HOSTED BY

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Journal of Sport and Health Science xx (2016) 1–9

www.jshs.org.cn

Original article

Wearable monitors criterion validity for energy expenditure in sedentary and light activities

Florez-Pregonero Alberto [a,*], Meckes Nathanael [b], Buman Mathew [c], Ainsworth Barbara [c]

[a] Departamento de Formación, Facultad de Educación, Pontificia Universidad Javeriana, Bogota, Distrito Capital 11001000, Colombia

[b] Department of Exercise and Sport Science, School of Behavioral and Applied Sciences, Azuza Pacific University, Azuza, CA 91702, USA

[c] School of Nutrition and Health Promotion, Arizona State University, Phoenix, AZ 85004, USA

Received 10 November 2015; revised 8 April 2016; accepted 5 July 2016

Available online

Abstract

Background: Wearable monitors (WMs) are used to estimate the time spent in sedentary behaviors (SBs) and light-intensity physical activities (LPAs) and their associated energy cost; however, the accuracy of WMs in measuring behaviors on the lower end of the intensity spectrum is unclear. The aim of this study was to assess the validity of 3 WMs (ActiGraph GT3X+; *activ*PAL, and SenseWear 2) in estimating the intensity of SB and LPA in adults as compared with the criterion measure of oxygen uptake measured by indirect calorimetry (oxygen uptake, VO_2).

Methods: Sixteen participants (age: 25.38 ± 8.58 years) wore the ActiGraph GT3X+, *activ*PAL, and SenseWear devices during 7 sedentary-to-light activities. VO_2 (mL/kg/min) was estimated by means of a portable gas analyzer, Oxycon Mobile (Becton, Dickinson and Company, Franklin Lakes, NJ, USA). All data were transformed into metabolic equivalents and analyzed using mean percentage error, equivalence plots, Bland-Altman plots, kappa statistics, and sensitivity/specificity.

Results: Mean percentage error was lowest for the *activ*PAL for SB (14.9%) and LPA (9.3%) compared with other WMs, which were >21.2%. None of the WMs fell within the equivalency range of ±10% of the criterion mean value. Bland-Altman plots revealed narrower levels of agreement with all WMs for SB than for LPA. Kappa statistics were low for all WMs, and sensitivity and specificity varied by WM type.

Conclusion: None of the WMs tested in this study were equivalent with the criterion measure (VO_2) in estimating sedentary-to-light activities; however, the *activ*PAL had greater overall accuracy in measuring SB and LPA than did the ActiGraph and SenseWear monitors.

Keywords: Accelerometers; Accuracy; Low intensity; Metabolic estimations; Objective measurement; Sedentary behaviors

1. Introduction

Sedentary behavior (SB) is an important determinant of health.[1] Accurate assessment of this behavior is useful for epidemiologic research and to evaluate changes for interventions and programs.[2] Self-report has been the most common method of quantifying SB; however, its validity is still under assessment.[3,4] Therefore, objective measurement with sophisticated wearable monitors (WMs) has emerged to overcome self-reporting biases, yet many challenges accompany their use.[5–9] To date, the treatment and understanding of the data obtained from WMs is still very limited.[5,10] Furthermore, most of the available WMs have been extensively evaluated for accuracy in estimating moderate-to-vigorous physical activity (MVPA) and not SB or light-intensity physical activity (LPA).

Because many of the adults from developed and developing countries spend most of their time in SB and LPA,[11] it is critical to assess the validity of WMs in measuring SB and LPA. Early work in understanding energy expenditure (EE) has described the lack of ability of WMs to measure EE in the sedentary-to-light intensity spectrum.[12] More recently, Calabro et al.[13] assessed the validity of a variety of WMs in estimating EE during light-to-moderate intensity activities, finding a percentage error ranging from 9.5 to 30.5. Even though their work provides important information for considering whether to use a WM when there is interest in tracking low-intensity activities, several questions remain regarding which are the most valid and reliable objective wearable measures of SB and LPA.

Peer review under responsibility of Shanghai University of Sport.

* Corresponding author.

E-mail address: alberto.florez.p@gmail.com (F.-P. Alberto)

http://dx.doi.org/10.1016/j.jshs.2016.10.005

Please cite this article in press as: Florez-Pregonero Alberto, Meckes Nathanael, Buman Mathew, Ainsworth Barbara, Wearable monitors criterion validity for energy expenditure in sedentary and light activities, Journal of Sport and Health Science (2016), doi: 10.1016/j.jshs.2016.10.005

Currently, there are many types of WM brands available (e.g., ActiGraph, ActiGraph, Pensacola, FL, USA; *activ*PAL, PAL Technologies Ltd., Glasgow, UK; and SenseWear 2 (Body Media, Pittsburgh, PA, USA)) to measure PA and SB that have been extensively evaluated for accuracy in estimating MV PA. However, their ability to estimate EE on the lower end of the intensity spectrum, such as for SB and LPA, is less well known. For example, the ActiGraph, a triaxial accelerometer, measures acceleration in 3 individual axes (vertical, anteroposterior, and mediolateral) and provides activity counts for separate and composite vector magnitude of these 3 axes; however, the primary determination of SB by the ActiGraph is often based on only 1 axis, using an intensity threshold of <100 counts per minute (cpm). There has been some concern about the accuracy of this threshold because it has underestimated sitting time by 5%. Although 150 cpm seems to be a more accurate cutoff point for the ActiGraph WM,[14] there are several proposed cpm thresholds that classify SB in different studies: 50 cpm,[15] 100 cpm,[16] 150 cpm,[14] and 500 cpm.[17] Another monitor is the *activ*PAL PA logger, a uniaxial accelerometer and inclinometer that identifies walking, sitting, standing, steps, and instantaneous cadence.[18] The *activ*PAL has shown accuracy in distinguishing sitting or lying down from standing postures and in classifying time stepping;[14,19] however, the estimated metabolic equivalent (MET) values from the *activ*PAL at various speeds (2–4 mph) are significantly different ($p < 0.0001$) from the criterion of oxygen uptake.[20] A third example of a monitor to measure SB and LPA is the SenseWear 2, which integrates information from a biaxial accelerometer and other physiological sensors (heat flux, temperature, and galvanic skin response) to provide estimates of EE using a proprietary algorithm.[21] This WM overestimates EE at various walking and running speeds ranging from 2 to 8 mph ($p < 0.0001$) as compared with the criterion of oxygen uptake (VO_2).[22]

The accuracy (validity) for each of these WMs in estimating EE during sedentary-to-light activities is unclear. One way to assess validity of the WM is to compare its outputs against a criterion measure (criterion validity). The criterion validity describes the relationship between WM outputs and physiological measures that reflect more directly the energy cost of the activity. Thus, the goal of this study was to examine the validity of 3 WMs (ActiGraph GT3X+, *activ*PAL, and SenseWear 2) in estimating intensity for sedentary-to-light activities in adults as compared with oxygen uptake measured in mL/kg/min. We hypothesized that the validity of EE estimates made by the tested WM (ActiGraph, *activ*PAL, and SenseWear 2) would be low because most of the WMs are validated for measuring MVPA but not SB or LPA.

2. Materials and methods

2.1. Participant information

A convenience sample of 16 participants ($n = 8$ men, $n = 8$ women) with an age range of 19–47 years (25.38 ± 8.58 years), body mass index range of 18.8–35.0 kg/m^2 (24.6 ± 4.6 kg/m^2), no contraindications for exercise (assessed with the PA readiness questionnaire (PAR-Q)),[23] and ability to walk unassisted on a motorized treadmill at 2.0 mph participated in the study. Prior to participation, all participants read and signed an informed consent document approved by the Arizona State University Institutional Review Board.

2.2. Procedures

Participants were instructed to avoid vigorous exercise the day before the testing and to eat their usual diet. Each participant performed 7 sedentary-to-light activities in a randomly assigned order. Activities close to the LPA threshold of 1.5 METs were selected based on values listed in the 2011 Compendium of Physical Activities.[24] Every activity was performed for 7 min, with 4 min of rest between activities. Participants were instructed to be silent during the monitoring periods. The activities were performed twice, with at least 24 h between trials. Participants were instructed to perform the activities as follows:

1. Treadmill walking at 1.0 mph (0.45 m/s), 1.5 mph (0.67 m/s), and 2.0 mph (0.90 m/s)—walk using their normal gate at each speed and not using the handrails for support.
2. Cleaning a kitchen (cleaning)—simulate cleaning a kitchen and dishes using a dry rag. Tasks included clearing dishes off a counter, simulating washing and drying dishes, placing dishes in a cupboard, and wiping the counter.
3. Standing while reading (reading)—stand in place and read a book silently.
4. Sitting while typing (typing)—sit at a computer to type a given paragraph. Participants were instructed to sit up straight and maintain that posture while typing.
5. Sitting while gaming (gaming)—be seated and quietly play a board game, which required the participant to put 5 objects in a defined order. Participants also rolled a die and moved their game piece a certain number of spaces based on their score from ordering the objects. Participants competed against the researcher to more accurately simulate playing a board game.

2.3. WMs

Each participant wore the 3 WMs under assessment and the criterion monitor simultaneously during the 7 selected activities. The criterion measure, oxygen uptake in mL/kg/min, was determined with the Oxycon Mobile portable metabolic unit (CareFusion, San Diego, CA, USA);[25] the unit was calibrated before each test according to the manufacturer's specifications.

The ActiGraph was worn on an elastic belt on the right hip. The ActiGraph was initialized to collect data at 30 Hz. The *activ*PAL was worn on the anterior and medial portion of the right thigh attached to the skin by hypoallergenic medical tape. The SenseWear 2 was worn on the left upper arm of the individual using the factory-provided elastic strap.

2.4. Data management and processing

Researchers kept a written record of the time when each activity was performed; for example, walking 1 mph was performed from 1:00 p.m. to 1:07 p.m. When data collection was complete, data were downloaded from each of the WMs to a

Table 1
Calculations used to obtain METs from monitors and the criterion measure.

Monitor	Original units	Equation used to calculate METs
Oxycon Mobile	mL/kg/min	mL/kg/min / 3.5
ActiGraph	Counts per minute (cpm)	1.439008 + (0.000795 × cpm)[35]
*activ*PAL	MET·h	MET·h / 60
SenseWear 2	METs	No conversion needed

Abbreviation: MET = metabolic equivalent.

desktop computer. Data from 2 trials performed by each of the 16 participants were included for data analysis, resulting in a maximum of 32 trials.

To ensure that a steady state of VO_2 had been attained during each activity and to avoid small discrepancies between start and stop times for each activity, the first 2 min (Min 1–2) and the final minute of data (Min 7) were dropped from the analysis. Accordingly, Min 3–6 of each activity were utilized to identify the activity intensity for each WM. This process yielded four 1 min epochs for each subject in each activity.

Capabilities for data summarizing and measurement units are different among the selected WMs; as a result, data output lengths were standardized to a 1-min epoch and the measurement units were standardized to METs. A MET is defined as the energy cost of a specific activity divided by a standard resting EE of 3.5 ml/kg/min. Table 1 summarizes how the measurement units for the criterion and the WM output values were transformed into METs.

2.5. Statistical analysis

Analyses were conducted by averaging the four 1-min epochs of each activity into 1 variable reflecting the average energy cost for the activity. The variables were stratified into 2 groups according to their MET values: SB (<1.5 METs: reading, typing, and gaming) and LPA (≥1.5 METs: walking 1 mph, walking 1.5 mph, walking 2 mph, and cleaning). Because each participant completed 2 trials for each activity, we performed a test-retest reliability analysis (intraclass correlation coefficient, ICC) for each WM prior to comparison to the criterion measure.

Mean percentage error (MPE) was calculated to assess the proportion of error for each of the 3 WMs relative to the criterion measure. MPE was calculated using the equation MPE = ((measured score − true score) / true score) × 100. The true score was the criterion value (VO_2 in METs), and the measured score was the MET value obtained from each WM. A positive MPE indicated a MET value overestimation for the WM, whereas a negative MPE indicated a MET value underestimation for the WM.[26]

Equivalency testing was used to examine whether the MET value for each of the WMs was statistically equivalent to the criterion MET value. Equivalence testing is an alternative approach to testing for significant differences between means.[27] Equivalence testing requires researchers to identify a clinically meaningful range (i.e., equivalence zone) that permits comparisons between the values for WM and the criterion values in the equivalence zone. If the full 90% confidence interval (CI) range of the WM lies within the equivalence zone, then it can be concluded (with $\alpha < 0.05$) that the WM value is equivalent to the criterion value. Based on previously published work,[28] we established ±10% of the criterion mean MET value as the equivalence zone; by choosing the same values, we will facilitate comparisons when needed.

Bland-Altman plots[29] were used to show the distribution of the error and to assess systematic variation between the criterion MET value and each WM's MET value. The Bland-Altman plot is a graphical method to compare 2 measurement techniques. In this method, the difference score between 2 measures (i.e., criterion MET value minus the WM's MET value) is plotted against the averages of the 2 measures. The error distribution can be observed within 3 horizontal reference lines that are drawn: mean difference (zero deviation line), upper limit of agreement (+1.96 standard deviation (SD)), and lower limit of agreement (–1.96 SD). To provide a statistical reference for systematic bias between the criterion MET value and each WM's MET value, the difference score between methods is regressed on the average of the 2 scores. Thus, the regression line provides information about whether the WM value becomes more or less accurate at varying levels of the criterion value. A flat regression line in the Bland-Altman plot indicates that the MET estimate of the WM varies in the same manner as the criterion value; a positive slope indicates that the WM is positively biased when compared with the criterion MET value; and a negative slope indicates that the WM is negatively biased when compared with the criterion MET value. The White test was used to examine the presence of heteroskedasticity.[30]

Kappa statistics was used to observe agreement between each WM and the criterion value for classifying activities while taking into account the agreement occurring by chance.[31] Data were dichotomous indicator variables for SB (0) or LPA (1). The kappa value interpretation is based on recommendations from Landis and Koch[32] as follows: 0–0.2 = slight agreement, 0.2–0.4 = fair agreement, 0.4–0.6 = moderate agreement, 0.6–0.8 = substantial agreement, and 0.8–1.0 = almost perfect agreement.

Sensitivity and specificity were calculated to measure the accuracy of the WM in classifying an activity as SB or LPA. Sensitivity is the proportion of true positives (i.e., correct MET category for the WM and the criterion value) that are correctly identified by the WM (true positive proportion). Sensitivity was calculated using the formula: sensitivity = true positives / (true positives + false negatives).[33] A sensitivity value close to 1 shows that the WM is able to accurately classify a high proportion of the activities into the correct category; a sensitivity value close to 0 indicates that the WM fails to classify activities into the correct category. Specificity refers to the proportion of true negatives (i.e., correct exclusion of the WM and the criterion value from the incorrect category) that are correctly classified by the WM (true negative proportion). Specificity was calculated using the formula: specificity = true negatives / (false positives + true negatives).[33] A specificity value close to 1 shows that the WM is able to exclude a high proportion of the activities from being classified into the incorrect category. A specificity value close to 0 indicates that the WM is unable to exclude activities from being classified into the incorrect

Please cite this article in press as: Florez-Pregonero Alberto, Meckes Nathanael, Buman Mathew, Ainsworth Barbara, Wearable monitors criterion validity for energy expenditure in sedentary and light activities, Journal of Sport and Health Science (2016), doi: 10.1016/j.jshs.2016.10.005

category. Significance was set at the $p < 0.05$ probability level. All analyses were performed using SPSS Version 21 (IBM Corporation, Armonk, NY, USA) and SAS Version 9.3 (SAS Institute Inc., Cary, NC, USA).

3. Results

ICC test-retest values were high for all WMs (0.94, 0.97, 0.99, and 0.85 for Oxycon Mobile, ActiGraph, *activ*PAL, and SenseWear 2, respectively). Table 2 present means, SDs, and 95% CIs in METs for the criterion and all WMs under assessment for SB and LPA, respectively. MPE is presented for the ActiGraph, *activ*PAL, and SenseWear WMs referenced to the criterion value. For both SB and LPA, MPEs were lowest for the *activ*PAL and highest for the SenseWear monitors. When SB and LPA were combined, MPEs were lowest for the ActiGraph and highest for the SenseWear.

Based on the equivalence plots listed in Fig. 1, none of the WMs (and their associated CIs) fell within the equivalency

Table 2
Measured MET values (mean ± SD) and MPE for SB and LPA (both <1.5 METs).

	*n**	M	SD	95%CI	MPE
Criterion					
SB					
Standing, reading	30	1.13	0.18	1.07–1.20	NA
Sitting, typing	31	1.25	0.17	1.19–1.32	NA
Sitting, board games	30	1.17	0.16	1.11–1.23	NA
All sedentary activities combined	91	1.18	1.18	1.18	NA
LPA					
Walking 1 mph	31	2.19	0.27	2.09–2.28	NA
Walking 1.5 mph	30	2.46	0.28	2.35–2.56	NA
Walking 1.5 mph	30	2.74	2.74	2.63–2.85	NA
Walking 1.5 mph	30	1.68	0.32	1.56–1.80	NA
All light activities combined	121	2.26	0.48	2.17–2.35	NA
ActiGraph					
SB					
Standing, reading	21	1.44	0.00	1.44–1.44	32.48
Sitting, typing	22	1.44	0.01	1.44–1.45	12.96
Sitting, board games	21	1.44	0.01	1.44–1.45	21.67
All sedentary activities combined	64	1.44	0.01	1.43–1.44	22.22
LPA					
Walking 1 mph	22	1.55	0.10	1.51–1.60	−29.88
Walking 1.5 mph	21	1.80	0.20	1.71–1.89	−27.42
Walking 1.5 mph	21	2.35	0.32	2.20–2.49	−15.71
Walking 1.5 mph	21	1.47	0.03	1.45–1.48	−11.20
All light activities combined	85	1.78	0.39	1.70–1.87	−21.15
activPAL					
SB					
Standing, reading	15	1.40	0.00	1.40–1.40	32.34
Sitting, typing	16	1.25	0.00	1.25–1.25	0.98
Sitting, board games	15	1.27	0.05	1.24–1.30	12.36
All sedentary activities combined	46	1.30	0.07	1.28–1.32	14.89
LPA					
Walking 1 mph	16	2.22	0.44	1.98–2.46	0.06
Walking 1.5 mph	16	2.94	0.50	2.68–3.21	27.10
Walking 1.5 mph	15	3.41	0.06	3.38–3.45	23.88
Walking 1.5 mph	15	1.43	0.04	1.41–1.46	−13.17
All light activities combined	62	2.50	0.81	2.29–2.71	9.30
SenseWear 2					
SB					
Standing, reading	29	1.07	0.07	1.05–1.10	−2.55
Sitting, typing	30	1.96	0.77	1.68–2.25	56.05
Sitting, board games	29	1.67	0.63	1.42–1.91	41.05
All sedentary activities combined	88	1.57	0.68	1.42–1.71	31.79
LPA					
Walking 1 mph	29	3.06	3.06	2.86–3.26	40.68
Walking 1.5 mph	29	3.45	0.47	3.27–3.63	41.14
Walking 1.5 mph	29	3.86	0.49	3.68–4.05	42.36
Walking 1.5 mph	29	3.04	0.74	2.76–3.32	82.18
All light activities combined	116	3.35	3.35	3.23–3.47	51.58

Note: *The number of valid data points is different owing to instrument error.
Abbreviations: CI = confidence interval; LPA = light-intensity physical activity; M = mean; METs = metabolic equivalents; MPE = mean percentage error; NA = not applicable; SB = sedentary behavior; SD = standard deviation.

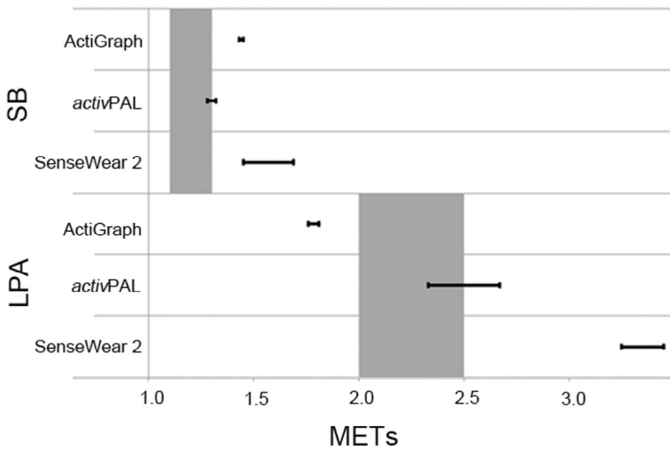

Fig. 1. Equivalence plots for sedentary behavior (SB) and light-intensity physical activity (LPA) compared with the criterion measure. Gray areas represent ±10% for the criterion mean (equivalence zone), and black bars represents 90% confidence interval for the test monitor. METs, metabolic equivalents.

range of ±10% for the criterion mean. The ActiGraph fell above the equivalence zone for SB and below the zone for LPA; the *activ*PAL provided estimates closest to the equivalency range for both SB and LPA; and the SenseWear 2 was over the equivalence range for both SB and LPA.

Bland-Altman plots (Fig. 2) revealed narrower levels of agreement for the WM when measuring SB (0.56, 0.55, and 1.62 METs for ActiGraph, *activ*PAL, and SenseWear 2, respectively) than when measuring LPA (1.42, 1.31, and 2.20 METs for ActiGraph, *activ*PAL, and SenseWear 2, respectively). For SB, the ActiGraph and the *activ*PAL had no pronounced variation across the intensity range, but the SenseWear 2 showed a slight cluster of data points below the mean difference line. The variation for LPA was greater for all the devices compared with the variation observed in SB; the ActiGraph had greater variation at higher intensity levels, with a negative slope indicating a negative bias for EE as the intensity levels increased. Heteroskedasticity was found for the *activ*PAL ($p = 0.11$) and SenseWear 2 ($p = 0.30$) for SB but not for LPA.

Table 3 shows the kappa statistics for agreement between the WM and the criterion measure to classify SB and LPA as well as results for sensitivity and specificity. There was a slight overall agreement among the instruments for measuring SB. For LPA, the agreement was fair for the ActiGraph and moderate for the *activ*PAL. When data for SB and LPA were combined, the agreement increased markedly. For SB, both the ActiGraph and the *activ*PAL had high sensitivity but low specificity. For LPA, both the ActiGraph and the *activ*PAL had fair sensitivity and good specificity; the SenseWear had good sensitivity but low specificity.

Table 3
Kappa statistics, sensitivity, and specificity for monitor MET values compared with criterion MET values.

	ActiGraph	*activ*PAL	SenseWear
SB			
Kappa statistics	−0.03 ($p = 0.02$)	0 (p = NA)	0.11 ($p = 0.08$)
Sensitivity	0.98	1.00	0.65
Specificity	0.00	0.00	0.66
LPA			
Kappa statistics	0.37 ($p = 0.01$)	0.53 ($p = 0.14$)	0 (p = NA)
Sensitivity	0.73	0.84	1.00
Specificity	1.00	1.00	0.00
Combined			
Kappa statistics	0.64 ($p = 0.06$)	0.80 ($p = 0.06$)	0.58 ($p = 0.05$)
Sensitivity	0.98	1.00	0.57
Specificity	0.68	0.81	0.98

Abbreviations: LPA = light-intensity physical activity; MET = metabolic equivalent; NA = not applicable; SB = sedentary behavior.

4. Discussion

The aim of the present study was to examine the accuracy of 3 WMs (ActiGraph GT3X+, *activ*PAL, and SenseWear 2) in estimating EE during SB and LPA in adults as compared with oxygen uptake measured by indirect calorimetry. The results showed overall low accuracy of the 3 WMs in estimating EE in METs. These findings emphasize the need for more refinements in the low spectrum of the EE measurements given the necessity of accurately estimating SB and/or LPA in relation to mortality and chronic diseases.[34] The analyses are relevant, because estimations made by the tested WMs are often used in PA research and commonly used to quantify behaviors in the lower range of intensity. Although the WM validity and reliability has been demonstrated in the moderate-to-vigorous EE spectrum, the tested WMs showed considerable limitations in measuring the metabolic cost of SB and LPA. For example, Calabró et al.[13] examined the validity of EE estimates during sedentary-to-moderate intensity activities for different monitors compared with the Oxycon Mobile. They reported a 25.5% and 22.2% underestimation for the ActiGraph and the *activ*PAL monitors, respectively. Their magnitude of underestimation is similar to what we found in the current study for the ActiGraph (21.2%) but differs for the *activ*PAL (9.3%). The discrepancies may be explained by the fact that their protocol included less-structured activities, which could have increased the amount of error for the *activ*PAL monitor. Similarly, Kozey-Keadle et al.[14] conducted a study to examine the validity of 2 monitors in classifying SB against direct observation as the criterion. They found that both the ActiGraph and the *activ*PAL monitors underestimated time spent in SB by 4.9% and 2.8%, respectively. They also tested the monitors for their ability to detect changes between sedentary and active pursuits. They found that the *activ*PAL was more precise in measuring time in SB and more sensitive in detecting reductions in sitting time. Even though the Kozey-Keadle et al. study[14] and the current study used different metrics (percentage bias *vs*. MPE) and outcomes (time spent in SB *vs*. EE), the results are similar in the sense that the *activ*PAL monitor performed best when estimating SB.

Possible explanations for the measurement error observed in the current study are that arm movements related to certain activities might cause the SenseWear not to differentiate arm movements during activities such as typing from free

Please cite this article in press as: Florez-Pregonero Alberto, Meckes Nathanael, Buman Mathew, Ainsworth Barbara, Wearable monitors criterion validity for energy expenditure in sedentary and light activities, Journal of Sport and Health Science (2016), doi: 10.1016/j.jshs.2016.10.005

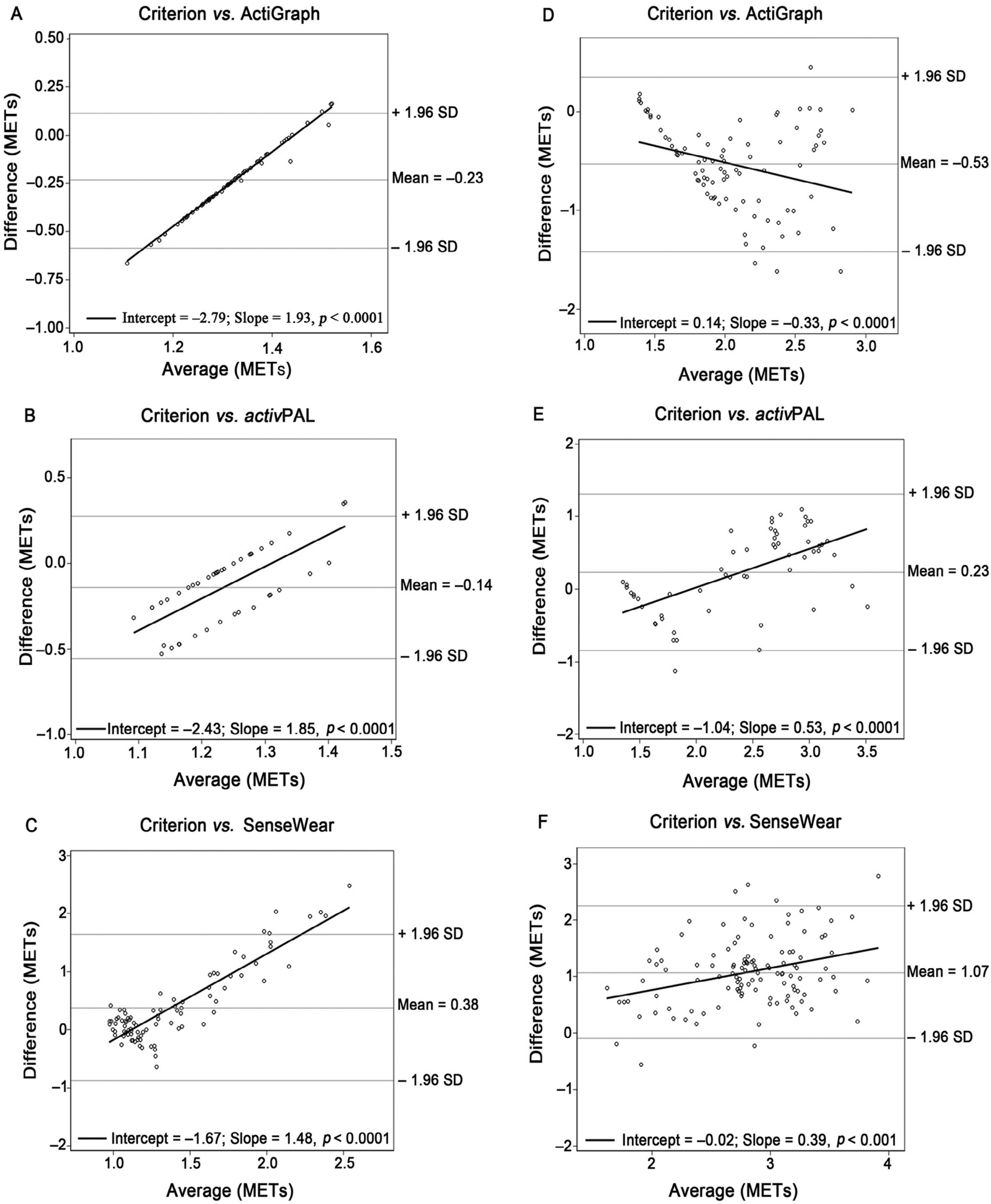

Fig. 2. Bland-Altman plots for sedentary behavior (SB) metabolic equivalents (METs) (A–C) and light-intensity physical activity (LPA) values (D–F) compared with the criterion value. SD, standard deviation.

Please cite this article in press as: Florez-Pregonero Alberto, Meckes Nathanael, Buman Mathew, Ainsworth Barbara, Wearable monitors criterion validity for energy expenditure in sedentary and light activities, Journal of Sport and Health Science (2016), doi: 10.1016/j.jshs.2016.10.005

ambulation. In addition, a small range of motion for the hip while walking at slow speeds on the treadmill may cause the ActiGraph to misclassify some LPAs as SBs. Another possible explanation of error measurement for the ActiGraph relates to the characteristics of the Freedson prediction equation used to estimate EE values.[35] The Freedson equation was validated using treadmill walking and running activities ranging from 3 to 9 mph. The EE estimates may be less accurate with lower-intensity activities. We acknowledge the availability of other equations to estimate METs from the ActiGraph counts per minute in adults;[36–41] however, we are unaware of an equation to estimate EE during SB-to-LPA for the ActiGraph. Thus, we chose Freedson's equation[35] given its common use in the field and its validity ($R^2 = 0.82$; Standard Error of the Estimate (SEE) = ±1.12 METs).

Among the 3 assessed WMs, only the *activ*PAL was located in the anteromedial thigh. The *activ*PAL showed the lowest amount of error compared with the other WMs (MPE = 14.89, MPE = 9.3 for SB and LPA, respectively). Thus, locating WMs in the lower limbs seemed to be more suitable when measuring SB and LPA. These findings need further confirmation by means of an WM placed in even more distal locations, such as the ankle. Alternatively, the *activ*PAL may include a more sensitive transducer than the ActiGraph or SenseWear WMs that may make it more suitable for measuring SB and LPA.[42] However, this assertion has to be tested, because we investigated the validity of the equations of each monitor not the transducer itself.

Interestingly, less error distribution was observed in the tested WMs for SB than for LPA. This was demonstrated by the Bland-Altman plots, which revealed narrower levels of agreement for SB than LPA. In particular, the lowest limits of agreement were for the *activ*PAL as compared with the other WMs (0.55 METs for the SB and 1.31 METs for LPA). On the other hand, the heteroskedasticity test revealed that the variance of the residuals was not homogeneous for the *activ*PAL and SenseWear in SB. These results suggest a systematic bias in the 2 WMs for assessing SB. In other words, the *activ*PAL and SenseWear WMs seem to have positive bias with heterogeneous error variation when assessing SB. Future research should be conducted to assess the sources of variability on the measurement of SB.

Although all WMs had high test-retest reliability (ICC = 0.94, ICC = 0.97, ICC = 0.99, and ICC = 0.85 for the Oxycon Mobile, ActiGraph, *activ*PAL, and SenseWear, respectively), all WMs showed low agreement with the criterion measure in classifying activities as either SB or LPA. When SB and LPA were combined into 1 category, the agreement of the WMs tended to be better than when evaluated separately (see kappa statistics in Table 3). In a similar manner, the accuracy of the devices seemed to improve when SB and LPA were combined (see sensitivity and specificity results in Table 3). This may have been due to having more data points with a greater range of values from low- to light-intensity EE. It may also be due to reduced variation in movement when performing SBs as compared with LPAs. These results highlight the importance of refining WM accuracy for the lower spectrum of the EE (sedentary to light).

When assessing the validity of WMs in measuring SB, researchers should pay special attention to criterion selection regarding the complexity of the behavior measured. Any waking behavior is characterized by an EE ≤1.5 METs while in a sitting or reclining posture.[43] As a consequence, and whenever possible, a combination of criterion measures should be considered (e.g., VO_2 for EE and direct observation for postural allocation). In the current study we aimed to examine the accuracy of WMs in estimating EE of sedentary-to-light activities; thus, a criterion such as the VO_2 was needed to detect small differences between SB and LPA. However, a more comprehensive approach in classifying SB and LPA should also include the assessment of posture to fully address the ability of a device to detect SB. Among the selected activities for the current study there was one activity, standing while reading, that we classified as an SB. We considered that the very low EE of 1.13 METs and the lack of motion made this activity more of an SB than an LPA. However, this assertion may not be applicable to everybody, because the EE for standing may differ according to individual characteristics.[44]

There were several strengths to this study. First, participants wore the 3 WMs and the Oxycon Mobile simultaneously so each activity could be monitored within a laboratory setting. Second, activities were randomized to prevent systematic bias in the measurement, which allowed the results to improve in accuracy. Last, activities were selected to be near the light-intensity activity threshold of >1.5 METs. This insured that activities performed would aid in understanding the accuracy of estimating EE in the lower end of the spectrum (sedentary to light).

With respect to the results for EE estimations for this study, it should be noted that the participants' resting metabolic rates were not measured, and we used the standard MET value of 3.5 mL/kg/min to estimate resting EE units. This may have introduced error, resulting in an overestimation of resting EE (10% and 15% for men and women, respectively) as reported in a recent systematic review.[45] We also observed that the MET values in the lower levels of the intensity spectrum for SB activities were quite homogeneous, which implies the need to reduce the variance to obtain substantial agreement. Additionally, we would like to note that the 7 sedentary-to-light activities may not be representative of the whole spectrum of sedentary-to-light EE; however, they were thoughtfully selected and randomly assigned to avoid introducing systematic error. We acknowledge that the participants comprised a convenience sample of healthy adults and that data were obtained in a laboratory setting with staged activities, limiting generalization of the results to other populations (e.g., older adults). Missing data were caused by problems with WM initialization and an inability to save data to a spreadsheet. Although missing data are a limitation, the data loss was random and did not represent a systematic bias.

5. Conclusion

As growing evidence demonstrates the associations between SB and morbidity and mortality, more research and refinements in EE estimations and in the ability of WMs to record SB-to-LPA is needed. Based on equivalency testing, none of the WMs tested in this study was equivalent with the criterion measure of

Please cite this article in press as: Florez-Pregonero Alberto, Meckes Nathanael, Buman Mathew, Ainsworth Barbara, Wearable monitors criterion validity for energy expenditure in sedentary and light activities, Journal of Sport and Health Science (2016), doi: 10.1016/j.jshs.2016.10.005

oxygen uptake for estimating EE in SB-to-LPA. However, among the WMs tested, the *activ*PAL had the highest overall criterion validity to measure both SB and LPA as compared with the ActiGraph and SenseWear WMs. This study provides support for the use of *activ*PAL in studying SB and LPA in healthy adults.

Acknowledgment

The authors would like to thank the participants of this study. There are no funding sources to state.

Authors' contributions

MN, BM, and AB contributed to design of the study and data collection. FPA, MN, BM, and AB contributed to data analysis and drafted the manuscript. All authors have read and approved the final version of the manuscript and agree with the order of presentation.

Competing interests

None of the authors declare competing financial interests.

References

1. Tremblay MS, Colley RC, Saunders TJ, Healy GN, Owen N. Physiological and health implications of a sedentary lifestyle. *Appl Physiol Nutr Metab* 2010;**35**:725–40.
2. Pate RR, O'Neill JR, Lobelo F. The evolving definition of "sedentary." *Exerc Sport Sci Rev* 2008;**36**:173–8.
3. Lynch BM, Friedenreich CM, Khandwala F, Liu A, Nicholas J, Csizmadi I. Development and testing of a past year measure of sedentary behavior: the SIT-Q. *BMC Public Health* 2014;**14**:899. doi:10.1186/1471-2458-14-899
4. Marshall AL, Miller YD, Burton NW, Brown WJ. Measuring total and domain-specific sitting: a study of reliability and validity. *Med Sci Sports Exerc* 2010;**42**:1094–102.
5. Ward DS, Evenson KR, Vaughn A, Rodgers AB, Troiano RP. Accelerometer use in physical activity: best practices and research recommendations. *Med Sci Sports Exerc* 2005;**37**(Suppl. 11):S582–8.
6. Healy GN, Clark BK, Winkler EA, Gardiner PA, Brown WJ, Matthews CE. Measurement of adults' sedentary time in population-based studies. *Am J Prev Med* 2011;**41**:216–27.
7. Lyden K, Kozey SL, Staudenmeyer JW, Freedson PS. A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. *Eur J Appl Physiol* 2011;**111**:187–201.
8. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc* 2008;**40**:181–8.
9. Chen KY, Bassett Jr DR. The technology of accelerometry-based activity monitors: current and future. *Med Sci Sports Exerc* 2005;**37**(11 Suppl.): S490–500.
10. Freedson P, Bowles HR, Troiano R, Haskell W. Assessment of physical activity using wearable monitors: recommendations for monitor calibration and use in the field. *Med Sci Sports Exerc* 2012;**44**(Suppl. 1):S1–4.
11. Dempsey PC, Owen N, Biddle SJ, Dunstan DW. Managing sedentary behavior to reduce the risk of diabetes and cardiovascular disease. *Curr Diab Rep* 2014;**14**:522. doi:10.1007/s11892-014-0522-0
12. Montoye HJ, Washburn R, Servais S, Ertl A, Webster JG, Nagle FJ. Estimation of energy expenditure by a portable accelerometer. *Med Sci Sports Exerc* 1983;**15**:403–7.
13. Calabró M, Lee JM, Saint-Maurice PF, Yoo H, Welk GJ. Validity of physical activity monitors for assessing lower intensity activity in adults. *Int J Behav Nutr Phys Act* 2014;**11**:119. doi:10.1186/s12966-014-0119-7
14. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson PS. Validation of wearable monitors for assessing sedentary behavior. *Med Sci Sports Exerc* 2011;**43**:1561–7.
15. Crouter SE, Clowers KG, Bassett Jr DR. A novel method for using accelerometer data to predict energy expenditure. *J Appl Physiol* 2006;**100**:1324–31.
16. Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR, et al. Amount of time spent in sedentary behaviors in the United States, 2003–2004. *Am J Epidemiol* 2008;**167**:875–81.
17. Silva P, Aires L, Santos RM, Vale S, Welk G, Mota J. Lifespan snapshot of physical activity assessed by accelerometry in Porto. *J Phys Act Health* 2011;**8**:352–60.
18. Grant PM, Ryan CG, Tigbe WW, Granat MH. The validation of a novel activity monitor in the measurement of posture and motion during everyday activities. *Br J Sports Med* 2006;**40**:992–7.
19. Hart TL, McClain JJ, Tudor-Locke C. Controlled and free-living evaluation of objective measures of sedentary and active behaviors. *J Phys Act Health* 2011;**8**:848–57.
20. Harrington DM, Welk GJ, Donnelly AE. Validation of MET estimates and step measurement using the ActivPAL physical activity logger. *J Sports Sci* 2011;**29**:627–33.
21. Johannsen DL, Calabro MA, Stewart J, Franke W, Rood JC, Welk GJ. Accuracy of armband monitors for measuring daily energy expenditure in healthy adults. *Med Sci Sports Exerc* 2010;**42**:2134–40.
22. King GA, Torres N, Potter C, Brooks TJ, Coleman KJ. Comparison of activity monitors to estimate energy cost of treadmill exercise. *Med Sci Sports Exerc* 2004;**36**:1244–51.
23. Shephard RJ. PAR-Q, Canadian home fitness test and exercise screening alternatives. *Sports Med* 1988;**5**:185–95.
24. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Basset Jr DR, Tudor-Locke C, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med Sci Sports Exerc* 2011;**43**:1575–81.
25. Perret C, Mueller G. Validation of a new portable ergospirometric device (Oxycon Mobile) during exercise. *Int J Sports Med* 2006;**27**:363–7.
26. Hartsell H, Fitzpatrick D, Brand R, Frantz R, Saltzman C. Accuracy of a custom-designed activity monitor: implications for diabetic foot ulcer healing. *J Rehabil Res Dev* 2002;**39**:395–400.
27. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokinet Biopharm* 1987;**15**:657–80.
28. Lee JM, Kim Y, Welk GJ. Validity of consumer-based physical activity monitors. *Med Sci Sports Exerc* 2014;**46**:1840–8.
29. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**1**:307–10.
30. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 1980;**48**:817–38.
31. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;**20**:37–46.
32. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.
33. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol* 2008;**56**:45–50.
34. Owen N, Healy GN, Matthews CE, Dunstan DW. Too much sitting: the population health science of sedentary behavior. *Exerc Sport Sci Rev* 2010;**38**:105–13.
35. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc* 1998;**30**:777–81.
36. Crouter SE, Kuffel E, Haas JD, Frongillo EA, Bassett Jr DR. Refined two-regression model for the ActiGraph accelerometer. *Med Sci Sports Exerc* 2010;**42**:1029–37.
37. Hendelman D, Miller K, Baggett C, Debold E, Freedson P. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Med Sci Sports Exerc* 2000;**32**(Suppl. 9):S442–9.
38. Swartz AM, Strath SJ, Bassett Jr DR, O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Med Sci Sports Exerc* 2000;**32**(Suppl. 9):S450–6.

39. Leenders NY, Nelson TE, Sherman WM. Ability of different physical activity monitors to detect movement during treadmill walking. *Int J Sports Med* 2003;**24**:43–50.
40. Yngve A, Nilsson A, Sjostrom M, Ekelund U. Effect of monitor placement and of activity setting on the MTI accelerometer output. *Med Sci Sports Exerc* 2003;**35**:320–6.
41. Brooks AG, Gunn SM, Withers RT, Gore CJ, Plummer JL. Predicting walking METs and energy expenditure from speed or accelerometry. *Med Sci Sports Exerc* 2005;**37**:1216–23.
42. Chen KY, Janz KF, Zhu W, Brychta RJ. Redefining the roles of sensors in objective physical activity monitoring. *Med Sci Sports Exerc* 2012; **44**(Suppl. 1):S13–23.
43. Sedentary Behaviour Research Network. Letter to the editor: standardized use of the terms "sedentary" and "sedentary behaviours." *Appl Physiol Nutr Metab* 2012;**37**:540–2.
44. Miles-Chan JL, Sarafian D, Montani JP, Schutz Y, Dulloo A. Heterogeneity in the energy cost of posture maintenance during standing relative to sitting: phenotyping according to magnitude and time-course. *PLoS One* 2013;**8**:e65827. doi:10.1371/journal.pone.0065827
45. McMurray RG, Soares J, Caspersen CJ, McCurdy T. Examining variations of resting metabolic rate of adults: a public health perspective. *Med Sci Sports Exerc* 2014;**46**:1352–8.

Please cite this article in press as: Florez-Pregonero Alberto, Meckes Nathanael, Buman Mathew, Ainsworth Barbara, Wearable monitors criterion validity for energy expenditure in sedentary and light activities, Journal of Sport and Health Science (2016), doi: 10.1016/j.jshs.2016.10.005