Calibration of Accelerometer Output for Adults

CHARLES E. MATTHEWS

Department of Medicine, Division of General Internal Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN

ABSTRACT

MATTHEWS, C. E. Calibration of Accelerometer Output for Adults. *Med. Sci. Sports Exerc.*, Vol. 37, No. 11(Suppl), pp. S512–S522, 2005. This paper reviews the collective experience of monitor calibration studies in adults and seeks to answer the following questions: What has been done? What have we learned? What could be done to further enhance the comparability of results from future calibration research? Calibration studies in adults have typically used oxygen consumption as a criterion measure, similar types of source activities, and linear regression to obtain prediction equations that calibrate the activity counts to measured activity intensity levels. However, the methodological diversity of these studies has produced a great deal of variation in the resulting prediction equations and cut points, even when using the same monitor. Thus, data obtained from a relatively robust activity monitoring technology that captures many dynamic physical activities reasonably well have been splintered by the calibration process into a wide range of summary measures that are much less comparable than they could otherwise be. This heterogeneity in calibration results reduces our ability to interpret data obtained from accelerometers between different research groups, across the life span, between populations, and probably between the different monitor types. This report reviews and critiques methods typically used for developing calibration equations and determining activity count cut points for identifying specific intensities of PA among adults, and it highlights the need for flexible research methods that can enhance the comparability of results from future calibration studies. **Key Words:** PHYSICAL ACTIVITY, EXERCISE, AMBULATORY MONITORING, BEHAVIOR, ENERGY EXPENDITURE

Since their introduction as an objective measure of free-living physical activity (PA) in the early 1980s (10), waist-mounted accelerometers have become a staple of the PA assessment repertoire. They have been used extensively in the validation of self-reported PA surveys, as outcome measures in intervention studies, and in research designed to identify the psychosocial and environmental correlates of PA behaviors. More recently, they have been used in population surveillance of PA patterns in the National Health and Nutrition Examination Survey (NHANES). The main advantage of objective PA measures is that they overcome some of the inherent limitations in self-report methods designed to capture information about the salient characteristics of these behaviors, including the type and purpose, frequency, intensity, and duration of the behaviors.

In contrast to self-report, using motion generated by PA as the critical signal, accelerometers capture a standardized description of most, but not all, of these characteristics. Accelerometers alone cannot provide contextual information about the type or purpose of specific activities and they are limited in their abilities to provide optimal information for certain activities, such as cycling or weight lifting. However, monitor data are free from random and systematic errors introduced by respondents and interviewers, and real-time data collection and automated data reduction provide a rich description of the PA profile of free-living humans. Recent reports indicating that objective measures provide greater precision in comparison to self-report (22,41) suggests that accelerometers have the potential to substantially reduce sample sizes in PA studies while retaining statistical power. Thus, once the logistical hurdles of monitor delivery and retrieval are overcome (see Trost et al. (36)), accelerometers offer a relatively simple, precise, and efficient method of PA assessment suitable for small clinical studies and epidemiological studies of intermediate size (e.g., $N <$ 5000).

The first generation of accelerometers recorded integrated acceleration data, saving this information in the form of an activity count (ct) that reflects the duration and intensity of activity in a given sampling interval (e.g., 1 min). In order to translate the raw activity counts data into more physiologically meaningful information, some type of calibration is required. The most common method for conducting calibration studies has been to compare activity counts and measured oxygen consumption during specific activities selected to mimic key activities of daily living. These studies have typically used dynamic activities such as walking and running to develop prediction equations because the relationship between body acceleration and PA energy expenditure (PAEE) from dynamic activities is fairly linear over much of the PAEE range (11,28). More recent studies have focused on activities that are more generalizable to the full range of activities encountered in daily life and have

Address for correspondence: Charles E. Matthews, Ph.D., Division of General Internal Medicine, Center for Health Services Research, Medical Center East, Suite 6100, Nashville, TN 37232-8300; E-mail: charles.matthews@Vanderbilt.edu.

0195-9131/05/3711(Suppl)-S512/0
MEDICINE & SCIENCE IN SPORTS & EXERCISE®
Copyright © 2005 by the American College of Sports Medicine
DOI: 10.1249/01.mss.0000185659.11982.3d

Copyright © American College of Sports Medicine. Unauthorized reproduction of this article is prohibited.

employed activities that are both dynamic and static in nature, as well as activities composed of mixed dynamic and static elements (1,4,14,34). In contrast to purely dynamic activities, static activities are those where the relationship between body acceleration and PAEE are not tightly coupled. Nearly all published calibration studies in adults have used oxygen consumption as an outcome measurement, similar types of source activities, and linear regression to obtain prediction equations that calibrate the activity counts to measured PAEE levels and to calculate intensity specific count cut points. However, use of a wide range of activity types and intensities in these studies has produced a great deal of variation in the resulting prediction equations and cut points, even when using the same monitor. In perhaps the most striking case of methodological dissonance, proposed moderate-intensity cut points for the ActiGraph differ by as much as 10-fold across studies in adults (14,28). The heterogeneity within these studies suggests that there is also probably substantial variation in calibration results between the different types of monitors. The name ActiGraph is employed to refer to the device formerly known as Computer Sciences and Application (CSA) and Manufacturing Technology, Inc (MTI) (ActiGraph, LLC, Fort Walton Beach, FL).

A vigorous research effort over the last decade has greatly advanced our understanding of the measurement properties of accelerometers for capturing free-living PA. However, an unanticipated result of this effort has been that substantive differences in the calibration equations and cut points within, and probably between, monitors have emerged. In effect, data obtained from a relatively robust technology that captures many dynamic physical activities reasonably well (e.g., walking and running), have been splintered by the calibration process into a wide range of summary measures that are much less comparable than they could, or perhaps, should be. Lack of consistency in calibration methodology diminishes our ability to interpret data obtained from accelerometers across the lifespan, between populations, between different research groups, and between the different monitor types.

In this context, this paper reviews the collective experience of monitor calibration studies in adults and answers the following questions: What has been done? What have we learned? What could be done to enhance the comparability of results obtained from future calibration studies? In so doing, this report focusing only on adults reviews and critiques methods typically used for developing calibration equations and determining activity count cut points for specific intensities of PA. It also highlights the need for flexible research methods that can enhance the comparability of results obtained from future calibration studies, and ultimately the free-living data captured in the field. It should be noted that this report does not review the technological and economic particulars of accelerometers that are currently on the market because this information is available in great detail in a number of excellent reviews (3,12,37,39) and in other papers in this supplement (9). Additionally, investigations selected for presentation in this report are not intended to reflect a comprehensive review of the now substantial literature on ambulatory monitoring. Rather, they have been selected to describe particular issues of importance in calibrating monitor output. Many excellent papers have not been included in this report, and this in no way diminishes their quality or contribution to the field.

WHAT HAS BEEN DONE?

Early calibration studies: the Caltrac. Perhaps it is not surprising that investigators who had devoted substantial time and energy to developing systematic methods for collecting self-reported PA information saw the need for more objective measurements of activity and worked to produce such methods (17,24,25). Montoye et al. (27,31), after making the observations that a) the integrated mechanical forces applied to force platform and the energy cost of the activity were strongly correlated and b) that bodily acceleration during dynamic activities measured at the waist mirrored the mechanical force applied to a force platform, were among the first to develop an accelerometer-based activity monitor. In making the case that accelerometer-based devices could provide useful measures of free-living PA, the Montoye group carefully outlined a number of issues that had to be addressed for this new method to be valid. These issues are still relevant today.

First, accelerations measured at the waist during dynamic activity must be strongly correlated with the energy cost of the activity. Second, energy expenditure derived from dynamic physical activities (e.g., walking, running) is a major source of overall PAEE. Third, information about activities that cannot be captured by the monitor (e.g., swimming), and activities with a substantial static component (e.g., weight lifting), could be captured by other means. They acknowledged that many common daily activities result in muscular contractions that expend energy, but are not associated with much physical motion (e.g., household chores). Thus, the static component of the activity signal in these complex movement patterns will not be adequately captured by a single monitor worn at the waist.

Under ideal laboratory conditions, the Montoye group quantified a strong linear relationship between activity counts and oxygen consumption ($r = 0.74$) across a wide range of activity levels (e.g., slow walking and running), but also reported a moderate amount of individual variation in this relationship (of about 6.6 $mL{\cdot}kg^{-1}{\cdot}min^{-1}$ or nearly 2 METs) (27). These results suggested that the device had the ability to differentiate between dynamic activities of lower and higher intensity, but in retrospect, perhaps less ability to differentiate between low and moderate intensity activities (e.g., 2 vs 4 METs), particularly for activities made up of complex movement patterns.

Predicting energy expenditure and identifying cut points. Following the development and production of the Caltrac (Caltrac Personal Activity Computer (Muscle Dynamics, Torrence, CA)) (27), a multitude of accelerometer-based instruments entered the market, such as the RT3 Triaxial Research Tracker (formally known as Tritrac-R3D

Copyright © American College of Sports Medicine. Unauthorized reproduction of this article is prohibited.

TABLE 1. Description of existing ActiGraph prediction equations and cut points.

Reference	Source Activities				Prediction Equation					Cut Points		
	Sit Rest	Walk	Run	Mixed Dynamic-Static	Units	Intercept	Slope	R^2	SEE	Inactive/Light	Moderate	Vigorous
Freedson (11) $N = 50$; men/women (avg ~24 yr)		2	1		METs	1.439	0.000795	0.82	1.12	0–1951	1952	5725
Nichols (28) $N = 30$; men/women (18–35 yr)		2	1		METs[a]	1.731	0.0007271	0.89	1.06	0–1576[b]	3285[b]	5677[b]
Yngve (42)		2	1		METs	0.751	0.0008198	0.86	1.10	0–2742	2743	6403
$N = 28$; men/women (avg ~23 yr) (see equation code below)[c]					METs	1.136	0.0008249	0.85	1.14	0–2259	2260	5896
					METs	1.004	0.0007587	0.89	0.96	0–2630	2631	6585
					METs	1.762	0.0007371	0.86	1.09	0–1679	1680	5750
Brage (7) $N = 12$; men (23–30 yr)		2	2		METs[a,d]	2.886	0.0007429 -0.02 ($\dot{V}O_2$)	0.89	0.91	0–1809	1810	5850
Hendelman (14)		4			METs	1.602	0.000638	0.59	0.87	0–2190	2191	6893
$N = 25$; men/women (30–50 yr)		4		6	METs	2.922	0.000409	0.35	0.96	0–190	191	7526
Swartz (34) N ~10/activity; men/women (19–74 yr)		4		24	METs	2.606	0.0006863	0.32	1.16	0–573	574	4945
Leenders (18) $N = 28$, men/women (avg ~24 yr)		5			METs	2.240	0.0006	0.74	0.53	0–1266	1267	6252

Inactive/light (1.0–2.9 METs); moderate (3.0–6.0 METs); vigorous (≥ 6.1 METs).
[a] METs calculated from original units by dividing by 3.5 mL·kg^{-1}·min^{-1}.
[b] Intensity definitions employed by Nichols; light (2.0–3.9 METs); moderate (4.0–6.9 METs); vigorous (≥ 7.0 METs); cut points are from field-based study.
[c] Yngve equation code: a. Hip − track; b. Hip − treadmill; c. Back − track; d. Back − treadmill.
[d] Brage cut points estimated using this equation and the mean fitness level of participants in study (61.6 mL·kg^{-1}·min^{-1}).

(StayHealthy, Inc., Monrovia, CA)), Actical (Mini Mitter Co., Inc., Bend, OR), and the BioTrainer (IM Systems, Baltimore, MD), to name but a few devices that are commercially available. Most of these technologies provided individual researchers access to the raw activity count data, which allowed them to conduct independent calibration studies. Because the objective of this review is to evaluate the collective experience of this work, it focuses primarily on commercially available monitors for which multiple calibration studies have been published. Although a number of informative and high-quality calibration studies have been conducted for the three-dimensional Triaxial Accelerometer for Movement Registration (Tracmor) (Philips, Eindhoven, The Netherlands) (5,6), these studies have not been included here because this device is not commercially available at this time. Similarly, at the time of this writing, there are few published calibration studies among adults for the Actical device.

Multiple reports on the ActiGraph (7,11,14,18,28,34,42) and the Tritrac-RT3 (10,14,18,29) have published prediction equations and/or count cut points to describe activity intensity from counts. A systematic review this group of studies may be informative for identifying strengths and weaknesses of our current calibration methods. The earliest work on these devices primarily focused on dynamic activities (e.g., stepping, walking, running) (10,11,29), whereas a series of studies conducted in the late 1990s carefully examined common activities of daily living that involved a mixture of dynamic and static elements (e.g., household chores, lawn and garden activities) (4,14,26,34,38). Tables 1 and 2 present additional descriptive information about the type and number of the source activities employed by these studies in the development of equations and subsequent count cut points. In order to enhance comparability between equations that were not developed using MET values as the dependent variable (e.g., mL·kg^{-1}·min^{-1}) (7,28,29), the

TABLE 2. Description of Tritrac-RT3 prediction equations and cut points.

Reference	Source Activities				Prediction Equation[a]					Cut Points[a]		
	Sit Rest	Walk	Run	Mixed Dynamic-Static	Units	Intercept	Slope	R^2	SEE	Inactive/Light	Moderate	Vigorous
Nichols (29) $N = 20$; men/women (18–35 yr)		2	1		METs[b]	0.793	0.000123	0.90	0.59	0–650/651–1771[c]	1772[c]	3455[c]
Hendelman (14)		4			METs	1.077	0.000187	0.78	0.62	0–1027	1028	2633
$N = 25$; men/women (30–50 yr)		4		6	METs	2.817	0.000111	0.39	0.93	0–166	167	2904
Leenders (18) $N = 28$, men/women (avg ~24 yr)		5			METs	1.514	0.00149	0.81	0.46	0–997	998	3013

Inactive/light (1.0–2.9 METs); moderate (3.0–6.0 METs); vigorous (≥ 6.1 METs).
[a] All prediction equations and cut points are based on the vector magnitude summary data obtained from the Tritrac-RT3.
[b] METs calculated from original units of kilocalones per kilogram per minute by multiplying by body mass (kg) and then dividing by resting energy expenditure estimated using the Harris–Benedict equation (13).
[c] Intensity definitions employed by Nichols; light (2.0–3.9 METs); moderate (4.0–6.9 METs); vigorous (≥ 7.0 METs).

Copyright © American College of Sports Medicine. Unauthorized reproduction of this article is prohibited.

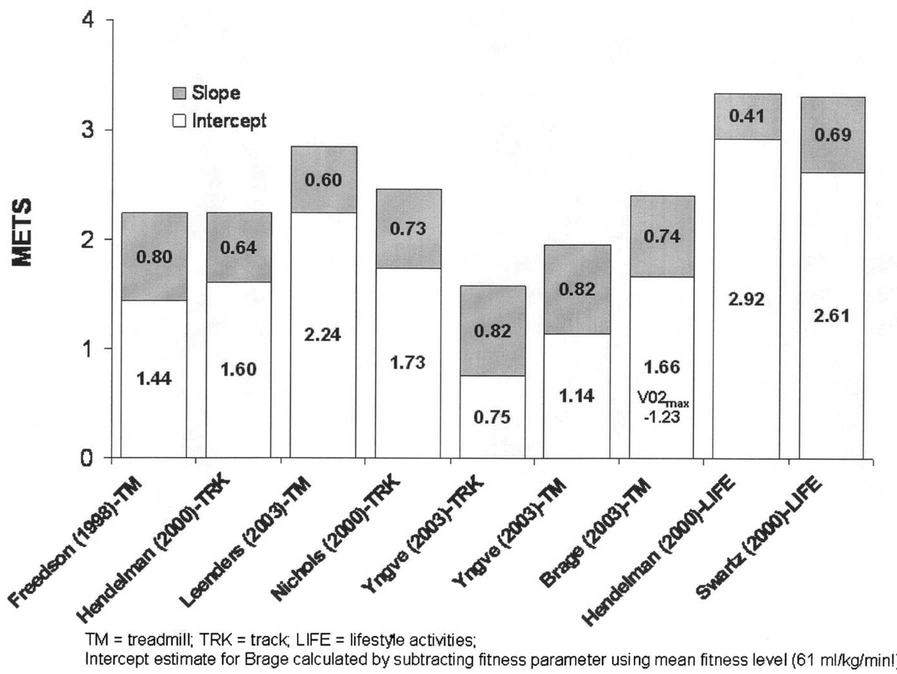

FIGURE 1—ActiGraph prediction equation MET values overall and separately for intercept and slope terms at 1000 ct·min^{-1}.

original regression equation parameters were transformed to METs in this tabular presentation.

Four of the ActiGraph studies examined both walking and running, and all these studies reported high levels of shared variance between oxygen consumption and activity counts (R^2 = 0.82 to 0.89). The SE of these regression models (SEE) were about 1 MET (Table 1). Two studies examined only walking (14,18) and these studies reported slightly lower R^2 values (R^2 = 0.59 and 0.74), and both had SEE below 1 MET. The studies of Swartz et al. (34) and Hendelman et al. (14) focused only on light- and moderate-intensity activities, particularly common lifestyle activities that were composed of activities with mixed dynamic and static elements. The R^2 values in these studies were lower (R^2 = 0.32 to 0.35), but SEE were also about 1 MET.

Moderate-intensity cut points (3–6 METs) obtained from six walking and running equations (7,11,42) for the ActiGraph ranged between 1680 and 2743 ct·min^{-1}, and the median was about 2100 ct·min^{-1}. In contrast, moderate-intensity cut points calculated from equations derived from a broad range of light to moderate lifestyle-oriented activities were much lower (191 and 574 ct·min^{-1}). Among studies that used both moderate and vigorous source activities, and 6 METs as an minimal threshold (7,11,42), the range of vigorous cut points was 5725–6585 ct·min^{-1}, and the median was about 5900 ct·min^{-1}. Table 2 presents similar information available for the Tritrac-RT3, and although there were only three studies to evaluate, among studies that used 3 METs as a minimal moderate intensity (14,18), the moderate cut point for vector magnitude values was about 1000 units. Vigorous cut points for studies using 6 METs as a minimum threshold ranged between 2633 and 3013 vector magnitude units. Moderate cut points were substantially lower for the equation derived from broad range of source activities (14), as compared with the moderate cut points derived only from walking (18).

Unfortunately, none of these studies attempted to characterize the low end of the activity spectrum by separating physical inactivity from light activity, nor did they employ measured resting energy expenditure values in regression modeling that would have helped anchor the intercept term of the equation on the low end of energy expenditure. Additionally, all of the studies reporting METs used an *estimate* of resting expenditure (1 MET ≈ 3.5 mL·kg^{-1}·min^{-1}) in their calculation of *measured* MET values.

To get a more tangible indication of the differences in the available equations for the ActiGraph monitor as described in Table 1, MET values produced by the equations were obtained by inserting a fixed activity level of 1000 ct·min^{-1}. Figure 1 presents results for overall energy expenditure levels and the separate contributions of both the slope and intercept terms of the equations to this value. MET values derived from the intercept terms are fixed and represent the number of METs each equation awards as a basal count value (i.e., 0 ct·min^{-1}). The slope are fairly similar across the five equations developed using walking and running activities (0.73–0.82 METs·1000 ct^{-1}). Interestingly, the equations of Hendelman et al. (14) and Leenders et al. (18) had the lowest slope values (0.60 and 0.64 METs·1000 ct^{-1}), and these studies did not employ a running or vigorous source activity that may have exerted a positive influence on the slope of the equation.

The two lifestyle equations of Hendelman et al. (14) and Swartz et al. (34) had the largest intercept terms (both >2.5 METs) and tended to have smaller slope terms (0.41–0.69 METs·1000 ct^{-1}) (Fig. 2), perhaps because no vigorous activities were included in the development of the equations. The impact of elevated intercept terms on predicting overall energy expenditure in these two equations is that energy expenditure during physical inactivity (0 ct·min^{-1}) will be vastly overestimated. When such equations are used to estimate count cut

Copyright © American College of Sports Medicine. Unauthorized reproduction of this article is prohibited.

points that would be consistent with moderate intensity activity (i.e., 3.0 METs), elevated intercept terms would be expected to produce moderate intensity count thresholds that are too low because fewer counts are needed to reach 3 METs. In field-based studies, use of moderate cut points that are too low would be expected to result in an overestimation of time spent in moderate activity.

WHAT HAVE WE LEARNED?

The extensive research effort within these calibration studies and studies that have evaluated the validity of the resulting prediction equations has solidified our understanding of the strengths and limitations of accelerometers for capturing important elements of free-living PA. First, accelerometers can largely do what they were originally intended to do. That is, they can provide useful objective information about physical *inactivity* and dynamic physical activities, such as walking, and running. Figure 2 emphasizes this point by presenting mean ActiGraph count data for different walking and running speeds as reported in eight different studies (7,11,14,16,28,34,38,42). Although the observation that activity counts increase with ambulatory speed is in no way novel, the fact that data in this figure were derived from eight different studies (seven groups), roughly 230 men and women, both track- and treadmill-based studies, different decades, and different continents, underscores the robustness of accelerometers as an objective measure of dynamic PA. It was not possible to directly evaluate the ability of the monitors to capture physical inactivity, but based on our experience, sitting and working quietly (e.g., reading, typing) rarely results in activity counts above 250 $ct \cdot min^{-1}$ from this device. Because this activity count level is well below the roughly 1000 $ct \cdot min^{-1}$ monitor output measured during walking at 2 mph (Fig. 3), it seems probable that physical inactivity is an activity pattern that is measurable by accelerometers. Calibration studies in children and adolescents have begun to characterize physical inactivity using accelerometers (30,35), but studies in adults are needed to identify optimal cut points that may differentiate physical inactivity from light-intensity activities. Both of these activity patterns have been shown to be important determinants of PAEE (20,40).

The second major observation is that a single waist mounted accelerometer cannot do it all. It cannot capture certain categories of activity that are highly static in nature or that are derived from movement patterns that contain both dynamic and static elements. One must accept the usual list of activities that are unlikely to be ascertained by such a monitor (e.g., swimming), as well as activities that are insufficiently captured by the single devices mounted at the waist (e.g., cycling, weight lifting). Given the relatively low prevalence of many of these activities, it is possible to capture this information by self-report and adjust accelerometer output accordingly. Other activities that are done frequently (e.g., household chores, light occupational activity) are of primary interest because of their contribution to overall energy expenditure. However, they are difficult to estimate accurately, which makes simple correction of accelerometer data through self-report problematic.

The development of the triaxial accelerometer was initially undertaken in an effort to overcome the inability of uniaxial devices to capture more complex movement patterns (2,6). However, it does not appear that the additional information obtained from three-dimensional (3-D) movement completely remedies the problem of underascertainment of activities with complex movement patterns, at least in the first generation of 3-D devices with relatively simplistic use of the 3-D data (2,6). For example, Welk et al. (38) examined measured energy expenditure to that predicted by three different monitors: the ActiGraph (uniaxial), Tritrac-RT3 (triaxial), and BioTrainer (bidirectional) (Fig.

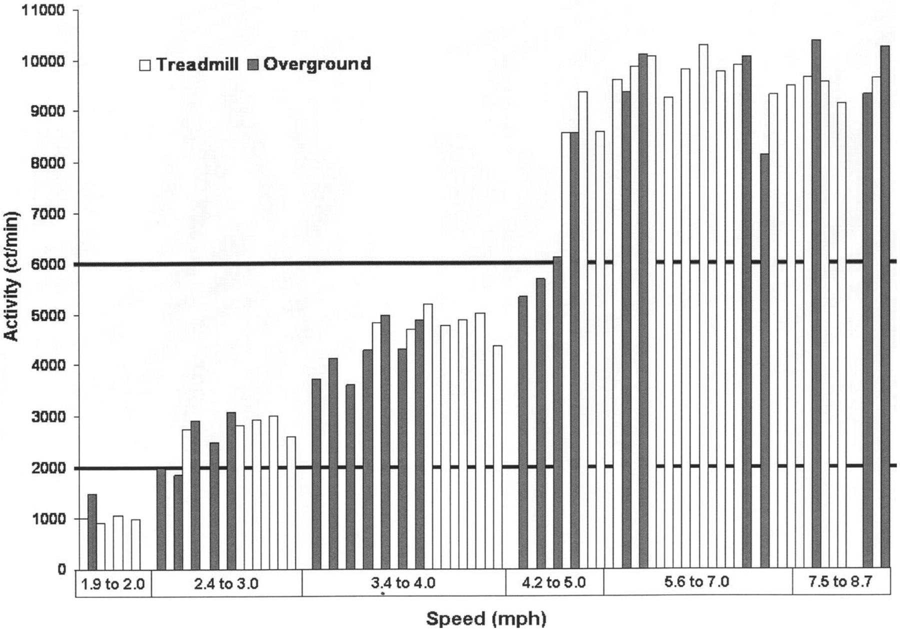

FIGURE 2—Mean ActiGraph counts by walking and running speed in eight studies.

Copyright © American College of Sports Medicine. Unauthorized reproduction of this article is prohibited.

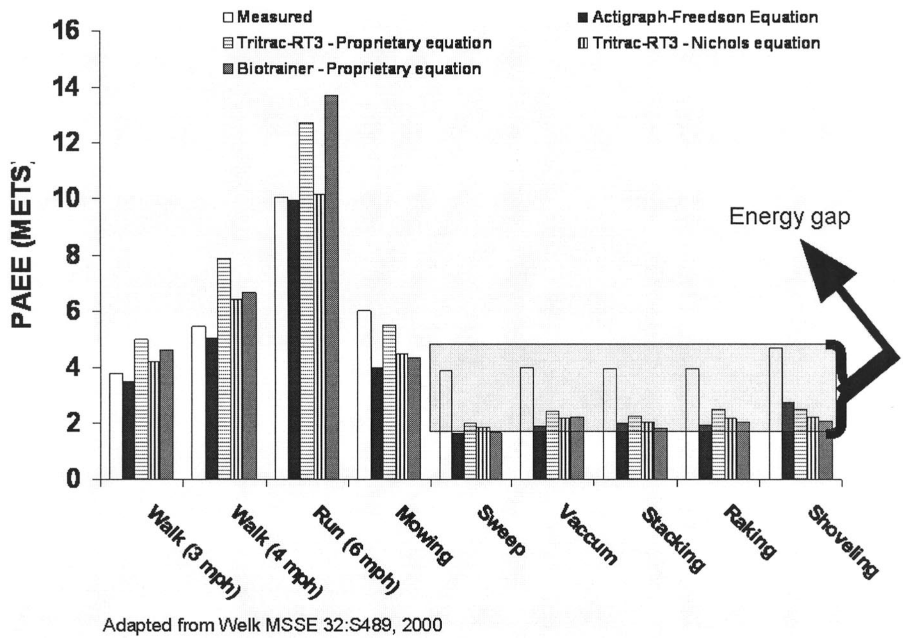

FIGURE 3—Measured vs predicted energy expenditure from different monitor technologies and prediction equations.

3). Expenditure estimates from all three units increased as expected during walking, running, and lawn mowing. However, all the monitors significantly underestimated the energy cost, by about 50%, of activities composed of more static and/or complex movement patterns. These data suggest that the assessment of multiple dimensions of movement do not provide enough extra information to fill the considerable "energy gap" experienced by these monitors (Fig. 3).

Although a single summary measure of the 3-D signal may not be sensitive enough to capture subtle differences in complex movement activities, only a few studies have attempted to extract more information from the 3-D signal. Chen and Sun (10) have suggested evidence that prediction of PAEE may be improved using more sophisticated modeling of the individual movement vectors from the 3-D signal (see also Chen and Bassett (9) in this supplement). Results of Campbell et al. (8) provide some support for such an approach. It is possible that more sophisticated modeling of the individual vectors (e.g., neural networks, pattern recognition) may provide additional and more accurate information that will allow improved PAEE prediction for the multidimensional compared to the uniaxial devices.

If researchers are primarily interested in capturing dynamic activities such as walking and running, even waist-mounted uniaxial monitors will capture relevant information. It would appear that current calibration approaches for the monitor technologies reviewed here will miss the static component of the activity signal, and this includes a portion of the signal in activities with more complex movement patterns. The magnitude of this gap in data obtained from the field will depend on the distribution of the dynamic and static elements of the activity signal in the population studied. Thus, an assumption initially highlighted by Montoye et al., that dynamic activities must represent a major source of overall PAEE for monitors to have utility, is still a critical issue today. Results from Strath et al. (33) in this supplement suggest that the magnitude of the underestimation in PAEE is roughly 13% in comparing a single waist-mounted ActiGraph to measured values. The challenge of the next generation of calibration studies for both first- and future-generation monitors that capture additional movement and or physiological signals (e.g., heat, HR, inclinometers) will be to fill in the gap in PAEE that is derived from static elements of activity behaviors.

Exploration of the heterogeneity in ActiGraph calibration studies. The ActiGraph device has been used repeatedly in calibration studies and six separate studies have produced 10 different prediction equations (Table 1). As noted previously, results from these studies are heterogeneous, particularly in contrasting studies of purely dynamic activities and studies that attempted to capture important lifestyle-oriented activities. In an effort to better understand this heterogeneity in results, we attempted to replicate these results in a simulation study.

Using data from Matthews et al. calibration of the Actillume monitor (Ambulatory Monitoring, Inc., Ardsley, NY) (23), we examined the influence of different combinations of source activities on the eventual linear regression equations obtained in a simple simulation exercise. These data were collected among healthy men (N = 7) and women (N = 12) aged 25–59 using a portable metabolic unit to measure oxygen consumption in a laboratory setting (23). The activities employed were sitting (e.g., reading, typing), box moving, stepping, and walking at 3.5, 4.25, and 5.0 km·h^{-1}. Values representing running were imputed by multiplying the available MET and activity count values at the fastest walking speed by a factor of 1.9. Figure 4A describes the idealized linear system derived from three types of source activities, sitting at rest, walking, and running.

Figure 4B represents the typical source data composed of walking and running used in the early ActiGraph calibration

Copyright © American College of Sports Medicine. Unauthorized reproduction of this article is prohibited.

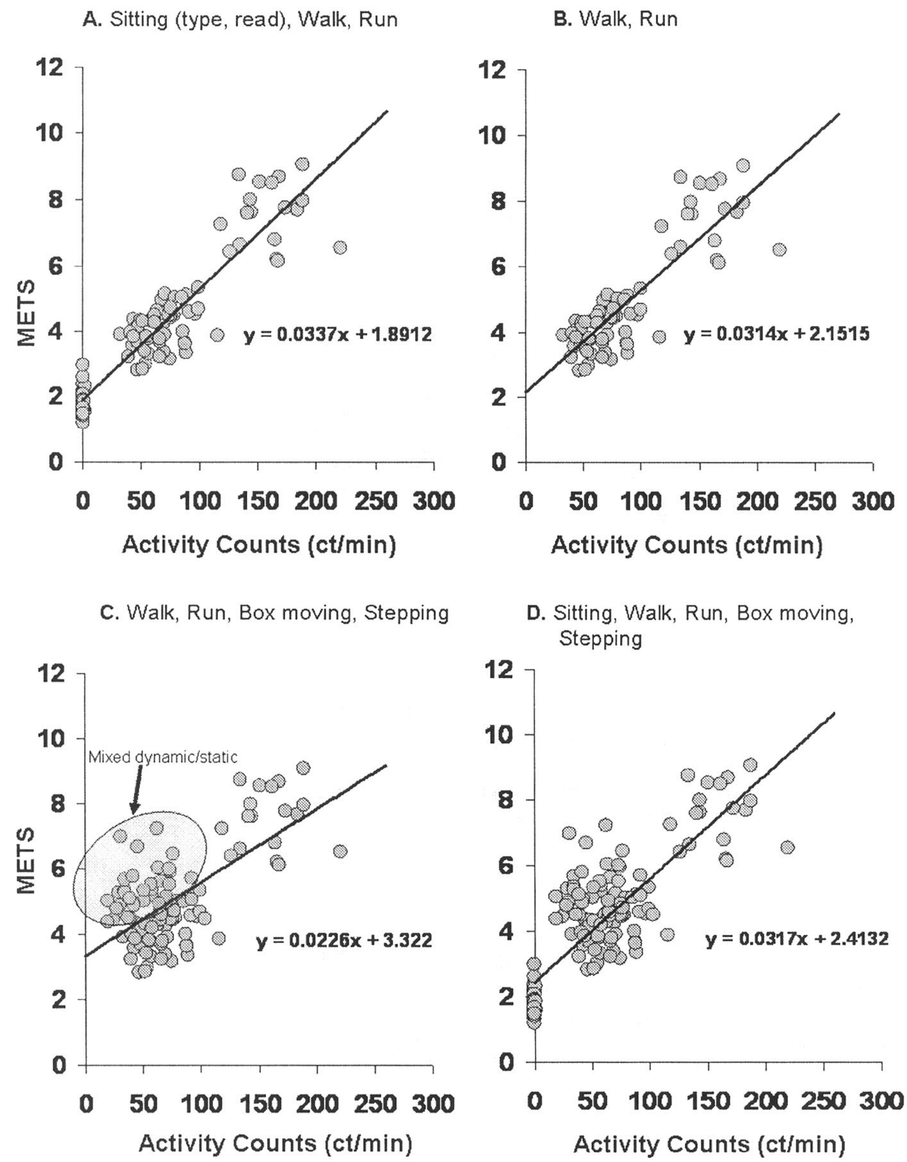

FIGURE 4—A–D: Influence of activity type and intensity on linear regression parameters.

studies. Comparing the resulting prediction equations between Fig. 4A and B indicates a modest influence on the equation when sitting values were employed in the development of the equation. When the resting values were included in the model, the value of the intercept term is slightly lower and the slope term is higher (Fig. 4A vs B). The addition of activities composed of more complex movement patterns to the walking and running data (Fig. 4C) results in the accumulation of data at lower count values for a given MET intensity compared to the walking and running data. When a linear model was fit to all these data (Fig. 4C), the equation obtained had an intercept term that was elevated and a slope that was depressed in comparison to the more strictly linear relationship described for the sitting, walking, and running data (Fig. 4A vs C). Adding back the sitting data in an effort to anchor the intercept term was successful in reducing it from 3.32 to 2.41 METs (Fig. 4C vs D). However, the resultant 2.41 MET value was still higher than that observed in the more idealized linear system described in Fig. 4A of 1.89 METs.

Not surprisingly, these simulations demonstrate that the source data used to estimate the linear regression equations have strong bearing on values of the parameters in these equations. Unfortunately, the equations with the most extreme results in this simulation exercise are representative of the same source data from which the current lifestyle-oriented prediction equations and subsequent moderate-intensity cut points were derived (14,34). The phenomenon that is evident in this simulation is also apparent in both the ActiGraph (Table 1) and Tritrac-RT3 equations and cut points (Table 2). In retrospect, it is clear that calibration methods used in the early studies that employed linear regression and a reasonably linear combination of activity (e.g., walking and running) may not be appropriate to use in the description of activities with more complex movement patterns. Thus, it appears that the type and intensity of activities chosen by investigators in these studies interacted with the central calibration tool selected to link PAEE with activity counts (linear regression) to produce the heterogeneity observed in these studies.

Copyright © American College of Sports Medicine. Unauthorized reproduction of this article is prohibited.

WHAT CAN BE DONE?

Future calibration studies among adults for the ActiGraph and Tritrac-RT3 devices as well as newer monitors and/or monitoring systems should address three major gaps that exist today. First, no single calibration study in adults for the ActiGraph and Tritrac-RT3 monitors has yet employed measurements of physical inactivity as well as measurements of light, moderate, and vigorous activity to determine activity count thresholds that describe time spent in each of these activity levels. Assessing the full spectrum of activity intensity is of considerable interest because lower intensity activities are increasingly believed to be important to overall PA energy expenditure (19,40,43). Motion sensors may be one of the few viable methods of assessing spontaneous activities because they occur more randomly during the day and they are probably poorly attended to cognitively, which limits opportunity for self-reported assessment.

Second, our understanding of the optimal methods for determining activity count thresholds for moderate activities that can capture both dynamic activities like walking, as well as common moderate-intensity lifestyle-oriented activities (e.g., household chores) still has clear gaps. The following example may be one potentially useful strategy for closing this gap. In a recent validation study of a short PA surveillance instrument (21), we embarked on a moderate-intensity cut point development strategy for the ActiGraph device that did not use linear regression as the central calibration tool but rather combined information from laboratory- and field-based studies to determine a moderate-intensity cut point that might capture the full range of moderate-intensity activities encountered in daily living. Briefly, using data collected by the laboratory of David Bassett at the University of Tennessee (4,32,34) and generously shared with our group, we examined percentile distributions (25th and 75th percentiles) of ActiGraph activity count values obtained during six common light-intensity activities (<2.9 METs) and six common moderate-intensity activities (2.9–3.6 METs). Using the average of the 75th percentile count values from the six light activities (826 ct·min^{-1}) to indicate an upper range of count values for light activities, and the average of the 25th percentile count values from the six moderate activities (558 ct·min^{-1}) to indicate a lower count range for moderate activities; we selected an intermediate value (690 ct·min^{-1}) as our candidate moderate count cut point. We also were fortunate to have access to data collected by Scott Strath (32) that were derived from direct field measurement of oxygen consumption over 6 h of relatively free-living PA in 10 young adults as well as data from interviewer-administered 24-h PA recalls in our target population (21). Using these two criterion measures, we cross-validated our candidate moderate-intensity cut point (i.e., 690 ct·min^{-1}) as well as two others selected to be approximately 10% above (760 ct·min^{-1}) and below (575 ct·min^{-1}) the candidate value. Thus, we compared estimates of time spent in moderate activity derived from the three cut points (575, 690, and 760 ct·min^{-1}) to the amount of time spent in free-living moderate-intensity activity as measured by two criterion measures. On average, the highest cut point of 760 ct·min^{-1} provided the most accurate group level estimate of time spent in moderate intensity activity in comparison to both criterion measures, although the individual variation around the group mean was considerable. Figure 5 shows the mean difference the between the cut points and time spent in measured free-living moderate intensity activity, which was about 60 min of the 6-h monitoring period.

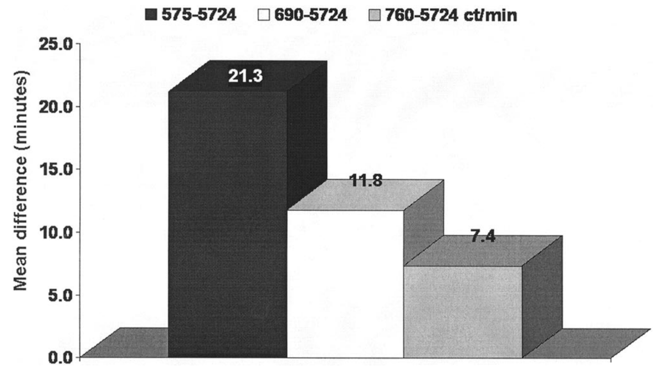

FIGURE 5—Field-based evaluation of candidate cut points.

Thus, our field-based cross-validation provided important information that allowed us to refine a cut point that was originally identified in the laboratory. We elected to use the 760 ct·min^{-1} cut point in our study. However, this count threshold is not perfect. Figure 6 presents the activity count distributions across the 28 light- and moderate-intensity activities from the Bassett lab and indicates that this threshold will inappropriately capture some light-intensity activities and miss some moderate-intensity activities for certain individuals. It is not likely that one cut point for a single waist-mounted accelerometer will be able to completely differentiate between all light- and moderate-intensity activities in free-living adults. However, these data suggest that this count threshold would capture the majority of time spent in the activities that were at or above 3.6 METs. All these activities were more dynamic in nature. These data suggest that a greater integration of lab- and field-based studies may be useful in future efforts to calibrate activity monitors to the actual activity patterns of free-living adults. More comprehensive research using a wider range of activities is needed to test and refine the cut point described here.

Finally, researchers and public health practitioners employ objective measures of activity in different projects to accomplish a variety of measurement objectives. These projects vary in the type of activity behavior(s) of greatest interest, the level of measurement error that is acceptable, and in the amount of burden to participants and the research team that is tolerable. Future calibration research should aid the development of a comprehensive repertoire of methods that fulfill a variety of research needs. For example, if researchers are interested only in measuring time spent walking in the context of an intervention study, calibration research could help them understand whether adding the cost and complexity of a combined motion and HR system or devices with multiple limb sensors would provide a better

Copyright © American College of Sports Medicine. Unauthorized reproduction of this article is prohibited.

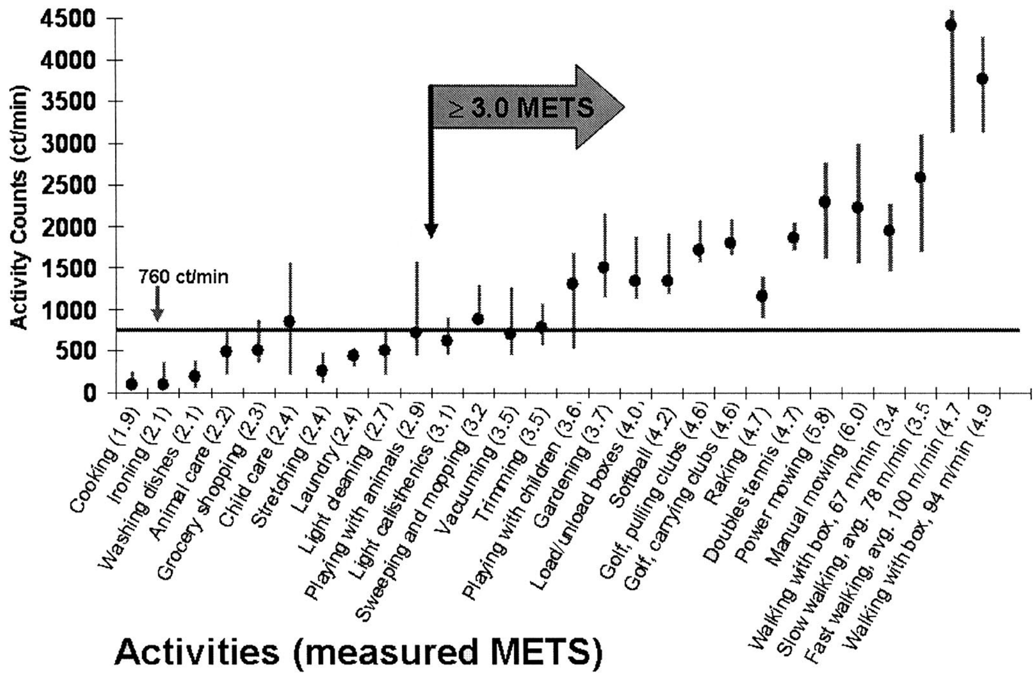

FIGURE 6—Moderate intensity cut point (760 ct·min^{-1}) across 28 light- to moderate-intensity activities (values are median (●) and 25th and 75th percentiles counts per minute).

estimate of walking time than a single waist-mounted accelerometer. While it now seems clear that using additional physiological and biomechanical signals to supplement whole-body acceleration will improve our estimates of PAEE (9,33) and the amount of time spent at different activity levels (20), the deluxe measurement systems may not always be necessary for certain measurement needs or feasible for certain types of studies.

SUMMARY AND FUTURE RESEARCH NEEDS

The first generation of waist-mounted accelerometers provide a robust method for describing major differences between physical inactivity and time spent in dynamic physical activities, such as walking and running. Most of the activity signal imparted by dynamic activities is captured by these devices, but the static elements of activities with complex movement patterns are likely to be underestimated. In reviewing available calibration studies for the ActiGraph and Tritrac–RT3 devices, we identified a number of high-quality studies that produced prediction equations and or activity count thresholds to describe moderate and vigorous activity levels. A comparison of these reports revealed considerable heterogeneity across studies that employed only dynamic activities (i.e., walking, running) and studies that employed walking and other lifestyle-oriented activities. Use of linear regression in these latter studies appears to have resulted in prediction equations and moderate-intensity cut points that may produce biased results for many common lifestyle activities. Alternate calibration approaches may be needed to identify optimal prediction equations and count thresholds for moderate-intensity lifestyle activities. A combination of laboratory- and field-based studies may enhance the appropriate calibration of accelerometer output to the underlying activity patterns in the target population. Future calibration studies should strive to provide results that can assist investigators assess the full spectrum of PA intensity. Objective monitoring methodologies are advancing rapidly to include acceleration and positional data for the torso and limbs (9,20), additional physiological signals associated with PA (e.g., HR and heat production) (15,33), in addition to whole-body movement. These new and more complex methodologies will probably increase the accuracy and precision of our measurements. However, they also will inevitably be more costly and are likely to increase the burden on participants who must put up with the devices for days at a time, and more complex devices may also increase the load on research staff who must collect and interpret these data. During the current period of evolution, it will be important to strive for the development of a knowledge base that fosters the use of a wide range of methods, both new and old, that are appropriate for a particular measurement need. A short list of future research needs follows:

- Develop prediction equations and activity count thresholds based on the full range of PA intensities, including physical inactivity (<2 METs), and light, moderate, and vigorous activities.
- Develop new approaches to determine optimal prediction equations and activity count thresholds for combinations of dynamic activities and other common activities with complex movement patterns.
- Seek to identify a range of valid objective monitoring methods for specific purposes that may vary in terms of cost and burden to participants and research staff.
- Describe the type and intensity of the most prevalent and frequently encountered physical activities in free-living humans in order to identify the most useful

Copyright © American College of Sports Medicine. Unauthorized reproduction of this article is prohibited.

source activities to employ in future calibration research.

- Delineate a core set of PA types and intensities for use in future calibration studies in order to facilitate comparability of calibration results across the lifespan and between different monitoring devices.

The results of the present study do not constitute endorsement by the author or ACSM of the products described in this paper. The author has no financial conflict of interest with any of the monitor manufacturers and has received no research funding from these companies.

This work was partially supported by the U.S. Centers for Disease Control and Prevention (CDC) under a cooperative agreement with the Prevention Research Center at the University of South Carolina, Columbia (Special Interest Project #13-2000, U48/CCU409664), as well as a Career Development Award from the National Cancer Institute (K07CA109363). The author thanks the following individuals: Patty Freedson, Russ Pate, and Barbara Ainsworth for their contributions in developing an alternate strategy for identifying a moderate-intensity cut point; Cara Hanby for her intellectual contribution to this work and her tireless administration of monitors to willing participants and for dealing with the occasionally messy data that are retrieved from them; David Bassett, Scott Strath, and Ann Swartz for generously sharing their data and insight; and Rick Troiano for useful feedback on an earlier version of this manuscript; and, finally, Anne Rodgers for her insightful editorial guidance.

REFERENCES

1. Ainsworth, B.E., D. R. Bassett, Jr., and S.J. Strath. et al. Comparison of three methods for measuring the time spent in physical activity. *Med. Sci. Sports Exerc.* 32:457–464, 2000.
2. Ayen, T. G., and H. J. Montoye. Estimation of energy expenditure with a simulated three-dimensional accelerometer. *J. Ambul. Monit.* 1:293–301, 1988.
3. Bassett, Jr. D. R. Validity and reliability issues in objective monitoring of physical activity. *Res. Q. Exerc. Sport* 71:S30–S36, 2000.
4. Bassett, Jr., D. R., B. E. Ainsworth, A. M. Swartz, S. J. Strath, W. L. O'Brien, and G. A. King. Validity of four motion sensors in measuring moderate intensity physical activity. *Med. Sci. Sports Exerc.* 32:471–480, 2000.
5. Bouten, C. V., K. T. Koekkoek, M. Verduin, R. Kodde, and J. D. Janssen. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Trans. Biomed. Eng.* 44:136–147, 1997.
6. Bouten, C. V., W. P. Verboeket-van de Venne, K. R. Westerterp, M. Verduin, and J. D. Janssen. Daily physical activity assessment: comparison between movement registration and doubly labeled water. *J. Appl. Physiol.* 81:1019–1026, 1996.
7. Brage, S., N. Wedderkopp, P. W. Franks, L. B. Andersen, and K. Froberg. Reexamination of validity and reliability of the CSA monitor in walking and running. *Med. Sci. Sports Exerc.* 35:1447–1454, 2003.
8. Campbell, K. L., P. R. Crocker, and D. C. McKenzie. Field evaluation of energy expenditure in women using Tritrac accelerometers. *Med. Sci. Sports Exerc.* 34:1667–1674, 2002.
9. Chen, K. Y., and D. R. Bassett. The technology of accelerometry-based activity monitors: current and future. *Med. Sci. Sports Exerc.* 37:S490–S500, 2005.
10. Chen, K. Y., and M. Sun. Improving energy expenditure estimation by using a triaxial accelerometer. *J. Appl. Physiol.* 83:2112–2122, 1997.
11. Freedson, P. S., E. Melanson, and J. Sirard. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med. Sci. Sports Exerc.* 30:777–781, 1998.
12. Freedson, P. S., and K. Miller. Objective monitoring of physical activity using motion sensors and heart rate. *Res. Q. Exerc. Sport* 71:S21–S29, 2000.
13. Harris, J. A., and F. G. Benedict. Biometric Studies of Basal Metabolism in Man.1919.
14. Hendelman, D., K. Miller, C. Baggett, E. Debold, and P. Freedson. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Med. Sci. Sports Exerc.* 32:442–449, 2000.
15. Jakicic, J. M., M. Marcus, K. I. Gallagher, et al. Evaluation of the SenseWear Pro Armband to assess energy expenditure during exercise. *Med. Sci. Sports Exerc.* 36:897–904, 2004.
16. King, G. A., N. Torres, C. Potter, T. J. Brooks, and K. J. Coleman. Comparison of activity monitors to estimate energy cost of treadmill exercise. *Med. Sci. Sports Exerc.* 36:1244–1251, 2004.
17. Laporte, R. E., H. J. Montoye, and C. J. Casperson. Assessment of physical activity in epidemiologic research: problems and prospects. *Public Health Rep.* 100:131–146, 1985.
18. Leenders, N. Y., T. E. Nelson, and W. M. Sherman. Ability of different physical activity monitors to detect movement during treadmill walking. *Int. J. Sports Med.* 24:43–50, 2003.
19. Levine, J. A., N. L. Eberhardt, and M. D.Jensen. Role of non-exercise activity thermogenesis in resistance to fat gain in humans. *Science* 283:212–214, 1999.
20. Levine, J. A., L. M. Lanningham-Foster, S. K. McCrady, et al. Interindividual variation in posture allocation: possible role in human obesity. *Science* 307:584–586, 2005.
21. Matthews, C. E., B. E. Ainsworth, C. L. Hanby, et al. Development and testing of a short physical activity recall questionnaire. *Med. Sci. Sports Exerc.* 37:986–994, 2005.
22. Matthews, C. E., B. E. Ainsworth, R. W. Thompson, and D. J. Bassett. Sources of variance in daily physical activity levels as measured by an accelerometer. *Med. Sci. Sports Exerc.* 34:1376–1381, 2002.
23. Matthews, C. E., J. R. Hebert, P. S. Freedson, E. J. Stanek, I. S. Ockene, and P. A. Merriam. Comparing physical activity assessment methods in the Seasonal Variation of Blood Cholesterol Levels Study. *Med. Sci. Sports Exerc.* 32:976–984, 2000.
24. Montoye, H. J. Estimation of habitual physical activity by questionnaire and interview. *Am. J. Clin. Nutr.* 24:1113–1118, 1971.
25. Montoye, H. J. Physical activity and health: an epidemiologic study of an entire community. In: *Physical Activity and Health: An Epidemiologic Study of an Entire Community*. Englewood Cliffs, NJ: Prentice Hall, 1975, pp. 13–28.
26. Montoye, H. J. Introduction: evaluation of some measurements of physical activity and energy expenditure. *Med. Sci. Sports Exerc.* 32:S439–S441, 2000.
27. Montoye, H. J., R. Washburn, S. Servais, A. Ertl, J. G. Webster, and F. J. Nagle. Estimation of energy expenditure by a portable accelerometer. *Med. Sci. Sports Exerc.* 15:403–407, 1983.
28. Nichols, J. F., C. G. Morgan, L. E. Chabot, J. F. Sallis, and K. J. Calfas. Assessment of physical activity with the Computer Science and Applications, Inc. accelerometer: Laboratory versus field validation. *Res. Q. Exerc. Sport* 71:36–43, 2000.
29. Nichols, J. F., C. G. Morgan, J. A. Sarkin, J. F. Sallis, and K. J. Calfas. Validity, reliability, and calibration of the Tritrac accelerometer as a measure of physical activity. *Med. Sci. Sports Exerc.* 31:908–912, 1999.
30. Puyau, M. R., A. L. Adolph, F. A. Vohra, and N. F. Butte. Validation and calibration of physical activity monitors in children. *Obes. Res.* 10:150–157, 2002.
31. Servais, S. B., J. G. Webster, and H. J. Montoye. Estimating Human Energy Expenditure Using an Accelerometer Device. *J. Clin. Eng.* 159–170,1984.
32. Strath, S. J., Bassett, Jr., D. R. D. L. Thompson, and A. M. Swartz. Validity of the simultaneous heart rate-motion sensor technique for measuring energy expenditure. *Med. Sci. Sports Exerc.* 34:888–894, 2002.
33. Strath, S. J., S. Brage, and U. Ekelund. Integration of physiological and accelerometer data to improve physical activity assessment. *Med. Sci. Sports Exerc.* 37:S563–S571.

Copyright © American College of Sports Medicine. Unauthorized reproduction of this article is prohibited.

34. Swartz, A. M., S. J. Strath, D. R. Bassett, Jr.,W. L. O'Brien G. A. King, and B. E. Ainsworth. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Med. Sci. Sports Exerc.* 32:450–456, 2000.
35. Treuth, M. S., K. Schmitz, D. J. Catellier, et al. Defining accelerometer thresholds for activity intensities in adolescent girls. *Med. Sci. Sports Exerc.* 36:1259–1266, 2004.
36. Trost, S. G., K. L. McIver, and R. R. Pate. Conducting accelerometer-based activity assessments in field-based research. *Med. Sci. Sports Exerc.* 37:S531–S543.
37. Welk, G. J. *Use of Accelerometry-Based Activity Monitors to Assess Physical Activity*. Champaign: Human Kinetics, 2002, pp. 125–141.
38. Welk, G. J., S. N. Blair, K. Wood, S. Jones, and R. W. Thompson. A comparative evaluation of three accelerometry-based physical activity monitors. *Med. Sci. Sports Exerc.* 32:489–497, 2000.
39. Westerterp, K. R. Physical activity assessment with accelerometers. *Int. J. Obes. Relat. Metab. Disord.* 23(suppl 3):S45–S49, 1999.
40. Westerterp, K. R. Pattern and intensity of physical activity. *Nature* 410:539–540, 2001.
41. Wong, M. Y., N. E. Day, J. A. Luan, K. P. Chan, and N. J. Wareham. The detection of gene-environment interaction for continuous traits: should we deal with measurement error by bigger studies or better measurement? *Int. J. Epidemiol.* 32:51–57, 2003.
42. Yngve, A., A. Nilsson, M. Sjostrom, and U. Ekelund. Effect of monitor placement and of activity setting on the MTI accelerometer output. *Med. Sci. Sports Exerc.* 35:320–326, 2003.
43. Zurlo, F., R. T. Ferraro, A. M. Fontvielle, R. Rising, C. Bogardus, and E. Ravussin. Spontaneous physical activity and obesity: cross-sectional and longitudinal studies in Pima Indians. *Am. J. Physiol.* 263:296–300, 1992.

Copyright © American College of Sports Medicine. Unauthorized reproduction of this article is prohibited.