Published in final edited form as:

Med Sci Sports Exerc. 2012 January ; 44(1 Suppl 1): S32–S38. doi:10.1249/MSS.0b013e3182399cf7.

CALIBRATION AND VALIDATION OF WEARABLE MONITORS

David R. Bassett Jr.¹, Ann V. Rowlands², and Stewart G. Trost³

¹University of Tennessee, Knoxville, TN

²University of Exeter, U.K

³Oregon State University, Corvallis, OR

Abstract

Background—Wearable monitors are increasingly being used to objectively monitor physical activity in research studies within the field of exercise science. Calibration and validation of these devices are vital to obtaining accurate data. This article is aimed primarily at the physical activity measurement specialist, although the end-user who is conducting studies with these devices also may benefit from knowing about this topic.

Best Practices—Initially, wearable physical activity monitors should undergo unit calibration to ensure inter-instrument reliability. The next step is to simultaneously collect both raw signal data (e.g., acceleration) from the wearable monitors and rates of energy expenditure, so that algorithms can be developed to convert the direct signals into energy expenditure. This process should use multiple wearable monitors, a large and diverse subject group, and should include a wide range of physical activities commonly performed in daily life (from sedentary to vigorous).

Future Directions—New methods of calibration now use “pattern recognition” approaches to train the algorithms on various activities, and they provide much better estimates of energy expenditure than were previously available with the single-regression approach. Once a method of predicting energy expenditure has been established, the next step is to examine its predictive accuracy by cross-validating it in other populations. In this paper, we attempt to summarize best practices for calibration and validation of wearable physical activity monitors. Finally, we conclude with some ideas for future research ideas that will move the field of physical activity measurement forward.

Keywords

accelerometer; heart rate; motion sensor; physical activity; energy expenditure; pattern recognition

INTRODUCTION

Wearable physical activity monitors are now widely used in research studies. These devices work by sensing either physiological or mechanical responses to bodily movement, and they use these signals to estimate variables that reflect physical activity. Calibration and validation of wearable activity monitors are the first steps to obtaining accurate and objective information. A detailed understanding of the best methods for doing this will ensure that the best possible data are collected. This paper is aimed primarily at the physical

---

Corresponding author: David R. Bassett, Jr., The University of Tennessee, 1914 Andy Holt Ave., Knoxville, TN 37919, Phone: 865-974-8766, FAX: 865-974-78981, dbassett@utk.edu.

CONFLICTS OF INTEREST

None declared.

activity measurement specialist, rather than the end-user seeking to use “off-the-shelf” devices for assessing physical activity.

This paper will discuss the process of calibrating and validating wearable physical activity monitors. First, we define validity and identify the types of validity analyses that are most appropriate for wearable activity monitors. Second, we review methods of calibration, paying attention to both “unit calibration” (to ensure inter-instrument reliability), and “value calibration” (i.e., conversion of raw signals into metabolic units) of wearable monitors. Third, we describe how to investigate the validity of wearable activity monitors, to determine whether they accurately estimate energy expenditure, time spent in various intensity categories, and the type of activities performed. We also discuss the strengths and weaknesses of various calibration methods. We conclude by summarizing what we consider to be best practices for calibrating and validating wearable activity monitors and priorities for future research.

Definitions of Terms

In 2004, at a conference on “Objective Monitoring of Physical Activity” held at the University of North Carolina, Welk (26) distinguished between two types of calibration in reference to wearable activity monitors. “Unit calibration” is performed to reduce inter-instrument variability and to ensure that individual activity monitors are correctly measuring the direct signals (acceleration, heart rate, body posture, heat flux). “Value calibration” of wearable monitors refers to the process used to convert the direct signals into other established measurement units. Value calibration is a validity issue, and is performed to ensure that a wearable monitor gives the intended values for outcome variables (i.e., “derived” variables).

Validity is defined as the extent to which an instrument measures what it is intended to measure (7). There are several different types of validity, including: criterion-referenced validity, content validity, and construct validity. Physical activity researchers are most interested in criterion-referenced validity with wearable physical activity monitors because the variables that we are attempting to measure are highly objective (as opposed to subjective). Content and construct validity have lesser value when it comes to establishing evidence of validity for wearable physical activity monitors; however, they are often used in validating physical activity questionnaires.

Criterion-referenced validity is comprised of two types: concurrent validity and predictive validity. Concurrent validity is determined by comparing or correlating data collected at the same time from a measure (wearable monitor) and a criterion measure. The criterion measure should be a gold standard or a measure with the highest accuracy and precision. Predictive validity is the extent to which a physical activity monitor is able to predict scores obtained using a criterion instrument.

CALIBRATION OF WEARABLE MONITORS

Unit Calibration of Direct Signals from Wearable Activity Monitors

Wearable monitors typically measure acceleration, physiological signals, body posture, or some combination of these factors. The validity of these direct signals can be checked by comparison to a “gold standard.” For accelerometers, a simple check on unit calibration would be to spin it in a circle with a known radius and frequency (RPM), so that it is exposed to a known acceleration. This technique can be used to verify that the accelerometer displays values within the manufacturer's stated tolerance limits. Unit calibration checks are important with some older activity monitors (e.g., ActiGraph [LLC, Fort Walton Beach, FL] 7164) that use cantilever beam sensors with analog filtering. They are not necessary with

newer monitors (e.g., ActiGraph model GT1M), which tend to use a direct compression sensor integrated into a solid state, micro electro mechanical system accelerometer with digital filtering. Due to tighter manufacturing tolerances with micro electro mechanical system accelerometers, the digital filter, and the initial unit calibration performed at the factory, the sensitivity varies little between the newer devices. They should remain calibrated for the life of the device, according to the manufacturers.

Although the newer micro electro mechanical system -based accelerometer technologies are more robust than their predecessors, they are not infallible. For example, little is known about the threshold detection levels of the existing accelerometer models, which could have a marked impact on sedentary time outcomes or wear-time algorithms that look for consecutive epochs of zeros (21). In addition, the robustness of the different manufacturers' unit calibrations are unclear. For example, it has not been established that a single-point unit calibration (one g-force and one frequency) is adequate to test the entire dynamic range. With these issues in mind, we still recommend that all accelerometer-based activity monitors undergo a unit calibration check with a mechanical shaker across a range of accelerations and frequencies before deployment.

Typically, unit calibration is not performed with wearable heart rate monitors and body posture devices. This is because the heart rate values obtained from heart monitors have previously been shown to be valid compared to electrocardiograms. Body posture monitors have been shown to be valid by placing them in different orientations (horizontal, vertical, tilted), and comparing this measurement to the output from the monitors. Thus, there does not appear to be a need for unit calibration with these devices.

Value Calibration of Wearable Activity Monitors

Value calibration of wearable monitors (e.g., metabolic calibration) refers to the process by which measurement researchers obtain data that allow them to convert direct signals from monitors into estimates of energy expenditure, time spent in various intensity categories, and activity type. This process involves collecting data on multiple individuals as they perform different activities and simultaneously collecting criterion data.

The data collected by accelerometer-based activity monitors are referred to as "activity counts." These activity counts are derived from the raw acceleration versus time curve. For instance, when an accelerometer is placed on a belt worn tightly around the waist, a sinusoidal acceleration versus time curve is observed. These acceleration data are filtered, full-wave rectified (meaning that the absolute value of the accelerations is used) and then integrated (i.e., the area under the curve is determined) over a pre-determined time period. The resultant activity counts are displayed for discrete time periods, or epochs (e.g., 1-min). The activity counts are then used to predict energy expenditure (9).

In 2004, Welk (26) summarized the principles for designing accelerometry-based value calibration studies. He noted that during calibration:

- The sample population must be representative of the intended population in demographics and size.
- The value calibration study should sample a sufficient number of wearable monitors because monitors have inter-unit variability.
- The activities included in the value calibration study should range in intensity from sedentary to vigorous and incorporate both ambulatory and non-ambulatory free-living activities typically performed by the intended population.

Historically, most value calibration studies on accelerometers have used a single linear regression approach. After collecting energy expenditure and activity count data on multiple individuals performing a range of physical activities, the relationship between these variables is plotted graphically and linear regression is used to determine the line of best fit. Once a single regression equation has been developed, the activity counts obtained by an individual performing an unknown activity can be used to estimate his or her energy expenditure. Activity count cutpoints denoting the dividing line between light- and moderate-intensity physical activity (3 metabolic equivalents [METs]) and moderate- and vigorous-intensity physical activity (6 METs) are typically identified. These cutpoints are then used to tally up the amount of time spent in light, moderate, and vigorous physical activity.

Montoye et al. (19) performed one of the first value calibration studies of a uniaxial accelerometer (a prototype to the Caltrac [Muscle Dynamics, Torrence, CA]). They had individuals perform level treadmill walking, treadmill running, bench stepping, knee bends, and floor touches. They used linear regression to determine the line of best fit relating acceleration to energy expenditure. Freedson et al. (13) used a similar value calibration approach with only treadmill walking and jogging. Hendelman et al. (15) and Swartz et al. (25) calibrated the ActiGraph by having subjects perform a variety of moderate-intensity lifestyle activities. In all, over a dozen regression equations have been developed for the ActiGraph alone (Table 1).

Because single regression equations cannot accurately determine energy expenditure across a wide range of activities, Crouter et al. (12) developed a two-regression equation that discriminates between walking/running and intermittent lifestyle activities based on the variability in counts across successive epochs. They calibrated the ActiGraph model 7164 on 20 different physical activities that ranged from seated rest to vigorous exercise. The method then uses one of two regression equations to predict energy expenditure, thus achieving a closer estimate than previous single regression models. Newer approaches to conducting value calibration studies that make use of “pattern recognition” have been developed, and some of them are even more accurate than Crouter’s two-regression model.

Pattern Recognition

Pattern recognition is a branch of artificial intelligence concerned with classifying or describing observations. The goal of this method is to classify data (or patterns) based on previous knowledge or statistical information extracted from the data. Pattern recognition requires the following: (a) a sensor that gathers the observations to be classified or described; (b) a feature extraction mechanism that computes numeric information from the observations; and (c) a scheme that performs the task of classifying or describing observations. The classification scheme is usually based on a set of patterns (input variables and desired outputs) that have previously been classified or described. This set of patterns is termed the training set. The machine learning strategy in this case is termed “supervised learning.” Machine learning can either use regression, in which case the resulting output function will be a continuous variable (e.g., energy expenditure), or it can use a statistical procedure known as “clustering,” in which case the output will be a category label.

Pattern recognition uses one of several approaches: statistical, syntactic, or neural. Statistical pattern recognition is based on statistical characteristics of the data. Syntactic pattern recognition is based on the structural interrelationships of features. Neural pattern recognition uses a computational method developed from artificial neural networks. No matter what approach is used, pattern recognition still requires “value calibration” in that models must be developed based on the relationship between activity counts and a direct measure of energy expenditure or physical activity intensity. The utility of pattern

recognition is highly dependent on the physical activities included in the calibration or training study. Rothney et al. (22) and Staudenmayer et al. (23) provide detailed descriptions of artificial neural network calibration studies. These studies demonstrate that pattern recognition has much greater accuracy than other methods of estimating energy expenditure using accelerometer-based activity monitors.

A number of “second generation” commercially available wearable monitors are already using pattern recognition. The Intelligent Device for Energy Expenditure and Activity (IDEEA) monitor takes data from an array of five accelerometers placed in different locations of the body, and uses this to predict energy expenditure and activity type based on an artificial neural network that was trained on a variety of activities (28,29). The SenseWear Armband (Bodymedia Inc., Pittsburgh, PA) measures acceleration, body temperature, and skin galvanic response and uses this to predict energy expenditure. It is repeatedly updated as new data collected with more subjects and more activities is used to train the artificial neural network.

Value Calibration of Heart Rate and Combined Heart Rate-Accelerometry Monitors

Value calibration of wearable heart rate monitors often involves constructing individual heart rate versus energy expenditure calibration curves. It is well known that heart rate is linearly related to oxygen uptake over a wide range of intensities, but heart rate is not a very good predictor of light-intensity physical activity. A typical method of individual calibration is the “flex method,” whereby the flex heart rate is identified as the average heart rate obtained during sitting, standing, and light exercise, and the linear relationship between heart rate and VO_2 is measured during a graded exercise test. Under free-living conditions, if an individual's heart rate falls below the flex heart rate, they are credited with 1.0 MET. Above the flex heart rate, the energy expenditure is estimated from the linear heart rate to energy expenditure (HR-EE) relationship (17).

Calibration studies of the combined heart rate-accelerometry method also have been conducted (4–6,14,24). Similar to the heart rate method, key parameters such as resting heart rate or even the individual HR-EE calibration curve must be assessed in order for this method to achieve optimal accuracy. That is, while generalized algorithms, which work for all people without the need for individual calibration, have been established, they do not have the same high level of accuracy as algorithms that use individual calibration data (5,6).

Strengths and Weaknesses of Calibration Methods

Each calibration method has strengths and weaknesses. The linear regression method is simple and easy to understand, which has led to its widespread adoption. Unfortunately, the large number of regression equations and devices is a major weakness that limits our ability to draw comparisons between studies. Adopting a single regression equation could solve this problem. However, a second and even bigger problem is that *no single regression equation accurately measures energy expenditure for all physical activities*. For example, equations developed on walking and jogging work reasonably well for those activities, but they severely underestimate the cost of most other activities (11).

New mathematical models, including hidden Markov models (20), artificial neural network (22), and classification trees (3), use the rich information contained in the acceleration versus time curve to arrive at even more accurate estimates of energy expenditure. However, a potential weakness with the current applications of pattern recognition is the reliance on data collected over 1 minute. Although models are being developed from acceleration data collected over 1 second, the parameters entered into the neural network require analysis of 1 minute's worth of 1-second data. Additionally, each minute of 1-second data is collected

from highly orchestrated and controlled calibration studies in which activity is being performed in a consistent fashion to ensure steady state energy expenditure. Approaches that use raw acceleration data collected at 20 to 30 Hz may not be subject to this limitation.

Currently, it is not known how these newer mathematical models will handle transitions in activities or the use of data collected over shorter periods that will allow one to effectively monitor more sporadic physical activity patterns (such as those exhibited by children). The issue of misclassification due to transitions from one activity type to another has surfaced with the Crouter two-regression approach (18) and this could be problematic for pattern recognition, given that artificial neural networks typically use inputs like measures of variability and auto-correlation of counts within a predetermined time period. Thus, future calibration studies may need to focus as much on transitions between activities as on the activities themselves. Modifications to the Crouter two-regression approach can address the transition limitation, but it is not certain whether similar solutions can be applied to pattern recognition.

An important calibration issue related to the use of an artificial neural network is the selection of inputs from the activity monitor used to predict activity type and/or intensity. Some researchers are using the raw acceleration data, while others are using activity counts recorded over 1-second epochs. There is no clear consensus on what parameters or signals to “extract” and feed into the neural network. The inputs may need to vary according to the prediction goal (activity type, METs, or the combination of both) and the population under study. This will be an important area for future research. The field may need to reach a consensus on what inputs are required to predict energy expenditure or physical activity intensity when using artificial neural networks or a related statistical classification technique for data reduction purposes.

Receiver Operating Characteristic (ROC) curves have been used to determine cut-off points for differing intensities of activity (sedentary, light, moderate, vigorous) in order to minimize false negatives and false positives, as discussed by Welk (26). This has the potential advantage of allowing the researcher to select cutpoints that maximize sensitivity at the cost of specificity or vice versa. This does not appear to have happened in practice to date, but these options may be appropriate for certain research questions. In addition to the ROC methodology for creating cutpoints, other methodologies include Decision Boundaries (16) and Reference Activity Calibration (10). Although these methods all have their advantages, they could further “muddy the water” regarding multiple cutpoints and limit comparability across studies.

One final point about new data reduction approaches is that they must be shown to be superior to previous approaches. For continuous variables, this may be accomplished by comparing standard error of the estimate (SEE) values, root mean square error, or 95% prediction intervals. However, new methods that involve broadly classifying the intensity of physical activity also should be compared to existing methods.

The strength of the individualized HR-EE calibration method is that it accounts for the fact that fit individuals (and younger individuals) have lower heart rate values than unfit (and older individuals), when they exercise at the same exercise intensity. However, a major weakness of this method is the need to conduct individual HR-EE calibration curves on each person, which limits its application in large-scale studies. Although this can be overcome by using generalized HR-EE regressions, the accuracy of the method decreases until it is in the same range as the best methods using accelerometer-based activity monitors. The HR-EE calibration method itself is also subject to errors that result from psychological stress, excitement, changes in ambient temperature, dehydration, and other variables.

VALIDATION OF WEARABLE MONITORS

Once calibration has taken place, the next step is to validate the wearable monitor by comparing it against a gold standard. It is important to note that researchers do not validate the measurement instrument (or physical activity monitor) *per se*. Instead, they validate the instrument in relation to the purpose for which it is being used. The practical implication is that a given wearable monitor (and associated data processing rules) may be valid for measuring one outcome variable but not another. For example, a waist-mounted uniaxial accelerometer may provide valid estimates of time spent in moderate-to-vigorous physical activity but not valid estimates of energy expenditure. Furthermore, the monitor may provide valid estimates of a given outcome in one group but not another. Estimates of time spent in different intensity categories may be valid for adults aged 20 to 50 years, but not children or adults over the age of 70 years. Estimates of total daily energy expenditure (TEE) may not be valid in any of the groups.

A typical scenario in which the intended use of a measurement instrument should be considered is the validation of accelerometer energy expenditure prediction equations. Researchers often attempt to evaluate the validity of accelerometer regression equations by correlating predicted and measured energy expenditure. However, in many cases, the equations were developed so as to derive cutpoints to broadly classify the intensity of the physical activity. In this case, validity should be judged by evaluating how well the cutpoints classify physical intensity categories, not whether they accurately predict energy expenditure for the time period. Similarly, when validating a device that measures the posture of a subject (sitting or standing), validity should be judged on the basis of how well it classifies sitting or standing. It is important to distinguish between “measurement error” (mis-measurement of something continuous) and “misclassification error” (when the target is categorical). The data analysis and statistical procedures required to validate continuous and categorical data differ markedly. Further details on this are presented in the paper by Staudenmayer et al. (8) in this supplement.

Validating Derived Variables from Wearable Monitors

Virtually all wearable monitors use direct signals to estimate derived variables, the primary one being energy expenditure. Several methods can be used to validate devices that predict energy expenditure:

- **Indirect calorimetry:** This method measures respiratory gas exchange (oxygen uptake and carbon dioxide production) to allow calculation of energy expenditure.
- **Room calorimetry:** This method involves a specially designed chamber that houses an individual for extended periods of time. Energy expenditure can be determined either through indirect calorimetry (gas exchange), or through direct calorimetry (heat production).
- **Doubly-labeled water (DLW):** In this method, a person drinks water containing stable isotopes of hydrogen and oxygen, which then equilibrate in the body. Carbon dioxide production is determined from the differential rate of elimination of ^2H and ^{18}O .
- **Direct observation:** In this method investigators keep a log of the time spent doing different activities, and the energy cost of these activities is then determined from the Compendium of Physical Activities (1).

Indirect calorimetry is an appropriate criterion for minute-by-minute energy expenditure. The Cosmed K4b² and the Jaeger Oxycon Mobile are examples of portable metabolic measurement systems that measure gas exchange (VO_2 and VCO_2) and calorie expenditure.

Because these breath-by-breath systems are more prone to error than are mixing chamber systems, a good practice is to periodically ensure the validity of instrument by measuring the oxygen cost on two to three individuals at rest and at work rates of 50, 100, 150, and 200 Watts on a cycle ergometer. The VO_2 measurements should be within 100 ml/min of expected values.

The DLW validation method is an excellent method for assessing TEE. It is important to note, however, that the DLW validation method suffers from a lack of temporally linked intensity information. In other words, it cannot provide any information on bout frequency, intensity, and duration of physical activity.

A relatively new derived variable is the type of physical activity, which can be validated against direct observation. Researchers are now using pattern recognition systems, such as artificial neural networks and classification trees, to predict types of activity (e.g., sitting, standing, walking, running, and bicycling). This is important because researchers want to know how much time people spend doing different types of activities. Knowing the type of activity could also lead to more accurate estimates of energy expenditure, because activity-specific predictions could then be applied.

BEST PRACTICES FOR WEARABLE MONITOR CALIBRATION AND VALIDATION

A wide range of physical activities should be used during calibration procedures. Given that most people spend the majority of the 24-hour day lying, sitting, and standing, it makes sense to include these activities. In addition, activities should span the entire range from sedentary activities to vigorous. Light, moderate, and vigorous activities that are typical of the types of activities performed by the population of interest should be included (26). It is helpful to think of physical activities as falling into several domains: transportation, housework, occupation, and leisure-time sports/recreation.

Core activities that should be examined include lying, sitting, standing, car driving, slow walking, brisk walking, bicycling, stair climbing, stair descending, slow running, and fast running. Other activities may include: television watching, vacuuming, sweeping/mopping, washing windows, washing dishes, doing laundry, lawn mowing, raking, one-on-one basketball, singles racquetball, and singles tennis (2,12,15,27). Selection of the activities used for metabolic calibrations has not been highly scientific up to this point. To take a more scientific approach, researchers may be able to use data from time use surveys or physical activity logs to get a more accurate picture of the most common activities performed by the population of interest.

The predictive validity of wearable monitors must be shown by cross-validation. One method of cross-validation involves dividing the data into two subsets. A researcher performs the metabolic calibration on one subset (the training set), and then validates the analysis on the other (the validation set). An alternative method of cross-validation procedure is the “leave-one-out” approach. This involves leaving out a single observation to use as the validation data, and using the remaining observations for metabolic calibration. This is repeated until every observation in the sample serves as the validation data one time.

Ideally, further cross-validations should be conducted in a simulated free-living environment across activities that are similar to, but different from, those included in the validity study. This maintains a degree of experimental control, but is more similar to what the monitor would cope with in a real-world setting. For example, a bank of tasks could be divided into categories; each category would contain tasks of a similar nature/intensity. A given number

from each category could be selected for calibration and different ones (but from the same general categories) could be put together into a simulated daily routine (punctuated by sitting time, TV viewing etc.) for the cross-validation. VO_2 could be assessed continually during the semi-structured routine using a portable system. The performance of the monitor over the entire period, as well as for individual tasks, could then be evaluated.

Summary

In calibrating and validating wearable activity monitors, it is helpful to keep these points in mind:

- Researchers must check to see that the wearable monitor is correctly measuring the direct signals (e.g., acceleration, heart rate) used to derive estimates of energy expenditure. In other words, researchers should conduct “unit calibration” studies to ensure that the signals are being properly monitored.
- When performing value calibrations of the wearable monitor, it is important to use a wide variety of activities, ranging from rest to vigorous exercise.
- The population used to calibrate the monitor should include both sexes, and a wide range of ages, adiposity classifications, and fitness levels.
- Validity of VO_2 measurements should be demonstrated by ensuring that the values obtained at fixed power outputs on a cycle ergometer are within known limits.
- Value calibration (i.e., “metabolic calibration”) studies that employ pattern recognition approaches will yield more accurate data than single-regression models because they take advantage of the rich mathematical information contained in the direct signals obtained by the wearable monitor.
- Monitors should be evaluated on the basis of their accuracy in estimating energy expenditure for many different activities, and for their ability to accurately estimate energy expenditure during free-living activity, over extended periods of time.
- Validation studies should use methods such as indirect calorimetry, DLW, and direct observation. Room calorimeters may be useful for light intensity, indoor activities, but severely restrict the types of activities that can be performed.
- Value calibrations of existing monitors should demonstrate whether the new approach offers any advantage over existing published calibrations.
- No single validation study can do it all; rather, multiple studies are needed to establish the validity of a wearable monitor.

FUTURE DIRECTIONS

Future research studies will need to answer the following questions: Do triaxial accelerometers improve predictive accuracy for determining energy expenditure compared to uniaxial accelerometers? Do pattern recognition approaches using raw data improve upon those using 1-second data? Which locations on the body provide the best predictions of energy expenditure? Do multiple sites provide greater predictive accuracy than single sites? Do pattern recognition approaches accurately measure energy expenditure and accurately classify activity type, when there are transitions between activities?

Future research should determine which demographic variables (e.g., age, height, weight, sex) are confounders and need to be used in conducting value calibrations of wearable monitors. In addition, researchers should consider whether it is appropriate to use VO_2 data for the criteria, when attempting to measure physical activity over short time periods (e.g.,

10 seconds). Finally, research laboratories should join together to design these studies and foster collaboration between groups in order to ensure a logical approach with comparability between the studies. Research groups should work together on these types of studies to attempt to increase the number of devices tested, to expand the serial number range tested, and to allow for different methods to be tested simultaneously.

Acknowledgments

Supported by National Institutes of Health grants 5 R21 CA122430 and 5 R01 HD55400. Results of the present study do not constitute endorsement by ACSM.

References

1. Ainsworth BE, Haskell WL, Whitt MC, et al. Compendium of Physical Activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc.* 2000; 32(9):S498–S516. [PubMed: 10993420]
2. Bassett DR, Ainsworth BE, Swartz AM, Strath SJ, O'Brien WL, King GA. Validity of 4 motion sensors in measuring moderate intensity physical activity. *Med Sci Sports Exerc.* 2000; 32(9):S471–S80. [PubMed: 10993417]
3. Bonomi AG, Plasqui G, Goris AH, Westerterp KR. Improving the assessment of daily energy expenditure by identifying types of physical activity using a single accelerometer. *J Appl Physiol.* 2009; 107(3):655–61. [PubMed: 19556460]
4. Brage S, Brage N, Franks PW, et al. Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. *J Appl Physiol.* 2004; 96(1):343–351. [PubMed: 12972441]
5. Brage S, Ekelund U, Brage N, et al. Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity. *J Appl Physiol.* 2007; 103(2):689–92.
6. Brage S, Franks PW, Luan J, Zhao JH, Ekelund U, Wareham NJ. Estimation of energy expenditure from accelerometry and heart rate during laboratory-based daily living activities. *J Appl Physiol.* 2009 In Press.
7. Carmines, EG.; Zeller, RA. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–017. Beverly Hills and London: Sage Publications; 1979. Reliability and validity assessment; p. 17
8. Staudenmayer J, Zhu W, Catellier D. Statistical considerations in the analysis of accelerometer-based activity monitor data. *Med Sci Sports Exerc.* 2010 In Press.
9. Chen KY, Bassett DR. The technology of accelerometry-based activity monitors: current and future. *Med Sci Sports Exerc.* 2005; 37(11):S490–500. [PubMed: 16294112]
10. Copeland JL, Esliger DW. Accelerometer assessment of physical activity in active, healthy older adults. *J Aging Phys Act.* 2009; 17(1):17–30. [PubMed: 19299836]
11. Crouter SE, Churilla JR, Bassett DR. Estimating energy expenditure using accelerometers. *Euro J Appl Physiol.* 2006; 98(6):601–12.
12. Crouter SE, Clowers KG, Bassett DR. A novel method of using accelerometer data to predict energy expenditure. *J Appl Physiol.* 2006; 100(4):1324–31. [PubMed: 16322367]
13. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc.* 1998; 30(5):777–81. [PubMed: 9588623]
14. Haskell WL, Yee MC, Evans A, Irby P. Simultaneous measurement of heart rate and body motion to quantitate physical activity. *Med Sci Sports Exerc.* 1993; 25(1):109–15. [PubMed: 8423743]
15. 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):S442–S9. [PubMed: 10993413]
16. Jago R, Zakeri I, Baranowski T, Watson K. Decision boundaries and receiver operating characteristic curves: new methods for determining accelerometer cutpoints. *J Sports Sci.* 2007; 25(8):937–44. [PubMed: 17474047]

17. Janz, KF. Use of heart rate monitors to assess physical activity. In: Welk, GJ., editor. *Physical Activity Assessments for Health-Related Research*. Champaign, IL: Human Kinetics; 2002. p. 143-61.
18. Kuffel EE, Crouter SE, Haas JD, Frongillo EA, Bassett DR. Validity of estimating minute-by-minute energy expenditure with accelerometry (abstract). *Med Sci Sports Exerc*. 2008; 40(5):S415.
19. Montoye HJ, Washburn R, Servais S, Ertyl A, Webster JG, Nagle FJ. Estimation of energy expenditure by a portable accelerometer. *Med Sci Sports Exerc*. 1983; 15(5):403-7. [PubMed: 6645869]
20. Pober DM, Staudenmayer J, Raphael C, Freedson PS. Development of novel techniques to classify physical activity mode using accelerometers. *Med Sci Sports Exerc*. 2006; 38(9):1626-34. [PubMed: 16960524]
21. Rothney MP, Apker GA, Song Y, Chen KY. Comparing the performance of three generations of ActiGraph accelerometers. *J Appl Physiol*. 2008; 105(4):1091-7. [PubMed: 18635874]
22. Rothney MP, Neumann M, Beziat A, Chen KY. An artificial neural network model of energy expenditure using non-integrated acceleration signals. *J Appl Physiol*. 2007; 103(4):1419-27. [PubMed: 17641221]
23. 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. [PubMed: 19644028]
24. Strath SJ, Bassett DR, Thompson DL. Simultaneous heart rate-motion sensor technique to estimate energy expenditure during lifestyle activities. *Med Sci Sports Exerc*. 2001; 33(12):2118-23. [PubMed: 11740308]
25. Swartz AM, Strath SJ, Bassett DR, O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Med Sci Sports Exerc*. 2000; 32(9):S450-456. [PubMed: 10993414]
26. Welk GJ. Principles of design and analyses for the calibration of accelerometry-based activity monitors. *Med Sci Sports Exerc*. 2005; 37(11):S501-11. [PubMed: 16294113]
27. Welk GJ, Blair SN, Wood K, Jones S, Thompson R. A comparative evaluation of three accelerometry-based physical activity monitors. *Med Sci Sports Exerc*. 2000; 32(9):S481-S8. [PubMed: 10993418]
28. Zhang K, Pi-Sunyer FX, Boozer CN. Improving energy expenditure estimation for physical activity. *Med Sci Sports Exerc*. 2004; 36(5):883-9. [PubMed: 15126725]
29. Zhang K, Werner P, Sun M, Pi-Sunyer FX, Boozer CN. Measurement of human daily physical activity. *Obes Res*. 2003; 11(1):33-40. [PubMed: 12529483]

Table 1

Characteristics of adult single-regression models using the Actigraph accelerometer, from Crouter et al. (11).

Actigraph					
	N	Equation	Activities	R ²	SEE
MET Predictions					
Freedson et al. (1998)	50	$1.439008 + (0.000795^* \text{ents} \cdot \text{min}^{-1})$	TM Walk (2)/run (1)	0.82	1.12
Hendelman et al. (2000)	25	$1.602 + (0.000638^* \text{ents} \cdot \text{min}^{-1})$	OG Walk (4 self-selected)	0.59	0.89
Hendelman et al. (2000)	25	$2.922 + (0.000409^* \text{ents} \cdot \text{min}^{-1})$	OG Walk (4 self-selected) and 6 lifestyle activities	0.35	0.96
Swartz et al. (2000)	70	$2.606 + (0.0006863^* \text{ents} \cdot \text{min}^{-1})$	OG Walk (2) and 26 Lifestyle activities	0.32	1.16
Leenders et al. (2003)	28	$2.240 + (0.0006^* \text{ents} \cdot \text{min}^{-1})$	TM Walk (5)	0.74	0.53
Yngve et al. (2003)	28	$1.136 + (0.0008249^* \text{ents} \cdot \text{min}^{-1})$	TM Walk (2)/run (1)	0.85	1.14
Yngve et al. (2003)	28	$0.751 + (0.0008198^* \text{ents} \cdot \text{min}^{-1})$	OG self-selected walk(2)/run (1)	0.86	1.10
Heil et al. (2003)	58	$(0.00171 \times \text{counts} \cdot \text{min}^{-1}) + (1.957 \times \text{height in cm}) - (0.000631 \times \text{counts} \cdot \text{min}^{-1} \times \text{height in cm}) - 1.883$	OG self-selected walk(2)/run (1)	0.71	0.59

Actigraph				
	N	Equation	Activities	SEE
Brooks et al. (2005)	72	$2.32 + (0.000289 \text{ counts} \cdot \text{min}^{-1})$	OG walk (1 self selected)	0.51
Brooks et al. (2005)	72	$3.33 + (0.000370 \text{ counts} \cdot \text{min}^{-1}) - (0.012 \text{ BM})$	OG walk (1 self selected)	0.61
Kcal$\cdot$min^{-1} Predictions				
Freedson et al. (1998)	35	$(0.00094 \text{ cnts} \cdot \text{min}^{-1}) + (0.1346 \text{ BM}) - 7.37418$	TM Walk(2)/run (1)	1.40
Brooks et al. (2005)	72	$3.377 + (0.000370 \text{ counts} \cdot \text{min}^{-1})$	OG walk (1 self selected)	0.17
Brooks et al. (2005)	72	$(0.000452 \text{ counts} \cdot \text{min}^{-1}) + (0.051 \text{ BM}) - 0.774$	OG walk (1 self selected)	0.77
VO$_2$ (ml$\cdot$kg$^{-1}\cdot$min^{-1}) Predictions				
Nichols et al. (2000)	60	$6.057359 + (0.002545 \text{ cnts} \cdot \text{min}^{-1})$	Walk(2)/jog (1)	0.89
				3.72

N, sample size; SEE, standard error of estimate; BM, body mass (kg); cnts $\cdot$ min $^{-1}$, counts per minute; TM, treadmill; OG, over-ground "Activity" column represents the types of activities used to develop the regression equation with the number of walking or running speeds in parenthesis.

*The manufacturer's equation is based on the work energy theorem.