



Published in final edited form as:

Med Sci Sports Exerc. 2011 May ; 43(5): 867–876. doi:10.1249/MSS.0b013e3181fc7162.

Comparative Validity of Physical Activity Measures in Older Adults

LISA H. COLBERT^{1,2}, CHARLES E. MATTHEWS⁵, THOMAS C. HAVIGHURST^{2,3},
KYUNGMIN KIM^{2,3}, and DALE A. SCHOELLER⁴

¹Department of Kinesiology, University of Wisconsin, Madison, WI

²Carbone Comprehensive Cancer Center, University of Wisconsin, Madison, WI

³Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI

⁴Department of Nutritional Sciences, University of Wisconsin, Madison, WI

⁵Nutritional Epidemiology Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD

Abstract

Purpose—To compare the validity of various physical activity measures with doubly labeled water (DLW)–measured physical activity energy expenditure (PAEE) in free-living older adults.

Methods—Fifty-six adults aged ≥ 65 yr wore three activity monitors (New Lifestyles pedometer, ActiGraph accelerometer, and a SenseWear (SW) armband) during a 10-d free-living period and completed three different surveys (Yale Physical Activity Survey (YPAS), Community Health Activities Model Program for Seniors (CHAMPS), and a modified Physical Activity Scale for the Elderly (modPASE)). Total energy expenditure was measured using DLW, resting metabolic rate was measured with indirect calorimetry, the thermic effect of food was estimated, and from these, estimates of PAEE were calculated. The degree of linear association between the various measures and PAEE was assessed, as were differences in group PAEE, when estimable by a given measure.

Results—All three monitors were significantly correlated with PAEE ($r = 0.48$ – 0.60 , $P < 0.001$). Of the questionnaires, only CHAMPS was significantly correlated with PAEE ($r = 0.28$, $P = 0.04$). Statistical comparison of the correlations suggested that the monitors were superior to YPAS and modPASE. Mean squared errors for all correlations were high, and the median PAEE from the different tools was significantly different from DLW for all but the YPAS and regression-estimated PAEE from the ActiGraph.

Conclusions—Objective devices more appropriately rank PAEE than self-reported instruments in older adults, but absolute estimates of PAEE are not accurate. Given the cost differential and ease of use, pedometers seem most useful in this population when ranking by physical activity level is adequate.

Keywords

ACCELEROMETER; PEDOMETER; QUESTIONNAIRE; DOUBLY LABELED WATER

Copyright © 2011 by the American College of Sports Medicine

Address for correspondence: Lisa H. Colbert, Ph.D., M.P.H., Department of Kinesiology, 2000 Observatory Dr., University of Wisconsin, Madison, WI 53706; lhcolbert@education.wisc.edu.

The authors report no conflicts of interest.

The results of this study do not constitute endorsement by the American College of Sports Medicine.

The valid assessment of physical activity is necessary for the advancement of knowledge regarding associations of physical activity with various health and/or disease end points. Although firm links between physical activity and specific conditions including coronary heart disease, high blood pressure, stroke, type 2 diabetes, and certain cancers have been established (35), questions remain as to the nature of the dose–response relationship for many end points. Historically, epidemiologic studies that have examined the physical activity and disease end point relationships have relied on self-report of the behavior. Given the potential for substantial overreporting of physical activity (34), it is possible that much lower volumes and/or intensities of activity are sufficient to decrease the risk of various diseases. The implication is that to detect an association with lower volume or intensities of activity, we need instruments that can appropriately measure activity at this end of the spectrum. This is particularly important for studies of older adults, who tend to be less physically active overall (21,34) and who are more likely to participate in lower-intensity activities, which can be difficult to report accurately (9).

In addition to the volume of physical activity, it is also important to know the type and intensity of the activity performed. Most objective monitors have the capability to capture the intensity of activity, allowing for the quantification of activities of moderate or greater intensity, as well as light-intensity activity and sedentary time, but the ability to capture the type of activity is limited. Self-report instruments do capture intensity and type of activity and can also capture the reason for the activity. Measurement errors, common with self-report, however, can lead to attenuation in the strength of association observed between physical activity and health outcomes of interest. That is, the regression coefficients that describe the association observed in a particular study may be smaller than the true association that would be observed if physical activity was measured with little or no error (10). This is less problematic for objective monitors when a sufficient number of days are assessed.

Validation of self-report instruments and objective monitors is difficult and thus often consists of within-individual measures of reproducibility. The difficulty is that there is no “gold standard” measure of free-living physical activity with which to compare the different measurement instruments. Doubly labeled water (DLW) is a useful method of measuring total, free-living energy expenditure (40). Combined with indirect calorimetric measures of resting metabolic rate (RMR) and an assumption about the thermic effect of food, this tool can provide reasonably accurate measures of free-living physical activity energy expenditure (PAEE) during a short period (i.e., 10–14 d).

In older adults, in particular, studies that have compared physical activity questionnaires to DLW measures of PAEE have reported correlations ranging from 0.16 for the Yale Physical Activity Survey (YPAS) to 0.8 with the Minnesota Leisure Time Physical Activity Questionnaire (3,14,19,28,32). Pedometers, accelerometers, and multi-sensor devices such as the SenseWear (SW) armband are all objective means of measuring physical activity; however, their ability to provide valid measures of activity in older adults has not been adequately addressed. Various accelerometers have been compared with DLW measures in older adults, with correlations on the order of 0.7–0.8 reported for the Caltrac, Lifecorder, and Tracmor devices (22,25). We are unaware of DLW and ActiGraph studies in older adults. The SW armband is a multisensor device that estimates time spent in physical activity as well as the energy cost of that activity. In a previous study of primarily younger adults, there was a strong correlation ($r = 0.70$) of DLW-measured and armband-estimated free-living PAEE (31). Studies of older adults in which different measurement tools have been compared with DLW are limited and have produced somewhat conflicting results (25,32).

Objective measurement may be useful to investigate dose–response relationships and/or to examine associations with lower levels of physical activity with less overall measurement error and should advance the utility of physical activity measurement in epidemiologic studies. These devices can capture activities of light intensity that are often poorly reported on physical activity questionnaires but common in older adults (22,30). Newer multisensor methods like that used in the SW armband may improve on standard accelerometers for the estimation of PAEE (17). Therefore, we sought to examine the relative validity of a variety of physical activity measurement tools, both objective and self-reported, in older adults. One of the primary limitations of a study comparing multiple assessment tools is that the statistical power to detect a difference between any two tools is limited because it involves multiple comparisons. The alternative of limiting the comparison to just two tools, however, is even less satisfying because it does not offer any data regarding the majority of assessment tools that are available to investigators. We therefore elected to preserve the power of the statistical analysis by testing the null hypothesis that all methods are equally valid against a rank ordering of the tools. Our primary aim focused on relative rather than absolute agreement between the measures of physical activity and PAEE derived from DLW, given that there are no pedometer-derived equations for predicting energy expenditure in older adults and that accelerometry equations have also not been developed specifically for this age group. Our secondary aim, however, was to compare the agreement between group estimates of PAEE when that was possible. Our *a priori* hypothesis was that the multisensor monitor would have the strongest relationship with DLW (ρ_1), that the subjective self-report measures would have the weakest relationship with DLW (ρ_4), and that the accelerometer (ρ_2) and pedometer (ρ_3) would fall between the other measures in relation to measured PAEE in older adults. The null hypothesis $H_0: \rho_1 = \rho_2 = \rho_3 = \rho_4$ was tested against the ordered alternative hypothesis that $H_1: \rho_1 > \rho_2 > \rho_3 > \rho_4$.

METHODS

Subjects and experimental design

Men and women were recruited for participation in this study from the greater Madison, WI, area through flyers posted at various community sites and word of mouth. Interested subjects called investigators and were screened about relevant inclusion and exclusion criteria. All subjects were 65 yr or older and were able to walk unassisted. Subjects were excluded if they reported any of the following: an implanted defibrillator or pacemaker, diabetes, an unstable thyroid condition, or the use of β -blockers, weight loss supplements, or oral steroids. The study protocol was approved by the University of Wisconsin Health Sciences Institutional Review Board, and all subjects provided written informed consent before study initiation.

Eligible and interested participants were scheduled for a study run-in period that was ~4 d long. Subjects were fitted with the study monitors, told to wear the monitors during waking hours (except for water-based activities), and given a log sheet on which to record the times they wore the monitors. At the close of the run-in period, logs and monitor data were examined qualitatively for compliance with the instructions, and subjects whose logs and monitor data were in close agreement in regard to time worn, who wore the monitors during waking hours, and who were willing to wear the monitors for 10 d were scheduled for their two study visits. At study visit 1, the DLW protocol was started after a minimum 8-h fast, and participants filled out three physical activity questionnaires and were fitted with the physical activity monitors and given logs. At study visit 2, 10 d later, subjects provided the final urine samples, returned the monitors and logs, completed the same three physical activity questionnaires and a basic demographic and health history questionnaire, and had their body fat and RMR measured. Subjects came to study visit two having fasted for at least 8 h and having refrained from vigorous exercise and alcohol consumption for at least 24 h.

A total of 70 men and women initially consented. None of the subjects were found to be noncompliant during the run-in period. Ten subjects dropped out of the study for the following reasons: lost interest/did not have time, $n = 4$; could not provide necessary urine samples, $n = 2$; developed health conditions between the run-in and study visits, $n = 2$; found to be ineligible, $n = 1$; reason unknown, $n = 1$; for a total sample size of 60. Of those 60 subjects, 4 were not included in this analysis for the following reasons: DLW samples that did not equilibrate ($n = 2$), a nonphysiologic value for the energy expenditure measure ($n = 1$), and an armband that did not collect data ($n = 1$), leaving 56 for analysis.

PAEE by doubly labeled water and indirect calorimetry

Total energy expenditure (TEE) was determined using the DLW method (16). Subjects provided a baseline urine sample and were given an oral dose containing an estimated $0.18 \text{ g}\cdot\text{kg}^{-1}$ total body water (TBW) of ^{18}O -labeled water and $0.16 \text{ g}\cdot\text{kg}^{-1}$ TBW of ^2H -labeled water. Urine samples were collected at 2, 3, and 4 h after dosing. Subjects were allowed a maximum of 500 mL of liquid at the 1-h post dose time point. At the second study visit, subjects provided two urine samples, 1 h apart. Urine samples were stored at -20°C until the stable isotope abundances of the physiologic samples were measured by isotope ratio mass spectrometry. Isotope dilution spaces were calculated by the plateau method according to Cole and Coward (5). Carbon dioxide production was calculated using equation A6 of Schoeller et al. (27) as modified by Racette et al. (24) and energy expenditure using the modified Weir equation (38). RMR was measured via indirect calorimetry using either a Deltatrac I or II respiratory gas analyzer (VIASYS Healthcare, Inc., SensorMedics, Yorba Linda, CA). Measurements took place between 08:00 and 09:00 a.m. after an overnight (≥ 8 h) fast, with subjects having refrained from exercise for at least 12 h. Subjects rested in a supine position in a quiet room free of bright light and at a comfortable temperature (20°C – 24°C) for at least 10 min before the 30-min respiratory gas collection period. The final 20 min of recordings were used to calculate RMR using the modified Weir equation (38). Periodic methanol burns were used to calculate correction factors, which were applied to the data to ensure equality between the two instruments. PAEE indices from DLW and RMR measures were calculated in several ways. A standard formula was used to calculate PAEE with the assumption of 10% of TEE as the thermic effect of food (i.e., $\text{PAEE} = 0.9\text{TEE} - \text{RMR}$). We also created a physical activity index (PAI), calculated as PAEE per body weight. Given the lack of agreement on the most appropriate way to adjust for the contribution of body weight to energy expenditure (20,26,29), we additionally examined physical activity level (PAL; i.e., TEE per RMR) and examined the residuals of PAEE regressed on body weight (PAEE_{adj}). Finally, we also used an analysis of covariance in which body weight was added to the models. All methods produced similar results and so we present only the data using PAEE, PAI, and PAEE_{adj} .

Physical activity monitors

Three different devices were used to objectively measure physical activity. A pedometer with 7-d memory (NL-2000; New-Lifestyles, Inc., Lee's Summit, MO) previously shown to be valid (8) was used to monitor steps per day and was worn on the left-hand side of a waist-worn elastic belt. Because of the limited memory of the device, pedometer data were only captured for the last 7 d of the 10-d monitoring period, and these days were used to calculate average steps per day. Accuracy for each pedometer and participant was checked using a 20-step test, and the monitor had to be within 1 step of the 20 counted steps. An ActiGraph GT1M uniaxial accelerometer (ActiGraph, LLC, Pensacola, FL) was used to measure steps per day and counts per day using 10-s epochs. The accelerometer was worn on the same belt as the pedometer but on the right-hand side of the body. An established algorithm was used to estimate nonwear periods using 60 min of 0 count per minute and 50-counts per minute thresholds (34). Wear time for the ActiGraph and pedometer was determined from the

ActiGraph data because both instruments were attached to the participants by the same method. For the estimation of PAEE from the ActiGraph, we used the Freedson and Crouter equations to estimate MET-minutes per day (7,11). Assuming 1 MET = 1 kcal·kg⁻¹·h⁻¹, we multiplied the MET-minutes per day estimated by the accelerometer equations by the subjects' body weight and divided by 60 to convert the estimates to kilocalories per day. Because our PAEE estimate from DLW had removed the resting EE underlying any movement, we also removed resting EE from the Freedson and Crouter-derived PAEE estimates by using the monitor wear time to determine the appropriate EE due to RMR to remove [e.g., Freedson estimate of PAEE - (RMR × wear time/24 h)].

The SenseWear Pro₃ Armband (SW; BodyMedia, Inc., Pittsburgh, PA) was used to measure steps per day, total daily energy expenditure, and PAEE using version 5.12 of the accompanying Innerview Research software. The armband calculated wear time itself.

For each monitor, any day with <10 h of wear time was excluded from the calculation of average steps, counts per day, or estimated energy expenditure. The activity logs were also examined for consistency of reported wear time with the monitors, and no substantial discrepancies were found as evaluated by the investigators.

Self-reported physical activity

Three questionnaires were used to assess self-reported physical activity. The CHAMPS (33) questionnaire was self-administered as designed, whereas the YPAS (9) and a modified version of the Physical Activity Scale for the Elderly (modPASE) (36) were interview-administered. In addition to weight-independent scoring of the questionnaires (i.e., kcal·kg⁻¹·d⁻¹), we calculated energy expenditure from the questionnaires (kcal·d⁻¹) using the average body weight across visits 1 and 2. For appropriate comparison with our measure of PAEE from doubly labeled water that had resting EE removed, we reduced the MET estimates of each activity by 1 MET before calculating EE.

CHAMPS is a comprehensive 41-item self-administered questionnaire that captures activities from various realms in a typical week during the past 4 wk and is described in the study of Stewart et al. (33). Using MET estimates developed specifically for this instrument (33), data from the active (nonsitting) behaviors on the instrument were scored and summarized. The YPAS queries the time spent during a typical week in the last month from a list of 27 different activities from various realms and is described in the study of DiPietro et al. (9). We scored the questionnaire as suggested; however, as with the CHAMPS and modPASE questionnaires, we reduced each activity by 1 kcal·min⁻¹ to subtract resting energy expenditure. For this questionnaire, we used body weight to convert the kilocalories per day estimates to kilocalories per kilogram per day. For the modPASE questionnaire, participants reporting participation in a given activity at least 10 times in the past 12 months are then further asked to report the frequency and duration spent in the following 12 activity types in the past 7 d (with appropriate examples given in each category): gardening or yard work, heavy housework, light housework, grocery shopping, laundry, stair climbing, walking for exercise, other walking, aerobics/calisthenics, weight training, high intensity exercises, or moderate-intensity exercises. Approximate MET values were assigned to each activity category (1), and the resultant summary score was calculated.

Statistical considerations and data analysis

The primary objective was to evaluate the validity of a variety of methods of physical activity measurement in elderly adults in comparison to a DLW measurement of energy expenditure. With 60 participants, any measure of PAEE could be estimated with a SE that was 0.13 times its sample SD, and the 95% confidence interval for the unknown mean

would have a length of 0.25 times the SD. We had 80% power to detect statistically significant single-correlation coefficients of 0.35 or higher with a two-tailed $\alpha = 0.05$.

The normality of the data was assessed, and subsequent analyses used both log-transformed measures as well as nonparametric statistics where appropriate. Intraclass correlations were calculated for repeat administrations of the three questionnaires as a measure of reliability and Bland–Altman analyses were performed. One subject did not complete both modPASE assessments, and this individual's data were excluded from the reliability analysis of that questionnaire. Steps per day as measured by the three objective devices were compared using both Spearman correlations and through Wilcoxon signed rank tests.

Both Pearson product–moment and Spearman rank correlations were calculated for the comparisons of each measure of physical activity to the various measures of PAEE. Similarly, we compared the questionnaire measures to the accelerometer measure of counts per day. As both the Pearson correlations of log-transformed data and the Spearman correlations of untransformed data produced similar results, the Spearman correlations are presented here. We present the root mean SE (RMSE) and the mean absolute percentage error (MAPE) for each comparison. RMSE is derived from the residual scatter in the dependent variable (y axis) that remains after removing the portion of the scatter explained by the relationship with the independent variable (x axis); it can be thought of as a measure of the average residual error (actually the square root of the average squared error). MAPE is also derived from the residual scatter, except that we are looking at the average of the absolute values of the residuals divided by the actual values, multiplied by 100; because it is an average of percentages, some find it more intuitive. Although not exactly equivalent, the RMSE is not dissimilar in nature to a SD, whereas the MAPE is not dissimilar in nature to a coefficient of variation, except that in this case each is describing the variance about the regression line. The Wilcoxon rank sum test was used to compare the median PAEE estimated by the various devices and questionnaires against the DLW-derived PAEE. Because a comparison between each of the six measures was made against DLW-derived PAEE, a Bonferroni adjustment was made such that $P < 0.008$ was considered to be significantly different. Bland–Altman analysis was also used to compare each of the measures to the mean PAEE (mean of device or questionnaire-estimated PAEE and DLW PAEE).

RESULTS

Characteristics of the 56 subjects in the current analysis are presented in Table 1. Subjects were predominantly female and White, with an average age of 74.7 yr. The median PAEE was 680 kcal·d⁻¹ (interquartile range (IQR) = 524–892), and the median PAL value was 1.72 (1.63–1.92). Table 2 presents compliance information from the objective monitoring devices. Monitor wear time for all devices was approximately 14 h·d⁻¹, and there were, on average, 10 valid days of information available for the ActiGraph and SW armband, and 7 valid days of information for the pedometer.

Intraclass correlation coefficients (ICC) indicate that the questionnaires demonstrated only moderate reliability in this population. ICC for the log-transformed CHAMPS, modPASE, and YPAS-derived kilocalories per day were 0.64, 0.60, and 0.73, respectively. Bland–Altman plots of the repeated administrations of the questionnaires are shown in Figure 1. The CHAMPS questionnaire was the most consistent with the least error at -11 ± 181 kcal·d⁻¹ (mean bias \pm SD). Both the modPASE and YPAS fared worse, with biases of 76 ± 354 and -78 ± 501 (mean \pm SD). All three questionnaires had poorer consistency as PAEE increased. The step count correlations between the accelerometer and pedometer, accelerometer and SW, and pedometer and SW were 0.88, 0.89, and 0.87, respectively.

Steps per day recorded by the pedometer, accelerometer, and SW armband were 7033 ± 2805 , 5917 ± 2171 , and 7022 ± 2417 (mean \pm SD), respectively. The accelerometer underestimated steps compared with both the pedometer (-1116 ± 179 steps per day (mean \pm SE), $P < 0.001$) and the SW (-1104 ± 130 steps per day, $P < 0.001$), whereas the difference in steps between the pedometer and SW was negligible (12 ± 184 steps per day, $P = 0.95$).

Table 3 presents the correlations between PAEE, body weight–adjusted PAEE (PAEE_{adj}), and PAI, and the various measures of physical activity. Although there is some variability, data show that regardless of the manner of expressing energy expenditure in physical activity, the various measures obtained from the objective monitors all had moderate correlations with our referent measures, in the range of 0.48–0.63. All the questionnaires had low correlations with measures of energy expenditure, in the range of 0.07–0.28. For all three questionnaires, the RMSE and MAPE for PAEE were larger than those for any of the objective measures. Use of the PAL produced similar correlations with the various measures, which are not presented here for simplicity. To test our *a priori* hypothesis about the relative validity of the measures, we used the correlations between PAEE and the SW armband (ρ_1), the pedometer (ρ_2), and then in separate comparisons: the Crouter and Freedson–estimated PAEE from the accelerometer (ρ_3) and the various questionnaire measures for subjective self-report results (ρ_4). We tested our null hypothesis ($H_0: \rho_1 = \rho_2 = \rho_3 = \rho_4$) against the ordered alternative hypothesis ($H_1: \rho_1 > \rho_2 > \rho_3 > \rho_4$), as outlined in the introduction. In general, we were able to reject the null hypothesis that all the assessment tools were equally correlated with DLW-estimated energy expenditure; however, the strength of ordering of the correlations that we found was mainly driven by which questionnaire we used for ρ_4 . When using the CHAMPS questionnaire in the comparative correlations, we did not reject our null hypothesis of equal correlations. Using the modPASE questionnaire produced results of varying significance depending on whether the Crouter ($P = 0.046$) or Freedson ($P = 0.075$) equations were used for the ActiGraph. With the YPAS questionnaire as the fourth variable in the comparison, we rejected the null hypotheses of equal correlations toward the ordered alternative. We note that ρ_1 , ρ_2 , and ρ_3 are fairly close; our rejection of equal correlations apparently depends on ρ_4 .

We also compared questionnaire measures to the accelerometer counts per day in the current study to allow comparison to previous studies in which accelerometers were used as the criterion measure (12,18,37). The CHAMPS, YPAS, and modPASE questionnaire–derived kilocalories per day had correlations of 0.52 ($P < 0.001$), 0.37 ($P < 0.01$), and 0.36 ($P < 0.01$), respectively, compared with the counts per minute of the ActiGraph.

In addition to relative agreement, we also sought to examine absolute estimates of average energy expenditure for the group for PA measures in which EE was estimable (all but the pedometer). Figure 2 compares the median PAEE from each of the physical activity measurement tools versus DLW-derived PAEE. The SW, Crouter equation with the ActiGraph, CHAMPS, and modPASE were all significantly different from DLW PAEE ($P < 0.008$). The Freedson equation with the ActiGraph data was not significantly different from PAEE (median underestimate = $125 \text{ kcal}\cdot\text{d}^{-1}$, IQR = -30 to $245 \text{ kcal}\cdot\text{d}^{-1}$). The YPAS was not significantly different from the DLW-derived value (median = $35 \text{ kcal}\cdot\text{d}^{-1}$, median = -212 to $385 \text{ kcal}\cdot\text{d}^{-1}$).

Figure 3 presents the Bland–Altman plots assessing agreement between each of the objective device or questionnaire-derived estimates of PAEE versus the mean PAEE. The mean bias and SD for the SW, Crouter equation from the ActiGraph, Freedson equation from the ActiGraph, CHAMPS, modPASE, and YPAS questionnaires, respectively, were -398 ± 241 , 342 ± 256 , -125 ± 209 , -419 ± 280 , -225 ± 401 , and $-35 \pm 499 \text{ kcal}\cdot\text{d}^{-1}$.

(mean \pm SD), demonstrating wide variability. Both the SW and CHAMPS tended to underestimate activity energy expenditure in individuals with higher levels of DLW PAEE.

DISCUSSION

Our results suggest that physical activity measures from all three objective monitoring devices correlate better with PAEE than any of the three questionnaires we examined in older adults, although statistically, the CHAMPS questionnaire was not significantly different from the monitors. In contrast to our hypothesis regarding the relative correlations between tools, the multisensor armband performed no better than the other objective monitors, and surprisingly, the pedometer did as well as the other two more sophisticated and expensive devices in regard to ranking physical activity. Although we did find moderate correlations between the various energy expenditure assessment tools with PAEE, absolute estimates of PAEE by the various devices had substantial error, and only for two of our six measures was the median PAEE of the group not different from our criterion measure. For the YPAS, which was good at the group level, there was a substantial error in individual estimation. With the possible exception of the Freedson equation used on the ActiGraph data, no measurement was good at both ranking the individuals and estimating the average PAEE of the group. These results suggest that, if total volume of activity energy expenditure is what is important to measure, a pedometer is the most cost-effective method for use in ranking older adults by physical activity level. Among the self-reported measures examined, the CHAMPS questionnaire seems to be the best choice for ranking individuals by PAEE, although absolute estimates of PAEE by this instrument at either the group or individual level were not accurate.

The reliability of the surveys we examined was found to be moderately strong based on ICC, similar to what has been reported previously for the CHAMPS and YPAS surveys (9,33). Given that the use of ICC may not be ideal when assumptions of normality or equality of variances are not met, as well as their sensitivity to the range of scores, we also examined reliability using Bland–Altman methodology. Results suggested that the CHAMPS questionnaire was the most consistent across the full range of PAEE.

Few previous studies have directly compared objective and self-reported physical activity measures against doubly labeled water in older adults. In older Japanese men, activity records were similarly correlated to doubly labeled water measures of energy expenditure ($r = 0.76$) as a Lifecorder accelerometer ($r = 0.83$) (25). Our results are consistent with those of a recent study examining the convergent and construct validity of the Actigraph GT1M, a Yamax pedometer, and the Zutphen questionnaire measure of physical activity in adults aged 65 and older (15). Although ours was a study of criterion validity, the main conclusions are consistent, in that both the accelerometer and pedometer compared better to health-related variables than did the questionnaire (15).

Studies comparing various accelerometers to doubly labeled water-estimated PAEE in older adults have compared favorably to PAEE in some (13,22,25) but not all studies (4). Our correlation of armband-estimated PAEE with DLW is consistent with a prior study in younger adults (31). And in two studies that have compared the SW and ActiGraph accelerometers versus either indirect calorimetry or the IDEEA monitor (a device that uses multiple sensors and pattern recognition to identify and measure physical activities), the armband has done somewhat better than the Actigraph device used (model 7164) (2,39). These studies were in young adults, however, and overall physical activity levels were likely much higher, and this can increase correlations by increasing the signal-to-noise ratio of the measures. In our study, the armband fared no better than did the other, more simple, objective devices, and this may be because the proprietary equations used by SW to

calculate PAEE were not developed specifically on populations of older adults. Perhaps with future versions of the SW software, estimates in the elderly will improve.

We propose that the limited time spent in activities of high intensity among older adults accounts for the consistency between the accelerometer and the pedometer when compared with measured energy expenditure of physical activity (34). In other words, because most of their movement occurs at the lower end of the intensity spectrum, there is little information to be gained by quantifying intensity of the total movements (i.e., a step is largely a step). This limited time spent in activities of higher intensity might also account for why the questionnaires did so poorly compared with the activity energy expenditure measures. It is clear from the graphs that even when little to no physical activity is reported on the questionnaires, the subjects are expending energy in physical activity. This is consistent with the postulation that unaccounted for activities of daily living on surveys may be one reason that these instruments do not correlate well with doubly labeled water (23). And in fact, the questionnaires were more strongly correlated with the ActiGraph counts than with the DLW PAEE. If, however, total PAEE, including that derived from activities of daily living is what is important for a given health outcome, and this may be particularly true in older adults, then questionnaires may not be adequate. Given the poor results of the questionnaires, and the cost differential between the objective monitors, a pedometer may be the tool of choice to measure the activity of older adults and has real potential in larger epidemiologic studies.

We chose to examine the relationship between the objective and questionnaire measures of physical activity with several different expressions of PAEE. In particular, body weight is highly related to PAEE, with heavier persons requiring greater energy expenditure to perform the same amount of work. When attempting to quantify the actual physical activity, it is important to try and remove the effect of body weight from the calculated PAEE, and there is considerable disagreement about how best to do this, nor is it clear that any method is sufficient (20,26,29). In light of this, we expressed physical activity both as weight-dependent and weight-independent measures when possible and compared these measures appropriately with PAEE, PAI, PAL, or PAEE residuals. Importantly, no matter the expression of PAEE, although the absolute correlations vary somewhat, the conclusions drawn would be the same.

In addition to relative agreement, we also examined the ability of the measures to estimate PAEE. This was not possible for the pedometer, and for the ActiGraph, we used equations that were not developed for use in older adults. Interestingly, the EE from the equations and the average counts-per-minute output were similarly correlated to PAEE. On a group level, the SW, CHAMPS, and modPASE significantly underestimated PAEE, whereas the Crouter equation overestimated PAEE. Only the Freedson equation and YPAS provided reasonable group averages; however, the IQR were large, such that there was substantial error in individual prediction. Although it is generally agreed that self-reporting results in an overestimation of actual activity levels (34), we did not find this to be true. We propose that this is because PAEE from DLW was our referent measure, which includes EE from any movement, not just structured or planned activities that are more easily recalled. It seems that there is a substantial component of PAEE in older adults that is not captured by most self-report instruments. This underreporting is consistent with evidence in older adults using the Minnesota Leisure Time Physical Activity Questionnaire (LTPA) (32). Consistent with our study, however, the YPAS has previously been shown to provide a reasonable estimate of PAEE at the group level (32). In terms of use of the ActiGraph to estimate PAEE, the Freedson equation, which was developed using a walking and jogging protocol, has generally been found to underestimate PAEE. However, when examined for accuracy of prediction in non-walking or running activities, it has been found to overestimate sedentary

activities such as lying, computer work, and standing (6), which may have resulted in the better estimate in our older adults.

Although the pedometer was as good on the relative scale as the other objective devices, it is limited by its inability to capture the parameters of the activity including frequency, intensity, and duration of bouts of activity. If, however, links to risk of various health or disease are identified with steps per day, then the pedometer has great public health appeal as an easily usable and affordable way to quantify activity and investigate dose–response relationships. This would have utility for researchers, clinicians who may be able to prescribe activity in steps per day, as well as the general public who could easily determine whether they were meeting guidelines by using a validated pedometer.

As with any validation study, we are limited by rapid changes in technology, and our conclusions are only applicable to the specific model devices and software used here. Our study may also be limited by the “volunteer effect,” in that our subjects may not be representative of older adults in general and additionally is limited by the predominantly female sample. Although their average PAL value indicates that they were active for their age, the ActiGraph counts per minute per day are consistent with NHANES survey data for people aged 65 yr and older (34). Regardless, we did have a wide range of activity represented, and our comparison of multiple measurements in the same study is a strength of this analysis.

In summary, objective measures of physical activity more appropriately ranked older adults by their PAEE than did self-report instruments. The pedometer, accelerometer, and the multisensor device all performed equally well, and as such, the pedometer may be the most cost-effective and user-friendly of the devices. Importantly, our data suggest that even when the instruments did an acceptable job at estimating PAEE at the group level, individual errors were large. However, if ranking of individuals is sufficient, and if knowledge of the context, type, frequency, intensity, or duration of the activity is necessary, the combined use of a pedometer with a questionnaire such as CHAMPS would be beneficial. Future studies of health and disease-related outcomes in older adults should consider the use of a pedometer for appropriately ranking participants by their PAEE.

Acknowledgments

The authors thank the valuable contributions of Heidi Walaski, Jeanne Stublaski, Barbara Woodhouse, Linda Harris, Howard Bailey, Brad Julius, Brent Johnson, Jaclyn Krupsky, Aimee Mastrangelo, Tim Shriver, and Jamie Cooper for the data collection and management.

This study was supported by R21AG025839 from the National Institutes of Health; P30 CA14520 from the National Cancer Institute, National Institutes of Health; 1UL1RR025011 from the Clinical and Translational Science Award program of the National Center for Research Resources, National Institutes of Health; and from the Graduate School, University of Wisconsin-Madison.

References

1. Ainsworth B, Haskell W, Leon A, et al. Compendium of physical activities: classification of energy costs of human physical activities. *Med Sci Sports Exerc.* 1993; 25(1):71–80. [PubMed: 8292105]
2. Berntsen S, Hageberg R, Aandstad A, et al. Validity of physical activity monitors in adults participating in free living activities. *Br J Sports Med.* 2010; 44:657–64. [PubMed: 18628358]
3. Bonnefoy M, Normand S, Pachiardi C, Lacour J, Laville M, Kostka T. Simultaneous validation of ten physical activity questionnaires in older men: a doubly labeled water study. *J Am Geriatr Soc.* 2001; 49:28–35. [PubMed: 11207839]

4. Choquette S, Chuin A, Lalancette D, Brochu M, Dionne I. Predicting energy expenditure in elders with the metabolic cost of activities. *Med Sci Sports Exerc.* 2009; 41(10):1915–20. [PubMed: 19727021]
5. Cole T, Coward W. Precision and accuracy of doubly labeled water energy expenditure by multipoint and two-point methods. *Am J Physiol.* 1992; 263:E965–73. [PubMed: 1443129]
6. Crouter SE, Churilla JR, Bassett DR Jr. Estimating energy expenditure using accelerometers. *Eur J Appl Physiol.* 2006; 98:601–12. [PubMed: 17058102]
7. Crouter SE, Clowers KG, Bassett DR Jr. A novel method for using accelerometer data to predict energy expenditure. *J Appl Physiol.* 2006; 100:1324–31. [PubMed: 16322367]
8. Crouter S, Schneider P, Karabulut M, Bassett D Jr. Validity of 10 electronic pedometers for measuring steps, distance, and energy cost. *Med Sci Sports Exerc.* 2003; 35(8):1455–60. [PubMed: 12900704]
9. DiPietro L, Caspersen CJ, Ostfeld AM, Nadel ER. A survey for assessing physical activity among older adults. *Med Sci Sports Exerc.* 1993; 25(5):628–42. [PubMed: 8492692]
10. Ferrari P, Friedenreich C, Matthews CE. The role of measurement error in estimating levels of physical activity. *Am J Epidemiol.* 2007; 166:832–40. [PubMed: 17670910]
11. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc.* 1998; 30(5):777–81. [PubMed: 9588623]
12. Friedenreich CM, Courneya KS, Neilson H, et al. Reliability and validity of the past year total physical activity questionnaire. *Am J Epidemiol.* 2006; 163:959–70. [PubMed: 16524954]
13. Gardner AW, Poehlman ET. Assessment of free-living daily physical activity in older claudicants: validation against the doubly labeled water technique. *J Gerontol Med Sci.* 1998; 53A:M275–80.
14. Goran MI, Poehlman ET. Total energy expenditure and energy requirements in healthy elderly persons. *Metabolism.* 1992; 41:744–53. [PubMed: 1619993]
15. Harris TJ, Owen CG, Victor CR, Adams R, Ekelund U, Cook DG. A comparison of questionnaire, accelerometer, and pedometer: measures in older people. *Med Sci Sports Exerc.* 2009; 41(7): 1392–402. [PubMed: 19516162]
16. International Atomic Energy Agency. Assessment of body composition and total energy expenditure in humans using stable isotope techniques. Vienna (Austria): IAEA; 2009. [cited 2010 June 2]. Human Series No. 3. Available from: <http://www.iaea.org/Publications/index.html>
17. King G, Torres N, Potter C, Brooks T, Coleman K. Comparison of activity monitors to estimate energy cost of treadmill exercise. *Med Sci Sports Exerc.* 2004; 36(7):1244–51. [PubMed: 15235333]
18. Macfarlane D, Lee C, Ho E, Chan K, Chan D. Convergent validity of six methods to assess physical activity in daily life. *J Appl Physiol.* 2006; 101:1328–34. [PubMed: 16825525]
19. Mahabir S, Baer D, Giffen C, et al. Comparison of energy expenditure estimates from 4 physical activity questionnaires with doubly labeled water estimates in postmenopausal women. *Am J Clin Nutr.* 2006; 84:230–6. [PubMed: 16825700]
20. Mâsse L, Fulton J, Watson K, Mahar M, Meyers M, Wong W. Influence of body composition on physical activity validation studies using doubly labeled water. *J Appl Physiol.* 2004; 96:1357–64. [PubMed: 14660509]
21. Matthews CE, Chen KY, Freedson PS, et al. Amount of time spent in sedentary behaviors in the United States, 2003–2004. *Am J Epidemiol.* 2008; 167:875–81. [PubMed: 18303006]
22. Meijer EP, Goris AHC, Wouters L, Westererp K. Physical inactivity as a determinant of the physical activity in the elderly. *Int J Obes.* 2001; 25:935–9.
23. Neilson H, Friedenreich C, Brockton N, Millikan R. Physical activity and postmenopausal breast cancer: proposed biologic mechanisms and areas for future research. *Cancer Epidemiol Biomarkers Prev.* 2009; 18:11–27. [PubMed: 19124476]
24. Racette S, Schoeller DA, Luke A, Shay K, Hnilicka J, Kushner R. Relative dilution spaces of ²H- and ¹⁸O-labeled water in humans. *Am J Physiol Endocrinol Metab.* 1994; 267:E585–90.
25. Rafamantanantsoa H, Ebine N, Yoshioka M, et al. Validation of three alternative methods to measure total energy expenditure against doubly labeled water method for older Japanese men. *J Nutr Sci Vitaminol.* 2002; 48:517–23. [PubMed: 12778890]

26. Schoeller D, Jefford G. Determinants of the energy costs of light activities: inferences for interpreting doubly labeled water data. *Int J Obes.* 2002; 26:97–101.
27. Schoeller D, Ravussin E, Schutz Y, Acheson K, Baertschi P, Jequier E. Energy expenditure by doubly labeled water: validation in humans and proposed calculation. *Am J Physiol.* 1986; 250:R823–30. [PubMed: 3085521]
28. Schuit A, Schouten E, Westerterp K, Saris W. Validity of the Physical Activity Scale for the Elderly (PASE): according to energy expenditure assessed by the doubly labeled water method. *J Clin Epidemiol.* 1997; 50:541–6. [PubMed: 9180646]
29. Schutz Y, Weinsier R, Hunter G. Assessment of free-living physical activity in humans: an overview of currently available and proposed new measures. *Obes Res.* 2001; 9:368–79. [PubMed: 11399784]
30. Shephard R. Limits to the measurement of habitual physical activity by questionnaires. *Br J Sports Med.* 2003; 37:197–206. [PubMed: 12782543]
31. St-Onge M, Mignault D, Allison D, Rabasa-Lhoret R. Evaluation of a portable device to measure daily energy expenditure in free-living adults. *Am J Clin Nutr.* 2007; 85:742–9. [PubMed: 17344495]
32. Starling R, Matthews D, Ades P, Poehlman E. Assessment of physical activity in older individuals: a doubly labeled water study. *J Appl Physiol.* 1999; 86:2090–6. [PubMed: 10368377]
33. Stewart AL, Mills KM, King AC, Haskell WL, Gillis D, Ritter PL. CHAMPS physical activity questionnaire for older adults: outcomes for interventions. *Med Sci Sports Exerc.* 2001; 33(7): 1126–41. [PubMed: 11445760]
34. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc.* 2008; 40(1):181–8. [PubMed: 18091006]
35. U.S. Department of Health and Human Services, Physical Activity Guidelines Advisory Committee. Physical Activity Guidelines Committee Report, 2008. Washington (DC): US Department of Health and Human Services; 2008. p. A2-3.
36. Washburn RA, Smith KW, Jette AM, Janney CA. The Physical Activity Scale for the Elderly (PASE): development and evaluation. *J Clin Epidemiol.* 1993; 46:153–62. [PubMed: 8437031]
37. Washburn R, Ficker J. Physical Activity Scale for the Elderly (PASE): the relationship with activity measured by a portable accelerometer. *J Sports Med Phys Fitness.* 1999; 39:336–40. [PubMed: 10726435]
38. Weir J. New methods for calculating metabolic rate with special reference to protein metabolism. *J Physiol.* 1949; 109:1–9. [PubMed: 15394301]
39. Welk G, McClain JJ, Eisenmann JC, Wickel EE. Field validation of the MTI ActiGraph and BodyMedia armband monitor using the IDEEA monitor. *Obesity.* 2007; 15:918–28. [PubMed: 17426327]
40. Westerterp K. Assessment of physical activity: a critical appraisal. *Eur J Appl Physiol.* 2009; 105:823–8. [PubMed: 19205725]

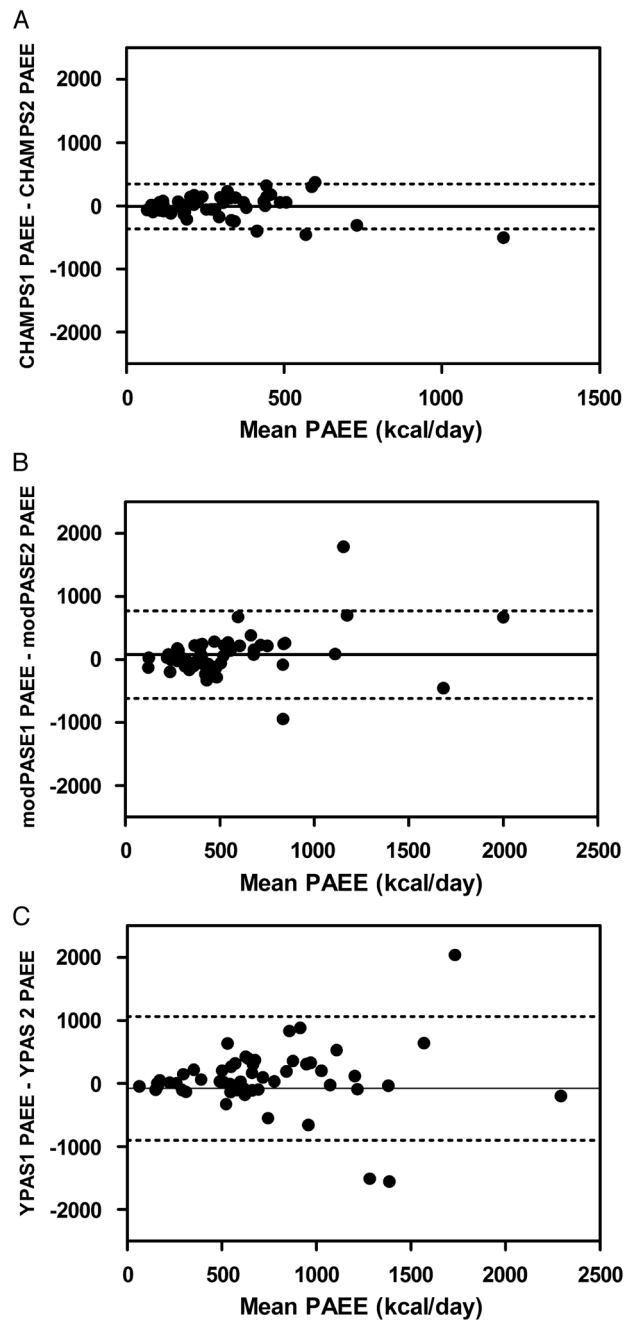


FIGURE 1. Bland–Altman plots of repeat administrations, 10 d apart, of the physical activity questionnaires. Mean differences in PAEE (first administration – second administration) are plotted against the mean PAEE ($\text{kcal}\cdot\text{d}^{-1}$) derived from CHAMPS (A), modPASE (B), and YPAS (C).

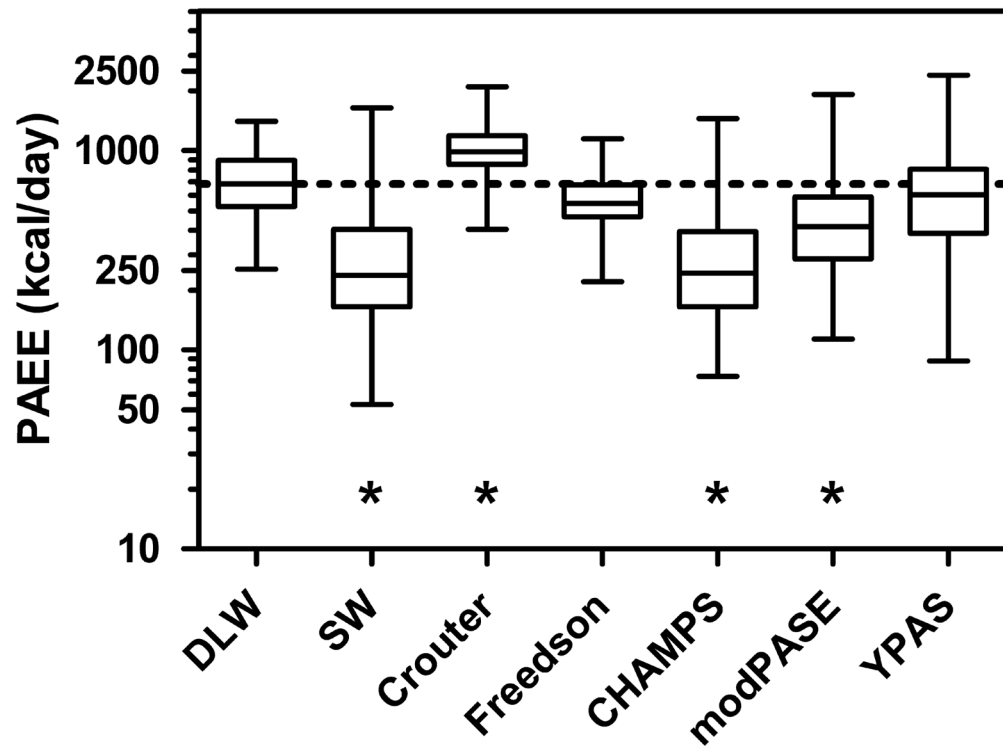


FIGURE 2.

Comparison of median PAEE ($\text{kcal}\cdot\text{d}^{-1}$) values from DLW versus SW armband, Freedson equation from ActiGraph, Crouter equation from ActiGraph, and CHAMPS, modPASE, and YPAS questionnaires. The y axis is presented in the log scale. The *solid line* represents median value, *box edges* are the 25th and 75th percentiles, and the *whiskers* indicate the range. The *dashed horizontal line* is at the median value ($680 \text{ kcal}\cdot\text{d}^{-1}$) for DLW.

*Significantly different from PAEE based on Wilcoxon rank sum test, $P < 0.008$ (Bonferroni adjustment: $0.05/6$ comparisons = 0.008).

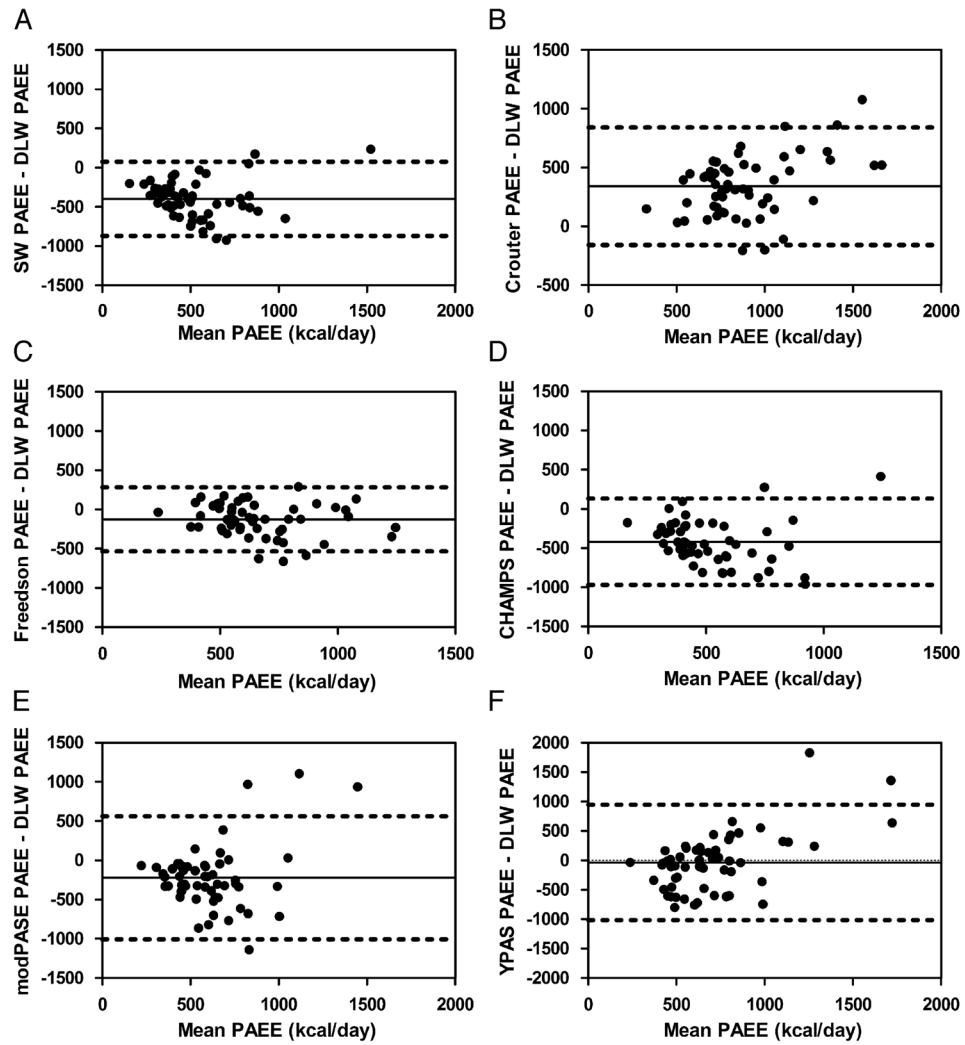


FIGURE 3. Bland–Altman plots of PAEE ($\text{kcal}\cdot\text{d}^{-1}$) estimated by different physical activity measures assessed for agreement with DLW-derived PAEE. Differences ($\text{PAEE}_{\text{predicted}} - \text{PAEE}_{\text{DLW}}$) are plotted against the mean PAEE (mean of $\text{PAEE}_{\text{predicted}}$ and PAEE_{DLW}) for the SW armband (A), Crouter equation from the ActiGraph (B), Freedson equation from the ActiGraph (C), CHAMPS questionnaire (D), modPASE questionnaire (E), and YPAS questionnaire (F).

TABLE 1

Subject characteristics and energy expenditure data ($n = 56$).^a

Characteristic	Value
Female (%)	79
White (%)	98
Age (yr)	74.7 ± 6.5
Body mass (kg)	69.2 ± 14.5
Body mass index (kg·m ⁻²)	25.8 ± 4.2
Current smoker ^b (%)	4
Education ^b (%)	
≤High school	21
College	44
≥Graduate work	35
Current disease ^b	
Arthritis	50
Back pain	10
Hip/knee pain	24
Osteoporosis	21
Lung disease	8
Cancer	15
Hypertension	35
TEE ^c (kcal·d ⁻¹)	2148 (1762–2420)
RMR ^c (kcal·d ⁻¹)	1165 (1062–1326)
PAEE ^c (kcal·d ⁻¹)	680 (524–892)
PAL ^c	1.72 (1.63–1.92)
PAI ^c (kcal·kg ⁻¹ ·d ⁻¹)	10.40 (7.76–12.11)

^aPresented as percent of group, mean ± SD, or median (IQR).

^bData are missing for smoking, education, and current diseases in four subjects, with the exception of hip/knee pain, which has five missing data points.

^cPAI, physical activity index; PAL, physical activity level; RMR, resting metabolic rate; TEE, total energy expenditure.

TABLE 2Physical activity monitor compliance data.^a

Measurement	Value
Pedometer	
Average wear time (h·d ⁻¹)	13.8 ± 1.2
Valid number of monitoring days	6.8 ± 0.7
ActiGraph	
Average wear time (h·d ⁻¹)	13.8 ± 1.2
Valid number of monitoring days	9.7 ± 0.9
SenseWear armband	
Average wear-time (h·d ⁻¹)	14.3 ± 1.6
Valid number monitoring days	9.7 ± 1.0

^aPresented as mean ± SD.

TABLE 3

Spearman correlation coefficients between physical activity monitors, questionnaire measures, and doubly labeled water-derived measures of PAEE.^{a,b}

Measure	PAEE ^a (kcal·d ⁻¹)			PAEE _{adj} ^a			PAI ^a (kcal·kg ⁻¹ ·d ⁻¹)					
	RMSE	MAPE	r	RMSE	MAPE	r	RMSE	MAPE	r			
SenseWear armband												
PAEE ^a (kcal·d ⁻¹)	210	26.8	0.479	<0.01								
Steps per day	212	26.3	0.564	<0.01	185	-75.2	0.604	<0.001	2.90	22.3	0.590	<0.001
ActiGraph												
Crouter EE (kcal·d ⁻¹)	185	22.5	0.600	<0.01								
Freedson EE (kcal·d ⁻¹)	202	24.4	0.489	<0.01								
Average counts per minute	206	25.2	0.559	<0.01	185	-48.1	0.584	<0.001	2.97	23.1	0.561	<0.001
Steps per day	212	26.0	0.585	<0.01	183	-103.4	0.618	<0.001	2.84	20.7	0.629	<0.001
Pedometer												
Steps per day	213	26.8	0.530	<0.01	182	-80.2	0.585	<0.001	2.82	21.3	0.597	<0.001
CHAMPS												
PAEE ^a (kcal·d ⁻¹)	242	30.4	0.278	0.04								
PAI ^a (kcal·kg ⁻¹ ·d ⁻¹)					226	3.1	0.225	0.096	3.42	29.0	0.234	0.083
modPASE												
PAEE ^a (kcal·d ⁻¹)	256	32.4	0.195	0.15								
PAI (kcal·kg ⁻¹ ·d ⁻¹)					234	-20.6	0.111	0.412	3.51	29.3	0.180	0.183
YPAS												
PAEE ^a (kcal·d ⁻¹)	252	32.8	0.074	0.59								
PAI ^a (kcal·kg ⁻¹ ·d ⁻¹)					231	-15.6	0.093	0.496	3.45	29.4	0.182	0.179

^aMAPE, mean absolute percent error (%); PAEE_{adj}, PAEE adjusted for body weight; PAI, physical activity index (= PAEE per body weight (kg)); RMSE, root mean squared error for univariate regression.

^bCorrelations are only presented for caloric expenditure when the units are equivalent (i.e., kcal·d⁻¹ or kcal·kg⁻¹·d⁻¹).