Validity of an Integrative Method for Processing Physical Activity Data

LAURA D. ELLINGSON[1,2], ISAAC J. SCHWABACHER[2,3], YOUNGWON KIM[1,4], GREGORY J. WELK[1], and DANE B. COOK[2,3]

[1]Department of Kinesiology, Iowa State University, Ames, IA; [2]Department of Kinesiology, University of Wisconsin-Madison, Madison, WI; [3]William S. Middleton Memorial Veterans Hospital, Madison, WI; and [4]MRC Epidemiology Unit, University of Cambridge, Cambridge, UNITED KINGDOM

ABSTRACT

ELLINGSON, L. D., I. J. SCHWABACHER, Y. KIM, G. J. WELK, and D. B. COOK. Validity of an Integrative Method for Processing Physical Activity Data. *Med. Sci. Sports Exerc.*, Vol. 48, No. 8, pp. 1629–1638, 2016. Accurate assessments of both physical activity and sedentary behaviors are crucial to understand the health consequences of movement patterns and to track changes over time and in response to interventions. **Purpose**: The study evaluates the validity of an integrative, machine learning method for processing activity monitor data in relation to a portable metabolic analyzer (Oxycon mobile [OM]) and direct observation (DO). **Methods**: Forty-nine adults (age 18–40 yr) each completed 5-min bouts of 15 activities ranging from sedentary to vigorous intensity in a laboratory setting while wearing ActiGraph (AG) on the hip, activPAL on the thigh, and OM. Estimates of energy expenditure (EE) and categorization of activity intensity were obtained from the AG processed with Lyden's sojourn (SOJ) method and from our new sojourns including posture (SIP) method, which integrates output from the AG and activPAL. Classification accuracy and estimates of EE were then compared with criterion measures (OM and DO) using confusion matrices and comparisons of the mean absolute error of log-transformed data (MAE ln Q). **Results**: The SIP method had a higher overall classification agreement (79%, 95% CI = 75%–82%) than the SOJ (56%, 95% CI = 52%–59%) based on DO. Compared with OM, estimates of EE from SIP had lower mean absolute error of log-transformed data than SOJ for light-intensity (0.21 vs 0.27), moderate-intensity (0.33 vs 0.42), and vigorous-intensity (0.16 vs 0.35) activities. **Conclusions**: The SIP method was superior to SOJ for distinguishing between sedentary and light activities as well as estimating EE at higher intensities. Thus, SIP is recommended for research in which accuracy of measurement across the full range of activity intensities is of interest. **Key Words:** ActiGraph, activPAL, SOJOURNS, MACHINE LEARNING, SEDENTARY, MEASUREMENT

Human movement is a complex and dynamic set of behaviors, which makes it challenging to fully understand and appreciate the associated effects on health. The nature, intensity, and posture of our daily movements interact in unique ways to influence outcomes. For example, the accumulation of 10-min bouts of moderate and vigorous activity is associated with numerous physical and psychological health benefits, whereas prolonged bouts of sedentary time are increasingly associated with negative cardiometabolic and mental health consequences (8,12,37). Measurement methods that can provide accurate assessments of the full range of intensities (from sedentary to vigorous) are critical for advancing research on the effect of these behaviors on health outcomes and for assessing change over time in the context of interventions (3,29).

There are several objective physical activity monitors that are being used in the field (38). These monitors offer multiple options for processing methods that provide varying levels of precision and accuracy for both estimating energy expenditure (EE) as well as classifying activities into different intensity categories. For example, data from the ActiGraph (AG) accelerometer, one of the most commonly used monitors, have been processed in a variety of ways, including differences in intensity cut points, wear time algorithms, and bout durations, as well as using data from the vertical axis only versus data from all three axes (15,23,27,31,34). These processing differences may result in disparate conclusions regarding the activity levels of participants and the associated health consequences. For example, methods that mistakenly classify standing still (light-intensity activity) as sedentary, as

Address for correspondence: Laura Ellingson, Ph.D., Department of Kinesiology, Iowa State University, 239 Forker, Ames, IA 50011; E-mail: ellingl@iastate.edu.
Submitted for publication November 2015.
Accepted for publication February 2016.
Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.acsm-msse.org).

0195-9131/16/4808-1629/0
MEDICINE & SCIENCE IN SPORTS & EXERCISE®

DOI: 10.1249/MSS.0000000000000915

seen in studies using the AG (26,27), are likely to inaccurately estimate the risks associated with sedentary time. This variation presents challenges for researchers with respect to determining which monitor(s) and associated processing method to use in their work.

Other monitoring technologies have shown promise for capturing some aspects of physical activity but limitations for others. Results from our recent study demonstrated notable differences in accuracy of several monitors across a range of free-living activities. The SenseWear Core Armband was found to be the most accurate for assessing moderate- and vigorous-intensity activity, whereas the activPAL (AP) was more accurate for light-intensity activity and sedentary behavior (25). The advantages of the AP for these lower-intensity components can be attributed to its positioning on the thigh, giving it a unique capability to distinguish the postural component of behavior (e.g., sitting vs standing) (26,28). Our past data showed limited utility for standard AG estimates of EE using cut points, as it was demonstrated to have the highest error rate of all monitors tested (25). However, new processing algorithms based on a machine learning neural network approach have shown promising results in the initial validation studies (24,27). Although there is no definitive choice for all applications, it is clear that pattern recognition technologies and methods such as these offer great promise for advancing research on physical activity assessment.

Machine learning techniques offer a distinct advantage over previous processing methods in that they permit the fitting of models that are too complex and general to be hypothesized or interpreted at a reasonable computational cost. These methods are particularly well suited to situations in which one would like to estimate a quantity of interest (e.g., EE) from a measured quantity of no intrinsic interest (e.g., count data from an accelerometer) and when relationships between predictors and outcomes are not strictly linear and one is not concerned about the mechanism underlying the relationship.

Although not widely used in physical activity research at present, these methods have demonstrated value for improving the estimation of activity intensity over the more common cut point methods that rely primarily on less flexible models such as linear regression (12,27,30,36). The sojourn (SOJ) method developed by Lyden and colleagues (24,27) and validated under free-living conditions is a particularly promising platform because it extracts segments of activity and inactivity based on their own natural time course and duration rather than by user-specified time sampling intervals (e.g., 15-s epochs). Briefly, this method inputs triaxial count data from the AG in 1-s epochs, partitions the data into bouts of at least 30 s, and applies to each bout a model fit with a hybrid machine learning approach including artificial neural networks in combination with decision tree analysis. However, because of its use of a hip-worn monitor that cannot detect posture, the SOJ method is unable to reliably differentiate between sedentary and light-intensity activities (27).

The present study advances research with SOJ (and machine learning methods in general) in two important ways. First, this study provides an independent evaluation of the accuracy of Lyden's SOJ three-axis method (27), which analyzed data collected using a different methodology. Second, this study determined the added utility of integrating data from the AP into the processing stream—henceforth referred to as the sojourns including posture (SIP) method. It is important to note that our use of the prefit model from Lyden et al. (27) in the SIP method eliminates the risk of overfitting, which would otherwise be a concern with such a flexible model. Consistent with best practices (5) and our past validation study (25), the accuracy of these methods is compared with one another as well as with measured METs and direct observation (DO). We hypothesize that the SOJ processing method would perform well but that the SIP method would be more accurate than SOJ both in terms of estimating EE (e.g., METs) as well as categorizing activity intensity (e.g., sedentary, light, moderate, and vigorous). This is a secondary data analysis using the same raw data set used in our recently published manuscript (25).

METHODS

Participants. Participants were 49 adults (63% female; average age = 23.9 ± 5.3 yr) with no mobility limitations and an average body mass index of 23.4 ± 3.5 kg·m^{-2}. Participants were recruited from the Iowa State University community via flyers and word of mouth. Exclusionary criteria were metal allergies or mobility limitations that precluded participation in any of the prescribed activities during testing.

Procedures. All procedures were approved by the institutional review board, and all participants read and signed written informed consent documents. Data collection procedures are thoroughly detailed in our previous publication (25). Briefly, each participant performed a series of 15 activities in a controlled laboratory setting. Activities were selected to represent common behaviors within each of the four basic intensity categories (sedentary, light, moderate, and vigorous) and included supine resting, sitting reading a book, sitting typing, sitting fidgeting, standing reading a book, standing typing, standing fidgeting, climbing stairs, throwing/catching a ball, stationary biking, walking at 2 and 3 mph, walking at 3 mph typing, and running at 4.5 and 5.5 mph. For all participants, activities were performed in the order listed previously in recognition that if activities were presented in a random order, EE for lower-intensity activities could potentially be overestimated when they were performed immediately after higher-intensity activities. Therefore, to avoid this type of systematic measurement error, we used this incremental approach rather than the random approach. Each activity was performed for 5 min, with 1-min resting intervals between different activities. While completing the activities, the participants concurrently wore the AG and AP and a portable indirect calorimetry system (Oxycon mobile [OM]). Specific details of the two monitors and indirect calorimetry are provided in the following sections.

Activity monitors. The ActiGraph GT3X+ (AG; ActiGraph LLC, Pensacola, FL) is a small (4.6 × 3.3 × 1.5 cm), light (19 g) triaxial accelerometer that records acceleration ranging between ±6*g*. It samples acceleration at a rate of 30–100 Hz, which is then digitized through a 12-bit analog-to-digital converter. The digital values are filtered via a band-pass filter at a range between 0.24 and 2.5 Hz. In the present study, the GT3X+ was placed on the right hip at the level of iliac crest. The GT3X+ was initialized at 100 Hz, and data were preprocessed into 1-s epochs using the ActiLife software (version 6.5.1) in preparation for use in the SOJ and SIP methods described in the next paragraph. Data from the AG is capable of outputting raw acceleration in three axes (vertical, anteroposterior, and lateral) as well as steps, and counts (a unitless metric of movement). For the purposes of this study, counts per second from each of the three axes were used further processing with SOJ and SIP. For the data presented in this article, the low-frequency extension (LFE) was used for the initial processing with the ActiLife software, as recommended by Cain and colleagues (6). However, examination of participants processed with and without this extension demonstrated that it had minimal influence on counts per second values and negligible differences in both MET estimates and intensity categorization from both the SOJ and the SIP methods. Although there is clear evidence that the use of the LFE makes a substantive difference for outcomes related to steps (39), it appears the neural network used to estimate METs in both SIP and SOJ are not significantly influenced by the LFE option. Thus, data processed with or without this filter will likely result in the detection of similar patterns of physical activity and sedentary behaviors.

The activPAL3^{TM} (AP; PAL Technologies Ltd, Glasgow, Scotland, UK) is a small (35 × 35 × 7 mm), light (15 g), capacitive, triaxial accelerometer that collects data in the range of ±2*g*. The AP digitizes acceleration data (sampled at 20 Hz) through an 8-bit analog-to-digital converter. In the current study, the AP was placed on the midpoint of the anterior surface of the right thigh. The proprietary AP software is capable of outputting the following parameters: time spent in sitting, standing, and stepping; steps; step cadence; and activity MET scores per hour. For the purposes of this study, the activity classification component (sitting, standing, and stepping) and the boundaries between different activities (e.g., change from standing to stepping) were used from the AP software's *Events.csv* output file.

AG and AP monitors were initialized and downloaded on the same computer such that their internal time stamps would match to facilitate postprocessing. Appropriate placement for both monitors was demonstrated, and each participant put the monitors on in front of one of the study personnel to ensure accuracy. Data from both monitors were recorded continuously during the 15 activities performed in the laboratory.

Criterion measures. Indirect calorimetry was used as the criterion for estimates of EE. The OM (CareFusion Corp, San Diego, CA) is a portable indirect calorimetry system that measures breath-by-breath respiratory gas exchange under laboratory settings as well as free-living environments. Previous research has demonstrated the validity and reliability of the OM (18,33), and it is commonly used as a criterion method to provide accurate values of EE. For the present study, gas exchange measurements were collected using Hans Rudolph Masks (Hans Rudolf, Inc., Kansas City, MO). Before each test, gas and volume calibrations were performed according to manufacturer's recommendations. The flowmeter was calibrated automatically. Acceptable percent error ranges for gas calibration and volume calibration were 3% and 2%, respectively. Oxygen consumption data ($mL \cdot kg^{-1} \cdot min^{-1}$) from the OM were downloaded for each minute of testing and averaged for each of the 15 activities.

Direct observation was used as the criterion for intensity classification agreement. Participants were observed by two research assistants during all laboratory procedures (e.g., performance of the 15 activities). Notes were taken throughout each session to document that each activity was performed as prescribed.

Data processing. Data collected from the three described methods (i.e., AG, AP, and OM) were temporally matched (by second) to facilitate analyses. Data from the AG were processed alone using R-software to run the SOJ three-axis method as detailed in the validation article by Lyden and colleagues (27). Briefly, SOJ uses the counts per second data from the vertical, anteroposterior, and medial–lateral acceleration signals of a hip-mounted AG (output of AG's ActiLife software) to identify candidate bout intervals by identifying periods of rapid acceleration and deceleration indicative of a change in activity, then iteratively merges bouts until all bouts are at least 30 s in duration. Next, a decision tree is used to classify the bouts as either activity or one of four types of inactivity. MET values between 1 and 1.7 are assigned to the inactive bouts based on classification only: sitting or lying still—1 MET, sitting with minimal movement—1.2 METs, standing still—1.5 METS, and standing with minimal movement—1.7 METS. A neural net is then used to estimate the MET values for the active bouts using counts per second from the *x*, *y*, and *z*-axes.

The SIP method modifies SOJ in two ways. First, using output of AP's proprietary software (*Events.csv* file), the posture transitions (sitting, standing, and stepping) identified by the thigh-mounted monitor are added to the list of candidate bout boundaries to be considered by SOJ, thus refining the MET estimation by providing the opportunity for decreasing the time over which EE is estimated. Crucially, this appears to correct a problem where SOJ can fail to find a boundary between two long bouts of different activities, forcing the neural net to estimate across a mixture of two distinct activities. Second, inactive bouts identified by SOJ are reclassified as sitting or standing based on the prevailing posture reported by the AP. This latter change alters the estimated EE for such bouts by ±0.5 MET so as to be either greater than or less than 1.5 depending on whether the AP indicates sitting (<1.5) or standing (≥1.5).

The SIP and the SOJ methods estimated METs throughout the 90-min protocol as one continuous event, and there was no information provided in the processing to indicate when activities started and stopped. The first and the last minutes of each activity were then excluded to remove noise captured during the transition periods. To get an estimate of the MET value associated with each activity, the second-by-second MET estimates for minutes 2–4 of the activity were then averaged. Similarly, for the OM, the minute-by-minute values ($mL \cdot kg^{-1} \cdot min^{-1}$) were averaged for minutes 2–4 of each activity.

As in our previous validation paper (25), to facilitate direct comparisons among the different measurement methods (resulting in different outcome units), we used MET_{RMR}, which is a MET score that takes into consideration one's resting metabolic rate (RMR). Participants' RMR values were estimated using the prediction equation developed by Schofield (35). MET_{RMR} values for the OM were obtained by dividing measured $\dot{V}O_2$ values ($mL \cdot kg^{-1} \cdot min^{-1}$) by estimated RMR values. For the SOJ and SIP methods, MET_{RMR} was calculated by multiplying average MET values for each 3-min period by 3.5 and then by dividing by the estimated RMR values. For DO, each activity was categorized by intensity (sedentary, light, moderate, or vigorous) using the estimates of EE found in the compendium of physical activities (1) and standard cut points (sedentary, <1.5; light, 1.5–2.99; moderate, 3.0–5.99; vigorous, ≥6).

Statistical analyses. Classification accuracy was assessed using confusion matrices (tables describing the performance of each method with respect to a criterion), and EE estimates for each method were compared via the log error relative to the criterion measure (ln Q). Rather than use root mean squares ln Q as recommended by Tofallis (37), we compute the mean absolute error of log-transformed data (MAE ln Q) because this value is more comparable with the mean absolute percent error (MAPE) for small errors (up to a factor of 100%; i.e., 1% MAPE corresponds to 0.01 MAE ln Q), while still avoiding the bias in the presence of larger errors associated with relying solely on MAPE, as discussed by Tofallis (37). We also calculated MAPE in the conventional way to make more direct comparisons with our previously published manuscript examining the accuracy of EE estimates from other widely used accelerometers and processing methods. Lastly, Bland–Altman plots were created for each intensity category (sedentary, light, moderate, and vigorous) to illustrate the agreement between methods and to identify any systematic biases.

RESULTS

Table 1 lists the 15 activities and their associated estimates for EE as well as activity intensity classification from each processing method (OM, SIP, and SOJ) and from the compendium. Table 2 and Figure 1 highlight differences in classification accuracy by intensity and by the specific activity, respectively. The SOJ method had an overall accuracy of 56.0% (95% CI = 52.5%–59.8%), whereas the SIP method had an overall accuracy of 78.7% (95% CI = 75.5%–81.6%). With respect to sensitivity, SIP was superior to SOJ across all intensities; SOJ method was particularly poor for light-intensity activity at 17.5% compared with SIP at 73.4%. SIP also had higher specificity for moderate- and vigorous-intensity activities.

Examinations of EE also suggest that the SIP method has some advantages over SOJ. As seen in Table 3, estimates of EE from SIP included in smaller MAE ln Q for light-, moderate-, and vigorous-intensity activities and comparable accuracy for sedentary activities. Similarly, MAPE was also lower in SIP as compared with SOJ, particularly for moderate- and vigorous-intensity activities. This is further demonstrated by the Bland–Altman plots in Figure 2, comparing MET estimates from SOJ and SIP to estimates from the OM. Overall, the Bland–Altman plots demonstrated that agreement between both SIP and SOJ methods with the OM was relatively good. However, the limits of agreement for SIP were narrower than for SOJ for moderate- and vigorous-intensity activities. For sedentary activities, both SIP and

TABLE 1. Estimates of EE (MET_{RMR}) and intensity category classification for each activity performed during laboratory testing.

Activity	OM		SOJ		SIP		Compendium Classification
	MET_{RMR} Estimate	Category	MET_{RMR} Estimate	Category	MET_{RMR} Estimate	Category	
Supine, resting	1.37 (0.19)	S	1.20 (0.18)	S	1.16 (0.14)	S	S
Sitting, reading	1.26 (0.21)	S	1.16 (0.14)	S	1.17 (0.16)	S	S
Sitting, typing	1.42 (0.23)	S	1.13 (0.15)	S	1.13 (0.18)	S	S
Sitting, fidgeting	1.46 (0.23)	S	1.13 (0.13)	S	1.21 (0.22)	S	S
Standing, reading	1.28 (0.22)	S	1.15 (0.18)	S	1.62 (0.11)	L	L
Standing, typing	1.39 (0.26)	S	1.14 (0.15)	S	1.61 (0.10)	L	L
Standing, fidgeting	1.72 (0.49)	L	1.34 (0.62)	S	1.76 (0.54)	L	L
Climbing stairs	4.71 (1.09)	M	4.94 (1.78)	M	5.04 (1.65)	M	M
Throwing/catching a ball	3.12 (0.93)	M	2.52 (1.56)	L	2.60 (1.64)	L	L
Stationary biking	4.91 (1.45)	M	3.43 (2.09)	M	3.48 (2.10)	M	M
Treadmill walking, 2 mph	3.74 (0.55)	M	5.16 (1.15)	M	5.37 (0.95)	M	L
Treadmill walking, 3 mph	4.71 (0.57)	M	5.70 (2.31)	M	5.76 (1.49)	M	M
Treadmill walking while typing, 3 mph	4.69 (0.89)	M	7.49 (3.71)	V	5.51 (1.26)	M	M
Treadmill running, 4.5 mph	8.70 (1.00)	V	9.02 (3.89)	V	9.79 (2.32)	V	V
Treadmill running, 5.5 mph	10.22 (0.99)	V	9.57 (3.81)	V	10.34 (1.58)	V	V

S, sedentary; L, light; M, moderate; V, vigorous.

TABLE 2. Confusion matrices and agreement statistics by intensity classification comparing SOJ (A) and SIP (B) to intensity categories using DO and the compendium of physical activities as the criterion measure.

(A)

	Reference			
Prediction	Sedentary	Light	Moderate	Vigorous
Sedentary	190	142	5	0
Light	6	43	31	4
Moderate	0	50	101	15
Vigorous	0	10	57	76
	Sedentary	Light	Moderate	Vigorous
Sensitivity	96.9%	17.6%	52.1%	80.0%
Specificity	72.5%	91.5%	87.8%	89.5%
Positive prediction value	56.4%	51.2%	60.8%	53.2%
Negative prediction value	98.5%	68.7%	83.5%	96.7%
Balanced accuracy	84.7%	54.5%	69.9%	84.7%

(B)

	Reference			
Prediction	Sedentary	Light	Moderate	Vigorous
Sedentary	182	3	1	0
Light	14	180	28	0
Moderate	0	53	119	1
Vigorous	0	9	47	94
	Sedentary	Light	Moderate	Vigorous
Sensitivity	92.8%	73.5%	61.0%	98.9%
Specificity	99.3%	91.4%	89.9%	91.2%
Positive prediction value	97.9%	81.1%	68.8%	62.7%
Negative prediction value	97.4%	87.2%	86.4%	99.8%
Balanced accuracy	96.1%	82.4%	75.5%	95.1%

SOJ slightly underestimated METs compared with OM (bias of 0.2). For light, moderate, and vigorous intensities, overall estimates were closer with differences centering around zero. However, for light and vigorous intensities, there was evidence of bias. For light intensity, both SOJ and SIP tended to underestimate METs when the OM estimates were low and overestimate when OM estimates were high. For vigorous intensity, the situation was reversed in that both SIP and SOJ tended to overestimate METs when the OM METs were low and underestimate when OM METs were high.

DISCUSSION

This study examined the validity of two machine learning methods for processing physical activity data in relation to

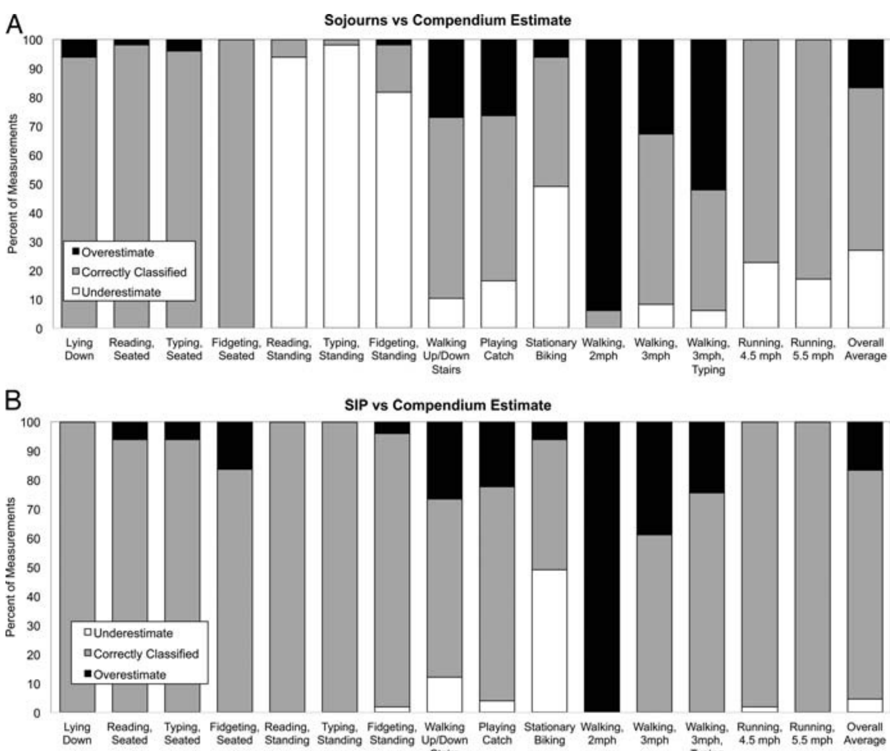

FIGURE 1—Classification accuracy for the SOJ (A) and SIP (B) methods for each of the 15 activities, using the compendium as the criterion.

TABLE 3. Indicators of agreement for EE (MET_{RMR} estimate) by activity intensity for SOJ and SIP methods in relation to the OM.

			MET_{RMR} Estimate, Mean (95% CI)	MAE ln Q, Mean (SD)	MAPE, Mean (SD)
Sedentary		OM	1.36 (1.02–1.69)	NA	NA
		SOJ	1.15 (0.94–1.36)	0.21 (0.05)	18.01 (10.48)
		SIP	1.17 (0.93–1.41)	0.21 (0.05)	17.56 (10.16)
Light		OM	1.46 (0.89–2.03)	NA	NA
		SOJ	1.21 (0.69–1.72)	0.27 (0.06)	20.88 (11.54)
		SIP	1.66 (1.24–2.08)	0.21 (0.05)	20.49 (18.14)
Moderate	Total*	OM	4.31 (3.16–5.46)	NA	NA
		SOJ	4.89 (2.72–7.05)	0.42 (0.13)	24.17 (23.91)
		SIP	4.63 (2.69–6.56)	0.33 (0.15)	17.19 (17.73)
	Ambulatory only	OM	4.46 (3.33–5.59)	NA	NA
		SOJ	5.83 (3.18–8.47)	0.38 (0.11)	37.22 (31.18)
		SIP	5.42 (3.53–7.31)	0.26 (0.07)	24.74 (20.18)
Vigorous		OM	9.46 (7.59–11.31)	NA	NA
		SOJ	9.31 (3.33–15.28)	0.35 (0.02)	24.62 (19.98)
		SIP	10.08 (6.61–13.55)	0.16 (0.04)	15.74 (14.17)

*Moderate total includes stationary biking and throwing and catching a ball in additional to walking and climbing stairs.

both DO and indirect calorimetry in a sample of healthy adults. Overall, both SIP and SOJ methods were relatively good at classifying activity intensity and also estimating EE, providing lower individual error estimates in comparison with other commonly used research-grade monitors and processing methods. Our previous validation study comparing the SenseWear Core Armband, the AP, and the AG processed with cut point method found that the Core Armband was the most accurate, with an overall MAPE of 31.4% (25). On the basis of the same data set, the SOJ and the SIP methods demonstrated overall MAPE values of 21.9% and 17.7%, respectively, suggesting that both methods are improvements over standard approaches. However, the SIP method was found to be superior to the SOJ method both for intensity classification at lower intensities (sedentary vs light) as well as for estimating EE at higher intensities (moderate and vigorous).

The SOJ method (based on the AG) improves upon previous processing methods, using data from all three axes to provide more accurate estimates of activity intensity than earlier cut point methods under free-living conditions (27). Instead of averaging intensity estimates across predetermined time intervals (e.g., 15-s or 1-min epochs), the SOJ method identifies individualized data-driven epochs by examining the data for bout boundaries (periods of rapid acceleration/deceleration). A neural network is then used to estimate EE averaged over these bouts. Thus, estimates are more accurate and precise than previous cut point methods because they are averaging across period of similar activity (24,27). However, because of its use of a single, hip-worn accelerometer, it is less capable of differentiating between sedentary-intensity (i.e., sitting) and light-intensity (i.e., standing still) activities. The AP, processed using its proprietary software, has been validated for accurately assessing posture (e.g., differentiating between sitting and standing) (26,28), and our group previously demonstrated its utility for accurately estimating EE for sedentary and light-intensity activities (25). However, it tends to provide an inaccurate estimate of EE for higher intensities using currently available algorithms (25).

Our SIP method was designed to combine the positive aspects of these two commonly used monitors and their respective validated processing methods to more accurately assess activity across intensities. The preliminary data presented here are encouraging and suggest that SIP can accurately assess a range of human movement from sitting or lying down to being vigorously active. It is notable that the SIP method, which is designed to better distinguish sitting from standing through the inclusion of postural data from the AP, improved on the SOJ method even for vigorous upright activities. This can only have been caused by the insertion of candidate bout boundaries, as there was no standing or sedentary time to reclassify, and indeed we observed several instances in our data in which the SOJ method failed to separate a rest period from an adjacent active bout, causing SOJ to underestimate the intensity of the active bout (see Figure, Supplemental Digital Content 1, Graphical representation of data from an individual participant in this study doing all 15 activities with 1-min transitions between activities, http://links.lww.com/MSS/A667). This suggests that improvements to the SOJ method's estimation of bout boundaries might be able to provide the benefits of the SIP method for higher intensities without the addition of the AP monitor. However, this possibility remains to be examined.

The inherent challenges in processing and interpreting accelerometry-based physical activity data have been well chronicled (3,20). The need to simultaneously examine physical activity and sedentary behaviors further compounds these challenges. Accurately differentiating between sedentary and light activity has proven particularly difficult (17,19,22). This is one area where the SIP methodology has a distinct advantage over previous methods because it integrates postural and acceleration data to more accurately differentiate light-intensity and sedentary activities.

To date, public health recommendations have focused primarily on promoting physical activity, and recommendations have been established to this effect for both children and adults (9,40). However, for most individuals, moderate- and vigorous-intensity physical activity makes up a small proportion of waking hours (<5%), with the vast majority of time being divided between sedentary and light-intensity behaviors. The recognition of this, in conjunction with

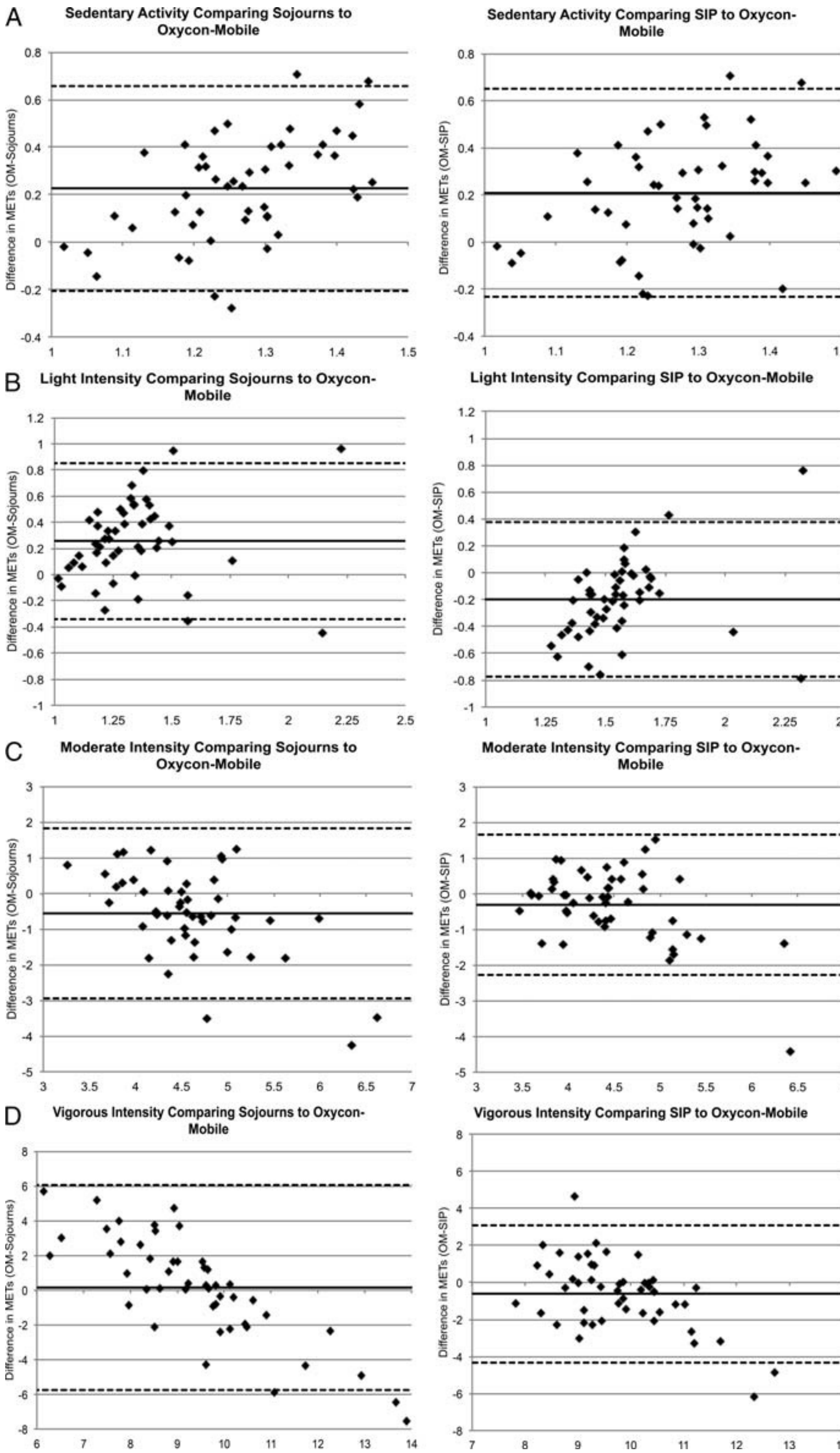

FIGURE 2—Bland–Altman plots comparing estimated EE for the SOJ and SIP methods using the OM as the criterion for sedentary (A), light-intensity (B), moderate-intensity (C), and vigorous-intensity (D) activities.

research demonstrating the negative effects of excessive sedentary time on health, has led to an interest in working toward developing recommendations for sedentary time as well as joint recommendations, including both physical activity and sedentary behaviors (10,11,14,32). Several countries have responded to this by recommending that their citizens try to break up prolonged periods of sedentary time when possible (2,7). The ability to accurately assess individual patterns of behavior would advance research on how to accumulate physical activity and sedentary time in a way that minimizes negative health consequences. Further, in the context of intervention trials, improvements in measurement accuracy would enable the evaluation of how lifestyle change influences allocation of time spent in different intensities (sedentary, light, moderate, and vigorous). For example, when physical activity increases, what happens to sedentary time? Or conversely, when sedentary time is reduced, what is it replaced by and do health outcomes vary based on this? Because of its accuracy across the range of activity intensities, the SIP method offers distinct advantages over other contemporary assessment methods.

It is important to note, however, that the incorporation of multiple sensors, as done in the SIP method, is not a novel concept. For example, Bassett and colleagues (4) used two AP monitors, one placed on the thigh as it typically done and the other placed on the ribcage, and were able to accurately classify posture (lying down, sitting, standing, and stepping) with 98%–100% accuracy in a laboratory setting. Ellis and colleagues (12) also used two monitoring devices incorporating a hip-worn AG and a GPS unit to classify a subset of free-living activities into one of four categories: sedentary, standing, ambulatory, and in a vehicle, using images from a small lanyard-mounted video camera worn around the neck as the criterion. Overall, the classification accuracy of these activities exceeded 85%. However, accurate classification for standing still and walking was substantially lower at 67% and 63%, respectively, indicating that differentiation between activities at the lower end of the intensity continuum remained problematic. Without discounting these impressive levels of accuracy, it is worth noting that neither of these protocols attempted to estimate EE, as was done in the present study with the SIP and SOJ methods, nor did these studies demonstrate an ability to correctly classify a range of activities from sedentary to vigorous intensities.

Several groups have also used multisensor methods with the goal of improving activity classification as well as being able to identify movement patterns associated with specific activities (e.g., cycling) (13,21). These systems typically provide a high level of accuracy that is superior to single-sensor systems and may develop into promising solutions for addressing measurement issues. However, at present, the systems themselves are often poorly suited for typical research studies in which activities will be measured over multiple days under free-living conditions and require higher-level processing skills for working with the data.

The primary limitation of the SIP method is the need for two activity monitors, the AG and the AP. This necessitates a larger cost but, more importantly, also may lead to issues with participant compliance because of the increased burden of wearing two devices. However, this method has been used in several studies (publications in process) across different populations, including college students, adults, older adults, and several patient groups (e.g., Parkinson's disease and fibromyalgia) with few difficulties regarding wear time compliance. Further, there is evidence of larger-scale trials successfully using two monitors for activity assessment (SenseWear Armband and AP), including multiple measures over several years demonstrating the feasibility and potential utility of two-monitor methods (16).

In addition to the limitations of the method itself, there are also several limitations inherent in our study. Our study sample was homogenous, including primarily young adults with normal BMI. As such, the SIP method should undergo additional testing in both younger and older individuals as well as those in the overweight and obese categories to determine its validity in these populations. RMR was not objectively measured; rather, it was estimated using the Schofield equation (35), a relatively common practice in validation studies (8,25,41). It was also notable that although in most cases the OM and the compendium were in agreement, there were several activities for which this was not the case. For example, walking at 2 mph was classified as light intensity when using the compendium and as moderate-intensity according to estimates from the OM. In those cases, we felt that it was better to keep the criterion measure consistent within each set of analyses (OM for MET estimation vs compendium for categorization accuracy). This choice was made because the OM-measured METs varied among participants such that for some individuals, an activity that is assumed to be a particular intensity (e.g., standing reading—light intensity) was not always within that range for all participants. As such, it would have been confusing to have a single activity classified as one intensity for some individuals and another for others. Lastly, this study was laboratory based and did not include a free-living component. However, SOJ has been validated using DO under free-living conditions (24,27) and SIP builds directly upon this platform. In addition, the 15 activities included in the protocol ranged from sedentary to vigorous to capture a range of potential activities individuals may engage in during normal daily life. However, additional testing of this method under free-living conditions including household activities will be an important next step.

In summary, our results demonstrate that both the SOJ and the SIP methods provide more accurate estimates of EE in comparison with commonly used cut point methods and were able to classify activities by intensity across a range of activities with a high degree of accuracy. The SIP method was either superior or equivalent to SOJ, demonstrating the utility of including two monitors that are designed to focus on assessing different aspects of physical activity behaviors across the range of intensities. Importantly, SIP was markedly better at discriminating between sedentary and light-intensity behaviors, which may only be differentiated by posture. Thus, SIP is recommended for research in which accurate differentiation between sedentary and light activities is necessary (e.g., sedentary interventions) and/or where accuracy of measurement across the full range of activity intensities is of primary interest.

The code for the SIP method is freely available online at https://www.github.com/ischwabacher/SIP. In addition to the SIP method itself, which is implemented in R, there are two wrapper scripts written in Python to display data in a user-friendly format and to generate a wide array of summary statistics. Installation instructions and documentation are also available.

The authors thank all the participants for their participation in the current study. They also thank Rachel Devick and Paul Hibbing for their assistance in data collection. Authors Y. Kim and G. J. Welk declare that they have had ongoing research support from Bodymedia/Jawbone to assist in cross validation of algorithms and in testing of new prototype monitors. The data reported here were collected as part of a larger study designed to test new monitoring technologies, but the new devices were not reported in the present study. The present study was conducted independently from Bodymedia/Jawbone, but the authors acknowledge that funding from Bodymedia/Jawbone supported the data collection. The SIP software detailed in this article was developed by Isaac Schwabacher in the laboratory of Dr. Cook at the University of Wisconsin-Madison. The contents of this study do not represent the views of the U.S. Department of Veterans Affairs or the United States Government. The findings of the current study do not constitute endorsement by the American College of Sports Medicine.

REFERENCES

1. Ainsworth BE, Haskell WL, Herrmann SD, et al. 2011 Compendium of physical activities: a second update of codes and MET values. *Med Sci Sports Exerc*. 2011;43(8):1575–81.
2. Australian Government Department of Health. Australia's Physical Activity and Sedentary Behaviour Guidelines [Internet]. [date unknown]; [cited 2015 Dec 17] Available from: http://www.health.gov.au/internet/main/publishing.nsf/content/health-pubhlth-strateg-phys-act-guidelines#apaadult.
3. Bassett DR. Device-based monitoring in physical activity and public health research. *Physiol Meas*. 2012;33(11):1769–83.
4. Bassett DR, John D, Conger SA, Rider BC, Passmore RM, Clark JM. Detection of lying down, sitting, standing, and stepping using two activPAL monitors. *Med Sci Sports Exerc*. 2014;46(10):2025–9.
5. Bassett DR Jr, Rowlands A, Trost SG. Calibration and validation of wearable monitors. *Med Sci Sports Exerc*. 2012;44(1 Suppl 1):S32–8.
6. Cain KL, Conway TL, Adams MA, Husak LE, Sallis JF. Comparison of older and newer generations of ActiGraph accelerometers with the normal filter and the low frequency extension. *Int J Behav Nutr Phys Act*. 2013;10:51.
7. Canadian Society for Exercise Physiology. Canadian Physical Activity Guidelines and Canadian Sedentary Behaviour Guidelines [Internet]. [date unknown]; [cited 2015 Dec 17] Available from: http://www.csep.ca/en/guidelines/get-the-guidelines.
8. Carter J, Wilkinson D, Blacker S, et al. An investigation of a novel three-dimensional activity monitor to predict free-living energy expenditure. *J Sports Sci*. 2008;26(6):553–61.
9. Centers for Disease Control and Prevention. Children | Physical Activity | DNPAO | CDC [Internet]. [date unknown]; [cited 2015 Sep 18] Available from: http://www.cdc.gov/physicalactivity/basics/children/.
10. Chaput JP, Carson V, Gray CE, Tremblay MS. Importance of all movement behaviors in a 24 hour period for overall health. *Int J Environ Res Public Health*. 2014;11(12):12575–81.
11. Dunstan DW, Howard B, Healy GN, Owen N. Too much sitting—a health hazard. *Diabetes Res Clin Pract*. 2012;97(3):368–76.
12. Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, Marshall S. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiol Meas*. 2014;35(11):2191–203.
13. Gao L, Bourke AK, Nelson J. Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. *Med Eng Phys*. 2014;36(6):779–85.
14. Hamilton MT, Healy GN, Dunstan DW, Zderic TW, Owen N. Too little exercise and too much sitting: inactivity physiology and the need for new recommendations on sedentary behavior. *Curr Cardiovasc Risk Rep*. 2008;2(4):292–8.
15. Ham SA, Reis JP, Strath SJ, Dubose KD, Ainsworth BE. Discrepancies between methods of identifying objectively determined physical activity. *Med Sci Sports Exerc*. 2007;39(1):52–8.
16. Hand GA, Shook RP, Paluch AE, et al. The energy balance study: the design and baseline results for a longitudinal study of energy balance. *Res Q Exerc Sport*. 2013;84(3):275–86.
17. Healy GN, Clark BK, Winkler EAH, Gardiner PA, Brown WJ, Matthews CE. Measurement of adults' sedentary time in population-based studies. *Am J Prev Med*. 2011;41(2):216–27.
18. Hodges LD, Brodie DA, Bromley PD. Validity and reliability of selected commercially available metabolic analyzer systems. *Scand J Med Sci Sports*. 2005;15(5):271–9.
19. Janssen X, Basterfield L, Parkinson KN, et al. Objective measurement of sedentary behavior: impact of non-wear time rules on changes in sedentary time. *BMC Public Health*. 2015;15(1):504.
20. Janssen X, Cliff DP. Issues related to measuring and interpreting objectively measured sedentary behavior data. *Meas Phys Educ Exerc Sci*. 2015;19(3):116–24.
21. Jiang Y, Larson JL. IDEEA activity monitor: validity of activity recognition for lying, reclining, sitting and standing. *Front Med*. 2012;7(1):126–31.
22. Kang M, Rowe DA. Issues and challenges in sedentary behavior measurement. *Meas Phys Educ Exerc Sci*. 2015;19(3):105–15.
23. Keadle SK, Shiroma EJ, Freedson PS, Lee I-M. Impact of accelerometer data processing decisions on the sample size, wear time and physical activity level of a large cohort study. *BMC Public Health*. 2014;14:1210.
24. Kim Y, Barry VW, Kang M. Validation of the ActiGraph GT3X and activPAL accelerometers for the assessment of sedentary behavior. *Meas Phys Educ Exerc Sci*. 2015;19(3):125–37.
25. Kim Y, Welk GJ. Criterion validity of competing accelerometry-based activity monitoring devices. *Med Sci Sports Exerc*. 2015;47(11):2456–63.
26. 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(8):1561–7.
27. Lyden K, Keadle SK, Staudenmayer J, Freedson PS. A method to estimate free-living active and sedentary behavior from an accelerometer. *Med Sci Sports Exerc*. 2014;46(2):386–97.
28. Lyden K, Kozey Keadle SL, Staudenmayer JW, Freedson PS. Validity of two wearable monitors to estimate breaks from sedentary time. *Med Sci Sports Exerc*. 2012;44(11):2243–52.
29. Matthews CE, Hagströmer M, Pober DM, Bowles HR. Best practices for using physical activity monitors in population-based research. *Med Sci Sports Exerc*. 2012;44(1 Suppl 1):S68–76.
30. Montoye AHK, Mudd LM, Biswas S, Pfeiffer KA. Energy expenditure prediction using raw accelerometer data in simulated free living. *Med Sci Sports Exerc*. 2015;47(8):1735–46.
31. Orme M, Wijndaele K, Sharp SJ, Westgate K, Ekelund U, Brage S. Combined influence of epoch length, cut-point and bout duration on accelerometry-derived physical activity. *Int J Behav Nutr Phys Act*. 2014;11(1):34.

32. Owen N, Sparling PB, Healy GN, Dunstan DW, Matthews CE. Sedentary behavior: emerging evidence for a new health risk. *Mayo Clin Proc*. 2010;85(12):1138–41.
33. Salier Eriksson J, Rosdahl H, Schantz P. Validity of the Oxycon Mobile metabolic system under field measuring conditions. *Eur J Appl Physiol*. 2012;112(1):345–55.
34. Santos-Lozano A, Santín-Medeiros F, Cardon G, et al. ActiGraph GT3X: validation and determination of physical activity intensity cut points. *Int J Sports Med*. 2013;34(11):975–82.
35. Schofield WN. Predicting basal metabolic rate, new standards and review of previous work. *Hum Nutr Clin Nutr*. 1985;39(1 Suppl): 5–41.
36. Staudenmayer J, He S, Hickey A, Sasaki J, Freedson P. Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. *J Appl Physiol (1985)*. 2015;119(4):396–403.
37. Tofallis C. A better measure of relative prediction accuracy for model selection and model estimation. *J Oper Res Soc*. 2015;66(8): 1352–62.
38. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. *Br J Sports Med*. 2014;48(13):1019–23.
39. Tudor-Locke C, Barreira TV, Schuna JM. Comparison of step outputs for waist and wrist accelerometer attachment sites. *Med Sci Sports Exerc*. 2015;47(4):839–42.
40. U.S. Department of Health and Human Services. Physical Activity Guidelines for Americans [Internet]. [date unknown]; Available from: http://www.health.gov/paguidelines/.
41. Westerterp KR. Physical activity and physical activity induced energy expenditure in humans: measurement, determinants, and effects [Internet]. *Front Physiol*. 2013;4; [cited 2015 Oct 5] Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3636460/.