Free-living cross-comparison of two wearable monitors for sleep and physical activity in healthy young adults

Nicola Cellini [a,b], Elizabeth A. McDevitt [b], Sara C. Mednick [b], Matthew P. Buman [c,*]

[a] Department of General Psychology, University of Padova, Via Venezia 8, 35131 Padova, Italy
[b] Department of Psychology, University of California, Riverside, 900 University Ave., Riverside, CA 92521, USA
[c] School of Nutrition & Health Promotion, Arizona State University, 550 N 3rd St, Phoenix, AZ 85004, USA

HIGHLIGHTS

- Sleep and physical activity/sedentary behaviors impact health.
- Validation of monitors for assessing the full 24 h spectrum of behaviors is lacking.
- We cross-compared two standard monitors to assess sleep and physical activity.
- The GT3X+ showed good agreement with AW-64 for assessing sleep.
- There was a lack of agreement between AW-64 and GT3X+ for physical activity.

ARTICLE INFO

ABSTRACT

There is a growing need for free-living monitoring of the full 24 h spectrum of behaviors with a single or integrated set of sensors. The validity of field standard wearable monitors in sleep and physical activity have yet to be assessed for the complementary behavior in the context of 24 h continuous monitoring. We conducted a free-living comparison study of the Actigraph GT3X+ (GT3X+) to assess sleep parameters as compared with the Actiwatch-64 (AW-64) and concurrently, the AW-64 to assess sedentary and physical activity behaviors as compared with the GT3X+. Thirty young adults (15 female, 19.2 ± 0.86 years) wore both monitors for 3 consecutive days and 2 consecutive nights. Agreement of sleep, sedentary, and physical activity metrics were evaluated using analyses of variance, intraclass correlation coefficients, Bland-Altman plots with associated confidence limits, mean absolute percentage of errors and equivalence tests. For sleep, the GT3X+ showed high agreement for total sleep time and sleep efficiency, but underestimated wakefulness after sleep onset and sleep onset latency relative to the AW-64. For sedentary behavior and physical activity, the AW-64 showed a moderate agreement for activity energy expenditure, but not for sedentary, light or moderate-vigorous physical activities relative to the GT3X+. Overall our results showed good agreement of the GT3X+ with AW-64 for assessing sleep but a lack of agreement between AW-64 and GT3X+ for physical activity and sedentary behaviors. These results are likely due to the monitor placement (wrist vs hip), as well as the algorithm employed to score the data. Future validation work of existing and emerging technologies that may hold promise for 24 h continuous monitoring is needed.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Recently, several cost-effective systems have been developed to objectively monitor sleep and physical activity in real-world environments. Accelerometry is a widely used ecological, non-invasive technology and cost-effective substitute for both polysomnography (PSG), the gold standard for sleep monitoring [1], and indirect calorimetry [2], the gold standard for physical activity. Accelerometry provides objective monitoring of sleep-wake rhythm [3] and physical activity behavior [4] in free-living settings based on the recording of motion. Body movements are recorded by an accelerometer, which can be worn on the wrist, ankle, or hip. For sleep assessment the device is typically worn on the non-dominant wrist, and this technique has shown acceptable agreement with PSG, ranging between 85% and 95% for the identification of sleep-wake epochs [5], and for a similar sleep-wake detection ability of systems with dry electrodes [6, 7], at least for healthy adults (but see [8–10] about limits of these devices with paediatric and sleep disorder populations). Similarly, new accelerometer

* Corresponding author.
E-mail addresses: cellini.nicola@gmail.com (N. Cellini), mcdevitt@ucr.edu (E.A. McDevitt), smednick@ucr.edu (S.C. Mednick), mbuman@asu.edu (M.P. Buman).

monitors show acceptable agreement with indirect calorimetry, a reliable measure of energy expenditure for waking activities, when worn in the hip [11].

Accelerometry systems may be used to concurrently assess sleep and physical activity behaviors in free-living settings. Measuring the full 24 h cycle with a single monitor would be a potentially convenient and cost-effective way to collect information about health and well-being [12]. Indeed, both sleep and physical activity/sedentary behaviors impact health [13], with both sedentary behaviors [14, 15] and poor sleep quality [16, 17] being associated with negative health outcomes. In addition, the dynamic interaction between these behaviors in the 24 h day mediates health outcomes. For example, higher sleep quality increases energy and reduces fatigue levels [18]. Reciprocally, greater physical activity ameliorates sleep quality [12, 18]. Moreover, the optimal combination between time spent sleeping and time spent in active behaviors (both light and moderate to vigorous physical activities) is associated with lower cardiovascular risk [19]. Thus, continuous, free-living monitoring of the full 24 h spectrum is needed to better understand the unique and combined impacts of these health behaviors.

However, current standard wearable monitors in sleep and physical activity are not well-validated for complementary behaviors during the 24 h period. For example, only a small number of studies have assessed the validity of the GT3X+ accelerometer (Actigraph, Pensacola, Florida, USA), a commonly used monitor for measuring physical activity and sedentary behaviors [20–22], to detect sleep-wake patterns against a concurrent PSG recording [23–25]. These studies showed that GT3X+ showed a systematic underestimation of wakefulness (i.e., both sleep latency and wake after sleep onset (WASO)), as well as a device placement effect. When the device was worn on the wrist, the typical position for sleep trackers, accuracy, sensitivity and specificity values were comparable to previous reports of other similar devices [23, 25], but not when worn on the hip [24, 25]. Overall, the wrist–worn GT3X+ appeared to be valid for detecting sleep/wake patterns in a laboratory setting. However, to our knowledge the ability of this device to monitor sleep in a free-living environment has not yet been studied.

In contrast, the Actiwatch-64 (AW-64; Phillips Respironics, Bend, Oregon, USA) is a validated and widely-used wearable monitor for sleep in both healthy and clinical populations [23, 26, 27], as well as children [8, 28]. This wrist-worn monitor is commonly used in clinical settings [29–31] and is one of the most trusted wearable monitors for sleep assessment. However, validation studies investigating the ability of this monitor to accurately assess physical activity behaviors in adults are lacking.

There is an important knowledge gap in understanding the accuracy of these commonly used wearable monitors for measuring the full 24 h spectrum of behaviors. The current study aimed to establish the agreement of these monitors in estimating both sleep and physical activity parameters under free-living conditions. Specifically, the aims are two-fold: a) to determine the agreement of the Actigraph (GT3X+) in measuring sleep parameters as compared with the Actiwatch-64 (AW-64) and b) to determine the agreement of the AW-64 in measuring sedentary and physical activity behaviors as compared with the GT3X+.

2. Method and materials

2.1. General study design

Study participants had no personal history of neurological, psychological, or other chronic illness. Participants reporting irregular sleep-wake schedules (i.e., reporting a habitual time in bed (TIB) longer or shorter than 7–9 h per night) or excessive levels of average daytime sleepiness (i.e., scores > 10), as measured by the Epworth Sleepiness Scale [32], were excluded from the study. Additionally, any participant currently experiencing symptoms of a sleep disorder determined during an in-person or phone interview with an experimenter were excluded

from the study. Each participant gave informed consent prior to participation in the study, which was approved by the University of California at Riverside Human Research Protections Program, and received financial compensation or course credit for participating in the study.

Participants came to the Sleep and Cognition Lab at UC Riverside to receive the two study monitors (AW-64 and GT3X+). The field-based standard criterion for sleep was the AW-64. The field-based standard criterion for physical activity and sedentary behaviors was the GT3X+. Both monitors were initialized from the same computer prior to the participant visit and the clocks were synchronized in order to compare their output. Each participant wore both monitors for 3 consecutive days and 2 consecutive nights. Participants were instructed to wear the GT3X+ on the hip during the day using an elastic waistband that could be worn underneath clothing, and to move the monitor to be worn on a wristband at night. While a full 24 h wrist-worn protocol was considered, the decision for hip placement during the day was made to accommodate our goal to provide the best field-based criterion measure of free-living physical activity and sedentary behavior. While wrist placement is emerging for physical activity assessment, currently the best measurement potential remains the hip location [4]. The AW-64 was worn continuously on the non-dominant wrist. Participants were instructed to wear the monitors at all times for the next three days and two nights, and to remove them only when there was a chance the monitor could be damaged, i.e., coming in contact with water or playing a high-impact sport.

2.2. Wearable monitors

The AW-64 (Phillips Respironics, Bend, Oregon, USA) consists of a piezoelectric accelerometer with a vertical acceleration sensitivity of 0.02 g, a sampling rate of 32 Hz and a storage capacity of 64 kb. The GT3X+ (Actigraph, LLC, Pensacola, FL, USA) contains a tri-axial accelerometer with a sensitivity of 0.05 g and the sampling rate ranges from 30 to 100 Hz. Both monitors allow the recording and storage of several weeks of sleep-wake activity data. Since the ActiLife 6.4.3 software (Actigraph, LLC Pensacola, Florida, USA) automatically scores sleep using 60-s epoch cycles, even if the GT3X+ is initialized to collect data in 30-s epochs or shorter, both devices were initialized to collect data in 1-min epochs.

2.3. Sleep data processing

AW-64 sleep-recordings were analyzed using Actiware 5.52.0003 (Phillips Respironics, Bend, Oregon, USA) software. Data were scored using a proprietary algorithm provided by the software with three different sensitivity threshold levels, which refer to the number of activity counts used to define wake: Low (20 activity counts/epoch), Medium (40 activity counts/epoch), and High (80 activity counts/epoch).

GT3X+ data were analyzed with ActiLife 6.4.3 using the Sadeh sleep scoring algorithm [33]. We scored our data with both the default setting (ACT) and the Low Frequency Extension (LFE) setting. The LFE option has been designed to lower the band-pass filter threshold for signal detection.

All analyses were confined to the period between lights off and lights on, which was defined by bed times and wake times reported by the participants [34, 35]. The following sleep parameters were examined for the two systems: total sleep time (TST), defined as the number of minutes scored as sleep between lights off and lights on; sleep onset latency (SL), the number of minutes between lights out and the first epoch scored as sleep; wake after sleep onset (WASO), the number of minutes scored as wake after sleep onset; and sleep efficiency (SE), the ratio between TST and total time spent in bed. For both monitors these parameters were directly extracted from the output of the respective software packages.

2.4. Physical activity data processing

GT3X+ physical activity recordings were analyzed using ActiLife 6.4.3 software. Data were scored using a common threshold-based algorithm in order to derive the following parameters: sedentary (<-100 counts/min [cpm]), light-intensity (100–1951 cpm), and moderate-vigorous physical activity (MVPA; 1952+ cpm) [36]. These parameters represent a continuum in energy expenditure activities from very low (i.e., sedentary behaviors, such as sitting/reclining with as during television viewing or workplace sitting) to moderate-to-vigorous physical activities (e.g., fast-paced walking, playing sports, hard physical work), passing through light-intensity activities (i.e., lifestyle behaviors, such as walking or doing housekeeping activities) [19]. We also estimate via ActiLife 6.4.3 software the total activity energy expenditure (AEE, kcal).

AW-64 physical activity recordings were analyzed using a custom-built SAS program that scored epoch-based acceleration counts derived from the Actiware 5.52.0003 software. Data were scored using a threshold-based algorithm for activity energy expenditure, sedentary, light-intensity, and moderate-vigorous physical activity derived from data collected at the wrist for the Actical monitor (a monitor functionally similar to the AW-64 built by the same manufacturer) [37].

All analyses were confined to periods not included in the sleep recordings to account for the full 24 h. A non-wear algorithm was applied to both GT3X+ and AW-64 data (i.e., 60 consecutive '0' counts without grace period) prior to scoring and no non-wear periods were identified in either dataset.

2.5. Data analysis

Sleep parameters were compared between the different monitors with a series of analysis of variance (ANOVA) with the different settings (Low, Medium, High, ACT, LFE) as within-subjects factors and Tukey HSD post-hoc tests. Intra-class correlation coefficients (ICCs), a widely used technique to assess the relative reliability between measurements [38], were calculated to measure the agreement between the systems. Specifically, we used ICCs (2,1) with absolute agreement, which are used to assess the agreement between two or more measurements. The ICCs (2,1) account for both agreement of performance between the two measures in the same individual (within-subject change), as well as change in average performance of participants as a group between the two measures (i.e., systematic change in mean). According to the Landis and Koch scale [39], we considered ICCs of 0–0.2 as slight agreement, 0.2–0.4 as fair agreement, 0.4–0.6 as moderate agreement, 0.6–0.8 as substantial agreement, and 0.8–1.0 almost perfect agreement. A $p < 0.05$ was considered statistically significant for all analyses.

Mean absolute percent errors (MAPE) were calculated as an indicator of overall measurement error. MAPEs were computed for each sleep parameter and for each setting as the average of absolute differences between the score of the GT3X+ and the AW-64 value divided by the AW-64 value, multiplied by 100. Similarly, MAPEs were computed for each physical activity measure and for each setting as the average of absolute differences between the score of the AW-64 and the GT3X+ value divided by the GT3X+ value, multiplied by 100.

We also tested whether the parameters estimated by one monitor were "equivalent" to the other using an equivalence test [40]. Following Lee et al. [40], for each parameter we computed the equivalence zone (EZ) as the ±10% of the mean of our criterion (e.g., TST score from the AW-64 High sensitivity) and the 90% confidence interval (CI) of the mean of the estimated parameter by the comparison monitor (e.g., TST score from the GT3X+ LFE setting). If the 90% CI of the estimated parameter fell within the EZ of the corresponding criterion, the two measures were considered equivalent. This equivalence test made it possible to determine if the sleep parameters estimated by the two GT3X+ settings were equivalent to the estimate from the criterion measures (the three AW-64 sensitivity settings). Similarly, we tested whether the physical activity parameters estimated by the three AW-64 sensitivity settings were equivalent to the estimate from the criterion measures (the two GT3X+ settings).

Lastly, as linear correlation is an inadequate analysis to evaluate the agreement between two measurement methods [41], we used Bland-Altman plots to visually evaluate the agreement of the sleep summaries derived from two monitors. This technique plots the difference score between two measures (the Bias; e.g., TST for GT3X+ LFE minus TST for AW-64 High sensitivity) against their averages. Here, since we are considering one specific device/setting in each analysis as the absolute reference, we decided to plot the difference score between two measures against the values of the reference device/setting (see [42]). We also computed the limits of agreement (i.e., the upper and the lower limits of the Bias measured as Bias ± 1.96 standard deviation of the Bias), which indicate the range where the difference between two recordings of a new individual should lie with 95% probability.

3. Results

3.1. Sample characteristics

The final sample included 30 young adults ($M_{age} = 19.24, SD = 0.95$, 15 F). Sample body mass index was $23.93 ± 4.56 \text{ kg/m}^2$ and it was composed by African American (N = 4; 13% of the sample), Asian (N = 12; 40% of the sample), Hispanic (N = 11; 36.7% of the sample) and Caucasian (N = 3; 10% of the sample) individuals. We lost 7 GT3X+ recordings (7.78%) in day 3 due to lack of compliance from participants. The final sample was composed of a total of 60 nights and 83 days of recording.

3.2. Sleep parameters

Sleep parameters are reported in Table 1.

The analysis showed a significant difference between settings for TST ($F_{4,236} = 38.59$; $p < 0.001$), with Tukey HSD tests showing a significantly higher TST scored by the GT3X+ default setting than AW-64 Medium ($p < 0.01$) and Low sensitivity ($p < 0.001$), whereas the GT3X+ LFE setting only scored higher TST than AW-64 Low ($p < 0.001$). Comparing AW-64 settings, Medium sensitivity showed a higher TST than Low sensitivity ($p < 0.001$), but lower than High sensitivity ($p < 0.001$). A similar pattern was observed for SE ($F_{4,236} = 44.04$; $p < 0.001$), with the AW-64 High setting scoring higher SE than AW-64

Table 1
Sleep measures (mean ± SD) for the two monitors.

	AW-64			GT3X+	
	Low	Medium	High	ACT	LFE
TST (min)	343.08 ± 60.65	368.38 ± 60.63	388.75 ± 61.55	382.63 ± 57.87	378.78 ± 57.41
SL (min)	23.17 ± 17.29	23.17 ± 17.29	23.17 ± 17.29	11.88 ± 10.77	12.4 ± 10.66
WASO (min)	106.77 ± 44.43	80.73 ± 39.20	61.1 ± 3.40	78.68 ± 36.89	81.98 ± 37.53
SE (%)	72.52 ± 9.50	77.98 ± 8.73	82.13 ± 8.47	80.80 ± 7.61	80.02 ± 7.71

Low: AW-64 low setting; Medium: AW-64 medium setting; High: AW-64 high setting; ACT: GT3X+ default setting; LFE: GT3X+ low extension frequency setting; TST: total sleep time; SL: sleep onset latency; WASO: wake after sleep onset; SE: sleep efficiency.

Table 2
Intra-class correlations between GT3X+ and AW-64 settings for sleep parameters.

			AW-64		
			Low	Medium	High
GT3X+	ACT	TST	0.66	0.76	0.77
		SL	0.15	0.15	0.15
		WASO	0.52	0.55	0.42
		SE	0.45	0.53	0.50
	LFE	TST	0.69	0.77	0.77
		SL	0.14	0.14	0.14
		WASO	0.55	0.57	0.42
		SE	0.48	0.56	0.50

Low: AW-64 low setting; Medium: AW-64 medium setting; High: AW-64 high setting; ACT: GT3X+ default setting; LFE: GT3X+ low extension frequency setting; TST: total sleep time; SL: sleep onset latency; WASO: wake after sleep onset; SE: sleep efficiency.

Medium ($p < 0.01$) and Low sensitivity ($p < 0.001$), whereas both GT3X+ settings differed significantly from AW-64 Low ($p < 0.001$). GT3X+ default setting also reported a higher SE than AW-64 Medium ($p < 0.001$). Again, Medium sensitivity showed higher SE than Low sensitivity ($p < 0.001$) but lower than High sensitivity ($p < 0.001$).

We also observed differences in WASO ($F_{4,236} = 36.26$; $p < 0.001$). The AW-64 High showed less WASO than both GT3X+ settings ($p < 0.001$) as well as the other AW-64 settings ($p's < 0.01$). Also, both GT3X+ settings and AW-64 Medium showed reduced WASO compared to AW-64 Low ($p's < 0.001$). The analyses also revealed a significant difference between settings for SL ($F_{4,236} = 21.55$; $p < 0.001$), with Tukey

HSD tests showing a shorter SL reported by the GT3X+ settings relative to AW-64 ($p's < 0.001$). Note that the three AW-64 settings have the same SL; this is due to the Actiware 5.52.0003 algorithm, which computes sleep onset independently from the sleep-wake discrimination: sleep onset depends only on the presence or absence of activity counts (i.e. sleep onset is computed as the first epoch of the first 10 min block of epochs that contained no more than one epoch in which the subject showed activity higher than 0). Given that in this study we used the same lights off time for each setting, and given that sleep onset is the same (the settings do not affect the activity counts), mean and standard deviation of SL are the same for each setting of AW-64. To explore whether clustering effects impacting any of the results (i.e., dependency among observations within participants), we re-ran all analyses using linear mixed models. The results were comparable to the ANOVA outputs and therefore not reported.

Overall, ICCs revealed good agreement between the two systems for TST, moderate for WASO and SE, and poor agreement for SL. The best relationship was observed between GT3X+ LFE and the AW-64 Medium sensitivity setting (Table 2).

Bland-Altman plots and limits of agreements revealed that GT3X+ settings underestimate SL relative to the AW-64 (ACT = −11.28 min; LFE = −10.77 min). This results in a systematic overestimation of the TST and SE by GT3X+ compared to the AW-64 Medium (TST ACT = 14.25 min; TST LFE = 10.40 min; SE ACT = 2.82%; SE LFE = 2.04%; see Fig. 1) and Low (TST ACT = 39.55 min; TST LFE = 35.70 min; SE ACT = 8.28%; SE LFE = 7.50%). In addition, the GT3X+ settings overestimated WASO compared to AW-64 High (ACT = 17.58 min;
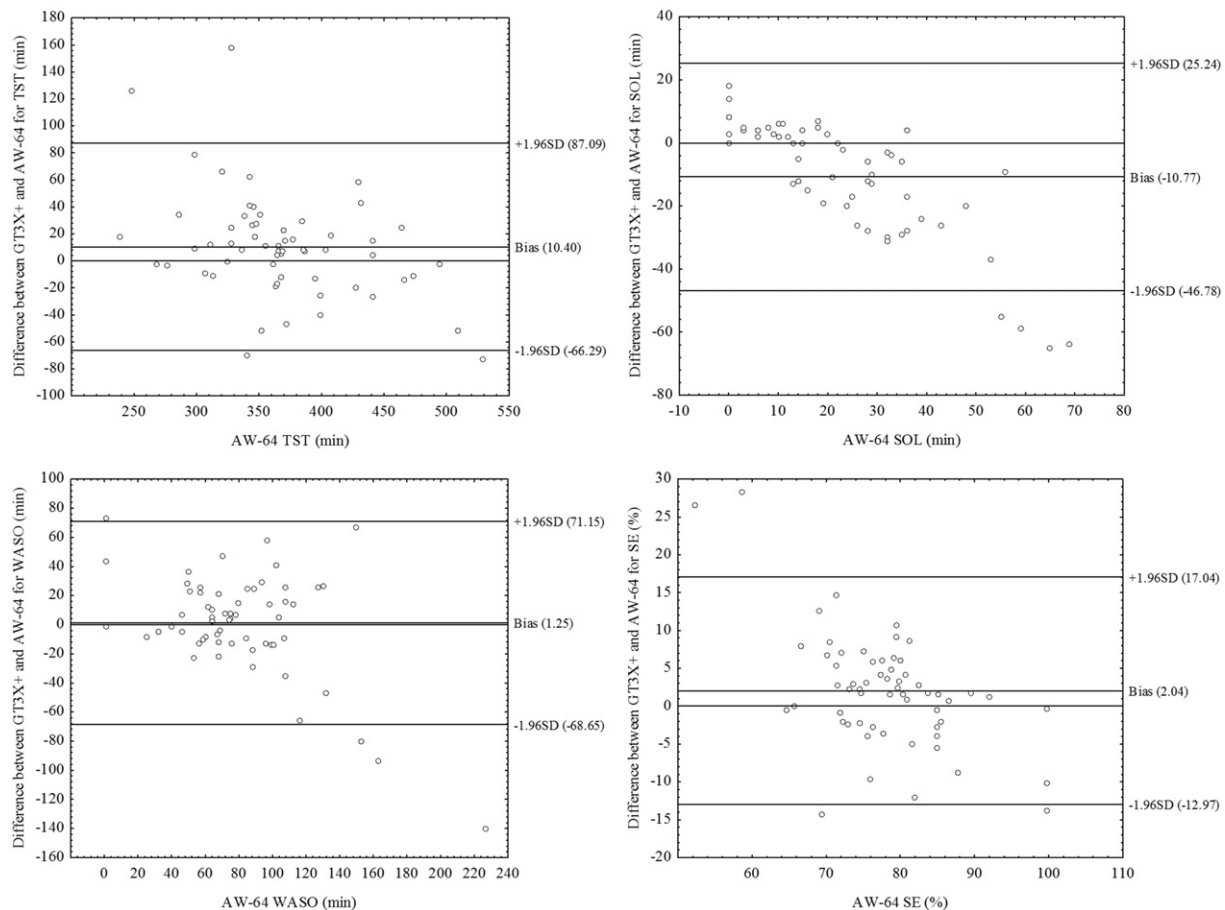
Fig. 1. Bland-Altman plots of GT3X+ LFE setting and AW-64 Medium sensitivity for sleep parameters. The y-axis indicates the differences between the GT3X+ score minus the AW-64 score, whereas the x-axis shows the average of their scoring. The Bias represents the mean difference between the devices for a specific parameter, with values above zero meaning an overestimation and the values below zero meaning an underestimation of the GT3X+ relative to AW-64. Upper and lower lines represent Bias ± 1.96 SD.

Table 3
Mean absolute percentage error (MAPE) of sleep parameters between GT3X+ and AW-64 settings.

			AW-64		
			Low	Medium	High
GT3X+	ACT	TST	11.53	3.87	1.57
		SL	48.71	48.71	48.71
		WASO	26.30	2.54	28.78
		SE	11.42	3.62	1.62
	LFE	TST	10.41	2.82	2.56
		SL	46.47	46.47	46.47
		WASO	23.21	1.55	34.18
		SE	10.34	2.61	2.58

Low: AW-64 low setting; Medium: AW-64 medium setting; High: AW-64 high setting; ACT: GT3X+ default setting; LFE: GT3X+ low extension frequency setting; TST: total sleep time; SL: sleep onset latency; WASO: wake after sleep onset; SE: sleep efficiency.

LFE = 20.88 min) and underestimated WASO when compared to AW-64 Low (ACT = −28.08 min; LFE = −24.74 min; see Supplemental Figs. 1–5).

MAPEs for the various parameters (computed as the average absolute value of the errors relative to three AW-64 settings) are reported in Table 3.

MAPEs were similar for both GT3X+ settings. The magnitude of error was high for WASO compared to AW-64 Low and High sensitivity, and relatively low for SE and TST for all three sensitivity settings, especially for Medium and Low sensitivity. SL showed the highest errors.

The equivalence test revealed that for both GT3X+ settings, 90% CI of TST and SE were within the EZ of the AW-64 High and Medium sensitivity settings (see Supplemental Fig. 6). WASO was slightly outside the EZ of the Medium settings (EZ: 72.66–88.81 min; 90% CI ACT: 70.82–86.54 min; 90% CI LFE: 73.99–89.97 min). No other parameter fell within the EZ.

3.3. Physical activity and sedentary behaviors

Physical activity parameters are reported in Table 4.

We observed a significant difference between settings for activity counts ($F_{4,328}$ = 218.14; p < 0.001) with Tukey HSD test showing significantly lower number of counts recorded by all three AW-64 sensitivities compared to the two GT3X+ settings for both parameters (p's < 0.001, Table 4). A similar, but reversed, pattern was observed for light intensity activity ($F_{4,328}$ = 185.89; p < 0.001), with a more minutes identified by the AW-64 compared to the GT3X+ settings (p's < 0.001). We also observed significant differences across the monitors for sedentary behavior ($F_{4,328}$ = 62.49; p < 0.001) and MVPA ($F_{4,328}$ = 309.52; p < 0.001), with a underestimation of epochs for all AW-64 sensitivities compared to both GT3X+ settings (p's < 0.001). Also, AW-64 Medium sensitivity identified a lower number of sedentary epochs compare to Low sensitivity (p < 0.001). The analysis of the AEE showed significant differences across the monitors ($F_{4,328}$ = 52.84; p < 0.001), with Tukey HSD test revealing that the AW-64 sensitivities significantly overestimate the AEE compared to both GT3X+ settings (p's ≤ 0.001). Again,

we controlled for possible clustering effects in a subsequent analysis and found the results to be comparable.

ICCs revealed very low agreement between the two systems for sedentary, counts per minute and light activity, no agreement with MVPA, whereas AEE showed a fair to moderate agreement. Note that the all the AW-64 sensitivities were more consistent with the GT3X+ LFE rather than the default setting (Table 5).

Bland-Altman plots and limits of agreement revealed that the AW sensitivities overestimate both sedentary, MVPA, AEE and counts per minutes, whereas light time was consistently underestimated (see Fig. 2 and Supplemental Figs. 7–11).

A high MAPE was observed for physical activity parameters, especially for MVPA for all AW-64 sensitivities (Table 6). The magnitude of errors was also very high for the Light, counts per minute, and AEE, whereas sedentary time showed moderate error. In general, among the three AW-64 sensitivities, the Medium sensitivity showed the lowest error compared to the GT3X+ settings.

The equivalence test revealed that, for all three AW-64 settings, 90% CI of the PA parameters were outside the equivalence zone (see Supplemental Fig. 12).

4. Discussion

In the current study we examined the cross-validity of two widely used accelerometers for measuring physical activity and sleep-wake behaviors in healthy young adults in a free-living environment. We conducted a continuous monitoring for 3 consecutive days and 2 consecutive nights in order to test whether two wearable monitors, individually validated for sleep and physical behaviors, might be interchangeable. Overall, we observed a lack of congruence between the monitors, with some differences between sleep and physical activity behaviors.

Specifically, the GT3X+ worn in the wrist showed some promise for assessing sleep in healthy adults in a similar way to the well-validated AW-64 device, except for sleep onset latency (SL). In fact, as suggested by mean value and Bland-Altman plots and limits of agreements, the GT3X+ tended to underestimate both SL and WASO, resulting in a consequent overestimation of total sleep time (TST) and sleep efficiency (SE). However, ICCs showed poor agreement only for sleep latency, whereas for the other variables the agreement ranged between 0.42 and 0.877. Similarly, the highest mean absolute percentage error was for SL, followed by WASO, and TST and SE had the lowest error. The same results were shown by the equivalence test, with both GT3X+ settings falling in the equivalence zone for TST and SE, with WASO only partially overlapping the equivalence zone. These results indicate wakefulness detection as the primary difficulty of the GT3X+ algorithm, with a particular deficit in identifying the sleep onset time. The poor ability of the GT3X+ to detect wakefulness is a common problem for accelerometry [10] and it is likely due, at least partially, to the algorithms employed to obtain sleep/wake parameters. For example, the GT3X+ software (ActiLife) derives sleep parameters as the sum of the single epochs defined as sleep or wake by the Sadeh algorithm [33]. Differently, Actiware, which is the software used by the AW-64 (as well as

Table 4
Physical activity parameters (mean ± SD) for the two monitors.

	AW-64			GT3X+	
	Low	Medium	High	ACT	LFE
Sedentary	379.04 ± 212.02	285.75 ± 173.02	321.37 ± 119.27	552.66 ± 250.00	561.11 ± 248.40
Light	482.63 ± 205.05	469.64 ± 205.17	466.65 ± 203.14	162.55 ± 85.95	165.43 ± 96.74
MVPA	3.72 ± 7.45	3.72 ± 7.44	3.66 ± 7.44	52.64 ± 34.08	50.51 ± 34.41
Counts/min	291.74 ± 124.05	318.36 ± 119.41	307.44 ± 123.19	713.58 ± 282.82	697.99 ± 278.31
AEE	511.80 ± 262.88	497.87 ± 259.27	493.55 ± 256.20	281.26 ± 265.01	296.68 ± 273.80

Low: AW-64 low setting; Medium: AW-64 medium setting; High: AW-64 high setting; ACT: GT3X+ default setting; LFE: GT3X+ low extension frequency setting; AEE: activity energy expenditure; MVPA: moderate-vigorous physical activity.

Table 5
Intra-class correlations between GT3X+ and AW-64 settings for physical activity measures.

			AW-64		
			Low	Medium	High
GT3X+	ACT	Sedentary	0.20	0.31	0.25
		Light	0.17	0.18	0.17
		MVPA	0.06	0.06	0.06
		Counts/min	0.11	0.14	0.14
		AEE	0.37	0.37	0.37
	LFE	Sedentary	0.21	0.31	0.24
		Light	0.18	0.19	0.18
		MVPA	0.06	0.06	0.06
		Counts/min	0.13	0.16	0.15
		AEE	0.39	0.39	0.40

Low: AW-64 low setting; Medium: AW-64 medium setting; High: AW-64 high setting; ACT: GT3X+ default setting; LFE: GT3X+ low extension frequency setting; AEE: activity energy expenditure; MVPA: moderate-vigorous physical activity.

Table 6
Mean absolute percentage error (MAPE) of physical activity parameters between GT3X+ and AW-64 settings.

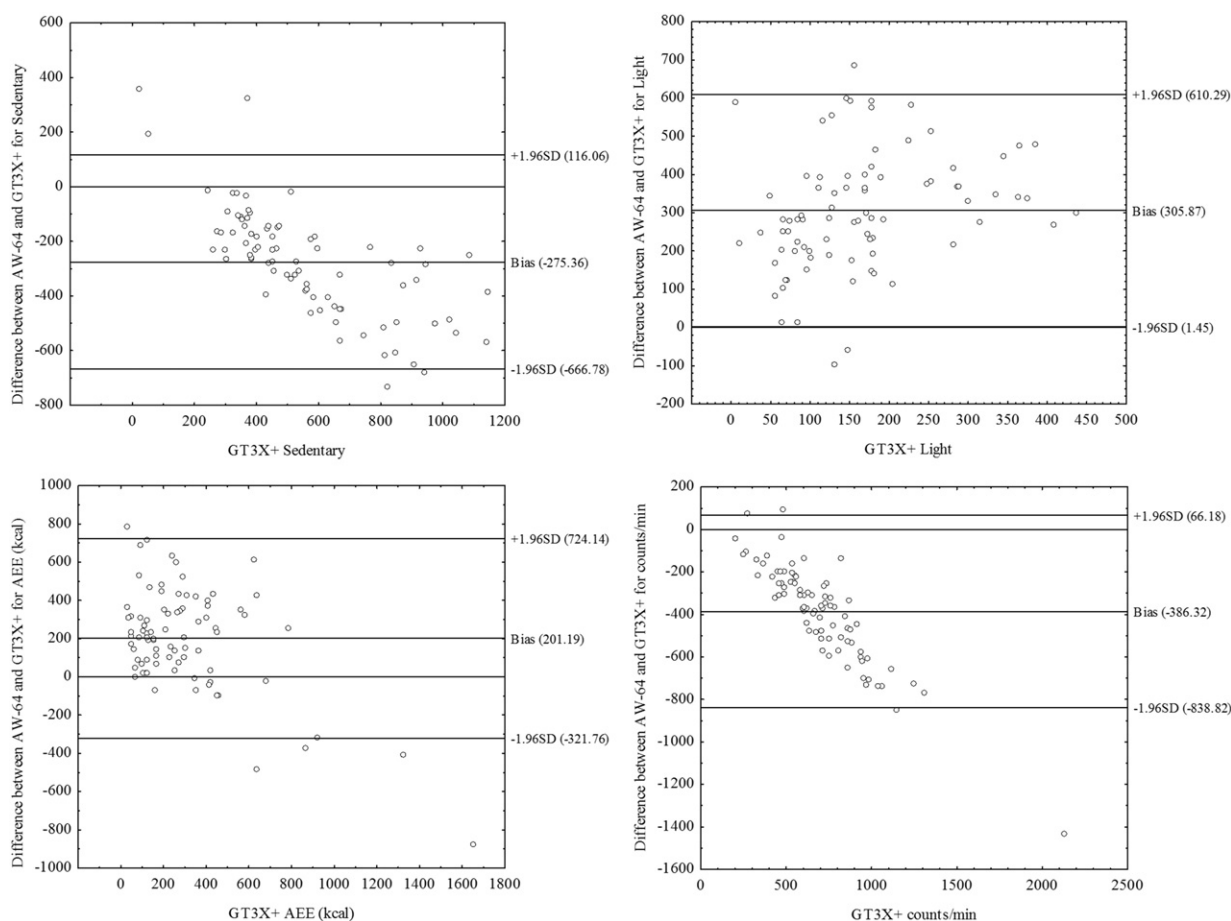|			AW-64		
			Low	Medium	High
GT3X+	ACT	Sedentary	45.81	93.41	71.97
		Light	66.32	65.39	65.18
		MVPA	1313.92	1337.17	1313.92
		Counts/min	144.60	124.14	132.10
		AEE	45.04	43.51	43.01
	LFE	Sedentary	16.39	19.06	20.10
		Light	65.72	64.77	64.56
		MVPA	1256.63	725.20	1256.63
		Counts/min	139.25	119.24	127.03
		AEE	42.03	40.41	39.89

Low: AW-64 low setting; Medium: AW-64 medium setting; High: AW-64 high setting; ACT: GT3X+ default setting; LFE: GT3X+ low extension frequency setting; AEE: activity energy expenditure; MVPA: moderate-vigorous physical activity.

the newest products from the same manufacture), uses a proprietary algorithm that applies a "PSG-derived correction factor" to improve sleep estimates beyond that of just summing the sleep/wake epochs (Actiware Sofware Actiwatch Instruction Manual).

Overall, the sleep results are consistent with our previous work showing a good reliability of the GT3X+ in assessing sleep when compared to concurrent PSG in a laboratory-setting [23]. In addition, the current study showed that in free-living settings where sleep periods longer than a daytime nap are most common, LFE seems to be slightly more comparable to the AW-64 settings than the default GT3X+

setting. This is probably due to the fact that LFE expands the lower end of the frequency range. This likely increases the count output at the lower frequency (which are usually considered outside the normal range), and it is consistent with previous work showing greater accuracy in detecting sedentary and light intensity activities using LFE than the default setting [43]. This is an important consideration for researchers who may wish to use the GT3X+ to measure sleep and wake activity across the 24 h day.

The AW-64 showed very poor agreement with the GT3X+ in detecting physical activity and sedentary behaviors. AW-64 showed moderate

Fig. 2. Bland-Altman plots of GT3X+ LFE setting and AW-64 Medium sensitivity for physical activity parameters. The y-axis indicates the differences between the AW-64 score minus the GT3X+ score, whereas the x-axis shows the average of their scoring. The Bias represents the mean difference between the devices for a specific parameter, with values above zero meaning an overestimation and the values below zero meaning an underestimation of the GT3X+ relative to AW-64. Upper and lower lines represent Bias ± 1.96 SD. Note that MVPA is not shown.

agreement in measuring AEE, but sedentary, light or moderate-vigorous physical activities were not comparable with the GT3X+ output. Analyses showed a dramatic overestimation of light intensity activity and underestimation of MVPA compared to the GT3X+. Also AEE was overestimated by the AW-64, but overall the ICCs, MAPE, and ANOVA indicate moderate agreement with the GT3X+, particularly with the LFE output. We observed some differences across AW-64 sensitivities with the Medium sensitivity resulting in the best estimation for physical activity and sedentary behavior.

The disagreement between monitors for sedentary/physical activity behaviors is likely due to two main reasons. First, the AW-64 and the newest model of the same brand are manufactured to be worn on the wrist and not on the hip. However, it has been shown that hip-worn monitors provide higher outputs than wrist-worn monitors [20]. Also, physical activity parameters are more accurate when an accelerometer is worn on the hip than on the wrist [11]. Given this, we used the GT3X+ placed on the hip as our reference for physical activity assessment. This decision was made in light of the extensive literature reporting about the validity of this product to measure both physical activity and sedentary behavior in a variety of populations, in both laboratory and free-living settings, when the device is fixed to the hip [22, 36, 44]. Second, we estimated sedentary, light-intensity, and moderate-vigorous physical activity using a published threshold-based algorithm derived from data of the wrist-worn Actical monitor [37], a device of the same manufacturer of AW-64.

In other words, our results indicated that both the monitor placement and the currently available algorithm may account for the differences we observed for physical behaviors and energy expenditure. Therefore, in order to use the AW-64 as a valid device to assess physical activity, new algorithms/threshold levels need to be developed.

The results of the current study should be interpreted in light of some limitations. As described above, the lack of uniform monitor placement (hip vs wrist) which may act as a confounding factor. Whereas the GT3X+ was placed on the hip during the day and on the wrist during sleep, the AW-64 was constantly worn on the wrist. However, while the GT3X+ has been designed to be worn either on the wrist, hip, arm, or ankle by default (see the device manual), the AW-64 is manufactured to be worn on the wrist. Also, even if the Actigraph has developed an algorithm to scale wrist-worn data to hip-worn estimations, recent studies using the Actigraph in children [45] as well as in adults [25, 46] have found that wrist and hip placements are not interchangeable. Specifically, hip placement is more precise for physical activity assessment [20] whereas wrist placement is more accurate when assessing sleep parameters [24]. Therefore, in the current study we decide to place the GT3X+ monitor in its recommended location for each condition (wrist for sleep and hip for physical activity), whereas AW-64 was worn in its default placement.

The current study also lacked gold standards for sleep (polysomnography) or physical activity (calorimetry). Therefore, even if the AW-64 and the GT3X+ are two valid and extensively used wearable monitors in the sleep and physical activity field respectively, there is likely substantial error in both measures. Also, in light of the lack of polysomnography, we used self-reported lights off and lights on to delimit our rest analysis windows. This method may have biased the sleep onset latency detection. However, it should be note that the bias was common across devices and settings. Moreover, this analysis is more appropriate for devices not equipped with event-markers, such as the GT3X+, for which researchers and clinicians have to rely on self-reported bed and wake time information. The lack of polysomnography screening means we cannot be completely sure of having not included participants with sleep problems, even if we carefully screened participants for excessive daytime sleepiness, sleep disorders, irregular sleep schedules and poor sleep via questionnaires and interview.

Notwithstanding these issues, this comparison is still useful given that these monitors are commonly used in epidemiological and intervention research to measure these behaviors in free-living contexts.

5. Conclusion

In conclusion, in the present study we tested whether two widely used and commercially-available actigraphs, both considered valid and reliable monitors in either the sleep or the physical activity field, might be interchangeable. Our results showed a good agreement of the GT3X+ with AW-64 for assessing sleep but a lack of congruence between AW-64 and GT3X+ for physical activity and sedentary behaviors. These results are probably due to the actigraph placement (wrist vs hip) as well as the algorithm employed to score the data. Future validation work of existing and emerging technologies that may hold promise for 24 h, continuous monitoring is needed. Moreover, considering that compliance is higher with wrist-worn actigraphs rather than with hip-worn monitors [4, 47], future feasibility studies that explore compliance and participant burden factors impacting 24 h monitoring with both wrist- and hip-worn monitors are warranted.

Conflicts of interest

Authors declare the absence of any conflicts of interest.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.physbeh.2016.01.034.

References

[1] J.M. Kelly, R.E. Strecker, M.T. Bianchi, Recent developments in home sleep-monitoring devices, ISRN Neurol. 2012 (2012) 768794.
[2] K.R. Westerterp, Physical activity and physical activity induced energy expenditure in humans: measurement, determinants, and effects, Front. Physiol. 4 (2013).
[3] M. Chesson Jr., M. Coleman, M. Lee-Chiong, D. Pancer, Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007, Sleep 30 (2007) 519–529.
[4] B. Ainsworth, L. Cahalin, M. Buman, R. Ross, The current state of physical activity assessment tools, Prog. Cardiovasc. Dis. 57 (2015) 387–395.
[5] A. Sadeh, Commentary: comparing actigraphy and parental report as measures of children's sleep, J. Pediatr. Psychol. 33 (2008) 406–407.
[6] N. Cellini, E.A. McDevitt, A.A. Ricker, K.M. Rowe, S.C. Mednick, Validation of an automated wireless system for sleep monitoring during daytime naps, Behav. Sleep. Med. 13 (2015) 157–168.
[7] L. Tonetti, N. Cellini, M. De Zambotti, M. Fabbri, M. Martoni, S. Fabregas, et al., Polysomnographic validation of a wireless dry headband technology for sleep monitoring, Physiol. Behav. 118 (2013) 185–188.
[8] L. Meltzer, H. Montgomery-Downs, S. Insana, C. Walsh, Use of actigraphy for assessment in pediatric sleep research, Sleep Med. Rev. 16 (2012) 463–475.
[9] S.P. Insana, D. Gozal, H.E. Montgomery-Downs, Invalidity of one actigraphy brand for identifying sleep and wake among infants, Sleep Med. 11 (2010) 191–196.
[10] A. Sadeh, The role and validity of actigraphy in sleep medicine: an update, Sleep Med. Rev. 15 (2011) 259–267.
[11] D. McMinn, R. Acharya, D.A. Rowe, S.R. Gray, J.L. Allan, Measuring activity energy expenditure: accuracy of the GT3X+ and actiheart monitors, Int. J. Exerc. Sci. 6 (2013) 5.
[12] M.P. Buman, A.C. King, Exercise as a treatment to enhance sleep, Am. J. Lifestyle Med. 4 (2010) 500–514.
[13] G. Atkinson, D. Davenne, Relationships between sleep, physical activity and human health, Physiol. Behav. 90 (2007) 229–235.
[14] P.T. Katzmarzyk, T.S. Church, C.L. Craig, C. Bouchard, Sitting time and mortality from all causes, cardiovascular disease, and cancer, Med. Sci. Sports Exerc. 41 (2009) 998–1005.
[15] A. Grøntved, F.B. Hu, Television viewing and risk of type 2 diabetes, cardiovascular disease, and all-cause mortality: a meta-analysis, J. Am. Med. Assoc. 305 (2011) 2448–2455.
[16] P. Bansil, E.V. Kuklina, R.K. Merritt, P.W. Yoon, Associations between sleep disorders, sleep duration, quality of sleep, and hypertension: results from the National Health and Nutrition Examination Survey, 2005 to 2008, J. Clin. Hypertens. 13 (2011) 739–743.

[17] M.P. Hoevenaar-Blom, A.M. Spijkerman, D. Kromhout, J.F. van den Berg, W.M. Verschuren, Sleep duration and sleep quality in relation to 12-year cardiovascular disease incidence: the MORGEN study, Sleep 34 (2011) 1487.

[18] J.M. Dzierzewski, M.P. Buman, P.R. Giacobbi, B.L. Roberts, A.T. Aiken-Morgan, M. Marsiske, et al., Exercise and sleep in community-dwelling older adults: evidence for a reciprocal relationship, J. Sleep Res. 23 (2014) 61–68.

[19] M.P. Buman, E.A. Winkler, J.M. Kurka, E.B. Hekler, C.M. Baldwin, N. Owen, et al., Reallocating time to sleep, sedentary behaviors, or active behaviors: associations with cardiovascular disease risk biomarkers, NHANES 2005–2006, Am. J. Epidemiol. (2013) kwt292.

[20] A. Rowlands, V. Stiles, Accelerometer counts and raw acceleration output in relation to mechanical loading, J. Biomech. 45 (2012) 448–454.

[21] K.M. Robusto, S.G. Trost, Comparison of three generations of ActiGraph™ activity monitors in children and adolescents, J. Sports Sci. 30 (2012) 1429–1435.

[22] S.G. Trost, K.L. McIver, R.R. Pate, Conducting accelerometer-based activity assessments in field-based research, Med. Sci. Sports Exerc. 37 (2005) S531–S543.

[23] N. Cellini, M.P. Buman, E.A. McDevitt, A.A. Ricker, S.C. Mednick, Direct comparison of two actigraphy devices with polysomnographically-recorded naps in healthy young adults, Chronobiol. Int. 30 (2013) 691–698.

[24] M. Zinkhan, K. Berger, S. Hense, M. Nagel, A. Obst, B. Koch, et al., Agreement of different methods for assessing sleep characteristics: a comparison of two actigraphs, wrist and hip placement, and self-report with polysomnography, Sleep Med. 15 (2014) 1107–1114.

[25] J.A. Slater, T. Botsis, J. Walsh, S. King, L.M. Straker, P.R. Eastwood, Assessing sleep using hip and wrist actigraphy, Sleep Biol. Rhythms. 13 (2015) 172–180.

[26] T.L. Rupp, T.J. Balkin, Comparison of Motionlogger Watch and Actiwatch actigraphs to polysomnography for sleep/wake estimation in healthy young adults, Behav. Res. Methods 43 (2011) 1152–1160.

[27] V. Natale, D. Léger, V. Martoni, V. Bayon, A. Erbacci, The role of actigraphy in the assessment of primary insomnia: a retrospective study, Sleep Med. 15 (2014) 111–115.

[28] M. Hyde, D.M. O'Driscoll, S. Binette, C. Galang, S.K. Tan, N. Verginis, et al., Validation of actigraphy for determining sleep and wake in children with sleep disordered breathing, J. Sleep Res. 16 (2007) 213–216.

[29] C.L. Comella, M. Morrissey, K. Janko, Nocturnal activity with nighttime pergolide in Parkinson disease: a controlled study using actigraphy, Neurology 64 (2005) 1450–1451.

[30] G.A. Dowling, J. Mastick, E.M. Hubbard, J.S. Luxenberg, R.L. Burr, Effect of timed bright light treatment for rest-activity disruption in institutionalized patients with Alzheimer's disease, Int. J. Geriatr. Psychiatry. 20 (2005) 738–743.

[31] G.A. Dowling, R.L. Burr, E.J. Van Someren, E.M. Hubbard, J.S. Luxenberg, J. Mastick, et al., Melatonin and bright-light treatment for rest–activity disruption in institutionalized patients with Alzheimer's disease, J. Am. Geriatr. Soc. 56 (2008) 239–246.

[32] M.W. Johns, A new method for measuring daytime sleepiness: the Epworth sleepiness scale, Sleep 14 (1991) 540–545.

[33] A. Sadeh, K. Sharkey, M. Carskadon, Activity-based sleep-wake identification: an empirical test of methodological issues, Sleep 17 (1994) 201–207.

[34] T. Morgenthaler, C. Alessi, L. Friedman, J. Owens, V. Kapur, B. Boehlecke, et al., Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007, Sleep 30 (2007) 519–529.

[35] M.J. Lambiase, K.P. Gabriel, Y.-F. Chang, L.H. Kuller, K.A. Matthews, Utility of actiwatch sleep monitor to assess waking movement behavior in older women, Med. Sci. Sports Exerc. 46 (2014) 2301–2307.

[36] P.S. Freedson, E. Melanson, J. Sirard, Calibration of the computer science and applications, Inc. accelerometer, Med. Sci. Sports Exerc. 30 (1998) 777–781.

[37] D.P. Heil, Predicting activity energy expenditure using the Actical® activity monitor, Res. Q. Exerc. Sport 77 (2006) 64–80.

[38] G. Atkinson, A.M. Nevill, Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine, Sports Med. 26 (1998) 217–238.

[39] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (1977) 159–174.

[40] J.-M. Lee, Y. Kim, G.J. Welk, Validity of consumer-based physical activity monitors, Med. Sci. Sports Exerc. 46 (2014) 1840–1848.

[41] J.M. Bland, D.G. Altman, Measuring agreement in method comparison studies, Stat. Methods Med. Res. 8 (1999) 135–160.

[42] S.P. Insana, S.S. Glowacki, H.E. Montgomery-Downs, Assessing the efficacy to conduct the multiple sleep latency test with actigraphy, Behav. Sleep. Med. 9 (2011) 257–265.

[43] M. Ried-Larsen, J.C. Brønd, S. Brage, B.H. Hansen, M. Grydeland, L.B. Andersen, et al., Mechanical and free living comparisons of four generations of the Actigraph activity monitor, Int. J. Behav. Nutr. Phys. Act. 9 (2012) 113.

[44] S. Kozey-Keadle, A. Libertine, K. Lyden, J. Staudenmayer, P.S. Freedson, Validation of wearable monitors for assessing sedentary behavior, Med. Sci. Sports Exerc. 43 (2011) 1561–1567.

[45] M.F. Hjorth, J.P. Chaput, C.T. Damsgaard, S.M. Dalskov, K.F. Michaelsen, I. Tetens, et al., Measure of sleep and physical activity by a single accelerometer: can a waist-worn Actigraph adequately measure sleep in children? Sleep Biol. Rhythms. 10 (2012) 328–335.

[46] S. Korpan, J. Schafer, K. Wilson, S. Webber, Effect of ActiGraph GT3X+ position and algorithm choice on step count accuracy in older adults, J. Aging Phys. Act. 23 (2014) 377–382.

[47] V.T. van Hees, F. Renström, A. Wright, A. Gradmark, M. Catt, K.Y. Chen, et al., Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer, PLoS ONE 6 (2011) e22922.