



OPEN

In-Clinic and Natural Gait Observations master protocol (I-CAN-GO) to validate gait using a lumbar accelerometer

Miles Welbourn^{1,3}, Paul Sheriff^{1,3}, Pirinka Georgiev Tuttle¹, Lukas Adamowicz¹, Dimitrios Psaltos¹, Amey Kelekar¹, Jessica Selig¹, Andrew Messere¹, Winnie Mei¹, David Caouette¹, Sana Ghafoor¹, Mar Santamaria¹, Hao Zhang¹, Charmaine Demanuele¹, F. Isik Karahanoglu¹ & Xuemei Cai^{1,2}✉

Traditional measurements of gait are typically performed in clinical or laboratory settings where functional assessments are used to collect episodic data, which may not reflect naturalistic gait and activity patterns. The emergence of digital health technologies has enabled reliable and continuous representation of gait and activity in free-living environments. To provide further evidence for naturalistic gait characterization, we designed a master protocol to validate and evaluate the performance of a method for measuring gait derived from a single lumbar-worn accelerometer with respect to reference methods. This evaluation included distinguishing between participants' self-perceived different gait speed levels, and effects of different floor surfaces such as carpet and tile on walking performance, and performance under different bouts, speed, and duration of walking during a wide range of simulated daily activities. Using data from 20 healthy adult participants, we found different self-paced walking speeds and floor surface effects can be accurately characterized. Furthermore, we showed accurate representation of gait and activity during simulated daily living activities and longer bouts of outside walking. Participants in general found that the devices were comfortable. These results extend our previous validation of the method to more naturalistic setting and increases confidence of implementation at-home.

Gait is the primary means of mobility for most of the population and is a strong indicator of overall health^{1–4}. The free-living 95th percentile stride velocity has been accepted as a primary endpoint by European regulators for patients with Duchenne Muscular Dystrophy (DMD) and has been correlated with existing clinical assessments such as the 6 min walk test (6MWT) Click or tap here to enter text. and the 4-stair climb test (4SC) with the added relevance of monitoring ambulatory function in the real-world⁵. An increase in gait speed over a one-year period is strongly correlated with significant reductions in both relative and absolute risk of death¹. Studies have also indicated that a decline in gait speed is associated with reduced cognition, diminished vitality, and an increased risk of falls and other injurious events⁶. Due to its predictive power in a wide variety of health conditions, gait speed has come to be known as “the sixth vital sign” alongside respiratory rate, heart rate, blood pressure, temperature, and oxygen saturation^{7–9}. Traditional measurements of physical performance, such as the 6MWT and the Short Physical Performance Battery (SPPB) where walking duration is measured with stop watch, are typically performed in clinical or laboratory settings using functional assessment tests and clinical grading scales. Such tests and scales may be unreliable due to their episodic administration, subjectivity of the rater, and the Hawthorne effect^{10,11}. Additionally, in a clinical trial setting, traditional assessments might require multiple clinic visits increasing patient burden and risk of attrition^{12,13}. Thus, there is a need to derive reliable and stable metrics of physical performance, such as gait speed in the 95th percentile and mean gait, in naturalistic environments, as an alternative to traditional assessments^{14,15}. The emergence of digital health technologies (DHTs) has opened opportunities to quantify gait, physical activity, and enabled the continuous monitoring of patients in free-living settings¹⁵. Furthermore, DHTs have demonstrated the potential to reduce the need for multiple clinic visits and patient burden by saving time^{16–18}. One example of DHTs are inertial sensors, which have

¹Pfizer, Inc, Cambridge, MA, USA. ²Tufts Medical Center, Boston, MA, USA. ³These authors contributed equally: Miles Welbourn and Paul Sheriff. ✉email: xuemei.cai@pfizer.com

been validated for their reliability and can be comfortably worn, often minimizing their visibility by resembling familiar articles of clothing like fitness trackers or watches^{15,19}.

Validation of DHTs to support novel digital endpoints can be successful through stakeholder collaboration in pre-competitive initiatives combining industry and academic sponsors, regulators, manufacturers, patients, and healthcare providers (HCPs)²⁰. Regulatory bodies and stakeholders have been actively investigating and advising on the procedures for incorporating the digital endpoints derived by DHTs within clinical use and necessary criteria for their acceptance as endpoints or biomarkers^{20–25}. The necessary steps towards acceptance and adoption of novel digital endpoints include the technical and analytical validation for accuracy and reliability, as well as, patient experience, clinical validation, and clinical meaningfulness^{26–30}. While the analytical validation of gait and activity endpoints are generally performed during a controlled in-lab environment, where traditional and reference methods can be measured simultaneously with the test methods for comparison, the validation of these endpoints during longer, naturalistic, and free living conditions has been more challenging^{7,31}. Nonetheless, multiple DHT validation studies have shown evidence of measuring reliable gait and activity related metrics in both pediatric and adult populations^{32–42}. Scientific consortiums such as Mobilise-DS^{43,44} through the Innovative Medicines Initiative (IMI) has focused not only on a unified approach for the validation of mobility measures across patient populations^{45,46}, but also understanding the acceptability of DHTs for remote monitoring⁴⁷ and research on patient-centricity to shape development⁴⁸.

We previously showed that a single lumbar-worn accelerometer can be used to reliably measure gait metrics during a walking task, and gait measured during free-living conditions captures useful information that cannot be obtained in controlled settings (e.g., clinic or laboratory)^{15,49}. In this study entitled In-Clinic and Natural Gait Observations (I-CAN-GO), we provide further evidence on the reliability and performance of gait metrics derived by our in-house developed gait algorithm (SciKit Digital Health (SKDH) gait module)⁴⁹ as a device-agnostic gait characterization tool during various naturalistic or simulated daily activity tasks using a different lumbar-worn accelerometer (ActiGraph CentrePoint watch). In order to quantify the sensitivity of gait metrics to detect naturalistic behavior, 1) we first validate that using a single lumbar-worn accelerometer can distinguish between participants' self-perceived gait speeds and the effects of different floor surfaces on gait speed, 2) we show that gait and activity during simulated free-living activities and longer bouts of walking can be reliably captured, suggesting accurate representation of gait metrics in free-living conditions, and 3) we evaluate device comfort and wearability, an important consideration for using wearable sensors in patient populations for extended durations^{19,50}, for tested devices in different wear locations.

Results

Different self-paced walking speeds can be accurately captured by single lumbar-worn accelerometer

During in-lab assessments, participants were asked to walk three laps on an instrumented mat with three self-paced walking speed tasks (i.e., at their natural, fast, and slow walking speeds). During these walking tasks, participants wore an Ambulatory Parkinson's Disease Monitoring (APDM) six-sensor set^{51,52}, as well as a single accelerometer at their lumbar location. Gait metrics were derived from 1) GAITRite using ProtoKinetics' Movement Analysis Software (PKMAS)⁵³ (as reference), 2) APDM six-sensor set method⁵¹ (as reference), and 3) single lumbar-worn accelerometer using the SKDH gait algorithm⁴⁹ (as comparator). Figure 1 depicts the comparison of gait speed between references and comparator devices/methods using Bland–Altman plots and correlations across various walking speeds. Although SKDH gait underestimated gait speed compared to APDM and GAITRite, as also reported in previous findings¹⁵, the Bland–Altman plots showed homogenous mean bias (GAITRite – SKDH gait m/s = 0.15 (26%), 0.19 (19%), 0.29 (22%); APDM – SKDH gait (m/s) = 0.05 (8%), 0.10 (10%), 0.18 (14%) for slow, normal, fast walking speeds, respectively). The variability of SKDH gait metrics were higher and intraclass correlation coefficients (ICC) varied between poor to good across the gait metrics between APDM (range across speeds [0.59–0.84]), and GAITRite (range across speeds [0.36 – 0.55]), mainly due to the mean differences (see Supplementary Tables 1–2 for summaries).

Mixed effects model was used to test if gait metrics differed between the walking tasks and if gait metrics can be reliably captured with a single accelerometer device across different walking speeds. The following model was tested for each gait metrics: methods (GAITRite/APDM/SKDH gait), task (slow/natural/fast), and their interaction were entered as independent variables (fixed effect), age was entered as covariate, participant as random effect. Figure 2 shows gait speed, cadence, stride duration, and stride length across multiple methods and tasks. A significant task effect was present for all gait metrics (gait speed $F(152,2) = 604.85$, $p < 0.0001$; cadence $F(152,2) = 504.27$, $p < 0.0001$; stride duration $F(152,2) = 362.37$, $p < 0.0001$; stride length $F(152,2) = 418.03$, $p < 0.0001$, persisted after false discovery rate (FDR) correction for multiple comparisons), therefore, gait metrics derived by different methods (GAITRite, APDM, SKDH gait) were able to differentiate between different walking speed tasks. Gait speed and stride length showed significant method effect (gait speed $F(152,2) = 44.73$, $p < 0.0001$; stride length $F(152,2) = 107.55$, $p < 0.0001$; persisted after FDR correction), e.g., overall gait speed was estimated significantly higher in GAITRite than APDM and SKDH gait (see Supplementary Table 3 for all comparisons), mainly driven by the stride length. (Fig. 2). Overall, temporal gait metrics, such as cadence and stride duration showed excellent agreement and minimum bias across different methods; and all methods were able to differentiate between different walking speeds.

The floor surface affects gait metrics

To assess if gait on different floor surfaces differed, we compared the gait parameters extracted from a walk on a carpet and tile floors. GAITRite was placed under the carpet to capture as reference, however, GAITRite data was not possible to be captured for tile floor walk. Figure 3a shows the Bland–Altman plots and correlations

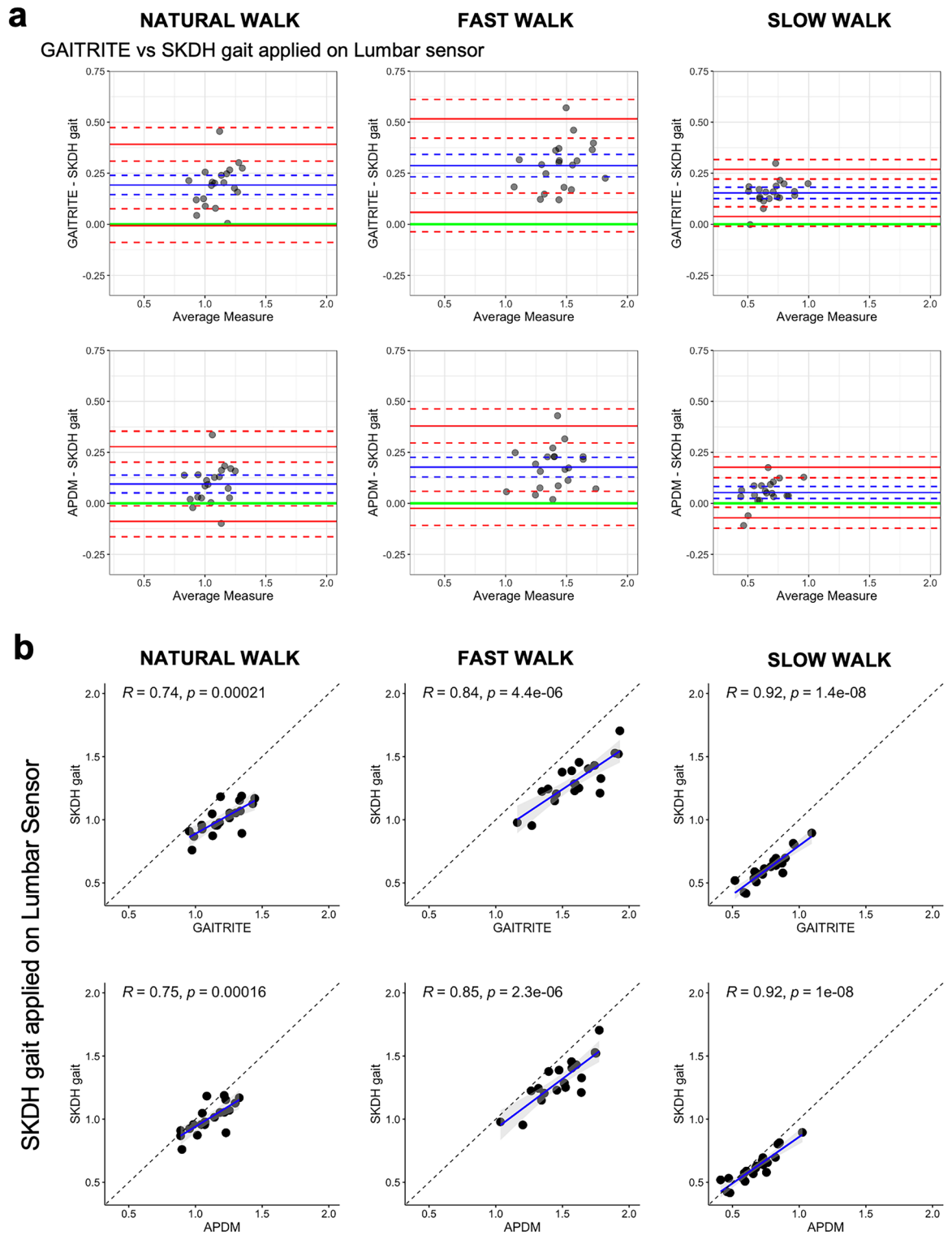


Figure 1. Gait speed validation for natural, fast, and slow in-lab walking tasks. Between GAITRite (reference), APDM 6 sensor-set (reference), and SKDH gait (comparator) applied on single lumbar-worn accelerometer across natural, fast, and slow a shows Bland–Altman plots, red lines represent the limits of upper lower bound and dotted lines represent the 95%CI, blue line represents the mean difference and dotted lines represent the 95%CI, green is 0 line, b correlations between reference and comparator methods with 95%CI in shaded area. The dotted line represent the equal line.

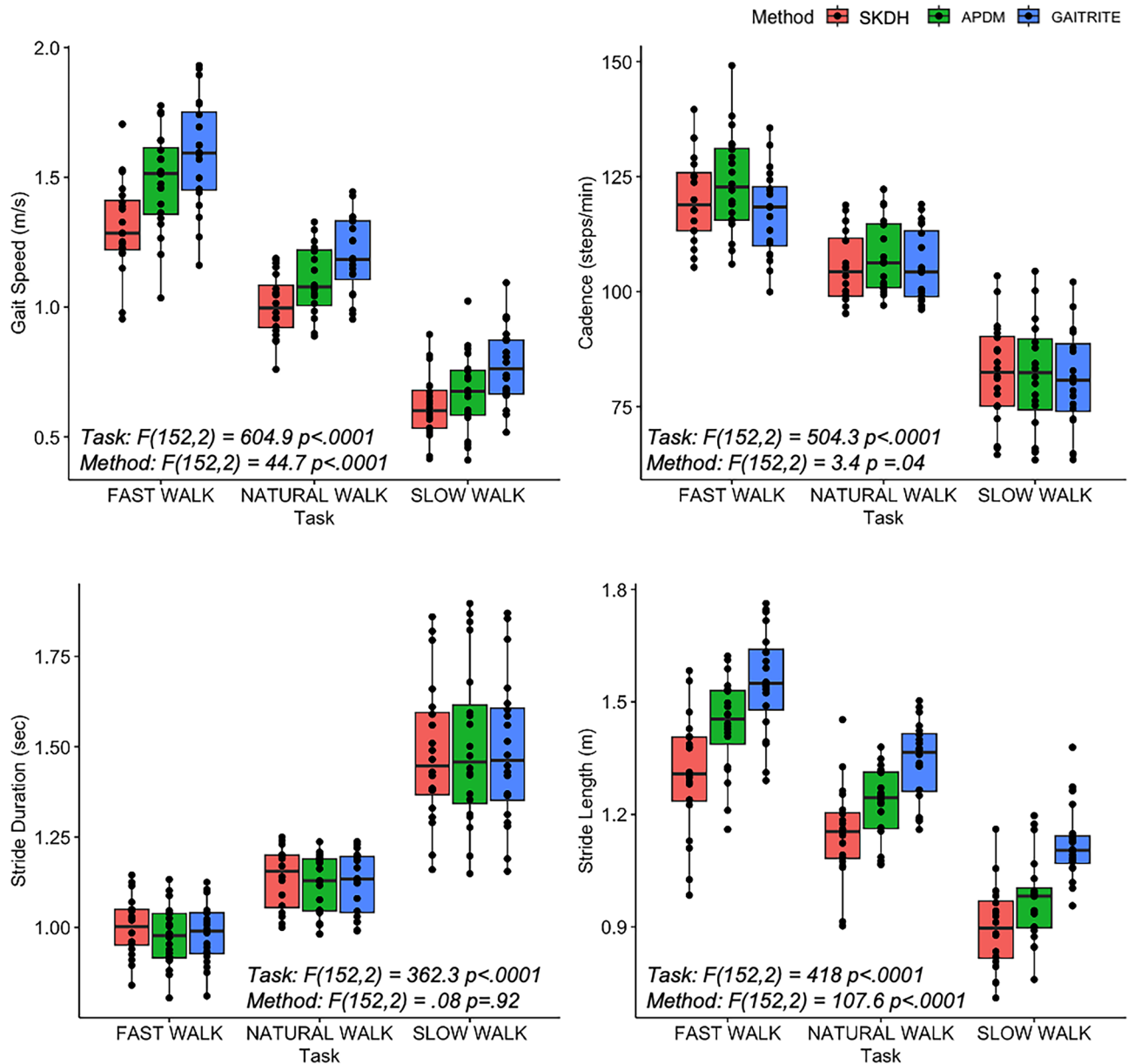


Figure 2. Different walking speeds were captured by gait metrics derived from all methods. All gait metrics showed significant task effect, gait speed and stride length had significant differences across methods that persisted FDR correction; gait speed and stride length were consistently underestimated compared to GAITRite. There was no interaction effects between tasks and methods which suggested error did not significantly variate across walking tasks and devices.

for gait speed, reflecting homogeneous mean bias and good correlations across methods. Like different walking speeds, we observed homogenous mean bias across different methods (GAITRite – SKDH gait (m/s) [= 0.17 (17%) for carpet walk; APDM – SKDH gait (m/s) = 0.08 (7%), 0.09 (8%), for carpet and tile floor walking speeds, respectively). Mixed effects model showed significant task effect for gait metrics, except stride length; and significant effect of method for all gait metrics (Fig. 3b). Gait speed was underestimated by SKDH gait compared to GAITRite and APDM for both floor surfaces. There was no interaction effect between tasks and methods. The changes in gait due to floor surface differences were captured by all methodologies; i.e., higher gait speed, higher cadence, and faster strides were observed during tile walk compared to carpet walk (see Supplementary Table 3).

Gait metrics during simulated activity blocks and longer walking can be derived accurately from single lumbar-worn accelerometer

In order to simulate naturalistic walking behavior, participants were asked to perform two sets of simulated activities in the lab, each consisted of combinations of walking, sitting, and standing tasks and lasted around 7 min, as well as 20 min of free walking outside the lab (around the block) attended by a facilitator. The APDM six-sensor set and lumbar-worn accelerometer sensor were worn simultaneously by the participants to validate

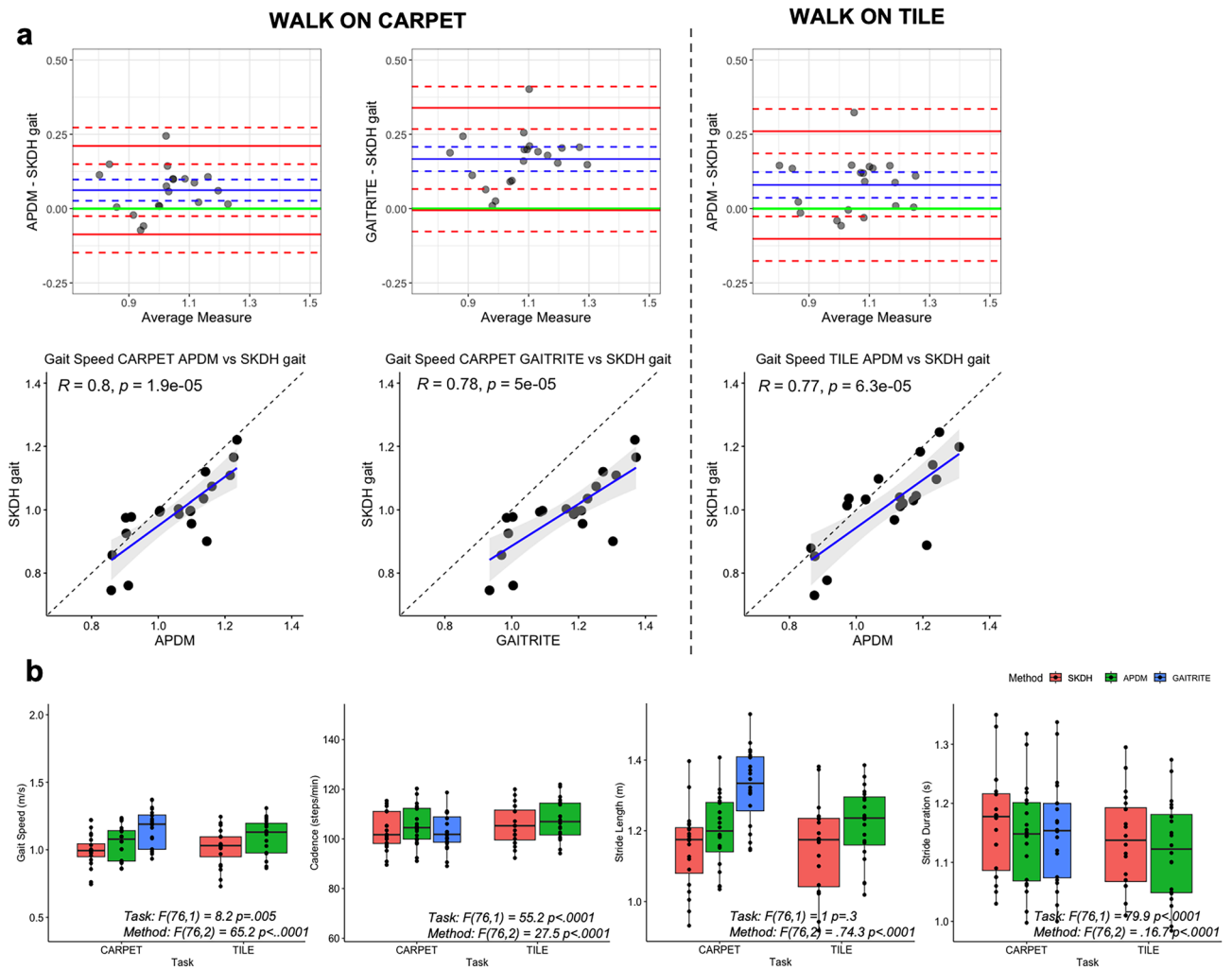


Figure 3. Gait metrics validation for carpet and tile floor walking. Due to the nature of the task, GAITRITE data could not be captured for tile floor walking. Similar characteristics across methodologies were observed compared to different walking speed tasks; a shows homogeneous bias between methods with moderate to excellent agreement and good correlations, b mixed effect models showed significant different across methods with SKDH gait for all gait metrics, and significant task effect, except stride length. SKDH gait underestimated gait speed compared to other methods; and all methods were able to capture the change of gait metrics due to different floor walk (i.e., gait speed was significantly higher for tile walk compared to carpet).

the gait metrics. Figure 4 a shows the Bland–Altman plots and correlations between gait metrics derived from APDM six-sensor set and SKDH gait from lumbar sensor. Gait speed shows good correlation and agreement between the two methods, with mean bias of 0 and 0.04 m/s (3.2%) for activity blocks and outside lab walk tasks. ICC values were 0.816 for activity block and 0.898 for the outside lab walk tasks. Other gait metrics also showed excellent agreement (see Supplementary Tables 1–2).

A mixed effects model was used to test if gait metrics differed between the simulated activity blocks and outside lab walks and if gait metrics can be reliably captured with a single accelerometer device across different walking speeds. The following model was tested for each gait metric: methods (GAITRITE/APDM/SKDH gait), task (activity blocks/outside lab walk), and their interaction were entered as independent variables (fixed effect), age was entered as covariate and participant as random effect. Figure. 4b depicts the gait speed and cadence across multiple methods and tasks. There was only a significant task effect present for all gait metrics (gait speed $F(57,1) = 6.97, p = 0.01$; cadence $F(57,1) = 4.2, p = 0.05$; stride duration $F(57,1) = 4.05, p = 0.05$; stride length $F(57,1) = 13.11, p = 0.0006$), which indicated simulated activity blocks showed only slightly higher cadence and lower gait speed compared to outside walk, due to shorter, smaller steps taken during the simulated activity blocks.

In addition to gait metrics, we investigated the total number of steps detected by APDM and SKDH gait applied on lumbar-worn sensor, and wrist-worn sensor⁵⁴ during simulated activity blocks and outside lab walk tasks. Figure 5 depicts the Bland–Altman plots and correlations of total number of detected steps during the 20 min outside lab walk task across three devices. On average mean (SD) = 2150 (299), 2076 (277), and 1894 (334) steps were detected during the 20 min walk with wrist-worn sensor, APDM six-sensor set, and SKDH gait

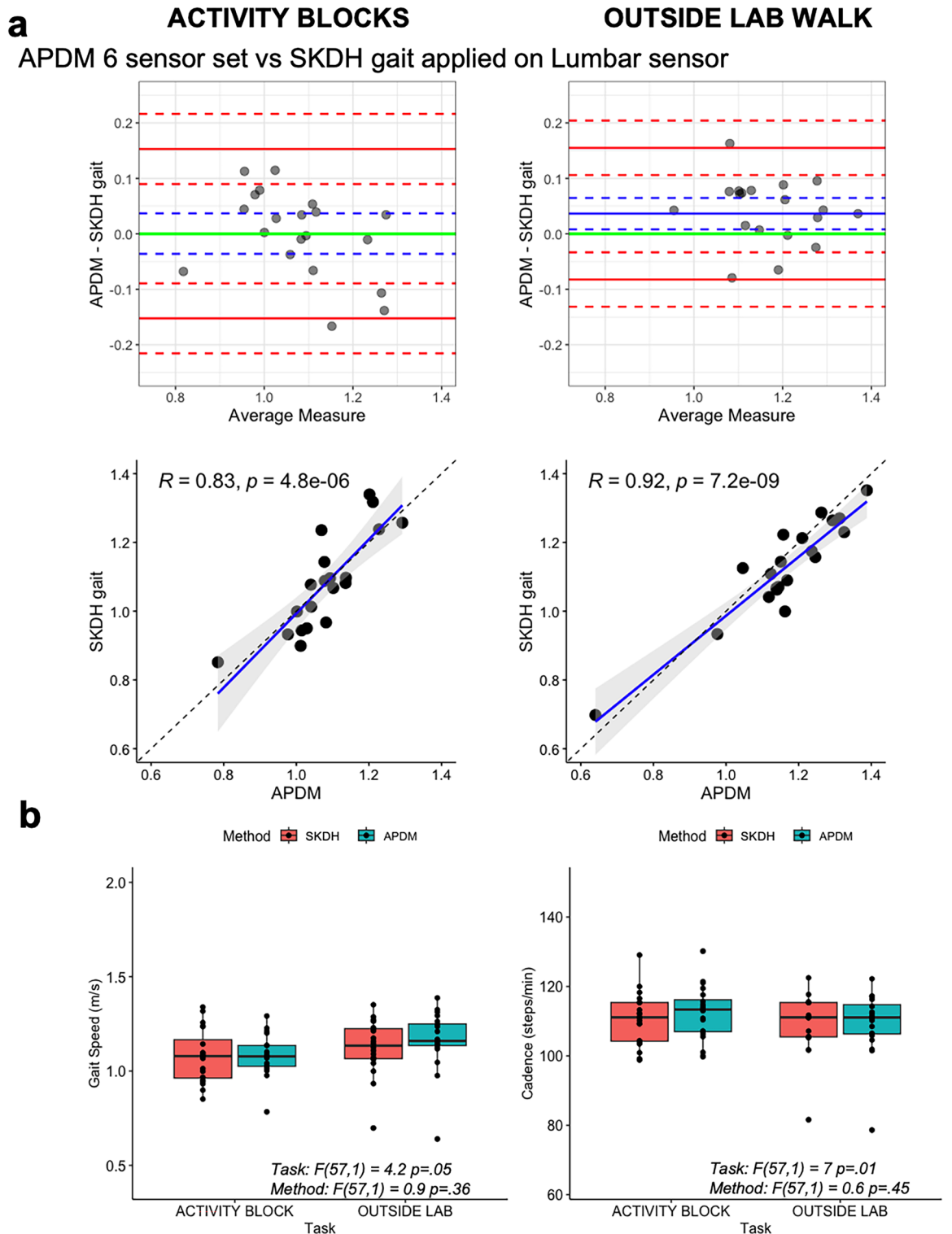


Figure 4. Gait speed validation for simulated activity blocks and 20 min outside lab walk tasks. Participants performed two simulated activity blocks, each lasted around 7 min, as well as, 20 min outside lab walk, where they were let to move/walk freely around the neighborhood. The gait metrics were derived by the APDM six-sensor set (reference) and SKDH gait applied on the lumbar-worn sensor (comparator). There was good agreement and correlation for gait speed as reflected by a Bland Altman and correlation plots between APDM six-sensor set and single lumbar-worn accelerometer. The mean gait speed across devices and tasks are depicted in b. The mixed effects model showed significant task effect between methods, whereas, no method effect existed, showing only slightly lower gait speed during activity blocks compared to outside lab walk.

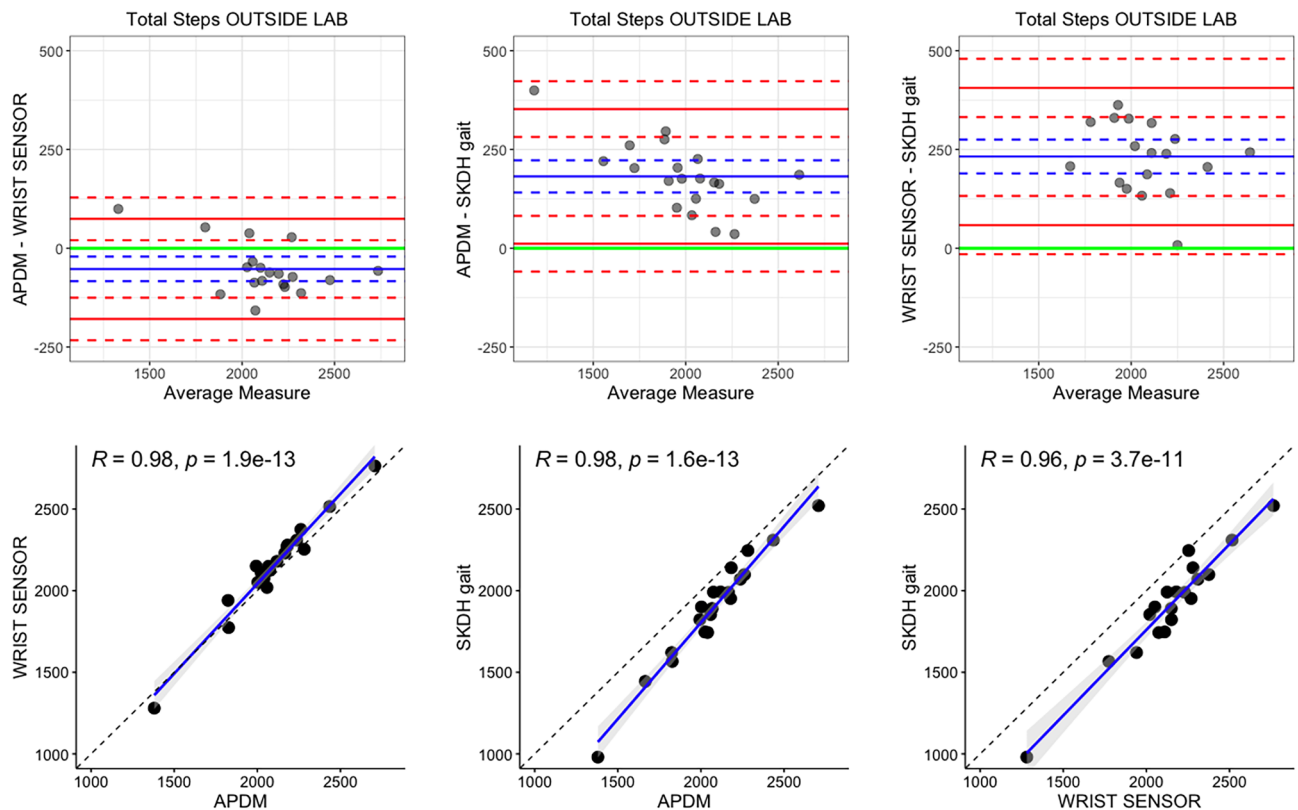


Figure 5. Comparison of total number of steps detected across multiple methods during 20 min outside walk task. The number of detected steps were compared between APDM six-sensor set, wrist-worn sensor, as well as SKDH gait applied on lumbar-worn sensor. The wrist-worn sensor detected the highest number of steps, APDM and SKDH gait applied on lumbar-worn sensor detected slightly lower number of steps, which is expected as SKDH gait and APDM prioritize gait characterization over step counts. All methods were in good agreement and had excellent correlation between each other.

applied on lumbar-worn ActiGraph sensor. The correlations and ICCs showed excellent agreement and correlations within the 20 min walk task across methods, whereas the ICCs were poor during the in-lab simulated activity blocks with high correlations (Supplementary Table 2 and Supplementary Fig. 1).

Simulated in-lab activity and 20 min walking compared to at home 7 days of data

We compared the gait speed derived from the single lumbar-worn accelerometer with SKDH gait during in-lab tasks resembling naturalistic walking; i.e., simulated activities and 20 min outside walk, and continuous monitoring for around one week. When compared using mixed effect model, there was a significant effect of task ($F(37,2) = 12.3$, $p = 0.0001$; $F(37,2) = 12.8$, $p = 0.0001$ for gait speed and 95th percentile of gait speed, respectively). The average at-home gait speed was significantly lower (paired t-test, $t(18) = -4.72$, $p < 10^{-3}$; $t(18) = -3.97$, $p < 10^{-3}$) and 95th percentile of at-home gait speed was significantly higher ($t(18) = 5.65$; $p < 10^{-4}$; $t(18) = 3.61$, $p = 0.002$) than both of the in-lab activities (Fig. 6a). Additionally, regression analyses were used to test if the in-lab gait speeds estimated the at home gait speed. The outside walk task gait speed showed trending pattern to predict the at home gait speed (adjusted $R^2 = 0.14$, $F(1,17) = 4.03$, $\beta = 0.33$, $p = 0.06$); gait speed during activity blocks predicted the at home gait speed (adjusted $R^2 = 0.33$, $F(1,17) = 9.7$, $\beta = 0.48$, $p = 0.006$), however, the explained variations were 14% and 33% (Fig. 6b). Similar behavior was observed for the relationships between gait speed during in-lab activities and 95th percentile of gait speed at home (see Supplementary Fig. 2). When we separated the at home gait speed across various bout duration (short: 10 s-30 s, medium: 30 s-60 s, long: > 60 s) we observed increasing gait speed with bout duration and gait speed derived from the medium and longer at-home bouts were in closer range with simulated activities and outside lab walking gait speed (Fig. 6c). The paired t-tests showed that the at-home gait speed during short bouts was significantly slower than the gait speed during in-lab activity blocks and 20 min walk tasks ($t(18) = -5.19$, $p < 10^{-4}$; $t(18) = -5.46$, $p < 10^{-4}$, respectively); gait speed during longer bouts was significantly higher than gait speed during in-lab activity blocks ($t(18) = 3.43$, $p = 0.003$), suggesting medium bouts were more reflective of in lab simulated activities, and both medium and longer bouts might be more relevant to outside lab walk. The regression analysis showed that the explained variance was relatively higher (adjusted $R^2 = 0.45$, $F(1,17) = 15.52$, $\beta = 0.79$, $p = 0.001$, Fig. 6d) between activity blocks and at home gait speed during medium duration bouts, and adding shorter and longer bouts to the model did not have a significant change. Outside lab walk and medium and longer bouts shared low variance (see Supplementary Fig. 2).

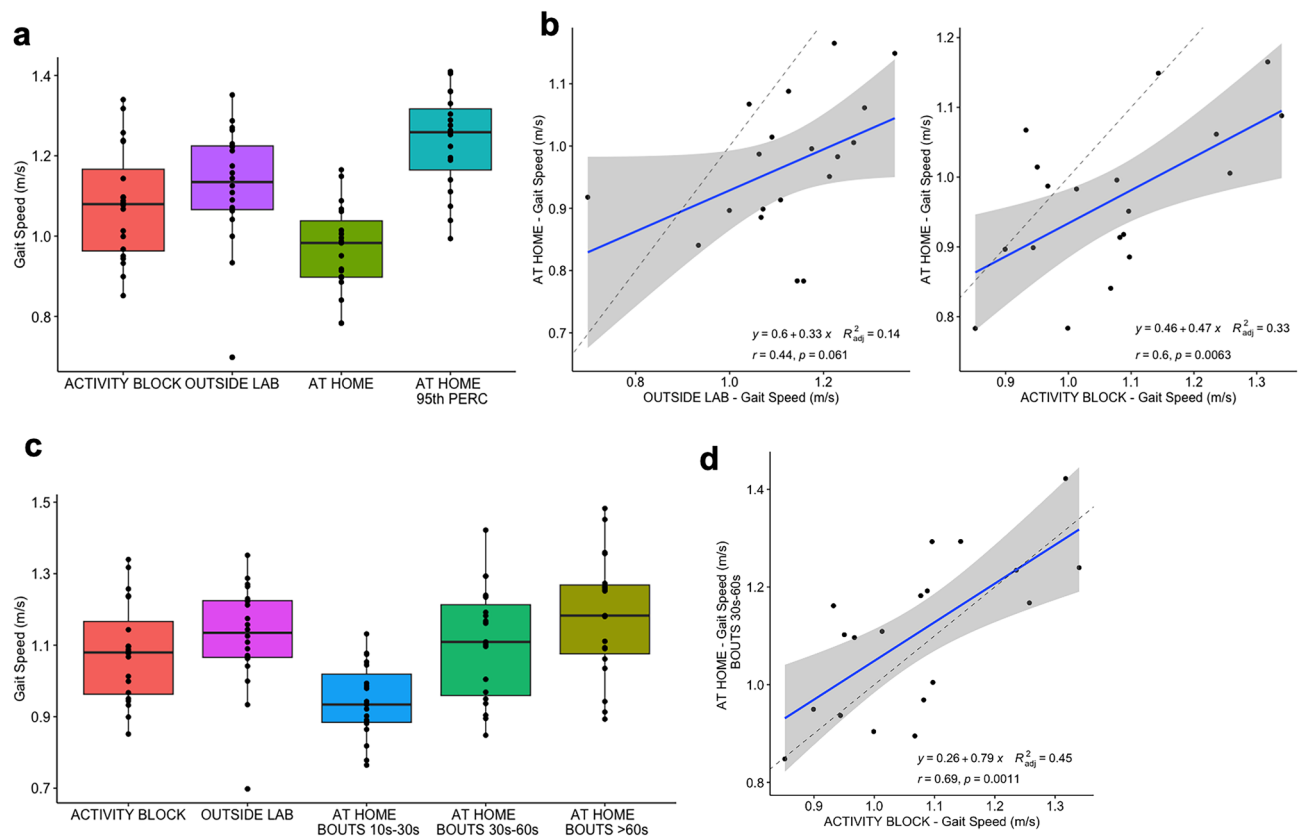


Figure 6. Comparison of naturalistic gait speed at home and in-lab tasks. **(a)** Both the average and 95th percentile gait speed at home was significantly different than in-lab gait speed. **(b)** The linear regression model showed the outside lab walk only explained 14% of the variance at-home ($b = 0.33$, and Pearson's $r = 0.44$), and activity blocks explained only slightly higher 33% of the variance at-home ($b = 0.47$, and Pearson's $r = 0.6$). **(c)** When broken down into increasing bout lengths, the at home gait speed spans in-lab gait speeds, and **(d)** at home gait speed during medium bout durations (30 s-60 s bouts) explained the most variance in gait speed during activity blocks (45%, ($b = 0.79$, and Pearson's $r = 0.69$). The dashed lines represent the equal line.

Comfort and wearability of sensors

The comfort of wearing sensors was rated by participants both during in-lab visits and after 1 week of home monitoring. Overall, sensors were found acceptable and comfortable to wear; 88.2% and all participants are comfortable to wear lumbar and wrist sensors at home, respectively (Fig. 7a; rating 3 and above, where 0 indicated tolerable enough and 5 indicated very comfortable). Similarly, participants rated how long they would be willing to wear the sensors continuously; 72.2 and 85% of the participants indicated they would wear lumbar and wrist sensors for more than 1 week (Fig. 7b). The comfort of sensors during in-lab visits was also positive: for both visits, participants reported positive experience with sensors' comfort (Fig. 7c).

For this study, during the at home monitoring period, the compliance per day was defined as 18 and 10 h of wear time for the wrist- and lumbar-worn sensors, respectively. The average compliant days for the wrist-worn sensor was mean (SD) 5.79 (0.63), with range between 4 to 7 days; and the average compliant days for the lumbar-worn sensor was mean (SD) 5.7 (0.87), with range between 3 to 7 days. Participants were also asked to report average sensor wear time during at home monitoring. The participant reported wear time was in concordance with the average percent wear time derived by the sensor data (Fig. 7d; on average 97.4 and 72% wear time were reported by the participants compared to the 96.5 and 64.3% wear time derived from the wrist and lumbar-worn sensor data, respectively). The lumbar sensor wear time is shorter than the wrist sensor as the participants were given the option to take the lumbar-worn sensor off during night time.

Discussion

We designed a master protocol for technology evaluation against reference standards to allow for rapid evaluation of digital health technologies that assess physical activity and gait in adults aged 18 and older. In this first cohort of this protocol, we investigated the use of the ActiGraph Centrepoint Insight Watch in both the wrist-worn position and the lumbar position during two in-lab assessments and a free-living period of approximately 7 days. We aimed to (1) further validate and evaluate the performance of a method for measuring gait relying on a single lumbar-worn accelerometer running SKDH gait with respect to reference methods relying on six devices (APDM), and an instrumented walkway (GAITRite); (2) test the ability of our in-house built SKDH gait algorithm to distinguish between participants' self-perception of different gait speed changes (fast walk,

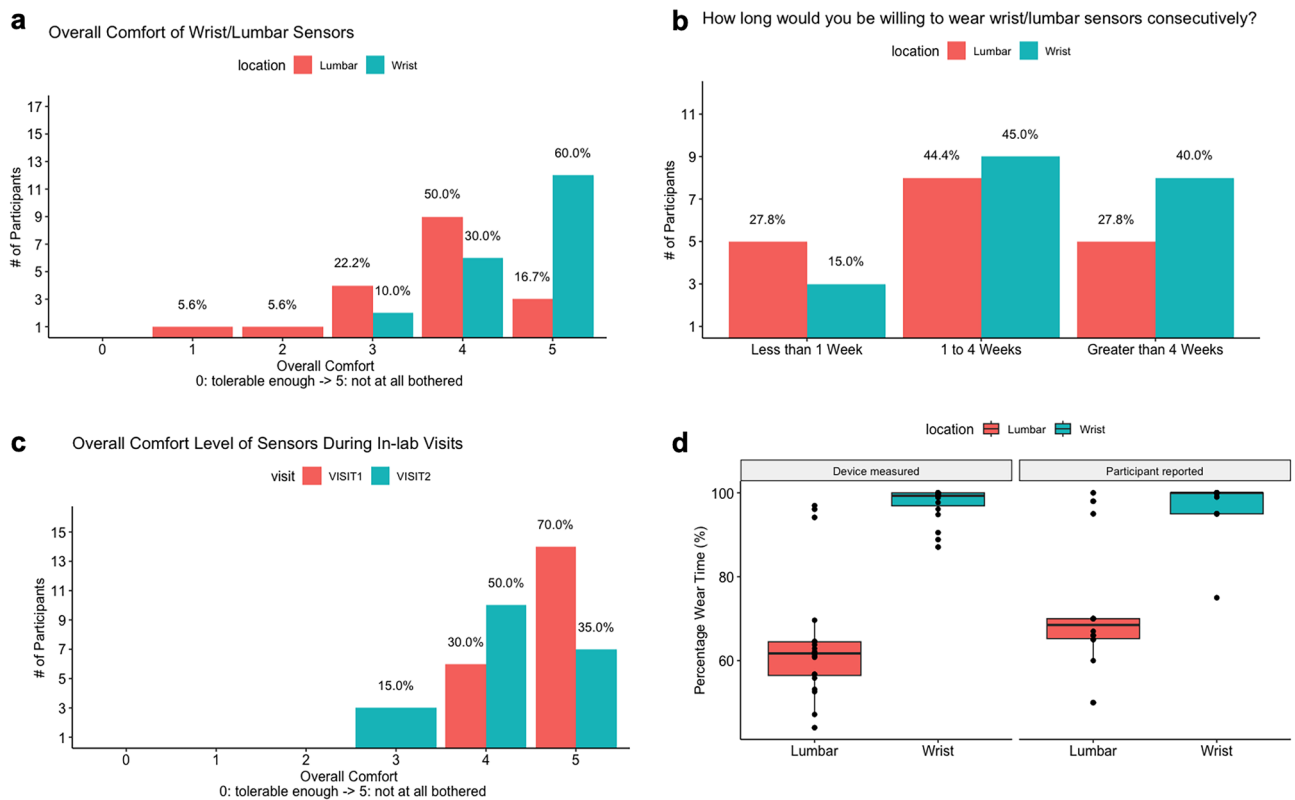


Figure 7. Overall comfort assessment of wearing the sensors at home and in-lab, and total home wear time. Participants were asked to assess the comfort of wearing the sensors by multiple questions; (a) shows results of when participants were asked after 1 week of home monitoring “Please rate the overall comfort of the lumbar and wrist- sensor at home where 0 means that they were only tolerable enough to participate in the study and 5 means that you were not at all bothered.”; (b) shows participants’ willingness to wear those devices for extended periods of time, where more than 70% and 85% responded they would wear lumbar and wrist devices more than 1 week; (c) plots the overall comfort level of sensors during 2 in-lab visits (question worded similarly to what is shown in Fig. 7a); and (d) shows the average wear time per day recorded by the sensors compared to the participants’ responses to the wear time question “What percentage of the time did you wear the wrist/lumbar sensors?” Both methods resulted in similar results, note that the participants were given the option to take the lumbar sensor off during the night.

slow walk, natural walk); (3) test SKDH gait performance against APDM during a 20 min free-walk as well as simulated activity block with 2 sitting, 2 standing, and 2 walking tests performed in random sequence to better estimate performance during future unsupervised ambulatory monitoring; (4) look at the effect of different floor surfaces such as carpet and tile on walking performance; and (5) compare to previous studies with different accelerometer devices to demonstrate SKDH gait is agnostic to specific accelerometer devices across studies when evaluating algorithm performance. Our results show that SKDH gait performance is reliable and accurate with homogenous bias with respect to the reference methods and provide further evidence of its use in free-living conditions with different sensors.

Gait has often been called the sixth vital sign⁹. Mobility and how well we walk has been shown to be correlated with long term health outcomes and recovery from disease. Accurately assessing mobility is thus of paramount importance. In previous studies, we have assessed the validity of sensor-based methods for estimating gait speed in a controlled setting using the GENEActiv Activ Insights watch^{15,55}. Technology is frequently changing both from hardware and software perspectives and being able to develop algorithms that are device-type agnostic is essential to the ability to continue to use digital health technologies to improve measurement capabilities in clinical trials and clinical use⁵⁶. The results of this study on both temporal and spatial metrics of gait derived using SKDH gait using the ActiGraph Centrepoint Watch show similar agreement as measured via ICC to reference standards used in previous studies¹⁵. Similarly to previous studies, temporal metrics are more accurate compared to spatial metrics, which rely on the physical inverted pendulum model for a single lumbar accelerometer. However, there remains good to excellent agreement between all measures of gait, concordant with previously published results with a different manufacturer-made sensor worn in the same position¹⁵.

An important feature of devices that monitor gait and physical activity is the ability to detect walking at all walking paces from slow to fast as would happen during bouts of naturalistic walking. In this study, participants’ self-perception of three different walking speeds (fast, slow, natural speed) can clearly be captured with accelerometers (Fig. 2). As the self-perceived walking speed increased, all methods consistently trended towards smaller (i.e., decrease in stride length) but longer (i.e., increase in stride duration) steps. Rather, gait speed, as a byproduct

of stride length and duration, might be the most promising gait metric, with larger effect size, to detect changes between different populations or therapeutic effect. There was a significant device effect on most measures as the overall gait speed is estimated significantly higher in the GAITRite than APDM and SKDH gait. This is likely due to the relatively higher accuracy of temporal vs spatial measures of gait using the SKDH gait model seen also in previous studies with an earlier version of the same algorithm¹⁵. Stride duration does not differ between the ActiGraph Centrepoint device running SKDH gait against the reference measures but there is a significant difference in stride length across the three devices. Overall, however, despite the device effect, the SKDH gait algorithm run on accelerometry data from ActiGraph Centrepoint Insight watch at the lumbar position was able to show clear differences between the three different speeds of walking of this cohort of participants in lab. Furthermore, the error measured across tasks appears to be consistent for each device which would also further give confidence towards measuring different walking paces at home using the same device consistently across a study or to measure therapeutic effect. Additionally, the correlation appears to be strongest between the reference standards and SKDH gait during the slow walk task; given that most pathological diseases will likely adversely affect mobility from the norm rather than improve it, this may be promising for future clinical applications.

Accurately assessing mobility at home is challenging as this may require special instrumented houses or rooms which is labor intensive, wearing multiple sensor sets continuously which is burdensome, or installing other specialized equipment in peoples' homes which presents operational and privacy challenges⁴⁵. This study incorporated longer duration walk activities and simulated activity blocks to replicate naturalistic behaviors while still undergoing reference measures. While it cannot completely replace validation in the home, these simulated blocks are longer and more consistent. We observed during the 20 min walks and the simulated activity blocks higher correlations and lower bias between gait speed as measured by SKDH gait with a single accelerometer against reference standards compared to during the walk tasks on the GAITRite. For step counts however, simulated activity blocks had lower ICCs; there was high correlation seen which could be related to bias. For the longer 20 min walk, the step counts had high ICC and high correlation across the devices likely due to the longer duration of the task and longer number of consecutive steps. Previously, we have published data, that replicates other studies, that stimulated walking on a GAITMat could not distinguish between younger and older participants in lab, likely related to a performance effect^{15,57,58} of being observed during walking activities. These simulated activities of daily living (ADLs) and longer 20 min walking tasks newly incorporated in this master protocol have a higher diversity of duration of walking bouts and walking speeds compared to traditional shorter and episodic walking tasks (e.g. 4 m or 30 m walking tasks). The SKDH gait algorithm continues to provide good performance under these more diverse simulated walking conditions that are closer to naturalistic walking bouts outside of the clinic setting. In fact, the performance of SKDH gait algorithm by ICC comparisons is higher for the simulated and 20 min walk tasks compared to the instrumented gait mat walking tasks which are of shorter length (Fig. 6b). Thus, the algorithm performance increases with longer duration and bouts of walking, which are likely closer to the diversity of walking tasks in the real world. Furthermore, the simulated activity blocks and the 20 min walk task capture the full range of different short, medium, and long walking bouts seen in the at home gait data (Fig. 6c).

As part of the study procedures, we also sought to determine if there was an effect of different types of floor surfaces on gait parameters. Based on both results from the APDM system and SKDH gait system we detected a statistically significant floor surface effect in gait metrics including cadence (steps/min), gait speed (m/sec), stride duration (sec). In general, participants walked faster on tile than on carpet, driven by stride duration as the stride length was comparable between carpet and tile floors. The ICC and correlation values between devices suggest good concordance between the test device and reference standard. From a measurement perspective, this is unlikely to impact measurements of efficacy and treatment effect as long as the participant is not being assessed frequently on a multitude of different floor surfaces. Additionally, with the same algorithm, over time, with longer bouts of monitoring continuously, effects from floor surfaces such as carpet or hardwood or tile will likely even out. However, as decentralized elements of trials continue to grow and assessments at home increase such as 6 min walk test evaluation, this has important implications for standardizing the room in which a participant may complete the assessment repeatedly as to not add an additional variable to the 6 min walk distance.

This study was conducted during the height of the COVID-19 pandemic and as a result the demographics of this cohort skew towards younger participants (Mean age 34 SD 9.06 years) which may affect the reports of participant acceptance of devices. However, because the measurements are around the devices in this study and not on a populational level, this does not affect the validation data but may make the study less generalizable in the future when trying to age-match this cohort to other studies with patients. It is important to note that this study serves as a proof-of-principle study around the analytical validation of the SKDH using in clinic tasks to reflect at home gait ranges and variations. With the small sample size, we are not generalizing the gait metrics on a populational level in this study but rather looking at the validity of a tool and methods to be used in future clinical validation studies with different patient populations. Also, as this is a Master protocol, in the future, populational level data could potentially be reported as the sample size for the overall study increases over time.

Future directions include applying SKDH gait to other form factors that are relevant to patient populations including chest accelerometry⁵⁹ which, in particular, may be of interest to patients with cardiovascular diseases as many implanted devices as well as externally worn ECG patches also contain an accelerometer placed in the sternal position. Additionally, there are some technology companies that have designed pendants that are worn around the neck that sit upon the chest for which this algorithm may be applied. This is important as the lumbar position for gait assessment comes with some downsides compared to wrist worn devices shown when participants are asked to consider the best type of form factor that does not intrude with day-to-day activities and comfort (Fig. 7). There were still 94% of participants willing to wear the wrist worn device for 1–4 weeks or longer than 4 weeks compared to 72% of participants willing to wear the lumbar device for 1–4 weeks or longer than 4 weeks. No digital health technology study can accomplish its objectives if participants are unable

or unwilling to wear the technology. Wearability is an important consideration for the future which may require different form factors coupled with statistical analysis of the data to investigate the amount of wear time required to assess for mobility in a particular study and target patient population to reduce participant burden.

While SKDH gait has been applied to adults as old as 85 years of age and other older populations with heart failure^{60,61}, osteoarthritis⁵⁶, and Parkinson's disease^{55,62,63}, a population that has been less well studied with digital health technologies includes children under the age of 18. This is a future direction to apply the knowledge learnt in validation of gait in adults to pediatric populations as well as to test the validity of the gait algorithm against reference standards in populations with physical activity limitations to evaluate the potential impact on algorithm reliability and gait speed variability. This dataset combined with other datasets collected from healthy control participants may also be used to look for construct validity by comparing these other populations against age-matched healthy controls and to evaluate sensitivity of certain gait endpoints such as gait speed in detecting differences induced by disease.

In current clinical trials, measures such as the 6MWT, 4SC, and the SPPB have been used as performance outcome assessments in the clinic to measure outcomes in interventional therapeutic trials. However, these assessments only provide a snapshot view of a participant's functional ability and may depend on a multitude of factors including length of travel to clinical trial site, motivation, and performance effect¹⁵. Digital health technology offers an opportunity to measure patients in their own life and provide continuous assessment of their day-to-day real life fluctuations. The patient voice has greatly contributed to modern drug development; in multiple therapeutic areas ranging from heart failure to chronic obstructive pulmonary disease (COPD) to DMD^{5,64,65} there is a high unmet need for measures that accurately represent how a patient is doing at home. The demand for validated clinical outcome assessments that can demonstrate a clinically meaningful therapeutic response in clinical trials is only growing. This study illustrates an example of how to analytically validate these types of quantitative endpoints and select digital health technologies to give confidence for further clinical validation in patient populations.

Conclusion

The rise in sophisticated digital health technologies has enabled reliable and continuous measurements of gait and activity in free-living environments opening new doors in the clinical trial and healthcare spaces. This study serves as a proof-of-principle study around the analytical validation of the SKDH algorithm using in-lab tasks to simulate the variations of activity and walking patterns that would be found at home. Our results show that SKDH gait performance is reliable and accurate with respect to the reference methods and provide further evidence of its use with different sensors. Participants reported high comfort and compliance with the DHT in this study. In the future, these gait and activity measurement findings can be leveraged when performing clinical validation and assessments of digital health technologies in patient populations.

Methods

Subjects

Twenty healthy individuals aged 25–61 years (mean (SD) age = 33.95 (9.06 years), 13F,

Table 1), a race distribution of 13 white-identifying/7 Asian-identifying, mean (SD) of height of 168.5 (7.08) cm and weight of 70.11 (17.57) kg, signed informed consent prior to participation in this study¹⁰. Participants did not have any significant health problems as determined by a medically qualified study doctor during initial intake. The study was reviewed and approved by the Advarra IRB (protocol number: Pro00043100) and was conducted in accordance with relevant guidelines and regulations including the Declaration of Helsinki. Informed consent was obtained from all participants.

Experimental procedure

Each participant took part in two instrumented in-lab assessments about a week apart and around a 7 day at-home portion in between. The in-lab portion was completed at the Pfizer Innovation Research Laboratory (PfIRe Lab) in Cambridge, Massachusetts. During both visits, participants wore a set of 6 inertial sensors (Opal,

Number of participants (#)	20
Sex (#, F/M)	13/7
Age (years, mean \pm sd) [min-max]	33.95 \pm 9.06 [25-61]
BMI (kg/m ² , mean \pm sd)	24.44 \pm 4.67
Height (cm, mean \pm sd)	168.5 \pm 7.08
Weight (kg, mean \pm sd)	70.11 \pm 17.57
Race (#)	13 White 7 Asian
Ethnicity (#)	3 Hispanic or Latino 17 Non Hispanic or Latino

Table 1. Participant Demographics. In total, 20 adult participants were included in the study, which was conducted at the Pfizer Innovation Research Laboratory (PfIRe Lab), MA. Participants did not have any significant health problems as determined by a medically qualified study doctor during initial intake.

APDM Inc., Portland, Oregon) consisting of tri-axial accelerometer, gyroscope, and magnetometer worn at the sternum, lumbar (approximately L4 position), and bilaterally on the wrists and the feet at a sampling rate of 128 Hz. Two tri-axial accelerometers (ActiGraph CentrePoint[™] Insight Watch, ActiGraph, Pensacola, Florida) were also placed during the in-lab visits, one on the non-dominant wrist and one at the lumbar (L4 position) using a belt. During scripted walking tasks, participants walked on an instrumented walkway (GAITRite[®], CIR Systems Inc., Franklin, New Jersey) to capture foot pressure data.

Participants completed a battery of tasks. First, they were asked to: walk at a self-selected “normal” walking speed, at a self-selected “slow” walking speed, and then at a self-selected “fast” walking speed for 3 laps over the 7 m mat. They were then asked to walk at a “normal” walking speed first on a thick carpet walkway and then on tile flooring to examine floor surface effect. The same 7 m mat was also used under the carpet but given the nature of the floor surface, GAITRite[®] could not be deployed on the tile walk task.

Six stations were set up where participants would complete 2 walking (tossing and retrieving beanbag, transferring blocks between stations), 2 sitting (reading a short story for a minute, reading a short passage aloud), and 2 standing tasks (rolling a dice 3 times and summing results, transferring paper crown from hook-to-hook) with station selection given at random as determined by the roll of a 6-sided dice. The participant would continue this process until 7 min passed and would repeat this a second time after a short break.

If they chose to and were deemed healthy-enough to do so, participants also participated in a 2 min elliptical task which served as a non-walking active task for algorithm training. Lastly, accompanied by study staff, participants completed a 20 min walk outside of the lab space wearing the 6 APDM Opals and 2 ActiGraph CentrePoint Insight watches at non-dominant wrist and lumbar locations. They then completed a self-developed comfort and wearability assessment (Supplementary Material: Comfort and Wearability Questions) and were instructed on how to wear the two ActiGraph CentrePoint Insight watches at home while going about their normal daily routines to capture free living data. The lumbar accelerometer was permitted to be removed during sleep, but participants were asked to wear the non-dominant wrist accelerometer continuously at home. They then returned to the site approximately 7 days later and repeated the battery of in lab testing and comfort and wearability assessment described above. In addition, participants answered comfort and wearability questions related to the devices they wore during the at-home monitoring period and their estimated wear time of these devices.

Algorithms and endpoint derivation

Three separate methods were used for estimating gait metrics during the performance of gait tasks in the laboratory. Reference gait metrics were estimated from data collected using an instrumented walkway using a vendor supplied proprietary algorithm (GAITRite Software version 4.8.8) with pressure data from the GAITRite[®] loaded into ProtoKinetics' Movement Analysis Software (PKMAS Software Version 5.09C2a) for final endpoint derivation. Six inertial sensors (Opal, APDM) estimated gait metrics using proprietary algorithms (APDM Mobility Lab V2.0)^{51,52} during all in-lab tasks including the scripted gait tasks, the activity blocks, and the 20-min walking activity. Gait metrics including step count were also derived from the lumbar mounted ActiGraph tri-axial accelerometer using our open-source algorithm (SciKit Digital Health gait module, v0.9.7). Step counts were derived from the wrist-worn ActiGraph tri-axial accelerometer using the algorithm published by Bagui *et al*⁵⁴.

Statistical analysis

The sample size is determined by ICC between gait speed obtained from the sensor data vs. gait speed obtained from the instrumented walkway (GAITRite[®], CIR Systems Inc., Franklin, New Jersey) with null hypothesis derived by the median value of the lower bound of ICC for all gait metrics (8 spatial and temporal metrics) using the data from previous study¹⁵. In order to reach ICC value of 0.72 with ICCnull value of 0.27 using one-tailed test with $\alpha = 0.05$ and power = 0.8, 17 participants were required, therefore, 20 participants were enrolled (anticipating 15% attrition).

Statistical analysis was performed in SAS version 9.4 and R version 4.2.2 (2022–10–31) with following main packages: “lme4” for linear mixed-effect regression, “BlandAltmanLeh” for Bland–Altman plots, and “psych” for ICC. For each in-lab task, the gait metrics or other activity-related variables were computed as the median values across the steps. The gait metrics were averaged across the in-lab visits for simplicity, as the data had high test–retest reliability and similar results were obtained when tested per each in-lab visit (see Supplementary Table 5). Agreement of gait metrics between different methods were characterized with Bland–Altman plots, accuracy metrics such as mean (percent) and mean absolute error, as well as ICC_{2,1} (two-way random effects, absolute agreement, with respect to single measurement). Reliability of estimated gait speed was assessed according to the following benchmarks: ICC < 0.4 indicates poor, 0.4–0.59 fair, 0.6–0.74 good, and 0.75–1 excellent reliability⁶⁶. Pearson's correlation coefficient was computed to test the similarity between gait metrics.

Linear mixed-effects models were used to test if gait metrics differed between the walking tasks and if gait metrics could be reliably captured with a single accelerometer device across different walking speeds as well as across simulated activities and outside lab walk tasks. The following model was tested for each gait metrics during multiple speed walking tasks: methods (GAITRite/APDM/SKDH gait as categorical variables), task (slow/natural/fast as categorical variables), and their interaction were entered as independent variable (fixed effect), age (as continuous variable) was entered as covariate, participant as random effect. For the comparison of simulated activity blocks and outside lab walks as well as the carpet vs tile walk tasks, the same model was used with methods (APDM/SKDH gait and GAITRite/APDM/SKDH gait), tasks (activity block/outside lab and carpet/tile), and their interaction as independent variables (fixed effect), age as covariate, and participant as random effect. The estimated least square means (lsmeans) or mean difference with 95% CI, F- or t-values, model fit parameters, and degrees of freedom are reported in supplementary materials as well as in the main figures when necessary. The uncorrected p-values are reported unless stated otherwise, then, correction for multiple comparisons was

performed using false discovery rate (FDR). Paired t-tests were performed within the in-lab tasks across methods, within a method across the in-lab tasks, and between visits. One participant's slow walk data for one visit was excluded from analysis, as there was an error in data collection for that task.

The at-home metrics were summarized with the following steps: non-compliant days (lumbar: 10 h of wear time, wrist: 18 h of wear time) and the first and last days of data monitoring (i.e., in-lab visit days) were excluded from the at-home analysis as participants performed an extensive amount of walking tasks during the in-lab visits. The metrics were then averaged/counted within days and averaged across days of wear. As suggested by multiple studies^{15,67–70}, the average within the monitoring period was computed only if there were at least 4 compliant days. One participant's lumbar at-home data was excluded as there were only 3 compliant days, and one participant's wrist data was lost due to device malfunction. For the gait metrics, only gait bouts that were longer than 10 s and with more than four detected gait cycles were included in the analysis to ensure robust gait characterization. The median across the steps within each gait bout was computed, then averaged across the gait bouts within each day, finally each daily summary was averaged across the monitoring window. Only for the wear time computation at-home, non-compliant days were included in the summaries, and transformed into percentages, so that comparison to comfort and wearability questionnaires could be performed more precisely. The comfort and wearability questionnaires responses are summarized using counts and percentages. Two participants had missing data for the at-home comfort and wearability questionnaire for the lumbar device due to an error in the clinical data capture platform setup.

Data availability

Upon request and subject to review, Pfizer will provide the data that support the findings of this study. Subject to certain criteria, conditions, and exceptions, Pfizer may also provide access to the related individual de-identified participant data. See <https://www.pfizer.com/science/clinical-trials/trial-data-and-results> for more information. Please contact the corresponding author, X.C. for the data request.

Received: 8 December 2023; Accepted: 15 July 2024

Published online: 29 August 2024

References

- Hardy, S. E., Perera, S., Roumani, Y. F., Chandler, J. M. & Studenski, S. A. Improvement in usual gait speed predicts better survival in older adults. *J. Am. Geriatr. Soc.* **55**, 1727–1734 (2007).
- Kan, G. A. V. et al. Gait speed at usual pace as a predictor of adverse outcomes in community-dwelling older people an International Academy on Nutrition and Aging (IANA) Task Force. *J. Nutr. Heal. Aging* **13**, 881–889 (2009).
- Hornyak, V., VanSwearingen, J. M. & Brach, J. S. Measurement of gait speed. *Top. Geriatr. Rehabilitation* **28**, 27–32 (2012).
- Pirker, W. & Katzenschlager, R. Gait disorders in adults and the elderly. *Wien. Klin. Wochenschr.* **129**, 81–95 (2017).
- European Medicines Agency. Final Qualification Opinion for Stride velocity 95th centile as primary endpoint in studies in ambulatory Duchenne Muscular Dystrophy studies. https://www.ema.europa.eu/en/documents/scientific-guideline/qualification-opinion-stride-velocity-95th-centile-primary-endpoint-studies-ambulatory-duchenne_en.pdf.
- Adam, C. E. et al. Change in gait speed and fall risk among community-dwelling older adults with and without mild cognitive impairment: A retrospective cohort analysis. *BMC Geriatr.* **23**, 328 (2023).
- Soltani, A. et al. Real-world gait speed estimation, frailty and handgrip strength: A cohort-based study. *Sci. Rep.* **11**, 18966 (2021).
- Cardona-Morrell, M. et al. Vital signs monitoring and nurse–patient interaction: A qualitative observational study of hospital practice. *Int. J. Nurs. Stud.* **56**, 9–16 (2016).
- Fritz, S. & Lusardi, M. White paper: “walking speed: The sixth vital sign”. *J. Geriatr. Phys. Ther.* **2001**(32), 46–49 (2009).
- McCambridge, J., Witton, J. & Elbourne, D. R. Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *J. Clin. Epidemiol.* **67**, 267–277 (2014).
- Landsberger, H. A. Hawthorne Revisited. Management and the Worker, its Critics and Developments in Human Relations in Industry. Cornell Studies in Industrial and Labor Relations, vol. IX. Ithaca, N.Y., Cornell Universit. **25**(2), 164–164. <https://doi.org/10.1017/S1373971900078768> (1959).
- Chang, B.-H., Hendricks, A. M., Slawsky, M. T. & Locastro, J. S. Patient recruitment to a randomized clinical trial of behavioral therapy for chronic heart failure. *BMC Méd. Res. Methodol.* **4**, 8 (2004).
- Nipp, R. D., Powell, E., Chabner, B. & Moy, B. Recognizing the financial burden of cancer patients in clinical trials. *Oncol.* **20**, 572–575 (2015).
- Martin, E. et al. 6th vital sign app: Testing validity and reliability for measuring gait speed. *Gait Posture* **68**, 264–268 (2019).
- Czech, M. D. et al. Age and environment-related differences in gait in healthy adults using wearables. *Npj Digit. Med.* **3**, 127 (2020).
- Crossley, G. H. et al. The CONNECT (Clinical Evaluation of Remote Notification to Reduce Time to Clinical Decision) trial the value of wireless remote monitoring with automatic clinician alerts. *J. Am. Coll. Cardiol.* **57**, 1181–1189 (2011).
- Slotwiner, D. et al. HRS expert consensus statement on remote interrogation and monitoring for cardiovascular implantable electronic devices. *Hear. Rhythm* **12**, e69–e100 (2015).
- Varma, N., Michalski, J., Stambler, B., Pavri, B. B. & Investigators, T. Superiority of automatic remote monitoring compared with in-person evaluation for scheduled ICD follow-up in the TRUST trial— testing execution of the recommendations. *Eur. Hear. J.* **35**, 1345–1352 (2014).
- Simblett, S. K. et al. Patients' experience of wearing multimodal sensor devices intended to detect epileptic seizures: A qualitative analysis. *Epilepsy Behav.* **102**, 106717 (2020).
- Colloud, S. et al. Evolving regulatory perspectives on digital health technologies for medicinal product development. *npj Digit. Med.* **6**, 56 (2023).
- Coran, P. et al. Advancing the use of mobile technologies in clinical trials: Recommendations from the clinical trials transformation initiative. *Digit. Biomark.* **3**, 145–154 (2019).
- Kruizinga, M. D. et al. Development of novel, value-based, digital endpoints for clinical trials: A structured approach toward fit-for-purpose validation. *Pharmacol. Rev.* **72**, 899–909 (2020).
- Stern, A. D. et al. Advancing digital health applications: Priorities for innovation in real-world evidence generation. *Lancet Digit. Heal.* **4**, e200–e206 (2021).
- Mathews, S. C. et al. Digital health: A path to validation. *NPJ Digit. Med.* **2**, 38 (2019).
- Torous, J., Stern, A. D. & Bourgeois, F. T. Regulatory considerations to keep pace with innovation in digital health products. *NPJ Digit. Med.* **5**, 121 (2022).

26. Rego, S. *et al.* Methods for the clinical validation of digital endpoints: Protocol for a scoping review abstract. *JMIR Res. Protoc.* **12**, e47119 (2023).
27. Bertha, A. *et al.* Incorporating digitally derived endpoints within clinical development programs by leveraging prior work. *NPJ Digit. Med.* **6**, 139 (2023).
28. Goldsack, J. C. *et al.* Verification, analytical validation, and clinical validation (V3): The foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). *NPJ Digit. Med.* **3**, 55 (2020).
29. Carthy, M. M. *et al.* from meaningful outcomes to meaningful change thresholds: A path to progress for establishing digital endpoints. *Ther. Innov. Regul. Sci.* **57**, 629–645 (2023).
30. Perry, B. *et al.* How much evidence is enough? Research sponsor experiences seeking regulatory acceptance of digital health technology-derived endpoints. *Digit. Biomark.* **7**, 45–53 (2023).
31. Picerno, P. *et al.* Wearable inertial sensors for human movement analysis: A five-year update. *Expert Rev. Méd. Devices* **18**, 79–94 (2021).
32. Kobsar, D. *et al.* Validity and reliability of wearable inertial sensors in healthy adult walking: A systematic review and meta-analysis. *J. Neuroeng. Rehabilitation* **17**, 62 (2020).
33. Werner, C. *et al.* Validity and reliability of the Apple health app on iPhone for measuring gait parameters in children, adults, and seniors. *Sci. Rep.* **13**, 5350 (2023).
34. Cudejko, T., Button, K. & Al-Amri, M. Validity and reliability of accelerations and orientations measured using wearable sensors during functional activities. *Sci. Rep.* **12**, 14619 (2022).
35. Lanovaz, J. L., Oates, A. R., Treen, T. T., Unger, J. & Musselman, K. E. Validation of a commercial inertial sensor system for spatio-temporal gait measurements in children. *Gait Posture* **51**, 14–19 (2017).
36. Rojer, A. G. M. *et al.* Robustness of in-laboratory and daily-life gait speed measures over one year in high functioning 61- to 70-year-old adults. *Gerontology* **67**, 650–659 (2021).
37. Shahar, R. T. & Agmon, M. Gait analysis using accelerometry data from a single smartphone: Agreement and consistency between a smartphone application and gold-standard gait analysis system. *Sens. (Basel, Switzerland)* **21**, 7497 (2021).
38. Soulard, J., Vaillant, J., Balaguier, R. & Vuillerme, N. Spatio-temporal gait parameters obtained from foot-worn inertial sensors are reliable in healthy adults in single- and dual-task conditions. *Sci. Rep.* **11**, 10229 (2021).
39. Álvarez, M. N. *et al.* Assessing falls in the elderly population using G-STRIDE foot-mounted inertial sensor. *Sci. Rep.* **13**, 9208 (2023).
40. Mao, Y., Ogata, T., Ora, H., Tanaka, N. & Miyake, Y. Estimation of stride-by-stride spatial gait parameters using inertial measurement unit attached to the shank with inverted pendulum model. *Sci. Rep.* **11**, 1391 (2021).
41. Cudejko, T., Button, K. & Al-Amri, M. Wireless pressure insoles for measuring ground reaction forces and trajectories of the centre of pressure during functional activities. *Sci. Rep.* **13**, 14946 (2023).
42. Lemay, J.-F. *et al.* Using wearable sensors to characterize gait after spinal cord injury: Evaluation of test–retest reliability and construct validity. *Spinal Cord* **59**, 675–683 (2021).
43. Rochester, L. *et al.* A roadmap to inform development, validation and approval of digital mobility outcomes: The mobilise-D approach. *Digit. Biomark.* **4**, 13–27 (2020).
44. Bonci, T. *et al.* An objective methodology for the selection of a device for continuous mobility assessment. *Sens. (Basel, Switzerland)* **20**, 6509 (2020).
45. Micó-Amigo, M. E. *et al.* Assessing real-world gait with digital technology? Validation, insights and recommendations from the Mobilise-D consortium. *J. Neuroeng. Rehabilitation* **20**, 78 (2023).
46. Salis, F. *et al.* A multi-sensor wearable system for the assessment of diseased gait in real-world conditions. *Front. Bioeng. Biotechnol.* **11**, 1143248 (2023).
47. Keogh, A. *et al.* Acceptability of wearable devices for measuring mobility remotely: Observations from the Mobilise-D technical validation study. *Digit. Heal.* **9**, 20552076221150744 (2023).
48. Clay, I. *et al.* Reverse engineering of digital measures: Inviting patients to the conversation. *Digit. Biomark.* **7**, 28–44 (2023).
49. Adamowicz, L., Christakis, Y., Czech, M. D. & Adamusiak, T. SciKit digital health: Python package for streamlined wearable inertial sensor data processing. *JMIR mHealth uHealth* **10**, e36762 (2022).
50. Majumder, S., Mondal, T. & Deen, M. J. Wearable sensors for remote health monitoring. *Sensors* **17**, 130 (2017).
51. Mancini, M. & Horak, F. B. Potential of APDM mobility lab for the monitoring of the progression of Parkinson's disease. *Expert Rev. Méd. Devices* **13**, 455–462 (2016).
52. Morris, R. *et al.* Validity of Mobility Lab (version 2) for gait assessment in young adults, older adults and Parkinson's disease. *Physiol. Meas.* **40**, 095003 (2019).
53. Egerton, T., Thingstad, P. & Helbostad, J. L. Comparison of programs for determining temporal-spatial gait variables from instrumented walkway data: PKmas versus GAITRite. *BMC Res. Notes* **7**, 542 (2014).
54. Bagui, S. *et al.* An improved step counting algorithm using classification and double autocorrelation. *Int. J. Comput. Appl.* **44**, 250–259 (2022).
55. Czech, M. *et al.* The impact of reducing the number of wearable devices on measuring gait in Parkinson disease: Noninterventonal exploratory study. *JMIR Rehabilitation Assist. Technol.* **7**, e17986 (2020).
56. Rose, M. J. *et al.* Reliability of wearable sensors for assessing gait and chair stand function at home in people with knee osteoarthritis. *Arthritis. Care Res.* <https://doi.org/10.1002/acr.25096> (2023).
57. Takayanagi, N. *et al.* Relationship between daily and in-laboratory gait speed among healthy community-dwelling older adults. *Sci. Rep.* **9**, 3496 (2019).
58. Hillel, I. *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. Act.* **16**, 6 (2019).
59. Camerlingo, N. *et al.* In-lab assessment of a sternum-worn accelerometer to measure gait speed in healthy individuals (P3-11.010). *Neurology* <https://doi.org/10.1212/wnl.000000000202383> (2023).
60. Pfizer. Assessing Wearable Activity in Relation to Endurance in Heart Failure (AWARE-HF). <https://www.pfizerclinicaltrials.com/awarehf.com>.
61. Weiss, R. G. Physical Activity, Cardiac and Skeletal Muscle Energetics in Healthy Subjects and HFpEF Patients. <https://clinicaltrials.gov/study/NCT04998877>.
62. Brand, Y. E. *et al.* Gait detection from a wrist-worn sensor using machine learning methods: A daily living study in older adults and people with Parkinson's disease. *Sensors* **22**, 7094 (2022).
63. Peraza, L. R. *et al.* an automatic gait analysis pipeline for wearable sensors: A pilot study in Parkinson's disease. *Sensors* **21**, 8286 (2021).
64. Buekers, J. *et al.* Laboratory and free-living gait performance in adults with COPD and healthy controls. *ERJ Open Res.* **9**, 00159–02023 (2023).
65. U.S. Food and Drug Administration. Treatment for Heart Failure: Endpoints for Drug Development Guidance for Industry. <https://www.fda.gov/media/128372/download>.
66. Cicchetti, D. V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess.* **6**, 284–290 (1994).

67. Mueller, A. *et al.* Continuous digital monitoring of walking speed in frail elderly patients: Noninterventional validation study and longitudinal clinical trial. *JMIR mHealth uHealth* **7**, e15191 (2019).
68. Herrmann, S. D., Barreira, T. V., Kang, M. & Ainsworth, B. E. How many hours are enough? Accelerometer wear time may provide bias in daily activity estimates. *J. Phys. Act. Heal.* **10**, 742–749 (2013).
69. Tudor-Locke, C. *et al.* Improving wear time compliance with a 24-hour waist-worn accelerometer protocol in the International study of childhood obesity, lifestyle and the environment (ISCOLE). *Int. J. Behav. Nutr. Phys. Act.* **12**, 11 (2015).
70. Burchartz, A. *et al.* Measurement of physical activity and sedentary behavior by accelerometry among a nationwide sample from the KIGGS and MoMo study: Study protocol. *JMIR Res. Protoc.* **9**, e14370 (2020).

Acknowledgements

Without participants there can be no research; the study team thanks all the people who participated in this study. The team would also like to acknowledge Stephanie An-Stoddard, Ma Bettina Asinias, Tomasz Adamusiak, Bunmi Williams, and Bhavna Adhin for study operations support, programming and data management support, and study monitoring.

Author contributions

Study concept and design: M.S., H.Z., C.D., F.I.K., X.C. Data collection and study execution: M.S., H.Z., X.C., P.G.T., D.P., A.K., J.S., A.M., D.C., S.G. SciKit Digital Health software: L.A. Analysis: F.I.K., L.A., D.P., W.L. Drafting of manuscript: P.S., M.W., F.I.K., X.C. Critical revision of the manuscript for important intellectual content: C.D., L.A., P.S., M.W., F.I.K., X.C., H.Z. All authors had full access to all the study data and approved this manuscript.

Funding

This study was sponsored by Pfizer.

Competing interests

M.W., P.S., M.S., H.Z., C.D., F.I.K., X.C., L.A., P.G.T., D.P., J.S., A.M., D.C., S.G. are current employees of Pfizer, Inc and hold stock and/or stock options. A.K. was a former employee of Pfizer, Inc. W.L. was a biostatistics intern at Pfizer, Inc at the time the study was conducted.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-67675-6>.

Correspondence and requests for materials should be addressed to X.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024