Agreement Between StepWatch3 and ActiGraph wGT3X+ for Measuring Step-Based Metrics in People With Peripheral Artery Disease

Pierre Jéhannin, Alexis Le Faucheur, Ségolène Chaudru, Aline Taoum, Guillaume Mahé, Pierre-Yves de Müllenheim

Agreement between StepWatch3 and ActiGraph wGT3X+ for measuring step-based metrics in people with peripheral artery disease

Pierre Jéhannin [1] Alexis Le Faucheur [2, 3] Ségolène Chaudru [4, 1] Aline Taoum 3 Guillaume Mahé [3, 4, 1] Pierre-Yves de Müllenheim [2, 5, *]

* *Auteur correspondant*

1 - CHU Pontchaillou [Rennes]
2 - APCOSS - Activité Physique, Corps, Sport et Santé
UCO - Université Catholique de l'Ouest
3 - M2S - Laboratoire Mouvement Sport Santé
4 - CIC - Centre d'Investigation Clinique [Rennes]
5 - IFEPSA - Institut Formation Education Physique et Sportive Angers

ABSTRACT

We investigated the agreement between StepWatch3™ (SW3) and ActiGraph™ wGT3X+ monitors for measuring step-based metrics in peripheral artery disease (PAD) patients and older adults. In 23 PAD and 38 older participants, we compared metrics obtained during an outdoor (400-m track) walking session (OWS) (step count) and a 7-day free-living period (step count and 60/30/5/1-min maximal or peak step accumulation) using the SW3 (ankle) and the wGT3X+ (hip) with the low frequency extension filter enabled (wGT3X+/LFE) or not (wGT3X+/N). During OWS, agreement was high, particularly for wGT3X+/LFE: correlations ≥0.98; median absolute percentage errors (MAPEs) <1%; significant equivalence using a ±15%-equivalence zone (EZ) or narrower. In free-living, no wGT3X+ method was equivalent to SW3 for step count. wGT3X+/LFE was equivalent to SW3 regarding all step accumulation metrics using a ±20%-EZ or narrower, with MAPEs <11%. The wGT3X+/LFE method is the best option for comparisons with SW3 in PAD and older adults.

Key-words. Physical activity, accelerometers, pedometers, step rate, intermittent claudication

INTRODUCTION

Lower extremity peripheral artery disease (PAD) represents an important public health problem that is associated with increasing disability and mortality and that affects by now more than 236 million people worldwide (Song et al., 2019). During the past twenty years, several studies conducted in people with PAD have used monitor-based physical activity measurement methods to investigate their usual physical or walking behavior and their relationships with clinical and functional outcomes, or to quantify the training load during home-based walking programs (de Müllenheim, Chaudru, Mahé, Prioux, & Le Faucheur, 2016). The StepWatch3™ (SW3) accelerometer-based pedometer (Orthocare Innovations™ / Modus Health™, LLC, Edmonds, WA) and the ActiGraph™ accelerometers of the GT3X generation (i.e. GT3X, GT3X+, wGT3X+; ActiGraph™, LLC, Pensacola, FL) are research-grade monitors that have been commonly used to count steps in healthy (Bassett, Toth, LaMunion, & Crouter, 2017) and PAD populations (de Müllenheim et al., 2016; Murrow et al., 2019; Parsons et al., 2016; Tew et al., 2015). These monitors are of particular interest in the field of physical activity measurement for health purpose since they have received US Food and Drug Administration clearance as a Class II medical device, and thus appear to be good candidates to be integrated into the standard-of-care for management of certain diseases (Bassett et al., 2017). Moreover, the SW3 is the most accurate pedometer for counting steps during walking (Bassett et al., 2017), and the monitors of the GT3X generation are currently used in several large-scale studies to prospectively investigate various relationships between physical activity and health (Bassett et al., 2017).

Because of the increasing number of studies using either the SW3 or the GT3X generation monitors in PAD, it is important to know how well these monitors provide comparable step-based metrics in this population to allow researchers to make meaningful comparisons between studies and to allow clinicians to better compare their own data that

would have been obtained with a given monitor (e.g., an SW3) with future clinical trials or normative data that would have be obtained using another monitor (e.g., an ActiGraph GT3X generation monitor). Most of the validity studies that have provided results allowing the comparisons between the SW3 and the GT3X generation monitors have been conducted using a treadmill or overground walking, in young adults (Feito, Bassett, & Thompson, 2012; Feito, Garner, & Bassett, 2015; Hickey, John, Sasaki, Mavilia, & Freedson, 2016), overweight and obese individuals (Feito et al., 2012), older adults with diverse disorders (Hergenroeder et al., 2018; Treacy et al., 2017; Webber & St John, 2016), people with Rett syndrome (Downs et al., 2015), and people with multiple sclerosis (Sandroff et al., 2014). Some studies have also conducted measurements during various physical activities (Hickey et al., 2016) and daily living (Feito et al., 2012; Feito et al., 2015; Webber & St John, 2016).

Results from the validity studies that have been conducted in healthy adults and people with various chronic disease conditions may not be directly applicable to people with PAD, because this population present gait modifications when lower-limb pain appears during walking due to arterial insufficiency, such as lower walking speed, shorter stride length, and asymmetric gait to favor the asymptomatic limb during stance phases (Gardner, Montgomery, Ritti-Dias, & Forrester, 2010). Because no study has compared the SW3 and GT3X generation monitors for counting steps in people with PAD, additional work is required to fill this gap. Moreover, the comparisons of these monitors in previous studies were performed by essentially considering total step counts. However, other step-based metrics that may present an interest in studying relationships between physical activity and health (Tudor-Locke et al., 2017) deserve comparisons across different monitors.

The main objective of this study was to investigate the agreement between the SW3 and the GT3X generation monitors (here a wGT3X+) for measuring relevant step-based metrics in people with PAD in two contexts of interest: an outdoor walking session as it could be

performed during a home-based walking program in PAD (Gardner, Parker, Montgomery, Scott, & Blevins, 2011; McDermott et al., 2013) and during which one would want to know the volume of physical activity performed; and a 7-day free-living period as usually performed to determine the daily walking pattern in PAD research (de Müllenheim et al., 2016). Secondary objectives included the followings: (i) explore the influence of PAD on the agreement between the monitors by conducting the same measurements and analyses in a healthy population of the same age as in PAD; (ii) explore the effect of pain occurrence on the agreement between the monitors in PAD during the outdoor level walking session; (iii) determine, where appropriate, correction models to make the step-based metrics from the SW3 and the wGT3X+ more comparable.

METHODS

Study design

To conduct the present study, we used data collected in two populations from two research projects: people with PAD from the "[Blinded for review]" project (NCT[Blinded for review]) and healthy older adults from the "[Blinded for review]" project (NCT[Blinded for review]). Both projects were approved by the local ethics committee of [Blinded for review]. All the PAD and healthy older participants gave their written informed consent to participate. The PAD participants were recruited during their medical appointment at our Vascular Medicine Unit ([Blinded for review]), whereas healthy older adults were recruited from seniors' physical activity associations. PAD participants had to be ≥18 years old, have an ankle-brachial index (ABI) ≤0.90 (or if resting ABI >0.90 and <1.00, a decrease in recovery ankle systolic pressure or in recovery ABI from treadmill exercise higher than 30 mmHg or 20%, respectively), have a maximal walking distance on the treadmill (modified Strandness test: 3.2 km/h, 10% grade) <500 m (as assessed during the medical appointment), and complain of exertional lower-limb pain that sometimes can begin at rest, causes the participant to stop walking, and relieves or lessens within 10 minutes of rest (confirmed during treadmill testing). A full description of the inclusion criteria for the PAD participants is provided elsewhere (Chaudru et al., 2019). The healthy older adults had to be ≥50 years old, have an ABI ≥1.00 and ≤1.40, have no pain in the lower limbs during walking, and have no functional limitation during treadmill walking (15 min at 3.2 km/h and 10% grade).

During the "[Blinded for review]" and "[Blinded for review]" protocols, the PAD and healthy older participants completed an outdoor walking session and a 7-day free-living period while wearing the SW3 monitor (initially released by Orthocare Innovations™ [LLC, Edmonds, WA], now by Modus Health™ [LLC, Edmonds, WA]), and an ActiGraph™ wGT3X+ monitor.

The outdoor walking session

The PAD participants completed an outdoor walking session that consisted of walking on a 400 meters flat athletic track during a 40 – 60 min period (walks and rests included) depending on the capabilities of the participants, while respecting the following instructions: (i) walking at their usual pace; (ii) walking up to maximal claudication pain forces to stop walking; (iii) recovering as usually done during daily life (Le Faucheur et al., 2008). During the session, the PAD participants were equipped with the SW3 and the wGT3X+ monitors. They also wore a DG100 GPS receiver (GlobalSat, Taipei, Taiwan) to obtain speed information. During the session, patients were asked to use a watch to push a specific marker button to record pain manifestations in the lower limbs while walking. Pain-free walking time was the time measured by the investigators using a stopwatch between the start of a walking bout and the time when the patient pushed the button of the watch to indicate the occurrence of pain during walking (Chaudru et al., 2019). Maximal walking time was the duration of a maximal walking bout detected using the GPS speed signals along with a validated data processing methodology as previously performed (Le Faucheur et al., 2008). The actual duration of the session was recorded by the investigators using a chronometer.

The healthy older participants completed an outdoor walking session that was divided into a first part at spontaneous walking speed and a second part at 4 km/h, with both parts performed on the same 400 meters flat athletic track as for the PAD participants. Each part consisted of a first walking period of 6 minutes followed by various randomized walking periods (i.e. 30 s, 1 min, 1.5 min, 2 min) separated by recovery periods. Using such a procedure, we aimed at: (i) having a walking session that would include different walking durations as in PAD participants, since healthy older participants are not limited; (ii) and having a walking speed close to that of the PAD participants (Gernigon et al., 2014). During the session, the

healthy older participants were equipped with the SW3 and the wGT3X+ monitors, and with a Qstarz BT-Q1000XT GPS receiver (Qstarz International Co., Ltd., Taipei, Taiwan) for measuring walking speed.

The 7-day free-living period

Each day of the 7-day free-living period, both the PAD and healthy older participants were asked to put on the monitors when they woke up and to remove them just before going to bed or during water-related activities (e.g., taking a shower). A booklet was given to the participants to provide any comment about the use of the monitors and to indicate when the monitors were removed during the day where appropriate. All the participants were equipped with the monitors at the end of the first visit of their respective protocol and they brought them back at the subsequent visit of the protocol that was scheduled at least seven full days after the first visit. Thus, for all participants, the 7-day free-living measurement was completed before the outdoor walking session.

The wearable monitors

SW3. The SW3 used in the present study is a small (7.5 × 5.0 × 2.0 cm) and light (38 g) accelerometer-based pedometer. For all the participants, this monitor was worn at the right ankle, except for one PAD participant who wore it at the left ankle due to preexisting skin irritation. The SW3 was set to record strides (i.e., ipsilateral steps) using 10-s epochs. The initialization of the SW3 was performed using the preprogrammed "quick start" settings of StepWatch software. The "Walking speed" and "Leg motion" settings were set based on the characteristics of the participants as recommended by the manufacturer.

wGT3X+. The wGT3X+ used in the present study is a small (4.6 × 3.3 × 1.5 cm) and light (19 g) tri-axial accelerometer. In all the participants, this monitor was worn at the right

hip, except for one PAD participant who wore it at the left hip so that it was placed on the same side as the SW3. The wGT3X+ was secured in a nylon pouch linked to a nylon belt and was set to record raw acceleration at a 30-Hz sampling rate (firmware version: 2.2.1 to 2.5.0). The initialization of the wGT3X+ was performed using Actilife 6 software. Following the measurement period, the wGT3X+ data were accumulated into 10-s epochs using Actilife 6 software with either the low frequency extension filter (wGT3X+/LFE) or the normal filter (wGT3X+/N) enabled.

GPS receivers. The DG100 GPS receiver (8.0 × 5.5 × 1.8 cm, weight ~60 g for the unit; Abraham et al. (2012), Noury-Desvaux et al. (2011)) and the Qstarz BT-Q1000XT GPS receiver (7.2 x 4.7 x 2 cm, weight ~65 g for the unit) were used to measure walking speed in the PAD and healthy older participants, respectively. The Qstarz BT-Q1000XT was used in the healthy older participants in place of the DG100 because the DG100 was no longer marketed. Both these GPS receivers have close accuracy for detecting walking and stopping bouts and estimating walking speed in environments with low level of obstruction (Taoum et al., 2021). The GPS receivers were set at a 1-Hz recording rate using a personal computer and the corresponding manufacturer's software (DG100: Data Logger Utility, version 1.1; Qstarz BT-Q1000XT: Q-Travel, version 1.53.000).

For both the outdoor walking session and the 7-day free-living period, all devices were systematically initialized using the Coordinated Universal Time to ensure the synchronization of the device recordings.

Data analysis

The timestamped SW3 strides, the wGT3X+/LFE and wGT3X+/N activity counts and steps, and the DG100 or Qstarz speed, were exported to spreadsheets for further analysis in R software (versions 3.5.3/3.6.0). For all step-related analyses, we aligned and synchronized the

three activity monitor data files (SW3, wGT3X+/LFE, wGT3X+/N) based on the timestamp of the files, and we obtained the SW3 steps by doubling the number of the SW3 strides.

The outdoor walking session. We calculated the total number of steps recorded from the activity monitor (SW3, wGT3X+/LFE, wGT3X+/N) over the entire walking session. Moreover, to investigate the effect of lower-limb pain symptoms occurrence on the agreement between the monitors in the PAD participants, we calculated from the activity monitor data files the total step count corresponding to walking without pain and to walking with pain during the first maximal walking bout of the session. Only the first walking bout was considered to attempt to capture the potential effect of pain on the agreement between the monitors while minimizing the potential additional effect of fatigue that could have been introduced throughout the repetition of the maximal walking bouts. To better interpret the potential differences in the step counts obtained from the different methods, we determined the mean walking speed adopted over the walking session as well as the mean walking speed related to walking without pain and to walking with pain during the first maximal walking bout of the session for the PAD participants. Speed data from the GPS devices were processed according to the method proposed by Le Faucheur et al. (2007) implemented in R software.

The 7-day free-living period. We analyzed the first seven full days of measurement. The 10-s epochs of the merged data files were collapsed into 1-min epochs and marked for wear and non-wear-time using the wGT3X+/LFE vector magnitude (VM) counts and the Choi et al. algorithm (Choi, Ward, Schnelle, & Buchowski, 2012) with the "PhysicalActivity" R package (version 0.2-2, 2018). Thereafter, a graphical inspection of the data related to each day was implemented to check that: (i) wear-time periods corresponded to waking-time periods only (and not to in-bed night-time periods, which could be the case when the participant forgot to remove the wGT3X+, and also possibly the SW3, before going to bed at night); (ii) each participant wore the SW3 in addition to the wGT3X+ at the beginning of her/his first out-of-

bed waking-time period of the day. All the epochs that were initially related to wGT3X+ wear-time were reconsidered as non-wear-time when they corresponded to in-bed night-time periods or when it was determined that both monitors were not simultaneously worn at the beginning of the day. This procedure aimed to remove "artificial" extra wGT3X+ steps compared to the SW3 and thus obtain more accurate comparisons of the SW3 and wGT3X+ step-based metrics. The details of the graphical inspection procedure are provided and illustrated in the supplemental materials (eText and eFigure). Finally, the 1-min epochs with visually abnormal step counts were removed (a 1-min epoch with 542 SW3 steps from a PAD participant dataset and another 1-min epoch with 290 SW3 steps from another PAD participant dataset were removed).

We considered a day as valid if there was both a wGT3X+/LFE wear time ≥600 min (Migueles et al., 2017) and an SW3 step count ≥1500 (Tudor-Locke, Barreira, & Schuna, 2015). Then, using the valid days and the wear time epochs, we calculated the daily means for the following step-based metrics from each method: the steps per day, the maximum step accumulation over a sliding window of 60, 30, and 5 continuous minutes, and the peak step accumulation over the top 60, 30, and 1 minute(s) of the day that could be continuous or discontinuous. These metrics have been chosen as they have been regularly used in PAD studies (de Müllenheim et al., 2016) or they might present an interest in epidemiologic studies (Tudor-Locke et al., 2017).

Statistical analysis

Descriptive statistics. Statistical analyses were conducted using R software. The quantitative variables are shown as means ± standard deviations (SD) when normally distributed (as inferred from Shapiro-Wilk tests) and as medians (25^{th} – 75^{th} percentiles)

otherwise. For consistency purposes with specific inferential analyses, some variables that were skewed are shown as geometric means ×/÷ factors for SD.

Agreement between the monitors. First, we conducted a correlation analysis. When the hypothesis of bivariate normality (tested using the Royston's method implemented in the "MVN" package (Korkmaz, Goksuluk, & Zararsiz, 2014)) concerning the SW3 and the wGT3X+ data (or the log-transformed data when necessary) was not rejected (P>0.05), we used the Pearson coefficient (R) and its corresponding 95% confidence interval (CI). When the bivariate normality hypothesis was rejected even when using the log-transformed data, the Spearman rho coefficient with its 95% jackknife empirical likelihood CI was computed using the "spearmanCI" package (version 1.0, 2018). Correlation coefficients with 95% CIs that did not include zero were considered as statistically significant.

Second, we calculated the median percent error (MPE) and the median absolute percent error (MAPE), with error computed as wGT3X+ minus SW3, as previously described (DeShaw et al., 2018). MPE and MAPE are shown along with their respective 95% bias-corrected and accelerated bootstrap (BCa) CIs (obtained with 999 bootstrap replicates) that were calculated using the "rcompanion" package (version 2.3.25, 2020) (Mangiafico, 2016). MPE and MAPE with CIs that did not include zero were considered statistically significant.

Third, we conducted 95% equivalence testing (Dixon et al., 2018). In the present study, the wGT3X+ was considered as significantly equivalent to the SW3 when the 90% CI for the wGT3X+ means fell into a given equivalence zone defined for the SW3 (±10%, ±15%, or ±20% of the SW3 mean score). While the ±10% equivalence zone has been classically used, the ±15% and ±20% equivalence zones have been considered as "viable targets" for the measurement of physical activity (35). The 90% CIs for the wGT3X+ means were calculated from a mixed model with an unstructured covariance matrix (Baguley, 2012). The model used

the considered step-based metric as the response variable, the wGT3X+ method (wGT3X+/LFE; wGT3X+/N) as a fixed predictor, and the participant as a random predictor.

Effect of lower-limb pain occurrence on the agreement between the monitors during outdoor walking in PAD. We studied the effect of walking pain on the agreement between the SW3 and the wGT3X+ for step counting in PAD using the data related to the first walking bout of the outdoor walking session. For the wGT3X+/LFE and wGT3X+/N methods, we calculated, for each participant, the difference of absolute percent error (APE) between walking with pain and walking without pain. Then, for a given wGT3X+ method, we computed the 95% BCa CI for the median of the APE differences. The percent errors related to walking without pain and with pain were also calculated to show the direction of the error (underestimation or overestimation) when walking without pain and with pain. We also calculated the 95% normal CI for the mean change in speed after pain occurrence, considering that the change would be substantial if the CI would fall outside the range of ±0.2 km/h (that is, two times the typical error related to the estimation of walking speed using the DG100 (Noury-Desvaux et al., 2011)).

Correction models. We developed correction models for the wGT3X+ as compared with the SW3 regarding the step-based metrics for which there was a clear need for correction, that is for which there was no significant equivalence for both wGT3X+ methods, even when using the widest equivalence zone tested (i.e., ±20% of the SW3 mean score). Linear regression models were constructed to make predictions of the considered SW3 metric from either the wGT3X+/LFE or the wGT3X+/N method, in the PAD and the healthy older participants separately. Each model and the corresponding R^2 coefficient and root-mean-squared error (RMSE) were computed using a leave-one-out procedure implemented using the "caret" package (version 6.0-85, 2020) (Kuhn, 2008).

RESULTS

Twenty-three PAD participants (3 women / 20 men; age: 60 ± 10 y, height: 1.69 ± 0.08 m, weight: 79.0 ± 16.2 kg, body mass index: 27.36 ± 4.71 $kg \cdot m^{-2}$, ABI: 0.61 (0.53 – 0.74)), and 38 healthy older participants (14 women / 24 men; age: 64 ± 6 y, height: 1.69 ± 0.10 m, weight: 68.7 ± 12.0 kg, body mass index: 23.89 ± 3.01 $kg \cdot m^{-2}$, ABI: 1.19 (1.00 – 1.38)), were studied. All PAD participants were limited during the treadmill test due to leg pain.

The results regarding the agreement between the monitors for step counting during the outdoor walking session are shown in Table 1 and Figure 1. The median (25th – 75th percentiles) duration of the session was of 52.5 (48.6 – 55.4) minutes. One PAD participant and three healthy older participants had incomplete step data due to lack of monitors availability. Moreover, the walking speed was missing for one healthy older participant due to an error when downloading the data from the GPS receiver. In the PAD participants, the wGT3X+/LFE and wGT3X+/N methods were both significantly and nearly perfectly correlated to the SW3 (R ≥0.99). However, as shown in Table 1, the wGT3X+/N method significantly underestimated the SW3 with an MPE of ~-5%, while the wGT3X+/LFE method reached an MPE and a MAPE inferior to 1%. Figure 1 shows that the wGT3X+/LFE and wGT3X+/N methods were significantly equivalent to the SW3 when considering equivalence zones of respectively ±15% and ±20% of the SW3 mean score. In the healthy older participants, for whom the session walking speed (4.9 [4.7 – 5.1] km/h) was slightly higher than in the PAD participants (4.2 ± 0.4 km/h), the wGT3X+ methods were also significantly and highly correlated to the SW3 (rho ≥0.97). The wGT3X+/N method significantly underestimated the SW3 with an MPE of only -1.4%, and the wGT3X+/LFE method, as for the PAD participants, reached an MPE and a MAPE inferior to 1% (Table 1). Both the wGT3X+ methods were significantly equivalent to the SW3 when considering the equivalence zone of ±10%.

Figure 2 shows the results regarding the effect of pain on the agreement between the monitors for step counting during the first walking bout of the outdoor walking session in the PAD participants. The results are related to 17 participants who indicated the occurrence of pain during the walking bout. For both the wGT3X+/LFE and the wGT3X+/N methods, the median of the APE differences between walking with pain and without pain was not significantly different from zero, suggesting no effect of pain on the agreement between the monitors for each of the wGT3X+ methods. However, we could observe that a few PAD participants had an APE that markedly increased when walking with pain in comparison with walking without pain with the wGT3X+/N method (Figure 2, panel B). Figure 2 (panel D) also shows that the change in walking speed was not substantial.

The results regarding the agreement between the monitors for step counting during daily living are shown in Table 2 and Figure 3. Twenty-one PAD participants had complete data after data analysis using graphical inspection. In the PAD participants, the wGT3X+ methods yielded significant and high correlations with the SW3 (R and rho ≥ 0.90) for all the step-based metrics. However, the wGT3X+/LFE method significantly overestimated the SW3 daily step count by more than 20%, while it reached a MAPE <11% for all the other step-based metrics, with a lower error for the metrics reflecting a higher intensity of walking. The wGT3X+/N method significantly underestimated the SW3 daily step count by 45%, and significantly underestimated all the SW3 step-based metrics by at least 15% except for the peak 1-min step accumulation with an underestimation by less than 10%. Both the wGT3X+ methods were not significantly equivalent to the SW3 for the daily step count, but the wGT3X+/LFE method was significantly equivalent to the SW3 for all the other step-based metrics when considering an equivalence zone of ±20%, and for the maximum 5-min step accumulation and the peak 60/30/1-min step accumulation when considering an equivalence zone of ±10% (Figure 3). The wGT3X+/N method was significantly equivalent to the SW3 only for the peak 1-min step

accumulation (using an equivalence zone of ±15%). In the healthy older participants, the results were similar to those obtained in the PAD participants with the following noticeable exceptions: correlations between the wGT3X+ methods and the SW3 were <0.80 for the peak 1-min step accumulation; the error for the wGT3X+/N method consistently seemed lower than in the PAD participants with a MAPE <15% for all the step-based metrics except the daily step count; the wGT3X+/N method was significantly equivalent to the SW3 for all the step-based metrics (when considering an equivalence zone of ±20%) except for the daily step count and for the maximum 60-min step accumulation.

Basing on the large disagreement between the monitors regarding daily step count, we constructed linear models to correct the wGT3X+/LFE or wGT3X+/N daily step count as compared with the SW3 daily step count (Table 3). While the wGT3X+/LFE and wGT3X+/N models reached quite similar accuracies in both the PAD and the healthy older participants, the wGT3X+/N model was the most accurate in the PAD participants, while this was the wGT3X+/LFE model that performed better in the healthy older participants.

DISCUSSION

The present study adds several new results to the current literature for comparing step-based metrics from the SW3 and the GT3X generation monitors (Downs et al., 2015; Feito et al., 2012; Feito et al., 2015; Hergenroeder et al., 2018; Hickey et al., 2016; Sandroff et al., 2014; Treacy et al., 2017; Webber & St John, 2016). Indeed, these results are unique in the PAD population, as well as in the general population regarding the comparisons of these monitors for several step-based metrics that may be of interest from clinical and epidemiological perspectives. Moreover, a particularity of the present study is that we provide results in two contexts where monitors are likely to be of interest for practitioners and researchers: an outdoor walking session where monitors can be used to quantify the training load; and a free-living period where monitors can be used to characterize the usual walking behavior.

Our results suggest that the GT3X generation monitors, when worn at the hip, provide step counts that can be easily compared to those of the SW3 in the context of an outdoor level walking session, regardless of the method used (i.e., enabling the LFE filter or not), in both the PAD and the healthy older populations represented in the present study. Thus, the presence of PAD did not seem to have a marked influence on the agreement between the monitors during the outdoor walking session as compared with the healthy older adults, although percent errors seemed slightly higher in the PAD participants when using the normal filter. This high level of agreement between the monitors, in particular in terms of percent errors, can be explained by the participants mean walking speeds that fell into the range of walking speeds (speeds ≥4 km/h or 67 m/min) for which the GT3X generation monitors can provide step count estimates that are close to those of the SW3 (Feito et al., 2015). However, these results should be used with caution because our PAD participants walked faster than that previously reported by Gernigon et al. (Gernigon et al., 2014) in a quite large sample of people with PAD ($N = 203$; median: 3.6

km/h). Moreover, our healthy older participants were likely to have a higher physical condition than in the normal population of the same age. These considerations are important because walking slower than 4 km/h is likely to be associated with the range of the combinations of acceleration frequency and amplitude thresholds for which Actilife software does not count steps, at least when the ActiGraph™ LFE filter is not enabled (John, Morton, Arguello, Lyden, & Bassett, 2018). When conducting measurements with the GT3X generation monitors during an outdoor walking session, enabling the LFE filter could be the best choice to get step counts that are comparable to those of the SW3. Indeed, the LFE filter allows to get closer estimates to those of the SW3 at walking speeds that are ≥4 km/h (Feito et al., 2015), and it allows to get step counts that are likely to remain close to those of the SW3 at slow walking speeds contrary to the normal filter (Feito et al., 2015).

When considering the daily living context, the level of the agreement between the monitors varied according to the investigated step-based metric and the use of the LFE filter or not along with the wGT3X+. No wGT3X+ method was significantly equivalent to the SW3 for measuring daily step count in the present study. This is consistent with the results from the study by Feito et al. (2015) who found a large positive MPE for the GT3X/LFE method (+36%) and a large negative MPE for the GT3X/N method (-25%) in young adults. As previously discussed (Toth et al., 2018), the LFE filter likely allows the capture of artifact movements during the day due to the extension of the range of the movement frequencies that the monitor can detect. On the opposite, the use of the normal filter does not allow to correctly capture all the daily living steps, likely because a large part of these steps may be performed at frequencies and at walking speeds that are too low (when walking at home for example) for the sensitivity of the normal filter. Although the GT3X monitors have been shown to be less accurate in step count estimation than the SW3, the latter has also been reported to show a lower accuracy during some daily-life activities such as cycling or running (Hickey et al., 2016; Karabulut,

Crouter, & Bassett, 2005). This could also contribute to the weaker between-monitor agreement during the free-living period in comparison with the outdoor walking session. Interestingly, a better agreement between the monitors was obtained for daily living regarding the step-based metrics related to the maximum and the peak step accumulation metrics in both our PAD and healthy older participants, particularly when using the LFE filter. Thus, while the daily step counts from GT3X generation monitors should be corrected to be compared with those of the SW3, it seems that no important correction is needed for the step accumulation-based metrics if the LFE filter is enabled, particularly regarding the maximum 5-min step accumulation and the peak (1 to 60-min) step accumulation metrics. The better agreement related to these last metrics (compared with the daily step count) and obtained when using the LFE filter enabled, could be explained by the fact that the steps taken in these cases reflect the best physical performances of the day in terms of step accumulation, which could be more likely related to actual walking periods, thus limiting the counting of artifact extra step counts in the calculation of the metric. When using the normal filter, the observed better agreement could be explained by the fact that the steps performed in these situations were likely to be related to the top part of the spectrum of the walking speeds adopted by the participant over the day, these speeds falling into the range of speeds for which the normal filter allows an accurate capture of steps.

The occurrence of pain during the outdoor walking sessions in the PAD participants did not modify the agreement between the monitors for step counting. This result could be related to a decrease in mean walking speed that was not substantial, leading to a mean walking speed of 4.0 km/h after pain occurrence. Thus, speed remained sufficiently high to not alter the ability of the wGT3X+ to capture steps, in particular when using the normal filter. However, two PAD participants had a marked increase in APE after pain occurrence (Figure 2) when using the normal filter, which could not be fully explained by a decrease in speed because other

participants with similar walking speeds and a similar decrease in walking speed after pain occurrence did not show such increases in APE. This suggests the potential influence of other gait-related mechanisms, such as a reduction in cadence or asymmetric swing and stance phases related to the painful limb (Gardner, Montgomery, et al., 2010), that could affect the capability of the GT3X generation monitors to detect steps, at least when using the normal filter, while the accuracy of the SW3 is more likely to remain high in patients with gait impairment (Sandroff et al., 2014; Webber & St John, 2016). Further studies are required to state the potential effects of such gait impairments on pedometer step counts accuracy.

The present study has confirmed the particular need to correct the GT3X generation monitor daily step count when we want to compare it with that of the SW3. We constructed separate models for the PAD and the healthy older participants. This was justified as the models we obtained showed different intercepts and slopes, meaning different correction needs depending on the population tested and the data filter used (LFE or normal). Our results suggest that if the daily step count is the only metric of interest in PAD, using the normal filter along with a GT3X generation monitor, and then correcting the daily step count with the proposed equation, would reach a better agreement with the SW3 daily step count as compared with the use of the LFE filter and the corresponding equation. However, if a GT3X generation monitor would be used with the intention to study other step-based metrics and to conduct comparisons with the SW3 in PAD, it would be more valuable to enable the LFE filter since there is better equivalence with the SW3 for various metrics when using this method as compared with the use of the normal filter. In the healthy older participants, using the LFE filter seems to be the best option to get step-based metrics comparable to those of the SW3 since the performance of the correction model for daily step count was slightly higher when using the LFE filter and the

agreement with the SW3 was better for all the other step-based metrics when using the LFE filter.

The main limit of the present study is that it was exploratory and based on convenience samples. Thus, several of our results should be confirmed with confirmatory studies with adequate statistical power. Other limits regarding secondary objectives could also be evocated. First, the overground walking session protocols implemented with the PAD and the healthy older participants were not the same, thus the comparisons between the results obtained for the PAD and the healthy older participants during the overground walking session protocol should be used with caution. Second, we were not able to distinguish the effect of speed from the potential effects of gait abnormalities and pain on the agreement between the monitors in the PAD participants. Future studies could investigate the independent and combined effects of walking speed, gait features, and pain on the agreement between various monitors in PAD, particularly using a range of low walking speeds, to provide information about the comparability of different monitors in PAD patients with various walking impairments levels. Third, the sex distributions in both the PAD and the healthy older participants were very unequal in terms of sex distribution, with substantially more men within each group and overall (13% of women in the PAD participants and 37% of women in the healthy older adults). It could be questioned whether this may affect the external validity of our results since it has been shown that women have gait features (e.g., walking speed) that are different from men (Frimenko, Goodyear, & Bruening, 2015; Gardner, Parker, et al., 2010), which could potentially affect monitor accuracy for counting steps. However, to our knowledge, no gender effect on monitor accuracy during walking has been reported (Leicht & Crowther, 2007, 2009; Shepherd, Toloza, McClung, & Schmalzried, 1999). Fourth, although we developed an approach to ensure a simultaneous wearing of both devices, we acknowledge that without an available validated algorithm to detect wear and non-wear periods from the SW3, it cannot

totally be ruled out that short periods of non-wear with the SW3 were missed. We are not aware of studies that proposed a more robust methodology when using the SW3 and the GT3X/+ concurrently (Feito et al., 2012; Feito et al., 2015). Fifth, our correction models to predict SW3 daily step count from wGT3X+ step count were developed using samples that could be considered as too small to provide accurate predictions across the spectrum of the predicted values, particularly from the perspective of some guidelines that have been proposed for developing a clinical prediction model (Riley et al., 2020). However, our analysis was exploratory only, and the statistics we have provided, in particular R^2 coefficients, could be valuable information for planning sample size in view of building an accurate prediction model (Riley et al., 2020). Moreover, some authors (Wanner, Martin, Meier, Probst-Hensch, & Kriemler, 2013) have shown that building a correction model using a relatively small sample (N=33) can be valuable for substantially reducing mean between-method difference for step counting at the group level. Importantly, the ability of our models to reduce mean difference between the SW3 and the wGT3X+ monitors for step counting should be tested in additional studies.

CONCLUSION

The SW3 and the wGT3X+ monitors provided comparable step counts during an outdoor walking session in our samples of PAD and healthy older participants. However, the monitors did not provide comparable daily step counts during a 7-day free-living measurement in both populations. Interestingly, when considering daily step accumulation metrics, the monitors became more comparable, with better equivalence for the metrics representing a higher walking intensity.

When using a GT3X generation monitor to describe daily physical behavior using various step-based metrics to get results comparable to those of the SW3, enabling the LFE filter is likely to be the best option to choose, both in PAD and in healthy older populations.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Abraham, P., Noury-Desvaux, B., Gernigon, M., Mahe, G., Sauvaget, T., Leftheriotis, G., & Le Faucheur, A. (2012). The inter- and intra-unit variability of a low-cost GPS data logger/receiver to study human outdoor walking in view of health and clinical studies. *PLoS One, 7*(2), e31338.

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for ANOVA. *Behav Res Methods, 44*(1), 158-175.

Bassett, D. R., Jr., Toth, L. P., LaMunion, S. R., & Crouter, S. E. (2017). Step counting: a review of measurement considerations and health-related applications. *Sports Med, 47*(7), 1303-1315.

Chaudru, S., Jehannin, P., de Müllenheim, P. Y., Klein, H., Jaquinandi, V., Mahe, G., & Le Faucheur, A. (2019). Using wearable monitors to assess daily walking limitations induced by ischemic pain in peripheral artery disease. *Scand J Med Sci Sports, 29*(11), 1813-1826.

Choi, L., Ward, S. C., Schnelle, J. F., & Buchowski, M. S. (2012). Assessment of wear/nonwear time classification algorithms for triaxial accelerometer. *Med Sci Sports Exerc, 44*(10), 2009-2016.

de Müllenheim, P. Y., Chaudru, S., Mahé, G., Prioux, J., & Le Faucheur, A. (2016). Clinical interest of ambulatory assessment of physical activity and walking capacity in peripheral artery disease. *Scand J Med Sci Sports, 26*(7), 716-730.

DeShaw, K. J., Ellingson, L., Bai, Y., Lansing, J., Perez, M., & Welk, G. (2018). Methods for activity monitor validation studies: an example with the Fitbit Charge. *J Meas Phys Behav, 1*, 130-135.

Dixon, P. M., Saint-Maurice, P. F., Kim, Y., Hibbing, P., Bai, Y., & Welk, G. J. (2018). A primer on the use of equivalence testing for evaluating measurement agreement. *Med Sci Sports Exerc, 50*(4), 837-845.

Downs, J., Leonard, H., Jacoby, P., Brisco, L., Baikie, G., & Hill, K. (2015). Rett syndrome: establishing a novel outcome measure for walking activity in an era of clinical trials for rare disorders. *Disabil Rehabil, 37*(21), 1992-1996.

Feito, Y., Bassett, D. R., & Thompson, D. L. (2012). Evaluation of activity monitors in controlled and free-living environments. *Med Sci Sports Exerc, 44*(4), 733-741.

Feito, Y., Garner, H. R., & Bassett, D. R. (2015). Evaluation of ActiGraph's low-frequency filter in laboratory and free-living environments. *Med Sci Sports Exerc, 47*(1), 211-217.

Frimenko, R., Goodyear, C., & Bruening, D. (2015). Interactions of sex and aging on spatiotemporal metrics in non-pathological gait: a descriptive meta-analysis. *Physiotherapy, 101*(3), 266-272.

Gardner, A. W., Montgomery, P. S., Ritti-Dias, R. M., & Forrester, L. (2010). The effect of claudication pain on temporal and spatial gait measures during self-paced ambulation. *Vasc Med, 15*(1), 21-26.

Gardner, A. W., Parker, D. E., Montgomery, P. S., Khurana, A., Ritti-Dias, R. M., & Blevins, S. M. (2010). Gender differences in daily ambulatory activity patterns in patients with intermittent claudication. *J Vasc Surg, 52*(5), 1204-1210.

Gardner, A. W., Parker, D. E., Montgomery, P. S., Scott, K. J., & Blevins, S. M. (2011). Efficacy of quantified home-based exercise and supervised exercise in patients with intermittent claudication: a randomized controlled trial. *Circulation, 123*(5), 491-498.

Gernigon, M., Le Faucheur, A., Noury-Desvaux, B., Mahe, G., Abraham, P., & Post-GPS Study Coinvestigators Group. (2014). Applicability of global positioning system for

the assessment of walking ability in patients with arterial claudication. *J Vasc Surg, 60*(4), 973-981 e971.

Hergenroeder, A. L., Barone Gibbs, B., Kotlarczyk, M. P., Kowalsky, R. J., Perera, S., & Brach, J. S. (2018). Accuracy of objective physical activity monitors in measuring steps in older adults. *Gerontol Geriatr Med, 4*, 1-5.

Hickey, A., John, D., Sasaki, J. E., Mavilia, M., & Freedson, P. (2016). Validity of activity monitor step detection is related to movement patterns. *J Phys Act Health, 13*(2), 145-153.

John, D., Morton, A., Arguello, D., Lyden, K., & Bassett, D. (2018). "What is a step?" Differences in how a step is detected among three popular activity monitors that have impacted physical activity research. *Sensors (Basel), 18*(4).

Karabulut, M., Crouter, S. E., & Bassett, D. R., Jr. (2005). Comparison of two waist-mounted and two ankle-mounted electronic pedometers. *Eur J Appl Physiol, 95*(4), 335-343.

Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal, 6*(2), 151-162.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 28*(5), 1-26.

Le Faucheur, A., Abraham, P., Jaquinandi, V., Bouye, P., Saumet, J. L., & Noury-Desvaux, B. (2007). Study of human outdoor walking with a low-cost GPS and simple spreadsheet analysis. *Med Sci Sports Exerc, 39*(9), 1570-1578.

Le Faucheur, A., Abraham, P., Jaquinandi, V., Bouye, P., Saumet, J. L., & Noury-Desvaux, B. (2008). Measurement of walking distance and speed in patients with peripheral arterial disease: a novel method using a global positioning system. *Circulation, 117*(7), 897-904.

Leicht, A. S., & Crowther, R. G. (2007). Pedometer accuracy during walking over different surfaces. *Med Sci Sports Exerc, 39*(10), 1847-1850.

Leicht, A. S., & Crowther, R. G. (2009). Influence of non-level walking on pedometer accuracy. *J Sci Med Sport, 12*(3), 361-365.

Mangiafico, S. S. (2016). Summary and analysis of extension program evaluation in R, version 1.18.1. URL https://rcompanion.org/.

McDermott, M. M., Liu, K., Guralnik, J. M., Criqui, M. H., Spring, B., Tian, L., . . . Rejeski, W. J. (2013). Home-based walking exercise intervention in peripheral artery disease: a randomized clinical trial. *JAMA, 310*(1), 57-65.

Migueles, J. H., Cadenas-Sanchez, C., Ekelund, U., Delisle Nystrom, C., Mora-Gonzalez, J., Lof, M., . . . Ortega, F. B. (2017). Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. *Sports Med, 47*(9), 1821-1845.

Murrow, J. R., Brizendine, J. T., Djire, B., Young, H.-J., Rathbun, S., Nilsson, K. R., Jr., & McCully, K. K. (2019). Near infrared spectroscopy-guided exercise training for claudication in peripheral arterial disease. *Eur J Prev Cardiol, 26*(5), 471-480.

Noury-Desvaux, B., Abraham, P., Mahe, G., Sauvaget, T., Leftheriotis, G., & Le Faucheur, A. (2011). The accuracy of a simple, low-cost GPS data logger/receiver to study outdoor human walking in view of health and clinical studies. *PLoS One, 6*(9), e23027.

Parsons, T. J., Sartini, C., Ellins, E. A., Halcox, J. P., Smith, K. E., Ash, S., . . . Jefferis, B. J. (2016). Objectively measured physical activity and sedentary behaviour and ankle brachial index: Cross-sectional and longitudinal associations in older men. *Atherosclerosis, 247*, 28-34.

Riley, R. D., Ensor, J., Snell, K. I. E., Harrell, F. E., Jr., Martin, G. P., Reitsma, J. B., . . . van Smeden, M. (2020). Calculating the sample size required for developing a clinical prediction model. *BMJ, 368*, m441.

Sandroff, B. M., Motl, R. W., Pilutti, L. A., Learmonth, Y. C., Ensari, I., Dlugonski, D., . . . Riskin, B. J. (2014). Accuracy of StepWatch and ActiGraph accelerometers for measuring steps taken among persons with multiple sclerosis. *PLoS One, 9*(4), e93511.

Shepherd, E. F., Toloza, E., McClung, C. D., & Schmalzried, T. P. (1999). Step activity monitor: increased accuracy in quantifying ambulatory activity. *J Orthop Res, 17*(5), 703-708.

Song, P., Rudan, D., Zhu, Y., Fowkes, F. J. I., Rahimi, K., Fowkes, F. G. R., & Rudan, I. (2019). Global, regional, and national prevalence and risk factors for peripheral artery disease in 2015: an updated systematic review and analysis. *Lancet Glob Health, 7*(8), e1020-e1030.

Taoum, A., Chaudru, S., de Mullenheim, P. Y., Congnard, F., Emily, M., Noury-Desvaux, B., . . . Le Faucheur, A. (2021). Comparison of activity monitors accuracy in assessing intermittent outdoor walking. *Med Sci Sports Exerc, in press*.

Tew, G. A., Humphreys, L., Crank, H., Hewitt, C., Nawaz, S., Al-Jundi, W., . . . Gorely, T. (2015). The development and pilot randomised controlled trial of a group education programme for promoting walking in people with intermittent claudication. *Vasc Med, 20*(4), 348-357.

Toth, L. P., Park, S., Springer, C. M., Feyerabend, M. D., Steeves, J. A., & Bassett, D. R. (2018). Video-Recorded Validation of Wearable Step Counters under Free-living Conditions. *Med Sci Sports Exerc, 50*(6), 1315-1322.

Treacy, D., Hassett, L., Schurr, K., Chagpar, S., Paul, S. S., & Sherrington, C. (2017). Validity of different activity monitors to count steps in an inpatient rehabilitation setting. *Phys Ther, 97*(5), 581-588.

Tudor-Locke, C., Barreira, T. V., & Schuna, J. M., Jr. (2015). Comparison of step outputs for waist and wrist accelerometer attachment sites. *Med Sci Sports Exerc, 47*(4), 839-842.

Tudor-Locke, C., Schuna, J. M., Jr., Han, H. O., Aguiar, E. J., Green, M. A., Busa, M. A., . . . Johnson, W. D. (2017). Step-based physical activity metrics and cardiometabolic risk: NHANES 2005-2006. *Med Sci Sports Exerc, 49*(2), 283-291.

Wanner, M., Martin, B. W., Meier, F., Probst-Hensch, N., & Kriemler, S. (2013). Effects of filter choice in GT3X accelerometer assessments of free-living activity. *Med Sci Sports Exerc, 45*(1), 170-177.

Webber, S. C., & St John, P. D. (2016). Comparison of ActiGraph GT3X+ and StepWatch step count accuracy in geriatric rehabilitation patients. *J Aging Phys Act, 24*(3), 451-458.

16 **TABLES**

17 **Table 1.** Agreement between the SW3 and wGT3X+ step counts related to the outdoor walking session

	People with PAD (*N* = 22)			**Older adults (*N* = 35)**		
	Session walking speed (km/h) : 4.2 ± 0.4			Session walking speed (km/h) : 4.9 (4.7 – 5.1)[a]		
	SW3	**wGT3X+/LFE**	**wGT3X+/N**	**SW3**	**wGT3X+/LFE**	**wGT3X+/N**
Mean ± SD (or ×/÷ factor)	3879 ± 1373	3879 ± 1370	3694 ± 1393	2455 ×/÷ 1.1	2480 ×/÷ 1.1	2406 ×/÷ 1.1
Corr. coeff. [95% CI]		1.00 [1.00; 1.00]	0.99 [0.98; 1.00]		*0.98 [0.94; 1.00]*	*0.97 [0.94; 1.00]*
MPE [95% CI], %		-0.1 [-0.3; 0.4]	-5.2 [-7.3; -2.8]		0.8 [0.2; 1.2]	-1.4 [-1.6; -1.2]
MAPE [95% CI], %		0.5 [0.2; 0.8]	5.2 [2.8; 7.0]		0.8 [0.3; 1.2]	1.4 [1.1; 1.6]

Note. Descriptive statistics are means ± SD, medians (25^{th} – 75^{th} percentiles), or geometric means ×/÷ factors for SD. *N* = Number of the participants available for the analysis; Corr. coeff. = correlation coefficient (values in italics are Spearman rho coefficients; otherwise, values are Pearson r coefficients); MPE = median percent error; MAPE = median absolute percent error; 95% CI = 95 % confidence interval.

[a]Based on 34 participants because data were lost for one participant.

18

19 **Table 2.** Agreement between the SW3 and wGT3X+ step-based metrics related to the 7-day free-living period

	People with PAD (*N* = 21) Number of valid days : 7 (6 – 7) Mean daily wear time (min) : 764 (725 – 840)			**Older adults (*N* = 32)** Number of valid days : 7 (6 – 7) Mean daily wear time (min) : 856 ± 57		
	SW3	**wGT3X+/LFE**	**wGT3X+/N**	**SW3**	**wGT3X+/LFE**	**wGT3X+/N**
Step count						
Mean ± SD (or ×/÷ factor)	9194 ×/÷ 1.5	11394 ×/÷ 1.4	5072 ×/÷ 1.6	15958 ×/÷ 1.4	19531 ×/÷ 1.3	10754 ×/÷ 1.5
Corr. coeff. [95% CI]		0.92 [0.80; 0.97][a]	0.94 [0.85; 0.97][a]		*0.96 [0.93; 1.00]*	*0.92 [0.84; 1.00]*
MPE [95% CI], %		20.5 [11.8; 27.5]	-44.9 [-51.1; -41.1]		21.9 [16.3; 25.3]	-31.6 [-36.5; -28.6]
MAPE [95% CI], %		20.5 [11.8; 27.5]	44.9 [39.2; 49.1]		21.9 [16.3; 25.3]	31.6 [28.6; 35.8]
Max 60-min steps accumulation (steps/min)						
Mean ± SD (or ×/÷ factor)	35.6 ± 10.5	38.9 ± 8.4	23.9 ± 8.9	67.1 ×/÷ 1.3	71.0 ×/÷ 1.3	57.2 ×/÷ 1.4
Corr. coeff. [95% CI]		0.91 [0.79; 0.96][a]	0.96 [0.89; 0.98]		0.96 [0.91; 0.98][a]	0.95 [0.90; 0.98][a]
MPE [95% CI], %		8.6 [2.3; 12.1]	-31.6 [-40.1; -29.3]		3.1 [1.4; 8.3]	-13.2 [-17.7; -10.0]
MAPE [95% CI], %		10.9 [4.1; 13.3]	31.6 [26.9; 35.7]		4.0 [1.9; 8.3]	14.5 [10.1; 17.7]
Max 30-min steps accumulation (steps/min)						

Mean ± SD (or ×/÷ factor)	46.1 ± 12.9	48.0 ± 10.4	32.9 ± 12.2	84.1 ×/÷ 1.2	86.3 ×/÷ 1.2	75 ×/÷ 1.4
Corr. coeff. [95% CI]		0.96 [0.90; 0.98]	0.96 [0.90; 0.98]		0.96 [0.92; 0.98][a]	0.94 [0.88; 0.97][a]
MPE [95% CI], %		4.7 [-1.3; 7.1]	-29.3 [-38.1; -22.9]		1.3 [-0.2; 2.3]	-10.0 [-11.8; -7.9]
MAPE [95% CI], %		5.2 [2.6; 8.9]	29.3 [22.6; 34.5]		1.7 [0.8; 2.4]	10.2 [8.6; 11.8]
Max 5-min steps accumulation (steps/min)						
Mean ± SD (or ×/÷ factor)	81.9 ± 15.2	80.7 ± 15.3	69.0 ± 19.1	108.5 ± 12.9	109.7 ± 15.6	103.8 ± 19.6
Corr. coeff. [95% CI]		0.99 [0.98; 1.00]	0.97 [0.91; 0.99]		0.90 [0.81; 0.95]	0.91 [0.81; 0.95]
MPE [95% CI], %		-0.9 [-1.8; -0.7]	-14.7 [-21.8; -10.4]		-0.2 [-0.6; 0.1]	-4.1 [-9.1; -2.6]
MAPE [95% CI], %		1.4 [0.9; 1.6]	14.7 [10.3; 18.6]		1.2 [0.3; 2.2]	5.0 [3.3; 9.5]
Peak 60-min steps accumulation (steps/min)						
Mean ± SD (or ×/÷ factor)	66.8 ± 11.4	64.9 ± 10.5	47.0 ± 13.5	94.9 ×/÷ 1.2	95.0 ×/÷ 1.2	82.8 ×/÷ 1.3
Corr. coeff. [95% CI]		0.95 [0.88; 0.98]	0.93 [0.83; 0.97]		0.94 [0.88; 0.97][a]	0.93 [0.86; 0.97]
MPE [95% CI], %		-1.5 [-4.8; -0.6]	-28.2 [-36.2; -25.7]		-0.1 [-1.2; 0.2]	-11.1 [-16.0; -7.6]
MAPE [95% CI], %		3.9 [0.8; 4.8]	28.2 [24.3; 33.3]		1.2 [0.6; 1.9]	11.5 [7.6; 16.9]
Peak 30-min steps accumulation (steps/min)						
Mean ± SD (or ×/÷ factor)	79.7 ± 12.1	77.2 ± 11.8	61.7 ± 15.9	105.0 ×/÷ 1.1	105.2 ×/÷ 1.1	96.6 ×/÷ 1.2

Human Kinetics, 1607 N Market St, Champaign, IL 61825

Corr. coeff. [95% CI]		0.97 [0.91; 0.99]	0.93 [0.84; 0.97]		0.91 [0.82; 0.96][a]	0.89 [0.79; 0.95]
MPE [95% CI], %		-2.2 [-3.5; -1.7]	-19.1 [-27.5; -17.4]		-0.2 [-0.8; 0.0]	-6.6 [-9.5; -3.7]
MAPE [95% CI], %		2.9 [1.4; 3.4]	19.1 [16.7; 26.7]		0.6 [0.4; 1.5]	7.8 [4.2; 10.4]
Peak 1-min steps accumulation (steps/min)						
Mean ± SD (or ×/÷ factor)	104.5 ± 9.5	103.1 ± 11.1	96.2 ± 14.8	119.3 ×/÷ 1.1	120.8 ×/÷ 1.1	116.7 ×/÷ 1.1
Corr. coeff. [95% CI]		0.99 [0.98; 1.00]	0.96 [0.90; 0.98]		*0.76 [0.44; 0.95]*	*0.75 [0.42; 0.95]*
MPE [95% CI], %		-1.1 [-2.1; -0.5]	-7.4 [-11.0; -4.4]		0.2 [-0.8; 1.9]	-2.2 [-3.7; -0.6]
MAPE [95% CI], %		1.3 [0.6; 1.4]	7.4 [3.1; 8.7]		1.9 [1.0; 2.6]	3.1 [1.8; 6.6]

Note. Descriptive statistics are means ± SD, medians (25^{th} – 75^{th} percentiles), or geometric means ×/÷ factor for SD. *N* = Number of the participants available for the analysis; 95% CI = 95 % confidence interval; Corr. coeff. = correlation coefficient (values in italics are Spearman rho coefficients; otherwise, values are Pearson r coefficients); MPE = median percent error; MAPE = median absolute percent error.

[a]Calculated after log transformation of the variables.

21 **Table 3.** Linear models for predicting SW3 daily step count from wGT3X+ daily step count

Method	Model for predicting SW3 steps/day	R^2	RMSE
People with PAD			
wGT3X+/LFE	-30.30477 + 0.82201 × steps/day	0.85	1663
wGT3X+/N	1704.7047 + 1.4690 × steps/day	0.89	1392
Older adults			
wGT3X+/LFE	-1897 + 0.9193 × steps/day	0.95	1298
wGT3X+/N	4903 + 1.013 × steps/day	0.93	1574

Note. Steps/day is the mean of the daily step count calculated using the valid days. R^2 coefficient and root-mean-squared error (RMSE) were computed using the leave-one-out method.

22

FIGURES CAPTIONS

Figure 1. Step counts from the SW3 and wGT3X+ monitors related to the outdoor walking session

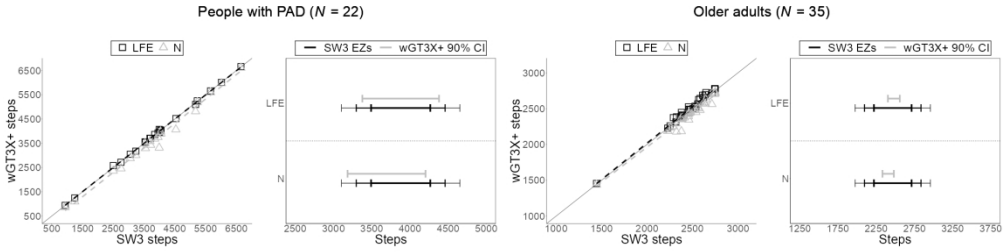
On the scatter plots, the solid line shows the identity, and the dashed lines show the regressions of the wGT3X+/LFE and wGT3X+/N data on the SW3 data respectively. On the equivalence plots, the thick, thin, and very thin black lines show the ±10%, ±15%, ±20% SW3 equivalence zones, respectively; LFE = lower frequency extension filter; N = normal filter; EZs = equivalence zones; 90% CI = 90% confidence interval.

Figure 2. Effect of pain occurrence on the agreement between the SW3 and wGT3X+ monitors and on walking speed

Panel A shows, for each wGT3X+ method, the individual percent errors (PE) for walking without pain and for walking with pain. Panel B shows, for each wGT3X+ method, the individual differences of absolute percent error (APE) between walking with pain and walking without pain (Change in APE = APE[with pain] minus APE[without pain]), along with the corresponding medians (grey bars) and 95% confidence intervals (black error bars). Panel C shows the individual walking speeds when walking without pain and walking with pain. Panel D shows the differences in walking speed (Change in speed = speed[with pain] minus speed[without pain]), with the mean difference and its corresponding 95% confidence interval. The grey zone depicts non-substantial change in speed. The results are related to 17 participants who indicated the occurrence of pain during the first walking bout.

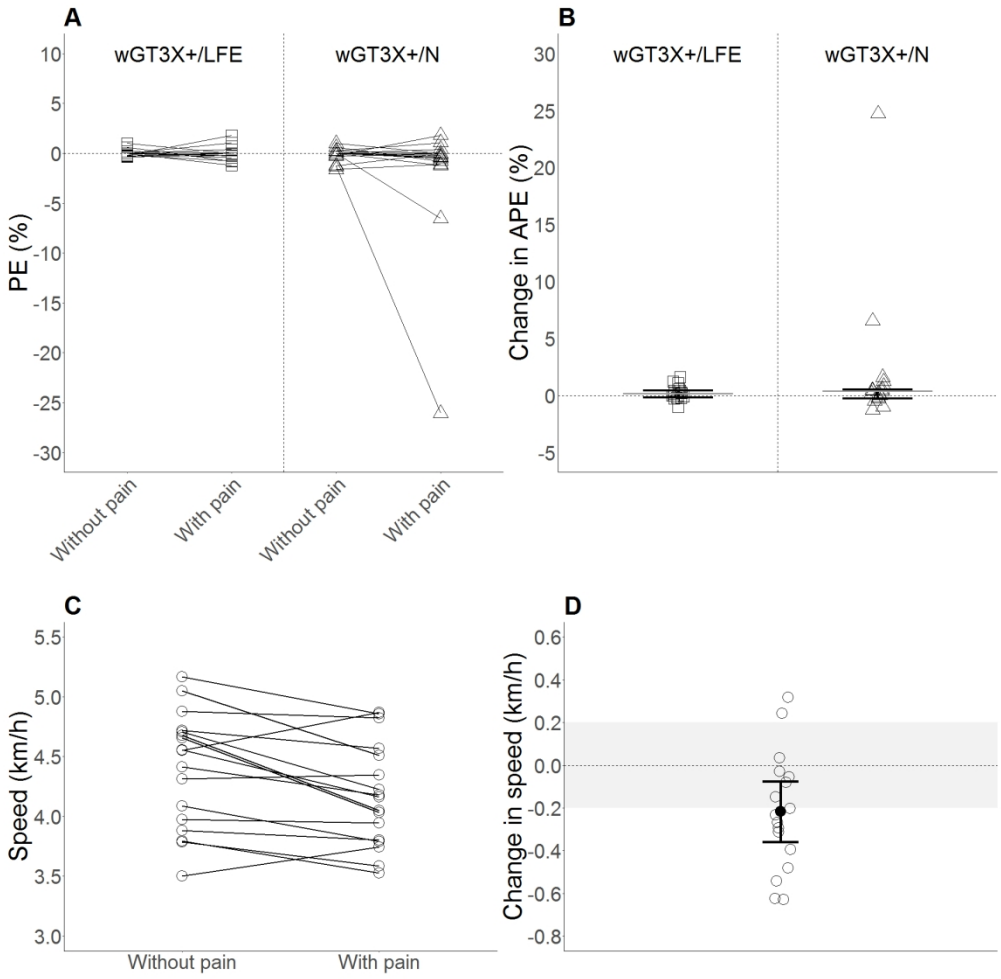
Figure 3. Step-based metrics from the SW3 and wGT3X+ monitors for the 7-day free-living period

47 On the scatter plots, the solid line shows the identity, and the dashed lines show the regressions
48 of the wGT3X+/LFE and wGT3X+/N data on the SW3 data respectively. On the equivalence
49 plots, the thick, thin, and very thin black lines show the ±10%, ±15%, ±20% SW3 equivalence
50 zones, respectively; LFE = lower frequency extension filter; N = normal filter; EZs =
51 equivalence zones; 90% CI = 90% confidence interval.



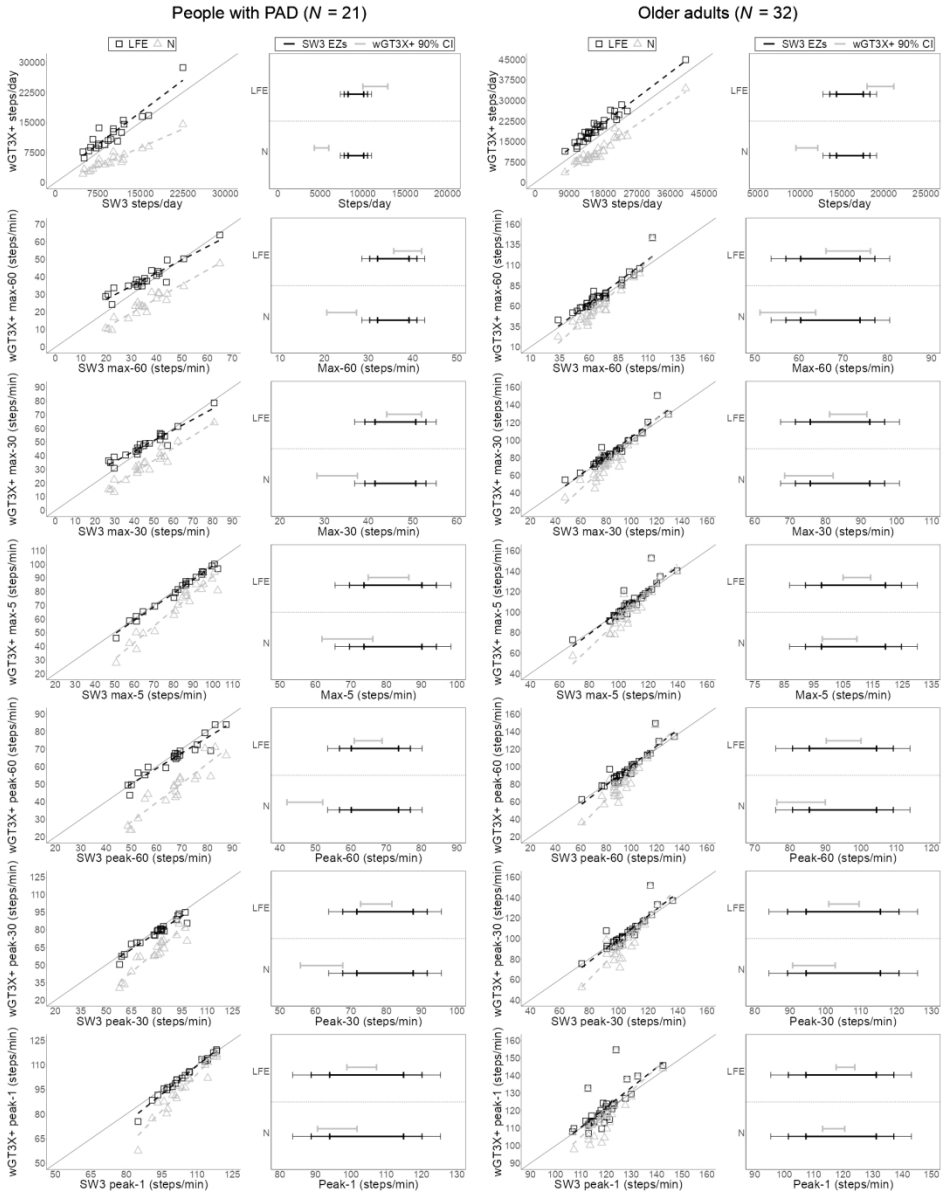
On the scatter plots, the solid line shows the identity, and the dashed lines show the regressions of the wGT3X+/LFE and wGT3X+/N data on the SW3 data respectively. On the equivalence plots, the thick, thin, and very thin black lines show the ±10%, ±15%, ±20% SW3 equivalence zones, respectively; LFE = lower frequency extension filter; N = normal filter; EZs = equivalence zones; 90% CI = 90% confidence interval.

179x49mm (300 x 300 DPI)



Panel A shows, for each wGT3X+ method, the individual percent errors (PE) for walking without pain and for walking with pain. Panel B shows, for each wGT3X+ method, the individual differences of absolute percent error (APE) between walking with pain and walking without pain (Change in APE = APE[with pain] minus APE[without pain]), along with the corresponding medians (grey bars) and 95% confidence intervals (black error bars). Panel C shows the individual walking speeds when walking without pain and walking with pain. Panel D shows the differences in walking speed (Change in speed = speed[with pain] minus speed[without pain]), with the mean difference and its corresponding 95% confidence interval. The grey zone depicts non-substantial change in speed. The results are related to 17 participants who indicated the occurrence of pain during the first walking bout

705x705mm (72 x 72 DPI)



On the scatter plots, the solid line shows the identity, and the dashed lines show the regressions of the wGT3X+/LFE and wGT3X+/N data on the SW3 data respectively. On the equivalence plots, the thick, thin, and very thin black lines show the ±10%, ±15%, ±20% SW3 equivalence zones, respectively; LFE = lower frequency extension filter; N = normal filter; EZs = equivalence zones; 90% CI = 90% confidence interval.

179x229mm (250 x 250 DPI)

Supplemental material

eText. Procedure to remove the initial wGT3X+ wear time periods where the comparison of the monitors was not appropriate

eFigure. Visualizations of the cases where the "wear" in-bed night-time epochs (panels A, B, and C) or the waking-time epochs (panel D) were reconsidered as "nonwear" after graphical inspection

eText. Procedure to remove the initial wGT3X+ wear time periods where the comparison of the monitors was not appropriate

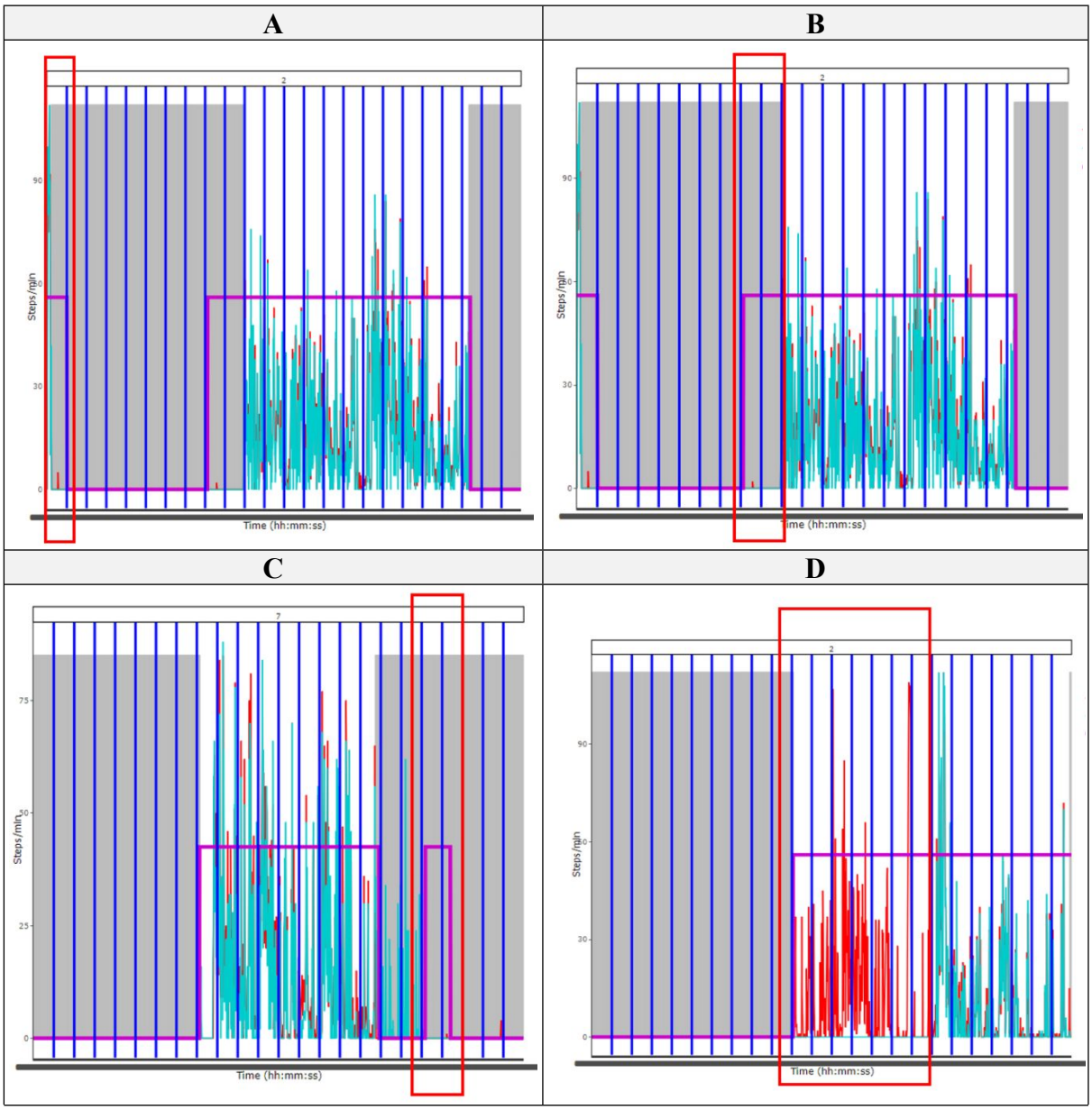
Before conducting the graphical inspection, in-bed time was determined using the "PhysActBedRest" R package (version 1.0, 2016) provided by Tracy et al. (Tracy, Acra, Chen, & Buchowski, 2018). In-bed time was marked into the merged data file using the Tracy et al. (Tracy et al., 2018) algorithm, the wGT3X+/LFE vector magnitude (VM) counts, and the optimal cut-points ($CP_0 = 50$; $CP_1 = 210$; $CP_2 = 350$) provided by Belletiere et al. (Bellettiere et al., 2019). The beginning of the in-bed night-time period was considered as the beginning of the in-bed time period visible at the end of the 24-h day (or as the beginning of the first in-bed time period of the following day if any) that appeared after the last out-of-bed time period of the day. Graphical inspection was conducted in R by plotting both the SW and wGT3X+/LFE steps, wGT3X+/LFE wear time, and in-bed time, against time (see eFigure below), and by using the "plotly" package. In the following three cases, all the concerned "wear" in-bed night-time epochs were reconsidered as "nonwear": (i) when an initial wear time period began in an out-of-bed time period and was prolonged into an in-bed night time period for ≥30 min (eFigure, panel A); (ii) when an initial wear time period began during an in-bed nighttime period more than 30 min before to be prolonged into the next out-of-bed time period (eFigure, panel B); (iii) when an initial wear time period both began and ended during an in-bed night-time period (eFigure, panel C). In addition, if the first out-of-bed time period of the day included more than 30 min of wGT3X+/LFE wear time without SW steps, the corresponding epochs were reconsidered as "nonwear" (eFigure, panel D).

References

Bellettiere, J., Zhang, Y., Berardi, V., Full, K. M., Kerr, J., LaMonte, M. J., . . . Di, C. (2019). Parameterizing and validating existing algorithms for identifying out-of-bed time using hip-worn accelerometer data from older women. *Physiol Meas, 40*(7), 075008.

Tracy, J. D., Acra, S., Chen, K. Y., & Buchowski, M. S. (2018). Identifying bedrest using 24-h waist or wrist accelerometry in adults. *PLoS One, 13*(3), e0194461.

eFigure. Visualizations of the cases where the "wear" in-bed night-time epochs (panels A, B, and C) or the waking-time epochs (panel D) were reconsidered as "nonwear" after graphical inspection



Note. The cyan lines, red lines, and purple lines, show the SW3 steps, the wGT3X+/LFE steps, and the wGT3X+/LFE wear time, respectively. The grey background shows the in-bed time. The vertical blue lines show each hour of the day. Red rectangles highlight the wear time periods to be removed. A: An initial wear time period began in an out-of-bed time period and was prolonged into an in-bed night time period for ≥30 min; B: An initial wear time period began during an in-bed nighttime period more than 30 min before to be prolonged into the following out-of-bed time period; C: An initial wear time period both began and ended during an in-bed night-time period; D: The first out-of-bed time period of the day with wear time included more than 30 min of wGT3X+/LFE steps without SW3 steps.