Modeling Physical Activity Outcomes from Wearable Monitors

DANIEL P. HEIL[1], SOREN BRAGE[2], and MEGAN P. ROTHNEY[3]

[1]Department of Health and Human Development, Movement Science/Human Performance Lab, Montana State University, Bozeman, MT; [2]Medical Research Council Epidemiology Unit, Institute of Metabolic Science, Cambridge, UNITED KINGDOM; and [3]GE Global Research Center, Niskayuna, NY

ABSTRACT

HEIL, D. P., S. BRAGE, and M. P. ROTHNEY. Modeling Physical Activity Outcomes from Wearable Monitors. *Med. Sci. Sports Exerc.*, Vol. 44, No. 1S, pp. S50–S60, 2012. Although the measurement of physical activity with wearable monitors may be considered objective, consensus guidelines for collecting and processing these objective data are lacking. This article presents an algorithm embodying best practice recommendations for collecting, processing, and reporting physical activity data routinely collected with accelerometry-based activity monitors. This algorithm is proposed as a linear series of seven steps within three successive phases. The Precollection Phase includes two steps. Step 1 defines the population of interest, the type and intensity of physical activity behaviors to be targeted, and the preferred outcome variables, and identifies the epoch duration. In Step 2, the activity monitor is selected, and decisions about how long and where on the body the monitor is to be worn are made. The Data Collection Phase, Step 3, consists of collecting and processing activity monitor data and is dependent on decisions made previously. The Postcollection Phase consists of four steps. Step 4 involves quality and quantity control checks of the activity monitor data. In Step 5, the raw data are transformed into physiologically meaningful units using a calibration algorithm. Step 6 involves summarizing these data according to the target behavior. In Step 7, physical activity outcome variables based on time, energy expenditure, or movement type are generated. Best practice recommendations include the full disclosure of each step within the algorithm when reporting monitor-derived physical activity outcome variables in the research literature. As such, those reading and publishing within the research literature, as well as future users, will have the best chance for understanding the interactions between study methodology and activity monitor selection, as well as the best possibility for relating their own monitor-derived physical activity outcome variables to the research literature. **Key Words:** ACCELEROMETRY, ACTIVITY MONITOR, ALGORITHM, ENERGY EXPENDITURE, MODERATE-TO-VIGOROUS, PHYSICAL ACTIVITY

Assessment of free-living physical activity (PA) has greatly improved with the use of wearable monitors to objectively measure one or more biosignals, such as positioning and acceleration of a limb or the body, heart rate, and various measures of temperature (e.g., positioning and acceleration of a limb or the body, heart rate, and various measures of temperature). A common problem with all current wearable monitors is processing and summarizing data into PA outcome variables once the data are collected. Specifically, different methods for handling the same data can result in dramatically different values for the same outcome variables (11,15). Given the number of users (e.g., biostatisticians, laboratory technicians, research scientists) making decisions that affect the collection and processing of free-living wearable monitor data, the lack of congruence between PA outcome measures from different monitors within the same study should be expected. Thus, although the measurement of PA with wearable monitors may be considered objective, consensus guidelines for collecting and processing these objective data are lacking (26).

The goal of this article is to provide best practice recommendations for the future collection, processing, and reporting of PA data routinely collected with accelerometry-based activity monitors. Common commercial examples of activity monitors include the ActiGraph (ActiGraph, Fort Walton Beach, FL), Tritrac RT3 (StayHealthy, Monrovia, CA), Actical (Philips Respironics, Bend, OR), and the Actiheart (Metrisense, Bend, OR). This article's focus on these types of wearable monitors is primarily a function of their prevalent use in clinical and research settings, as well as their historical significance to the field of free-living PA assessment.

The target audience for this article includes anyone responsible for making decisions about collecting, processing, or summarizing wearable monitor data before writing a report, research manuscript, or grant proposal. All of these users will have made decisions that influence the values of the PA outcome variables.

A STANDARD SEVEN-STEP ALGORITHM

The best practice recommendations described here are embodied in a series of functional and analytical steps used to predict PA outcome variables with wearable monitors, called a data collection and processing algorithm, or simply an algorithm. The essence of this algorithm is to translate the information measured from one or more biosignals into a summarized variable or variables that predict one or more types of PA outcome variables (i.e., based on time, energy expenditure, or activity type). Although the steps described in this algorithm should be similar for many wearable monitors, even when collecting different biosignals, the options available within each step will depend on each device. This article will focus on what we are calling a standard seven-

step algorithm (Fig. 1), a linear series of steps that cascade from a Precollection Phase (Steps 1–2) to the Collection (Step 3) and Postcollection phases (Steps 4–7). The presentation of this algorithm is intended to serve as one example of how to conceptualize an algorithm and is targeted at users who are planning to collect data soon.

Step 1: Define the Data Collection Strategy

The first step of the PA assessment algorithm is to clearly define the criteria on which a wearable monitor is selected, which should include, at a minimum: 1) the population of interest, 2) the intensity and type of PA behaviors that are targeted, 3) the preferred PA outcome variables, and 4) the epoch duration. Once these criteria have been defined,
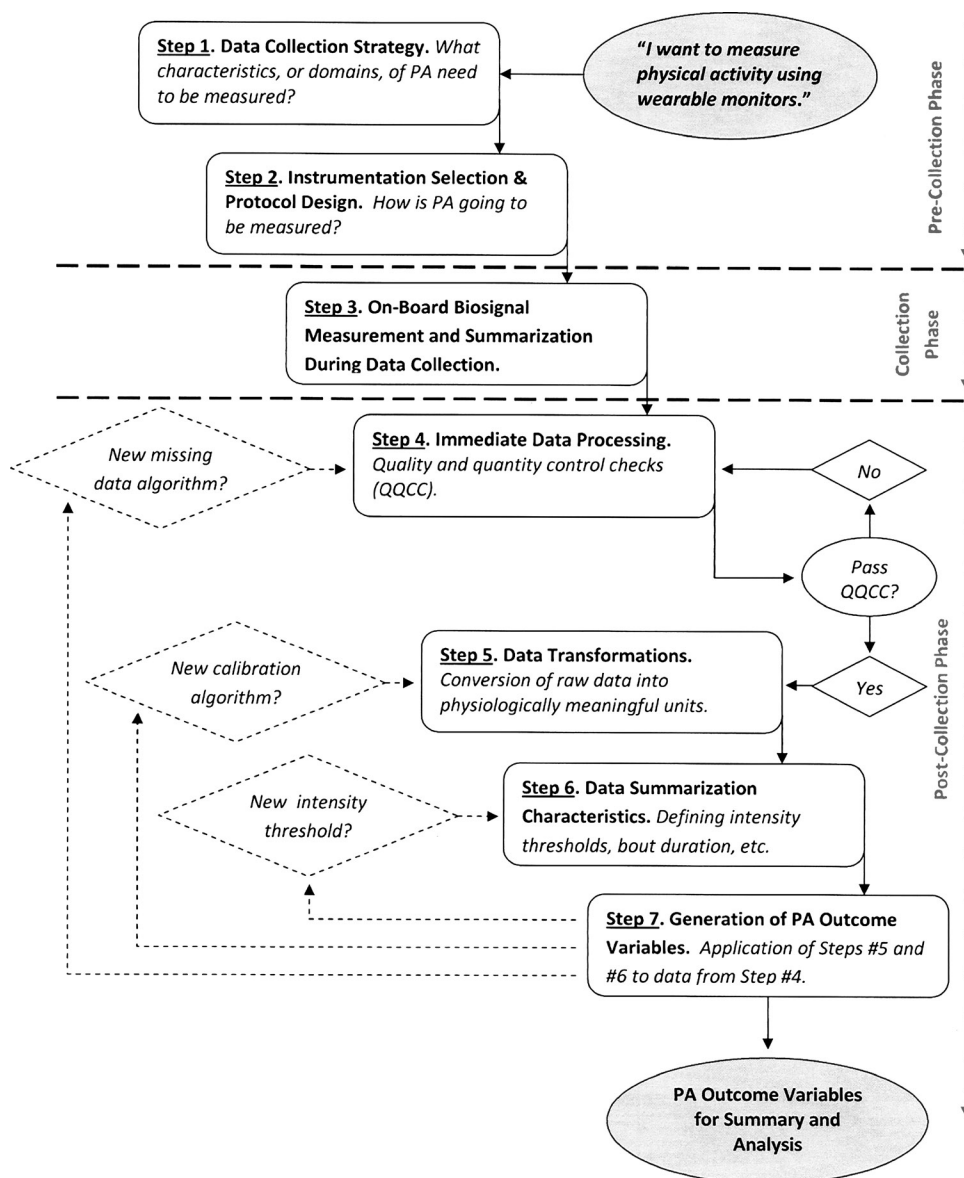

FIGURE 1—The standard seven-step algorithm conceptualizing the collection, processing, and summarization of PA data collected with wearable monitoring systems. As presented, the algorithm applies most accurately to the use of accelerometry-based activity monitors for free-living PA assessments. The *solid-lined arrows* depict this series of steps using traditional activity monitor data, whereas the *dashed-lined arrows* depict how previously collected data can be reevaluated when new postcollection data processing components become available in the literature.

the user will be ready to select an instrument that can meet the majority of the study's needs.

Population of interest. The population of interest should be defined in terms of characteristics such as sex, age, adiposity, disease pathology (e.g., diabetes, cardiovascular disease), or other grouping characteristic of interest (e.g., functional capacity, living environment, socioeconomic, or employment status). Each of these descriptions may be an important determinant of other PA algorithm parameters, such as monitor wearing location (Step 2), the availability of data transformation algorithms (Step 5), and data summarization characteristics (Step 6), or the expected types and intensity of PA behaviors to be monitored.

Intensity and type of PAs. The four domains that traditionally characterize any PA include duration, frequency, intensity (i.e., sedentary, light, moderate, and vigorous), and activity type (or mode). Although most activity monitors are theoretically capable of operating over a broad range of intensities, research has traditionally been biased toward identifying moderate-to-vigorous intensity PAs (MVPAs). For instance, a common health promotion goal has been to change participants' behavior so that MVPA increases regardless of the type of activity. Physical activity intervention studies also have focused on increasing the occurrence of particular PA behaviors, such as the amount of habitual walking or daily lifestyle activities that also satisfy MVPA health promotion guidelines (3,9,28). Some weak evidence suggests that traditional activity monitors may be used to grossly categorize locomotor versus nonlocomotor PAs (6,10). However, an accurate classification of activity types (e.g., sitting, standing, walking, lifestyle activities) or the actual identification of specific activities (standing quietly vs standing and washing dishes) will require the use of more sophisticated wearable monitors, such as those based on artificial neural networks (24,25,33,34).

Most recently, activity monitors have been applied to the study of sedentary behaviors without an adjustment for acceleration range sensitivity (19). Although this strategy is convenient because monitors are already available to study MVPAs, it also may compromise the accuracy of PA outcome variables focused on sedentary behaviors. Regardless, the burden is on the user to determine whether the study should focus on identifying the intensity of PAs, identifying or classifying PA types, or some combination of both, as well as which instrument satisfies these criteria for the population of interest.

Physical activity outcome variable(s) of interest. The most important factors dictating choice of biosignal measurement should be the user's ability to accurately transform that biosignal into PA outcome variables of interest and the feasibility of measuring that biosignal within the population of interest. Table 1 can be used as a starting point for guiding activity monitor users from the PA behavior-related question to an outcome variable as derived from an activity monitor. Research questions related to measuring the specific types of activity (e.g., Tae Kwon Do), however, should be considered an emerging field that will evolve dramatically in the next 5–10 yr.

Epoch duration. An epoch is a user-defined time interval over which the activity monitor information is summarized. The background details on these computations have been described (5) and reviewed (27) previously. This choice is intimately related to the choice of PA outcome variables and thus must be defined before data collection. The traditional epoch length for energy expenditure–based studies in adults has been 60 s, which is a direct reflection of the standard 60-s sample interval used with indirect calorimetry procedures to measure submaximal steady-state oxygen consumption ($\dot{V}O_2$). Most published calibration studies have used the 60-s epoch to establish relationships between activity monitor outputs (the result of Step 3) with energy expenditure. Some research in children, in contrast, has shown that shorter epochs (≤ 5 s) better capture their behavior preference for short bursts of vigorous intensity activity (17). In contrast, wearable monitors that rely on neural network models simply use the raw data collected at high frequency (10–40 Hz) as modeling inputs to the PA algorithm rather than summarizing these data into epochs. Thus, although steady-state energy expenditure–based outcomes may be represented reasonably well with

TABLE 1. Linking the PA outcome variables of interest with the types of information typically derived from wearable monitoring (WM) systems and the PA behavior-related question or topic.

PA Behavior-Related Question/Topics	Variables Derived from WM Systems	Examples of PA Outcome Variables	Example Commercial WMs for Assessment[a]
Have the ACSM/AHA guidelines for weekly PA been satisfied?	PA intensity as determined by O_2, METs, EE, or AEE.	Time in MVPA	ActiGraph, Actical, Actiheart, Tritrac RT3, Sensewear Pro 2
PA intervention study to increase the number of 10-min or longer bouts of moderate-intensity PA.	PA intensity as determined by O_2, METs, EE, or AEE.	Number of MVPA bouts; time in MVPA	
Influence of energy expenditure on energy balance and body weight maintenance?	Energy expenditure as determined by EE or AEE.	TEE and PAL	ActiGraph, Actical, Actiheart, Tritrac RT3, Sensewear Pro 2
Walking activity intervention study	PA activity classification or identification	Time spent walking; total steps from pedometry	IDEEA Digiwalker Pedometer
Lifestyle intervention study	PA activity classification or identification	Time spent performing lifestyle activities	IDEEA

[a] Company name and location information for commercial brand names listed are provided in the text.

AEE, activity energy expenditure (EE − RMR, or RMR); BMR, basal metabolic rate; EE, energy expenditure; METs, metabolic equivalents ($\dot{V}O_2/3.5$); PAL, physical activity level (TEE/BMR); TEE, total daily energy expenditure; $\dot{V}O_2$, oxygen uptake.

60-s epochs, outcomes related to activity type are almost completely lost. As a general recommendation, activity monitor data should be collected over the shortest possible epoch to retain as much information as possible about the original PA-related biosignal.

Step 2: Select Monitor and Design Protocol

Users are asked to combine criteria from Step 1 with additional methodological criteria to select an activity monitor. Additional criteria include how the monitoring system can (or must) be worn, the amount of time the monitor should be worn, and timing constraints on when the activity monitor is actually worn. Collectively, the decisions (Steps 1 and 2) play a critical role in all downstream algorithm steps because they constrain the available data and analysis strategies to which a user will have access.

Monitor wearing location. The primary issues driving the selection of the monitor wearing location are wearing compliance and design characteristics of the monitor itself. Some commercial devices, such as the Sensewear Pro 2 Armband (BodyMedia, Pittsburgh, PA) and the IDEEA monitor (Intelligent Device for Energy Expenditure and Physical Activity; MiniSun LLC, Fresno, CA) are worn only in one way owing to the sensors' characteristics. In contrast, many of the watch-sized activity monitors have been evaluated for wear at several different locations, the most common of which is the waistline (i.e., hip-worn monitors) (27). Although the hip offers a theoretical advantage (measuring acceleration near the subject's center-of-mass), historically, wearing compliance has been poor. This is especially true in large studies or studies with long monitoring durations, where significant personal contact by the study team may not be possible. Unfortunately, an accurate assessment of the scope and magnitude of this compliance problem is not possible because researchers have been reluctant to report these data along with the data successfully collected. Responding, in part, to this compliance problem, other wearing locations for the same hip-worn monitors, especially the ankle or wrist, have been evaluated for use in free-living PA assessments (10–12). Limb-worn monitors may be contraindicated for populations at risk for swelling in the extremities, while some people simply dislike wearing a watch-like band for long periods.

Monitor wearing duration. Behavioral reliability refers to the minimum number of days that monitors should be worn to ensure that daily averages for PA outcome variables accurately reflect habitual PA. A previous review indicates that 3–5 d of monitoring for adults, and 7 d for children, should be sufficient to reliably estimate habitual PA with hip-worn monitors (27). However, the very premise of the current article is that different data processing algorithms applied to the same data set will result in different PA outcome values. Mâsse et al. (15), for example, used four different PA algorithms to analyze the same activity monitor data set on adults and found that nearly every outcome variable was influenced regardless of intensity category (sedentary, light, or MVPA). Clearly, this topic needs to be addressed more systematically in the research literature to better understand the potential interaction of wear duration and location on measures of behavioral reliability.

Timing of monitor wear. Studies should be designed so that monitors can be worn continuously while avoiding periods in which individuals' activity behavior is altered from their "typical" 7-d weekly routine (e.g., holidays, vacations, scheduled surgery). The phrase "continuous wearing" often refers to a specific predetermined period in which everyone wears the monitors. Alternatively, subjects could be given a window of time, such as 4 wk, in which to continuously wear the activity monitor for a period of 7–10 d.

Step 3: Collect and Process On-board Biosignals

Decisions made within Steps 1 and 2 cannot be altered once data collection with an activity monitor has begun. The on-board processing and summarizing of biosignals by commercial monitors actually involve both collecting and processing data. On-board data summarization generally means converting a relatively high-frequency "raw" signal into a single positive value for each user-defined epoch. In the case of activity monitors, a common result of this transformation is an activity count value. This processing step can be avoided only if raw biosignal data are stored by the monitor, but this currently is not an option for many commercial activity monitors. Each brand of monitor has a slightly different approach to this data-filtering step, and the peer-reviewed literature offers substantial insights into the filtering strategies used by the makers of the most common monitors (5). Very little can be done to change this on-board data processing routine, but it is important for users to realize that this is the first data processing step and thus may affect downstream analyses because the information stored by the activity monitor will be less than what is actually measured (i.e., the sample interval, or epoch, is always less than the monitor's sampling frequency of the biosignal). This feature and its limitations need to be understood at the time of activity monitor selection.

Step 4: Conduct Immediate Data Processing

This processing step has three goals: 1) identify raw data points that may not accurately represent the PA behavior, 2) decide how to handle these data, and 3) fully disclose the process and outcomes for these steps when publishing. Collectively, these quality and quantity control checks are performed before Step 5. Outcomes from Step 4 include the reporting of the total number of data files successfully collected, the number of files not included with those used for statistical analyses, a summary of reasons why files were not included in the analyses (e.g., failed monitors, subjects did not wear the monitor, not enough complete measurement days), and user errors (e.g., new files were saved with the same name as older files). This will help

future users evaluate the success of the methodology, as well as understand the challenges to expect when using this instrumentation in the future.

Quality control checks. Sometimes called "data cleaning," this step involves identifying spuriously high and/or low values according to predetermined threshold values using a combination of quantitative and qualitative assessment techniques. A common *quantitative* check has been to screen the raw data, either visually or with automated methods, to identify data points that exceed a predetermined threshold. For example, activity monitor data exceeding a 20,000-counts-per-minute threshold has been used to indicate spuriously high data points or monitor failure. The exact value of these thresholds will vary between monitoring devices and may be established by physiological expectations, commonly reported values from the literature, or known limitations to the measurement device itself. An effective *qualitative* method includes the visual assessment of time-based graphic plots. Regardless of what software is used to generate these plots, the experienced user can visually identify spurious biosignal values, unexpected periods of nonwear, and monitor malfunctions.

Quantity control checks. Two common activity monitor quantity control checks are 1) identifying nonwear times and 2) ensuring that enough data were collected to satisfy the minimal requirements for behavioral reliability. Nonwear times are periods within a monitoring day during which the activity monitor was purposely removed and then reattached by the subject, when the activity monitor was accidentally disconnected from the subject, or periods when the activity monitor was not worn enough to represent a full monitoring day. Nonwear times create missing or incomplete data, meaning that the recorded activity monitor data do not fully describe the subject's actual PA behavior for the intended monitoring period (4). While the analytical procedures for handling missing data have been described previously
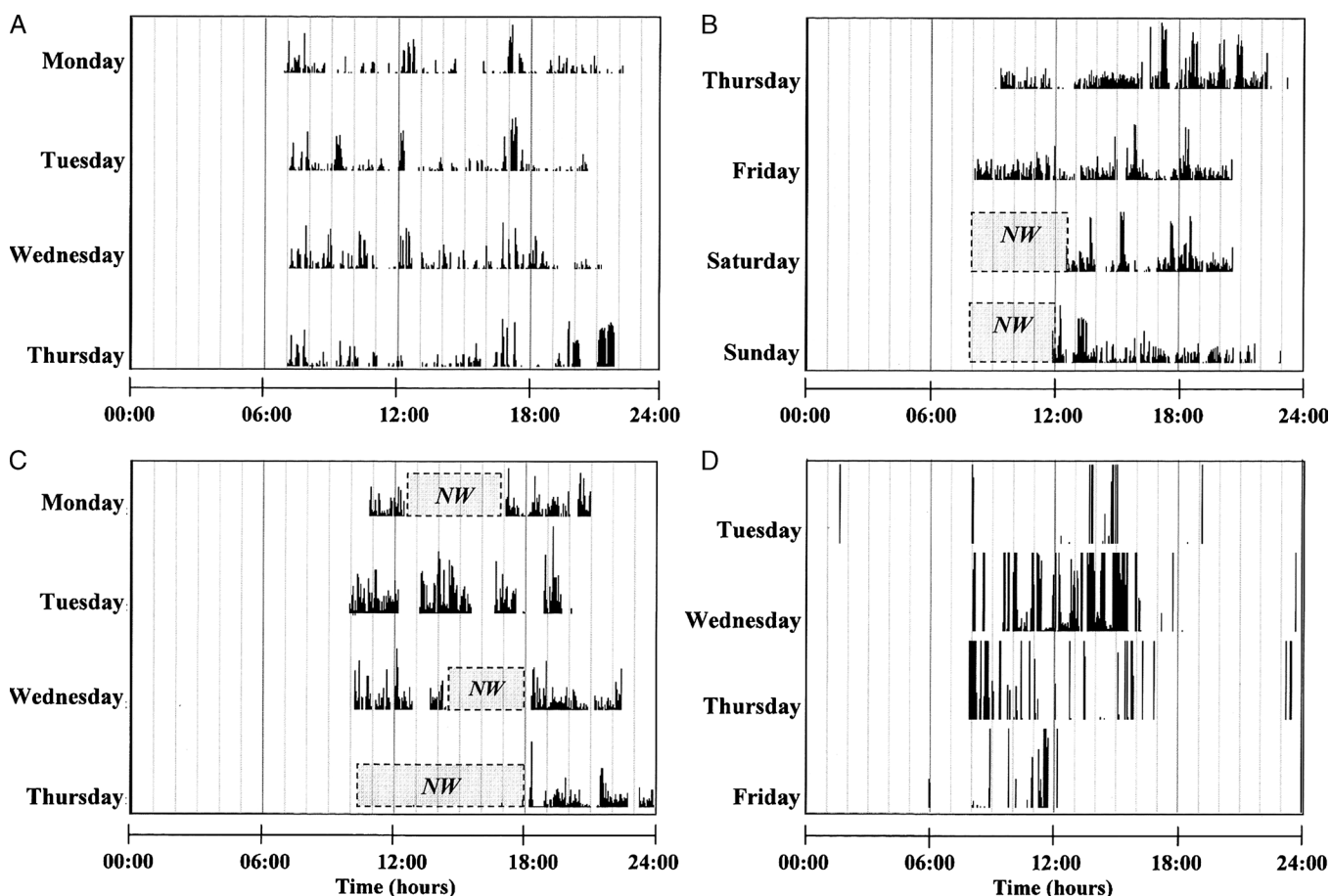

FIGURE 2—Depiction of several common quantity and quality control issues that are easily observed with a time-based graphic plot of raw data from an accelerometry-based activity monitor. Each graph shows four successive days from a 7-d collection period using a hip-worn activity monitor. Each graph is for a different subject, but all subjects were part of the same PA intervention study. Plot A is an example of someone with perfect compliance (i.e., good monitor wearing habits) during all 4 d. This subject's habits include similar monitor wear time (14.5–16 h·d^{-1}), similar monitor donning time each day (7:00 a.m.), and no discernable nonwear (NW) periods or short wear days. Plot B, in contrast, shows a person with good compliance during the weekdays but evidence of short wear days on the weekend (i.e., late to don and early to doff monitors). Weekend donning was after noon, while weekday donning was 3–4 h earlier. Plot C is a good example of someone with mixed nonwear time issues, with evidence of multiple 1–1.5 h nonwear times on 1 d, extended nonwear times (>3 h) on two other days, and a fourth "short wear" day (Thursday). Plot D shows what a monitor failure can look like, a result that likely has nothing to do with the subject's wearing habits. Although some of these data patterns may resemble PA, data are saturated, which means that a large range of true data points have recorded the same spuriously high value.

(4,14,29), the underlying principle of this step is that these procedures should be planned, systematic, and thoroughly described in formal reports.

Several common decision rules for handling missing data have been used as a starting point to decide whether more advanced data imputation procedures are needed. It is common, for instance, to allow a single 1- to 2-h midday period of apparent nonwear time during each measurement day. Another common rule is to establish a minimum amount of continuous activity monitor wear time to define a full monitoring day. Described as such, this procedure allows for an individual to have different monitor wear times each day and thus accounts for some variances in day-to-day sleep patterns. Large deviations from a full monitoring day can be identified visually as short wear days (e.g., 6 h vs the usual 10–12 h). A related issue is the minimum number of wear days needed to ensure sufficient behavior reliability (from Step 2). Thus, although the standard for reliability was established in Step 2, the verification of satisfying the reliability standard occurs in Step 4. Using data from the research literature (32), Figure 2 illustrates several types of the quality and quantity control issues from a study collecting activity monitor data at the hip during seven successive days (31). These examples include perfect compliance (Fig. 2A), nonwear times on weekends versus weekdays (Fig. 2B), multiple nonwear time issues (Fig. 2C), and output from a faulty activity monitor (Fig. 2D).

Step 5: Transform Data

The raw activity monitor data are typically transformed into one or more new physiologically meaningful variables. These transformations, hereafter called calibration algorithms, have used a variety of analytical techniques, such as simple linear regression (7,20), multiple regression (22,23), which are multistep evaluations that involve the use of threshold values, one or more regression lines (2,6,10), and artificial neural networks (25,33,34). Derived from what the research literature calls calibration studies (8,16), these calibration algorithms typically transform the raw activity monitor data into predicted energy expenditure or are used to identify activity types. Although much of the earlier PA research literature has previously emphasized the reporting of summarized raw activity monitor data (e.g., reporting summed counts for an intensity category), this is no longer considered appropriate because only studies that have used the exact same monitor can be compared directly. Transforming the raw data into a commonly reported physiological energy expenditure unit, however, allows monitors based on current and new technologies to be compared directly. Given that the development and use of calibration algorithms is a highly active research field, the burden is truly on the user to determine what algorithms are available and considered appropriate for each wearable monitor by the current research literature.

Transformation to energy expenditure units. Previous calibration study reviews (8,16) have highlighted the fact that no universal standard exists for transforming raw activity monitor data into units of energy expenditure. For example, calibration studies in children have generated calibration algorithms that predict oxygen uptake $\dot{V}O_2$, or mL·kg^{-1}·min^{-1}), metabolic equivalents (METs, or $\dot{V}O_2/3.5$), total energy expenditure (TEE, or kcal·kg^{-1}·min^{-1}), PA energy expenditure (PAEE, or TEE − RMR [resting metabolic rate]), and the PA ratio (PAR, or TEE/BMR [basal metabolic rate]). Adult calibration studies have been just as varied but with greater emphasis on predicting METs and $\dot{V}O_2$. Although calibration algorithms have historically been dominated by MET prediction algorithms for adults and children, recent research has tended to focus on predicting PAEE.

Complicating this translation step are issues such as the fact that each calibration algorithm is specific to each brand of activity monitor and the wearing location at which the activity monitor was originally validated (i.e., calibrated). Moreover, each calibration algorithm is considered valid only for populations that are similar to that for which the algorithm was originally derived. Researchers interested in using the Actical monitor, for instance, can predict PAEE in both adults and children at several wearing locations (10,22), but users wanting to predict PA levels can only do so in children with the Actical worn at the hip (23). Thus, issues related to population and EE variables of interest, monitor wearing location, available calibration algorithms from calibration studies, and choice of activity monitor brand (discussed in Step 2) are inexorably linked.

Transformation to activity type. Because of the limitations with PA inferences when predicting energy expenditure, recent calibration algorithms have begun to incorporate some degree of activity type classification and/or identification. The simplest of these algorithms have used various approaches for fitting different regression lines to locomotion versus nonlocomotion calibration data (6,10). Artificial neural network algorithms, sometimes called pattern recognition software, are "trained" to identify activity types based on numerous channels of information to either predict energy expenditure and/or predict specific types of activity (classification) or the activity itself (identification) (24,25,33,34).

Proprietary calibration algorithms. Some commercial monitors use software to automatically transform the raw data after they have been downloaded from the computer. From the perspective of the seven-step algorithm (Fig. 1), proprietary algorithms combine Steps 3 and 5, may eliminate the ability to screen raw data (Step 4), and effectively limit, or even eliminate, some of the typical Step 6 postprocessing options (Fig. 3). An advantage of using monitors with proprietary algorithms is that many computational portions of the seven-step algorithm are automated, which simplifies the task of processing and summarizing PA data. A disadvantage, of course, is that the user will have limited choices about how to handle the data after collection. Another major disadvantage is that it may not be possible to reexamine the raw data when the monitor manufacturer,
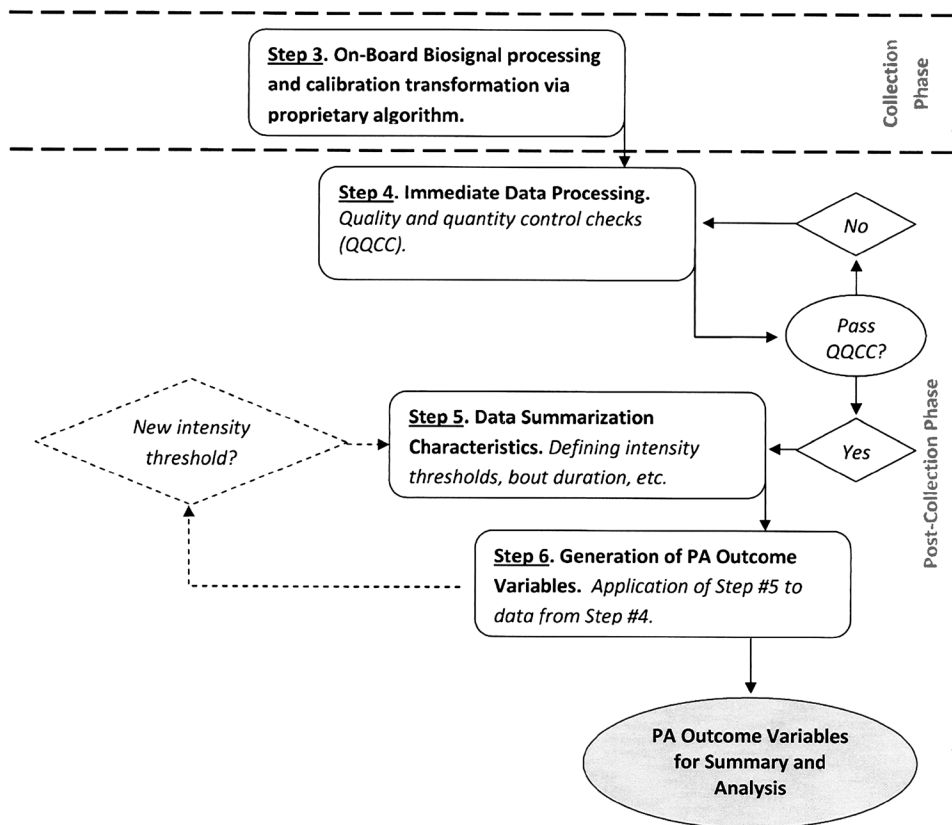
FIGURE 3—A modified PA data analysis and summarization algorithm for wearable monitors that incorporates a *proprietary* data transformation algorithm at Step 3. This depiction represents a combination of Steps 3 and 5 shown in the standard seven-step algorithm (Fig. 1). The most noticeable influence of the proprietary algorithm is a reduced number of postcollection analytic options because the original raw data are lost within the Step 3 transformations. Note that Steps 1 and 2 are the same as those presented for the standard seven-step algorithm.

or the research literature, provides new or updated calibration algorithms.

Step 6: Summarize Data Characteristics

Completing Step 5 results in a string of energy expenditure–based numbers. In Step 6, these numbers are typically summarized according to intensity thresholds and the definition of a bout, both of which define the dose of PA experienced by the body (9). General conclusions from this section include 1) the traditional use and development of cut points to summarize activity monitor data be avoided in future studies, 2) the traditional use of 1-min bouts to determine outcome variables should be avoided when the data are intended to relate to PA guidelines in adults, and 3) the recognition of MVPA bouts within activity monitor data using an algorithm requires the use of both time–intensity and MET-minute bout definitions (9).

Intensity thresholds. Physical activity intensity is traditionally classified as one of four intensity thresholds: sedentary, light, moderate, and vigorous (9,18). Given that it is now recommended that activity monitor data be preferentially transformed into units of PAEE, it seems prudent to reinterpret the traditional MET-based intensity thresholds for use with PAEE data. Because PAEE is computed as energy expenditure minus RMR, PAEE-based thresholds

will depend on population differences in measured RMR. When expressed in units relative to body mass (e.g., $mL \cdot kg^{-1} \cdot min^{-1}$ or $kcal \cdot kg^{-1} \cdot min^{-1}$), measured RMR will tend to decrease with maturation through childhood, decrease with advancing age through adulthood, and decrease with an increase in adiposity independent of age. Thus, an emerging component of the standard algorithm is the need for population-specific PAEE-based intensity thresholds. Some specific values for both MET- and PAEE-based intensity thresholds as reported in the literature are provided in Table 2.

A common goal within PA-related physiological and epidemiological studies is the need to relate raw activity monitor data to the time spent within the MVPA intensity range. There are two general analytical approaches to this issue. The standard approach is to collect, screen, transform, and then summarize data according to a defined PA bout and MVPA intensity threshold (i.e., ≥3.0 METs)—i.e., the standard seven-step algorithm (Fig. 1). A large portion of the research literature, however, follows the traditional approach, which is characterized by avoiding the transformation of activity monitor data with a calibration algorithm (skip Step 5) by simply associating an activity monitor cut point with the MVPA intensity threshold. The activity monitor data are then summarized in the same manner as

TABLE 2. Intensity thresholds based on METs and AEE reported in the literature for adults and children.

Intensity Threshold Categories	Sedentary	Light	Moderate	Vigorous
MET based				
Intensity thresholds				
Adults[a]	<1.6 METs	1.6–2.9 METs	3.0–5.9 METs	≥6.0 METs
AEE based				
Intensity thresholds				
Children[b]	<0.0100	0.0100–0.0390	0.0400–0.0900	≥0.1000
Adults[c]	<0.0075	0.0075–0.0309	0.0310–0.0831	≥0.0832
Adults[d]	<0.0147	0.0147–0.0384	0.0385–0.0895	≥0.0896

The MET-based thresholds are those commonly cited in literature for use with adults. Units for METs are dimensionless, whereas those for AEE are in kilocalories per kilogram per minute (kcal·kg^{-1}·min^{-1}).
MET (dimensionless) = oxygen consumption $\dot{V}O_2$, mL·kg^{-1}·min^{-1})/(3.5 mL·kg^{-1}·min^{-1}); AEE = EE (kcal·kg^{-1}·min^{-1}) − RMR (kcal·kg^{-1}·min^{-1}).
[a] Intensity thresholds reported by both Haskell et al. (9) and Pate et al. (19).
[b] Thresholds were used for classification of AEE for a mix of overweight and nonoverweight ethnically diverse children (7–18 yr) (23).
[c] Thresholds were based on physically active, healthy, nonoverweight Caucasian adults (18–49 yr) (10).
[d] Thresholds were based on healthy overweight and obese African American women (26–59 yr) (13).

that described for the standard approach (Steps 6 and 7). Use of the traditional approach requires at least three major assumptions: 1) the relationship between activity monitor data and the transformed data can be modeled with simple linear regression throughout the intensity range of interest, 2) an energy expenditure–based definition of a bout can be accurately recognized using nontransformed activity monitor data, 3) activity monitor data and measures of energy expenditure are relatively similar regardless of the type of activities being performed. With regard to the first assumption, many calibration studies have reported linear relationships between activity monitor data and measures of energy expenditure (7,20), but most of the recent calibration studies have consistently described these relationships with nonlinear regression, multiple regression, or multiple nonoverlapping linear regression (2,6,10,22,23). Because the second assumption regarding bout definitions relies on the validity of the linearity assumption, the second assumption also is invalid. The third assumption, as thoroughly discussed in the literature (8,16), also is invalid because the relationship between raw activity monitor data and energy expenditure depends on more than just PA intensity (i.e., the linearity assumption). Other factors include population demographics (e.g., age, sex, and body size) and the types of activities on which the monitor cut point evaluation was based (i.e., treadmill walking and jogging vs simulated household activities). Although the scope and impact of the cut point problem (8,16), as well as alternative strategies for computing cut points (30), have been thoroughly acknowledged in the literature, an appropriate alternative analytical strategy has yet to be well defined.

Physical activity bout definition. Public health recommendations for PA define a bout as a minimum of 10 successive minutes (9,28), but this definition has not been consistently applied to the evaluation of activity monitor data. In fact, the traditional use of 1-min recording epochs

seems to have been used as the *de facto* bout definition. It is likely that the common use of the 1-min bout definition has been prevalent because of the ease with which summary variables are determined by statistical or spreadsheet programs versus the need to recognize and summarize 10-min bouts. However, when the goal of evaluating activity monitor data is to relate PA behavior with PA guideline compliance (9,28), the use of a 1-min bout definition should be completely avoided in favor of the 10-min definition. According to Haskell et al. (9), there are two complementary methods for recognizing an MVPA bout: 1) time–intensity, which defines an activity bout as lasting ≥10 consecutive minutes at a moderate intensity (≥3.0 METs), and 2) MET-minutes, which defines an activity bout as the accumulation of ≥30 MET·min (10-min bout × 3.0 METs) over ≥10 consecutive minutes.

The potential complexity of PA bout recognition is illustrated by five PA bouts shown in Figure 4. The first bout (B_1) shows several 1-min PA spikes above 3 METs, which may be clinically relevant for specific PA behaviors, such as fidgeting (14). However, they are not considered relevant to the accumulation of MVPA bouts according to PA guidelines. The classic square-wave MVPA bout (Fig. 4, B_2; 3.5 METs for 10 min, or 35 MET·min) satisfies both time–intensity and MET-minute definitions, but free-living PA behavior often demonstrates 1- to 2-min breaks in activity bouts (i.e., bout interruptions, such as stopping at a signal crossing, or tying a shoelace), as well as the natural variability around the average intensity. Although not explicitly outlined by past or present PA guidelines (9,19,28), both bout interruptions and PA variability are accounted for within the MVPA bout definitions. For example, a 10-min square-wave PA bout with a 2-min bout interruption
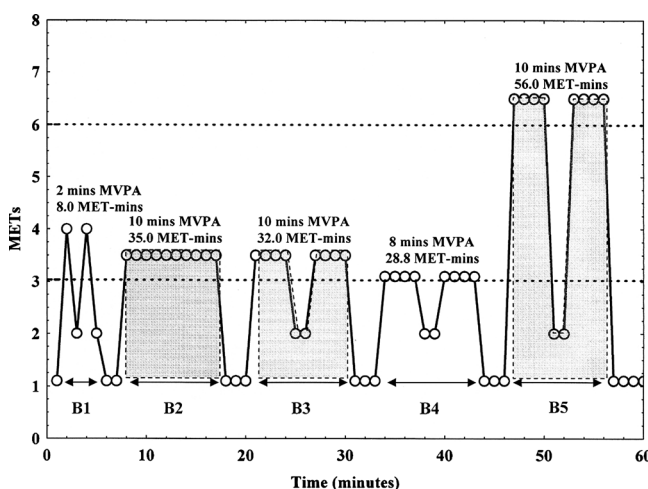

FIGURE 4—Graphic illustration of five PA bouts (B_1–B_5) where only those that are shaded (B_2, B_3, and B_5) actually qualify as MVPA bouts according to the PA guidelines for adults. Each of the five bouts illustrates the interaction of bout duration, average bout intensity, and bout interruptions relative to the MVPA bout definition. *Horizontal dashed lines* at 3.0 and 6.0 METs represent the most common PA intensity thresholds for moderate and vigorous intensities in adults, respectively.

(Fig. 4, B_3) satisfies both satisfies neither the time-intensity definition nor the MET-minute definition (32 MET·min) bout definitions. Thus, despite the obvious interruption of the bout, the PA guidelines would recognize this as an MVPA bout. In contrast, a similar bout at a lower intensity (10 min at 3.0-MET average) satisfies neither the time-intensity definition nor the MET-minute definition (28.8 MET·min) because of the same 2-min bout interruption (Fig. 4, B_4). When PA bouts average at much higher intensities (Fig. 4, B_5), interruptions have a relatively small influence on both bout definitions. Thus, 1- to 2-min bout interruptions should be allowed when defining MVPA bouts in activity monitor data as recommended previously (15,29), so long as both time–intensity and MET-minute definitions are satisfied. It should be noted that, while the interaction of bout duration, bout intensity, and bout interruptions seems reasonable when explained with overly simplified illustrations (Fig. 4), actual activity monitor and energy expenditure data tend to have considerably more variability as a result of natural variation within the most repetitive of motions. Thus, a rigorous definition of an activity bout is critical to the eventual summarization of outcome variables.

Step 7: Generate Physical Activity Outcome Variables

Finally, the data summarization characteristics (Step 6) are applied to the transformed data from Step 5 to generate PA outcome variables. Generally, four types of outcome variables are of interest. These are variables based on 1) movement, 2) time, 3) energy expenditure, and 4) activity type. Currently, the best strategy is to minimally report both time- and energy expenditure–based outcome variables using a commonly reported metric (e.g., MVPA min·wk^{-1}) so that the results can be directly compared with as many activity monitor studies as possible. However, many studies using existing data sets, as well as in-place longitudinal studies, also may need to report movement-based variables.

Movement-based variables. The movement-based term is used to indicate that an activity monitor has recorded data because of sensed movement without any attempt to transform the data into physiological units. Traditionally, movement-based PA variables reported in the literature are derived from activity monitors and reported as daily (counts per day) or intensity-based summed counts (MVPA counts). Future activity monitor studies, however, should not rely solely on the use of count-based outcome variables because the unit of measurement is specific to the brand of activity monitor being used and typically rely on the use of activity monitor cut points. For those needing to report their activity monitor data as a movement-based variable (e.g., a longitudinal study in progress), a reasonable compromise would be to include other variable types (i.e., those based on time, energy expenditure, or activity type). In the near future, commercial monitors may allow users to return

to a movement-based outcome variable with the adoption of a common metric, such as acceleration in standard units.

Time-based variables. The time-based outcome variables are one of the most commonly reported because of the direct relationship to PA guidelines (9,28). For example, accumulated time spent within specific PA categories, such as average MVPA minutes per day or minutes per day of sedentary time, are common metrics that can be derived from any commercial monitor.

Energy expenditure–based variables. The energy expenditure–based outcome variables are increasingly common as monitors are incorporated into longitudinal weight management studies with targeted PA components. Common variables can include both daily (total kcal·d^{-1} or PAEE per day) and intensity-based (TEE or PAEE during MVPA) outcomes.

Activity type–based variables. This type of variable summarizes a metric relative to a type of activity (e.g., summed steps by a pedometer for locomotion-based activities) or a specific activity (e.g., weekly time spent jogging). Activity type–based variables are less frequently reported as a PA outcome variable because the most commonly used monitors have been unable to accurately identify specific activities or types of activities. However, as described in the next section, this is an emerging area in the development of new activity monitors that takes advantage of more sophisticated technology and algorithms. Users should expect activity type–based variables to become more prevalent in the reporting of PA outcome variables.

FUTURE DIRECTIONS

Although the dominant trend with activity monitors is to predict energy expenditure, the classification of activity types and identification of specific activities are also possible with pattern recognition and machine learning approaches. Such approaches have been addressed in the literature (24,25,33,34), but their use is not yet widespread. This is probably due to a combination of poor feasibility for long-term monitoring and a relatively high cost per monitoring system. However, these types of systems are likely to become cheaper and less invasive for long-term monitoring and will likely replace the ubiquitous use of single-channel monitors for smaller clinical and research studies. Regardless, these advanced systems are already being used as field-based criterion measures for predicting both energy expenditure and activity types (11,31).

The new generation of wearable monitoring systems will likely be small, easy to wear for extended periods, and have the capacity to store raw data instead of filtered or processed data. These characteristics are as much a function of what systematic advances in technology are allowing at a reasonable price, as what past experience with simpler devices has taught researchers about what is really needed in a wearable monitor. Manufacturers of the earliest activity

monitors, for example, created the user-defined epoch as a means of summarizing a complex analog biosignal while not exhausting the activity monitor's on-board memory. As newer monitors emerged onto the market, the old epoch-based model was replicated because it was viewed as an acceptable (and successful) procedure to users. However, it is now generally recognized that this epoch-based system of activity monitor data processing and storage actually limits the accuracy of predicting energy expenditure and almost eliminates the ability to predict activity types. Thus, the newest monitors with the most potential will be those with the ability to store the raw biosignal during long periods (at least seven successive days) so that predicting both EE and activity types is possible.

Analytical procedures for postprocessing of data will likely become considerably more sophisticated and complex, so much so in fact that common use of proprietary or open-source software will become necessary for routine PA data evaluations. Indeed, the availability of such data processing software is generally considered a limitation to the common adoption of advanced analytical strategies by activity monitor users. The process of identifying PA bouts, or PA doses, from activity monitor data also will evolve to more accurately reflect the dynamic changes in energy expenditure observed with true free-living behavior. Lastly, the influence of body size on the expression of energy expenditure–based outcome measures from activity monitors has been almost completely ignored by the research literature. Although it is well known that body size does not proportionally influence the energy cost of locomotion versus nonlocation activities (21), it is likely that daily measures of energy expenditure as outcome variables also will vary nonproportionally with body size (1).

Finally, although the standard seven-step algorithm has been described here as most applicable to most of today's commercial activity monitors, the essence of the algorithm will be the same for emerging technologically and analytically advanced wearable monitors. Decisions will always have to be made within a Precollection Phase, for example, and these decisions will affect the nature and precision of the downstream PA outcome variables. A Data Collection Phase in which decisions from the previous phase are merged with the actual collection of data according to a predefined plan also always will exist. A Postcollection Phase also will exist, but users should expect that the sophistication of options available would continue to evolve at an accelerating pace.

CONCLUSIONS

The standard seven-step algorithm presented in this article should be viewed as a starting point, rather than the end point, for conceptualizing and implementing the collection, processing, and summarization of objective PA data with wearable monitors. It is worth noting that the application of this type of algorithm is intended for the development of new activity monitor studies rather than an attempt to dictate changes to studies and analyses already in progress. In addition, the seven-step algorithm is intended to be a detailed example of how to conceptualize this process and not the only possible algorithm.

Assuming the use of the seven-step (or similar) algorithm and given the dependence of PA outcome variables on the interaction of upstream decisions in the algorithm, best practice reporting standards should include a full disclosure of the algorithm used to conceptualize, collect, transform, and summarize data into PA outcome variables. Using the seven-step algorithm and the preceding discussion of activity monitors as an example, this disclosure should minimally include:

- *Precollection Phase*—Population of interest, the type and intensity of PA behaviors targeted by the investigator, the intended PA outcome variables, how an epoch was defined, monitoring wearing characteristics (wearing location, duration, and timing of monitor wear).
- *Collection Phase*—A complete description of the on-board biosignal collection and processing characteristics that ultimately dictate the raw activity monitor data available to users. This should minimally include functional characteristics of the biosignal monitor itself, the sampling frequency of the monitor's analog signal, and any other filtering or summarizing characteristics of the biosignal.
- *Postcollection Phase*—The type and outcome of data quality and quantity control checks, the use of calibration algorithms (published or proprietary), specific definition used for data summarization characteristics (e.g., intensity thresholds, bout definition), and the rationale for the choice of reported PA outcome variables.

REFERENCES

1. Abbott RA, Davies PSW. Correcting physical activity energy expenditure for body size in children. *Ann Hum Biol*. 2004;31(6): 690–4.
2. 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):343–51.
3. Brooks GA, Butte NF, Rand WM, Flatt JP, Caballero B. Chronicle of the Institute of Medicine physical activity recommendation: how a physical activity recommendation came to be among dietary recommendations. *Am J Clin Nutr*. 2004;79(5): S921–30S.
4. Catellier DJ, Hannan PJ, Murray DM, et al. Imputation of missing data when measuring physical activity by accelerometry. *Med Sci Sports Exerc*. 2005;37(11 suppl):S555–62.
5. Chen KY, Bassett DR. The technology of accelerometry-based activity monitors: current and future. *Med Sci Sports Exerc*. 2005; 37(11 suppl):S490–500.

6. Crouter SE, Clowers KG, Bassett DR. A novel method for using accelerometer data to predict energy expenditure. *J Appl Physiol*. 2006;100(4):1324–31.

7. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc., accelerometer. *Med Sci Sports Exerc*. 1998;30(5):777–81.

8. Freedson P, Pober D, Janz KF. Calibration of accelerometer output for children. *Med Sci Sports Exerc*. 2005;37(11 suppl):S523–30.

9. Haskell WL, Lee I, Pate RR, et al. Physical activity and public health: Updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Med Sci Sports Exerc*. 2007;39(8):1423–34.

10. Heil DP. Predicting activity energy expenditure using the Actical® activity monitor. *Res Q Exerc Sport*. 2006;77(1):64–80.

11. Heil DP, Bennett GG, Bond KS, Webster MD, Wolin KY. Influence of activity monitor location and bout duration on free-living physical activity. *Res Q Exerc Sport*. 2009;80(3):424–33.

12. Heil DP, Hymel AM, Martin CK. Predicting free-living energy expenditure with hip and wrist accelerometry versus doubly labeled water. *Med Sci Sport Exerc*. 2009;41(5):447.

13. Heil DP, Whitt-Glover MC, Brubaker PH, Mori Y. Influence of moderate intensity cut-point on free-living physical activity outcome variables. *Med Sci Sports Exerc*. 2007;39(5 suppl):S185.

14. Johannsen DL, Ravussin E. Spontaneous physical activity: relationship between fidgeting and body weight control. *Curr Opin Endocrinol Diabetes Obes*. 2008;15(5):409–15.

15. Mâsse LC, Fuemmuler BF, Anderson CB, et al. Accelerometer data reduction: a comparison of four reduction algorithms on select outcome variables. *Med Sci Sports Exerc*. 2005;37(11 suppl):S544–54.

16. Matthews CE. Calibration of accelerometer output for adults. *Med Sci Sports Exerc*. 2005;37(11 suppl):S512–22.

17. Nilsson A, Ekelund U, Yngve A, Sjostrom M. Assessing physical activity among children with accelerometers using different time sampling intervals and placements. *Pediatr Exerc Sci*. 2002; 14:87–96.

18. Pate RR, O'Neill JR, Lobelo F. The evolving definition of "sedentary." *Exerc Sport Sci Rev*. 2008;36(4):173–8.

19. Pate RR, Pratt M, Blair SN, et al. Physical activity and public health. A recommendation from the Centers for Disease Control and the American College of Sports Medicine. *J Am Med Assoc*. 1995;273(5):402–7.

20. Pfeiffer KA, Mciver KL, Dowda M, Almeida MJCA, Pate RR. Validation and calibration of the Actical accelerometer in preschool children. *Med Sci Sports Exerc*. 2006;38(1):152–7.

21. Prentice AM, Goldberg GR, Murgatroyd PR, Cole TJ. Physical activity and obesity: problems in correcting expenditure for body size. *Int J Obes*. 1996;20(7):688–91.

22. Puyau MR, Adolph AL, Vohra FA, Butte NF. Validation and calibration of physical activity monitors in children. *Obes Res*. 2002;10(3):150–7.

23. Puyau MR, Adolph AL, Vohra FA, Zakeri I, Butte NF. Prediction of activity energy expenditure using accelerometers in children. *Med Sci Sports Exerc*. 2004;36(9):1625–31.

24. Rothney MP, Neumann M, Beziat A, Chen KY. An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *J Appl Physiol*. 2007;103(4):1419–27.

25. 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.

26. Troiano RP. A timely meeting: objective measurement of physical activity. *Med Sci Sport Exerc*. 2005;37(11 suppl):S487–9.

27. Trost SG, Mciver KL, Pate RR. Conducting accelerometer-based activity assessment in field-based research. *Med Sci Sport Exerc*. 2005;37(11 suppl):S531–43.

28. US Department of Health and Human Services. *2008 Physical Activity Guidelines for Americans. 2008*. Washington (DC): US Department of Health and Human Services; 2008.

29. Ward DS, Evenson KR, Vaughn A, Roders AB, Troiano RP. Accelerometer use in physical activity: best practices and research recommendations. *Med Sci Sports Exerc*. 2005;37(11 suppl):S582–8.

30. Welk GJ. Principles of design and analyses for the calibration of accelerometry-based activity monitors. *Med Sci Sports Exerc*. 2005;37(11 suppl):S501–11.

31. Welk GJ, McClain JJ, Eisenmann JC, Wickel EE. Field validation of the MTI ActiGraph and BodyMedia armband monitor using the IDEEA monitor. *Obesity (Silver Spring)*. 2007;15(4): 918–28.

32. Whitt-Glover MC, Hogan PE, Lang W, Heil DP. Pilot study of a faith-based physical activity program among sedentary blacks. *Prev Chron Dis*. 2008;5(2):1–9.

33. Zhang K, Pi-Sunyer FX, Boozer CN. Improving energy expenditure estimation for physical activity. *Med Sci Sports Exerc*. 2004; 36(5):883–9.

34. Zhang K, Werner P, Sun M, Pi-Sunyer FX, Boozer CN. Measurement of human daily physical activity. *Obes Res*. 2003;11(1): 33–40.