

# The Number of Steps for Representative Real-World, Unsupervised Walking Data Using a Shoe-Worn Inertial Sensor

Jesse M. Charlton<sup>ID</sup>, Calvin Kuo, *Member, IEEE*, and Michael A. Hunt<sup>ID</sup>

**Abstract**—Inertial measurement units are now commonly used to quantify gait in healthy and clinical populations outside the laboratory environment, yet it is unclear how much data needs to be collected in these highly variable environments before a consistent gait pattern is identified. We investigated the number of steps to reach consistent outcomes calculated from real-world, unsupervised walking in people with ( $n=15$ ) and without ( $n=15$ ) knee osteoarthritis. A shoe-embedded inertial sensor measured seven foot-derived biomechanical variables on a step-by-step basis during purposeful, outdoor walking over seven days. Univariate Gaussian distributions were generated from incrementally larger training data blocks (increased in 5 step increments) and compared to all unique testing data blocks (5 steps/block). A consistent outcome was defined when the addition of another testing block did not change the percent similarity of the training block by more than 0.01% and this was maintained for the subsequent 100 training blocks (equivalent to 500 steps). No evidence was found for differences between those with and without knee osteoarthritis ( $p=0.490$ ), but the measured gait outcomes differed in the number of steps to become consistent ( $p<0.001$ ). The results demonstrate that collecting consistent foot-specific gait biomechanics is feasible in free-living conditions. This supports the potential for

shorter or more targeted data collection periods that could reduce participant or equipment burden.

**Index Terms**—Free-living gait, inertial sensor, motion capture, foot biomechanics, consistent data, osteoarthritis.

## I. INTRODUCTION

**I**N-LABORATORY, optical motion capture has long been regarded as the gold-standard method to measure biomechanical variables related to gait. Though an accurate and powerful tool, the inherent limitations of these systems and environments may limit the external validity of the data [1]. This is partly a result of the non-typical walking surfaces, unfamiliar conditions, and the short, linear walking bouts associated with laboratory gait analysis.

In recent years, the use of body-worn inertial measurement units (IMUs) has opened the doors for examining movement in free-living settings, outside the laboratory. Though the use of IMUs to measure human motion is far from new, nearing 3 decades of maturity [2], the exponential increase in their use is largely due to the improvements in accessibility, size, and the cost of IMU systems [3]. The field of gait biomechanics continues to embrace this rapidly developing technology [4], now using it to record vast amounts of data over extended timeframes, both inside and outside of the traditional laboratory setting [5]. This is also true for clinical populations, such as those with knee osteoarthritis [6], where longitudinal gait monitoring could provide important insights into functional aspects of the disease or monitoring the progress of a gait-related rehabilitation program.

Due to the wearability of IMU systems, it is far easier to collect many thousands of steps spread over days or weeks, compared to typical laboratory-based collections which usually consist of less than 10 steps of over-ground walking, or up to several hundred when a treadmill is used [3]. Longitudinal, real-world IMU datasets have great potential to provide rich information on an individual's gait biomechanics in natural environments, capturing the inherent variability associated with real-world gait. With that, we can characterize a gait pattern that represents the persons walking in normal daily conditions while capturing day-to-day fluctuations. Since indefinite data collections are generally not possible, limited by factors such as patient burden, data storage, and device battery capacity, it is important to establish when enough data is collected. One approach to determining this is identifying

Manuscript received 15 September 2022; revised 21 November 2022 and 1 February 2023; accepted 16 February 2023. Date of publication 1 March 2023; date of current version 8 March 2023. This work was supported by the Natural Sciences and Engineering Research Council of Canada under Grant RGPIN-2021-02484. The work of Jesse M. Charlton was supported in part by the Canadian Institutes of Health Research Doctoral Research Award and in part by the University of British Columbia 4-Year Fellowship. The work of Calvin Kuo was supported by the University of British Columbia Killam Postdoctoral Award. The work of Michael A. Hunt was supported in part by the Canadian Institutes of Health Research and in part by the Michael Smith Health Research British Columbia. (*Corresponding author: Jesse M. Charlton.*)

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the University of British Columbia Clinical Research Ethics Board under Application No. H19-02323.

Jesse M. Charlton is with the School of Biomedical Engineering and the School of Kinesiology, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: jchar32@alum.ubc.ca).

Calvin Kuo is with the School of Biomedical Engineering, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: calvin.kuo@ubc).

Michael A. Hunt is with the Department of Physical Therapy, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: michael.hunt@ubc.ca).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TNSRE.2023.3250612>, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2023.3250612

how much data must be recorded such that the outcome of interest for a given individual is consistent (i.e., no longer changes when more datapoints are added). In the present study, we conceptually defined a consistent outcome in this way, where adding more data to a dataset does not appreciably change the distribution of an outcome. Having a guideline of this nature could assist with lowering the burden on study participants, researchers and the equipment involved in real-world data collections by reducing overcollection of data.

Several studies have sought to identify the length of gait assessment recordings to establish consistent outcomes. In the context of gait variability research, treadmill walking has been shown to require 400 steps before accurate estimates of step time or step width variability could be made [7]. Another study assessed over-ground walking both in a hallway and over 48 hours of unsupervised activity using a waist mounted IMU [8]. The authors suggested that several days of recording were required to establish consistent gait data, but further details were not provided. More recently, Benson et al. [9] identified the number of individual outdoor running bouts that were needed to observe consistent outcomes from a waist-mounted IMU (e.g., cadence, pelvis kinematics). The results indicated that four running bouts (~15-20 mins each, equating to approximately 2550 steps per bout at a cadence of 170) was enough. While these studies suggest that consistent gait outcomes can be identified, the applicability of the results to real-world, unsupervised walking remains unclear.

The influence of variability (or lack of) arising from person-related factors could affect the amount of data that is needed in these real-world data collections. Unfortunately, none of these studies considered the effects of clinical populations who may have altered gait variability, potentially requiring different sized datasets from health counterparts. Knee osteoarthritis is particularly known to affect various aspects of gait biomechanics including kinematic, kinetic, and variability outcomes, which can differ based on the severity of the disease [10], [11] or with respect to healthy adults [12], [13], [14]. This population could act as a useful model to identify potential differences in data collection requirements when gait-influencing musculoskeletal disease is present in one's population of interest.

Placing the IMU on the low back is most common in the extant literature that has examined gait feature validity and reliability [15], [16]. However, placement on other body segments may have more relevance for certain populations or clinical applications. For example, measuring or modifying the orientation of the foot while walking and running has relevance to clinical rehabilitation of knee pathologies like osteoarthritis or patellofemoral pain [17], [18]. Foot kinematics at the start and end of the stance phase also have relevance, both with respect to quantifying impacts [19] or foot trajectories, which may be influenced by variable terrain underfoot [21]. Moreover, IMUs located on the foot result in very reliable spatiotemporal outcomes, and in some cases can outperform IMUs mounted on the low-back in laboratory conditions [15]. Therefore, our study objective was to determine the number of steps that need to be collected to establish consistent gait outcomes from unsupervised, real-world and purposeful

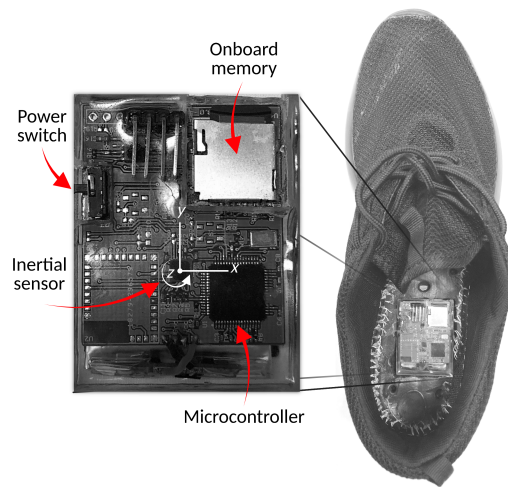


Fig. 1. Inertial sensor module embedded in the sole of a generic walking shoe. The entire module was encased in epoxy resin for durability. A relief was cut into the sole such that the sensor press fit tightly within and was flush with the inside of the shoe (illustrated by cut-away image). Created with BioRender.

walking in populations with and without knee osteoarthritis, using a shoe-embedded inertial sensor.

## II. METHODS

Two distinct participant samples were recruited for the present study: a group with symptomatic knee osteoarthritis (KOA,  $n=15$ ) and a healthy adult group (HA,  $n=15$ ). The KOA group was included to examine the effects of symptomatic musculoskeletal disease on the consistency of gait outcomes. All participants were required to be 19 years of age or older, able to walk for 30 minutes, and fit into shoe sizes between women's US5 and men's US13. Further eligibility for the HA group required them to be free of pain, injury, or surgery in the lower extremities in the last 12 months. Inclusion in the KOA group required that the participants were  $\geq 50$  years of age, exhibited radiographic evidence of predominantly medial compartment knee osteoarthritis (Kellgren & Lawrence grade 2 or greater) [22], and self-reported average knee pain over the past week ( $\geq 1/10$ ) on an 11-point numerical rating scale with anchors of 0 being "no pain at all" and 10 being the "worst pain imaginable". Any individual who required a gait aid, had a history of knee surgery or joint replacement, had recent (within past 6 months) corticosteroid injections, inflammatory arthritis, were on a waitlist for knee joint replacement, or had any other condition that affected normal gait were ineligible to participate. Informed consent was obtained prior to data collection, and the study was approved by the Institutional Clinical Research Ethics Board (H19-02323).

The data in this study were collected using a custom designed IMU that was embedded in the sole of an athletics shoe [23], hereafter called the sensorized shoe (Fig 1). The module consisted of a magneto-inertial sensor (MPU-9150, InvenSense, CA, USA) equipped with a 3-axis accelerometer (AC; signal range:  $\pm 4g$ ), 3-axis gyroscope (GYR; signal range:  $\pm 500^\circ/s$ ), and 3-axis magnetometer (signal range:  $\pm 1200\mu T$ ). A microprocessor (STMicroelectronics,

STM32F401; clock rate=84 MHz, Switzerland) sampled the sensor data at 100Hz. The sensor's y-axis axis was aligned with the long axis of the shoe (mid point of the heel counter to the midpoint of the toe box), the x-axis was aligned to point mediolaterally, and the z-axis pointed in a downward direction following the right-hand rule. The data were recorded to an onboard 8 gigabyte microSD (microSDHC Class 4, Transcend, China) and the Qi-wireless charging equipped lithium-ion battery allowed for up to 16 hours of continuous data collection. An on-off switch was manually operated to power the device before and after a walking bout.

Seven days of unsupervised, real-world walking were collected using the sensorized shoes on the dominant limb (HA group) or the affected (unilateral osteoarthritis) or most symptomatic (bilateral osteoarthritis) limb (KOA group). The participants were asked to walk in their usual manner and at their preferred pace during all collections. Instructions were to walk with the sensorized shoes for a minimum of 20 minutes per day during their normal walking activity and in their normal walking environments. However, more walking was permitted if desired. We specifically requested that they not wear the shoes during short and sporadic walking, such as doing chores around the home or while doing office work, but instead to only wear the shoes during extended bouts of walking, such as walking exercise, walking to the store, or similar.

The datasets extracted from the seven days of real-world walking were processed using custom scripts (MATLAB, v2021b, Mathworks Inc., Natick USA). First, the datasets were classified to identify data that corresponded with walking and non-walking activity. To do this we split the datasets into segments based on the mid-swing foot pitch peak. A single data segment from a known bout of over-ground walking (collected from a walking bout in the hallways adjacent to the laboratory) was used as a representative walking segment and was compared to all other segments by a cross-correlation algorithm. Briefly, the signal energy of the Euclidean norm acceleration signal (squared magnitude of the signal) was calculated for every data segment, and each signal was cross correlated with the representative walking segment's signal. For every data segment, the percent difference between the signal energy and the maximum cross correlation was calculated. All segments with a percent difference  $\leq 27\%$  were retained as walking segments, while all others were removed from the dataset. The 27% difference threshold was the optimal value for separating walking from non-walking data based on our development of the classification [24] using outdoor, linear walking data from a previously published study [25] (See Supplementary File 1 for details).

Next, we identified gait events within each walking segment based on the most accurate signal variables from a previously published automated process [26]. Each segment spanned mid-swing to mid-swing; therefore, the heel-strike (HS), start of foot flat, mid-foot flat, end of foot flat, and toe-off (TO) were calculated. Each data segment contained a single stance phase (described as a "step" hereafter), from which we calculated seven gait-related outcomes detailed in Table I.

TABLE I  
EXTRACTED GAIT OUTCOME DEFINITIONS

Outcome	Time	ABBR.	Unit	Definition
Peak Acceleration Norm	Heel strike	AC-HS	m/s <sup>2</sup>	The maximum value of the Euclidean norm of acceleration at $\pm 5$ frames around gait event
	Toe off	AC-TO		
Peak Angular Rate Norm	Heel strike	GYR-HS	°/s	The maximum value of the Euclidean norm of angular rate at $\pm 5$ frames around gait event
	Toe-off	GYR-TO		
Stance Time	Heel strike to Toe-off	ST	s	Time from heel strike to toe-off, the stance phase of gait
Foot Strike Angle	Heel strike	FSA	°	Sagittal plane angle of the foot in the inertial frame of reference (+ values indicate the toe is higher than the heel)
Foot Progression Angle	15-50% of stance	FPA	°	Angle of the foot in the horizontal plane relative to the direction of walking (+ values indicate the toe is more medial than the heel, or toe-in)

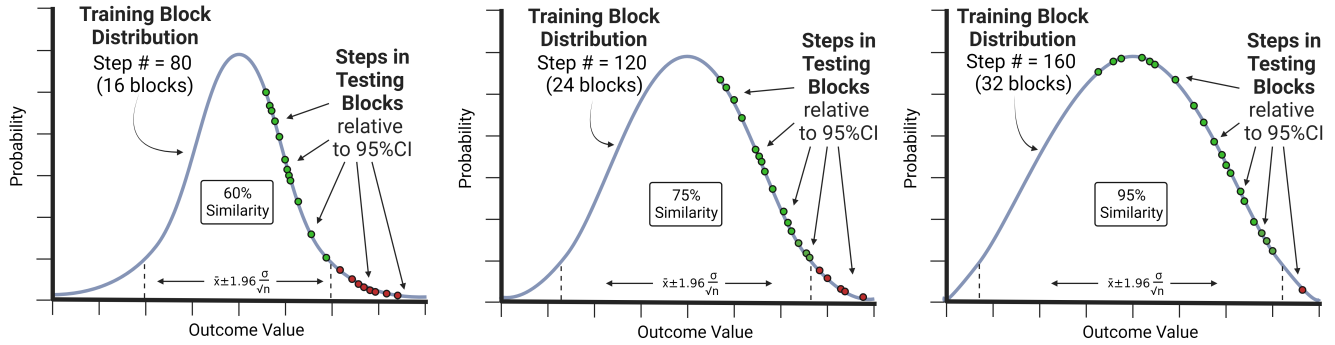
We adapted methods from previous research [9] to identify when an outcome was consistent, where adding more walking data would not change the outcome's distribution. This consistency point was calculated by comparing the similarity between a training block (comprising an increasing number of steps used to form the training block) with all other unique sets of steps (testing blocks) that were not in the training block. A univariate normal distribution was fit to the training block containing the data from a single gait outcome and the 95% confidence intervals (CI) were calculated. The percent similarity was calculated based on the proportion of testing block steps with an outcome value that was within the 95% CI bounds of the training block (Fig 2).

Testing blocks were constructed by grouping all a participant's steps into unique blocks of five. For example, a participant with  $n=100$  steps would have 20 ( $n \div 5$ ) testing blocks. Larger and larger training blocks were then constructed by combining testing blocks in increasing amounts (training block 1 = 5 steps, training block 2 = 10 steps, etc.). For a given training block size, the mean percent similarity score was calculated for every unique combination between the training and testing blocks. This was repeated for every training block size, generating a mean percent similarity score for each training block increment.

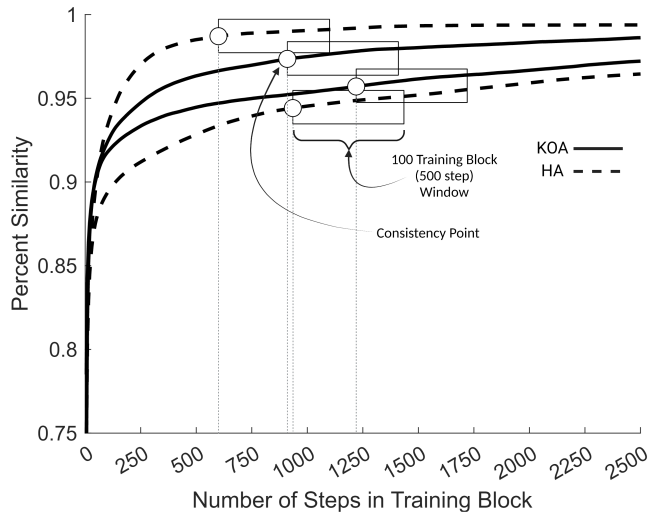
To define when an outcome was consistent, we adapted definitions and recommendations from previous research. Benson et al. [9] examined several running-related outcomes collected across 10 running bouts (approximately 2550 steps per bout) and defined this consistency point (which the authors termed "stability") as the minimum number of running bouts in which adding another running bout to the training block did not change the percent similarity of that block more than 5%.



### Example Percent Similarity comparisons between training and testing blocks



**Fig. 2.** Illustrating the training block distributions, 95% CI and relation to testing block step comparisons. As the number of steps within the training block increases, a greater percent of the testing step outcomes are within the 95% CI resulting in a larger percent similarity value. For each training block size, comparisons were made to all testing block steps not in the training block. The figure depicts an example of what an increase in the training block size may do to the similarity between the distribution and testing block step outcomes. Created with BioRender.



**Fig. 3.** Illustrating the consistency point criteria from four representative participants, two from the HA group (dashed) and two from the KOA group (solid), for one gait outcome. The rectangle represents the moving window of 100 training blocks (equivalent to 500 steps, 100 blocks  $\times$  5 steps per block). The circle identifies the first training block of the moving window when the percent similarity change is  $<0.01\%$  for the entirety of the 100 training block moving window. Created with BioRender.

For the present study, we defined the consistency point when a moving window of 100 consecutive training blocks (equivalent to 500 steps) had a percent similarity slope that was  $<0.01\%$  to ensure the percent similarity had plateaued (Fig 3). The small slope threshold in our definition reflects the much smaller data size increment (5 steps) compared to the running bouts in previous analyses ( $>2000$  steps) [9]. Finally, we recorded the number of steps in the training block when this threshold was met. As the resulting dataset of step counts was right-censored due to some participant-outcome combinations not reaching the consistency point, we accounted for this by replacing these censored data with the participants total number of steps recorded plus 5 (five was used as the resolution for step counts was in intervals of five). This uncensored dataset was used in all subsequent analyses.

Our primary analysis was calculating the median (IQR; interquartile ranges) of the consistency points for each study group and gait outcome, and the percentage similarity at each consistency point. However, we were secondarily interested in statistical differences between the groups and gait outcomes. Therefore, we fit generalized linear mixed effects models using a negative binomial distribution and log link function (Poisson modelling exhibited poor fit and overdispersion). Fixed effects included the group (2 levels), the gait outcomes (7 levels), and their interaction, while a random intercept on the participant variable was included. The AC-HS outcome and the HA group were set as the referent levels for the models. The Wald  $\chi^2$  statistics were used to examine the main effects (group, gait outcomes, interaction) for significance and the model parameters were individually assessed using Wald's Z statistic (see Supplementary File 2 for model parameters results). A significant main effect prompted examination of contrasts among the levels of that effect. We adjusted these contrasts for multiplicity using a Bonferroni correction. The number of steps to the consistency point for each outcome and group are reported as the raw step count values. Alpha was set at 0.05 for all tests (Bonferroni corrected alpha = 0.0024 for the gait outcome contrasts). Statistical analyses were completed in R stats (v. 3.6.0) [27], [28], [29].

### III. RESULTS

The real-world walking datasets consisted of median (min, max) 9159 (4180, 24495) total steps over the seven-day assessment period; the HA group recorded 10964 (4180, 23312) steps, while the KOA group recorded 9156 (5216, 24495) steps. The HA group was a mean (SD) 29.5 (8.2) years of age, 1.72 (0.09) m tall, had a body mass index of 23.8 (2.7)  $\text{kg}/\text{m}^2$ , and was comprised of 10 (67%) males. The KOA group was 67.4 (6.9) years of age, 1.67 (0.09) m tall, had a body mass index of 26.7 (8.2)  $\text{kg}/\text{m}^2$ , and had 4 (27%) males. There were nine with mild, four with moderate, and two with severe structural knee osteoarthritis and they reported average knee pain while walking over the week prior to data collection equal to 4 (2), where 0 represented no pain and 10 represented the



**TABLE II**  
PARTICIPANT MEAN (SD) GAIT OUTCOME VARIABLES BY GROUP

Outcome	Overall, N = 30	HA, n = 15	KOA, n = 15
AC-HS (m/s <sup>2</sup> )	43.7 (6.7)	45.7 (5.6)	41.7 (7.3)
AC-TO (m/s <sup>2</sup> )	28.1 (4.6)	29.7 (4.4)	26.6 (4.3)
GYR-HS (°/s)	480.3 (53.9)	492.2 (46.0)	468.4 (59.9)
GYR-TO (°/s)	504.7 (23.1)	509.5 (20.9)	499.9 (24.8)
ST (s)	0.65 (0.05)	0.64 (0.03)	0.67 (0.06)
FSA (°)	21.5 (4.7)	22.9 (2.8)	20.1 (5.9)
FPA (°)	-9.2 (5.8)	-9.1 (5.7)	-9.2 (6.1)

Abbreviations: AC-HS, peak Euclidean norm of acceleration at heel strike; AC-TO, peak Euclidean norm of acceleration at toe-off; GYR-HS, peak Euclidean norm of angular velocity at heel strike; GYR-TO, peak Euclidean norm of angular velocity at toe-off; ST, stance time; FSA, foot strike angle; FPA, foot progression angle.

worst pain imaginable. The gait outcome values are reported in Table II for reference, though no comparisons between groups were performed on these data.

The number of steps to the consistency point are reported in Table III and illustrated in Fig 4 for all 30 participants split by group and gait outcome. Across all gait outcomes and participants, the mean (SD) of the percent similarity at the consistency point was 93.9% (2.8), 93.7% (2.6), and 93.8% (3.0) for all participants, the HA group, and the KOA group, respectively. Not all participant-outcome combinations were consistent enough to meet our criteria, and the number of individuals that did not reach the consistency point differed between gait outcomes, with the median number of participants not reaching consistency being 8 (AC-HS=11, AC-TO=9, GYR-HS=9, GYR-TO=4, ST=8, FSA=1, FPA=0).

The number of steps needed for consistent outcomes did not differ between groups ( $\chi^2 = 0.098$ ,  $df=1$ ,  $p=0.754$ ), but there were significant differences between the gait outcomes ( $\chi^2 = 52.42$ ,  $df=6$ ,  $p<0.001$ ). Across both groups and all outcomes of interest, the median (IQR) number of steps to reach the consistency point was 1140 (725, 3230), while the minimum and maximum were 250 and 23237, respectively.

Contrasts between the gait outcomes ( $\alpha=0.0024$ ) highlighted that the FSA and FPA outcomes required significantly lower steps to reach the consistency point compared to all other outcomes ( $p<0.001$ ) except for the GYR-TO outcome (FSA:  $p=0.006$ , FPA:  $p=0.206$ ). All contrast comparisons are reported in Supplementary File 2 Table II. The censored model results and contrasts can be found in Supplementary File 3.

#### IV. DISCUSSION

The collection of real-world, unsupervised data is becoming more viable and common place in the field of gait biomechanics. While it is an important tool to advance the field and our understanding of gait, there are many methods and measurement questions that still need addressing. Our study

estimated the number of steps required to observe consistent outcomes when using a shoe-embedded IMU to record real-world walking. Our primary finding suggests that the number of steps during purposeful walking required to see consistency differed based on the outcome being measured. The median number of steps to obtain consistent data was between 615 and 3720 when looking across all gait outcomes. Assuming a cadence of 58 steps per minute for a given limb based on data from adults with and without knee osteoarthritis [30], a data collection need only comprise anywhere between 11 and 65 minutes of walking, depending on the outcome of interest. This is well within feasible data collection bounds, especially in the context of unsupervised, real-world collection spread over days or weeks. These findings can guide study design and potentially minimize the over-collection of data that could be problematic for clinical populations who may have limited mobility or capacity to walk due to pain. Notably, this analysis does not replace a traditional sample size estimate which is bounded by the particular statistical modelling approach that would be used in the study. Instead, these consistency estimates are complementary to sample size estimates and should be considered before new data collection protocols are conducted.

The amount of data needed to see consistent outcomes differed based on the underlying signal. We did not observe significant differences among the AC, GYR, and ST outcomes, all requiring more steps to reach consistency than the FSA and FPA outcomes. All of these outcomes had individual participants not reach a consistency point, leading to higher median step count estimates, larger interquartile ranges, and ultimately less ability to detect differences. A larger sample size, both in population and in the amount of gait data collected, would be necessary to more accurately quantify the number of steps to reach consistency in these outcomes. The GYR-TO outcome only had 4 individuals who did not reach consistency while also exhibiting a potentially (not statistically significant) lower median step count. It is possible that the differences in rotational and linear motion of the foot across the gait cycle can result due to variable terrain and features of the built world. The terrain underfoot, such as a sidewalk, gravel, or dirt trail can lead to differences in foot trajectories in the horizontal and vertical planes [21], while walking direction changes can affect foot trajectories, as well [31], [32]. Ultimately, gathering enough data to characterize a participant's foot dynamics may demand different amounts of walking data depending on the outcome, and significantly more than measures of foot orientation.

Foot segment orientation outcomes readily met the consistency criteria we set in this study. The FSA was consistent with a median of 662.5 steps across all participants, and only 1 participant did not reach our threshold. This suggests that the foot pitch angle at HS is consistently patterned over time despite variable terrain conditions. As our data were specific to walking, it is unclear if these results would transfer to running contexts where FSA may have more relevance [33]. While more steps were needed to observe consistent FPA (985 steps) data compared to the FSA, every participant reached the threshold. The higher relative variability in FPA

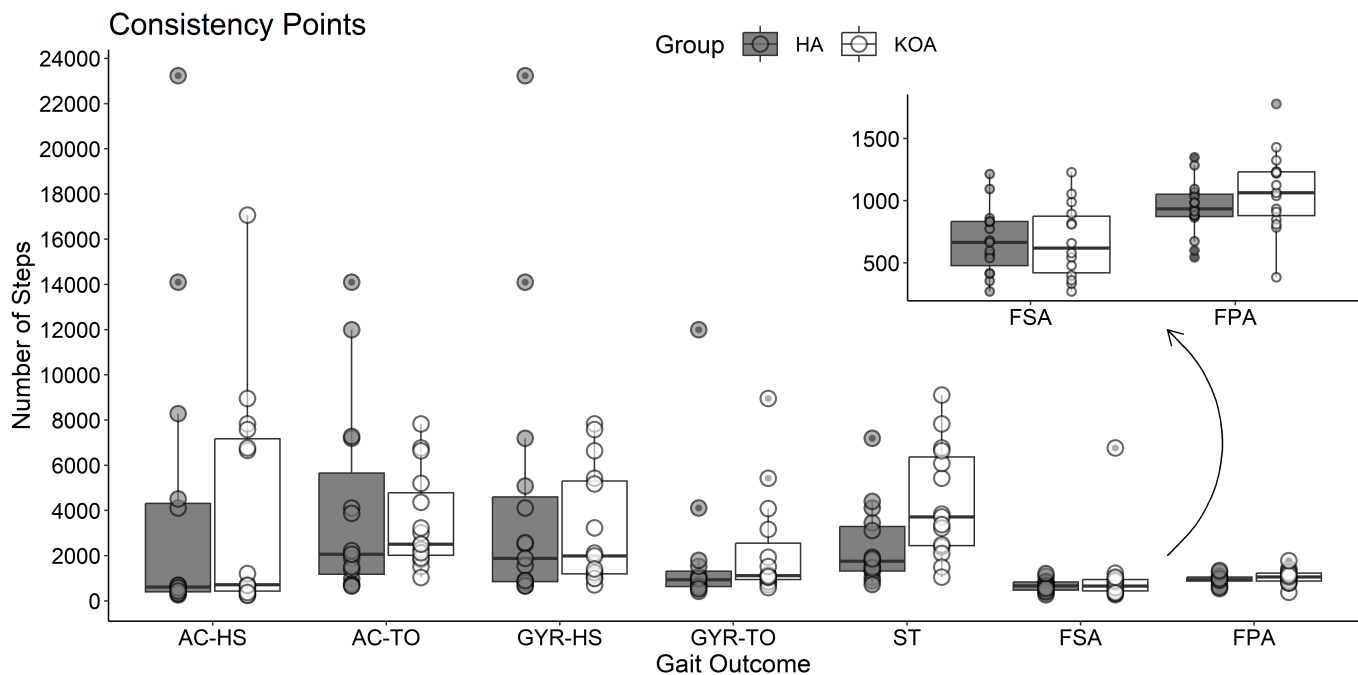


Fig. 4. The number of steps to observe consistent data for each group and gait outcome combination. Circles represent individual participant data while the box and whisker plots indicate the median (horizontal line), interquartile range (box) and 150% of the interquartile range (whisker). Circles with dots within indicate data beyond the 150% of the interquartile range.

magnitudes (Table II) relative to the other gait outcomes may be a direct product of walking direction changes that can alter FPA magnitude on a step-to-step basis [32]. Despite this variability, it appears the FPA follows a regular within-subject pattern that is more consistent than the linear acceleration, angular velocity, or spatiotemporal outcomes we examined.

In populations with knee osteoarthritis, altering the FPA is being investigated as a rehabilitation strategy to improve the loading environment within the knee joint [17]. Most of the research to date has examined the FPA in laboratory settings though, limiting our knowledge of how changes to the FPA are integrated into daily walking. Based on our analysis, consistent FPA required 985 steps, which equates to approximately 17 minutes of walking with an ipsilateral cadence of 58 steps per minute. Considering that typical optical motion capture-based laboratory data collections can exceed this amount of collection time in total, when including participant preparation and calibration (despite only recording less than 10 steps for analysis), collecting this amount of data in real-world settings using IMUs is certainly feasible. However, not all data that are recorded during unsupervised data collections will be walking, so we suggest using the estimates from the present study as minimum guidelines.

The influence knee osteoarthritis was statistically inconsequential in our results. While there was not a main effect for group, there did appear to be more steps to see consistent stance time outcomes for the KOA group. Previous investigations have shown conflicting evidence for differences in the variability of spatiotemporal and kinematic gait outcomes between groups with and without knee osteoarthritis, or among groups with knee osteoarthritis [11], [12], [13], [33].

However, these studies were conducted over one laboratory testing session which cannot capture daily fluctuations in osteoarthritis disease or gait. If the fluctuations are separated in time by a large amount, more steps over longer timeframes may be needed to see consistent variables in this population. Generally, our results indicate that the time commitment for collecting consistent data are not appreciably different between people with and without knee osteoarthritis. This supports the generalizability of our findings beyond healthy adults.

The results of our study are limited in several ways. First, we conducted this analysis on healthy adults and people living with knee osteoarthritis, which may exhibit gait characteristics that are not generalizable to other populations (e.g., children, people living with other musculoskeletal or neurological pathologies). Further, it is likely that our data arose from periods of time (or individuals) who had less severe symptoms overall given the requirements of the walking study. Second, our results are specific to foot mounted IMUs, which likely limits generalizability to IMU signals gathered from other body segments or those that are derived from multiple sensors (e.g., joint angles). It should be noted that there may have been clipping of the vertical acceleration signals at heel strike due to the sensor dynamic range, though an exploratory post-hoc analysis found no observable effects on the AC-HS results. Third, our sensorized shoe design involved fitting the IMU into a relief that was cut into the shoe sole. This ensured the IMU was correctly and consistently positioned each time, but meant participants had to walk in the shoes we provided. Externally mounted IMU systems could allow for walking in participants' preferred shoes, but this would likely result in greater movement artefact and inherent variability

TABLE III  
NUMBER OF STEPS TO OBSERVE CONSISTENT GAIT OUTCOMES

Outcome	Overall, N=30		HA Group, n=15		KOA Group, n=15	
	Reached consistency n (%)	Median (IQR)	Reached consistency n (%)	Median (IQR)	Reached consistency n (%)	Median (IQR)
AC-HS (m/s <sup>2</sup> )	19 (63)	690 (391, 6738)	10 (67)	615 (400, 4314)	9 (60)	720 (432, 7172)
AC-TO (m/s <sup>2</sup> )	21 (70)	2390 (1598, 4990)	10 (67)	2065 (1182, 5656)	11 (73)	2510 (2015, 4782)
GYR-HS (°/s)	21 (70)	1935 (995, 5156)	10 (67)	1880 (848, 4600)	11 (73)	1990 (1195, 5306)
GYR-TO (°/s)	26 (87)	1060 (722, 1732)	13 (87)	935 (638, 1310)	13 (87)	1120 (948, 2550)
ST (s)	22 (73)	2800 (1504, 4329)	12 (80)	1755 (1322, 3288)	10 (67)	3720 (2440, 6364)
FSA (°)	29 (97)	663 (431, 854)	15 (100)	665 (478, 832)	14 (93)	660 (440, 942)
FPA (°)	30 (100)	985 (869, 1196)	15 (100)	935 (872, 1052)	15 (100)	1065 (880, 1232)

Median (interquartile range; IQR) for the number of steps to reach the consistency point and the number of participants who did so as n (%), split by gait outcome and group. Abbreviations: AC-HS, peak Euclidean norm of acceleration at heel strike; AC-TO, peak Euclidean norm of acceleration at toe-off; GYR-HS, peak Euclidean norm of angular velocity at heel strike; GYR-TO, peak Euclidean norm of angular velocity at toe-off; ST, stance time; FSA, foot strike angle; FPA, foot progression angle.

which may lead to different results than we observed. Fourth, we did not parse our walking data into different terrain conditions (level ground, inclines, declines, stairs etc.), nor did we collect other external data (e.g., weather, obstacles, urban vs rural settings etc.) or internal factors (participant affect, occupation etc.) which may impact walking behaviour. Doing so could capture a broader range of walking conditions but may lead to observing consistent outcomes more readily than if the analysis was limited to specific terrain/activity conditions and consistent external/internal factors. However, limiting the conditions in this way could support more direct comparisons between studies in the future. Fifth, our data arose from unsupervised, outdoor walking for longer periods of time, which may not generalize to short, sporadic walking bouts. Finally, not all participants reached our definition of consistency for each outcome. This may have resulted from people walking in distinctly different environmental conditions across the week of data collection. Restricting analyses to specific terrain, as noted above, may reduce this issue in future studies.

Based on our findings and the above-mentioned limitations, there are several areas in need of further examination. In order to capture the upper end of step counts needed to gather consistent data in some participants, we need to collect larger datasets over greater timeframes (beyond 7 days). This should also be done across a more representative population of people with knee osteoarthritis, comprising the spectrum of clinical and structural severity. Comparisons with healthy, age-matched individuals will strengthen our disease-specific understanding as well. Finally, with larger datasets there will be support for appropriately powered subgroup analyses that can identify who, why, and to what extent some people are more likely to require larger data collection periods to capture consistent data.

## V. CONCLUSION

Sensor-based gait analysis is quickly becoming a popular method to collect gait data. With this, real-world, unsupervised gait monitoring is becoming increasingly feasible, providing

researchers and clinicians with diverse and highly relevant datasets. However, guidelines regarding how data should be collected are lacking. To our knowledge, this is the first study to estimate the amount of data needed in real-world, unsupervised walking research to see consistent gait-related outcomes arising from populations with and without knee osteoarthritis. Our results support the rapidly expanding research space related to disease- and rehabilitation-specific gait biomechanics in this population, and the use of IMUs to conduct data collections outside the laboratory [6]. Researchers should consider what outcomes they wish to investigate in these highly variable walking environments, as the amount of data needed differed by outcome. Overall, our results are promising in that the number of steps to see consistent outcomes were well within feasible data collection timeframes, though the number of participants who did not reach consistency suggests more work is necessary. This ultimately supports continued research using IMUs to measure gait dynamics in real-world settings, which can support the capture of long-term and highly relevant gait data in healthy and clinical populations.

## APPENDIX

Supplementary File 1: Gait signal classification algorithm

Supplementary File 2: Model results

Supplementary File 3: Censored data model results

Supplementary File 4: Censored data median step counts

The study data and statistical analysis are available at: <https://doi.org/10.5683/SP3/PZKWTB>

## REFERENCES

- [1] I. Hillel et al., "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. Activity*, vol. 16, no. 1, p. 6, Dec. 2019.
- [2] P. Picerno, "25 years of lower limb joint kinematics by using inertial and magnetic sensors: A review of methodological approaches," *Gait Posture*, vol. 51, pp. 239–246, Jan. 2017.
- [3] P. B. Shull, W. Jirattigalachote, M. A. Hunt, M. R. Cutkosky, and S. L. Delp, "Quantified self and human movement: A review on the clinical impact of wearable sensing and feedback for gait analysis and intervention," *Gait Posture*, vol. 40, no. 1, pp. 11–19, 2014.



- [4] A. Muro-De-La-Herran, B. Garcia-Zapirain, and A. Mendez-Zorrilla, "Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications," *Sensors*, vol. 14, no. 2, pp. 3362–3394, 2014.
- [5] L. C. Benson, C. A. Clermont, E. Bošnjak, and R. Ferber, "The use of wearable devices for walking and running gait analysis outside of the lab: A systematic review," *Gait Posture*, vol. 63, pp. 124–138, Jun. 2018.
- [6] D. Kobsar et al., "Wearable inertial sensors for gait analysis in adults with osteoarthritis—A scoping review," *Sensors*, vol. 20, no. 24, p. 7143, Dec. 2020.
- [7] T. M. Owings and M. D. Grabiner, "Measuring step kinematic variability on an instrumented treadmill: How many steps are enough?" *J. Biomechanics*, vol. 36, no. 8, pp. 1215–1218, Aug. 2003.
- [8] G. Cola, M. Avvenuti, A. Vecchio, G.-Z. Yang, and B. Lo, "An on-node processing approach for anomaly detection in gait," *IEEE Sensors J.*, vol. 15, no. 11, pp. 6640–6649, Nov. 2015.
- [9] L. C. Benson, N. U. Ahamed, D. Kobsar, and R. Ferber, "New considerations for collecting biomechanical data using wearable sensors: Number of level runs to define a stable running pattern with a single IMU," *J. Biomechanics*, vol. 85, pp. 187–192, Mar. 2019.
- [10] J. L. Astephen, K. J. Deluzio, G. E. Caldwell, and M. J. Dunbar, "Biomechanical changes at the hip, knee, and ankle joints during gait are associated with knee osteoarthritis severity," *J. Orthopaedic Res.*, vol. 26, no. 3, pp. 332–341, Mar. 2008.
- [11] R. M. Kiss, "Effect of severity of knee osteoarthritis on the variability of gait parameters," *J. Electromyogr. Kinesiol.*, vol. 21, no. 5, pp. 695–703, Oct. 2011.
- [12] C. A. Clermont and J. M. Barden, "Accelerometer-based determination of gait variability in older adults with knee osteoarthritis," *Gait Posture*, vol. 50, pp. 126–130, Oct. 2016.
- [13] J. H. Park, H. Lee, J.-S. Cho, I. Kim, J. Lee, and S. H. Jang, "Effects of knee osteoarthritis severity on inter-joint coordination and gait variability as measured by hip-knee cyclograms," *Sci. Rep.*, vol. 11, no. 1, p. 1789, Jan. 2021.
- [14] K. Mills, M. A. Hunt, and R. Ferber, "Biomechanical deviations during level walking associated with knee osteoarthritis: A systematic review and meta-analysis," *Arthritis Care Res.*, vol. 65, no. 10, pp. 1643–1665, Oct. 2013.
- [15] D. Kobsar et al., "Validity and reliability of wearable inertial sensors in healthy adult walking: A systematic review and meta-analysis," *J. NeuroEngineering Rehabil.*, vol. 17, no. 1, pp. 1–21, May 2020.
- [16] F. A. Storm, A. Cesareo, G. Reni, and E. Biffi, "Wearable inertial sensors to assess gait during the 6-minute walk test: A systematic review," *Sensors*, vol. 20, no. 9, p. 2660, May 2020.
- [17] R. Richards, J. C. van den Noort, J. Dekker, and J. Harlaar, "Gait retraining with real-time biofeedback to reduce knee adduction moment: Systematic review of effects and methods used," *Arch. Phys. Med. Rehabil.*, vol. 98, no. 1, pp. 137–150, Jan. 2017.
- [18] J. L. Roper et al., "The effects of gait retraining in runners with patellofemoral pain: A randomized trial," *Clin. Biomechanics*, vol. 35, pp. 14–22, Jun. 2016.
- [19] I. Eitzen, J. Renberg, and H. Førevik, "The use of wearable sensor technology to detect shock impacts in sports and occupational terrains: A scoping review," *Sensors*, vol. 21, no. 15, p. 4962, Jul. 2021.
- [20] D. B. Kowalsky, J. R. Rebula, L. V. Ojeda, P. G. Adamczyk, and A. D. Kuo, "Human walking in the real world: Interactions between terrain type, gait parameters, and energy expenditure," *PLoS ONE*, vol. 16, no. 1, Jan. 2021, Art. no. e0228682.
- [21] J. H. Kellgren and J. S. Lawrence, "Radiological assessment of osteoarthritis," *Ann. Rheumatic Diseases*, vol. 16, no. 4, pp. 494–502, Dec. 1957.
- [22] H. Xia, J. Xu, J. Wang, M. A. Hunt, and P. B. Shull, "Validation of a smart shoe for estimating foot progression angle during walking gait," *J. Biomechanics*, vol. 61, pp. 193–198, Aug. 2017.
- [23] J. Charlton, H. Xia, P. Shull, J. J. Eng, L. C. Li, and M. A. Hunt, "Multi-day monitoring of foot progression angles during unsupervised, real-world walking in people with and without knee osteoarthritis," *SSRN Electron. J.*, pp. 1–27, 2022, doi: [10.2139/ssrn.4173272](https://doi.org/10.2139/ssrn.4173272).
- [24] J. M. Charlton, H. Xia, P. B. Shull, and M. A. Hunt, "Validity and reliability of a shoe-embedded sensor module for measuring foot progression angle during over-ground walking," *J. Biomechanics*, vol. 89, pp. 123–127, May 2019.
- [25] B. Mariani, H. Rouhani, X. Crevoisier, and K. Aminian, "Quantitative estimation of foot-flat and stance phase of gait using foot-worn inertial sensors," *Gait posture*, vol. 37, no. 2, pp. 229–234, 2013.
- [26] *R: A Language and Environment for Statistical Computing*, RC Team, R Foundation for Statistical Computing, Vienna, Austria, 2019.
- [27] F. Hartig, "DHARMA: Residual diagnostics for hierarchical (multi-level/mixed) regression models," Regensburg, Germany, Tech. Rep., 2021.
- [28] M. E. Brooks et al., "glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling," *R J.*, vol. 9, no. 2, pp. 378–400, 2017, 2017.
- [29] J. F. Hafer, S. G. Provenzano, K. L. Kern, C. E. Agresta, J. A. Grant, and R. F. Zernicke, "Measuring markers of aging and knee osteoarthritis gait using inertial measurement units," *J. Biomechanics*, vol. 99, Jan. 2020, Art. no. 109567.
- [30] G. Courtine and M. Schieppati, "Human walking along a curved path. II. Gait features and EMG patterns," *Eur. J. Neurosci.*, vol. 18, no. 1, pp. 191–205, Jul. 2003.
- [31] G. Courtine and M. Schieppati, "Human walking along a curved path. I. Body trajectory, segment orientation and the effect of vision," *Eur. J. Neurosci.*, vol. 18, no. 1, pp. 177–190, Jul. 2003.
- [32] M. O. Almeida, I. S. Davis, and A. D. Lopes, "Biomechanical differences of foot-strike patterns during running: A systematic review with meta-analysis," *J. Orthopaedic Sports Phys. Therapy*, vol. 45, no. 10, pp. 738–755, Oct. 2015.
- [33] T. Alkjaer et al., "Gait variability and motor control in people with knee osteoarthritis," *Gait Posture*, vol. 42, no. 4, pp. 479–484, Oct. 2015.