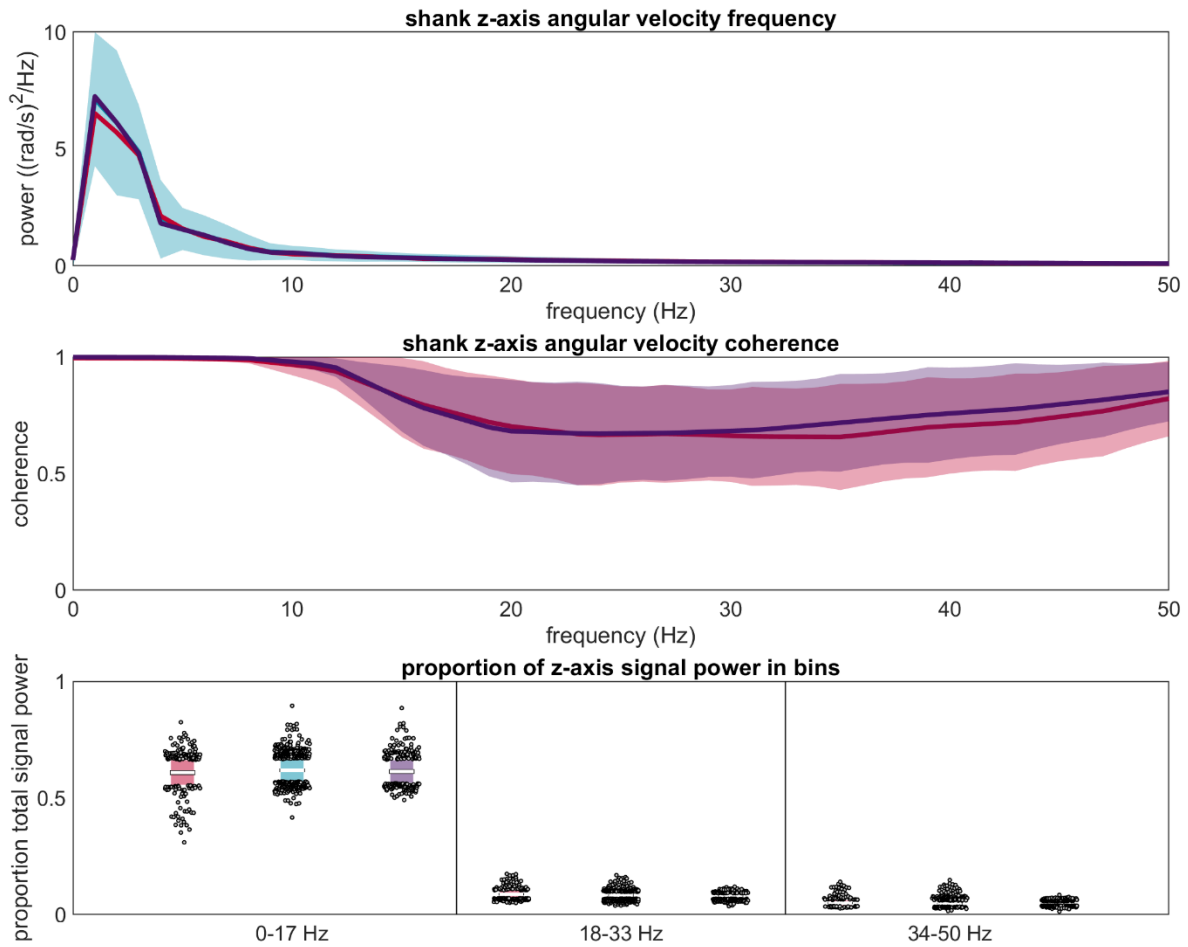
5S.2.2. Shank angular velocity



**Figure 5S.22: Frequency analysis of shank x-axis angular velocity.** (Top) Power spectral density from 0 to 50 Hz. The blue line represents the mean for the reference IMU, the blue shaded area represents  $\pm 1$  SD, and the red and purple lines represent the anterior-proximal and posterior-proximal misplaced IMUs. (Middle) Magnitude squared coherence between each misplaced IMU and the reference IMU. Plotted as mean  $\pm 1$  SD. (Bottom) Proportion of total signal power observed in the 0-17 Hz, 18-33 Hz, and 34-50 Hz bins. Central white line represents the mean across trials, dark shaded area represents  $\pm 1.96$  SEM (confidence interval; difficult to see due to small size), light shaded area represents  $\pm 1$  SD, and grey dots represent trials falling outside  $\pm 1$  SD.

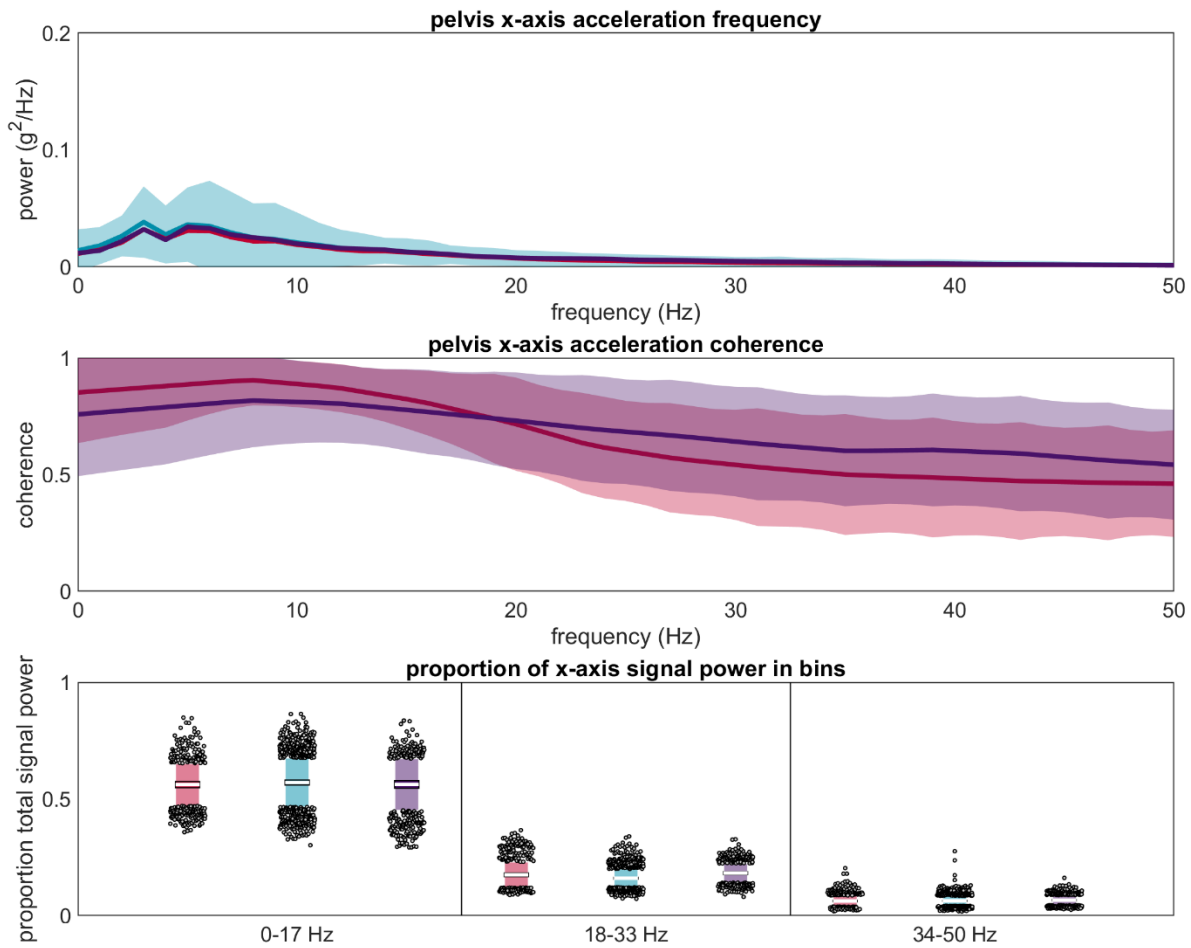


**Figure 5S.23: Frequency analysis of shank y-axis angular velocity.** (Top) Power spectral from 0 to 50 Hz. The blue line represents the mean for the reference IMU, the blue shaded area represents  $\pm 1$  SD, and the red and purple lines represent the anterior-proximal and posterior-proximal misplaced IMUs. (Middle) Magnitude squared coherence between each misplaced IMU and the reference IMU. Plotted as mean  $\pm 1$  SD. (Bottom) Proportion of total signal power observed in the 0-17 Hz, 18-33 Hz, and 34-50 Hz bins. Central white line represents the mean across trials, dark shaded area represents  $\pm 1.96$  SEM (confidence interval; difficult to see due to small size), light shaded area represents  $\pm 1$  SD, and grey dots represent trials falling outside  $\pm 1$  SD.

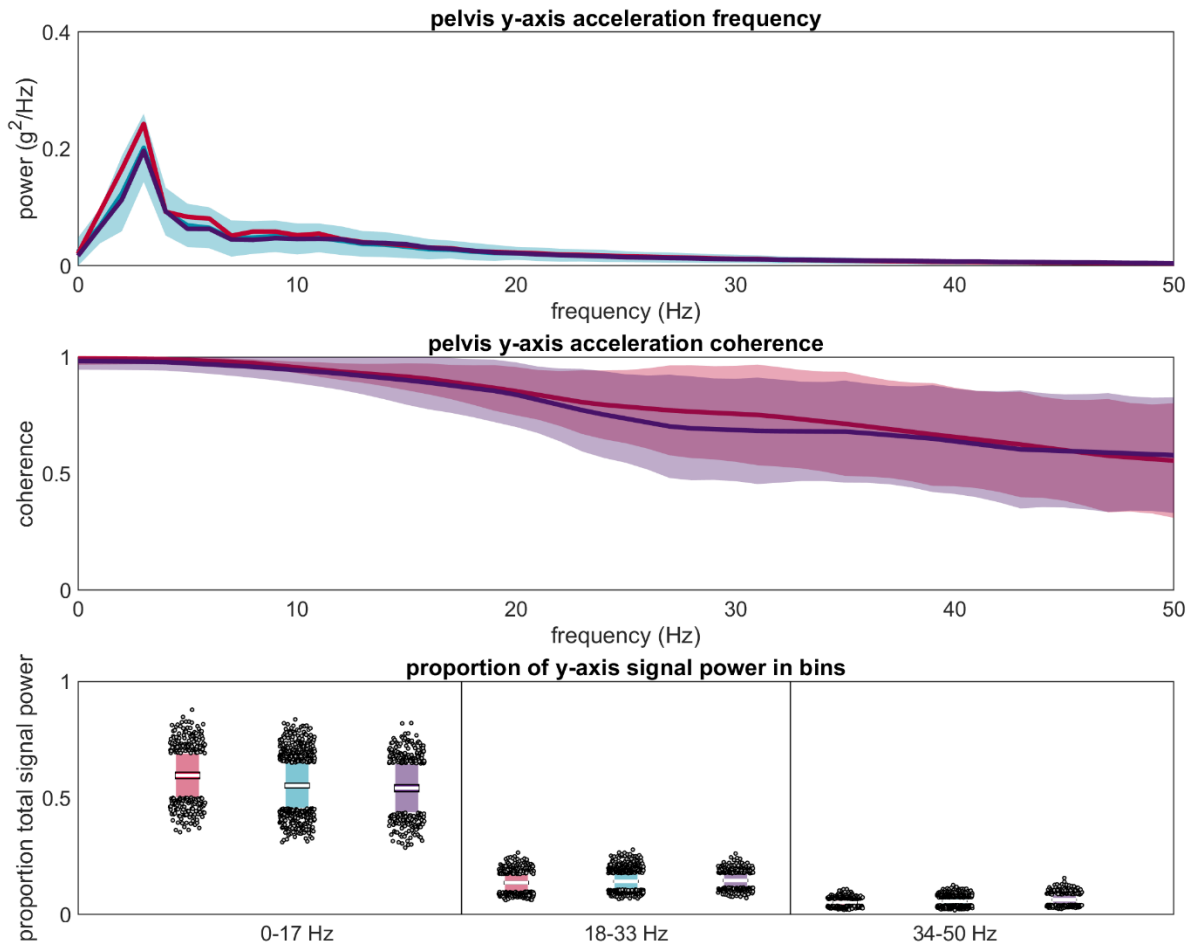


**Figure 5S.24: Frequency analysis of shank z-axis angular velocity.** (Top) Power spectral from 0 to 50 Hz. The blue line represents the mean for the reference IMU, the blue shaded area represents  $\pm 1$  SD, and the red and purple lines represent the anterior-proximal and posterior-proximal misplaced IMUs. (Middle) Magnitude squared coherence between each misplaced IMU and the reference IMU. Plotted as mean  $\pm 1$  SD. (Bottom) Proportion of total signal power observed in the 0-17 Hz, 18-33 Hz, and 34-50 Hz bins. Central white line represents the mean across trials, dark shaded area represents  $\pm 1.96$  SEM (confidence interval; difficult to see due to small size), light shaded area represents  $\pm 1$  SD, and grey dots represent trials falling outside  $\pm 1$  SD.

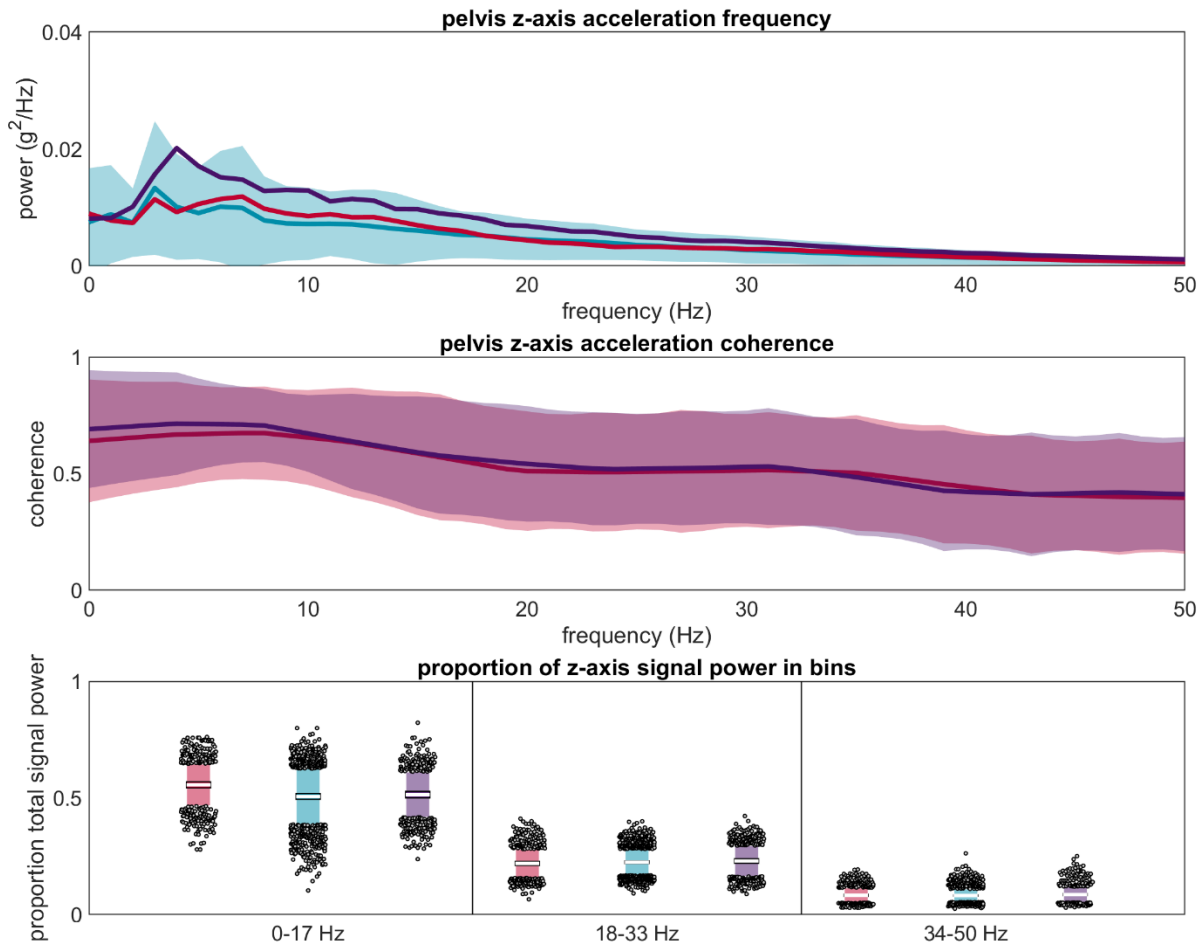
### 5S.2.3. Pelvis acceleration



**Figure 5S.25: Frequency analysis of pelvis x-axis acceleration.** **(Top)** Power spectral density from 0 to 50 Hz. The blue line represents the mean for the reference IMU, the blue shaded area represents  $\pm 1$  SD, and the red and purple lines represent the anterior-proximal and posterior-proximal misplaced IMUs. **(Middle)** Magnitude squared coherence between each misplaced IMU and the reference IMU. Plotted as mean  $\pm 1$  SD. **(Bottom)** Proportion of total signal power observed in the 0-17 Hz, 18-33 Hz, and 34-50 Hz bins. Central white line represents the mean across trials, dark shaded area represents  $\pm 1.96$  SEM (confidence interval; difficult to see due to small size), light shaded area represents  $\pm 1$  SD, and grey dots represent trials falling outside  $\pm 1$  SD.



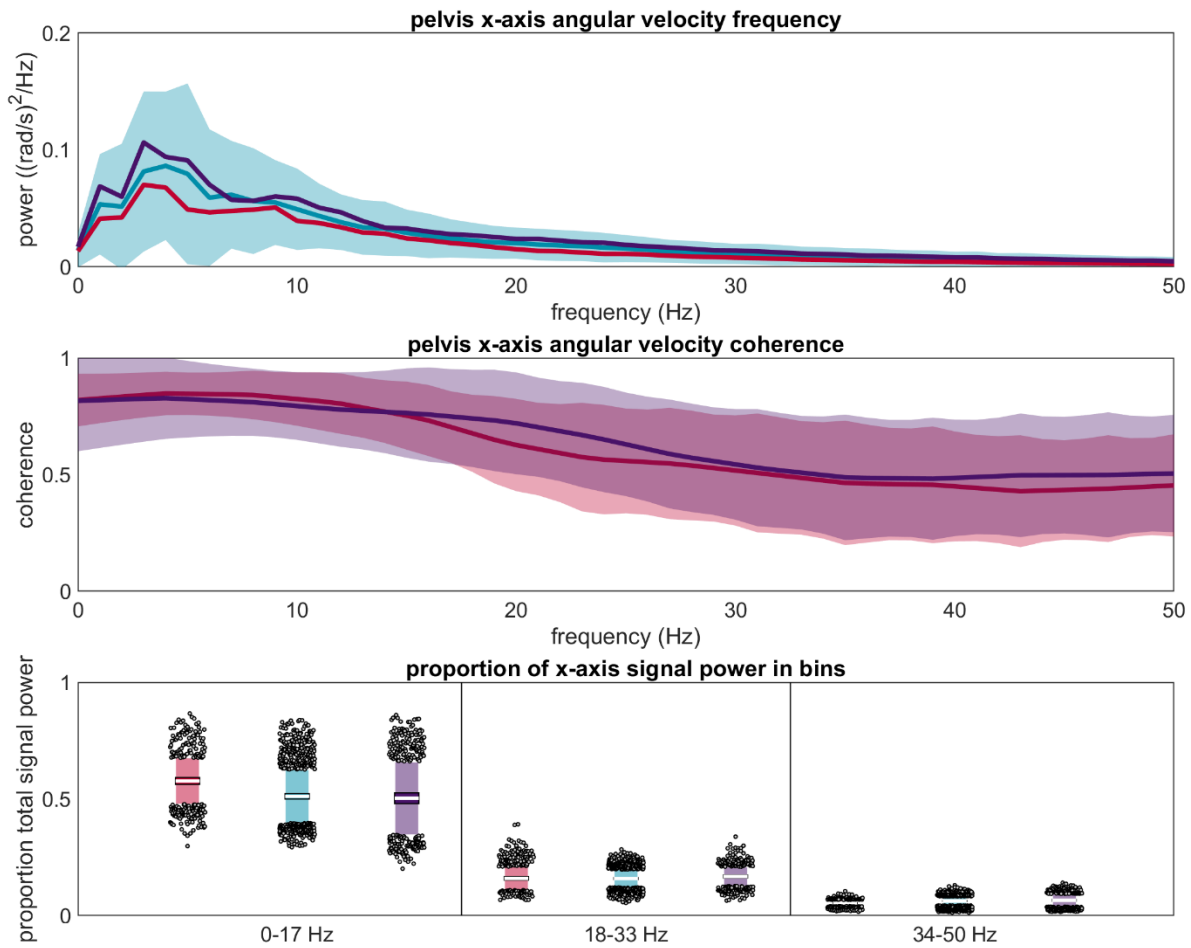
**Figure 5S.26: Frequency analysis of pelvis y-axis acceleration. (Top)** Power spectral density from 0 to 50 Hz. The blue line represents the mean for the reference IMU, the blue shaded area represents  $\pm 1$  SD, and the red and purple lines represent the anterior-proximal and posterior-proximal misplaced IMUs. **(Middle)** Magnitude squared coherence between each misplaced IMU and the reference IMU. Plotted as mean  $\pm 1$  SD. **(Bottom)** Proportion of total signal power observed in the 0-17 Hz, 18-33 Hz, and 34-50 Hz bins. Central white line represents the mean across trials, dark shaded area represents  $\pm 1.96$  SEM (confidence interval; difficult to see due to small size), light shaded area represents  $\pm 1$  SD, and grey dots represent trials falling outside  $\pm 1$  SD.



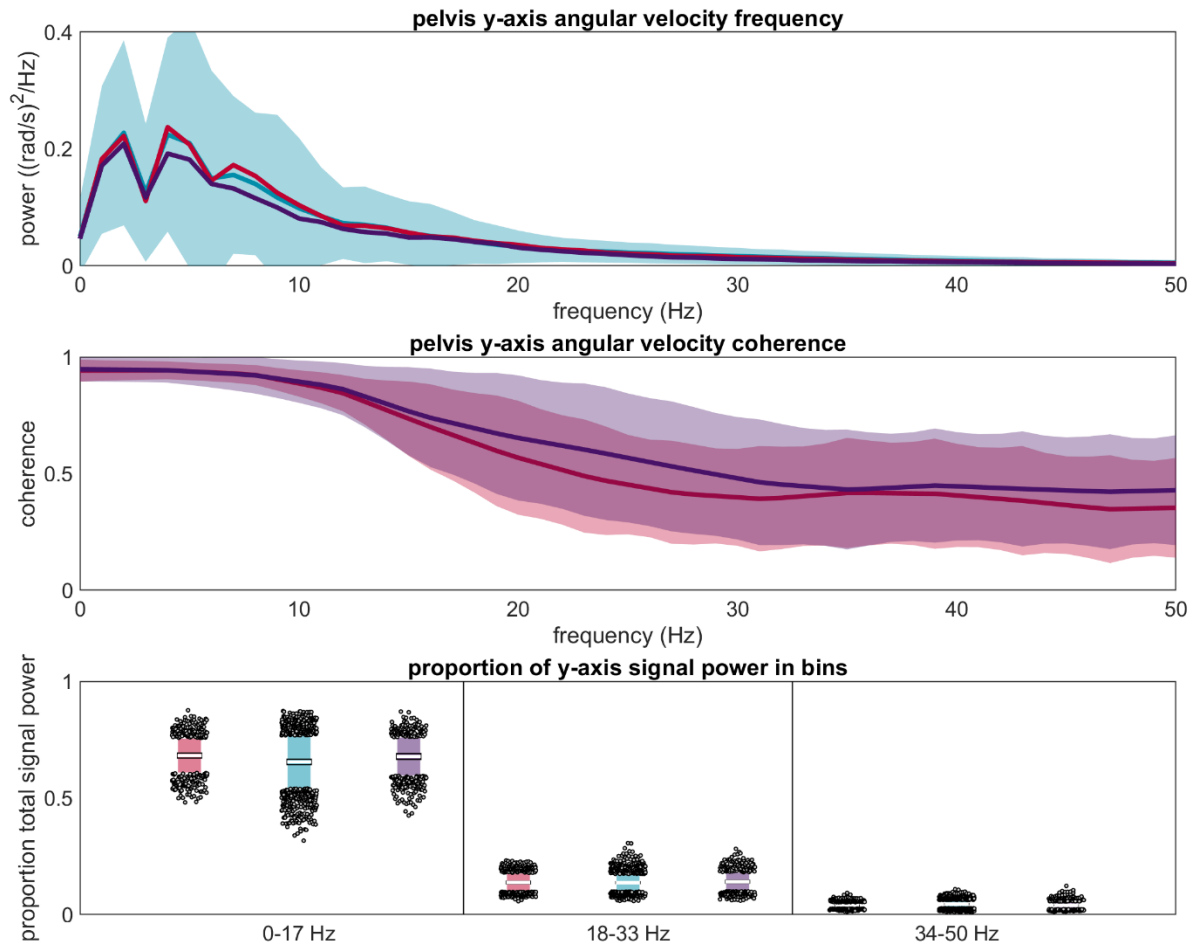
**Figure 5S.27: Frequency analysis of pelvis z-axis acceleration. (Top)** Power spectral density from 0 to 50 Hz. The blue line represents the mean for the reference IMU, the blue shaded area represents  $\pm 1$  SD, and the red and purple lines represent the anterior-proximal and posterior-proximal misplaced IMUs. **(Middle)** Magnitude squared coherence between each misplaced IMU and the reference IMU. Plotted as mean  $\pm 1$  SD. **(Bottom)** Proportion of total signal power observed in the 0-17 Hz, 18-33 Hz, and 34-50 Hz bins. Central white line represents the mean across trials, dark shaded area represents  $\pm 1.96$  SEM (confidence interval; difficult to see due to small size), light shaded area represents  $\pm 1$  SD, and grey dots represent trials falling outside  $\pm 1$  SD.



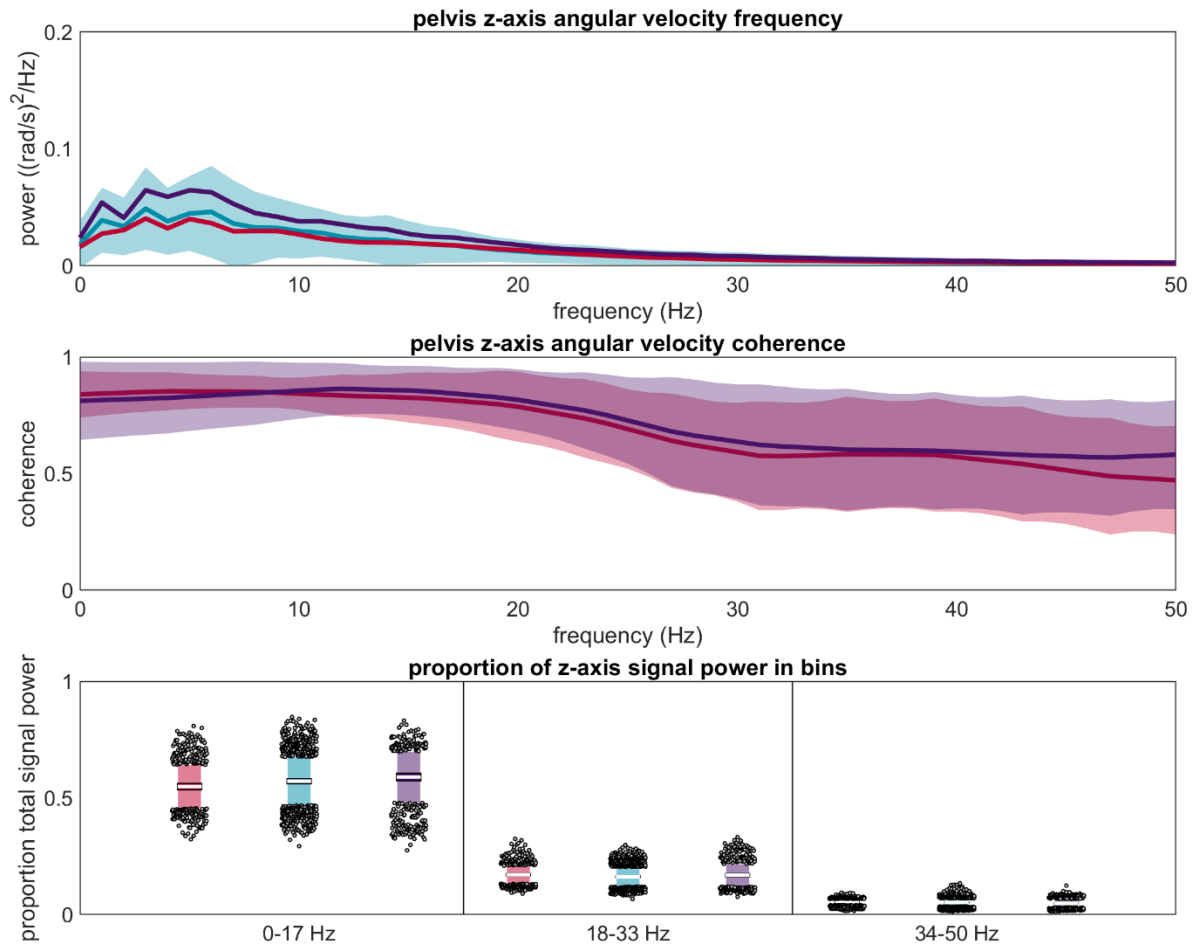
5S.2.4. Pelvis angular velocity



**Figure 5S.28: Frequency analysis of pelvis x-axis angular velocity.** (Top) Power spectral density from 0 to 50 Hz. The blue line represents the mean for the reference IMU, the blue shaded area represents  $\pm 1$  SD, and the red and purple lines represent the anterior-proximal and posterior-proximal misplaced IMUs. (Middle) Magnitude squared coherence between each misplaced IMU and the reference IMU. Plotted as mean  $\pm 1$  SD. (Bottom) Proportion of total signal power observed in the 0-17 Hz, 18-33 Hz, and 34-50 Hz bins. Central white line represents the mean across trials, dark shaded area represents  $\pm 1.96$  SEM (confidence interval; difficult to see due to small size), light shaded area represents  $\pm 1$  SD, and grey dots represent trials falling outside  $\pm 1$  SD.

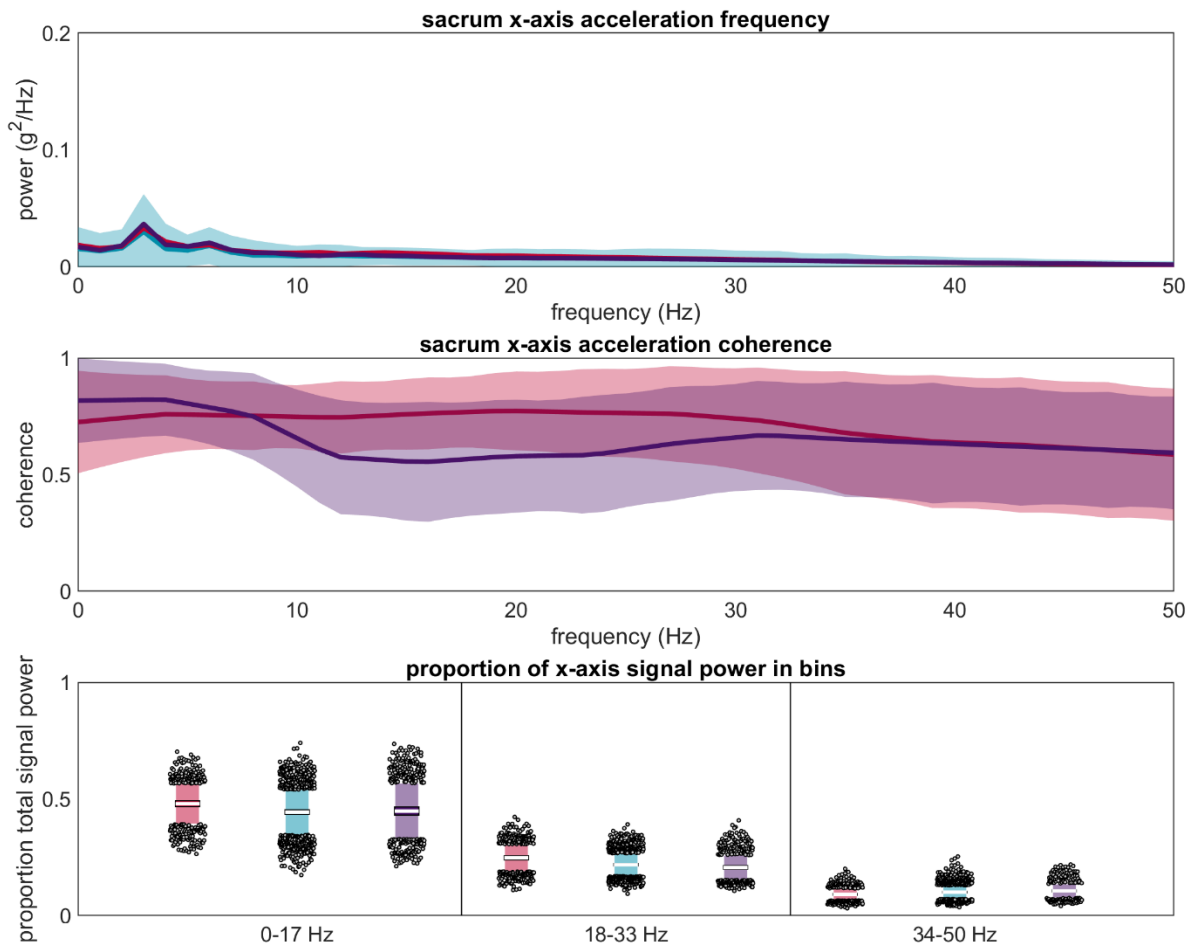


**Figure 5S.29: Frequency analysis of pelvis y-axis angular velocity.** (Top) Power spectral density from 0 to 50 Hz. The blue line represents the mean for the reference IMU, the blue shaded area represents  $\pm 1$  SD, and the red and purple lines represent the anterior-proximal and posterior-proximal misplaced IMUs. (Middle) Magnitude squared coherence between each misplaced IMU and the reference IMU. Plotted as mean  $\pm 1$  SD. (Bottom) Proportion of total signal power observed in the 0-17 Hz, 18-33 Hz, and 34-50 Hz bins. Central white line represents the mean across trials, dark shaded area represents  $\pm 1.96$  SEM (confidence interval; difficult to see due to small size), light shaded area represents  $\pm 1$  SD, and grey dots represent trials falling outside  $\pm 1$  SD.

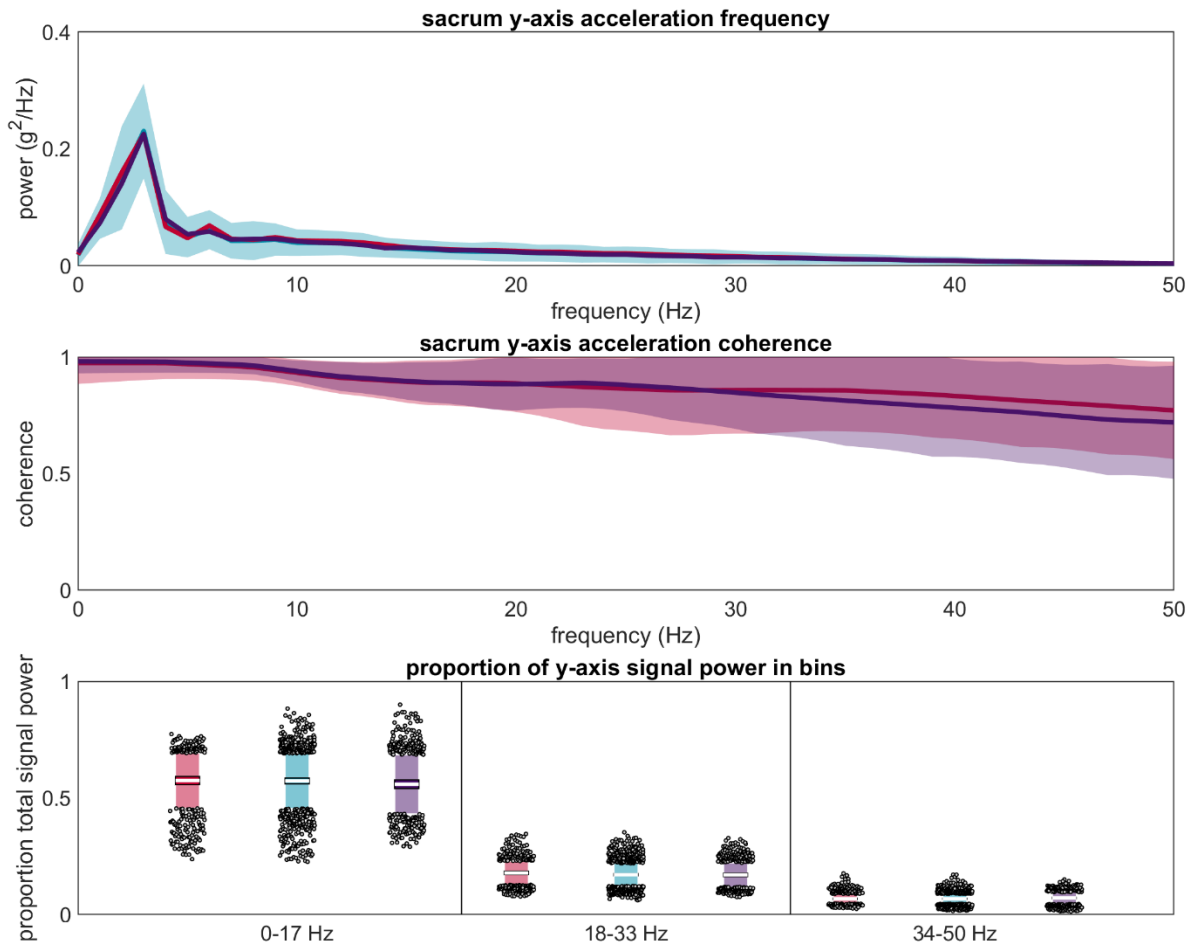


**Figure 5S.30: Frequency analysis of pelvis z-axis angular velocity. (Top)** Power spectral density from 0 to 50 Hz. The blue line represents the mean for the reference IMU, the blue shaded area represents  $\pm 1$  SD, and the red and purple lines represent the anterior-proximal and posterior-proximal misplaced IMUs. **(Middle)** Magnitude squared coherence between each misplaced IMU and the reference IMU. Plotted as mean  $\pm 1$  SD. **(Bottom)** Proportion of total signal power observed in the 0-17 Hz, 18-33 Hz, and 34-50 Hz bins. Central white line represents the mean across trials, dark shaded area represents  $\pm 1.96$  SEM (confidence interval; difficult to see due to small size), light shaded area represents  $\pm 1$  SD, and grey dots represent trials falling outside  $\pm 1$  SD.

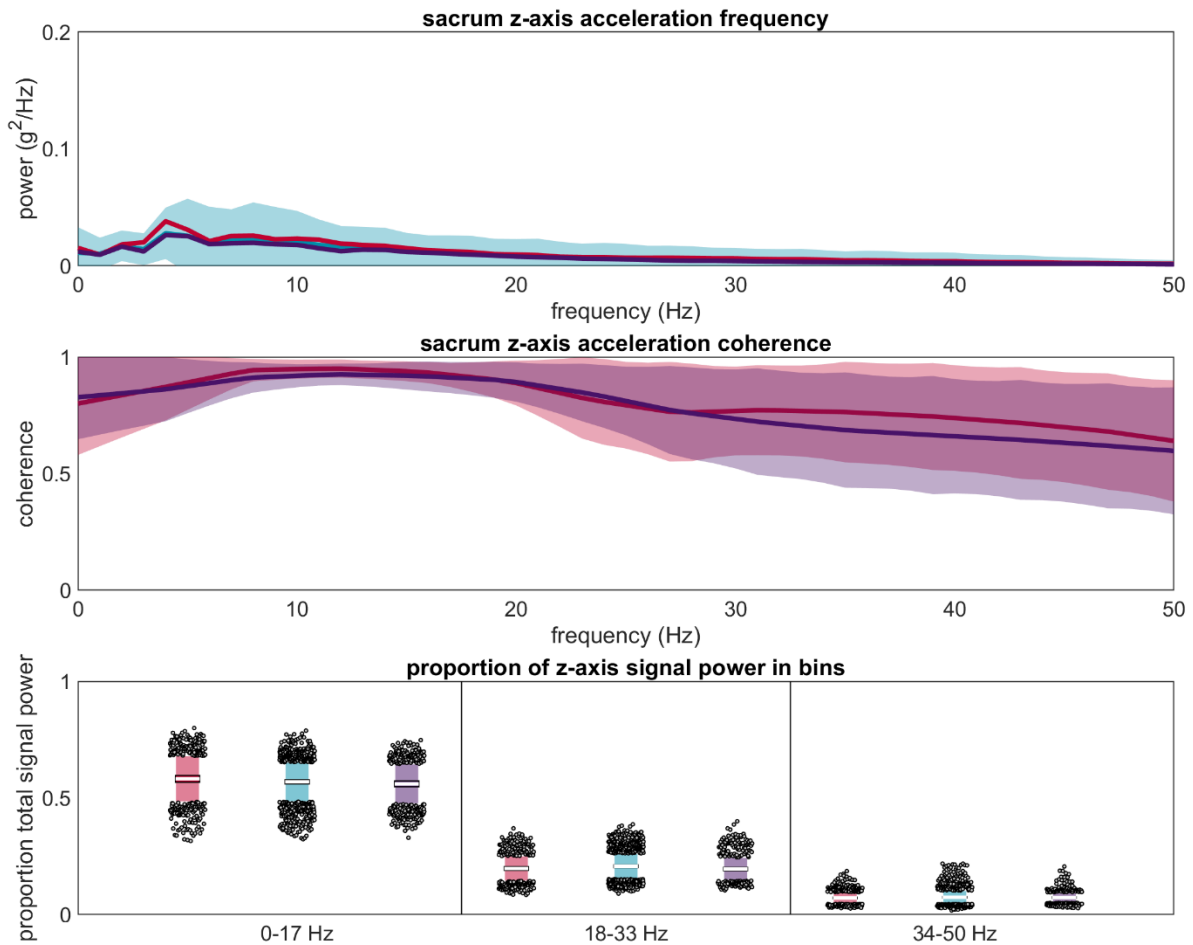
5S.2.5. Sacrum acceleration



**Figure 5S.31: Frequency analysis of sacrum x-axis acceleration.** (Top) Power spectral density from 0 to 50 Hz. The blue line represents the mean for the reference IMU, the blue shaded area represents  $\pm 1$  SD, and the red and purple lines represent the left-proximal and right-proximal misplaced IMUs. (Middle) Magnitude squared coherence between each misplaced IMU and the reference IMU. Plotted as mean  $\pm 1$  SD. (Bottom) Proportion of total signal power observed in the 0-17 Hz, 18-33 Hz, and 34-50 Hz bins. Central white line represents the mean across trials, dark shaded area represents  $\pm 1.96$  SEM (confidence interval; difficult to see due to small size), light shaded area represents  $\pm 1$  SD, and grey dots represent trials falling outside  $\pm 1$  SD.

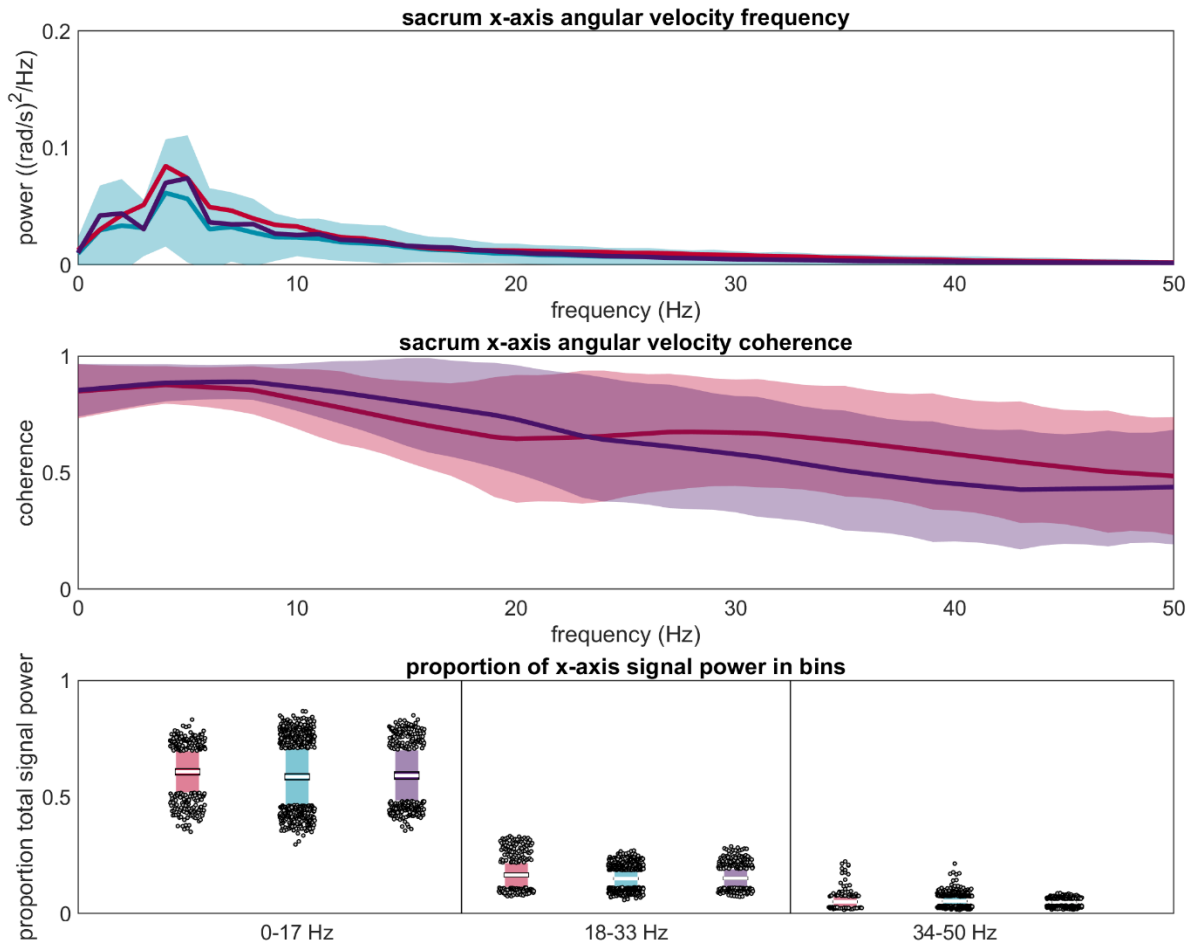


**Figure 5S.32: Frequency analysis of sacrum y-axis acceleration. (Top)** Power spectral density from 0 to 50 Hz. The blue line represents the mean for the reference IMU, the blue shaded area represents  $\pm 1$  SD, and the red and purple lines represent the left-proximal and right-proximal misplaced IMUs. **(Middle)** Magnitude squared coherence between each misplaced IMU and the reference IMU. Plotted as mean  $\pm 1$  SD. **(Bottom)** Proportion of total signal power observed in the 0-17 Hz, 18-33 Hz, and 34-50 Hz bins. Central white line represents the mean across trials, dark shaded area represents  $\pm 1.96$  SEM (confidence interval; difficult to see due to small size), light shaded area represents  $\pm 1$  SD, and grey dots represent trials falling outside  $\pm 1$  SD.

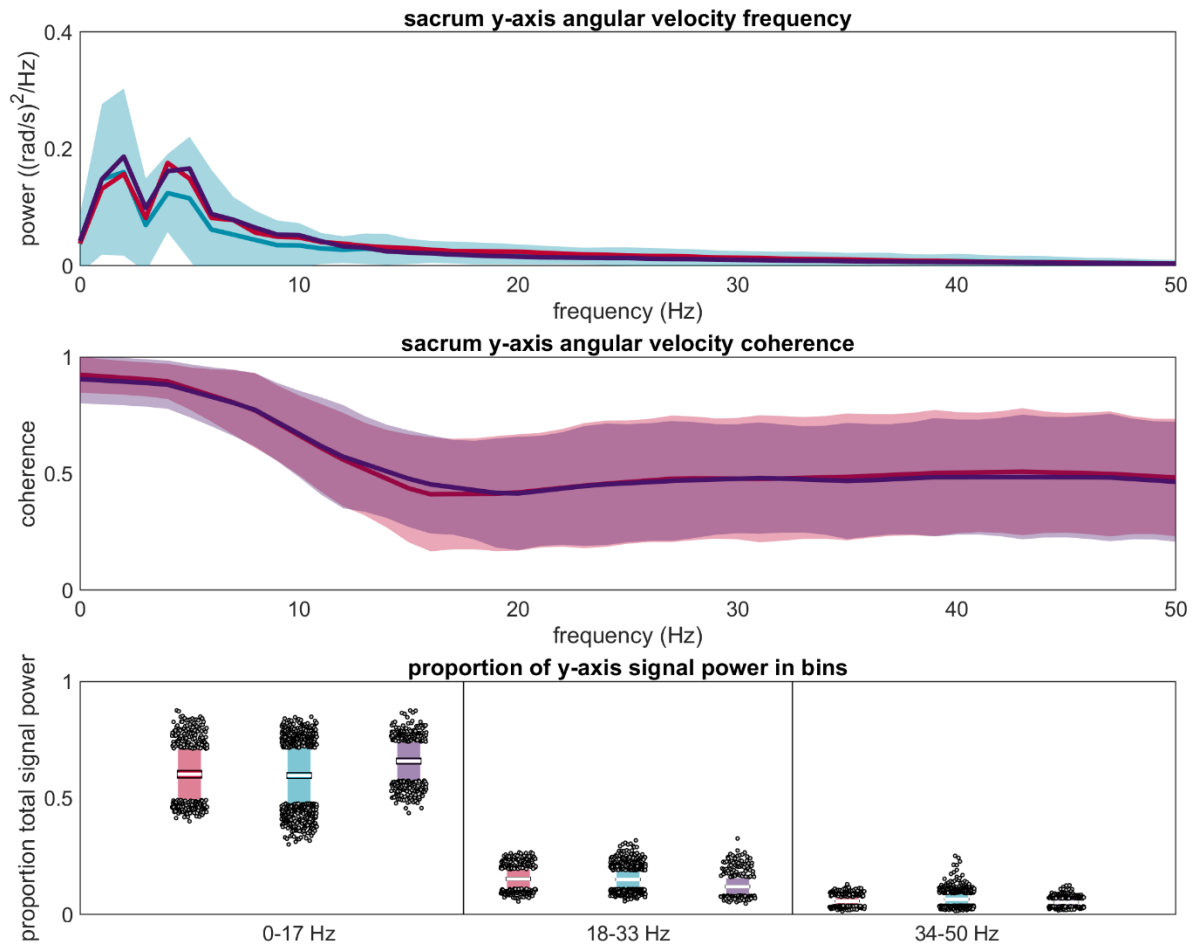


**Figure 5S.33: Frequency analysis of sacrum z-axis acceleration. (Top)** Power spectral density from 0 to 50 Hz. The blue line represents the mean for the reference IMU, the blue shaded area represents  $\pm 1$  SD, and the red and purple lines represent the left-proximal and right-proximal misplaced IMUs. **(Middle)** Magnitude squared coherence between each misplaced IMU and the reference IMU. Plotted as mean  $\pm 1$  SD. **(Bottom)** Proportion of total signal power observed in the 0-17 Hz, 18-33 Hz, and 34-50 Hz bins. Central white line represents the mean across trials, dark shaded area represents  $\pm 1.96$  SEM (confidence interval; difficult to see due to small size), light shaded area represents  $\pm 1$  SD, and grey dots represent trials falling outside  $\pm 1$  SD.

5S.2.6. Sacrum angular velocity

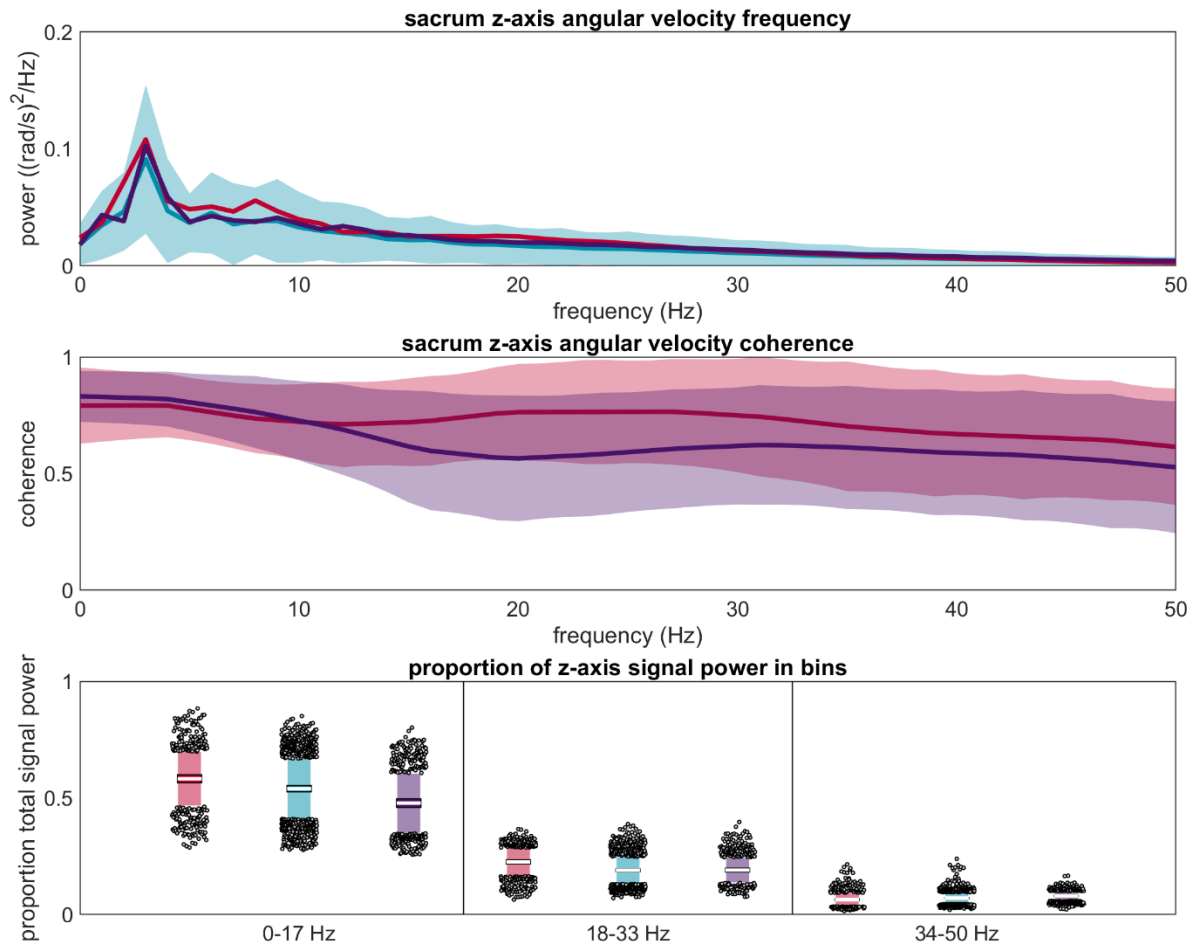


**Figure 5S.34: Frequency analysis of sacrum x-axis angular velocity. (Top)** Power spectral density from 0 to 50 Hz. The blue line represents the mean for the reference IMU, the blue shaded area represents  $\pm 1$  SD, and the red and purple lines represent the left-proximal and right-proximal misplaced IMUs. **(Middle)** Magnitude squared coherence between each misplaced IMU and the reference IMU. Plotted as mean  $\pm 1$  SD. **(Bottom)** Proportion of total signal power observed in the 0-17 Hz, 18-33 Hz, and 34-50 Hz bins. Central white line represents the mean across trials, dark shaded area represents  $\pm 1.96$  SEM (confidence interval; difficult to see due to small size), light shaded area represents  $\pm 1$  SD, and grey dots represent trials falling outside  $\pm 1$  SD.



**Figure 5S.35: Frequency analysis of sacrum y-axis angular velocity. (Top)** Power spectral density from 0 to 50 Hz. The blue line represents the mean for the reference IMU, the blue shaded area represents  $\pm 1$  SD, and the red and purple lines represent the left-proximal and right-proximal misplaced IMUs. **(Middle)** Magnitude squared coherence between each misplaced IMU and the reference IMU. Plotted as mean  $\pm 1$  SD. **(Bottom)** Proportion of total signal power observed in the 0-17 Hz, 18-33 Hz, and 34-50 Hz bins. Central white line represents the mean across trials, dark shaded area represents  $\pm 1.96$  SEM (confidence interval; difficult to see due to small size), light shaded area represents  $\pm 1$  SD, and grey dots represent trials falling outside  $\pm 1$  SD.



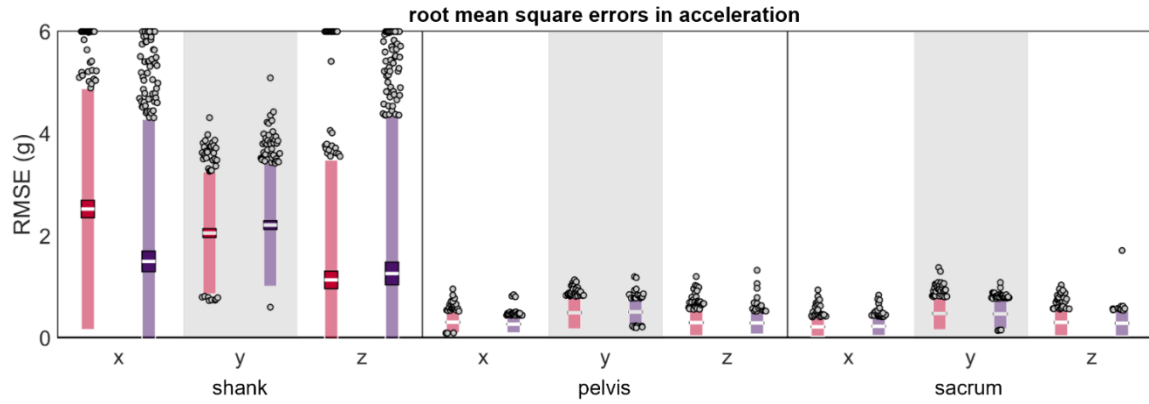


**Figure 5S.36: Frequency analysis of sacrum z-axis angular velocity. (Top)** Power spectral density from 0 to 50 Hz. The blue line represents the mean for the reference IMU, the blue shaded area represents  $\pm 1$  SD, and the red and purple lines represent the left-proximal and right-proximal misplaced IMUs. **(Middle)** Magnitude squared coherence between each misplaced IMU and the reference IMU. Plotted as mean  $\pm 1$  SD. **(Bottom)** Proportion of total signal power observed in the 0-17 Hz, 18-33 Hz, and 34-50 Hz bins. Central white line represents the mean across trials, dark shaded area represents  $\pm 1.96$  SEM (confidence interval; difficult to see due to small size), light shaded area represents  $\pm 1$  SD, and grey dots represent trials falling outside  $\pm 1$  SD.

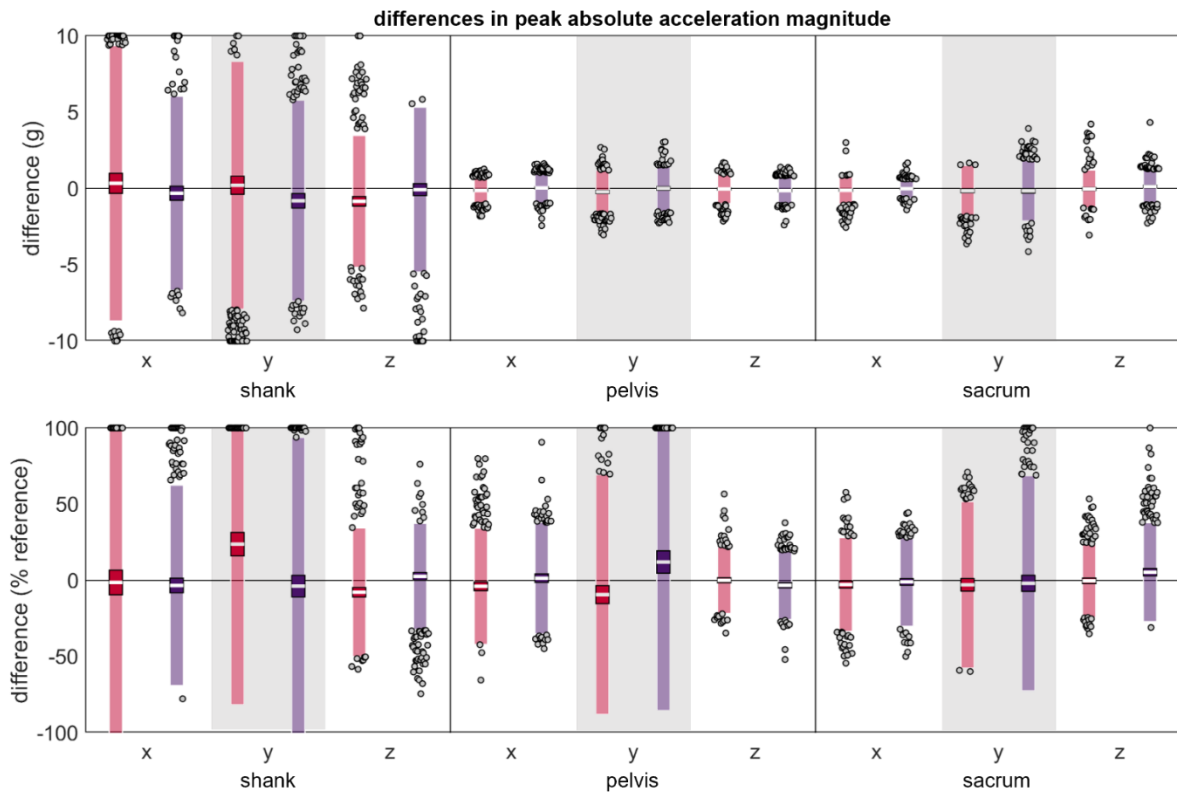
### 5.3 Wearable coordinate system results

The analyses presented in the main paper were repeated on the same data expressed in the wearable coordinate system (WCS) as opposed to the segment coordinate system (SCS) (see Chapter 3S for a detailed description of these coordinate systems). Analysis of derived outcome metrics could not be repeated due to difficulties in consistently identifying gait events with the WCS, making a direction

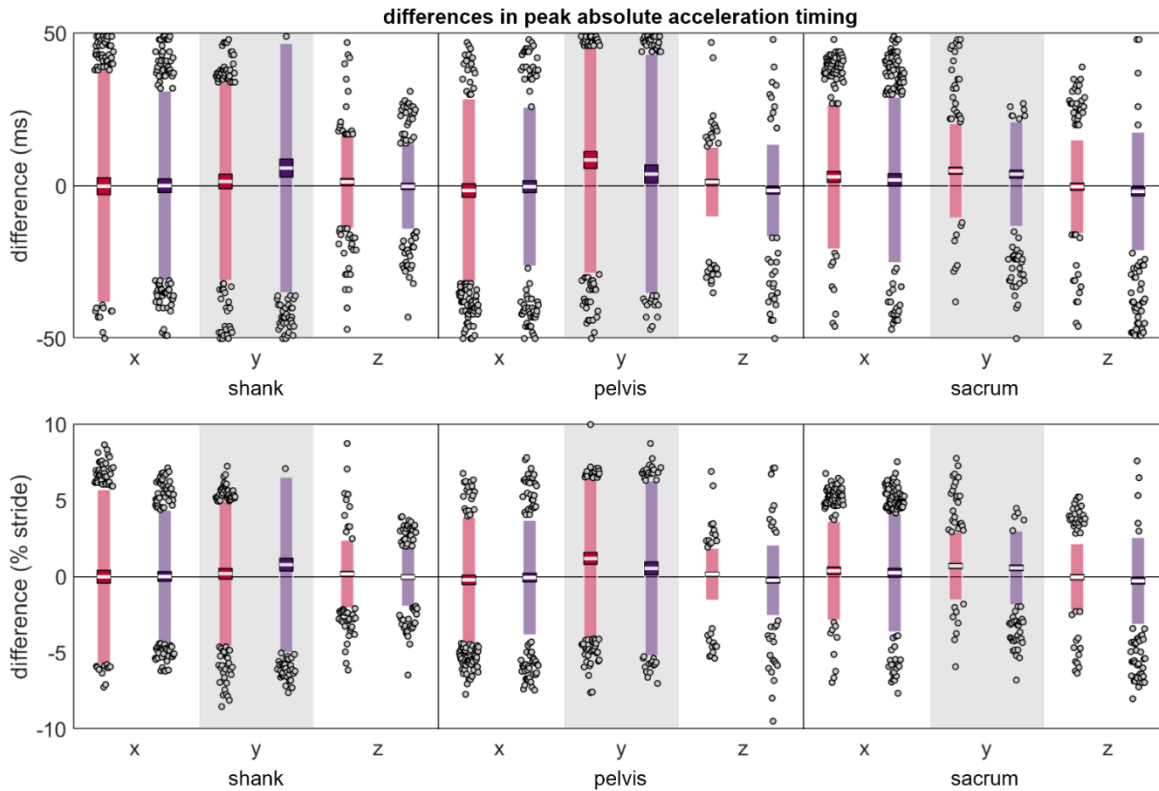
comparison between WCS and SCS impossible for those variables. Overall, the analyses revealed that misplacement tended to cause greater differences in the WCS than the SCS, particularly at the shank.



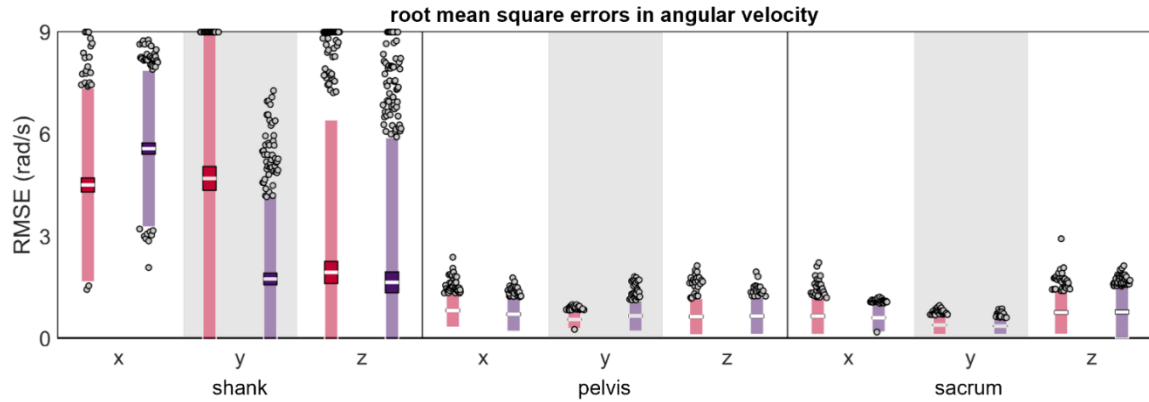
**Figure 5S.37: Wearable coordinate system root mean square error in acceleration.** RMSE between *a* measured simultaneously by the reference IMU and the misplaced IMU (red represents anterior- or left-proximal misplacement; purple represents posterior- or right-proximal misplacement). The white line represents the RMSE mean across all trials, the dark-colored box represents the confidence interval about the mean ( $\pm 1.96$  SEM), the light-colored box represents the limits of agreement ( $\pm 1.96$  SD), and the grey dots represent trials falling outside the limits of agreement (with trials falling outside the axis limit plotted at the limit).



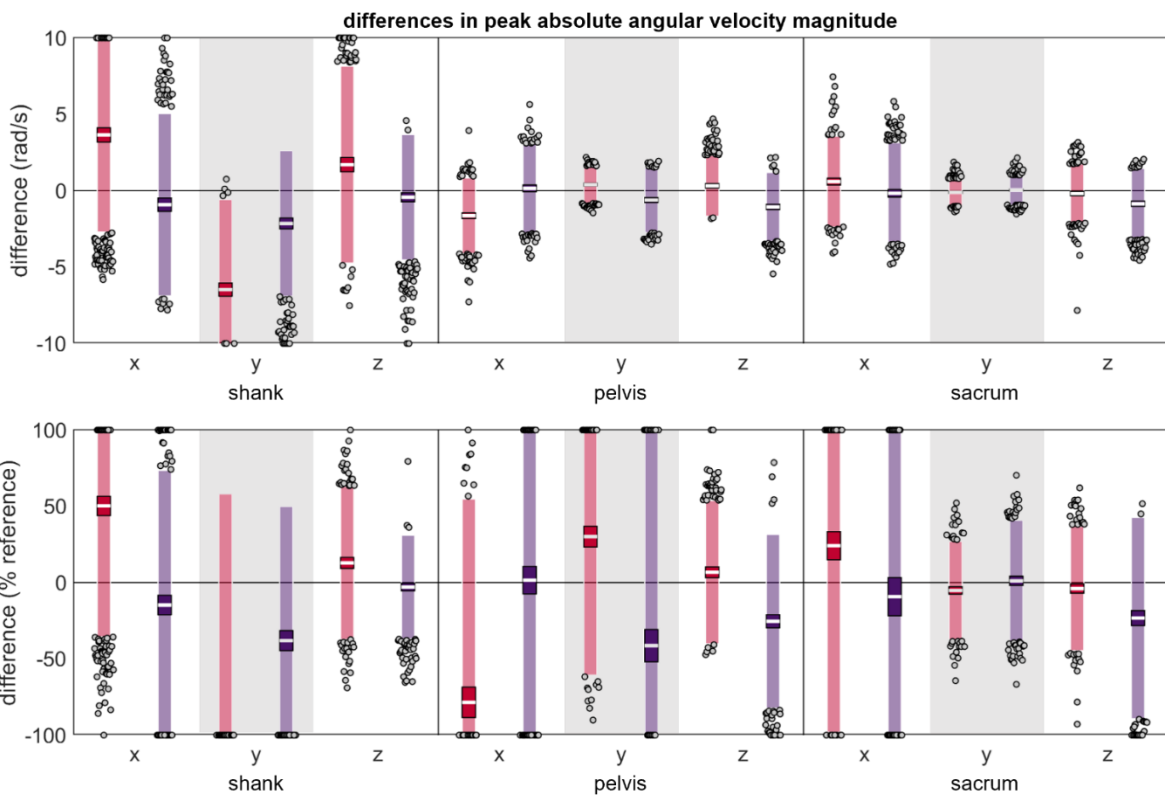
**Figure 5S.38: Wearable coordinate system differences between peak  $|a|$  magnitudes.** (A) absolute and (B) normalized (red represents anterior- or left-proximal misplacement; purple represents posterior- or right-proximal misplacement). The white line represents the mean observed difference across trials (bias), the dark-colored box represents the confidence interval about the mean ( $\pm 1.96$  SEM), the light-colored box represents the limits of agreement ( $\pm 1.96$  SD), and the grey dots represent trials falling outside the limits of agreement (with trials falling outside the axis limits plotted at the limit). Positive differences indicate the misplaced  $|a|$  magnitude was greater than the reference  $|a|$  magnitude while negative differences indicate the misplaced  $|a|$  magnitude was less than the reference  $|a|$  magnitude.



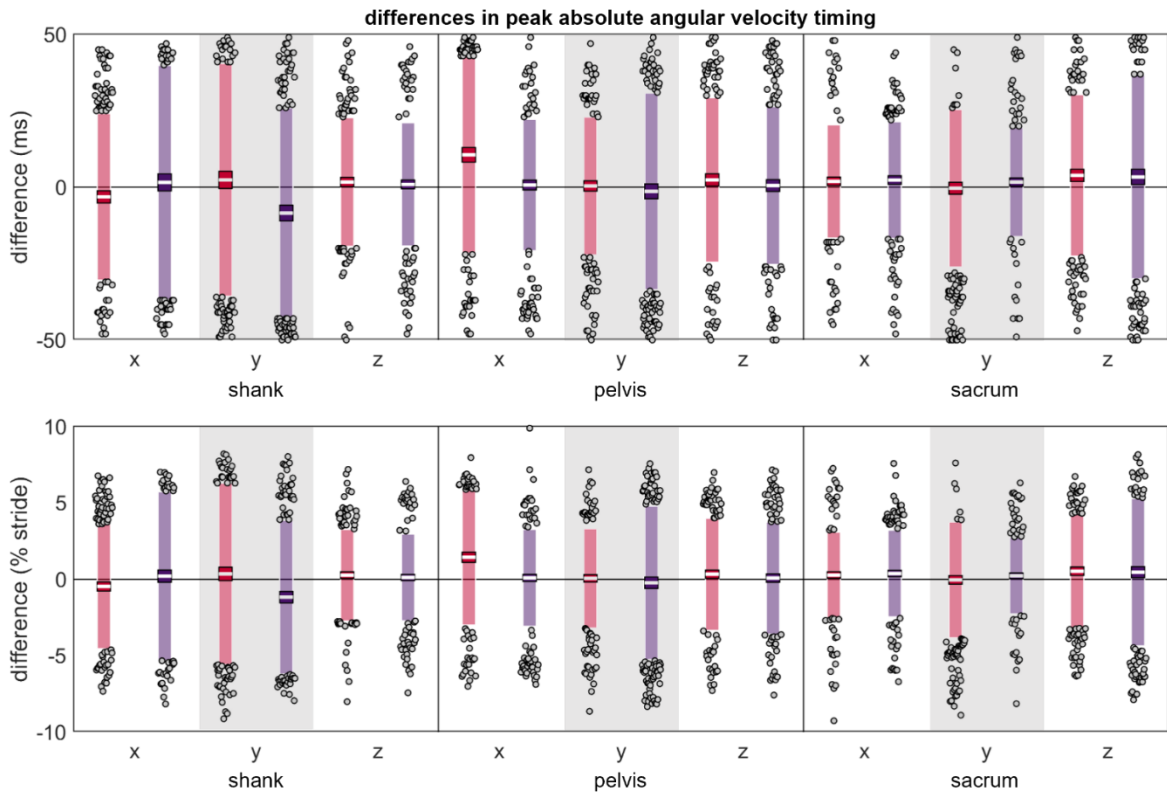
**Figure 5S.39: Wearable coordinate system differences between peak  $|a|$  timings.** (A) absolute and (B) normalized differences between peak  $|a|$  timings measured simultaneously by the reference IMU and the misplaced IMU (red represents anterior- or left-proximal misplacement; purple represents posterior- or right-proximal misplacement). The white line represents the mean observed difference across trials (bias), the dark-colored box represents the confidence interval about the mean ( $\pm 1.96$  SEM), the light-colored box represents the limits of agreement ( $\pm 1.96$  SD), and the grey dots represent trials falling outside the limits of agreement (with trials falling outside the axis limits plotted at the limit). Positive differences indicate the misplaced  $|a|$  peak occurred after the reference peak while negative differences indicate the misplaced  $|a|$  peak occurred before the reference peak.



**Figure 5S.40: Wearable coordinate system root mean square error in angular velocity.** RMSE between  $\omega$  measured simultaneously by the reference IMU and the misplaced IMU (red represents anterior- or left-proximal misplacement; purple represents posterior- or right-proximal misplacement). The white line represents the RMSE mean across all trials, the dark-colored box represents the confidence interval about the mean ( $\pm 1.96$  SEM), the light-colored box represents the limits of agreement ( $\pm 1.96$  SD), and the grey dots represent trials falling outside the limits of agreement (with trials falling outside the axis limit plotted at the limit).



**Figure 5S.41: Wearable coordinate system differences between peak  $|\omega|$  magnitudes.** (A) absolute and (B) normalized differences between peak  $|\omega|$  magnitudes measured simultaneously by the reference IMU and the misplaced IMU (red represents anterior- or left-proximal misplacement; purple represents posterior- or right-proximal misplacement). The white line represents the mean observed difference across trials (bias), the dark-colored box represents the confidence interval about the mean ( $\pm 1.96$  SEM), the light-colored box represents the limits of agreement ( $\pm 1.96$  SD), and the grey dots represent trials falling outside the limits of agreement (with trials falling outside the axis limits plotted at the limit). Positive differences indicate the misplaced  $|\omega|$  magnitude was greater than the reference  $|\omega|$  magnitude while negative differences indicate the misplaced  $|\omega|$  magnitude was less than the reference  $|\omega|$  magnitude.



**Figure 5S.41: Wearable coordinate system differences between peak  $|\omega|$  timings. (A) absolute and (B) normalized differences between peak  $|\omega|$  timings measured simultaneously by the reference IMU and the misplaced IMU (red represents anterior- or left-proximal misplacement; purple represents posterior- or right-proximal misplacement). The white line represents the mean observed difference across trials (bias), the dark-colored box represents the confidence interval about the mean ( $\pm 1.96$  SEM), the light-colored box represents the limits of agreement ( $\pm 1.96$  SD), and the grey dots represent trials falling outside the limits of agreement (with trials falling outside the axis limits plotted at the limit). Positive differences indicate the misplaced  $|\omega|$  peak occurred after the reference peak while negative differences indicate the misplaced  $|\omega|$  peak occurred before the reference peak.**

## 6. Cumulative loading does not relate to injury development over a 12-week prospective period

### 6.1. Introduction

Running is a relatively accessible way to achieve healthy levels of physical activity, providing cardiovascular health benefits that reduce morbidity and mortality [24]. Unfortunately, running is also associated with high rates of musculoskeletal overuse injury: As many as 33 running-related injuries (RRIs) occur per 1000 hours of running [25]. With millions of runners world-wide [9], these RRIs create health and economic burdens, and prevent people from maintaining healthy levels of physical activity [28, 31]. Thus, preventing RRIs is of considerable interest. Unfortunately, despite much high-quality research, the incidence of RRI remains high and new injury prevention strategies are needed [39, 25].

Most RRIs are overuse injuries caused by microtrauma that accumulates when musculoskeletal structures are exposed to repeated loading (e.g., stresses/strains). Trauma occurs when the combined number and magnitude of loads exceeds the threshold a structure can sustain [46, 59]. *In vivo*, it is difficult to measure or model this structure loading and microtrauma accumulation [66, 69]. To overcome this difficulty, biomechanists often assume that external loads imposed on the body are highly correlated with structure loading and therefore relate to injury risk. Based on this assumption, a frequent target of investigation in RRI research has been ground reaction forces measured in-lab (GRFs; external forces imposed on the foot by the ground with each step). Unfortunately, evidence relating GRF and RRI is inconsistent and knowing a runner's in-lab GRF appears insufficient to predict and prevent RRIs [68, 41].

One potential reason for this inconsistency is that assuming that GRFs are highly correlated with structure loading may be incorrect: Although GRF contributes to structure loading, structure loading is

also dictated by structure material properties, muscle forces, and other factors. Thus, the relation between GRF and structure loading is confounded and may be non-linear [69]. Despite these confounds, GRFs have explained up to 94% of the variance in loading of some structures during level-ground running [70]. Thus, during level-ground running, GRFs may provide a reasonable surrogate of structure loading and injury risk.

Another potential reason GRFs have not consistently predicted RRI in past research, is that previous research has focused on relations between RRIs and GRF magnitudes determined from a ‘representative’ stance created by averaging a few stances observed in a lab setting [73]. Such research is constrained in several ways: (1) loading data obtained in labs may not accurately represent loading in the real world; (2) loading may change throughout a repetitive task; for example, fatigue can alter biomechanics across long runs; and (3) given the importance of repetitive loading in RRI, metrics based on a single representative stance tell an incomplete story that is likely insufficient to predict RRI [77]. Therefore, extrapolation of lab-based, single-stance GRFs does not realistically represent the number or magnitude of GRFs actually experienced by runners, potentially explaining why GRFs have not consistently predicted RRIs in previous lab-based research.

To overcome these lab-based limitations, Kiernan et al. calculated GRF ‘profiles’ (combinations of load numbers and magnitudes) over long time periods in real-world settings by following NCAA Division 1 Track runners for two months of in-the-field training [198]. For every run during this two-month training period, Kiernan et al. calculated the number of right foot strikes and their GRF peak magnitudes. GRF numbers and magnitudes were then weighted based on human musculoskeletal structural properties to create a metric of ‘cumulative loading’ [84]. Using this approach, Kiernan et al. were able to collect a large amount of GRF data (>400 runs and >2 million strides) and found that injured runners had significantly higher cumulative loading than uninjured runners. This finding suggests that GRF numbers and magnitudes—when measured in the field—may predict RRI. This finding

demonstrates the value of collecting real-world GRF profiles (load number-magnitude combinations) over single representative lab-based GRFs.

To build on this proof-of-concept and test the hypothesis that GRF magnitudes and numbers measured in the field can be used to calculate cumulative loads that predict RRI, the current study uses a larger and more diverse free-running sample in combination with improved methodology [207, 296, 324]. These methodological advancements improve GRF estimation and gait event identification and validate all methods across different foot strike kinematics, running surfaces, and running speeds thus providing more confidence in any findings. With these improvements, the present study provides a well validated, robust, and accurate comparison of GRF magnitudes and numbers across runners who do and do not develop RRIs. Results from this study overcome limitations of previous in-lab measurements and provide ecologically valid estimates of real-world load magnitudes and numbers. Further, in comparison to single representative stances, measuring GRFs from each step across the entire study duration will better capture the repetitive loading that leads to overuse injury and the range and variance of GRF magnitudes and numbers within- and between-runners.

This approach will allow us to explore the relative importance of GRF magnitudes and numbers to prospective injury development and combine them into a weighted metric of cumulative loading. If field-measured cumulative loads are associated with injury, this approach will allow us to index risk as a function of GRF number and magnitude combinations. Such findings could be used to monitor and provide feedback on cumulative loading in runners and prevent RRI by avoiding 'high risk' training.

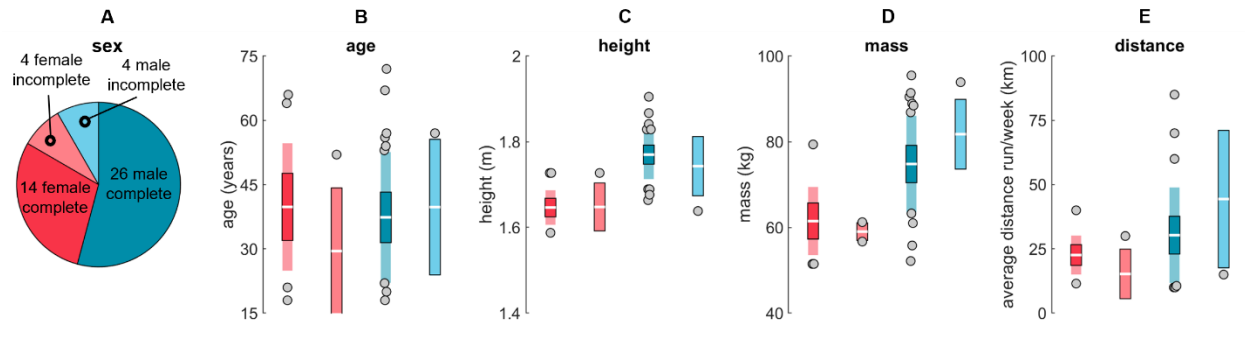


## 6.2. Methods

### 6.2.1. Sample size calculations and participants

Sample size calculations were based on previous work which found that 33 of 101 participants had  $\geq 1$  week where they missed  $\geq 4$  days of training due to RRI during a 26-week prospective study [118, 117]. Adjusting for a 54% shorter 12-week study duration, we estimated that  $\sim 15\%$  of participants would receive an RRI that caused  $\geq 4$  days of time-loss during  $\geq 1$  week. Using this percentage to calculate an uninjured/injured group allocation ratio, cumulative loading effect sizes calculated from previous work [198], a power of 0.8, and an alpha of 0.05, we calculated that a total sample size of 24 participants was required [325]. Eight percent attrition was observed in similar prospective research [118, 117]. To account for attrition or the possibility that fewer injuries would occur than predicted, we conservatively over-sampled and recruited double the subjects required.

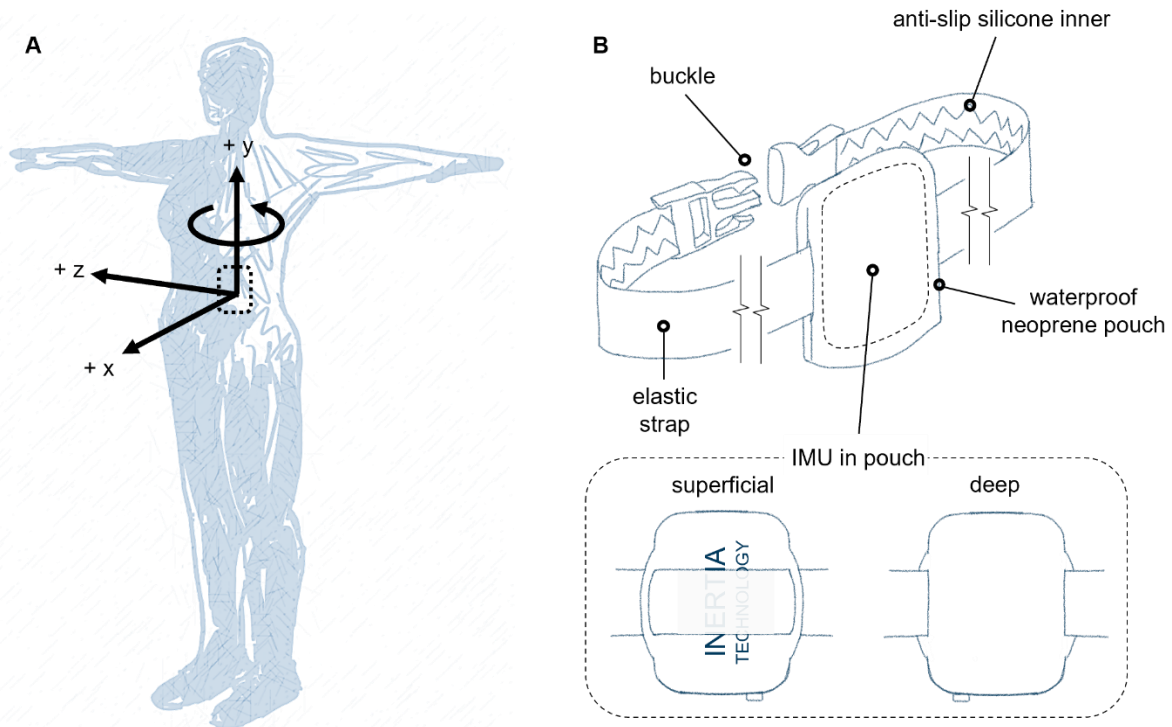
Thus, 48 participants  $\geq 18$  years old were recruited from the University of California Davis, local running clubs, and the community at large (18 female; 30 male; 0 non-binary; age  $37.60 \pm 15.02$  years; Figure 6.1). All participants reported (1) running  $\geq 16.09$  km per week for  $\geq 1$  year, (2) not having visited a medical professional for an RRI for  $\geq 1$  year, (3) willingness to wear an IMU on a belt every time they ran for 12-weeks, and (4) willingness to track their running and injury with a weekly survey for 12-weeks. Participants provided written informed consent and procedures were approved by the University of California, Davis Institutional Review Board. Participants were compensated with 10 USD at baseline and at weeks 1, 2, 7, and 12 of the study for a total of 50 USD.



**Figure 6.1: Participant details.** Sample (A) sex, (B) age, (C) height, (D) mass, and (E) self-reported average distance run per week for male (blue) and female (red) participants that did (darker colors) and did not (lighter colors) complete the entire 12-week experiment. The white horizontal line represents the mean, the darker central box represents  $\pm 95\%$  confidence interval ( $\pm 1.96$  SEM) around the mean, the lighter outside box represents  $\pm 1$  SD around the mean (not easily visible for incomplete participants due to low sample sizes increasing size of confidence intervals). Gray dots represent participants outside  $\pm 1$  SD.

### 6.2.2. Baseline data collection

Participants had their height and weight measured, completed an interview describing their running and injury history (including previous experience with races, previous injury, etc.; see Chapter 6S), and provided their daily running distances for the previous 4-weeks of running. Participants were trained to fix an inertial measurement unit (IMU; Inertia Technology ProMove MINI;  $\pm 16$  g primary,  $\pm 100$  g secondary,  $\pm 2000$  deg/s, 200 Hz) to their sacrum using an elastic belt with an anti-slip silicone inner and a waterproof neoprene pouch for the IMU (Figure 6.2). Participants were then trained to collect simple static and dynamic calibration movements (standing still then rotating forward and backward about the z-axis of the ankle joint like an inverted pendulum) that could be used to create a segment coordinate system.



**Figure 6.2: IMU placement and belt design. (A)** IMU placement and coordinate conventions. A coordinate system was defined as anterior (+x), proximal (+y), and medial-lateral (with right defined as +z) and right-left lateral bending, left-right axial rotation, and posterior-anterior tilt were defined about the x, y, and z axes with the right hand rule [169]. **(B)** Belt design and IMU fixation. Elastic straps with anti-slip silicone inners were threaded through the IMU within a waterproof neoprene pouch. These straps were buckled around the waist such that IMUs were placed on the superior aspect of the sacrum in line with the spine.

In-lab foot strike angle and ground reaction force data were collected as participants ran a 25 m runway with an embedded force plate (Kistler 9281, Kistler Group, Winterthur, Switzerland; 1000 Hz) while wearing their own shoes. Running speed was recorded using two custom-built laser speed gates, placed 2.5 m on each side of force plate center. Participants struck the force plate five times per side and were encouraged to progress through the range of speeds they typically use while running from slowest to fastest. Foot kinematics were recorded using a conventional camera (Exilim EX-FH25, Casio; 120 Hz) and foot strike angle was calculated by subtracting a neutral standing foot angle (Kinovea 0.9.5) [270].

Participants then (1) completed a detailed walk-through of the online survey (Qualtrics, Provo, Utah) they would receive each week to track their running and RRI, (2) were given a hard-copy tracking sheet to help them record details for the survey (Chapter 6S), (3) were given an IMU, charger, and belt, and (4) were sent an email summarizing the experiment, their action items as a participant, and a schedule of subsequent visits with the experimenter.

### *6.2.3. Prospective survey data*

After leaving the lab, participants began a 12-week prospective data collection. Each Sunday a link to an online survey was emailed to the participants (Chapter 6S). If the survey was not completed, automatic email reminders were sent out 24 and 48 h after the initial email. If a participant still had not responded after 36 h they were contacted by an experimenter. Participants reported:

(1) injury in three discrete ways:

(a) as pain (0-10 Likert scale and location),

(b) as number of training days lost and/or modified due to RRI, and

(c) as medical visits due to RRI,

(2) daily running distance, duration, and rating of perceived exertion (RPE) [326],

(3) daily non-steroidal anti-inflammatory drug use,

(4) heart rate mean, max, and variability and the device used to record it<sup>2</sup>

(5) stress across the week (0-10 Likert scale),

---

<sup>2</sup> No heart monitors were provided, however, if participants had their own, they were encouraged to provide these details.

- (6) weight,
- (7) daily sleep quality (0-10 Likert scale),
- (8) daily perception of recovery/readiness (0-10 Likert scale),
- (9) whether they wore their IMU during their run,
- (10) any other physical activity, and
- (11) any free form comments they thought were important to note

Masses on days between the day the survey was completed and the next reported mass were interpolated to give an estimate of day-by-day participant mass. These day-by-day masses were used for estimation of GRFs (see below).

#### *6.2.4. Prospective IMU data*

Participants were instructed to wear the IMU they had been given on every run for the entire 12-week period. During that period, they were instructed to maintain their typical running habits and not modify their running in any way for the experiment with one exception: Participants were asked to perform calibration movements at the start of every run. These calibration movements consisted of standing straight and still then rotating forward and backward about the z-axis of the ankle joint like an inverted pendulum. During this calibration and throughout the run, the IMU collected triaxial magnetometer, accelerometer, and gyroscope data at 200 Hz.

Each IMU had a 16 GB onboard capacity and was capable of storing > 200 hours of data at 200 Hz. Thus, the IMUs were likely capable of storing all 12-weeks of data; however, participants were required to exchange their IMU at/around weeks 1, 2, 7, and 12. During these exchanges all data were

downloaded from the IMU. These exchanges were used to ensure that the IMUs were in working order and that the participants were recording data properly, minimizing the potential for data loss.

#### *6.2.5. Attrition management*

If participants repeatedly (> 2 times) failed to exchange the IMU, had large discrepancies between the running reported in their survey and the files recorded on their IMU, or did not complete their survey, they were removed from the study by the experimenter. Participants could also voluntarily withdraw from the study due to the burden involved (or for any other reason). If participants did not run or record data for a week but still completed the survey it was deemed part of the normal variability in their running habits and thus an important feature of the data we wanted to collect (e.g., on vacation, lay-off after large race, injured).

#### *6.2.6. IMU data processing*

IMU data were filtered with a 4th-order 50-Hz low-pass Butterworth filter. The orientation of the IMU was calculated based on constraints in the fixation system (the belt limited the IMU to four potential orientations) and the observed mean  $\sim$ vertical acceleration and ratio of mean  $\sim$ anterior versus mean  $\sim$ posterior peaks. After determining the orientation, data were re-expressed in a coordinate system defined as  $\sim$ anterior (+x),  $\sim$ proximal (+y), and  $\sim$ medial-lateral (with right defined as +z) with  $\sim$ right-left lateral bending,  $\sim$ left-right axial rotation, and  $\sim$ posterior-anterior tilt defined about the x, y, and z axes with the right hand rule [169] (Figure 6.2). The gravity component was removed from accelerations by subtracting 1 g from the y axis.

Next, epochs of steady state movement with at least 10 consecutive peaks of at least 1 g occurring at least 210 ms apart were identified. Stances during each steady state running epoch were

identified by estimating initial and terminal contact gait events via the ‘Auvinet’ method [162, 207]. Then, for each stance, the second (‘active’) peak vertical GRF was estimated from y-axis accelerations using the ‘Veras’ method [198, 255]. Finally, GRF second peak magnitudes falling outside the acceptable range for human running (< 1.25 bodyweights) were eliminated [327] and epochs from a given day of running were concatenated together for further analysis.

### 6.2.7. IMU outcome metrics

For each day, the total number of stances was calculated. The GRF second peak magnitude was scaled with an allometric scaling factor predicted from dimensionality theory (Equation 6.1) then the mean was calculated for each day.

**Equation 6.1: Allometric scaling.**

$$L_s = P_s m^A$$

Where  $L_s$  is the allometrically-scaled ‘load’ for stance  $s$ ,  $P_s$  is the GRF second peak magnitude estimated for stance  $s$  in N,  $m$  is the participant’s mass in kg, and  $A$  is the allometric scaling factor -0.67 [328].

To calculate cumulative weighted load (CWL) for each day, Equation 6.2 was adapted from Edwards [143] and Firminger et al. [329].

**Equation 6.2: Cumulative weighted load.**

$$CWL = \left( \sum_{s=1}^{s=n} (L_s^b) \right)^{\frac{1}{b}}$$

Where  $s$  is a stance,  $n$  is the total number of stances for a given day,  $L_s$  is the load (allometrically-scaled GRF second peak magnitude) for stance  $s$  raised to  $b$ , and  $b$  is 9.3 based on the power law line of best fit

that relates applied strain magnitudes to the number of cycles to failure for human Achilles tendon [6, 84]. Finally, the result is raised to the power of  $\frac{1}{b}$  to reduce the magnitude [143, 329].

#### 6.2.8. Analyses

Participants who did and did not report a time-loss injury were divided into injured and uninjured groups, respectively. Mean self-reported session rating of perceived exertion/day, self-reported distance/day, self-reported pain, IMU-estimated steps/day, IMU-estimated allometrically-scaled loads, and IMU-estimated cumulative weighted loads/day were compared across injured and uninjured groups via independent samples t-tests with significance set at  $p \leq 0.05$ . For the uninjured group these means were taken across the duration of the study. In contrast, for the injured group, these means were taken from the first day of the study until they first reported a time loss injury.

In an alternate analysis, cumulative weighted loads were entered into equations that have been proposed to account for the structural properties and dynamic remodeling response of the musculoskeletal system in order to capture injury risk (Table 6.1). Although many variations of these 'risk models' have been proposed, in an investigation of 19 of these variations, a previous simulation study [330] suggested they are highly correlated; thus, only the subset that was least correlated and best able to categorize runners that developed injury was used here. This subset consisted of (1) ACWR (6 day:21 day), (2) strain, (3) exponentially weighted moving average (10 day), and (4) week-to-week change (see Table 6.1). For each of these four risk models, average uninjured and injured scores across the duration of data collection were compared via independent samples t-test. To further visualize any potential differences between injured and uninjured runners, the mean and standard deviation were calculated for the uninjured runners across the study duration and individual injured runner's scores were plotted against these values.



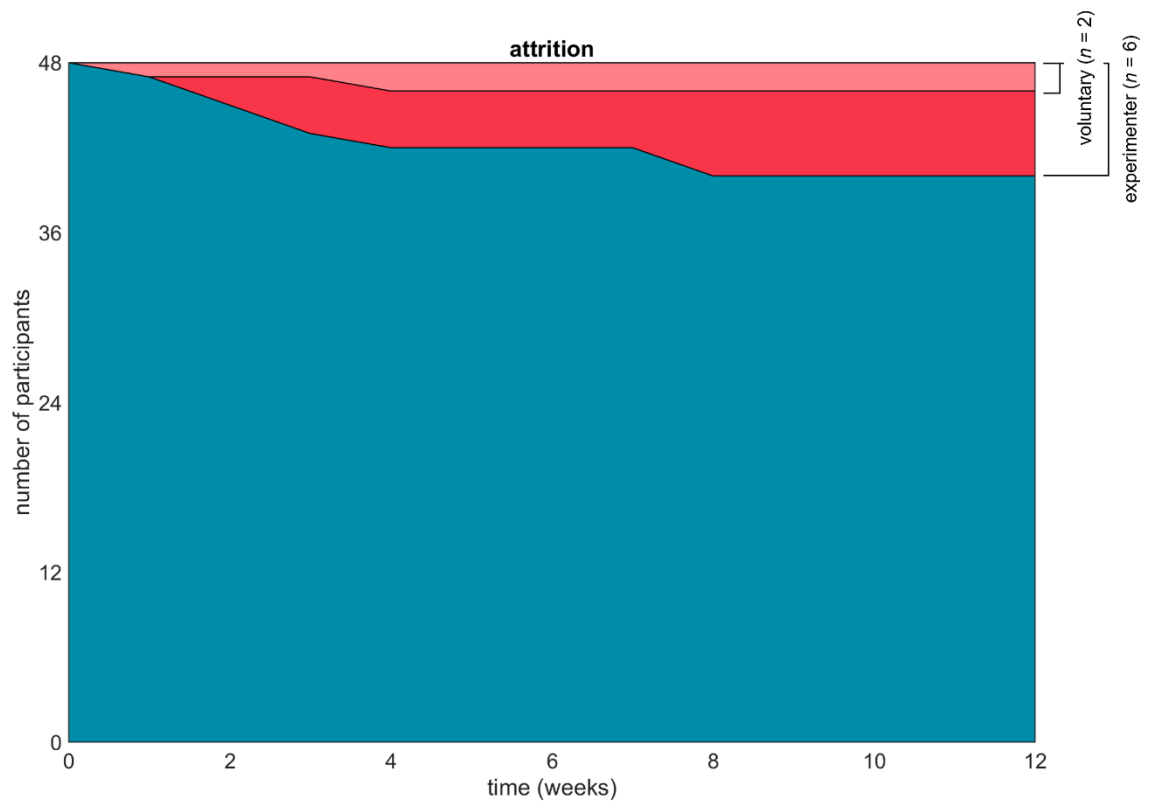
**Table 6.1: ‘Risk models’ proposed to predict overuse injury.** Equations proposed to predict overuse injury in the literature. Although many variations of these equations have been proposed a smaller subset was used here.

Risk model	Equation	Details
Mean daily workload	$\bar{L}_x = \frac{\sum_{t=0}^{t=-x+1} L_t}{x}$	$L_t$ is the loading for a given day, $t = 0$ is the current day, and $x$ is the total number of days
Cumulative workload	$C_x = \sum_{t=0}^{t=-x+1} L_t$	Where $x = \{7, 14, 21, 28\}$
Exponentially-weighted moving average (EWMA)	$E_{F,0} = FL_0 + (1 - F)E_{F,-1}$	Where $F = \{\frac{1}{20}, \frac{1}{10}, \frac{1}{5}, \frac{1}{2}\}$ and $E_{F,-1}$ is yesterday’s EWMA
Acute:Chronic Workload Ratio (ACWR)	$R_{x:y} = \frac{\sum_{t=0}^{t=-x+1} L_t}{\sum_{t=0}^{t=-y+1} L_t/y/x}$	Where $x = \{3, 6, 7\}$ and $y = \{21, 28\}$
Uncoupled ACWR	$U_{x:y} = \frac{\sum_{t=0}^{t=-x+1} L_t}{\sum_{t=-7}^{t=-y+1} L_t/y/x}$	Where $x = 7$ and $y = 28$
Exponential ACWR	$X = \frac{E_{\frac{1}{7},0}}{E_{\frac{1}{28},0}}$	Where $E_{\frac{1}{7},0}$ is EWMA for 1 week and $E_{\frac{1}{28},0}$ is EWMA for 4 weeks
Week-to-week change	$\Delta W = \left  C_7 - \sum_{t=-7}^{t=-13} L_t \right $	This week’s minus last week’s cumulative loads
Training monotony	$M = \bar{L}_7 / \sqrt{\frac{\sum_{t=0}^{t=-6} (L_t - \bar{L}_7)^2}{7}}$	One week’s mean divided by standard deviation
Training strain	$S = MC_7$	One week’s cumulative workload multiplied by monotony

## 6.3. Results

### 6.3.1. Compliance and attrition

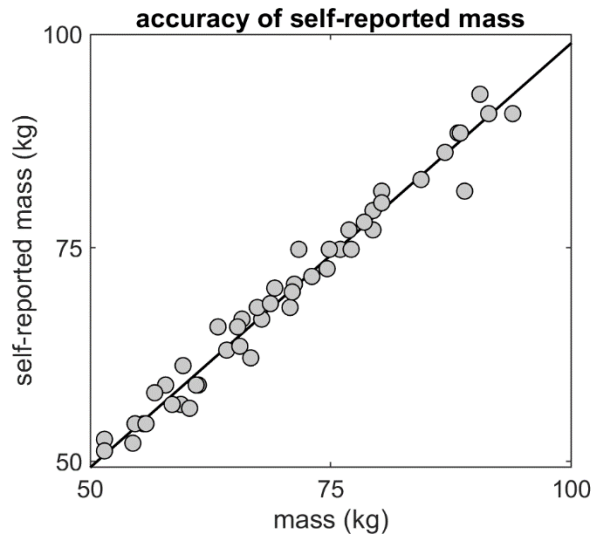
Across the 12-week prospective period, two participants voluntarily withdrew due to the burden of the study. An additional six participants were removed by the experimenter for failing to complete 2 consecutive surveys ( $n = 5$ ) or failure to accurately report/record data ( $n = 1$ ) (i.e., IMU and survey data were mismatched). The 40 remaining participants had 100% survey compliance.



**Figure 6.3: Attrition.** Stacked area showing the number of participants that withdrew voluntarily (pink;  $n = 2$ ), were withdrawn by the experimenter (red;  $n = 6$ ), and that completed the experiment (blue;  $n = 40$ ) over time in weeks.

### 6.3.2. Accuracy of self-reported mass

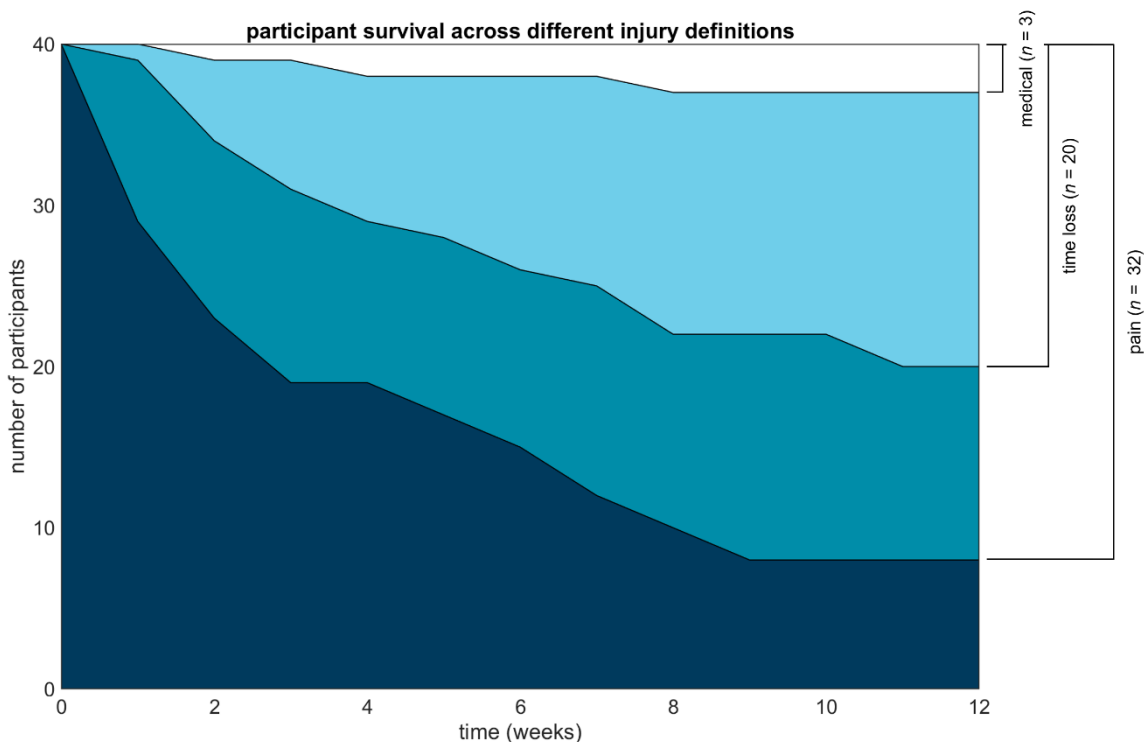
Mass was recorded twice at baseline: once via self-report and once by the experimenter. This procedure was done to assess the accuracy of self-reported masses across the prospective study duration (which are critical to accurately estimating GRF from recorded accelerations). Results suggest that self-reported mass accurately reflected participants' true mass: Mean errors between self-reported and true mass were  $1.58 \pm 1.38$  kg and the coefficient of determination ( $r^2$ ) was 0.99.



**Figure 6.4: Self-reported vs. true mass at baseline.** Participants' self-reported masses were highly accurate suggesting that self-reported mass from the prospective period can be confidently used to estimate GRF.

### 6.3.3. Injury

Injury was recorded via weekly survey in three different ways: (1) an injury that required a medical visit, (2) an injury that required time off from running, and (3) an injury that resulted in pain. Across the 12-week study duration three participants reported a medical injury, 20 reported a time loss injury, and 32 reported a pain injury. For the remainder of this Chapter, the term “injury” will be defined as a time loss injuries. This definition yields equal numbers of participants in the injured and uninjured groups ( $n = 20$  each).



**Figure 6.5: Injury development across time and definitions.** The number of participants that remained uninjured across time using different definitions of injury. Light blue represents a medical definition, blue represents a time loss definition, and dark blue represents a pain definition.

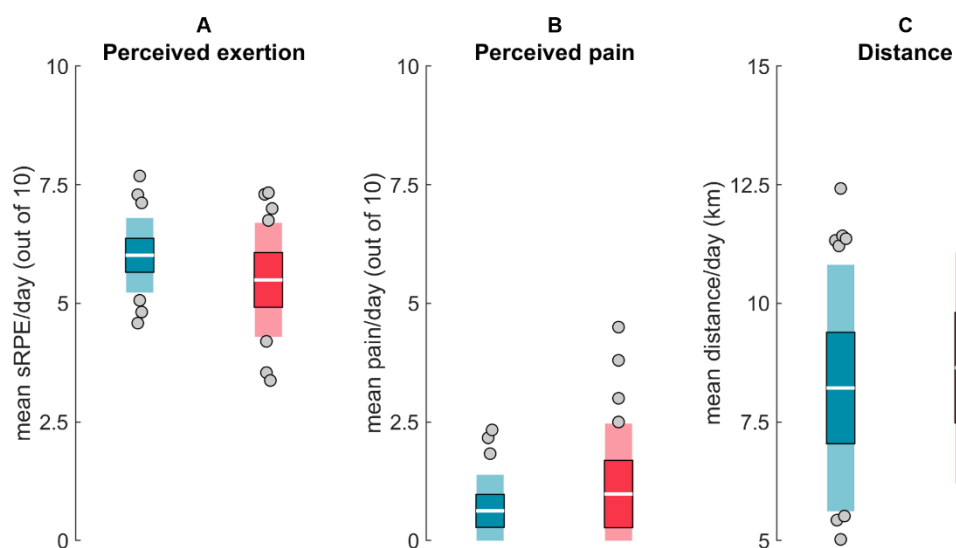
#### 6.3.4. Self-reported outcome variables across groups

Self-reported outcome variables were compared across injured and uninjured groups to evaluate whether they were predictive of injury (Table 6.2 and Figure 6.6). Of these, only the percentage of days that non-steroidal anti-inflammatory drugs (NSAIDs) were used varied significantly across groups at  $p \leq 0.05$ , with the injured group using NSAIDs more often than the uninjured group in the lead up to injury.

**Table 6.2: Self-reported outcome variables across groups.** Participants that did and did not sustain a time loss injury are divided into the injured and uninjured groups, respectively ( $n = 20$  each). Values represent mean  $\pm$  SD across the duration of the study for the uninjured group and across all time points preceding the time-loss injury for the injured group. Bolded values highlighted in pink indicate significance at  $p < 0.05$  with no correction for multiple comparisons.

Variable	Uninjured ( $n = 20$ )	Injured ( $n = 20$ )
Age (years)	35.68 $\pm$ 12.87	42.06 $\pm$ 17.98
Sex (% male)	57.89%	70.59%

sRPE (out of 10)	6.01 ± 0.80	5.45 ± 1.30
Pain (out of 10)	0.63 ± 0.77	1.8 ± 1.31
Sleep quality (out of 10)	6.76 ± 1.07	7.14 ± 1.13
Recovered (out of 10)	7.24 ± 1.52	7.20 ± 1.23
Stress (out of 10)	4.94 ± 1.62	4.71 ± 2.18
Mass (kg)	70.49 ± 10.65	68.80 ± 13.53
Distance run (km per day)	8.22 ± 2.61	8.98 ± 2.56
Running duration (minutes per day)	79.73 ± 23.79	79.46 ± 21.34
Running speed (m/s)	3.08 ± 0.90	3.07 ± 0.56
<b>NSAID use (% days used)</b>	<b>3.89 ± 1.77 %</b>	<b>11.15 ± 9.30 %</b>
Mean heart rate (BPM)	149.77 ± 13.00	153.22 ± 19.58
Max heart rate (BPM)	169.70 ± 12.75	170.61 ± 21.43

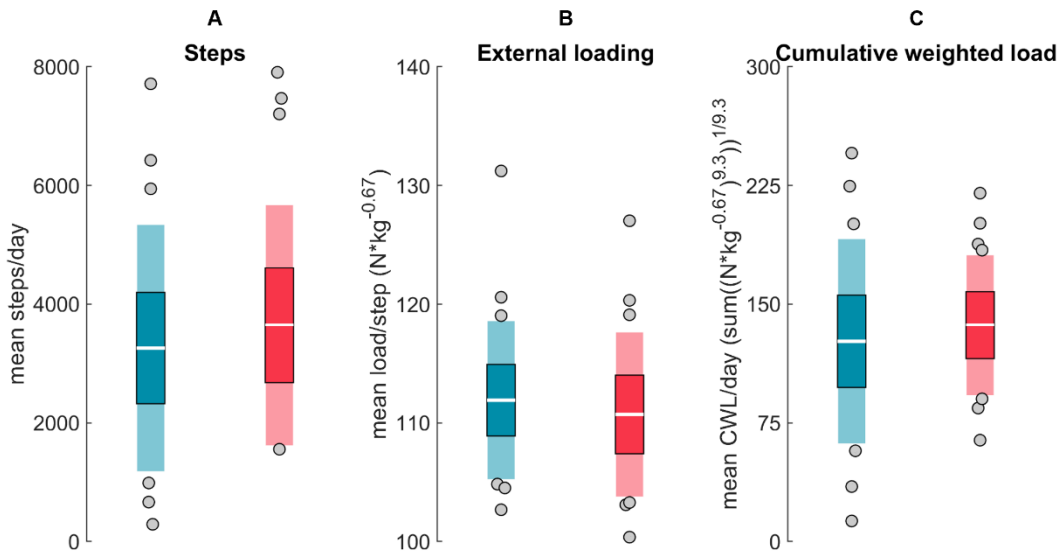


**Figure 6.6: Self-reported outcome variables across injured and uninjured runners. (A)** Mean session rating of perceived exertion from 0 (no exertion) to 10 (maximal exertion) across the study duration. **(B)** Mean perceived pain from 0 (no pain) to 10 (worst pain imaginable) across the study duration. **(C)** Mean self-reported distance run per day (in km) across the study duration. Uninjured group in blue. Injured group in red. The white horizontal line represents the mean, the darker central box represents ± 95% confidence interval (± 1.96 SEM) around the mean, the lighter outside box represents ± 1 SD around the mean, and gray dots represent participants outside ± 1 SD.

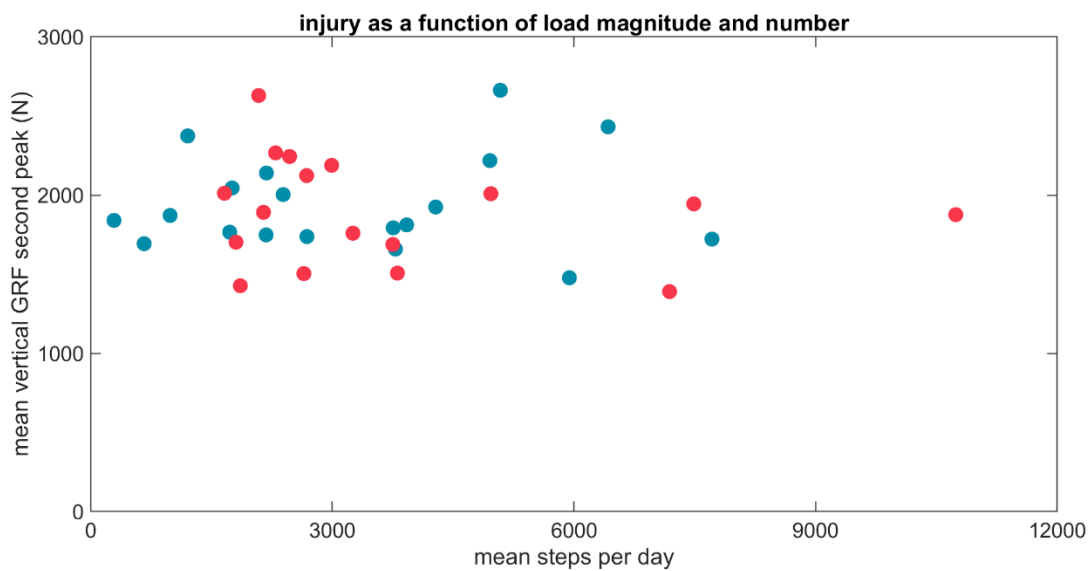
### 6.3.5. IMU-derived outcome variables across groups

The number of steps per day, the external load per step, and the cumulative load per day were compared across injured and uninjured groups (Figure 6.7). None of these outcomes varied significantly

across groups at  $p \leq 0.05$  and no clear pattern emerged when plotting injured and uninjured participants' load magnitudes and numbers (Figure 6.8)



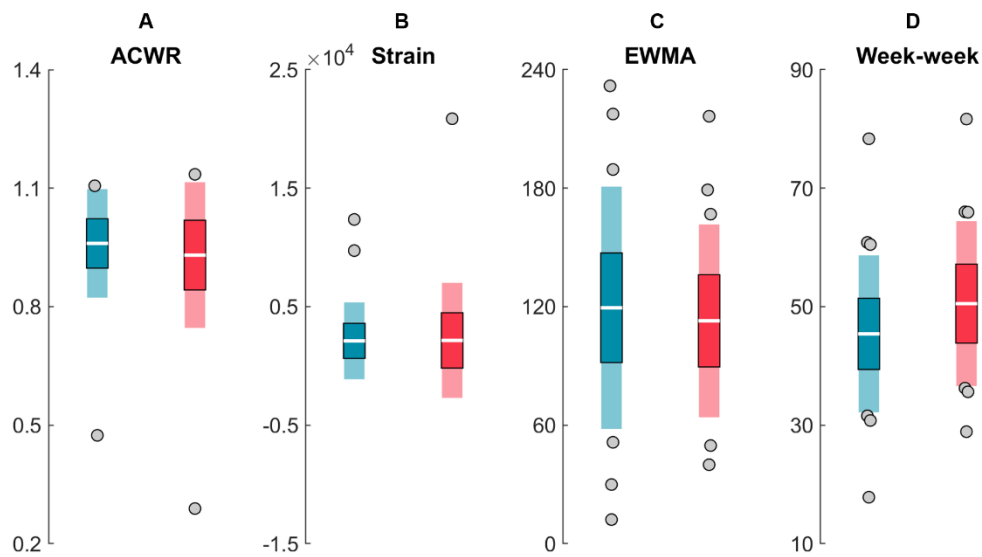
**Figure 6.7: IMU-estimated outcome variables across injured and uninjured runners. (A)** Mean number of steps run per day across the study duration. **(B)** Mean allometrically-scaled external load (GRF second peak magnitude) across all recorded steps. **(C)** Mean cumulative weighted load (CWL) per day across the study duration. Uninjured group in blue. Injured group in red. The white horizontal line represents the mean, the darker central box represents  $\pm 95\%$  confidence interval ( $\pm 1.96$  SEM) around the mean, the lighter outside box represents  $\pm 1$  SD around the mean, and gray dots represent participants outside  $\pm 1$  SD.



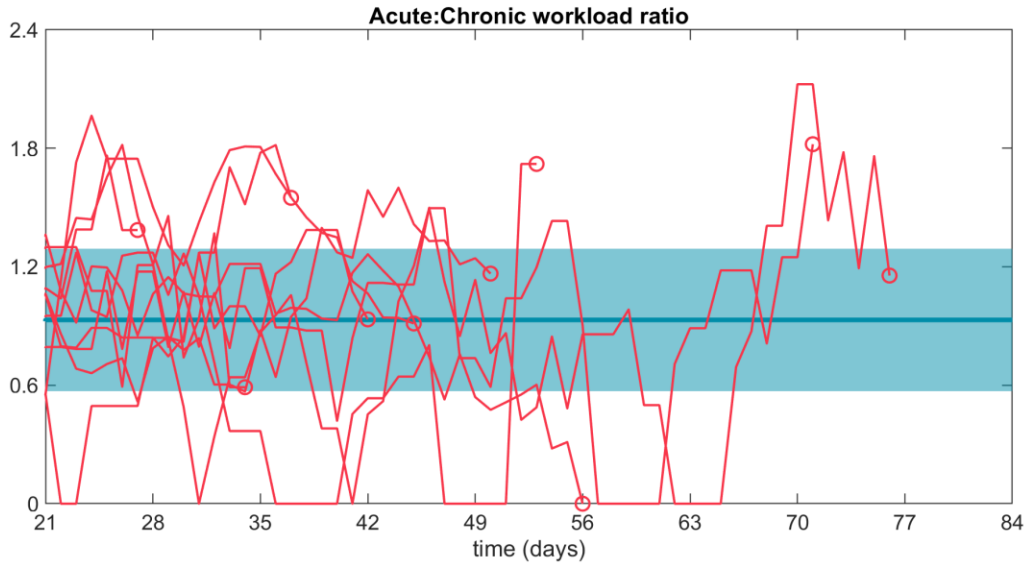
**Figure 6.8: Injury status as a function of load magnitudes and numbers.** Uninjured participants are plotted in blue, injured participants are plotted in red.

### 6.3.6. Risk models across groups

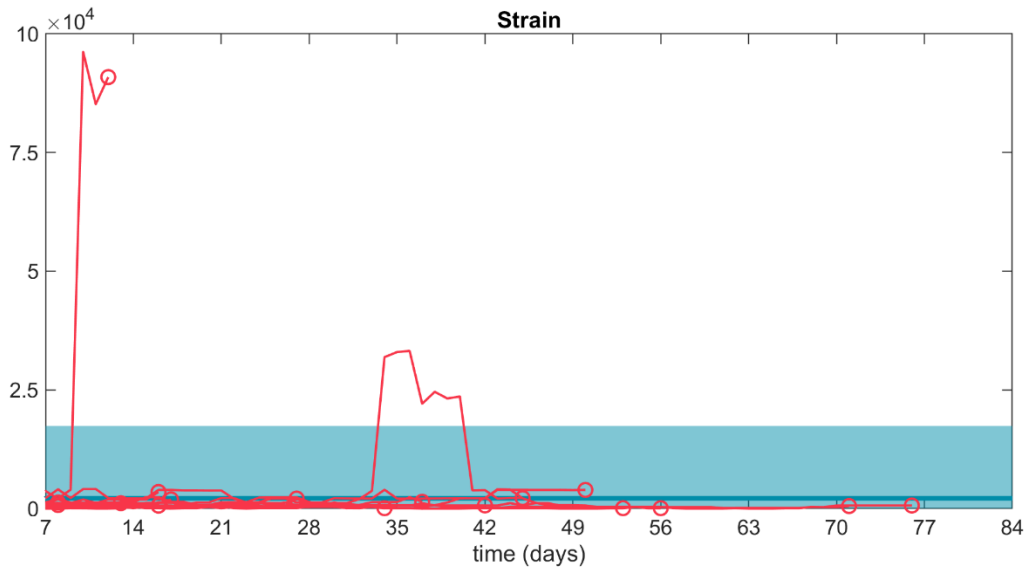
Cumulative weighted loads for each day were entered into four potential risk models: (1) ACWR (6:21), (2) strain, (3) exponentially-weighted moving average (1/10), and (4) week-to-week change. Independent samples t-tests comparing scores for injured and uninjured groups did not reveal any significant differences between any of these scores at  $p < 0.05$  (Figure 6.8). Patterns were also visualized by comparing injured participants scores over time against the mean and variance observed for uninjured runners (Figures 6.9-6.12).



**Figure 6.9: Risk model scores across injured and uninjured runners.** (A) Acute (6 day):Chronic (21 day) cumulative weighted loading. Unitless. (B) Training strain estimated from cumulative weighted loading. Expressed in original units  $((N \cdot \text{kg}^{-0.67})^{9.3})^{1/9.3}$ . (C) 10-day exponentially-weighted moving average of cumulative weighted loading. Expressed in original units  $((N \cdot \text{kg}^{-0.67})^{9.3})^{1/9.3}$ . (D) Week-to-week change in cumulative weighted loading. Expressed in original units  $((N \cdot \text{kg}^{-0.67})^{9.3})^{1/9.3}$ . Uninjured group in blue. Injured group in red. The white horizontal line represents the mean, the darker central box represents  $\pm 95\%$  confidence interval ( $\pm 1.96$  SEM) around the mean, the lighter outside box represents  $\pm 1$  SD around the mean, and gray dots represent participants outside  $\pm 1$  SD.

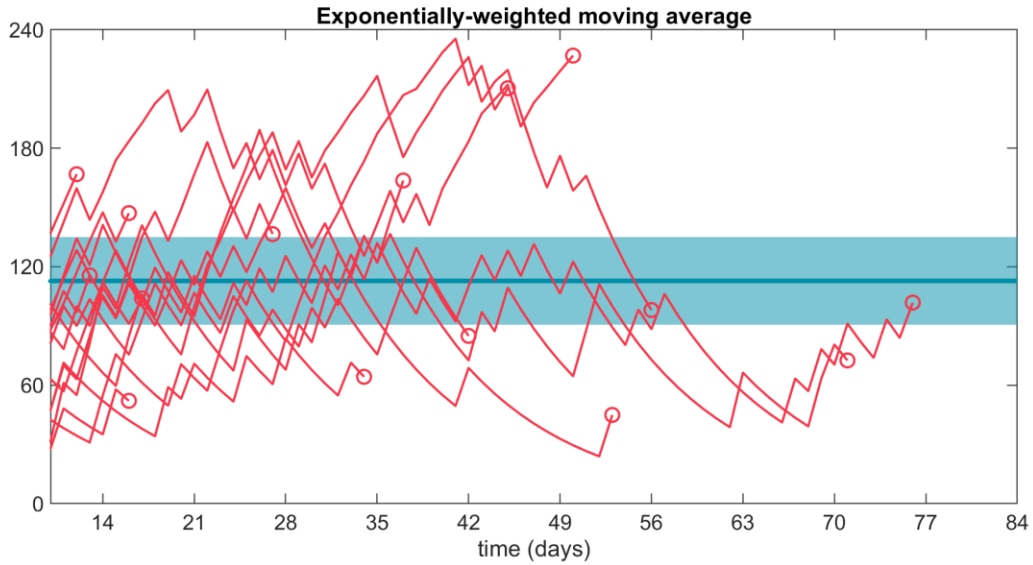


**Figure 6.10: Acute (6 day):Chronic (21 day) cumulative weighted loading over time.** Mean uninjured participant ACWR (dark blue line)  $\pm$  1.96 SD (blue shaded area) and individual injured participant ACWR scores (red lines) leading to first report of a time-loss injury (red circle). Unitless.

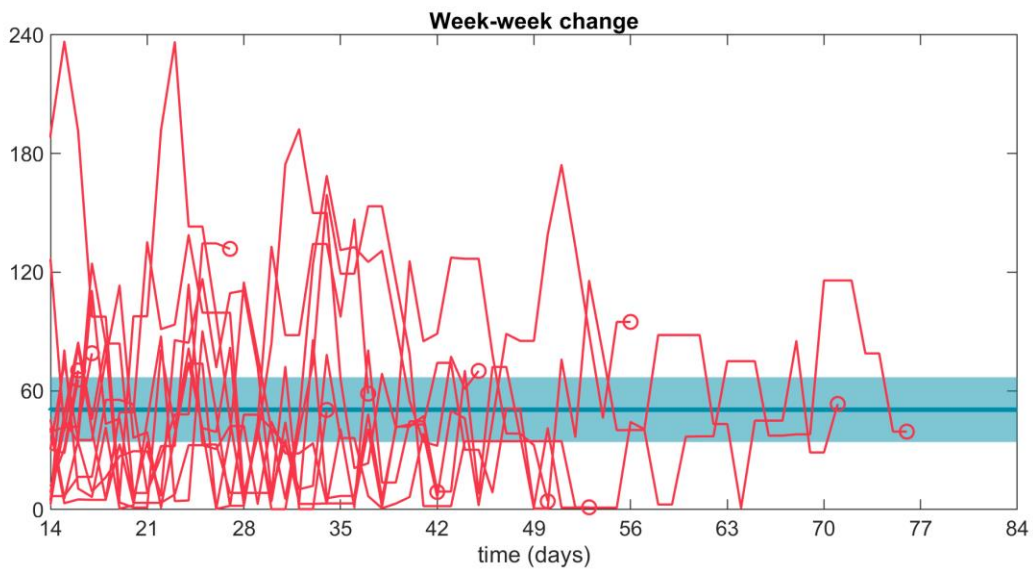


**Figure 6.11: Training strain over time.** Mean uninjured participant strain (dark blue line)  $\pm$  1.96 SD (blue shaded area) and individual injured participant strain scores (red lines) leading to first report of a time-loss injury (red circle). Expressed in original units  $((N \cdot kg^{0.67})^{9.3})^{1/9.3}$ .





**Figure 6.12: 10-day exponentially weighted moving average over time.** Mean uninjured participant EWMA (dark blue line)  $\pm$  1.96 SD (blue shaded area) and individual injured participant EWMA scores (red lines) leading to first report of a time-loss injury (red circle). Expressed in original units  $((N \cdot \text{kg}^{-0.67})^{9.3})^{1/9.3}$ .



**Figure 6.13: Week-to-week change in cumulative loading over time.** Mean uninjured participant week-week change (dark blue line)  $\pm$  1.96 SD (blue shaded area) and individual injured participant week-week change scores (red lines) leading to first report of a time-loss injury (red circle). Expressed in original units  $((N \cdot \text{kg}^{-0.67})^{9.3})^{1/9.3}$ .

## 6.4. Discussion

This study tests the hypothesis that real-world loading profiles (combinations of loading numbers and magnitudes) relate to the development of future injuries [100, 83]. To collect these profiles, 40 runners wore an inertial measurement unit (IMU) that collected their accelerations every time they ran for 12 weeks. Recorded accelerations were used to estimate each time the foot contacted the ground (load number) and the vertical ground reaction force second peak magnitude (external load magnitude). These two variables were weighted based off human musculoskeletal properties and combined to calculate the cumulative weighted load (CWL)—a surrogate of the musculoskeletal damage accrued during a run [143]. Over 10.3 million steps were recorded, making this the largest known study of external loading and running injury, and uniquely positioning it to overcome the limitations of previous lab-based cross-sectional and retrospective studies and shed light on the relation between external loading and running injury.

In contrast to the proof-of-concept study that pioneered this approach (presented in Chapter 2), the current study found no significant differences between the loading profiles of injured and uninjured runners. Indeed, neither the number of loads (stances), magnitude of loads (vertical GRF second peak magnitudes), nor CWLs significantly differed across groups. This failure to replicate the difference in load magnitudes and CWLs found in the original study may be due to several factors. First, the small sample size and exploratory nature of the original study ( $n = 9$ ) may have led to a Type I error (false positive) and the current results may better reflect the true nature of the relation (or lack thereof) between running injury and external loading operationalized as the vertical GRF second peak magnitude. This would suggest that future research and prevention efforts should focus on potential predictors of running injury other than vertical GRF second peak magnitude.

For example, some evidence suggests that the magnitude and/or rate of the first peak of the vertical GRF is more predictive of RRI than the magnitude of the second peak [68, 41]. We were not able to explore this hypothesis as existing methods to estimate vertical GRF first peak magnitude and rate of loading have low precision that would undermine any results [296]. Another method that has been advocated to operationalize external loading is the integral of the GRF waveform. Taking the integral captures potentially important information about load duration and loading throughout the entire gait cycle and thus may be a better predictor of RRI than a single-point magnitude [329]. While it is possible to estimate the integral of the GRF waveform from observed accelerations [283, 296], the computational processing time to do so here was prohibitive. The relation between RRI and external loads as operationalized as integrated GRF waveforms should, however, be explored in a future re-analysis of these data.

Ultimately, the most proximal causes of RRI will be internal loading, stresses, and strains. Unfortunately, these variables are currently difficult to measure/model and thus increasing access to these measurements should be a priority of overuse injury research. Further, even with accurate quantification of internal loading, the influence of these loads will be moderated by an individuals' musculoskeletal structure mechanical properties. We attempted to capture the general relation between these properties and loading by using an allometric scaling factor based on dimensionality theory [328], however, this is a very general approach that does not capture the inter-participant variation that is likely important in moderating the relation between loading and injury [69, 331]. Thus, another target of future research should be to develop methods to more easily measure/model musculoskeletal structure mechanical properties across participants (e.g., [332, 333]).

A second reason for the observed failure to replicate may be due to difficulty in appropriately defining and quantifying injury. Injury can be operationalized using different criteria such as pain, time-

loss, and medical visits [26, 114, 27, 115, 119] and each of these operationalizations may evince different relations with one another and with potential predictive variables [117, 118]. Although we recorded all three injury definitions, an *a priori* decision was made to use a time-loss definition. This decision was made due to previous findings that pain and injury are not related [116] (consistent with the current findings; Figure 6.6B) and due to a medical definition being too conservative to yield an adequately sized injured group (e.g., only three of 40 participants here sustained a medical injury). Another consideration when defining “injury” is the decision to consider an injury a function of *all* loading. Many RRIs are unilateral [39] and loading during running is often asymmetrical (e.g., [130]). Thus, future reanalysis of these data should use the reported locations of injury (see Chapter 6S) to define the side of the body that an injury occurred on and then relate only the ipsilateral loading to that injury.

A third possibility for the current null result is that the methods used here have obscured the relation between external loading and injury and led to a Type II error (false negative). Although the procedures used here are common across IMU-based research in the field, I do not believe they constitute “best practices.” Several key processing steps developed and validated across Chapters 3-5 (see Chapter 3S.4) were omitted here due to constraints in computational processing time: saturation and drift were not corrected, sensor fusion was not used to estimate orientation and subtract gravity, and data were left in a wearable coordinate system that is vulnerable to misplacement and between-participant differences in posture. Further, the ‘Veras’ method was used to estimate GRFs here because it provided the best balance of estimation accuracy and low computational processing time, however, superior methods are available. For example, the ‘Pogson’ methods provide a more accurate method with estimations that are more robust to differences in running speed and foot strike angle (an important factor in this study where runners’ behaviors and environments are unconstrained). The

computational cost of that method was, however, ~250x that of the Veras method and thus it was not used.

Another critical difference between the methods used in the proof-of-concept and those used here, was the lack of accounting for data missingness. From comparing the mean self-reported running distances in Figure 6.6C and the IMU-estimated number of steps in Figure 6.7A a large discrepancy can be observed: Mean self-reported distances for the uninjured and injured groups are 8.22 and 8.64 km respectively, at ~762 steps per km [53] we expect the mean number of steps to be ~6264 and ~6584; however, they are much lower at 3260.72 and 3646.00 respectively. This difference suggests a large amount of data missingness caused by participants forgetting to wear the IMU, turn the IMU on, charge the IMU, etc. These types of errors are common in this type of prospective research and can be corrected for by identifying and imputing missing data [108]. The omission of this missing data does, however, draw the current findings into question and future work should reprocess these data to address these methodological constraints and deal with missingness.

One final source that may potentially account for the observed difference between the current study and the proof-of-concept is that participants in the proof-of-concept were Division I running athletes (university student-athletes at the highest competitive level). These participants had training regimes that were highly constrained by their coach's prescriptions and thus there was relatively little inter-participant variability in the number of strides and the distances run. In contrast, the inclusion criteria for the current study was to run a minimum of  $\geq 16.09$  km per week (on average). Thus, the current sample is a much more heterogeneous group of runners with much more variability in the number of runs, steps per run, intensity of runs, time between runs, etc. The inclusion of this variability was intentional given the importance that each of these variables can play in RRI causation, it does, however, pose analytical challenges. Several approaches have been proposed to address these challenges by using 'risk models' to quantify the risk of injury over time (see Table 6.1). Although these

models are controversial [126, 330] they represent an important effort to identify the interplay between positive and negative remodeling and the critical time periods over which load accumulates to cause RRI. In the current study, none of the risk models evaluated differed across injured and uninjured runners. This may be due to one of (or a combination of) the reasons mentioned in the previous paragraphs or may be due to a disconnect between these models and the biological response to loading. Improved models may be developed with greater fidelity to these responses. For example, by calculating a daily probability of injury in a given structure based on the probability of positive and negative remodeling as a function of estimated structure loading, number of strides, and structure adaptation (e.g., [53]), or by applying Miner's Rule to calculate cumulative damage (e.g., [85]). Enacting these potential methods requires further research to determine mathematical associations between external loading conditions, internal structure loading, microtrauma accumulation, the temporal healing response, and RRI.

Despite the concerns and limitations presented above, the present finding that injured runners did not report higher pain replicates previous research and confirms that runners are not sensitive to impending injury. This finding underscores the need to continue developing objective metrics that can forecast injury. Indeed, given the high incidence of RRI and the burden that RRIs impose, these efforts remain critical [89, 39, 44]. Thus, the IMU-derived objective metrics used here represent an important effort in the field. The current method's capability to measure over 10 million steps in the real world demonstrates its ability to address several constraints in current biomechanics research. In the past, biomechanics research has largely been restricted to lab and clinical settings. This restriction has undermined the ecological validity of findings given lab-based observations may not be generalizable to real world behavior [73]. Further, the constraints of the lab have prevented accurate measurement of the number of loads actually experienced by runners – a variable theorized to play a critical role in RRI

causation [76, 61]. Thus, the method presented here is an important step toward ecologically valid measures of biomechanical loading magnitudes *and* numbers.

## **6.5. Conclusions**

In sum, no significant differences were observed in the loading profiles of runners who developed injury across a 12-week prospective period and those who did not. These results stand in contrast to an earlier proof-of-concept study and join a muddled literature. Future work is, however, needed to address a number of methodological limitations in the processing and analysis of the current results. After this work has been conducted it will provide strong evidence to either support or refute the hypothesis that external loading relates to running-related injury development. In either case, these results can help shape future biomechanics research by either guiding biomechanists to focus on alternative or moderating variables (e.g., internal loading, musculoskeletal structure mechanical properties) or by suggesting that measuring real-world external loading profiles is a viable method to predict and prevent injuries. Such results would help refine injury prediction models and provide the evidence necessary to develop adaptive feedback and training prescriptions that account for the mechanics and loading profiles of individual runners.

## **6.6. Funding**

This research was funded in part by the International Society of Biomechanics Matching Dissertation Grant (2021-2022). DK was supported by a scholarship from the Natural Sciences and Engineering Research Council of Canada (CGSD3-502358-2017), the Maury L Hull Endowed Fellowship for Musculoskeletal Biomechanics Research, and a University of California Davis Professors for the Future Fellowship.

## 6S: Supplement to Chapter 6

### 6S.1: Baseline interview

- What gender do you identify as? What is your biological sex?
- How old are you?
- How much do you weigh (in pounds)?
- How many years have you been running?
- Have you ever run competitively? If so, at what level?
- Have you ever run in any races? If so, what distances and what were your best times?
- How many times do you run per week on average?
- How many miles do you run per week on average?
- How much time do you spend running per week on average (give best estimate)?
- On what surfaces do you typically run (e.g., grass, concrete, dirt, etc.)?
- Do you warm-up before you run? If so, please describe what you do?
- What is the make, model, and size of the shoes you are currently wearing?
- How many miles do you have on these shoes?
- How many other pairs of shoes do you rotate through?
- Do you do any other physical activities regularly? If so, what activity, how often, and at what level of skill or competition (e.g., recreationally, intramurals, collegiate, professional)?
- Have you had a running-related injury or an injury to your legs/lower back in the last year? If so, please describe the injury.
  - Did you see a medical professional for this injury?
  - Did you reduce or alter your running because of this injury?



- Does this injury cause you any current pain? If so, please rate the pain from 0 (no pain) to 10 (worst pain imaginable).
- Do you take non-steroidal anti-inflammatory drugs such as aspirin, ibuprofen, indomethacin, etc.? If so, then on average how much do you take (e.g. 200 mg) and how often (e.g. before every run, once a week, less than once a week)?

### **6S.2: Weekly survey**

Each Sunday an email containing a link to an online survey was sent to participants. If the survey was not completed, then automatic reminder emails were sent on Monday and Tuesday. If the survey was still not completed by Wednesday, an experimenter manually contacted the participant. An example of this survey can be completed here:

[https://ucdavis.co1.qualtrics.com/jfe/form/SV\\_39FUucksE40fnHD](https://ucdavis.co1.qualtrics.com/jfe/form/SV_39FUucksE40fnHD) (accessed February 14th 2024).

Alternatively, all the survey questions are listed below. Note, the survey used a branching logic and only presented some sections depending on the answers to previous questions (e.g., if a respondent did not report running on Monday, they would not be asked to report their Monday running distance).



## **HUMAN PERFORMANCE LAB**

### **RUNNING SURVEY**

#### **WEEK 0 of 12 -- DEMO/TEMPLATE**

In the following weekly survey you will be asked to provide information about your running habits [over the last 7 days](#) (Monday through Sunday). Please be as accurate and detailed as possible in your responses - the more accurate and detailed your weekly reports, the better we will be able to identify the causes of running-related injuries.

To ensure your responses are as accurate as possible, we encourage you to record your activity every day on the paper log we provided.

The survey should take approximately 5-10 minutes to complete.

How much do you **WEIGH** this week (in pounds)?

Do you feel **STRESSED** this week?

*Stress means a situation in which you felt tense, restless, nervous, or anxious or were unable to sleep at night because your mind was troubled.*

Not at all      0      1      2      3      4      5      6      7      8      9      10      Very much



Which days did you run this week?

*Select all that apply.*

- Monday
- Tuesday
- Wednesday
- Thursday
- Friday
- Saturday
- Sunday

Please answer the following questions for your runs on **MONDAY...**

*If you ran more than once, report the total distance, time, and effort as if it were a single run.*

---

How would you rate the quality of your **SLEEP** before your run?

*Please think about how many hours of sleep you got, how easily you fell asleep, how often you woke up during the night (except to go to the bathroom), whether you woke up earlier than you had to in the morning, and how refreshing your sleep was. Drag and slide the bar to the appropriate number. You can see the number you have selected on the slider as you move it.*

Terrible 0      1      2      Poor 3      4      Fair 5      6      7      Good 8      9      Excellent 10



Were you completely **RECOVERED** from any previous physical activity?

*Consider any feelings of exhaustion, fatigue, or soreness. Drag and slide the bar to the appropriate number. You can see the number you have selected on the slider as you move it.*

Very poor recovery 0      1      2      3      Reasonable recovery 4      5      6      7      8      Very good recovery 9      10



Did you **WEAR YOUR ACTIVITY MONITOR** on your run?

Yes

No

---

How many **MILES** did you run?

MILES

---

How many **MINUTES** did you run?

MINUTES

---

What was your overall **EFFORT** during the run?

*Considering 0 as rest and 10 as the hardest you've ever exercised, rate the effort of your run. Drag and slide the bar to the appropriate number. You can see the number you have selected on the slider as you move it.*

Rest

No exertion

0

1

2

3

4

5

6

7

8

Extremely hard

Maximal exertion

9

10



In addition to running, did you participate in any **OTHER PHYSICAL ACTIVITY**

*Very briefly describe any physical activity you participated in over the week (e.g., "2 days recreational weightlifting", "1 day swim training for triathlon", or "NCAA Div 1 soccer game")*

Have you experienced any **PAIN OR INJURY** in your **LOWER LIMBS** (pelvis, legs, feet) this week?

Yes

No

Was there a **SINGLE TRAUMATIC EVENT** that caused this injury?

*For example, rolling your ankle, falling, or getting hit*

Did you seek **MEDICAL TREATMENT** from a medical professional for this pain or injury?

*For example, did you visit a doctor, physiotherapist, athletic therapist, chiropractor, etc.*

Please rate the **PAIN** you are experiencing on a scale from 0 (no pain) to 10 (worst pain imaginable).

*Drag and slide the bar to the appropriate number. You can see the number you have selected on the slider as you move it.*

No pain 0 1 2 3 4 5 6 7 8 9 10 Worst pain imaginable

Pain



Has the pain/injury that you are experiencing in this area caused you to **STOP OR ALTER YOUR RUNNING?**

Which days did you **STOP** running this week due to this pain or injury?

*i.e., you WOULD have gone for a run but decided not to due to pain or injury. Select all that apply*

Monday

Tuesday

Wednesday

Thursday

Friday

Saturday

Sunday

If you still ran even though you had pain or injury, which days did you **ALTER** your running due to the pain or injury?

*i.e., you WOULD have run further or faster but decided not to due to pain or injury. Select all that apply*

Monday

Tuesday

Wednesday

Thursday

Friday

Saturday

Sunday

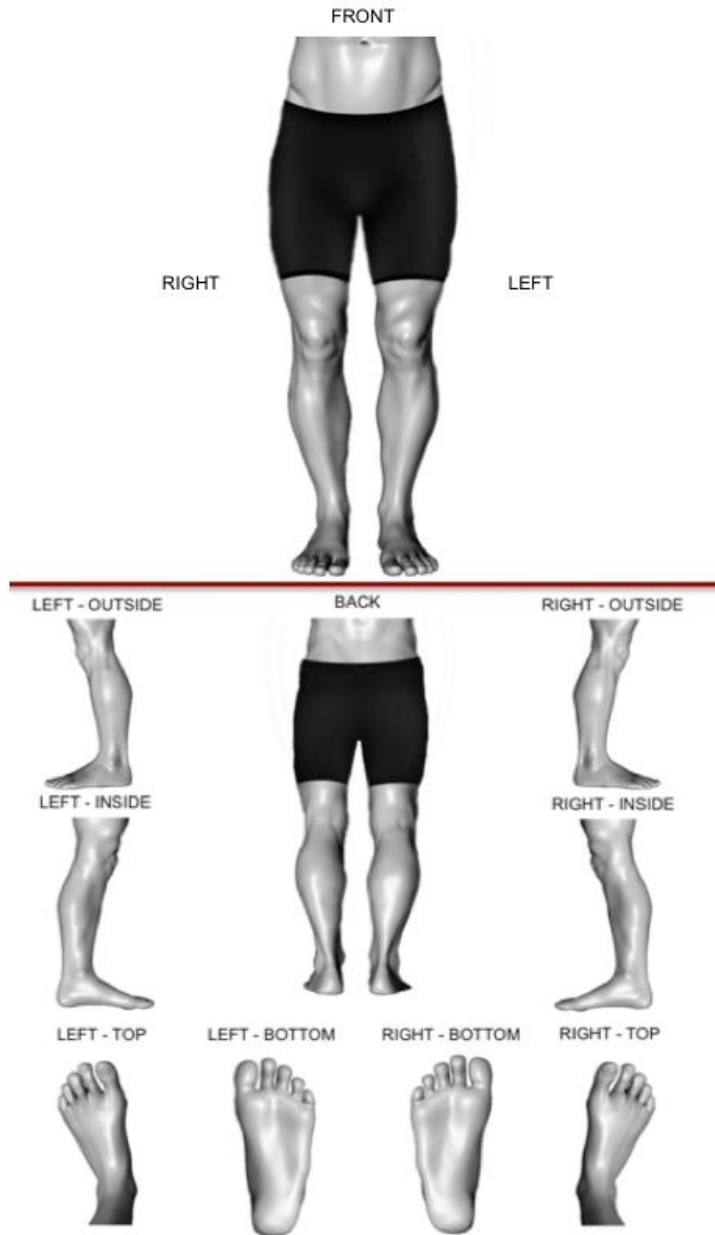


**WHERE** are you experiencing this pain or injury?

Use the following images of a runner and **click where the pain or injury is.**

**You can click up to 10 times** if the pain or injury extends through more than one area.

If you clicked in the wrong location by accident you can click a point to remove it; or, you can click and hold to drag to a new location.



If applicable, select any days that you used a **NON-STEROIDAL ANTI-INFLAMMATORY DRUG (NSAID)**

NSAIDs are widely used and include Aspirin (Bayer, Bufferin, Ecotrin, St. Joseph), Ibuprofen (Advil, Motrin), Naproxen (Aleve, Anaprox DS, Naprosyn), and Celecoxib (Celebrex).

	Used NSAID
Monday	<input type="checkbox"/>
Tuesday	<input type="checkbox"/>
Wednesday	<input type="checkbox"/>
Thursday	<input type="checkbox"/>
Friday	<input type="checkbox"/>
Saturday	<input type="checkbox"/>
Sunday	<input type="checkbox"/>

Did you monitor your **HEART RATE** during the week?

<input type="radio"/> Yes	<input type="radio"/> No
---------------------------	--------------------------

Fill out your **AVERAGE** Heart Rate, your **MAXIMUM** Heart Rate, and your Heart Rate **VARIABILITY** for any days that you monitored them.

*Leave fields blank if you did not monitor a certain parameter or day.*

	Average	Max	Variability
Monday	<input type="text"/>	<input type="text"/>	<input type="text"/>
Tuesday	<input type="text"/>	<input type="text"/>	<input type="text"/>
Wednesday	<input type="text"/>	<input type="text"/>	<input type="text"/>
Thursday	<input type="text"/>	<input type="text"/>	<input type="text"/>
Friday	<input type="text"/>	<input type="text"/>	<input type="text"/>
Saturday	<input type="text"/>	<input type="text"/>	<input type="text"/>
Sunday	<input type="text"/>	<input type="text"/>	<input type="text"/>

---

**WHAT DEVICE** did you use to monitor your Heart Rate?

Are there any additional notes or details you would like to provide?

### 6S.3: Weekly running log

UC Davis Human Performance Lab

#### Weekly Running Log

If needed, you can email any questions to [experimenter@ucdavis.edu](mailto:experimenter@ucdavis.edu)

What was your weight this week? \_\_\_\_\_

Do you feel stressed this week? Rate from 0 (not at all) to 10 (very much): \_\_\_\_\_

Stress means a situation in which you felt tense, restless, nervous, or anxious or were unable to sleep at night because your mind was troubled.

261

	RUN X – no run ✓ – run A – altered due to injury S – skipped due to injury	MONITOR X – forgot/didn't turn on monitor ✓ – wore monitor	SLEEP How well did you sleep last night? 0 (terrible) to 10 (excellent)	RECOVERY Were you completely recovered from any previous physical activity? 0 (very poor recovery) to 10 (very good recovery)	DIST Miles	TIME Minutes	EFFORT What was your overall effort? 0 (rest/no exertion) to 10 (max exertion)	NSAID	HEART RATE		
									Avg	Max	Var
Mon											
Tue											
Wed											
Thu											
Fri											
Sat											
Sun											

Did you experience any running-related pain or injury this week? \_\_\_\_\_

If so, was it caused by a single traumatic event? \_\_\_\_\_

Did you seek medical treatment for running-related pain or injury? \_\_\_\_\_

Rate the severity of the pain from 0 (no pain) to 10 (worst pain imaginable) \_\_\_\_\_

## 7. Conclusions

Running injuries affect millions of Americans each year. These injuries impose serious and costly health and economic burdens. Thus, reducing running injury has been a target of decades of biomechanics research. Despite these research efforts, the incidence of running injury has not appreciably decreased and easy-to-measure biomechanical metrics of running injury risk remain elusive. A major contributor to this lack of success is likely the disconnect between theories of running injury causation—in which injury is caused by the accumulation of microtrauma across many repeated loading events—and the way running injury has been studied—in lab environments where biomechanics have been represented as a single magnitude calculated from the average of a handful of gait cycles.

In Chapter 2 of this dissertation, I introduced a novel method that overcame these lab-based limitations. This first-of-its kind study used small wearable devices to estimate the magnitude and number of external loads (operationalized as maximum ground reaction forces and number of strides) across two months of training for 9 Division I running athletes. Results demonstrated that runners who developed injuries had higher external loads and cumulative loads (loads weighted by an exponent to account for musculoskeletal structure properties then summed across strides) than runners that did not develop an injury. These promising results underscored the potential of this method to move the field of biomechanics out of the lab and into the real world where ecologically valid measurements can be taken of the number and magnitude of loads that runners actually experience—and that potentially lead to injury.

Although exciting, this methodology was on the cutting edge of a new and rapidly developing area of biomechanics research: The use of wearable technology in biomechanics is exploding but no clear guidelines have yet emerged to guide the field, and few standards or publicly available pipelines

exist (see Chapter 5.1 for a more detailed discussion). As such, a number of limitations existed in this novel method. Critically, only unilateral maximum ground reaction forces could be estimated, no validation of force or gait event estimation had been conducted across different runners or running conditions, and no information existed on how sensor misplacement could impact data—even though this type of misplacement is likely in wearables-based prospective research. To systematically address these limitations, I present a series of studies in Chapters 3, 4, and 5 where typical lab-based measurements and measurements from a wearable inertial measurement unit were taken from 74 participants as they ran across a range of different conditions. These data were used to compare 18 different methods of estimating gait events, 27 different methods of estimating vertical ground reaction forces, and how these and other metrics were affected by a 5 cm change in sensor placement. To help guide the field, code to execute each of these methods was made available online and recommendations were made to guide future use.

Finally, applying these validated methods in Chapter 6, this dissertation replicated the initial proof-of-concept study in a large heterogenous sample across a longer period of time. Currently the methods of this study are limited due to the computational demands of executing several processing steps and correcting for missingness; thus, the present results should not be overinterpreted. The general approach illustrated by this study collectively with the earlier chapters of this dissertation does, however, demonstrate a novel way to measure running and injury, provide the tools to execute that approach, and show mixed support for the hypothesis that the magnitude and number of loads experienced by runners relates to RRI development.

## 8. References

- [1] D. B. Burr, A. G. Robling and C. H. Turner, "Effects of biomechanical stress on bones in animals," *Bone*, vol. 30, no. 5, pp. 781-786, 2002.
- [2] J. H. Shepherd and H. R. C. Screen, "Fatigue loading of tendon," *International Journal of Experimental Pathology*, vol. 94, no. 4, pp. 260-270, 2013.
- [3] C. Rolf, "Overuse injuries of the lower extremity in runners," *Scandinavian Journal of Medicine and Science in Sports*, vol. 5, no. 4, pp. 181-190, 1995.
- [4] H. M. Frost, "A brief review for orthopedic surgeons: fatigue damage (microdamage) in bone (its determinants and clinical implications)," *Journal of Orthopaedic Science*, vol. 3, no. 5, pp. 272-281, 1998.
- [5] B. Weightman, D. J. Chappell and E. A. Jenkins, "A second study of the tensile fatigue properties of human articular cartilage," *Annals of the Rheumatic Diseases*, vol. 37, no. 1, pp. 58-63, 1978.
- [6] T. A. Wren, D. P. Lindsey, G. S. Beaupre and D. R. Carter, "Effects of creep and cyclic loading on the mechanical properties and failure of human Achilles tendons," *Annals of Biomedical Engineering*, vol. 31, no. 6, pp. 710-717, 2003.
- [7] D. R. Carter and W. E. Caler, "A cumulative damage model for bone fracture," *Journal of Orthopaedic Research*, vol. 3, no. 1, pp. 84-90, 1985.
- [8] J. Yang, A. Tibbetts, T. Covassin, G. Cheng, S. Nayar and E. Heiden, "Epidemiology of Overuse and Acute Injuries Among Competitive Collegiate Athletes," *Journal of Athletic Training*, vol. 47, no. 2, pp. 198-204, 2012.
- [9] The Outdoor Foundation, "2016 Outdoor Participation Report," Washington, DC, 2016.
- [10] Sports and Fitness Industry Association, "2017 Sports, Fitness, and Leisure Activities Topline Participation Report: Definitive Source for Sports Participation in America," Silver Spring, MD, 2017.
- [11] National Sporting Goods Association, "Sports Participation: 2016 Edition," Mount Prospect, IL, 2016.
- [12] Nielsen Scarborough, "Number of people who went jogging or running in the last 12 months in the United States from spring 2008 to spring 2017 (in millions)," Statista - The Statistics Portal, 2018. [Online]. Available: <https://www.statista.com/statistics/227423/number-of-joggers-and-runners-usa/>. [Accessed 27 March 2018].
- [13] National Federation of State High School Associations, "2016-2017 High School athletics participation survey," Indianapolis, IN, 2017.
- [14] US Census Bureau, "Statistical abstract of the United States: 2012, Section 26. Arts, Recreation, and Travel," 2012.
- [15] Running USA, "2017 National Runner Survey," Troy, MI, 2017.
- [16] Running USA, "Marathon Report," 15 June 2017. [Online]. Available: <http://www.runningusa.org/marathon-report-2017>. [Accessed 12 April 2018].
- [17] Road Runners Club of America, "History of Road Runners Club of America," [Online]. Available: [https://www.rrca.org/docs/default-source/default-document-library/rrca\\_detailed\\_history\\_from\\_handbook.pdf?sfvrsn=cc91c30d\\_2](https://www.rrca.org/docs/default-source/default-document-library/rrca_detailed_history_from_handbook.pdf?sfvrsn=cc91c30d_2). [Accessed 12 April 2018].

- [18] D. E. R. Warburton, P. T. Katzmarzyk, R. E. Rhodes and R. J. Shephard, "Evidence-informed physical activity guidelines for Canadian adults," *Applied Physiology, Nutrition, and Metabolism*, vol. 32, no. Suppl. 2E, pp. S16-S68, 2007.
- [19] World Health Organization, "Global Health Risks: Mortality and the burden of disease attributable to selected major risks," WHO Press, Geneva, Switzerland, 2009.
- [20] US Department of Health and Human Services, Center for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, "Physical activity and health: A report of the Surgeon General," US Department of Health and Human Services, Atlanta, GA, 1996.
- [21] S. G. Wannamethee and A. G. Shaper, "Physical activity in the prevention of cardiovascular disease: An epidemiological perspective," *Sports Medicine*, vol. 31, pp. 101-114, 2001.
- [22] C. M. Friedenreich, "Physical activity and cancer prevention," *Cancer Epidemiology, Biomarkers and Prevention*, vol. 10, pp. 287-301, 2001.
- [23] M. E. Martinez, G. E. D. Spiegelman, D. J. Hunter, W. C. Willet and G. A. Colditz, "Leisure-time physical activity, body size, and colon cancer in women. Nurses' Health Study Research Group," *Journal of the National Cancer Institute*, vol. 89, no. 13, pp. 948-955, 1997.
- [24] W. M. Haskell, I. Lee, R. R. Pate, K. E. Powell, S. N. Blair, B. A. Franklin, C. A. Macera, G. W. Heath, P. D. Thompson and A. Bauman, "Physical activity and public health: Updated recommendations for adults from the American College of Sports Medicine and the American Heart Association," *Medicine and Science in Sports and Exercise*, vol. 39, no. 8, pp. 1423-1434, 2007.
- [25] S. Vidbaek, A. M. Bueno, R. O. Nielsen and S. Rasmussen, "Incidence of running-related injuries per 1000 h of running in different types of runners: A systematic review and meta-analysis," *Sports Medicine*, vol. 45, pp. 1017-1026, 2015.
- [26] R. Bahr, "No injuries but plenty of pain? On the methodology for recording overuse symptoms in sports," *British Journal of Sports Medicine*, vol. 43, no. 13, pp. e966-972, 2009.
- [27] B. Clarsen, G. Myklebust and R. Bahr, "Development and validation of a new method for the registration of overuse injury epidemiology: the Oslo Sports Trauma Research Centre (OSTRC) Overuse Injury Questionnaire," *British Journal of Sports Medicine*, vol. 47, no. 8, pp. 495-502, 2013.
- [28] R. O. Nielsen, L. Ronnow, S. Rasmussen and M. Lind, "A prospective study on time to recovery in 254 injured novice runners," *PLoS One*, p. e99877, 2014.
- [29] B. Marti, J. P. Vader, C. E. Minder and T. Abelin, "On the epidemiology of running injuries. The 1984 Bern Grand-Prix study," *American Journal of Sports Medicine*, vol. 16, no. 3, pp. 285-294, 1988.
- [30] J. P. Koplan, R. B. Rothenberg and E. L. Jones, "The natural history of exercise: a 10-yr follow-up of a cohort of runners," *Medicine and Science in Sports and Exercise*, vol. 27, no. 8, pp. 1180-1184, 1995.
- [31] L. C. Hespanhol Junior, W. van Mechelen, E. Postuma and E. Verhagen, "Health and economic burden of running-related injuries in runnerstraining for an event: A prospective cohort study," *Scandinavian Journal of Medicine and Science in Sports*, vol. 26, no. 9, pp. 1091-1099, 2016.
- [32] D. Smits, B. Huisstede, E. Verhagen, H. van der Worp, B. Kluitenberg, M. van Middelkoop, F. Hartgens and F. Backx, "Short-term absenteeis and health care utilization due to lower extremity injuries among novice runners: A prospective cohort study," *Clinical Journal of Sports Medicine*, vol. 26, no. 6, pp. 502-509, 2016.
- [33] W. van Mechelen, "Running injuries: A review of the epidemiological literature," *Sports Medicine*, vol. 14, no. 5, pp. 320-335, 1992.



- [34] S. D. Walter, L. E. Hart, J. M. McIntosh and J. R. Sutton, "The Ontario cohort study of running-related injuries," *Archives of Internal Medicine*, vol. 149, pp. 2561-2564, 1989.
- [35] T. C. Valovich McLeod, L. C. Decoster, K. J. Loud, L. J. Micheli, J. T. Parker, M. A. Sandrey and C. White, "National athletic trainers' association position statement: Prevention of pediatric overuse injuries," *Journal of Athletic Training*, vol. 46, no. 2, pp. 206-220, 2011.
- [36] R. Fokkema, R. de Vos, J. van Ochten, J. Verhaar, I. Davis, P. Bindels, S. Bierma-Zeinstra and M. van Middelkoop, "Preventing running-related injuries using evidence-based online advice: the design of a randomised-controlled trial," *BMJ Open Sport & Exercise Medicine*, vol. 3, no. 1, p. e000265, 2017.
- [37] J. Hoerberigs, "Factors related to the incidence of running injuries: A review," *Sports Medicine*, vol. 13, no. 6, pp. 408-422, 1992.
- [38] B. Nigg and M. Bobbert, "On the potential of various approaches in load analysis to reduce the frequency of sports injuries," *Journal of Biomechanics*, vol. 23, no. Suppl. 1, pp. 3-12, 1990.
- [39] A. D. Lopes, L. C. Hespanhol Junior, S. S. Yeung and L. O. Pena Costa, "What are the main running-related musculoskeletal injuries? A systematic review," *Sports Medicine*, vol. 42, no. 10, pp. 891-905, 2012.
- [40] A. Hulme, R. Nielsen, T. T. E. Verhagen and C. Finch, "Risk and protective factors for middle- and long-distance running related injury: A systematic review," *Sports Medicine*, vol. 47, no. 5, pp. 869-886, 2017.
- [41] H. van der Worp, J. W. Vrielink and S. W. Bredeweg, "Do runners who suffer injuries have higher vertical ground reaction forces than those who remain injury free? A systematic review and meta-analysis," *British Journal of Sports Medicine*, vol. 50, no. 8, pp. 450-457, 2016.
- [42] M. van der Worp, D. ten Haat, R. van Cingel, A. de Wijer, M. Nijhuis-van der Sanden and J. Staal, "Injuries in runners; a systematic review on risk factors and sex differences," *PLoS ONE*, vol. 10, no. 2, p. e0114937, 2015.
- [43] S. Messier, D. Martin, S. Mihalko, E. Ip, P. DeVita, W. Cannon, M. Love, D. Beringer, S. Saldana, R. Fellin and J. Seay, "A 2-year prospective cohort study of overuse running injuries: The runners and injury longitudinal study (TRAILS)," *The American Journal of Sports Medicine*, vol. 46, no. 9, pp. 2211-2221, 2018.
- [44] S. Videbaek, A. Bueno, R. Nielsen and S. Rasmussen, "Incidence of Running-Related Injuries Per 1000 h of running in Different Types of Runners: A Systematic Review and Meta-Analysis," *Sports Medicine*, vol. 45, no. 7, pp. 1017-1026, 2015.
- [45] D. Burr, R. Milgrom, W. Boyd, G. Higgins, E. Robin and J. Radin, "Experimental stress fractures of the tibia biological and mechanical aetiology in rabbits," *Journal of Bone and Joint Surgery. British Volume*, vol. 72, pp. 370-375, 1990.
- [46] D. Burr, C. Turner, P. Naick, M. Forwood, W. Ambrosius, M. Hasan and R. Pidaparti, "Does microdamage accumulation affect the mechanical properties of bone?," *Journal of Biomechanics*, vol. 31, no. 4, pp. 337-345, 1998.
- [47] T. Keller and D. Spengler, "Regulation of bone stress and strain in the immature and mature rat femur," *Journal of Biomechanics*, vol. 22, no. 11-12, pp. 1115-1127, 1989.
- [48] M. Schaffler, E. Radin and D. Burr, "Long-term fatigue behavior of compact bone at low strain magnitude and rate," *Bone*, vol. 11, no. 5, pp. 321-326, 1990.
- [49] D. Nunamaker, D. Butterweck and M. Provost, "Fatigue fractures in thoroughbred racehorses: relationships with age, peak bone strain, and training," *Journal of Orthopaedic Research*, vol. 8, no. 4, pp. 604-611, 1990.
- [50] S. Mori and D. Burr, "Increased intracortical remodeling following fatigue damage," *Bone*, vol. 14, no. 2, pp. 103-109, 1993.

- [51] T. Wren, G. Beaupre and D. Carter, "A model for loading-dependent growth, development, and adaptation of tendons and ligaments," *Journal of Biomechanics*, vol. 31, no. 2, pp. 107-114, 1998.
- [52] W. B. Edwards, D. Taylot, T. J. Rudolphi, J. C. Gillette and T. R. Derrick, "Effects of stride length and running mileage on a probabilistic stress fracture model," *Medicine and Science in Sports and Exercise*, vol. 41, no. 12, pp. 2177-2184, 2009.
- [53] W. B. Edwards, D. Taylor, T. J. Rudolphi, J. C. Gillette and T. R. Derrick, "Effects of running speed on a probabilistic stress fracture model," *Clinical Biomechanics*, vol. 25, no. 4, pp. 372-377, 2010.
- [54] T. Wren, G. Beaupre and D. Carter, "Tendon and ligament adaptation to exercise, immobilization, and remobilization," *Journal of Rehabilitation Research & Development*, vol. 37, no. 2, pp. 217-224, 2000.
- [55] C. Pattin, W. Caler and D. Carter, "Cyclic mechanical property degradation during fatigue loading of cortical bone," *Journal of Biomechanics*, vol. 29, no. 1, pp. 69-79, 1996.
- [56] K. Choi and S. Goldstein, "A comparison of the fatigue behavior of human trabecular and cortical bone tissue," *Journal of Biomechanics*, vol. 25, no. 12, pp. 1371-1381, 1992.
- [57] D. Carter and W. Caler, "Cycle-dependent and time-dependent bone fracture with repeated loading," *Journal of Biomechanical Engineering*, vol. 105, no. 2, pp. 166-170, 1983.
- [58] D. Carter, W. Caler, D. Spengler and V. Frankel, "Uniaxial fatigue of human cortical bone. The influence of tissue physical characteristics," *Journal of Biomechanics*, vol. 14, no. 7, pp. 461-470, 1981.
- [59] H. Schechtman and D. Bader, "Fatigue damage of human tendons," *Journal of Biomechanics*, vol. 35, no. 3, pp. 347-353, 2002.
- [60] H. Schechtman and D. Bader, "In vitro fatigue of human tendons," *Journal of Biomechanics*, vol. 30, no. 8, pp. 829-835, 1997.
- [61] M. L. Bertelsen, A. Hulme, P. J. R. K. Brund, H. Sorenson, C. F. Finch, E. T. Parner and R. O. Nielsen, "A framework for the etiology of running-related injuries," *Scandinavian Journal of Medicine and Science in Sports*, vol. 27, no. 11, pp. 1170-1180, 2017.
- [62] C. Maganaris and J. Paul, "In vivo human tendon mechanical properties," *Journal of Physiology*, vol. 521, no. 1, pp. 307-313, 1999.
- [63] T. Muramatsu, T. Muraoka, D. Takeshita, Y. Kawakami and Y. F. T. Hirano, "Mechanical properties of tendon and aponeurosis of human gastrocnemius muscle in vivo," *Journal of Applied Physiology*, vol. 90, no. 5, pp. 1671-1678, 1985.
- [64] L. Lanyon, W. Hampson, A. Goodship and J. Shah, "Bone deformation recorded in vivo from strain gauges attached to the human tibial shaft," *Acta Orthopaedica*, vol. 46, no. 2, pp. 256-268, 1975.
- [65] D. Burr, C. Milgrom, D. Fyhrie, M. Forwood, M. Nyska, A. Finestone, S. Hoshaw, E. Saiag and A. Simkin, "In vivo measurement of human tibial strains during vigorous activity," *Bone*, vol. 18, no. 5, pp. 405-410, 1996.
- [66] D. Fyhrie, C. Milgrom, S. Hoshaw, A. Simkin, S. Dar, D. Drumb and D. Burr, "Effect of fatiguing exercise on longitudinal bone strain as related to stress fracture in humans," *Annals of Biomedical Engineering*, vol. 26, no. 4, pp. 660-665, 1998.
- [67] B. Beynnon, J. Howe, M. Pope, R. Johnson and B. Fleming, "The measurement of anterior cruciate ligament strain in vivo," *International Orthopaedics*, vol. 16, no. 1, pp. 1-12, 1992.
- [68] A. A. Zadpoor and A. A. Nikooyan, "The relationship between lower extremity stress fractures and the ground reaction force: a systematic review," *Clinical Biomechanics*, vol. 26, no. 1, pp. 23-28, 2011.

- [69] R. H. Miller and J. Hamill, "Computer simulation of the effects of shoe cushioning on internal and external loading during running impacts," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 12, no. 4, pp. 481-490, 2009.
- [70] E. Matejevich, L. Branscombe, L. Scott and K. Zelik, "Ground reaction force metrics are not strongly correlated with tibial bone load when running across speeds and slopes: Implications for science, sport and wearable tech," *PLoS One*, vol. 14, no. 1, p. e0210000, 2019.
- [71] R. H. Miller, W. B. Edwards, S. C. E. Brandon, A. M. Morton and K. J. Deluzio, "Why don't most runners get knee osteoarthritis? A case for per-unit-distance loads," *Medicine and Science in Sports and Exercise*, vol. 46, no. 3, pp. 572-579, 2013.
- [72] C. R. Firminger and W. B. Edwards, "The influence of minimalist footwear and stride length reduction on lower-extremity running mechanics and cumulative loading," *Journal of Science and Medicine in Sport*, vol. 19, no. 12, pp. 975-979, 2016.
- [73] T. P. Andriacchi and A. E. J., "Studies of human locomotion: Past, present, and future," *Journal of Biomechanics*, vol. 33, pp. 1217-1224, 2000.
- [74] R. H. Miller, J. L. Lowry, S. Meardon and J. C. Gillette, "Lower extremity mechanics of iliotibial band syndrome during an exhaustive run," *Gait and Posture*, vol. 26, no. 3, pp. 407-413, 2007.
- [75] S. Meardon, J. Hamill and T. R. Derrick, "Running injury and stride time variability over a prolonged run," *Gait and Posture*, vol. 33, no. 1, pp. 36-40, 2011.
- [76] R. O. Nielsen, I. Buist, H. Sorenson, M. Lind and S. Rasmussen, "Training errors and running related injuries: a systematic review," *The International Journal of Sports Physical Therapy*, vol. 7, no. 1, pp. 58-75, 2012.
- [77] A. Hreljac, "Impact and overuse injuries in runners," *Medicine and Science in Sports and Exercise*, vol. 36, no. 5, pp. 845-849, 2004.
- [78] S. L. James, B. T. Bates and C. R. Osternig, "Injuries to runners," *American Journal of Sports Medicine*, vol. 6, pp. 40-50, 1978.
- [79] Burr, "Bone exercise and stress fractures," *Exercise and Sport Science Reviews*, vol. 25, no. 1, pp. 171-194, 1997.
- [80] A. Hreljac, R. N. Marshall and P. A. Hume, "Evaluation of lower extremity overuse injury potential in runners," *Medicine and Science in Sports and Exercise*, vol. 32, no. 9, pp. 1635-1641, 2000.
- [81] L. Malisoux, R. O. Nielsen, A. Urhausen and D. Theisen, "A step towards understanding the mechanisms of running-related injuries," *Journal of Science and Medicine in Sports*, vol. 18, no. 5, pp. 523-528, 2015.
- [82] R. H. Miller, W. B. Edwards and K. J. Deluzio, "Energy expended and knee joint load accumulated when walking, running, or standing for the same amount of time," *Gait and Posture*, vol. 41, no. 1, pp. 9-11, 2014.
- [83] J. Petersen, H. Sorenson and R. O. Nielsen, "Cumulative loads increase at the knee joint with slow-speed running compared to faster running: a biomechanical study," *Journal of Orthopaedic and Sports Physical Therapy*, vol. 45, no. 4, pp. 316-322, 2015.
- [84] M. Baggaley and W. B. Edwards, "Effect of running speed on Achilles tendon injury potential: Use of a weighted impulse measure," *Medicine and Science in Sports and Exercise*, vol. 49, no. 5S, p. 139, 2017.
- [85] R. H. Miller, "Joint loading in runners does not initiate knee osteoarthritis," *Exercise and Sport Science Reviews*, vol. 45, no. 2, pp. 87-95, 2017.
- [86] J. M. Neugebauer, D. A. Hawkins and B. L., "Estimating youth locomotion ground reaction forces using an accelerometer-based activity monitor," *PLoS One*, vol. 7, no. 10, p. e48182, 2012.
- [87] J. M. Neugebauer, K. H. Collins and D. A. Hawkins, "Ground reaction force estimates from ActiGraph GT3X+ hip accelerations," *PLoS One*, vol. 9, no. 6, p. e99023, 2014.

- [88] J. B. Lee, R. B. Mellifont and B. J. Burkett, "The use of a single inertial sensor to identify stride, step, and stance durations of running gait," *Journal of Science and Medicine in Sport*, vol. 13, no. 2, pp. 270-273, 2010.
- [89] R. N. van Gent, D. Siem, M. van Middelkoop, A. G. van Os, S. M. A. Bierma-Zeinstra and B. W. Koes, "Incidence and determinants of lower extremity running injuries in long distance runners: a systematic review," *British Journal of Sports Medicine*, vol. 41, pp. 469-480, 2007.
- [90] R. Johnson, "Common running injuries of the leg and foot," *Minnesota Medicine*, vol. 66, no. 7, pp. 441-444, 1983.
- [91] J. Lysholm and J. Wyklander, "Injuries in runners," *American Journal of Sports Medicine*, vol. 15, no. 2, pp. 168-171, 1987.
- [92] A. F. Renstrom, "Mechanism, diagnosis, and treatment of running injuries," *Instructional Course Lectures*, vol. 42, pp. 225-234, 1993.
- [93] P. G. Weyand, M. Kelly, T. Blackadar, J. Darley, S. R. Oliver, N. E. Ohlenbusch, S. W. Joffe and R. W. Hoyt, "Ambulatory estimates of maximal aerobic power from foot-ground contact times and heart rates in running humans," *Journal of Applied Physiology*, vol. 91, no. 1, pp. 451-458, 2001.
- [94] B. Auvinet, E. Gloria, G. Renault and E. Barrey, "Runner's stride analysis: Comparison of kinematic and kinetic analyses under field conditions," *Science and Sports*, vol. 17, no. 2, pp. 92-94, 2002.
- [95] A. Wixted, D. Billing and D. James, "Validation of trunk mounted inertial sensors for analysing running biomechanics under field conditions, using synchronously collected foot contact data," *Sports Engineering*, vol. 12, pp. 207-212, 2010.
- [96] E. Bergamini, P. Picerno, H. Pillet, F. Natta, P. Thoreux and V. Camomilla, "Estimation of tempoal parameters during sprint running using a trunk-mounted inertial measurement unit," *Journal of Biomechanics*, vol. 45, no. 6, pp. 1123-1126, 2012.
- [97] M. Buccheit, A. Gray and J. Morin, "Assessing stride variables and vertical stiffness with GPS embedded accelerometers: Preliminary insights for the monitoring of neuromuscular fatigue on the field," *Journal of Sports Science and Medicine*, vol. 14, pp. 698-701, 2015.
- [98] D. Kiernan, R. H. Miller, B. S. Baum, H. J. Kwon and J. K. Shim, "Amputee locomotion: Frequency content of prosthetic vs. intact limb vertical ground reaction forces during running and the effects of filter cut-off frequency," *Journal of Biomechanics*, vol. 60, pp. 248-252, 2017.
- [99] C. L. Coolbaugh and D. A. Hawkins, "Standardizing accelerometer-based activity monitor calibration and output reporting," *Journal of Applied Biomechanics*, vol. 30, no. 4, pp. 594-597, 2014.
- [100] P. R. Cavanagh and M. A. Lafortune, "Ground reaction forces in distance running," *Journal of Biomechanics*, vol. 13, pp. 397-406, 1980.
- [101] C. F. Munro, D. I. Miller and A. J. Fuglevand, "Ground reaction forces in running: A reexamination," *Journal of Biomechanics*, vol. 20, no. 2, pp. 147-155, 1987.
- [102] P. R. Cavanagh and R. Kram, "Stride length in distance running: velocity, body dimensions, and added mass effects," *Medicine and Science in Sports and Exercise*, vol. 21, no. 4, pp. 467-479, 1989.
- [103] K. R. Williams, R. Snow and C. Arguss, "Changes in distance running kinematics with fatigue," *International Journal of Sport Biomechanics*, vol. 7, pp. 138-162, 1991.
- [104] B. De Wit, D. De Clercq and P. Aerts, "Biomechanical analysis of the stance phase during barefoot and shod running," *Journal of Biomechanics*, vol. 33, no. 3, pp. 269-278, 2000.
- [105] P. G. Weyand, D. B. Sternlight, M. Bellizzi and S. Wright, "Faster top running speeds are achieved with greater ground forces not more rapid leg movements," *Journal of Applied Physiology*, vol. 89, no. 5, pp. 1991-1999, 2000.

- [106] A. Leskinen, K. Hakkinen, M. Virravirta, J. Isolehto and H. Kyrolainen, "Comparison of running kinematics between elite and national-standard 1500-m runners," *Sports Biomechanics*, vol. 8, no. 1, pp. 1-9, 2009.
- [107] P. G. Weyand, R. F. Sandell, D. N. L. Prime and M. W. Bundle, "The biological limits to running speed are imposed from the ground up," *Journal of Applied Physiology*, vol. 108, no. 4, pp. 950-961, 2010.
- [108] D. J. Catellier, P. J. Hannan, D. M. Murray, C. L. Addy, T. L. Conway, S. Yang and J. C. Rice, "Imputation of missing data when measuring physical activity by accelerometry," *Medicine and Science in Sports and Exercise*, vol. 37, no. 11 Suppl, pp. S555-S562, 2005.
- [109] D. S. Ward, K. R. Evenson, A. Vaughn, A. B. Rodgers and R. P. Troiano, "Accelerometer use in physical activity: Best practices and research recommendations," *Medicine and Science in Sports and Exercise*, vol. 37, no. 11 Suppl, pp. S582-S588, 2005.
- [110] D. B. Rubin, *Multiple imputation for nonresponse in surveys*, Hoboken, NJ: John Wiley & Sons, 1987.
- [111] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no. 1, pp. 289-300, 1995.
- [112] H. G. Davis, *Conservative Surgery*, New York, NY: D. Appleton & Co., 1867.
- [113] J. Wolff, *The law of bone remodelling*. Translation of *Das Grestz der Transformation der Knochen*, Berlin, Germany: Springer-Verlag, 1892.
- [114] I. Buist, S. W. Bredeweg, K. A. P. M. Lemmink, W. van Mechelen and R. Diercks, "Predictors of running-related injuries in novice runners enrolled in a systematic training program," *The American Journal of Sports Medicine*, vol. 38, no. 2, pp. 273-280, 2010.
- [115] B. Clarsen and R. Bahr, "Matching the choice of injury/illness definition to study setting, purpose and design: one size does not fit all!," *British Journal of Sports Medicine*, vol. 48, no. 7, pp. 510-512, 2014.
- [116] L. C. Hespanhol Junior, L. O. Pena Costa and A. D. Lopes, "Previous injuries and some training characteristics predict running-related injuries in recreational runners: a prospective cohort study," *Journal of Physiotherapy*, vol. 59, no. 4, pp. 263-269, 2013.
- [117] D. Kiernan, J. K. Shim and R. H. Miller, "Vertical ground reaction force magnitudes and rates not positively correlated with prospective running injury," in *Proceedings of the 39th American Society of Biomechanics*, Columbus, OH, 2015.
- [118] D. Kiernan, R. Krupenevich, J. K. Shim and R. H. Miller, "Baseline correlates of running injury: Hip hypermobility but not lower limb strength relates to future running injury," *Medicine and Science in Sports and Exercise*, vol. 48, no. 5, p. 170, 2016.
- [119] C. W. Fuller, J. Ekstrand, A. Junge, T. E. Andersen, R. Bahr, J. Dvorak, M. Hagglund, P. McCorry and W. H. Meeuwisse, "Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries," *Scandinavian Journal of Medicine and Science in Sports*, vol. 16, no. 2, pp. 83-92, 2006.
- [120] C. Coolbaugh, S. Raymond and D. Hawkins, "Feasibility of a dynamic web guidance approach for personalized physical activity prescription based on daily information from wearable technology," *Journal of Medical Internet Research: Research Protocols*, vol. 4, no. 2, p. e67, 2015.
- [121] S. M. Cain, R. S. McGinnis, S. P. Davidson, R. V. Vitalia, N. C. Perkins and S. McLean, "Quantifying performance and effects of load carriage during a challenging balancing task using an array of wireless inertial sensors," *Gait & Posture*, vol. 43, pp. 65-69, 2016.
- [122] R. Willy, L. Buchenic, K. Rogacki, J. Ackerman, A. Schmidt and J. Wilson, "In-field gait retraining and mobile monitoring to address running biomechanics associated with tibial stress fracture," *Scandinavian Journal of Medicine & Science in Sports*, vol. 26, pp. 197-205, 2016.

- [123] A. Gruber, M. Paquette, K. Camelio, D. Powell and J. Vollmar, "Multiscale entropy of center of mass acceleration as a measure for cumulative running fatigue," in *Proceedings of the 41st American Society of Biomechanics*, Boulder, CO, 2017.
- [124] M. Ruder, S. Jamison, A. Tenforde, M. Hannan and I. Davis, "Relationship of footstrike pattern and landing impacts during a marathon race," *Medicine and Science in Sports and Exercise*, vol. 49, no. 5, p. 136, 2017.
- [125] P. Blanch and T. Gabbett, "Has the athlete trained enough to return to play safely? The acute:chronic workload ratio permits clinicians to quantify a player's risk of subsequent injury," *British Journal of Sports Medicine*, vol. 50, no. 8, pp. 471-475, 2016.
- [126] L. Lolli, A. Batterham, R. Hawkins, D. Kelly, A. Strudwick, R. Thorpe, W. Gregson and G. Atkinson, "Mathematical coupling causes spurious correlation within the conventional acute-to-chronic workload ratio calculations.," *British Journal of Sports Medicine*, vol. 53, no. 15, pp. 921-922, 2017.
- [127] S. Grimston, J. Engsberg, R. Kloiber and D. Hanley, "Bone mass, external loads, and stress fracture in female runners," *International Journal of Sport Biomechanics*, vol. 7, no. 3, pp. 293-302, 1991.
- [128] S. Messier, S. Davis, W. Curl, R. Lowery and R. Pack, "Etiologic factors associated with patellofemoral pain in runners," *Medicine and Science in Sports and Exercise*, vol. 23, no. 9, pp. 1008-1015, 1991.
- [129] R. Ferber, I. MacClay-Davis, J. Hamill, C. Pollard and K. McKeown, "Kinetic variables in subjects with previous lower extremity stress fractures," *Medicine and Science in Sports and Exercise*, vol. 34, p. S5, 2002.
- [130] R. A. Zifchok, I. Davis and J. Hamill, "Kinetic asymmetry in female runners with and without retrospective tibial stress fractures," *Journal of Biomechanics*, vol. 39, no. 15, pp. 2792-2797, 2006.
- [131] J. Hay and J. Reid, *Anatomy, mechanics, and human motion*, Englewood Cliffs: Prentice-Hall, 1993.
- [132] J. Mercer, J. Dolgan, J. Griffin and A. Bestwick, "The physiological importance of preferred stride frequency during running at different speeds," *Journal of Exercise Physiology*, vol. 11, no. 3, pp. 26-31, 2008.
- [133] P. Weyand, D. Sternlight, M. Bellizzi and S. Wright, "Faster top running speeds are achieved with greater ground forces not more rapid leg movements," *Journal of Applied Physiology*, vol. 89, pp. 1991-1999, 2000.
- [134] J. Hunter, R. Marshall and P. McNair, "Interaction of step length and step rate during sprint running," *Medicine and Science in Sports and Exercise*, vol. 36, no. 2, pp. 261-271, 2004.
- [135] L. Smith, S. Preece, D. Mason and C. Bramah, "A comparison of kinematic algorithms to estimate gait events during overground running," *Gait & Posture*, vol. 41, no. 1, pp. 39-43, 2015.
- [136] R. Fellin, W. Rose, T. Royer and I. Davis, "Comparison of methods for kinematic identification of footstrike and toe-off during overground and treadmill running," *Journal of Science and Medicine in Sport*, vol. 13, pp. 646-650, 2010.
- [137] A. Hreljac and N. Stergiou, "Phase determination during normal running using kinematic data," *Med Biol Eng Comput*, vol. 38, no. 5, pp. 503-506, 2000.
- [138] J. Sinclair, C. Edmundson, D. Brooks and S. Hobbs, "Evaluation of kinematic methods of identifying gait events during running," *Int J of Sports Sci and Eng*, vol. 5, no. 3, pp. 188-192, 2011.
- [139] B. Nigg, R. De Boer and V. Fisher, "A kinematic comparison of overground and treadmill running," *Medicine and Science in Sports and Exercise*, vol. 27, no. 1, pp. 98-105, 1995.

- [140] P. Riley, J. Dicharry, J. Franz, U. Croce, R. Wilder and D. Kerrigan, "A kinematics and kinetic comparison of overground and treadmill running," *Medicine and Science in Sports and Exercise*, vol. 40, no. 6, pp. 1093-1100, 2008.
- [141] I. Hillel, E. Gazit, A. Nieuwboer, L. Avanzino, L. Rochester, A. Cereatti, U. Della Croce, M. Rikkert, B. Bloem, E. Pelosin, S. Del Din, P. Ginis, N. Giladi, A. Mielman and J. Hausdorff, "Is every-day walking in older adults more analogous to dual-task walking or to usual walking? Elucidating the gaps between gait performance in the lab and during 24/7 monitoring," *Eur Rev Aging Phys Act*, vol. 16, no. 6, pp. 1-12, 2019.
- [142] A. Schache, P. Blanch, D. Rath, R. Starr and K. Bennell, "A comparison of overground and treadmill running for measuring the three-dimensional kinematics of the lumbo-pelvic-hip complex," *Clinical Biomechanics*, vol. 16, no. 8, pp. 667-680, 2001.
- [143] W. Edwards, "Modeling Overuse Injuries in Sport as a Mechanical Fatigue Phenomenon," *Exercise and Sport Science Reviews*, vol. 46, no. 4, pp. 224-231, 2018.
- [144] R. El Kati, S. Forrester and P. Fleming, "Evaluation of pressure insoles during running," *Procedia Engineering*, vol. 2, pp. 3053-3058, 2010.
- [145] R. W. Willy, "Innovations and pitfalls in the use of wearable devices in the prevention and rehabilitation of running related injuries," *Physical Therapy in Sport*, vol. 29, pp. 26-33, 2018.
- [146] V. Camomilla, E. Bergamini, S. Fantozzi and G. Vannozzi, "Trends Supporting the In-Field Use of Wearable Inertial Sensors for Sport Performance Evaluation: A Systematic Review," *Sensors*, vol. 18, no. 3, p. 873, 2018.
- [147] I. Moore and R. Willy, "Use of Wearables: Tracking and Retraining in Endurance Runners," *Current Sports Medicine Reports*, vol. 18, no. 12, pp. 437-444, 2019.
- [148] J. Mizrahi, O. Verbitsky, E. Isakov and D. Daily, "Effect of fatigue on leg kinematics and impact acceleration in long distance running," *Human Movement Science*, vol. 19, pp. 139-151, 2000.
- [149] J. Mercer, B. Bates, J. Dufek and A. Hreljac, "Characteristics of shock attenuation during fatigued running," *Journal of Sports Science*, vol. 21, no. 11, pp. 911-919, 2003.
- [150] B. Purcell, J. Channells, D. James and R. Barrett, "Use of accelerometers for detecting foot-ground contact time during running," in *Proc. SPIE 6036, BioMEMS and Nanotechnology II, 603615*, Brisbane, Australia, 2006.
- [151] K. Aminian, B. Najafi, C. Bula, P. Leyvraz and P. Robert, "Spatio-temporal parameters of gait measured by an ambulatory system using miniature gyroscopes," *Journal of Biomechanics*, vol. 35, pp. 689-699, 2002.
- [152] K. O'Donovan, B. Greene, D. McGrath, R. O'Neill, A. Burns and B. Caulfield, "SHIMMER: A new tool for temporal Gait analysis," in *31st Annual International Conference of the IEEE EMBS*, Minneapolis, Minnesota, 2009.
- [153] D. McGrath, B. Greene, K. O'Donovan and B. Caulfield, "Gyroscope-based assessment of temporal gait parameters during treadmill walking and running," *Sports Engineering*, vol. 15, pp. 207-213, 2012.
- [154] B. Greene, D. McGrath, R. O'Neill, K. O'Donovan, A. Burns and B. Caulfield, "An adaptive gyroscope-based algorithm for temporal gait analysis," *Med Biol Eng Comput*, vol. 48, pp. 1251-1260, 2010.
- [155] J. Sinclair, S. Hobbs, L. Protheroe, C. Edmundson and A. Greenhalgh, "Determination of gait events using an externally mounted shank accelerometer," *Journal of Applied Biomechanics*, vol. 29, no. 1, pp. 118-122, 2013.
- [156] N. Whelan, R. Healy, I. Kenny and A. Harrison, "A Comparison of Foot Strike Events Using the Force Plate and Peak Impact Acceleration Measures," in *International Society of Biomechanics in Sport 33*, Poitiers, France, 2015.

- [157] M. Norris, I. Kenny and R. Anderson, "Comparison of accelerometry stride time calculation methods," *Journal of Biomechanics*, vol. 49, pp. 3031-3034, 2016.
- [158] M. Schmidt, C. Rheinlander, K. Nolte, S. Wille, N. Wehn and T. Jaitner, "IMU- based determination of stance duration during sprinting," *Procedia Engineering*, vol. 147, pp. 747-752, 2016.
- [159] K. Aubol and C. Milner, "Foot contact identification using a single triaxial accelerometer during running," *Journal of Biomechanics*, vol. 105, p. 109768, 2020.
- [160] C. Fadillioglu, B. Stetter, S. Ringhoff, F. Krafft, S. Sell and T. Stein, "Automated gait event detection for a variety of locomotion tasks using a novel gyroscope-based algorithm," *Gait & Posture*, vol. 81, pp. 102-108, 2020.
- [161] M. Bach, N. Dominici and A. Daffertshofer, "Predicting vertical ground reaction forces from 3D accelerometry using reservoir computers leads to accurate gait event detection," *Frontiers in Sports and Active Living*, vol. 4, p. <https://doi.org/10.3389/fspor.2022.1037438>, 2022.
- [162] B. Auvinet, E. Gloria, G. Renault and E. Barrey, "Runner's stride analysis: comparison of kinematic and kinetic analyses under field conditions," *Science & Sports*, vol. 17, pp. 92-94, 2002.
- [163] J. Lee, R. Mellifont and B. Burkett, "The use of a single inertial sensor to identify stride, step, and stance durations of running gait," *Journal of Science and Medicine in Sport*, vol. 13, pp. 270-273, 2010.
- [164] E. Bergamini, P. Picerno, H. Pillet, F. Natta, P. Thoreux and V. Camomilla, "Estimation of temporal parameters during sprint running using a trunk-mounted inertial measurement unit," *Journal of Biomechanics*, vol. 45, no. 6, pp. 1123-1126, 2012.
- [165] L. Benson, C. Clermont, R. Watari, T. Exley and R. Ferber, "Automated Accelerometer-Based Gait Event Detection During Multiple Running Conditions Lauren," *Sensors*, vol. 19, p. 1483, 2019.
- [166] J. Reenalda, M. Zandbergen, J. Harbers, M. Paquette and C. Milner, "Detection of foot contact in treadmill running with inertial and optical measurement systems," *Journal of Biomechanics*, vol. 121, p. 110419, 2021.
- [167] United States Geological Survey, "Gravity anomaly map of the continental United States," [Online]. Available: <https://mrdata.usgs.gov/gravity/map-us.html#home>. [Accessed 5 June 2019].
- [168] National Geodetic Survey, "NGS surface gravity prediction," National Oceanic and Atmospheric Administration, [Online]. Available: [https://www.ngs.noaa.gov/cgi-bin/grav\\_pdx.prl](https://www.ngs.noaa.gov/cgi-bin/grav_pdx.prl). [Accessed 5 June 2019].
- [169] G. Wu and P. Cavanagh, "ISB recommendations for standardization in the reporting of kinematic data," *Journal of Biomechanics*, vol. 28, no. 10, pp. 1257-1261, 1995.
- [170] A. R. Altman and I. S. Davis, "A Kinematic Method for Footstrike Pattern Detection in Barefoot and Shod Runners," *Gait and Posture*, vol. 35, no. 2, pp. 298-300, 2012.
- [171] S. Madgwick, A. Harrison and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *IEEE International Conference on Rehabilitation Robotics*, 2011.
- [172] S. Ludwig, K. Burnham, A. Jimenez and P. Touma, "Comparison of attitude and heading reference systems using foot mounted MIMU sensor data: basic, Madgwick, and Mahony".
- [173] S. Ludwig and K. Burnham, "Comparison of Euler Estimate using Extended Kalman Filter, Madgwick and Mahony on Quadcopter Flight Data," in *International Conference on Unmanned Aircraft Systems*, 2018.



- [174] McGinnis and Perkins, "A Highly Miniaturized, Wireless Inertial Measurement Unit for Characterizing the Dynamics of Pitched Baseballs and Softballs," *Sensors*, vol. 12, pp. 11933-11945, 2012.
- [175] S. M. Cain, "IMUs (Inertial Measurement Units): Unboxing the black box," in *41st Annual Meeting of the American Society of Biomechanics*, Boulder, CO, 2017.
- [176] B. Carstensen, J. Simpson and L. Gurrin, "Statistical Models for Assessing Agreement in Method Comparison Studies with Replicate Measurements," *The International Journal of Biostatistics*, vol. 4, no. 1, p. 16, 2008.
- [177] P. Myles and J. Cui, "Using the Bland–Altman method to measure agreement with repeated measures," *British Journal of Anaesthesia*, vol. 99, no. 3, pp. 309-311, 2007.
- [178] S. Searle, "An overview of variance component estimation," *Metrika*, vol. 42, pp. 215-230, 1995.
- [179] J. Bland and D. Altman, "Measuring agreement in method comparison studies," *Statistical methods in medical research*, vol. 8, no. 2, pp. 135-160, 1999.
- [180] D. Kiernan and D. Hawkins, "Bad Vibes? Preliminary data on shock attenuation in injured and uninjured runners," in *American Society of Biomechanics*, Atlanta, Georgia, 2020.
- [181] A. Patoz, L. Thibault, B. Breine, C. Gindre and D. Malatesta, "Comparison of different machine learning models to enhance sacral acceleration-based estimations of running stride temporal variables and peak vertical ground reaction force," *Sports Biomechanics*, vol. 6, pp. 1-17, 2023.
- [182] M. Falbriard, A. Soltani and K. Aminian, "Running Speed Estimation Using Shoe-Worn Inertial Sensors: Direct Integration, Linear, and Personalized Model," *Frontiers in Sports and Active Living*, vol. 3, 2021.
- [183] S. Yang, C. Mohr and Q. Li, "Ambulatory running speed estimation using an inertial sensor," *Gait & Posture*, vol. 34, no. 4, pp. 462-466, 2011.
- [184] T. Shiang, T. Hsieh, Y. Lee, C. Wu, M. Yu, C. Mei and I. Tai, "Determine the Foot Strike Pattern Using Inertial Sensors," *Journal of Sensors*, vol. vol. 2016, p. Article ID 4759626, 2016.
- [185] M. Shorten and D. Winslow, "Spectral analysis of impact shock during running," *International Journal of Applied Biomechanics*, vol. 8, pp. 288-304, 1992.
- [186] J. Hammill, T. Derrick and K. Holt, "Shock attenuation and stride frequency during running," *Human Movement Science*, vol. 14, pp. 45-60, 1995.
- [187] T. Derrick, J. Hamill and G. Caldwell, "Energy absorption of impacts during running at various stride lengths," *Medicine and Science in Sports and Exercise*, vol. 30, pp. 128-135, 1998.
- [188] T. Derrick, D. DeReu and S. McLean, "Impacts and kinematic adjustments during an exhaustive run," *Medicine and Science in Sports and Exercise*, vol. 34, pp. 998-1002, 2002.
- [189] J. Luo, B. Jing and S. Jinhua, "Application of the wavelet transforms on axial strain calculation in ultrasound elastography," *Progress in Natural Science*, vol. 16, no. 9, pp. 942-947, 2006.
- [190] M. Lafortune, "Three-dimensional acceleration of the tibia during walking and running," *Journal of Biomechanics*, vol. 24, no. 10, pp. 877-886, 1991.
- [191] R. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. D, pp. 35-45, 1960.
- [192] R. Mahony, T. Hamel and J. Pflimlin, "Nonlinear complementary filters on the special orthogonal group," *IEEE Transactions on Automatic Control*, vol. 53, no. 5, pp. 1203-1218, 2008.
- [193] S. Madgwick, "An efficient orientation filter for inertial and inertial/magnetic sensor arrays," x-io, Bristol, UK, 2010.
- [194] B. Nigg and W. Liu, "The effect of muscle stiffness and damping on simulated impact force peaks during running," *Journal of Biomechanics*, vol. 32, pp. 849-856, 1999.

- [195] S. Bredeweg, B. Kluitenberg, B. Bessem and I. Buist, "Differences in kinetic variables between injured and noninjured novice runners: a prospective cohort study," *Journal of Science and Medicine in Sport*, vol. 16, no. 3, pp. 205-210, 2013.
- [196] K. Crossley, K. Bennell, T. Wrigley and B. Oakes, "Ground reaction forces, bone characteristics, and tibial stress fracture in male runners," *Medicine and Science in Sports and Exercise*, vol. 31, no. 8, pp. 1088-1093, 1999.
- [197] L. Malisoux, P. Gette, N. Delattre, A. Urhausen and D. Theisen, "Spatiotemporal and Ground-Reaction Force Characteristics as Risk Factors for Running-Related Injury: A Secondary Analysis of a Randomized Trial Including 800+ Recreational Runners," *American Journal of Sports Medicine*, vol. 50, no. 2, pp. 537-544, 2022.
- [198] D. Kiernan, D. Hawkins, M. Manoukian, M. McKallip, L. Oelsner, C. Caskey and C. Coolbaugh, "Accelerometer-based prediction of running injury in National Collegiate Athletic Association track athletes," *Journal of Biomechanics*, 2018.
- [199] L. Ceyskens, R. Vanelderden, C. Barton, P. Malliaras and B. Dingenen, "Biomechanical Risk Factors Associated with Running-Related Injuries: A Systematic Review," *Sports Medicine*, vol. 49, no. 7, pp. 1095-1115, 2019.
- [200] M. Pohl, J. Hamill and I. Davis, "Biomechanical and anatomic factors associated with a history of plantar fasciitis in female runners," *Clinical Journal of Sports Medicine*, vol. 19, no. 5, pp. 372-376, 2009.
- [201] D. Kuhman, M. Paquette, S. Peel and D. Melcher, "Comparison of ankle kinematics and ground reaction forces between prospectively injured and uninjured collegiate cross country runners," *Human Movement Science*, vol. 47, pp. 9-15, 2016.
- [202] R. Nagahara, M. Mizutani, A. Matsuo, H. Kanehisa and T. Fukunaga, "Association of Sprint Performance With Ground Reaction Forces During Acceleration and Maximal Speed Phases in a Single Sprint," *Journal of Applied Biomechanics*, vol. 34, no. 2, pp. 104-110, 2016.
- [203] J. Morin, J. Slawinski, S. Dorel, E. de Villareal, A. Couturier, P. Samozino, M. Brughelli and G. Rabita, "Acceleration capability in elite sprinters and ground impulse: Push more, brake less?," *Journal of Biomechanics*, vol. 48, no. 12, pp. 3149-3154, 2015.
- [204] G. Rabita, S. Dorel, J. Slawinski, E. Saez de Villareal, A. Couturier, P. Samozino and J. Morin, "Sprint mechanics in world-class athletes: a new insight into the limits of human locomotion," *Scandinavian Journal of Medicine and Science in Sports*, vol. 25, no. 5, pp. 583-594, 2015.
- [205] R. Nagahara, M. Mizutani, A. Matsuo, H. Kanehisa and T. Fukunaga, "Association of Sprint Performance With Ground Reaction Forces During Acceleration and Maximal Speed Phases in a Single Sprint," *Journal of Applied Biomechanics*, vol. 34, no. 2, pp. 1-20, 2017.
- [206] T. Dos'Santos, C. Thomas, P. Jones and P. Comfort, "Mechanical Determinants of Faster Change of Direction Speed Performance in Male Athletes," *Journal of Strength and Conditioning Research*, vol. 31, pp. 696-705, 2017.
- [207] D. Kiernan, K. Dunn-Siino and D. Hawkins, "Unsupervised Gait Event Identification with a Single Wearable Accelerometer and/or Gyroscope: A Comparison of Methods across Running Speeds, Surfaces, and Foot Strike Patterns," *Sensors*, vol. 23, no. 11, p. 5022, 2023.
- [208] R. Alcantara, W. Edwards, G. Millet and A. Grabowski, "Predicting continuous ground reaction forces from accelerometers during uphill and downhill running: a recurrent neural network solution," *PeerJ*, vol. 4, no. 10, p. e12752, 2022.
- [209] A. Voloshina and D. Ferris, "Biomechanics and energetics of running on uneven terrain," *Journal of Experimental Biology*, vol. 218, pp. 711-719, 2015.
- [210] s. T. P. Kipp and R. Kram, "Ground reaction forces during steeplechase hurdling and waterjumps," *Sports Biomechanics*, vol. 16, pp. 152-165, 2017.

- [211] C. Whiting, S. Allen, J. Brill and R. Kram, "Steep ( $30^\circ$ ) uphill walking vs. running: COM movements, stride kinematics, and leg muscle excitations," *European Journal of Applied Physiology*, vol. 120, pp. 2147-2157, 2020.
- [212] C. Johnson, J. Outerlyers, S. Jamison, A. Tenforde, M. Ruder and I. Davis, "Comparison of Tibial Shock during Treadmill and Teal-World Running," *Medicine & Science in Sports & Exercise*, vol. 52, no. 7, pp. 1557-1562, 2020.
- [213] B. Davis, J. Perry, D. Neth and K. Waters, "A Device for Simultaneous Measurement of Pressure and Shear Force Distribution on the Plantar Surface of the Foot," *Journal of Applied Biomechanics*, vol. 14, pp. 93-104, 1998.
- [214] M. Razian and M. Pepper, "Design, development, and characteristics of an in-shoe triaxial pressure measurement transducer utilizing a single element of piezoelectric copolymer film.," *IEEE Trans Neural Syst. Rehabil. Eng.*, vol. 11, pp. 288-293, 2003.
- [215] M. Koch, L.-K. Lunde, M. Ernst, S. Knardahl and K. Veiersted, "Validity and reliability of pressure-measurement insoles for vertical ground reaction force assessment in field situations," *Applied Ergonomics*, vol. 53, pp. 44-51, 2016.
- [216] C. Price, D. Parker and C. Nester, "Validity and repeatability of three in-shoe pressure measurement systems," *Gait & posture*, vol. 46, pp. 69-74, 2016.
- [217] A. Cordero, H. Koopman and F. van der Helm, "Use of pressure insoles to calculate the complete ground reaction forces," *Journal of Biomechanics*, vol. 37, no. 9, pp. 1427-1432, 2004.
- [218] G. Burns, J. Zandler and R. Zernicke, "Wireless insoles to measure ground reaction forces: Step-by-step validity in hopping, walking, and running," *ISBS Proceedings Archive*, vol. 35, no. 1, p. 255, 2017.
- [219] T. Liu, Y. Inoue, K. Shibata, K. Shiojima and M. Han, "Triaxial joint moment estimation using a wearable three-dimensional gait analysis system," *Measurement*, vol. 47, pp. 125-129, 2014.
- [220] D. Fong, Y. Chan, Y. Hong, P. Yung, K. Fung and K. Chan, "Estimating the complete ground reaction forces with pressure insoles in walking," *Journal of Biomechanics*, vol. 41, pp. 2597-2601, 2008.
- [221] Y. Jung, M. Jung, K. Lee and S. Koo, "Ground reaction force estimation using an insole-type pressure mat and joint kinematics during walking," *Journal of Biomechanics*, vol. 47, no. 11, pp. 2693-2699, 2014.
- [222] G. Faber, I. Kingma, H. Martin Schepers, P. Veltink and J. van Dieen, "Determination of joint moments with instrumented force shoes in a variety of tasks," *Journal of Bioemchanics*, vol. 43, pp. 2848-2854, 2010.
- [223] C. Liedtke, S. Fokkenrood, J. Menger, H. van der Kooij and P. Veltink, "Evaluation of instrumented shoes for ambulatory assessment of ground reaction forces," *Gait & Posture*, vol. 26, pp. 39-47, 2007.
- [224] T. Liu, Y. Inoue and K. Shibata, "A wearable ground reaction force sensor system and its application to the measurement of extrinsic gait variability," *Sensors*, vol. 10, no. 11, pp. 10240-10255, 2010.
- [225] Y. Guo, F. Storm, Y. Zhao, S. Billings, A. Pavic, C. Mazza and L. Guo, "A New Proxy Measurement Algorithm with Application to the Estimation of Vertical Ground Reaction Forces Using Wearable Sensors," *Sensors*, vol. 17, no. 10, p. 2181, 2017.
- [226] H. Lim, B. Kim and S. Park, "Prediction of Lower Limb Kinetics and Kinematics during Walking by a Single IMU on the Lower Back Using Machine Learning," *Sensors*, vol. 20, no. 130, 2020.
- [227] G. Leporace, L. Batista and J. Nadal, "Prediction of 3D ground reaction forces during gait based on accelerometer," *Research on Biomedical Engineering*, vol. 34, no. 3, pp. 211-216, 2018.

- [228] A. Karatsidis and M. Schepers, "Estimation of Ground Reaction Forces and Moments During Gait Using Only Inertial Motion Capture," *Sensors*, vol. 17, no. 1, p. 75, 2016.
- [229] F. Wouda, M. Giuberti, G. Bellusci, E. Maartens, J. Reenalda, B. van Beijnum and P. Veltink, "Estimation of vertical ground reaction forces and sagittal knee kinematics during running using three inertial sensors," *Frontiers in Physiology*, vol. 9, pp. 1-14, 2018.
- [230] E. Dorschky, M. Nitschke, C. Martindale, A. van den Bogert, A. Koelewijn and B. Eskofier, "CNN-Based Estimation of Sagittal Plane Walking and Running Biomechanics From Measured and Simulated Inertial Sensor Data," *Frontiers in Bioengineering and Biotechnology*, vol. 8, 2020.
- [231] W. Johnson, A. Mian, M. Robinson, J. Verheul, D. Lloyd and J. Alderson, "Multidimensional Ground Reaction Forces and Moments From Wearable Sensor Accelerations via Deep Learning," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 1, pp. 289 - 297, 2021.
- [232] S. Higgins, L. Higgins and S. Vallabhajosula, "Site-specific Concurrent Validity of the ActiGraph GT9X Link in the Estimation of Activity-related Skeletal Loading," *Medicine & Science in Sports & Exercise*, vol. 53, no. 5, pp. 51-959, 2021.
- [233] V. Stiles, P. Griew and A. Rowlands, "Use of Accelerometry to Classify Activity Beneficial to Bone in Premenopausal Women," *Medicine & Science in Sports & Exercise*, vol. 45, no. 12, p. 23532361, 2013.
- [234] J. Denoth, "Load on the locomotor system and modeling," in *Biomechanics of Running Shoes*, Champaign, IL, Human Kinetics, 1986, pp. 78-84.
- [235] E. Hennig and M. Lafortune, "Relationships between Ground Reaction Force and Tibial Bone Acceleration Parameters," *Journal of Applied Biomechanics*, vol. 7, no. 3, pp. 303-309, 1991.
- [236] E. Hennig, T. Milani and M. Lafortune, "Use of Ground Reaction Force Parameters in Predicting Peak Tibial Accelerations in Running," *Journal of Applied Biomechanics*, vol. 9, no. 4, pp. 306-314, 1993.
- [237] A. Greenhalgh, J. Sinclair, L. Protheroe and N. Chockalingam, "Predicting Impact Shock Magnitude: Which Ground Reaction Force Variable Should We Use?," *International Journal of Sports Science and Engineering*, vol. 6, pp. 225-231, 2012.
- [238] J. Zhang, W. An, I. Au, T. Chen and R. Cheung, "Comparison of the correlations between impact loading rates and peak accelerations measured at two different body sites: Intra- and inter-subject analysis," *Gait & Posture*, vol. 46, pp. 53-56, 2016.
- [239] P. Van den Berghe, J. Six, J. Gerlo, M. Leman and D. De Clercq, "Validity and reliability of peak tibial accelerations as real-time measure of impact loading during over-ground rearfoot running at different speeds," *Journal of Biomechanics*, vol. 86, pp. 238-242, 2019.
- [240] D. Thiel, J. Shepherd, H. Espinosa, M. Kenny, K. Fischer, M. Worsley, A. Matsuo and T. Wada, "Predicting Ground Reaction Forces in Sprint Running Using a Shank Mounted Inertial Measurement Unit," *Proceedings*, vol. 2, no. 6, p. 199, 2018.
- [241] A. Tenforde, T. Hayano, S. Jamison, J. Outerleys and I. Davis, "Tibial Acceleration Measured from Wearable Sensors Is Associated with Loading Rates in Injured Runners," *PMR*, vol. 12, pp. 679-684, 2020.
- [242] D. Raper, J. Witchalls, E. Phillips and M. Drew, "Use of a tibial accelerometer to measure ground reaction force in running: A reliability and validity comparison with force plates," *Journal of Science and Medicine in Sport*, vol. 21, no. 1, pp. 84-88, 2017.
- [243] M. Bobbert, H. Schamhardt and B. Nigg, "Calculation of vertical ground reaction force estimates during running from positional data," *Journal of Biomechanics*, vol. 24, no. 12, pp. 1095-1105, 1991.
- [244] A. Rowlands and V. Stiles, "Accelerometer counts and raw acceleration output in relation to mechanical loading," *Journal of Biomechanics*, vol. 45, pp. 448-454, 2012.

- [245] G. Cavagna, L. Komarek and S. Mazzoleni, "The mechanics of sprint running," *Journal of Physiology*, vol. 217, pp. 709-721, 1971.
- [246] D. Wundersitz, K. Netto, B. Aisbett and P. Gastin, "Validity of an upper-body-mounted accelerometer to measure peak vertical and resultant force during running and change-of-direction tasks," *Sports Biomechanics*, vol. 12, no. 4, pp. 403-412, 2013.
- [247] R. Gurchiek, R. McGinnis, A. Needle, J. McBride and H. van Werkhoven, "The use of a single inertial sensor to estimate 3-dimensional ground reaction force during accelerative running tasks," *Journal of Biomechanics*, vol. 61, pp. 263-268, 2017.
- [248] R. Alcantara, E. Day, M. Hahn and A. Grabowski, "Sacral acceleration can predict whole-body kinetics and stride kinematics across running speeds," *PeerJ*, vol. 9, p. e11199, 2021.
- [249] C. Napier, X. Jiang, C. MacLean, C. Menon and M. Hunt, "The use of a single sacral marker method to approximate the centre of mass trajectory during treadmill running," *Journal of Biomechanics*, vol. 108, p. 109886, 2020.
- [250] S. Gard, S. Miff and A. Kuo, "Comparison of kinematic and kinetic methods for computing the vertical motion of the body center of mass during walking," *Human Movement Science*, vol. 22, p. 597-610, 2004.
- [251] L. Gullstrand, K. Halvorsen, F. Tinmark, M. Eriksson and J. Nilsson, "Measurements of vertical displacement in running, a methodological comparison," *Gait & Posture*, vol. 30, pp. 71-75, 2009.
- [252] A. Ranavolo, R. Don, A. Cacchio, M. Serrao, M. Paoloni, M. Mangone and V. Santilli, "Comparison between Kinematic and Kinetic Methods for Computing the Vertical Displacement of the Center of Mass during Human Hopping at Different Frequencies," *Journal of Applied Biomechanics*, vol. 24, pp. 271-279, 2008.
- [253] M. Whittle, "Three-dimensional motion of the center of gravity of the body during walking," *Human Movement Science*, vol. 16, pp. 347-355, 1997.
- [254] A. Ancillao, S. Tedesco, J. Barton and B. O'Flynn, "Indirect Measurement of Ground Reaction Forces and Moments by Means of Wearable Inertial Sensors: A Systematic Review," *Sensors*, vol. 18, no. 8, p. 2564, 2018.
- [255] L. Veras, F. Diniz-Sousa, G. Boppre, A. Resende-Coelho, E. Moutinho-Ribeiro, V. Devezas, H. Santos-Sousa, J. Preto, J. Vilas-Boas, L. Machado, J. Oliveira and H. Fonseca, "Mechanical loading prediction through accelerometry data during walking and running," *European Journal of Sport Science*, vol. 15, pp. 1-18, 2022.
- [256] K. Janz, R. Smita, H. Baumann and J. Schulz, "Measuring Children's Vertical Ground Reaction Forces with Accelerometry during Walking, Running, and Jumping: The Iowa Bone Development Study," *Pediatric Exercise Science*, vol. 15, no. 1, pp. 34-43, 2003.
- [257] A. Garcia, C. Langenthal, R. Angulo-Barroso and M. Gross, "A Comparison of Accelerometers for Predicting Energy Expenditure and Vertical Ground Reaction Force in School-Age Children," *Measurement in Physical Education and Exercise Science*, vol. 8, no. 3, pp. 119-144, 2004.
- [258] N. Nedergaard, J. Verheul, B. Drust, T. EtcHELLS, P. Lisboa, M. Robinson and J. Vanrenterghem, "The feasibility of predicting ground reaction forces during running from a trunk accelerometry driven mass-spring-damper model," *PeerJ*, vol. 6, p. e6105, 2018.
- [259] R. Deerie, P. Robberechts, P. Van Den Berghe, J. Gerlo, D. De Clercq, V. Segers and J. Davis, "Tibial Acceleration-Based Prediction of Maximal Vertical Loading Rate During Overground Running: A Machine Learning Approach," *Frontiers in Bioengineering and Biotechnology*, vol. 8, p. 33, 2020.
- [260] D. Sharma, P. Davidson, P. Muller and R. Piche, "Indirect Estimation of Vertical Ground Reaction Force from a Body-Mounted INS/GPS Using Machine Learning," *Sensors*, vol. 21, no. 4, p. 1553, 2021.

- [261] N. Jie-Han, D. Gouwanda, A. Gopalai and C. Zheng, "Estimation of vertical ground reaction force during running using neural network model and uniaxial accelerometer," *Journal of Biomechanics*, vol. 76, pp. 269-273, 2018.
- [262] S. Donahue and M. Hahn, "Estimation of gait events and kinetic waveforms with wearable sensors and machine learning when running in an unconstrained environment," *Scientific Reports*, vol. 13, no. 1, p. 2339, 2023.
- [263] P. Davidson, H. Virekunnas, D. Sharma, R. Piche and C. Cronin, "Continuous Analysis of Running Mechanics by Means of an Integrated INS/GPS Device," *Sensors*, vol. 19, no. 6, p. 1480, 2019.
- [264] E. Charry, W. Hu, M. Umer, A. Ronchi and S. Taylor, "Study on Estimation of Peak Ground Reaction Forces using Tibial Accelerations in Running," in *IEEE 8th International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, Melbourne, Australia, 2013.
- [265] U. Meyer, D. Ernst, S. Schott, C. Riera, J. Hattendorf, J. Romkes, U. Granacher, B. Gopfert and S. Kriemler, "Validation of two accelerometers to determine mechanical loading of physical activities in children," *Sports Science*, vol. 33, pp. 1702-1709, 2015.
- [266] D. Kiernan, B. Ng and D. Hawkins, "May the force be with you: Acceleration-based estimates of vertical ground reaction forces during running," in *67th American College of Sports Medicine*, San Francisco, CA, 2020.
- [267] B. Kim and S. Park, "Estimation of 3D GRF during running from a single sacral measurement using artificial neural network," in *Dynamic Walking*, Hawley, PA, 2020.
- [268] M. Pogson, J. Verheul, M. Robinson, J. Vanrenterghem and L. Paulo, "A neural network method to predict task- and step-specific ground reaction force magnitudes from trunk accelerations during running activities," *Medical Engineering and Physics*, vol. 78, pp. 82-89, 2020.
- [269] E. Day, R. Alcantara, M. McGeehan, A. Grabowski and M. Hahn, "Low-pass filter cutoff frequency affects sacral-mounted inertial measurement unit estimations of peak vertical ground reaction force and contact time during treadmill running," *Journal of Biomechanics*, vol. 119, p. 110323, 2021.
- [270] A. R. Altman and I. S. Davis, "Prospective comparison of running injuries between shod and barefoot runners," *British Journal of Sports Medicine*, vol. 50, no. 8, pp. 476-480, 2016.
- [271] T. Blackmore, R. W. Willy and M. W. Creaby, "The high frequency component of the vertical ground reaction force is a valid surrogate measure of the impact peak," *Journal of Biomechanics*, vol. 49, no. 3, pp. 479-483, 2016.
- [272] C. Milner, R. Ferber, C. Pollard, J. Hamill and I. Davis, "Biomechanical factors associated with tibial stress fracture in female runners," *Medicine and Science in Sports and Exercise*, vol. 38, no. 2, pp. 323-328, 2006.
- [273] B. Carstensen, "Replicate Measurements," in *Comparing Clinical Measurement Methods: A Practical Guide*, Hoboken, New Jersey, John Wiley & Sons. Ltd, 2010, pp. 49-65.
- [274] B. Carstensen, L. Gurrin, C. Ekstrom and M. Figurski, "Package 'MethComp'," 12 October 2022. [Online]. Available: <https://cran.r-project.org/web/packages/MethComp/MethComp.pdf>. [Accessed 23 August 2023].
- [275] A. Miltko, C. Milner, D. Powell and M. Paquette, "The influence of surface and speed on biomechanical external loads obtained from wearable devices in rearfoot strike runners," *Sports Biomechanics*, vol. 10, pp. 1-15, 2022.
- [276] A. Norsteh, E. Hajihosseini, S. Emami and H. Mahmoudi, "Assessing Thoracic and Lumbar Spinal Curvature Norm: A Systematic Review," *Physical Treatments*, vol. 9, no. 4, pp. 183-191, 2019.

- [277] T. Keller, A. Weisberger, J. Ray, S. Hasan and D. Spengler, "Relationship between vertical ground reaction force and speed during walking, slow jogging, and running," *Clinical Biomechanics*, vol. 11, no. 5, pp. 253-259, 1996.
- [278] J. Neugebauer and M. Lafiandra, "Predicting Ground Reaction Force from a Hip-Borne Accelerometer during Load Carriage," *Medicine & Science in Sports & Exercise*, vol. 50, no. 11, pp. 2369-2374, 2018.
- [279] E. Charry, W. Hu, M. Umer, A. Ronchi and S. Taylor, "Study on estimation of peak Ground Reaction Forces using tibial accelerations in running," in *IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, Melbourne, VIC, Australia, 2013.
- [280] D. Wundersitz, K. Netto, B. Aisbett and P. Gastin, "Validity of an upper-body-mounted accelerometer to measure peak vertical and resultant force during running and change-of-direction tasks," *Sports Biomechanics*, vol. 12, no. 4, pp. 403-412, 2013.
- [281] U. Meyer, D. Ernst, S. Schott, C. Riera, J. Hattendorf, J. Romkes, U. Granacher, B. Gopfert and S. Kriemler, "Validation of two accelerometers to determine mechanical loading of physical activities in children," *Journal of Sports Science*, vol. 33, pp. 1702-1709, 2015.
- [282] Kiernan, Ng and Hawkins, "May the force be with you: Acceleration-based estimates of vertical ground reaction forces during running," *TBD*, vol. TBD, no. TBD, p. TBD, TBD.
- [283] M. Pogson, J. Verheul, M. Robinson, J. Vanrenterghem and P. Lisboa, "A neural network method to predict task- and step-specific ground reaction force magnitudes from trunk accelerations during running activities," *Medical Engineering & Physics*, vol. 78, pp. 82-89, 2020.
- [284] E. Day, R. Alcantara, M. McGeehan, A. Grabowksi and M. Hahn, "Low-pass filter cutoff frequency affects sacral-mounted inertial measurement unit estimations of peak vertical ground reaction force and contact time during treadmill running," *Journal of Biomechanics*, vol. 119, p. 110323, 2021.
- [285] S. Higgins, L. Higgins and S. Vallabhajosula, "Site-specific Concurrent Validity of the ActiGraph GT9X Link in the Estimation of Activity-related Skeletal Loading," *Medicine & Science in Sports & Exercise*, vol. 53, no. 5, pp. 951-959, 2021.
- [286] D. Fong and C. YY, "The use of wearable inertial motion sensors in human lower limb biomechanics studies: A systematic review," *Sensors*, vol. 10, no. 12, pp. 11556-11565, 2010.
- [287] D. Raper, J. Witchalls, E. Phillips and M. Drew, "Use of a tibial accelerometer to measure ground reaction force in running: A reliability and validity comparison with force plates," *Journal of Science and Medicine in Sport*, vol. 21, no. 1, pp. 84-88, 2018.
- [288] T. Owings and M. Grabiner, "Measuring step kinematic variability on an instrumented treadmill: how many steps are enough?," *Journal of Biomechanics*, vol. 36, no. 8, pp. 1215-1218, 2003.
- [289] A. Karatsidis and M. Schepers, "Estimation of Ground Reaction Forces and Moments During Gait Using Only Inertial Motion Capture," *Sensors*, vol. 17, no. 75, p. s17010075, 2016.
- [290] N. Chambon, N. Delattre, N. Gueguen, E. Berton and G. Rao, "Shoe drop has opposite influence on running pattern when running overground or on a treadmill," *European Journal of Applied Physiology*, vol. 115, no. 5, pp. 911-918, 2015.
- [291] C. Johnson, J. Outerleys, S. Jamison, A. Tenforde, M. Ruder and I. Davis, "Comparison of Tibial Shock during Treadmill and Real-World Running," *Medicine & Science in Sports & Exercise*, vol. 52, no. 7, pp. 1557-1562, 2020.
- [292] M. Paquette and D. Melcher, "Impact of a Long Run on Injury-Related Biomechanics with Relation to Weekly Mileage in Trained Male Runners," *Journal of Applied Biomechanics*, vol. 33, no. 3, pp. 216-221, 2017.
- [293] M. Janssen, J. Scheerder, E. Thibaut, A. Brombacher and S. Vos, "Who uses running apps and sport watches? Determinants and consumer profiles of event runners' usage of

- running-related smartphone applications and smart watches," *PLoS ONE*, vol. 12, p. e0181167, 2017.
- [294] M. Pobiruchin, J. Suleder, R. Zowalla, M. Wiesner, G. Wilson, M. Mauriello and F. Cena, "Accuracy and Adoption of Wearable Technology Used by Active Citizens: A Marathon Event Field Study.," *JMIR mHealth and uHealth*, vol. 5, p. e24, 2017.
- [295] C. Clermont, L. Duffett-Leger, B. Hettinga and R. Ferber, "Runners' Perspectives on 'Smart' Wearable Technology and Its Use for Preventing Injury.," *International Journal of Human Computer Interaction*, vol. 36, pp. 31-40, 2019.
- [296] D. Kiernan, B. Ng and D. Hawkins, "Acceleration-Based Estimation of Vertical Ground Reaction Forces during Running: A Comparison of Methods across Running Speeds, Surfaces, and Foot Strike Patterns," *Sensors*, vol. 23, no. 21, p. 8719, 2023.
- [297] M. Falbriard, A. Soltani and K. Aminian, "Running Speed Estimation Using Shoe-Worn Inertial Sensors: Direct Integration, Linear, and Personalized Model," *Frontiers in Sports and Active Living*, vol. 3, p. 585809, 2021.
- [298] V. Hernandez, D. Dadkhah, V. Babakeshizadeh and D. Kulic, "Lower body kinematics estimation from wearable sensors for walking and running: A deep learning approach," *Gait & Posture*, vol. 83, pp. 185-193, 2021.
- [299] J. Reenalda, E. Maartens, J. Buurke and A. Gruber, "Kinematics and shock attenuation during a prolonged run on the athletic track as measured with inertial magnetic measurement units," *Gait & Posture*, vol. 68, pp. 155-160, 2019.
- [300] J. Reenalda, E. Maartens, L. Homan and J. Buurke, "Continuous three dimensional analysis of running mechanics during a marathon by means of inertial magnetic measurement units to objectify changes in running mechanics," *Journal of Biomechanics*, vol. 49, no. 14, pp. 3362-3367, 2016.
- [301] M. Ruder, S. Jamison, A. Tenforde, F. Mulloy and I. Davis, "Relationship of Foot Strike Pattern and Landing Impacts during a Marathon," *Med Sci Sports Exerc*, vol. 51, no. 10, pp. 2073-2079, 2019.
- [302] C. Melo, F. Carpes, T. Vieira, T. Mendes, L. de Paula, M. Chagas, G. Peixoto and A. Pereira de Andrade, "Correlation between running asymmetry, mechanical efficiency, and performance during a 10 km run," *Journal of Biomechanics*, vol. 109, p. 109913, 2020.
- [303] F. Meyer, M. Falbriard, B. Mariani, K. Aminian and G. Millet, "Continuous Analysis of Marathon Running Using Inertial Sensors: Hitting Two Walls?," *International Journal of Sports Medicine*, vol. 42, no. 13, pp. 1182-1190, 2021.
- [304] A. Gruber, J. McDonnell, J. Davis, J. Vollmar, J. Harelak and M. Paquette, "Monitoring Gait Complexity as an Indicator for Running-Related Injury Risk in Collegiate Cross-Country Runners: A Proof-of-Concept Study," *Frontiers in Sport and Active Living*, vol. 21, no. 3, p. 630975, 2021.
- [305] D. Moran, R. Evans, Y. Arbel, O. Luria, A. Hadid, R. Yanovich, C. Milgrom and A. Finestone, "Physical and psychological stressors linked with stress fractures in recruit training," *Scandinavian Journal of Medicine and Science in Sports*, vol. 23, no. 4, pp. 443-450, 2013.
- [306] A. DeJong Lempke, J. Hart, D. Hryvniak, J. Rodu and J. Hertel, "Running-Related Injuries Captured Using Wearable Technology during a Cross-Country Season: A Preliminary Study," *Translational Journal of the American College of Sports Medicine*, vol. 8, no. 1, pp. 1-6, 2023.
- [307] C. Wood and K. Kipp, "Use of Audio Biofeedback to Reduce Tibial Impact Accelerations During Running," *Journal of Biomechanics*, vol. 47, no. 7, pp. 1739-1741, 2014.
- [308] A. Giraldo-Pedroza, W. Lee, W. Lam, R. Coman and G. Alici, "Effects of Wearable Devices with Biofeedback on Biomechanical Performance of Running—A Systematic Review," *Sensors*, vol. 20, no. 22, p. 6637, 2020.



- [309] J. Morris, D. Goss, E. Miller and I. Davis, "Using real-time biofeedback to alter running biomechanics: A randomized controlled trial," *Translational Sports Medicine*, vol. 3, no. 1, pp. 63-71, 2020.
- [310] B. Van Hooren, J. Goudsmit, J. Restrepo and S. Vos, "Real-time feedback by wearables in running: Current approaches, challenges and suggestions for improvements," *Journal of Sports Sciences*, vol. 38, no. 2, pp. 214-230, 2020.
- [311] C. Napier, J. Esculier and M. Hunt, "Gait retraining: out of the lab and onto the streets with the benefit of wearables," *British Journal of Sports Medicine*, vol. 51, no. 23, pp. 1642-1643, 2017.
- [312] T. Tan, D. Chiasson, H. Hu and P. Shull, "Influence of IMU position and orientation placement errors on ground reaction force estimation," *Journal of Biomechanics*, vol. 97, p. 109416, 2019.
- [313] M. Ruder, M. Hunt, J. Charlton and C. Tse, "Validity and reliability of gait metrics derived from researcher-placed and self-placed wearable inertial sensors," *Journal of Biomechanics*, vol. 142, p. 111263, 2022.
- [314] S. Chen, J. Brantley, T. Kim, S. Ridenour and J. Lach, "Characterizing and Minimizing Sources of Error in Inertial Body Sensor Networks," *International Journal of Autonomous and Adaptive Communications Systems*, vol. 6, no. 3, pp. 253-271, 2013.
- [315] L. Sara, J. Outerleys and C. Johnson, "The effect of sensor placement on measured distal tibial accelerations during running," *Journal of Applied Biomechanics*, vol. 39, no. 3, pp. 1-5, 2023.
- [316] C. Napier, R. Willy, B. Hannigan, R. McCann and C. Menon, "The Effect of Footwear, Running Speed, and Location on the Validity of Two Commercially Available Inertial Measurement Units During Running," *Frontiers in Sports and Active Living*, vol. 3, p. 643385, 2021.
- [317] C. Decker, N. Prasad and G. Kawchuk, "The reproducibility of signals from skin-mounted accelerometers following removal and replacement," *Gait & Posture*, vol. 34, no. 3, pp. 432-434, 2011.
- [318] R. Mason, L. Pearson, G. Barry, F. Young, O. Lennon, A. Godfrey and S. Stuart, "Wearables for Running Gait Analysis: A Systematic Review," *Sports Medicine*, vol. 53, no. 1, pp. 241-268, 2023.
- [319] A. Lucas-Cuevas, A. Encarnacion-Martinez, A. Camacho-Garcia, S. Llana-Belloch and P. Perez-Soriano, "The location of the tibial accelerometer does influence impact acceleration parameters during running," *Journal of Sport Science*, vol. 35, no. 17, pp. 1734-1738, 2017.
- [320] M. Lafortune and E. Hennig, "Contribution of angular motion and gravity to tibial acceleration," *Medicine and Science in Sports and Exercise*, vol. 23, no. 3, pp. 360-363, 1991.
- [321] C. Johnson, J. Outerleys, A. Tenforde and I. Davis, "A comparison of attachment methods of skin mounted inertial measurement units on tibial accelerations," *Journal of Biomechanics*, vol. 113, p. 110118, 2020.
- [322] M. Lafortune, E. Hennig and G. Valiant, "Tibial shock measured with bone and skin mounted transducers," *Journal of Biomechanics*, vol. 28, pp. 989-993, 1995.
- [323] A. H. Gruber, K. A. Boyer, T. R. Derrick and J. Hamill, "Impact shock frequency components and attenuation in rearfoot and forefoot running," *Journal of Sport and Health Science*, vol. 3, no. 2, pp. 113-121, 2014.
- [324] D. Kiernan, Z. Katzman, D. Hawkins and B. Christiansen, "A 0.05 m change in inertial measurement unit placement alters time and frequency domain metrics during running," *Sensors*, vol. TBD, no. TBD, p. TBD, TBD.
- [325] E. Erdfelder, F. Faul and A. Buchner, "GPOWER: A general power analysis program," *Behavior Research Methods, Instruments, & Computers*, vol. 28, pp. 1-11, 1996.

- [326] P. MR, N. C, W. RW and S. T, "Moving Beyond Weekly "Distance": Optimizing Quantification of Training Load in Runners," *Journal of Orthopaedic & Sports Physical Therapy*, vol. 50, no. 10, pp. 564-569, 2020.
- [327] J. Nilsson and A. Thorstensson, "Ground reaction forces at different speeds of human walking and running," *Acta Physiol Scand*, vol. 136, pp. 217-227, 1989.
- [328] D. Mullineaux, C. Milner, I. Davis and J. Hamill, "Normalization of Ground Reaction Forces," *Journal of Applied Biomechanics*, vol. 22, no. 3, pp. 230-233, 2006.
- [329] C. Firminger, M. Asmussen, S. Cigoja, J. Fletcher, B. Nigg and W. Edwards, "Cumulative Metrics of Tendon Load and Damage Vary Discordantly with Running Speed," *Medicine & Science in Sports & Exercise*, vol. 49, no. 5S, pp. 1549-1556, 2020.
- [330] D. Kiernan, A. Sherrill and D. Hawkins, "Predicting overuse injury: Meaningful surrogates of damage and their treatment over time," in *University of California Davis Biomedical Engineering Graduate Group Conference*, Davis, CA, 2019.
- [331] Bruce, Baggaley, Khassetarash, Haider and Edwards, "Tibial-fibular geometry and density variations associated with elevated bone strain and sex disparities in young active adults," *Bone*, vol. 161, p. 116443, 2022.
- [332] O. Bruce, M. Baggaley, L. Welte, M. Rainbow and W. Edwards, "A statistical shape model of the tibia-fibula complex: sexual dimorphism and effects of age on reconstruction accuracy from anatomical landmarks," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 25, no. 8, pp. 875-886, 2022.
- [333] C. Firminger, I. Haider, O. Bruce, J. Wannop, D. Stefanyshyn and W. Edwards, "Are subject-specific models necessary to predict patellar tendon fatigue life? A finite element modelling study," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 25, no. 7, pp. 729-739, 2022.
- [334] B. Clarsen, O. Ronsen, G. Myklebust, T. J. Florenes and R. Bahr, "The Oslo Sports Trauma Research Center questionnaire on health problems: a new approach to prospective monitoring of illness and injury in elite athletes," *British Journal of Sports Medicine*, vol. 48, no. 9, pp. 754-760, 2014.