Validity of Consumer-Based Physical Activity Monitors

JUNG-MIN LEE[1], YOUNGWON KIM[2], and GREGORY J. WELK[2]

[1]*School of Health, Physical Education and Recreation, University of Nebraska at Omaha, Omaha, NE;* [2]*Department of Kinesiology, Iowa State University, Ames, IA*

ABSTRACT

LEE, J.-M., Y. KIM, and G. J. WELK. Validity of Consumer-Based Physical Activity Monitors. *Med. Sci. Sports Exerc.*, Vol. 46, No. 9, pp. 1840–1848, 2014. **Background**: Many consumer-based monitors are marketed to provide personal information on the levels of physical activity and daily energy expenditure (EE), but little or no information is available to substantiate their validity. **Purpose**: This study aimed to examine the validity of EE estimates from a variety of consumer-based, physical activity monitors under free-living conditions. **Methods**: Sixty (26.4 ± 5.7 yr) healthy males (n = 30) and females (n = 30) wore eight different types of activity monitors simultaneously while completing a 69-min protocol. The monitors included the BodyMedia FIT armband worn on the left arm, the DirectLife monitor around the neck, the Fitbit One, the Fitbit Zip, and the ActiGraph worn on the belt, as well as the Jawbone Up and Basis B1 Band monitor on the wrist. The validity of the EE estimates from each monitor was evaluated relative to criterion values concurrently obtained from a portable metabolic system (i.e., Oxycon Mobile). Differences from criterion measures were expressed as a mean absolute percent error and were evaluated using 95% equivalence testing. **Results**: For overall group comparisons, the mean absolute percent error values (computed as the average absolute value of the group-level errors) were 9.3%, 10.1%, 10.4%, 12.2%, 12.6%, 12.8%, 13.0%, and 23.5% for the BodyMedia FIT, Fitbit Zip, Fitbit One, Jawbone Up, ActiGraph, DirectLife, NikeFuel Band, and Basis B1 Band, respectively. The results from the equivalence testing showed that the estimates from the BodyMedia FIT, Fitbit Zip, and NikeFuel Band (90% confidence interval = 341.1–359.4) were each within the 10% equivalence zone around the indirect calorimetry estimate. **Conclusions**: The indicators of the agreement clearly favored the BodyMedia FIT armband, but promising preliminary findings were also observed with the Fitbit Zip. **Key Words:** VALIDATION, ACTIVITY MONITOR, PHYSICAL ACTIVITY, ENERGY EXPENDITURE

Accelerometers have become the standard method for assessing physical activity (PA) in field-based research (23). They are small, noninvasive, and easy to use, and they provide an objective indicator of PA over extended periods. They have been used almost exclusively for research, but advances in technology have led to the emergence of new consumer-based activity monitors designed for use by individuals interested in fitness, health, and weight control. Examples include the BodyMedia FIT (BMF), the Fitbit, the DirectLife (DL), the Jawbone Up (JU), the NikeFuel Band (NFB), and the Basis B1 Band (BB). The development of these consumer-based monitors has been driven in large part by the increased availability of low-cost accelerometer technology in the marketplace. The refinement of other technology (e.g., Bluetooth) and the increased sophistication of personalized social media applications have also spurred the movement. These new accelerometry-based monitors provide consumers with the ability to estimate PA and energy expenditure (EE) and track data over time on Web sites or through cell phone applications.

Other technologies have also been adapted to capitalize on consumer interest in health and wellness. Pedometers developed originally to measure steps have been calibrated to estimate EE and to store data over time (1). Global positioning system monitors, developed primarily for use in navigation, are now marketed to athletes and recreation enthusiasts to monitor speed and EE from the activity. HR monitors, originally marketed to athletes, have also been modified and marketed to appeal to most recreational athletes interested in health and weight control. Although the functions and features vary, all of these devices attempt to provide users with an easy way to objectively monitor their PA and EE over time.

The increased availability of monitoring technology provides consumers with options for PA self-monitoring, but these tools may also have utility for applied field-based research or intervention applications designed to promote PA in the population. However, little or no information is available to substantiate the validity of these consumer-based activity monitors under free-living conditions. It is important to formally evaluate the validity of these various

Address for correspondence: Jung-Min Lee, Ph.D., School of Health, Physical Education and Recreation, 6001 Dodge Street, Omaha, NE 68182; E-mail: jungminlee@unomaha.edu.
Submitted for publication October 2013.
Accepted for publication January 2014.

0195-9131/14/4609-1840/0
MEDICINE & SCIENCE IN SPORTS & EXERCISE®

DOI: 10.1249/MSS.0000000000000287

devices so consumers, fitness professionals, and researchers can make informed decisions when choosing one of the monitors. Research on PA assessment has progressed by continually evaluating new technologies and approaches against existing tools. The present study adds new information to the literature by formally evaluating the validity of eight different consumer-based, activity-monitoring technologies under semistructured free-living conditions, with estimates of EE from a portable metabolic analyzer as the criterion measure.

METHODS

Participants

Sixty healthy men (n = 30) and women (n = 30) volunteered to participate in the study. Participants did not have major diseases or illnesses, did not use medications that would affect their body weight or metabolism, and were nonsmokers determined by the self-report health history questionnaire. Individuals were recruited from within the university and surrounding community through posted announcements and word of mouth. Approval from the institutional review board of Iowa State University was obtained before beginning this study. Participants were aware of the procedures and purpose of the study before they signed the informed consent document.

Instruments

Oxycon mobile 5.0. The Oxycon mobile 5.0 (OM; Viasys Healthcare Inc., Yorba Linda, CA) is a portable metabolic analyzer that allows the measurement of oxygen consumption under free-living conditions and was used in this study as the criterion measure. In a recent validation study, the OM provided similar metabolic parameters ($\dot{V}_E$, $\dot{V}O_2$, and $\dot{V}CO_2$) compared with the Douglas bag method. The mean differences reported in the study were in all cases less than 5% (18). The expired gases were collected using Hans Rudolph masks (Hans Rudolf, Inc., Kansas City, MO). Volume and gas calibrations were performed before each trial by following manufacturer's instructions.

BodyMedia FIT. The BMF (BodyMedia Inc., Pittsburgh, PA) is a consumer version of a research-based armband monitor known as the SenseWear Armband. The SenseWear is an innovative, multisensor activity monitor that integrates movement data from a three-dimensional accelerometer with various heat-related variables (e.g., heat flux) and galvanic skin response to estimate EE. The BMF uses the same technology as the SenseWear device. However, it is designed to facilitate personal self-monitoring and weight control. The device comes with a watch interface and can connect wirelessly through Bluetooth with Smartphone apps for data monitoring. The monitor has rechargeable batteries that can be used to collect and store data for 2 wk. Data can be downloaded through a USB cable and viewed through a personalized Web-based software tool (ProConnect) to monitor results over time. The software also features an integrated tool for reporting calorie intake, which enables participants to track energy balance, and to set and monitor weight loss goals; the software interface also enables users to connect with health coaches for guidance and support. Numerous studies (5) have supported the validity of the SenseWear, yet studies to date have not evaluated the BMF monitor.

DirectLife. The DL (DirectLife, Philips Lifestyle Incubator, Amsterdam, The Netherlands) is a triaxial accelerometry-based monitor, based on a previously developed research device called the Tracmor (3). This device is a small (3.2 × 3.2 × 0.5 cm), light-weight (12.5 g) instrument. The DL is waterproof to a 3-m depth and has a battery life of 3 wk with an internal memory that can store data for up to 22 wk. The features of the DL have been designed to enhance wearability and reduce the interference of the monitoring system with spontaneous activity behavior. A personal Web page that provides statistics, tips, and activity ideas allows participants to track their estimated EE based on their activities. The validity of the Tracmor has been supported, but the DL has not been tested to date.

Fitbit One (FO). The FO (Fitbit Inc., San Francisco, CA) is a triaxial, accelerometry-based device that can measure steps taken, floors climbed, distance traveled, calories burned, and sleep quality. This monitor is a small (48.0 × 19.3 × 9.6 mm), light-weight (8 g) instrument. The FO has a 5- to 10-d battery life and an internal memory that can store data for up to 23 d. The unique feature of the FO is a wireless function that makes it possible to automatically upload data to the Web site without synchronizing the monitor to the computer. The Fitbit Ultra has been tested against estimates from a room calorimeter. However, was found to significantly underestimate total EE (8).

Fitbit Zip (FZ). The FZ (Fitbit Inc., San Francisco, CA) is a triaxial accelerometer that can measure steps taken, distance traveled, and calories burned. This monitor is smaller (35.6 × 28.9 × 9.6 mm) than the FO but has an expanded battery life—approximately 4–6 months—and is slightly less expensive.

Jawbone UP Band. The JU (Jawbone, San Francisco, CA) is a wrist-worn, three-dimensional, accelerometry-based device that can assess sleep patterns and PA patterns throughout the day. The JU corresponds with an iOS device (iPhone 3GS or higher) via a 3.5-mm standard cable to synchronize data. The JU is water resistant up to 1 m and has a battery lifespan of 10 d. No research has been published on the JU.

NikeFuel Band. The NFB (Nike Inc., Beaverton, OR) is a wrist-worn, three-dimensional, accelerometry-based device, which assesses body movement, steps taken, distance, and calories burned. Data can be synchronized to the Nike+ Connect (Web site) via the clasp, which doubles as a USB cable or the accompanying application for an iOS device (iPhone) using Bluetooth. The NFB's battery lasts up to 4 d, and the band uses a series of 100 mini-LED lights to provide a clear presentation of PA data (i.e., steps, distance, and activity EE). No published research has been reported on the NFB.

Basis B1 Band. The BB (Basis Science Inc., San Francisco, CA) is a wrist watch–style activity monitor with multiple sensors that integrates movement data from a triaxial accelerometer with various heat-related variables, such as skin surface temperature, ambient temperature, and galvanic skin response to estimate EE. The unique feature of the BB is its advanced optical sensing technology, which accurately measures HR and blood flow. The battery in the BB lasts up to 5 d, and the BB is also reported to be waterproof. In addition, it includes a digital watch, packed in an LCD touch screen interface. No published research has been reported on the BB.

ActiGraph GT3X+(AG3X). The AG3X (ActiGraph, Pensacola, FL) is the most commonly used accelerometers for the assessment of PA under free-living conditions. It is marketed exclusively as a research instrument and has been used in numerous studies to provide objective estimates of PA. The latest version of the AG3X features a triaxial accelerometer. The AG3X is not a consumer device. However, it is included in the study for comparison purposes.

Procedures

Participants reported to the laboratory twice. On the first visit, they were instructed on the characteristics of the study before signing an informed consent and completing a self-report health history. Anthropometric measures were obtained at the beginning of the data collection session. Standing height was measured to the nearest 0.1 cm using a wall mounted Harpenden stadiometer (Harpenden, London, UK) using standard procedures. Body mass was measured with participants in light clothes and bare feet on an electronic scale (Seca 770) to the nearest 0.1 kg. The body mass index was calculated as weight (kg) / height squared (m^2). The percentage of body fat was assessed, using a handheld Bioimpedence Analysis devices (Omron, Shelton, CT). After anthropometric measurements, the participants were asked to lay down in bed for 10 min and then fitted with the potable metabolic analyzer (i.e., OM) to measure resting EE (REE) for 15 min. The estimated REE was expressed as kilocalories per minute by dividing the total EE value by 15. The REE measurement was performed in the morning (i.e., 6:00–9:00 a.m.) after a 10-h fast, following previously published guidelines (6).

For the second visit (i.e., 1 wk after the first visit), the participants were fitted with the portable metabolic analyzer and eight different types of activity monitors. The BMF monitor was worn on the nondominant arm. The DL monitor was worn on the chest with a necklace. The NFB and the JU were worn on the left wrist, and the BB was worn on the right wrist. All other monitors (i.e., FO, FZ, and ACT) were positioned along the belt according to the manufacturer's instructions. All instruments were synchronized and initialized using the participant's personal information (age, gender, height, weight, handedness, and smoker/nonsmoker) before the measurements. The test was performed at various times of day; however, participants were asked to abstain from eating and exercise for 4 h before the test. Each participant then performed an activity routine that included 13 different activities and lasted 69 min.

Participants performed each activity for 5 min, except the activities on the treadmill, which were 3 min. There was a 1-min break between each activity to facilitate transitions and tracking of data. Oxygen consumption ($\dot{V}O_2$) was simultaneously measured throughout the routine with an OM metabolic cart. These activities were categorized into four distinct PA types: 1) sedentary (reclining, writing at a computer), 2) walking (treadmill walking at 2.5 mph, treadmill brisk walking at 3.5 mph, self-paced overground walking, and self-paced overground walking with 15 kg backpack), 3) running (treadmill jogging at 5.5 mph, treadmill running at 6.5 mph), and 4) moderate-to-vigorous activities (ascending and descending stairs, stationary bike, elliptical exercise, Wii tennis play, and playing basketball with researchers). One BB, DL, AG3X, and two NFB data were excluded from the final data analysis because of the delay of the Web site connection and initialization error.

Most of the consumer-based activity monitors do not provide direct access to the raw data, so estimates of EE were obtained directly from the associated Web sites for each monitor. The consumer devices also do not typically provide access to raw (e.g., minute-by-minute) data; therefore, the total estimates of EE across the entire period were used for the analyses. The AG3X allows easy access to the raw movement counts, and then data from this monitor were processed using standard methods and aggregated to produce estimates for the same period. The latest Freedson algorithm (2011) was used to obtain the estimate EE.

Data Analyses

Breath-by-breath data from the indirect calorimetry were aggregated to provide average minute-by-minute values to facilitate integration with the estimates of EE from each monitor. Evaluation of the entire monitoring period was necessitated by the limitations of some of the software applications that do not report data on a minute-by-minute basis (several provided only estimates of total EE). The primary statistical analyses involved evaluating overall group differences in EE estimates from the eight methods across the entire monitoring period (69-min trial). Although this prevents an analysis of individual activities, the evaluation over the full monitoring period provides a more ecologically valid assessment of what the monitors do under real-world conditions. Many validation studies have focused exclusively on the point estimates of individual PA, but the most important consideration is how the devices perform during a sustained period of monitoring.

Each monitor uses different outcome measures to summarize the data. Several of the monitors provide estimates of activity EE (AEE; i.e., NFB, DL, JU, and AG3X); however, several others report estimates of total EE (i.e., BMF, FO, FZ, and BB). To provide comparable estimates, it was necessary

TABLE 1. Physical characteristics of male (n = 30) and female (n = 30) subjects.

	Male		Female	
	Mean ± SD	Range	Mean ± SD	Range
Age (yr)	28.6 ± 6.4	18.0–43.0	24.2 ± 4.7	18.0–38.0
Height (cm)	176.1 ± 5.4	166.4–186.5	166 ± 7	154.2–187.0
Weight (kg)	75.4 ± 9.5	56.3–93.1	60.3 ± 8.6	47.6–85.2
Body fat (%)	17.7 ± 6.2	5.7–31.7	20.4 ± 5.8	8.3–35.6
Body mass index (kg·m^{-2})	24.3 ± 2.6	19.5–28.0	21.8 ± 2.7	18.1–31.2

to add REE to the AEE values from some of the estimates. Each individual's measured REE (expressed in kilocalories per minute) was added to the estimated AEE value for monitors that reported this outcome instead of TEE. This ensured that we had comparable outcome measures of TEE for all monitors.

Descriptive analyses were conducted to examine associations with the criterion measure. Pearson correlations were computed to examine overall group-level associations. Mean absolute percent errors (MAPE) were also calculated to provide an indicator of overall measurement error. MAPE were computed as the average of absolute differences between the activity monitors and the OM value divided by the OM value, multiplied by 100. This is a more conservative estimate of error that takes into account both overestimation and underestimation because the absolute value of the error is used in the calculation.

A novel, statistical approach used in this study was the use of "equivalence testing" (9,24) to statistically examine measurement agreements between the activity monitors and the OM. In traditional hypothesis testing, the focus is on testing for a significant difference. Failing to reject the null hypothesis (e.g., that two methods are not different) allows one to report that there is no evidence of a difference. However, this does not necessarily imply that the estimates are equivalent (11). Using an equivalence test, it is possible to determine whether a method is "significantly equivalent" to another method (i.e., OM). With this type of analyses, it is important to specify an appropriate equivalence zone before testing. There is no definitive standard, but we selected a 10% error zone. With a 95% equivalence test (i.e., an alpha of 5%), an estimate is considered to be equivalent to the criterion-measured value (with 95% precision) if the 90% confidence interval (CI) for a mean of the estimated EE falls into the proposed equivalence zone (i.e., ±10% of the mean) of the measured EE from OM. The estimated EE and measured EE data across all monitors and the 90% CI for means of the estimated and measured EE were obtained from a mixed ANOVA to control for participants' level clustering.

To further evaluate individual variations in a more systematic way, Bland–Altman plots with corresponding 95% limits of agreement and fitted lines (from regression analyses between mean and difference) with their corresponding parameters (i.e., intercept and slope) were presented. A fitted line that provides a slope of 0 and an intercept of 0 exemplifies perfect agreement. The root mean square error (RMSE) and the percentage of the RMSE relative to the measured value were calculated for each device to enable comparisons with previous studies.

RESULTS

Descriptive statistics for the sample population are provided in Table 1. Participants' ages ranged between 18 and 43 yr. The body mass index and the percentage of body fat ranged between 19.5 and 28.0 kg·m^{-2} and between 5.7% and 31.7%, respectively.

Table 2 provides descriptive statistics (means ± SD) for all of the different monitors compared with the measured values from the OM. The measured value was 356.9 ± 67.6 kcal, and the estimates from the monitors ranged from a low of 271.1 ± 53.8 kcal (BB) to a high of 370.1 ± 51.5 kcal (JU).

Table 3 shows the correlation coefficients (r) between indirect calorimetry (i.e., OM) and consumer activity monitors. The strongest relationship between the OM and the monitors were seen for the BMF (r = 0.84) and the two Fitbit monitors (FO: r = 0.81 and FZ: r = 0.81). These monitors were also highly correlated with one another (BMF vs FO: r = 0.90). The correlation coefficients for the other monitors ranged from r = 0.14 to 0.73 when compared with the criterion measure (i.e., OM)

Figure 1 shows the MAPE for the various monitors (computed as the average absolute value of the errors relative to the OM). The magnitude of errors was least for the BMF (9.3%), followed by the FZ (10.1%) and the FO (10.4%). Error rates for the other monitors ranged from 12.2% to 23.5%.

The use of equivalence testing made it possible to determine whether the EE estimates from the monitors were equivalent to the estimate from the criterion measures (OM). The calculated 90% CI for the estimates from the monitors were compared with the computed equivalence zone for the OM. The estimated EE from the BMF, FZ, and Nike+ Fuel

TABLE 2. Estimated total EE (kcal) with added measured REE.

	N	Mean ± SD	Minimum	Maximum	RMSE (kcal)
OM	60	356.9 ± 67.6	263.3	594.0	0
BB	59	271.1 ± 53.8	137.5	397.5	68.0
NFB[a]	58	350.2 ± 41.8	281.0	488.0	64.0
DL[a]	59	320.4 ± 51.8	231.0	481.8	47.0
FO	60	330.9 ± 55.0	248.0	470.0	40.1
FZ	60	370.1 ± 51.5	275.0	526.0	40.8
JU[a]	60	333.8 ± 66.1	218.3	535.8	45.8
BMF	60	338.9 ± 59.4	250.9	533.8	36.8
ActiGraph[a]	59	326.2 ± 64.7	170.0	496.6	47.1

[a]Added measured REE.
OM, Oxycon Mobile; RMSE, root mean square error.

APPLIED SCIENCES

TABLE 3. Correlation matrix with added measured REE.

	OM	BB	NFB[c]	DL[c]	FO	FZ	JU[c]	BMF	ActiGraph[c]
OM	1	0.136	0.346[a]	0.729[a]	0.808[a]	0.807[a]	0.741[a]	0.842[a]	0.722[a]
BB		1	0.254	0.122	0.309[b]	0.161	0.135	0.240	0.174
NFB			1	0.361[a]	0.353[a]	0.218	0.401[a]	0.308[b]	0.402[a]
DL				1	0.720[a]	0.642[a]	0.729[a]	0.756[a]	0.768[a]
FO					1	0.868[a]	0.745[a]	0.884[a]	0.796[a]
FZ						1	0.741[a]	0.895[a]	0.772[a]
JU							1	0.797[a]	0.648[a]
BMF								1	0.818[a]
ActiGraph									1

[a]Correlation is significant at the 0.01 level (two-tailed).
[b]Correlation is significant at the 0.05 level (two-tailed).
[c]REE was added to the estimates.

were significantly equivalent to the measured EE from the OM. This is shown by the fact 90% CI for the estimated EE from the three monitors were completely within the equivalence zone of the measured EE (lower bound = 321.2 kcal, upper bound = 392.6 kcal). Plots showing the distribution of error for all monitors are shown in Figure 2.

Bland–Altman plot analyses showed the distribution of error and assist with testing for proportional systematic bias in the estimates. The plots show the residuals of the various EE estimates on the y-axis (OM − estimates) relative to the mean of two methods (x-axis). The plots (see Fig. 3) revealed the narrowest 95% limits of agreement for the BMF (difference = 143.3) and slightly higher values for the FO (difference = 155.9) and FZ (difference = 156.8). Values were higher still for the DL (difference = 182.9), the JU (difference = 188.6), the AG3X (difference = 193.4), the Nike+ Fuel (difference = 259.1), and the BB (difference = 327.2). A tighter clustering of data points about the mean for BMF, FZ, FO, and JU and less overall error were observed compared with the measured EE values. The slopes for the fitted line were not significant for BMF (slope = −0.13, P = 0.071), JU (slope = 0.03, P = 0.800), AG3X (slope = −0.05, P = 0.650), and BB(slope = −0.42, P = 0.080). This suggests no significant patterns of proportional systematic bias with these monitors. However, significant bias was observed for the Nike+ Fuel (slope = −0.68, P = 0.001), DL (slope = −0.31, P = 0.003), FZ (slope = −0.29, P = 0.001), and FO (slope = −0.22, P = 0.010).

DISCUSSION

The present study investigated the accuracy of a variety of consumer-based activity monitors for estimating EE in healthy adults under semistructured free-living conditions. The results showed favorable outcomes for the estimation of EE from some, but not all, of the various consumer-based activity monitors. With the exception of the BB, the majority of the monitors yielded reasonably accurate estimates of EE compared with the OM values (within approximately 10%–15% error). Of the eight monitors tested, the BMF had the highest correlations with OM (r = 0.84), the smallest MAPE value (9.3%), the smallest RMSE value (36.8 kcal), the lowest 95% limits of agreement (143 kcal), and no evidence of proportional bias.

The favorable results for the BMF device show that the consumer monitor provides similar validity as the established SenseWear Mini or Core monitor. A recent doubly labeled water study (4) demonstrated the SenseWear Mini yielded EE estimates within 22 kcal·d^{-1}, based on group averages. The

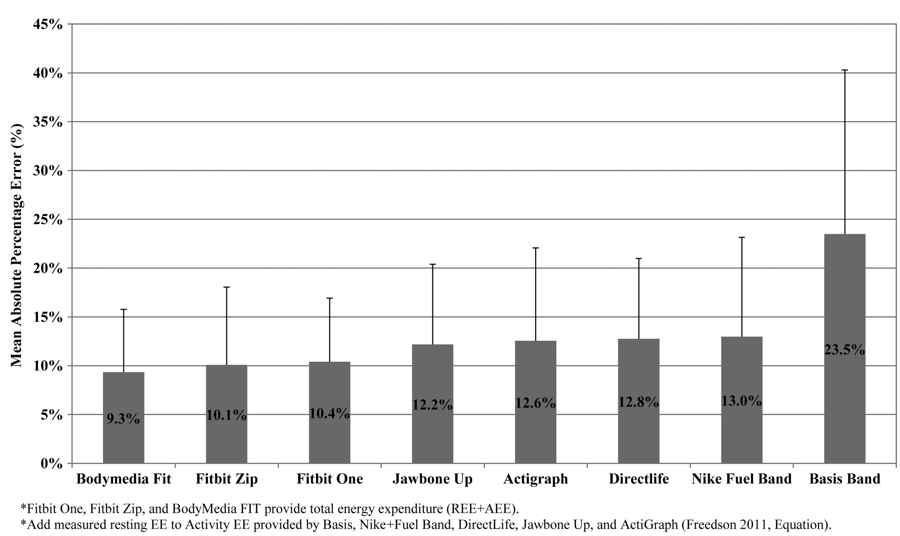


FIGURE 1—Mean absolute percentage error (±SD) for all monitors with measured REE (n = 60).

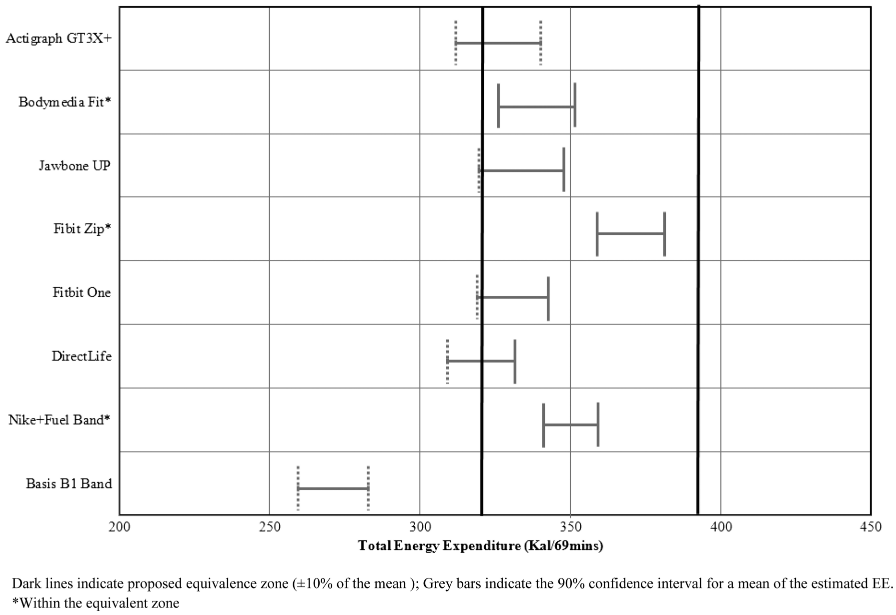


FIGURE 2—Results from 95% equivalence testing for agreement in total estimated EE between OM and all monitors.

absolute error rates in the doubly labeled water study were approximately 8% for both the BodyMedia Mini and the earlier SenseWear monitor. The error rates in the present study were comparable (~9%), which indicates the consumer monitor is providing similar accuracy as the established research device. The criterion measures in two studies (4,5) were different. However, the results were also consistent with other laboratory studies that have supported the validity of the SenseWear monitor. The robust predictive accuracy of the BMF (and the SenseWear) likely stems from the incorporation of both movement data and heat-related data in the prediction algorithm. All other devices used a single input source to evaluate the associated EE with activity (e.g., steps, counts, HR, ambulatory speed).

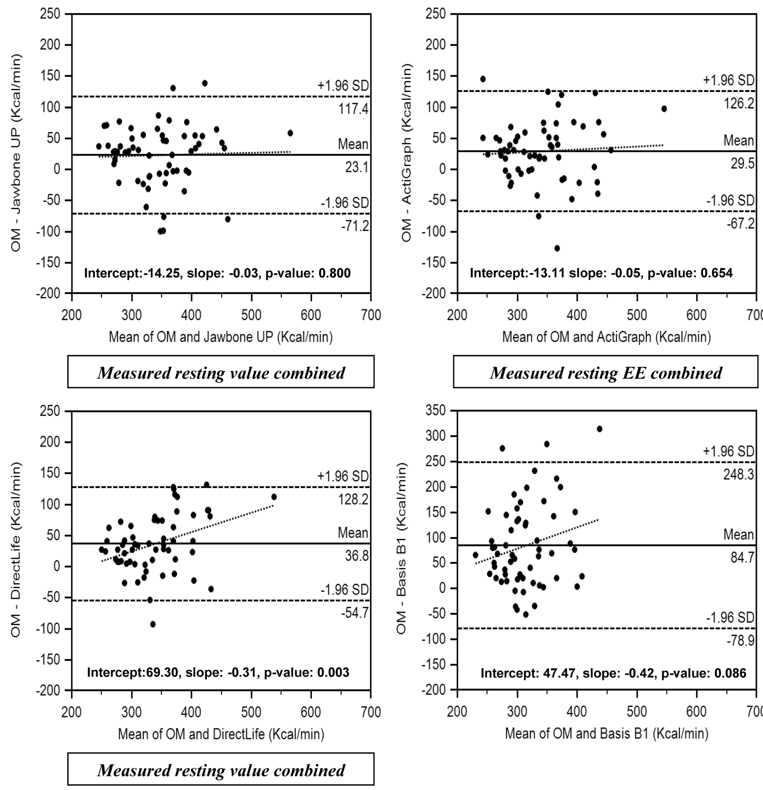
Although the BMF yielded the best overall results, the Fitbit monitors also performed well in this study. The Fitbit monitors had high correlations with OM and also with the BMF monitors. The Fitbit monitors also had similar values for MAPE (FO = 10.08% and FZ = 10.08%) and RMSE (FO = 40.11 kcal and FZ = 40.75 kcal) and slightly higher limits of agreement in the Bland–Altman plots as the BMF.

As mentioned, few studies have assessed any of these devices. However, some comparison can be made with the findings from a recent study (8) that reported the accuracy of some of the same consumer-based activity monitors compared with a room calorimeter. The reported RMSE% error of 17.9% for the DL is similar to the value obtained in the present study (14%). However, the comparison study (8) reported an RMSE% error of 28% for the Fitbit Ultra and 27% for the AG3X, using the standard Freedson equation. These values were considerably higher than those reported in our study. However, the mean RMSE was reduced to 12.9% for Fitbit monitor after the performed activities were manually entered into the Web-based software. The additional information about the type of activity likely allowed a more appropriate algorithm to be used for the estimation. However, no information was typed in to facilitate the estimation in our study, and it yielded similar RMSE values (FO = 15.2% and FZ = 15.4%). It is not clear how the Fitbit monitors work, but estimates may be enhanced by the inclusion of an altimeter sensor to capture altitude changes. This additional sensor may assist in capturing the increased energy cost of some activities (e.g., stair climbing). However, it is premature to draw a firm conclusion about the overall accuracy of the Fitbit monitors until additional testing is performed.

A unique advantage of the present study is the inclusion of an established research monitor into the protocol for comparison. The AG3X has been used in hundreds of studies and provides a useful comparison for the consumer models. In the present study, the AG3X provided similar EE estimates (MAPE = 12.6%) relative to the consumer-based activity monitor. It is noteworthy that consumer-based monitors perform similarly (or better) than the AG3X—especially because the values reported here for the AG3X are considerably better than past research. Previous studies (7,15,20) with the ActiGraph have shown MAPE values ranging from 4.5% to 29.4% for estimating EE or METs. The present study used a newly developed equation (Freedson 2011) and a new version of the ActiGraph monitor (GT3X+), so the improvements in the present study may reflect these changes. Direct comparisons between old and new ActiGraph monitors may be needed to more clearly determine whether the new features have contributed to improvements in accuracy.

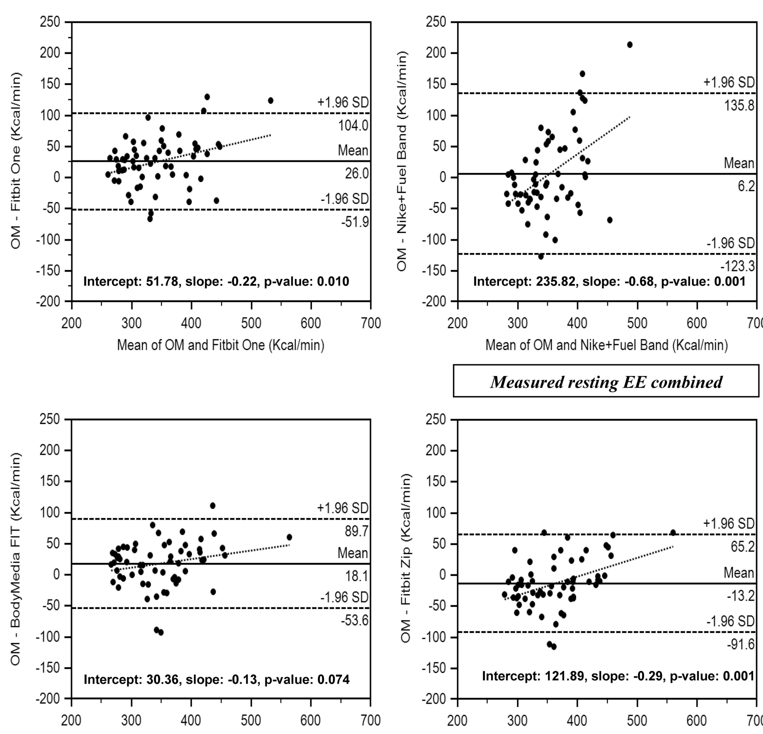
Overall, the performance of these consumer-based monitors is quite impressive, as most had MAPE values between 10% and 15%. The performance is especially noteworthy, considering the diverse range of activities tested in the study.

APPLIED SCIENCES



Regression based limits of agreement for differences in energy expenditure determined by the activity monitors and the Oxycon Mobile.

*Abbreviations: OM, Oxycon Mobile; EE, energy expenditure.



Regression based limits of agreement for differences in energy expenditure determined by the activity monitors and the Oxycon Mobile.

*Abbreviations: OM, Oxycon Mobile; EE, energy expenditure.

FIGURE 3—Bland–Altman plots using measured REE.

The protocol was designed to include typical activities that would be reflective of normal adult behavior, but accelerometry-based monitors generally have a hard time capturing activities (i.e., upper body movement, cycling activity, and weight-bearing activity). It is possible that the monitors overestimated some activities and underestimated others. However, the overall estimates were reasonable, considering the inherent challenges of assessing PA. Previous research (2,22) has consistently demonstrated higher correlations with O_2 (r = 0.85 – 0.93) under laboratory conditions in contrast to lower correlations under free-living conditions (r = 0.48 – 0.59). The correlations in the present study ranged from 0.13 to 0.84 and were generally consistent with the values reported with other research-based monitors. The reason free-living activities are more difficult to assess is because daily activities include a considerable amount of upper body movements that may not be captured by a monitor (i.e., weight lift, gardening, and vacuuming). In addition, the equations developed for traditional accelerometers (i.e., ActiGraph and ActiCal) have typically used treadmill equations that have been shown to underestimate EE with an estimated range from 31% to 67% lower than measured values (12,21). Additional research is clearly warranted to compare results with other research monitors, with different activities, sample populations, and criterion measures.

The results of this study add to the existing literature on accelerometry-based activity monitors and also provide new insights about these seven consumer-based monitors. Previous research (13,16) has demonstrated clear limitations using standard accelerometry-based activity monitors for assessing EE under free-living conditions. The limitation would seem challenging to overcome because there is no single regression equation that can be used in accelerometry-based devices to adequately capture the EE cost for all activities. However, the tendency for "reasonable" accuracy for many of the monitors suggests that the monitors may be using more robust pattern recognition approaches than previously appreciated. As described, the BMF is based on the existing pattern recognition algorithms used in the established SenseWear monitor. However, the comparable performance of some of the other accelerometry-based monitors suggests that these other devices may be using similar (or analogous) machine learning techniques that enable classification of underlying activity patterns. A previous study documented that pattern recognition techniques improved the overall EE estimate of the ActiGraph by up to 1.19 METs compared with the Freedson regression equation (19).

A challenge when interpreting the present results is that there were some seemingly discrepant findings in the outcomes. The Nike+ Fuel, for example, was found to produce accurate group-level estimates (based on the equivalency test), but low correlations were observed between the Nike+ Fuel and the OM. It is hard to reconcile how the monitor can produce accurate group-level estimates and still have low correlations. Therefore, caution should be used when interpreting these findings with the Nike+ Fuel. The BMF had consistently strong outcomes as well as strong correlations so stronger confidence can be placed in these outcomes and the validity of the monitor.

The monitors tested in the present study were not marketed as research-grade monitors, but the present study generally supports the relative utility (and accuracy) of the various monitoring technologies. It is unreasonable to expect consumer-based monitors to match the utility of other research-based devices because they are developed for different purposes and with different constraints (e.g., ease of use and keeping costs low). However, it is important for researchers, fitness professionals, and consumers to at least be aware of the relative accuracy of the various monitors so that it can be factored into decisions when selecting devices. The popularity of these devices with consumers will likely lead to increased use (and new research possibilities), so it is important to continue evaluating different aspects of these tools.

A key question in this regard is the relative utility of these devices for promoting PA behavior. Consumer-based monitors are developed primarily to facilitate self-monitoring and behavior change so features such as comfort, convenience, and functionality may ultimately be more important to consumers. To date, little work has been performed on usability or effects on changing behavior. A study using the PAM monitor (18) demonstrated significant increases in moderate PA in youth after a 3-month intervention (411 $min \cdot wk^{-1}$, 95% CI = 1–824, P = 0.04). In boys, the intervention groups showed a relative reduction in sedentary time compared with the control group (-1801 $min \cdot wk^{-1}$, 95% CI = -3545 to -57, P = 0.04). This applied intervention study demonstrates that consumer monitors, such as the PAM, may have utility for promoting PA behavior. It was not possible to systematically evaluate the features of the various devices in the present study, but the various monitoring technologies and Web sites all proved to be useable and intuitive. Additional research is clearly needed to evaluate the relative utility of these consumer-based devices for motivating adults to be more physically active.

The study provided new insights about these monitors, but it does have some limitations. The sample population included only healthy, young individuals within the normal range of body weight and body fat. Therefore, we cannot generalize these findings to other age groups or body sizes. In this study, we also did not assess the reliability of the activity monitors. Poor reliability can negatively impact validity, but solid-state construction has dramatically improved the reliability of most commercially available monitors. In terms of equivalent testing, no agreement has been made on acceptable ranges of the equivalence zone. In this study, ±10% of the mean of the OM was used as a lower/upper boundary of the equivalence zone. However, more supportive research is needed to create agreed-upon consensus on an equivalence zone.

In conclusion, the present study supports the validity of the more established BMF platform while also providing preliminary support for the FZ. Results with the NFB must be viewed with caution because of the somewhat discrepant findings (i.e., good agreement but low correlations and proportional systemic bias). Taken collectively, the results of the study demonstrate good potential for almost all of the models because the results were generally similar to, if not better than, the results from the established ActiGraph monitor in terms of measuring EE. An advantage of this new line of consumer-based activity monitors is that they offer additional online feedback and are less obtrusive than standard research-grade devices. The monitors also provide goal setting features, tracking tools, and other applications (e.g., social networking links) that provide additional value to consumers and potentially for behaviorally focused research applications.

This study was supported by the Department of Kinesiology at Iowa State University as Pease Family Doctoral Research Award. None of the authors have a professional relationship with companies or manufacturers who might benefit from the results of the present study.

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

REFERENCES

1. Bassett DR, Ainsworth BE, Swartz AM, Strath SJ, O'Brien WL, King GA. Validity of four motion sensors in measuring moderate intensity physical activity. *Med Sci Sports Exerc*. 2000;32(9 Suppl): S471–80.
2. Bassett DR Jr. Validity and reliability issues in objective monitoring of physical activity. *Res Q Exerc Sport*. 2000;71(2 Suppl):S30–6.
3. Bonomi AG, Plasqui G, Goris AH, Westerterp KR. Estimation of free-living energy expenditure using a novel activity monitor designed to minimize obtrusiveness. *Obesity (Silver Spring)*. 2010; 18(9):1845–51.
4. Calabro MA, Stewart JM, Welk GJ. Validation of pattern-recognition monitors in children using doubly labeled water. *Med Sci Sports Exerc*. 2013;45(7):1313–22.
5. Calabro MA, Welk GJ, Eisenmann JC. Validation of the SenseWear Pro Armband algorithms in children. *Med Sci Sports Exercise*. 2009;41(9):1714–20.
6. Compher C, Frankenfield D, Keim N, Roth-Yousey L. Best practice methods to apply to measurement of resting metabolic rate in adults: a systematic review. *J Am Diet Assoc*. 2006;106(6): 881–903.
7. Crouter SE, Kuffel E, Haas JD, Frongillo EA, Bassett DR Jr. Refined two-regression model for the ActiGraph accelerometer. *Med Sci Sports Exerc*. 2010;42(5):1029–37.
8. Dannecker KL, Sazonova NA, Melanson EL, Sazonov ES, Browning RC. A comparison of energy expenditure estimation of several physical activity monitors. *Med Sci Sports Exerc*. 2013; 45(11):2105–12.
9. Dixon PM, Pechmann JHK. A statistical test to show negligible trend. *Ecology*. 2005;86(7):1751–6.
10. Drenowatz C, Eisenmann JC. Validation of the SenseWear Armband at high intensity exercise. *Eur J Appl Physiol*. 2011;111(5): 883–7.
11. Hauck WW, Anderson S. A new statistical procedure for testing equivalence in two-group comparative bioavailability trials. *J Pharmacokinet Biopharm*. 1984;12(1):83–91.
12. Hendelman D, Miller K, Baggett C, Debold E, Freedson P. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Med Sci Sports Exerc*. 2000;32(9 Suppl):S442–9.
13. Hendrick P, Boyd T, Low O, et al. Construct validity of RT3 accelerometer: a comparison of level-ground and treadmill walking at self-selected speeds. *J Rehabil Res Dev*. 2010;47(2):157–68.
14. Johannsen DL, Calabro MA, Stewart J, Franke W, Rood JC, Welk GJ. Accuracy of armband monitors for measuring daily energy expenditure in healthy adults. *Med Sci Sports Exer*. 2010;42(11): 2134–40.
15. John D, Tyo B, Bassett DR. Comparison of four ActiGraph accelerometers during walking and running. *Med Sci Sports Exerc*. 2010;42(2):368–74.
16. Lee KY, Macfarlane DJ, Cerin E. Comparison of three models of ActiGraph accelerometers during free living and controlled laboratory conditions. *Eur J Sport Sci*. 2013;13(3):332–9.
17. Rosdahl H, Gullstrand L, Salier-Eriksson J, Johansson P, Schantz P. Evaluation of the Oxycon Mobile metabolic system against the Douglas bag method. *Eur J Appl Physiol*. 2010;109(2):159–71.
18. Slootmaker SM, Chinapaw MJ, Seidell JC, van Mechelen W, Schuit AJ. Accelerometers and Internet for physical activity promotion in youth? Feasibility and effectiveness of a minimal intervention. *Prev Med*. 2010;51(1):31–6.
19. Staudenmayer J, Pober D, Crouter S, Bassett D, Freedson P. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *J Appl Physiol*. 2009;107(4):1300–7.
20. Trost SG, Way R, Okely AD. Predictive validity of three ActiGraph energy expenditure equations for children. *Med Sci Sports Exerc*. 2006;38(2):380–7.
21. Welk GJ, Almeida J, Morss G. Laboratory calibration and validation of the Biotrainer and Actitrac activity monitors. *Med Sci Sports Exerc*. 2003;35(6):1057–64.
22. Welk GJ, Blair SN, Wood K, Jones S, Thompson RW. A comparative evaluation of three accelerometry-based physical activity monitors. *Med Sci Sports Exerc*. 2000;32(9 Suppl):S489–97.
23. Welk GJ, Schaben JA, Morrow JR Jr. Reliability of accelerometry-based activity monitors: a generalizability study. *Med Sci Sports Exerc*. 2004;36(9):1637–45.
24. Wellek S. *Testing Statistical Hypotheses of Equivalence*. Boca Raton (FL): Chapman & Hall/CRC; 2003, xv, 284 p.