A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations

Kate Lyden,
Department of Kinesiology, University of Massachusetts, Amherst, 30 Eastman Lane, Amherst, MA 01003, (508) 963-4821 (ph), (413) 545-2906 (fx), klyden@kin.umass.edu

Sarah L. Kozey,
Department of Kinesiology, University of Massachusetts, Amherst, 30 Eastman Lane, Amherst, MA 01003

John W. Staudenmeyer, and
Department of Math and Statistics, University of Massachusetts, Amherst, Lederle Graduate Research Tower, Amherst, MA 01003

Patty S. Freedson
Department of Kinesiology, University of Massachusetts, Amherst, 30 Eastman Lane, Amherst, MA 01003

Abstract

Numerous accelerometers and prediction methods are used to estimate energy expenditure (EE). Validation studies have been limited to small sample sizes in which participants complete a narrow range of activities and typically validate only one or two prediction models for one particular accelerometer.

Purpose—To evaluate the validity of nine published and two proprietary EE prediction equations for three different accelerometers.

Methods—277 participants completed an average of 6 treadmill (TRD) (1.34, 1.56, 2.23 m·sec^{-1} each at 0% and 3% grade) and 5 self-paced activities of daily living (ADLs). EE estimates were compared to indirect calorimetry. Accelerometers were worn while EE was measured using a portable metabolic unit. To estimate EE, 4 ActiGraph prediction models were used, 5 Actical models, and 2 RT3 proprietary models.

Results—Across all activities, each equation underestimated EE (bias −0.1 to −1.4 METs and −0.5 to −1.3 kcals, respectively). For ADLs EE was underestimated by all prediction models (bias −0.2 to −2.0 and −0.2 to −2.8, respectively), while TRD activities were underestimated by seven equations, and overestimated by four equations (bias −0.8 to 0.2 METs and −0.4 to 0.5 kcals, respectively). Misclassification rates ranged from 21.7% (95% CI 20.4%, 24.2%) to 34.3% (95% CI 32.3%, 36.3%), with vigorous intensity activities being most often misclassified.

Discussion—The prediction equations did not yield accurate point estimates of EE across a broad range of activities, nor were they accurate at classifying activities across a range of intensities (light < 3 METs, moderate 3–5.99 METs, vigorous 6 METs). Current prediction techniques have many limitations when translating accelerometer counts to EE.

Keywords

Accelerometers; energy expenditure prediction equations; objective measurement of physical activity; activity classification

Introduction

The role of physical activity (PA) in promoting health and preventing chronic disease has long been established. However, accurately quantifying or measuring PA remains a challenge to researchers and clinicians. Large-scale epidemiological studies, field-based research and clinical trials have traditionally relied on subjective methods such as questionnaires, self-report diaries and interviews. Such methods however, have proven inaccurate, with individuals tending to over-report time spent in PA (Sallis et al. 2000). In order to accurately quantify PA and elucidate the dose-response relationship between PA and health outcomes, researchers have turned to objective measurement tools. Specifically, accelerometers have emerged as the device of choice to measure free living PA.

Accelerometers offer minimal subject burden, versatility, and relative cost efficiency. However, once researchers decide to use accelerometers they are faced with two immediate challenges. First, they must choose which of the many commercially available devices is best suited for their research. For example, the ActiGraph, Actical and RT3 are three commonly used accelerometers. Each of these devices produces a "count" value as their output. The way in which this count value is generated depends on a unique set of technical specifications (e.g. A/D conversion scale, frequency filtering range, number of axes sensitive to acceleration etc.) distinctive for each monitor. Thus, although counts have traditionally been considered the universal accelerometer output used in PA research, technical differences render counts an arbitrary, unit-less output that is not comparable across monitors. The second challenge facing researchers is that for each device several different regression models are available to predict energy expenditure (EE) from accelerometer output (counts). Using the first generation of what is currently the ActiGraph, Freedson et al. (1998) developed the first linear regression prediction model to estimate energy expenditure from accelerometer counts. It was a relatively simple calibration study in which 50 subjects performed 3 treadmill activities ranging from brisk walking to slow jogging. The accelerometer was positioned on the anterior supra iliac spine in an attempt to capture vertical acceleration of the center of mass. This model was based on the principle that vertical acceleration is linearly related to energy expenditure during locomotion. However this relationship breaks down at higher running speeds (Cavagna et al. 1976) and does not translate to non-locomotive activities. Since then many single and multi, linear and non-linear equations have been developed for each monitor. In addition, several multi-step techniques have been developed which rely on the activity intensity, the variation of the movement, or the type of movement to determine the appropriate prediction equation to use (Klippel & Heil 2003; Crouter et al. 2006a; Crouter et al. 2008). Although these more sophisticated methods have been developed, all current regression techniques collect and average accelerometer counts over a specified period of time, usually one-minute. The averaged count value is used to estimate energy expenditure given the relationship dictated by the prediction equation. In other words, energy expenditure is a function of the average counts·min^{-1} and is often expressed as specific point estimates of energy expenditure (e.g. 4 kcals), as a rate of energy expenditure (e.g. 4 METs), or as light (< 3 METs), moderate (3–5.99 METs,) or vigorous (6 METs) intensity activity.

Due to the numerous commercially available accelerometers and published prediction techniques, there is a great deal of confusion in the literature when predicting energy expenditure from accelerometer counts. The many prediction models often produce widely different estimates of time spent at various intensities of PA with no clear indication of which estimate is correct. Although calibration studies often use a form of cross validation to assess the success of their newly developed model, the sample on which the model is tested is often very similar to the sample from which it was produced. In addition, the test sample usually performs the same (or similar) activities as the activities from which the

model was developed. As a result, the reported performance scores do not provide an accurate representation of how the model will perform for the general population and across a broad range of activity types and intensities. Although several studies have attempted to assess monitor accuracy and identify valid prediction techniques (Crouter et al. 2006b; Rothney et al. 2008), they have been limited by small sample sizes in which subjects perform a narrow range of activities. Additionally, studies often focus on only one monitor or prediction technique, making comparison between studies impossible. There has yet to be a single comprehensive evaluation that uses a large sample with an extensive range of participant characteristics (e.g. age, height, weight, BMI, physical activity level, and race/ ethnicity) to assess the validity of the most popular prediction models and activity monitoring devices for a wide range of common household, locomotion and sporting activities. Therefore, the purpose of this study was to conduct a comprehensive evaluation of commonly used prediction models on a large diverse population. Specifically, this study evaluated the validity of nine published and two proprietary energy expenditure prediction equations using the ActiGraph, the Actical and the RT3 activity monitors. Additionally, activity intensity classification accuracy using these prediction models was evaluated.

Methods

Subjects

Two hundred and seventy seven healthy men and women between the ages of 20–60 years were recruited from the Amherst, Massachusetts area. Each participant completed an informed consent document, a health history questionnaire, the Physical Activity Readiness Questionnaire (PAR-Q) and a questionnaire to evaluate habitual physical activity status. Before completing the study protocol, female participants over 50 yrs and male participants over 40 yrs were screened for cardiovascular disease risk with a physician-supervised 12-lead ECG stress test to 90% of age-predicted maximum heart-rate according to the American College of Sports Medicine Guidelines for Exercise Testing (2009). Participants were excluded if they had any contraindications to exercise, were taking medication altering metabolic rate or if the physician identified any cardiovascular abnormalities that potentially prevented them from safely completing the activity protocol.

Anthropometric and Metabolic Measurements

Prior to testing, participants' height and weight were measured using a stadiometer and a physician's scale and body mass index (BMI) was calculated. Blood pressure was measured using the "OSZ 5 easy" automatic blood pressure cuff (Welch Allyn, Inc, Arden, NC) and participants were excluded from the study if their blood pressure exceeded 140 mm systolic and 90 mm diastolic. Resting Metabolic Rate (RMR) was measured using the Med Gem Analyzer (HealtheTech, Inc, Golden, CO). The Med Gem is a hand-held indirect calorimeter that calculates energy expenditure based on a modified Weir equation and uses a fixed respiratory exchange ratio of 0.85 (HealtheTech 2003). The MedGem has been shown to be a valid device for measuring resting metabolic rate compared to the gold-standard Douglas bag method (Nieman et al. 2003). Following a 4-hr restriction of food, caffeine and exercise, participants rested quietly for 15 minutes in the supine position. RMR was measured while the participant remained supine.

Activity Protocol

The activity protocol consisted of two routines performed in random order; treadmill activities (TRDs) (Part A) and activities of daily living (ADLs) (Part B). Each activity was performed for 7 minutes (except ascending and descending the stairs) with 4 minutes rest between each bout. For any activity, if heart rate exceeded that which was safely established

during the stress test, or if the participant was unable to safely complete the activity (e.g. treadmill speed too fast), the activity was stopped and eliminated from analysis.

Part A

Participants performed six treadmill activities at 3 speeds (1.34, 1.56, 2.23 m·sec^{-1}), each at 0% and 3% grade. The order of activities was randomized across subjects.

Part B

Activities of daily living consisted of common household and sporting activities. Five ADLs were performed at a self-selected pace. Three ADLs were performed by each subject; ascend stairs, descend stairs and moving a 6 kg box. The remaining two activities were randomly selected from a catalog of 14 possible activities, including cleaning the room, dusting, gardening, laundry, mopping, mowing, painting, raking, sweeping, trimming, vacuuming, washing dishes, basketball and tennis. These activities represent common household, leisure time and sporting activities. Common ADLs (ascending and descending the stairs and moving a weighted object) were chosen as being representative of the spectrum of activities that people perform.

Accelerometers

The accelerometers were worn on a belt positioned around the participants' hips. The ActiGraph accelerometer was positioned on the non-dominant (right handed participants wore the accelerometers on the left hip etc) in line with the anterior superior iliac spine, the Actical was positioned directly posterior to the ActiGraph, and the RT3 was positioned directly anterior to the ActiGraph.

ActiGraph

The ActiGraph accelerometer (model GT1M) (ActiGraph, LLC, Fort Walton Beach, FL) is a uniaxial accelerometer that measures movement in the vertical plane. The monitor is small in size and lightweight, $5.1 \times 3.8 \times 1.5$ cm and 42.6 gm, respectively. It is sensitive to accelerations from 0.05–2.0 G's and has a band limited frequency of 0.25–2.5 Hz. The ActiGraph samples at a rate of 10 Hz and the signal is digitized by an 8 bit A/D converter. Each signal is summed over a user specified time interval (epoch) and activity counts are stored. The ActiGraph was initialized to collect data in one-second epochs and results were downloaded directly to a PC compatible computer using a USB cable.

Actical

The Actical (Mini Mitter Co., Inc., Bend, OR) is an omni-drectional accelerometer that is $28 \times 27 \times 10$ mm in size and weighs 17 g. It measures accelerations in the range of 0.05–2.0 G's and has a band limited frequency of 0.5–3.0 Hz. The Actical samples data at a rate of 32 Hz and can be initialized to collect data from 15-second – one-minute epochs. For this study, the Actical was initialized to collect data in 15-second epochs and results were downloaded directly to a PC compatible computer.

RT3

The RT3 accelerometer (StayHealthy, Inc., Monrovia, CA) is a triaxial monitor that measures acceleration in three orthogonal dimensions. It is the size of a pager, $71 \times 56 \times 28$ mm in size and 65 g. The RT3 provides triaxial vector data in activity counts. The sensor range, sampling frequency and the linear regression algorithm used by the manufacturer's software are proprietary. For this study, the RT3 was initialized to record data in one-second epochs and the vector magnitude (triaxial vector data) was used to predict EE. Data from the RT3 were downloaded directly to a PC compatible computer.

Indirect Calorimetry

During each activity, oxygen consumption was measured using a portable metabolic measurement system (Oxycon Mobile; Cardinal Health, Yorba Linda, California). The Oxycon Mobile is a portable respiratory gas exchange system that measures ventilation and expired concentrations of oxygen and carbon dioxide and estimates energy expenditure using a modified Weir equation (Weir et al. 1949). Its lightweight (2 kg) and wireless transmission system allow the Oxycon Mobile to be used in a non-laboratory setting. For each activity ventilation and expired gas concentrations were collected breath-by-breath and energy expenditure measured from the Oxycon Mobile served as the criterion measure to which to compare energy expenditure estimated from the prediction equations. Immediately prior to each activity routine (TRD and ADLs) a two-point (0.2 and 2.0 $L \cdot s^{-1}$) air flow calibration was performed using the automatic flow calibrator, and the gas analyzers were calibrated using a certified gas mixture of 16 % O_2, 4.01% CO_2. The Oxycon Mobile system is a valid device for measuring VO_2 (Perret et al. 2006; Rosdahl et al. 2009). Compared to the Douglas Bag, the Oxycon Mobile produced accurate and reliable estimates of V_E, VO_2 and VCO_2 during maximal and sub-maximal cycling. (Rosdahl et al. 2009).

Prediction Equations

Nine published and two proprietary regression models were examined. For information on each model, including features of their development, see Table 1. Equations most commonly used in research were chosen for analysis, and they are defined (and analyzed) exactly as they are published.

Data Analysis and Reduction

For each activity, the first 120 seconds were eliminated to ensure steady state had been reached and the last 10 seconds were eliminated to minimize any researcher error in timing synchronization between the monitor and the metabolic measurements. After elimination of the first 120 seconds and last 10 seconds, the remaining data needed to be at least 30 seconds in order to be included in the analyses. Thus activity data ranged from 30 seconds to 290 (7 minutes minus 130 seconds) seconds in length. In order to ensure steady state was reached within 120 seconds, we assessed the differences in oxygen consumption for minutes 2 and 3 vs. the last minutes of activity. For two activities (gardening and trimming), METs decreased about 8% (0.3 METs). For all other activities, the changes were less than 5%. Thus, 120 seconds was a sufficient time to establish steady state during these activities.

Monitor Data

For each activity, accelerometer data were converted to average counts·min^{-1} and entered appropriately into each equation to predict energy expenditure. For example, if an activity was performed for 290 seconds accelerometer data were averaged over 4.8 minutes. Each activity was then classified as light (<3 METs), moderate (3–5.99 METs) or vigorous (6 METs) intensity. For the equations that predict EE in kcals (Freedson kcal, Heil AEE, and RT3 Proprietary), kcals were first converted to METs and then classified. For the Crouter et al. ActiGraph and Actical two-regression methods, the accelerometer count coefficient of variation (CV) was determined for each minute of activity to direct counts to the appropriate equation. For the ActiGraph model, the CV for each minute was determined by using six 10-second epochs per minute (CV = standard deviation (SD)/mean). For the Actical model the CV was determined using 4 consecutive 15-second epochs. For both models, each minute of activity was assigned a CV and directed to the appropriate equation.

Indirect Calorimetry Data

Average measured VO_2 was determined and converted to relative VO_2 ($ml \cdot kg^{-1} \cdot min^{-1}$) and then to METs. Relative VO_2 was converted to METs by dividing by 3.5 $ml \cdot kg^{-1} \cdot min^{-1}$ for all analyses except for those pertaining to the RT3 proprietary models. For RT3 analyses, measured VO_2 was converted to relative VO_2, gross EE (GEE) ($kcals \cdot min^{-1}$), activity EE (AEE) ($kcals \cdot min^{-1}$ – RMR) and to METs using measured RMR rather than the standard 3.5 $ml \cdot kg^{-1} \cdot min^{-1}$. For each equation, predicted EE was compared to either measured METs (determined using either 3.5 $ml \cdot kg^{-1} \cdot min^{-1}$ – ActiGraph and Actical; or measured RMR – RT3), GEE or AEE.

Despite measuring RMR as part of the protocol, we defined RMR for the ActiGraph and Actical analyses as the commonly used standard 3.5 $ml \cdot kg^{-1} \cdot min^{-1}$. Although recent evidence suggests this standard measure significantly underestimates RMR for specific sub-groups (Byrne et al. 2005; Kozey et al., 2010), each of the prediction models evaluated (except RT3 models) were developed using 3.5 $ml \cdot kg^{-1} \cdot min^{-1}$ as a standard baseline measure. Because it was our intent to evaluate the models, and not to address the differences in using measured RMR compared to the standard 3.5 $ml \cdot kg^{-1} \cdot min^{-1}$, we used each model in the way in which it was developed. Furthermore, this is how these models are commonly used in the field, especially in large epidemiologic studies that do not have the means to measure RMR for all participants. Since it is our intent to provide a comprehensive report that can be used as a resource for researchers deciding which activity monitor and prediction model is best suited for their research, this approach is the most useful and widely applicable. For the development of the proprietary RT3 models, we are uncertain if the standard 3.5 $ml \cdot kg^{-1} \cdot min^{-1}$ or measured RMR was used. However, the RT3 proprietary equations estimate RMR as part of their prediction models, thus leading us to use measured RMR in the RT3 analyses. By using measured RMR in these analyses we believe we increased the likelihood that the RT3 models would be successful at predicting EE given the recent evidence that estimated RMR is more closely related to measured RMR than to the standard 3.5 $ml \cdot kg^{-1} \cdot min^{-1}$ (Byrne et al. 2005; Kozey et al., 2010).

Statistical Analysis

All statistical analyses were performed using the free and open source computing language and statistics package R (2009). For each prediction equation, predicted EE was compared to measured EE using a repeated measures mixed model. For each individual activity, treadmill activities, activities of daily living, and across all activities combined, the mixed models were used to assess the average difference between predicted EE and measured EE (bias). A negative bias (predicted EE-measured EE) indicates an underestimation of EE by the prediction model; a positive bias corresponds to an overestimation of EE by the prediction model. Ninety-five percent confidence intervals (CI) were also established from the mixed models and were used to determine significance. If the upper and lower confidence intervals spanned 0, then predicted EE was *not* significantly different from measured EE at $\alpha=0.05$. To describe the magnitude of the difference between measured and predicted EE the root mean squared error (RMSE) was also determined for each activity, treadmill activities, activities of daily living, and across all activities combined. Although bias and 95% CI's were used to determine significance it is essential to consider both the bias and RMSE when evaluating the validity of a prediction model. The bias is used to give an indication of whether the model under- or over- estimates EE. However, an overall bias close to 0 can be deceiving. For example, if a model considerably underestimates EE for activities of daily living, but considerably overestimates EE for treadmill activities these divergent errors will essentially cancel each other out, resulting in a small bias that may indicate the prediction model produces an EE that is *not* significantly different from measured EE. The RMSE measures the square root of the average squared difference between predicted and measured

EE. This is similar to average of the absolute value of the differences. Thus we will consider both bias and RMSE when interpreting the results.

Activity intensity classification was described using misclassification rates and 95% CI. Kappa statistics were used to describe the level of agreement between actual activity intensity classification and predicted activity intensity classification.

Results

The total possible number of activities was 3047 (277 participants × 11 activities). One hundred twenty seven activities were eliminated due to the participant being unable to perform the activity for the minimum time needed for analysis (30 seconds) (e.g. treadmill speed too fast) or researcher discretion (e.g. participant heart rate exceeding peak HR on graded exercise test), for a total of 2920 activities performed. Of the 2920 activities performed, 145 were eliminated due to Oxycon malfunction (e.g. Oxycon Mobile sample tube occlusion) or insufficient VO_2 data. Errors in monitor initialization, downloading or equipment malfunction, led to the deletion of 30, 179, and 390 activities for the ActiGraph, Actical and RT3 analyses, respectively. Sample size and physical characteristics for the participants for these analyses are reported in Table 2.

Figure 1 illustrates the biases (predicted EE – measured EE) of each model across all activities, for treadmill activities combined and for activities of daily living combined. The models tend to underestimate EE, with activities of daily living being underestimated to a greater degree than treadmill activities. In general, the ActiGraph models were more accurate at estimating EE for lower intensity activities while the Actical models were slightly better at estimating EE for higher intensity activities. Both the RT3 gross and activity EE prediction models tend to underestimate activities of daily living and graded treadmill activities while level treadmill activities tended to be overestimated. Table 3–Table 5 report the bias (95% confidence interval) and the RMSE for each individual activity, for treadmill activities combined, for activities of daily living combined and across all activities combined.

For the ActiGraph MET prediction models (Table 3) RMSE ranged from 0.5 METs (Freedson; dishes, Swartz; dusting, Crouter; dusting) to 6.2 METs (Freedson; ascend stairs). Bias ranged from −5.9 METs (Freedson; ascend stairs) to 2.1 METs (Crouter; descend stairs). ActiGraph MET prediction models underestimated EE (negative bias) 72% of the time. For the ActiGraph kcal prediction models (Table 3) RMSE ranged from 1.4 kcals (Freedson; 1.34 m·sec^{-1} 3% gr and 1.56 m·sec^{-1} 3% gr) to 7.4 kcals (Freedson; ascend stairs). Bias ranged from −7.1 kcals (Freedson; ascend stairs) to 0.7 kcals (Freedson; 1.34 m·sec^{-1} 0% gr). ActiGraph kcal prediction models underestimated EE 81% of the time.

For the Actical MET prediction models (Table 4) RMSE ranged from 0.5 METs (Crouter; dusting) to 5.9 METs (Klippel & Heil; ascend stairs). Bias ranged from −5.7 METs (Klippel & Heil; ascend stairs) to 2.7 METs (Crouter; descend stairs). Actical MET prediction models underestimated EE (negative bias) 79% of the time. For the Actical kcal prediction models (Table 4) RMSE ranged from 0.7 kcals (Heil 1R and Heil 2R; 1.34 m·sec^{-1} 0% gr) to 8.0 kcals (Heil 1R; ascend stairs). Bias ranged from −7.6 kcals (Heil 1R; ascend stairs) to 0.9 kcals (Heil 1R; 2.23 m·sec^{-1} 0% gr). Actical kcal prediction models underestimated EE 85% of the time.

For the RT3 prediction models (Table 5) RMSE ranged from 1.0 kcals (RT3 Gross EE and RT3 Activity EE; dishes) to 7.9 kcals (Gross EE and RT3 Activity EE; ascend stairs). Bias ranged from −7.6 kcals (RT3 Gross EE and RT3 Activity EE; ascend stairs) to 2.1 kcals

(RT3 Gross EE; descend stairs). RT3 prediction models underestimated EE (negative bias) 73% of the time.

Figure 2 illustrates the rate at which each model misclassified activity intensity. Across all intensities, misclassification rates ranged from 21.7% (95% CI 20.4%, 24.2%) (kappa statistic = 0.57) to 34.3% (95% CI 32.3%, 36.3%) (kappa statistic = 0.40), with vigorous intensity activities being most often misclassified.

Discussion

Although accelerometers are used extensively to assess physical activity, there has yet to be a *comprehensive* independent validation of EE and MET prediction models that use accelerometer output as the predictor variable. Several studies have attempted to investigate their accuracy, however they are limited by small sample sizes that are not representative of the population. This study is unique due to its large diverse sample size, the wide range of activities performed, the use of three commercially available accelerometers and the **simultaneous** comparison of 11 discrete prediction models on data independent from which they were developed. Similar to previous research our findings indicate that the ActiGraph, Actical and RT3 do not produce accurate point estimates of EE across a broad range of activities (Crouter et al. 2006b; Rothney et al. 2008). Additionally, no equation is accurate at classifying activities across all intensities (light < 3 METs, moderate 3–5.99 METs, vigorous 6 METs), with vigorous intensity activities being the most frequently misclassified.

ActiGraph

The ActiGraph is the most commonly used activity monitor and numerous published prediction techniques have been used to translate activity counts to EE. The Freedson MET and kcal equations, developed in 1998, have been extensively studied and tend to underestimate activities of daily living and vigorous treadmill activities (Crouter et al. 2006b; Rothney et al. 2008). This under-prediction is likely due to the fact both equations were developed on a small sample, where participants performed only three treadmill activities. In the current study, we observed this under-prediction for the Freedson MET and kcal equations. The Freedson MET equation under-predicted EE for all ADL (Bias −2.0 METs; 95% CI −2.1, −1.9) and TRD (Bias −0.8 METs; 95% CI −0.8, −0.7) activities except slow (1.34 m·sec^{-1}) and medium paced walking (1.56 m·sec^{-1}) on level ground and descending the stairs. The Freedson MET model appears to be most accurate for predicting EE for level treadmill activities (range RMSE 0.6 to 1.8 METs) and light intensity ADLs that require minimal lower body movement (dishes, dusting, laundry) (range RMSE 0.6 to 0.9 METs).

To address the Freedson MET model's consistent underestimation of EE for moderate-vigorous treadmill activities and ADLs, researchers began developing prediction equations on a wider range of activities, including activities of daily living. Swartz et al. (2000) employed a protocol consisting of 2 over-ground walking and 26 lifestyle activities (including household and sport activities) to produce a new linear regression model. In our sample, this model improved MET estimates (compared to the Freedson MET model) for all ADLs combined, for all TRDs combined and across all activities combined. However, the increased accuracy was predominantly due to the improved estimates of moderate intensity activities, while low intensity activities were considerably overestimated. The y-intercept of this linear model is 2.606 indicating that at 0 counts (sedentary behavior) an individual's EE is 2.606 METs, about 1.5 METs higher than RMR. Thus, activities performed between 1–2.6 METs will always be overestimated. In the current study, only 3 activities, (dishes, laundry and dusting) had a measured EE less than 2.6 METs. If more sedentary-light

activities were tested we likely would have seen a higher rate of EE overestimation. This lack of sensitivity to changes in sedentary and light activity is of considerable importance given the recent evidence that most Americans spend more than half of their waking hours engaged in sedentary behavior (< 1.5 METs) (Mathews et al 2008) and the subsequent public health focus on reducing sedentary behavior as a means to reduce many chronic disease risk factors.

In addition to its consistent overestimation of light intensity activities, the Swartz model, like the Freedson MET model, underestimated vigorous intensity ADLs, such as basketball, tennis and ascending the stairs. Thus, using a wider range of activities in its calibration process, the Swartz model was successful at improving EE estimates for moderate intensity ADLs, while minimally improving estimates for vigorous intensity ADLs. Additionally, the use of such a large y-intercept (2.606 METs) virtually *eliminates* the possibility of accurately estimating sedentary-light intensity activities.

These data indicate that linear regression models perform well when evaluating activities similar to those from which they were developed and it appears that EE estimates could improve if different regression equations were used for activities that exhibit distinctive properties (e.g. movement patterns or intensities), such as rhythmic locomotion activities and unconstrained activities of daily living. This realization led to the development of a two-regression model in which the variability in accelerometer counts is used to determine the type of activity performed (Crouter et al. 2006a). Counts are then directed into either a lifestyle or locomotion equation to predict METs. In addition, the two-regression model employs an inactivity threshold which assigns a value of 1 MET to activities with an average count value of < 50 counts·min^{-1}. The inactivity threshold is meant to provide better estimates of the low intensity activities that are often overestimated by single linear regression models. Our data indicate this new approach that uses a feature of the signal output to direct counts to one of two equations improves EE estimation for all ADLs combined, compared to the traditional single linear regression techniques of Freedson et al. (1998) and Swartz et al (2000). Perhaps more promising than the improved EE estimation of ADLs, is the range of intensities that were accurately predicted. The Crouter method performed well for activities ranging from 2.5–8.3 METs. The improvement across a wider range of intensities is likely due to the non-linear cubic function used to estimate EE for lifestyle activities. Non-linear regressions use more free parameters to model the relationship between counts and EE; they do not assume a single, "straight line" relationship across a range of intensities. On the other hand, the exponential curve used to estimate EE for locomotion activities did not improve EE estimates across all treadmill activities combined (RMSE 1.7 METs) compared to the Freedson and Swartz MET prediction equations (RMSE 1.5 and 1.3 METs, respectively). There are often two problems associated with more complicated, non-linear relationships such as exponential or cubic models. They sometimes do not transport to other data sets as well as simpler models, and they often do not extrapolate well to activities that are outside the range of counts from which they were developed. Despite the added challenges of a more complicated model, and its poor performance on treadmill activities, the Crouter method shows promise for distinguishing locomotion and lifestyle activities, as well as accurately estimating EE across a range of intensities.

Actical

Similar to the Actigraph, Actical prediction techniques tend to underestimate EE across a range of activities, with activities of daily living being considerably more underestimated compared to treadmill activities. Three of the five Actical prediction models evaluated are two-regression models. The Klippel & Heil 2R MET model and the Heil 2R kcal model, are two-regression models that were developed in an attempt to improve the single regression

predictions of EE (Klippel & Heil 1R and Heil 1R) across a range of activity intensities. These two-regression models use activity "intensity" to direct accelerometer counts to one of two regressions. This technique seems reasonable given that most prediction models are fairly accurate at predicting EE for activities within a narrow intensity range. Theoretically, if counts are directed to a regression model that is better suited to predict EE for that specific intensity range, an improvement in the accuracy of EE estimation should be observed. However, there is an inherent problem with both the Klippel & Heil 2R MET model and the Heil 2R kcal model – both models use count cut-points to distinguish activity intensity. In the current study, the average $count \cdot min^{-1}$ for raking was 202.8, while the average $count \cdot min^{-1}$ for descending the stairs was 3245, however these two activities have very similar average energy expenditure values, 5.2 and 5.0 $kcal \cdot min^{-1}$, respectively. These data clearly demonstrate that two activities of similar intensity can have drastically different count values due to the nature of the activities. Based on their count values these two activities would be classified as different intensities and directed to different prediction equations, resulting in inaccurate estimates of EE. Due to these limitations, the 2R models did not improve EE estimates compared to the 1R models. The Klippel & Heil 2R MET model improved EE estimates by an average of 0.1 METs across all activities, while the Heil 2R kcal model improved EE estimates by an average of 0.1 kcals across all activities. These data further illustrate the limitations of static regression models and their inability to accurately estimate EE across a range of activity intensities.

Similar to the Crouter two regression model used for the ActiGraph, the Crouter Actical two-regression model performed well across a broader range of intensities compared to other single and two-regression models. Again, this is likely due to the use of two non-linear regressions to model the relationship between counts and EE instead of assuming a single linear relationship. The Crouter Actical model was slightly more accurate at estimating EE for ADLs (RMSE 2.5 METs) compared to Klippel & Heil 1R and 2R MET prediction equations (RMSE 3.0 and 2.9 METs, respectively). However the Crouter model was considerably less accurate for TRD activities (RMSE 2.1 METs) compared to Klippel & Heil 1R and 2R MET prediction equations (both have an RMSE of 1.1 METs).

RT3

The RT3 activity monitor has not been studied as extensively as other commercially available monitors. The prediction equations most often used, and those examined in this study, are proprietary and can only be used through the RT3 software. Thus it is not possible to ascertain specific features of the equation or its development. It is also important to note that the technical specifications of the RT3 are considerably different than both the ActiGraph and the Actical. It is a tri-axial accelerometer, sensitive to acceleration in all three planes of movement. As a result, during a given activity or situation, the RT3 has the potential to register a much larger degree of acceleration. One would expect that these specification differences could improve the underestimation of EE exhibited by other monitors. Conversely, the RT3 models significantly underestimated EE during ADLs (Gross EE Bias; −1.6 kcals; 95% CI −1.8, −1.4: RMSE 3.8 kcals) and (Activity EE Bias; −1.7 kcals; 95% CI −1.9, −1.5: RMSE 3.8 kcals). Overall, the RT3 models overestimated treadmill activities (Gross EE Bias; 0.5 kcals; 95% CI 0.4, 0.7: RMSE 1.8 kcals) (Activity EE Bias; 0.5 kcals; 95% CI 0.3, 0.6: RMSE 1.8 kcals). However this overestimation was predominantly due to the large overestimation of *level* treadmill activities, while *graded* treadmill activities remained underestimated. This trend is similar to what is seen in both the ActiGraph and Actical accelerometers. In the current study, the RT3 tri-axial accelerometer does not appear to improve estimates of EE. This could be due to factors in the calibration process, the precision of the accelerometer, or factors related to the multi-axis monitor.

Previous findings suggest that contributions from each axis are not accurately represented in the 3-axes quantification of acceleration (Howe et al. 2009).

Activity Intensity Classification

In surveillance research, researchers are often not interested in point estimates of EE, but rather how well the monitor output distinguishes among light (< 3 METs), moderate (3–5.99 METs) or vigorous (6 METs) intensities. Large scale epidemiologic studies, including the National Health and Nutrition Examination Study (NHANES), are increasingly relying on accelerometers as an objective measurement of physical activity (Troiano et al. 2007). Often the primary goal of such studies is to understand an individual's habitual physical activity level. By understanding an individual's habitual physical activity a number of research and clinical outcomes can be elucidated, such as if an individual is meeting the physical activity guidelines, the health outcomes associated with a specific dose of physical activity, or an individual's compliance to a specific lifestyle intervention.

Similar to previous studies (Crouter et al. 2006b, Rothney et al. 2008) however, this study found that no prediction technique, for any monitor, accurately classifies physical activity across all intensity categories. Figure 2 illustrates the rate of activity intensity misclassification. Moderate intensity activity was the least often misclassified (range; 8.9–34.3%), while vigorous activity was most often misclassified (range; 28.2–54.5%). The higher rate of vigorous activity misclassification was likely due to a number of factors. Using current prediction techniques, a single accelerometer positioned on an individual's hip, 1) does not sufficiently account for the EE produced by upper body movements, 2) is not able to differentiate the terrain on which an individual is moving and thus cannot account for the increased EE associated with walking at an incline or ascending stairs, and 3) is not able to account for the increased EE associated with carrying a load. Additionally, sedentary and light intensity activities were often classified as moderate intensity. This error is due to prediction equations having a y-intercept as high as 2.6, meaning that at 0 counts (no acceleration), estimated EE is 2.6 METs. The insensitivity in distinguishing between sedentary/light and moderate intensities is important given the recent focus on decreasing sedentary behaviors and accumulating short bouts of moderate activity as a means to elicit health benefits (PAGAC 2008; Healy et al. 2008). Equations with a lower y-intercept were more sensitive to light intensity activity; however they tended to underestimate time spent in moderate and vigorous intensities. This inconsistency illustrates the persistent challenge of accurately predicting and classifying EE across a broad range of activity types and intensities with current accelerometer prediction techniques.

Standard 3.5ml·kg^{-1}·min^{-1} vs. Measured RMR

It is important to point out the errors reported in this paper are not due to our method of analysis; using the standard 3.5 ml·kg^{-1}·min^{-1} to establish criterion METs from oxygen consumption data. Data show the standard 3.5 ml·kg^{-1}·min^{-1} is not an accurate estimate of RMR for certain subgroups of the population (e.g. overweight) (Byrne et al. 2005; Kozey et al. 2010). This issue is important, and there are potential benefits for using measured RMR when future studies develop new prediction methods. However, the models assessed in the present study were *developed* using the standard 3.5 ml·kg^{-1}·min^{-1}. Although the use of measured RMR in the calibration process may improve prediction models, the purpose of the current paper was to validate existing methods. When we performed additional analyses to compare the predicted METs to METs calculated using measured RMR, performance deteriorates; when using the measured RMR, RMSE always increases, and it increases by 0.37 METs on average. This is perhaps not surprising since prediction models perform best when used in a manner that closely resembles their calibration. That said, it is important not

to lose sight of the fact that the original methods were calibrated using the 3.5 ml·kg^{-1}·min^{-1} standard, which is scientifically somewhat suspect.

Summary

Since 1998 and Freedson et al's initial calibration study, accelerometer prediction models have continuously evolved in an attempt to improve EE estimates. Each generation of prediction models appears to address one or more flaws inherent to its previous model, only to create or fail to account for additional errors. The Swartz model (2000) addressed the underestimation of lifestyle activities by the Freedson model (1998), but any observed improvements were at the expense of overestimating low intensity activities. Two-regression models, such as those by Klippel and Heil (2003) and Heil et al. (2006), attempted to improve estimates across a range intensities from light to vigorous by using one regression for light activities and a different regression for moderate-vigorous activities. These models however, relied on count·min^{-1} to determine intensity, a method inherently flawed and described in detail above. Crouter et al (2006a) and (2008) recognized this flaw and used a more sophisticated feature of the acceleration signal (coefficient of variation [CV]) to distinguish locomotion and lifestyle activities and direct count·min^{-1} to either a lifestyle or locomotion specific equation. Crouter et al (2006a) and (2008) also attempted to improve estimates of METs by using more complex non-linear regressions. Although Crouter's method appears to be successful at determining locomotion and lifestyle activities by using the count CV, the use of more complex regressions may limit this technique's validity when applied to independent data sets. Figure 3 summarizes these errors and the progression of prediction techniques from 1998 to the present.

We believe the underlying cause of the limitations noted above, is the fact that current techniques use a single integrated accelerometer signal averaged over time as the sole input into accelerometer prediction equations. In other words, the rich features of the signal are not used, thus patterns of movement are not considered in the translation of accelerometer counts to energy expenditure. For example, a treadmill activity performed for 10 minutes at a steady intensity could result in the same accelerometer output as a lifestyle activity (performed for the same time) that requires variable movement patterns. For these activities the acceleration signal is very different, but when averaged over time, produces a similar accelerometer output.

Figure 4 shows sample data from one subject. Second-by-second counts for level walking (1.34 m·sec^{-1}) (panel 1) and moving boxes (panel 2) are averaged over 7 minutes (shown in the solid gray line). Despite very different second-by-second data, these activities produce very similar counts·min^{-1}, 2198.5 and 2204.7, respectively. As a result, both activities will yield similar estimates of EE and classifications of activity intensity. It is clear that walking on a treadmill and performing a lifestyle activity such as moving boxes produce very different patterns of acceleration, however current regression techniques fail to recognize and model these differences.

Staudenmayer et al. (2009) has begun to address these issues by developing two artificial neural networks to estimate METs and identify activity type using more complex features of the accelerometer signal. Although this technique shows promise for substantially improving accelerometer based physical activity measurement (Staudenmayer et al 2009), it was developed using a relatively small sample of subjects (the subjects and data from Crouter et al, 2006a) and further refinement and development is ongoing.

Limitations

This study is influenced by one main limitation – activities were not performed in a free-living environment. The appeal of accelerometers is that they offer a minimally burdensome means to objectively measure physical activity during free-living conditions. Although the pace of activities was self-selected (except treadmill activities) and participants were encouraged to perform activities as they would in their "everyday lives," they were not performed in a true free-living environment. Activities were performed for 7-minutes each, allowing participants to reach steady state. In a true free-living environment the time spent in each activity is likely much more variable, with some activities lasting only a few seconds. As a result, much of free-living activity is not performed under steady state conditions, and because current accelerometer prediction techniques use average counts·min^{-1} to estimate EE, it is reasonable to assume that these techniques would perform even more poorly under free-living conditions. Testing accelerometers and prediction models in a true free-living environment would shed light on field-based research. However the procedures necessary to conduct such a validation remain complicated and sometimes impractical. This study attempted to bridge the gap by creating "free-living" activities within a laboratory setting.

The issue of how to handle 0 counts was not assessed in the present paper. At 0 counts, the device is registering no acceleration, and thus it is likely that the participant is seated, is involved in some sort of sedentary behavior that requires no movement, or has removed the monitor. As stated earlier, however, some regression models have y-intercepts as high as 2.6, meaning sedentary and some light intensity activities are considerably overestimated. The existing methods (and future methods) would benefit from explicitly adjusting their methodology to address 0 counts. In fact, some researchers have developed their own *ad hoc* methods to handle 0 counts (Matthews et al 2008), but we are unaware of an established methodology that has been empirically derived and is consistently used. As a result, evaluating such a method in combination with published prediction techniques is beyond the scope of this report. Users of the existing methods are advised to use caution when dealing with 0 counts and to consider the specific information sought and the population being assessed when deciding how to handle 0 counts.

Conclusion

In summary, current prediction techniques tend to underestimate energy expenditure, with the underestimation being greater for ADLs than TRD activities. Additionally, this study highlights the tendency of current prediction techniques to perform well within a specified range of intensities and/or specific activity types. These ranges are often indicative of the activities from which equation was developed. Similarly, current prediction equations are not accurate at classifying activity intensity, with vigorous intensity activity being most often misclassified.

In conclusion, accelerometers are a promising tool to objectively measure physical activity, however current data processing techniques fail to realize the potential of accelerometers for providing accurate estimates of energy expenditure and estimation of time spent in light, moderate, and vigorous intensities. This investigation illustrates the numerous limitations of current regression techniques when translating accelerometer output to physiologically meaningful energy expenditure metrics, including 1) the fixed, single relationship assumed between counts and EE when using linear regression models, 2) the insensitivity of these models to accurately distinguish sedentary and light activities, 3) the insufficient translation of regression models, especially non-linear models, to data sets independent from the development data set and 4) the reliance on a single integrated accelerometer signal averaged over time and subsequent elimination of the rich features of the signal. Future

research should focus on developing more sophisticated data processing techniques to estimate energy expenditure from accelerometer output.

Acknowledgments

References

1. American College of Sports Medicine. ACSM's resource manual for guidelines for exercise testing and prescription. Philadelphia: Lippincott Williams & Wilkens; 2009.

2. Bassett DR Jr, Ainsworth BE, Swartz AM, Strath SJ, O'Brien WL, King GA. Validity of four motion sensors in measuring moderate intensity physical activity. Med Sci Sports Exerc. 2000; 32 Suppl:S471–S480. [PubMed: 10993417]

3. Byrne NM, Hills AP, Hunter GR, Weinsier RL, Schultz Y. Metabolic equivalent: one size does not fit all. J Appl Physiol. 2005; 99(3):1112–1119. [PubMed: 15831804]

4. Cavagna GA, Thys H, Zamboni A. The sources of external work in level walking and running. J Physiol. 1976; 262:639–657. [PubMed: 1011078]

5. Crouter SE, Clowers KG, Bassett DR Jr. A novel method for using accelerometer data to predict energy expenditure. J Appl Physiol. 2006a; 100:1324–1331. [PubMed: 16322367]

6. Crouter SE, Churilla JR, Bassett DR Jr. Estimating Energy Expenditure using accelerometers. Eur J Appl Physiol. 2006b; 98:601–612. [PubMed: 17058102]

7. Crouter SE, Bassett DR Jr. A new two-regression model for the Actical accelerometer. Br. J. Sports Med. 2008; 42:217–224. [PubMed: 17761786]

8. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc. 1998; 30:777–781. [PubMed: 9588623]

9. HeltheTech. MedGem User Manual. Golden, CO; 2003. p. 3-7.

10. Healy GN, Wijndaele K, Dunstan DW, Shaw JE, Salmon J, Zimmet PZ, Owen N. Objectively measured sedentary time, physical activity, and metabolic risk: the Australian Diabetes, Obesity and Lifestyle Study (AusDiab). Diabetes Care. 2008; 31(2):369–371. [PubMed: 18000181]

11. Heil DP. Predicting activity energy expenditure using the Actical activity monitor. Res Q Exerc Sport. 2006; 77:64–80. [PubMed: 16646354]

12. Howe CA, Staudenmayer JW, Freedson PS. Accelerometer prediction of energy expenditure: Vector magnitude vs. vertical axis. Med Sci Sports Exerc. 2009; 41(12):2199–2206. [PubMed: 19915498]

13. Klippel NJ, Heil DP. Validation of energy expenditure prediction algorithms in adults using the Actical electronic activity monitor. Med Sci Sports Exerc. 2003; 35:S284.

14. Kozey SL, Lyden K, Staudenmeyer JW, Freedson PS. Errors of MET estimates of physical activities using 3.5 ml·kg^{-1}·min^{-1} as the baseline oxygen consumption. J Physical Activity and Health. 2010; 7(4):508–516.

15. Matthews CE. Calibration of accelerometer output for adults. Med Sci Sports Exerc. 2005; 37 Suppl:S512–S522. [PubMed: 16294114]

16. Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR, Troiano RP. Amount of time in sedentary behaviors in the United States, 2003–2004. AM J Epidemiol. 2008; 167(7): 875–881. [PubMed: 18303006]

17. Nieman DC, Trone GA, Austin MD. A new handheld device for measuring resting metabolic rate and oxygen consumption. J Am Diet Assoc. 2003; 103(5):588–592. [PubMed: 12728217]

18. Perret C, Mueller G. Validation of a new portable ergospirometric device (Oxycon Mobile) during exercise. Int J Sports Med. 2006; 27(5):363–367. [PubMed: 16729377]

19. Physical Activity Guidelines Advisory Committee, 2008. Physical Activity Guidelines Advisory Committee Report. Washington, DC: U.S. Department of Health and Human Services; 2008.

20. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2009. ISBN 3-900051-07-0, URL http://www.R-project.org

21. Rosdahl H, Gullstrand L, Salier-Eriksson J, Johansson P, Schantz P. Evaluation of the Oxycon Mobile metabolic system against the Douglas bag method. Eur J Appl Physiol. 2009 published ahead-of-print.

22. Rothney MP, Schaefer EV, Neumann MM, Choi L, Chen KY. Validity of physical activity intensity predictions by ActiGraph, Actical and RT3. Obesity. 2008; 16(8):1946–1952. [PubMed: 18535553]

23. Sallis JF, Saelens BE. Assessment of physical activity by self-report: Status, limitations and future directions. Res Q Exerc Sport. 2000; 71(2):1–14. [PubMed: 10763516]

24. Stay Healthy Inc. RT3 User Manuel; Version 1.2. Available via www.stayhelthy.com

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):1300–1307. [PubMed: 19644028]

26. Swartz AM, Strath SJ, Basset DR Jr, 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 Suppl:S450–S456. [PubMed: 10993414]

27. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc. 2008; 40(1):180–188.

28. Weir, JBde. New methods for calculating metabolic rate with special reference to protein metabolism. J Physiol. 1949; 109:1–9. [PubMed: 15394301]

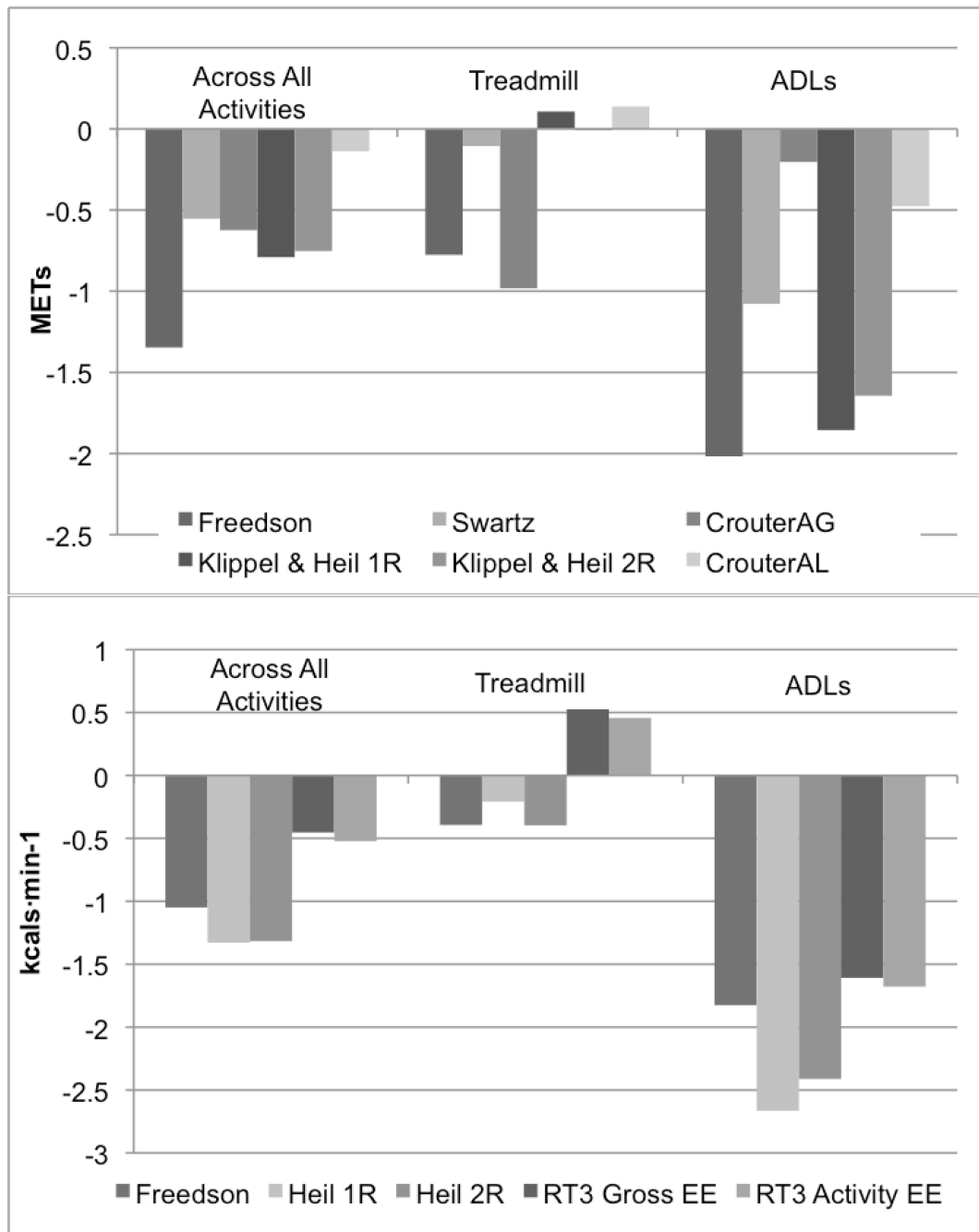**Fig. 1.**
Panel 1 shows the bias for MET prediction equations across all activities, for TRDs and ADLs. Panel 2 shows the bias for kcal prediction equations across all activities, for TRDs and ADLs. All predicted EE values were significantly different from indirect calorimetry except when treadmill activities were analyzed using the Klippel and Heil (2003) 2R MET equation
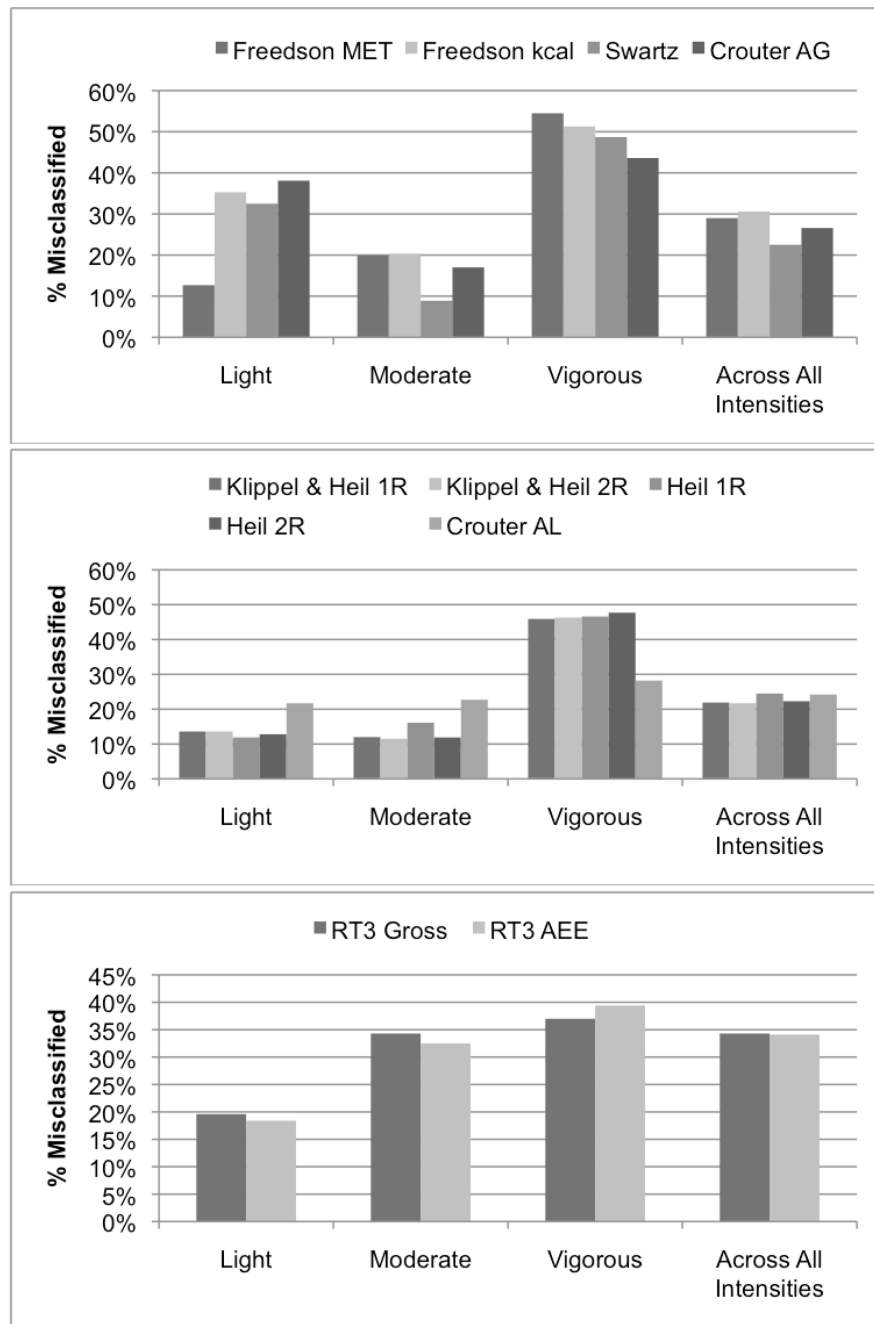
Fig. 2.
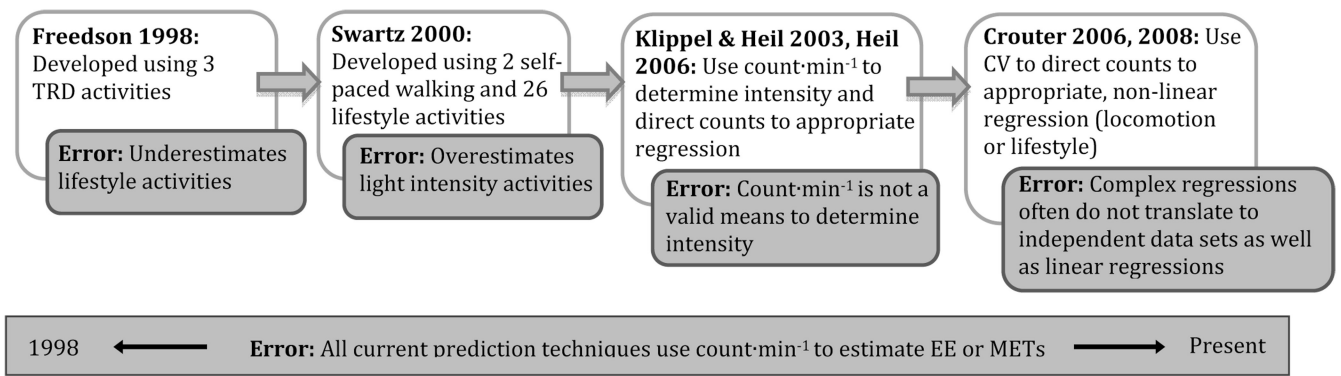Activity intensity misclassification for ActiGraph, Actical and RT3 prediction equations, respectively

Freedson 1998: Developed using 3 TRD activities

Error: Underestimates lifestyle activities

Swartz 2000: Developed using 2 self-paced walking and 26 lifestyle activities

Error: Overestimates light intensity activities

Klippel & Heil 2003, Heil 2006: Use count·min^{-1} to determine intensity and direct counts to appropriate regression

Error: Count·min^{-1} is not a valid means to determine intensity

Crouter 2006, 2008: Use CV to direct counts to appropriate, non-linear regression (locomotion or lifestyle)

Error: Complex regressions often do not translate to independent data sets as well as linear regressions

1998 ⟵ **Error:** All current prediction techniques use count·min^{-1} to estimate EE or METs ⟶ Present

Fig. 3.
Summary of accelerometer energy expenditure prediction equations from 1998 to present.
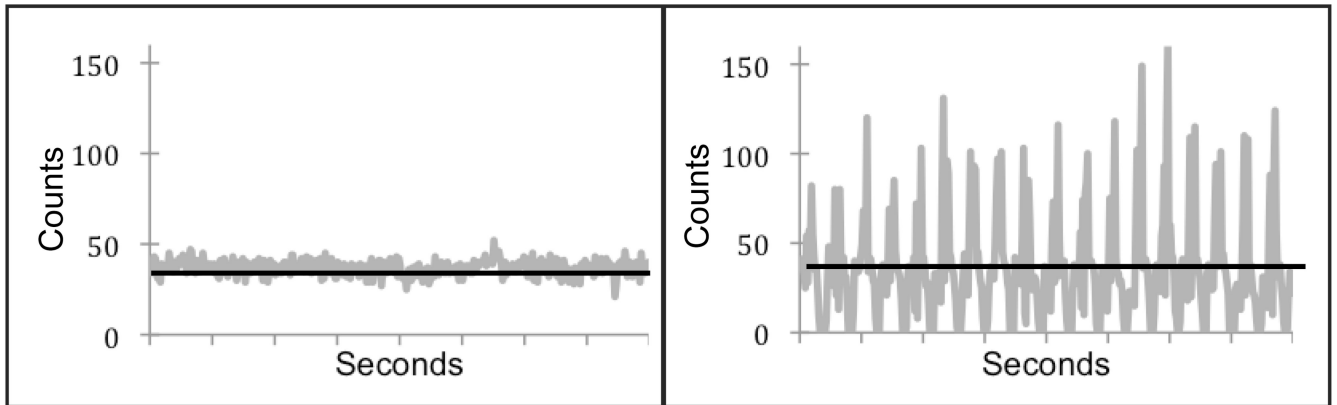
Fig. 4.
Counts·sec^{-1} for level walking (Panel 1) and moving boxes (Panel 2) over 7 minutes. When averaged these data produce similar counts·min^{-1}, 2198.5 and 2204.7, respectively.

Table 1

Prediction Models

Prediction Model	N	Accelerometer	Equation	EE Metric Predicted
Freedson et al. 1998	50	ActiGraph	$1.439008 + (0.000795 \times \text{cnts·min}^{-1})$	METs
Freedson et al. 1998	50	ActiGraph	$(0.00094 \times \text{cnts·min}^{-1}) + (0.1346\ BW) - 7.37418$	kcals
Swartz et al. 2000	70	ActiGraph	$2.606 + (0.0006863 \times \text{cnts·min}^{-1})$	METs
Crouter et al. 2006	48	ActiGraph	**cnts·min^{-1} ≤ 50; EE = 1 MET** **50 < cnts·min^{-1} and CV ≤ 10;** $2.379833 \times (\exp(0.00013529 \times \text{cnts·min}^{-1}))$ **50 < cnts·min^{-1} and CV = 0 or > 10;** $2.330519 + (0.001646 \times \text{cnts·min}^{-1}) - (1.2017 \times 10^{-7} \times (\text{cnts·min}^{-1})^2) + (3.3779 \times 10^{-12} \times (\text{cnts·min}^{-1})^3)$	METs
Klippel and Heil 2003 (1R)	24	Actical	**cnts·min^{-1} ≤ 50; EE = 1 MET** **50 < cnts·min^{-1} < 350;** EE = 1.83 METs **350 < cnts·min^{-1};** $2.826 + (0.0006526 \times \text{cnts·min}^{-1})$	METs
Klippel and Heil 2003 (2R)	24	Actical	**cnts·min^{-1} ≤ 50; EE = 1 MET** **50 < cnts·min^{-1} < 350;** EE = 1.83 METs **350 < cnts·min^{-1} < 1200;** $1.935 + (0.003002 \times \text{cnts·min}^{-1})$ **1200 < cnts·min^{-1};** $2.768 + (0.0006397 \times \text{cnts·min}^{-1})$	METs
Heil et al. 2006 (1R)	24	Actical	**50 < cnts·min^{-1} < 350;** EE = 0.007565 kcals·min^{-1} **350 < cnts·min^{-1};** $0.02779 + ((1.143\text{E-}5) \times \text{cnts·min}^{-1})$	kcals
Heil et al. 2006 (2R)	24	Actical	**50 < cnts·min^{-1} < 350;** EE = 0.007565 kcals·min^{-1} **350 < cnts·min^{-1} < 1200;** $0.01217 + ((5.268\text{E-}5) \times \text{cnts·min}^{-1})$ **1200 < cnts·min^{-1};** $0.02663 + ((1.107\text{E-}5) \times \text{cnts·min}^{-1})$	kcals
Crouter et al. 2006	48	Actical	**cnts·min^{-1} ≤ 10; EE = 1 MET** **10 < cnts·min^{-1} and CV ≤ 13;** $2.55095 \times (\exp(0.00013746 \times \text{cnts·min}^{-1}))$ **10 < cnts·min^{-1} and CV > 13;** $1.466072 + 0.210755 \times (\text{Ln}(\text{cnts·min}^{-1})) - 0.0595362 \times (\text{Ln}(\text{cnts·min}^{-1})^2) + 0.0157002 \times (\text{Ln}(\text{cnts·min}^{-1})^3)$	METs
RT3 Gross Proprietary (9)	Unknown	RT3	Proprietary	kcals
RT3 Activity EE Proprietary (9)	Unknown	RT3	Proprietary	kcals

N = participants used in development of model; EE = energy expenditure; cnts = counts; CV = Coefficient of Variation

Table 2

Physical Characteristics of Participants (mean ± SD (range)).

	All *N = 274*	Men *N = 135*	Women *N = 139*
Age *(years)*	38.3 ± 12.4 (20–60)	37.8 ± 12.7 (20–60)	38.8 ± 12.2 (20–58)
Body Mass *(kg)*	72.6 ± 14.8 (47.0–130.0)	79.2 ± 13.3 (52.3–130.0)	66.1 ± 13.2 (47.0–123.6)
Height *(cm)*	170.8 ± 9.6 (149.0–193.5)	177.5 ± 6.8 (160.0–193.5)	164.2 ± 7.1 (149.0–190.5)
BMI *(kg/m²)*	24.8 ± 4.2 (17.6–42.2)	25.1 ± 3.7 (17.6–42.2)	24.5 ± 4.5 (18.6–41.8)
Activities Analyzed	2745	1377	1368

BMI = Body Mass Index

Table 3

ActiGraph Bias (95% CI); RMSE

	Freedson (MET)		Swartz (MET)		Crouter (MET)		Freedson (kcal)	
	Bias (95% CI)	RMSE	Bias (95% CI)	RMSE	Bias (95% CI)	RMSE	Bias (95% CI)	RMSE
Across All Activities	−1.4 (−1.4, −1.3)	2.3	−0.6 (−0.6, −0.5)	2.0	−0.6 (−0.7, −0.6)	2.0	−1.1 (−1.2, −1.0)	2.9
Treadmill Activities	−0.8 (−0.8, −0.7)	1.5	−0.1 (−0.2, −0.1)	1.3	−1.0 (−1.1, −0.9)	1.7	−0.4 (−0.5, −0.3)	1.9
Activities of Daily Living	−2.0 (−2.1, −1.9)	3.1	−1.1 (−1.2, −0.9)	2.6	−0.2 (−0.3, −0.1)	2.3	−1.8 (−2.0, −1.7)	3.7
Walking 1.34 msec^{-1} 0% gr	0.0* (−0.1, 0.1)	0.6	0.8 (0.8, 0.9)	1.0	0.0* (−0.1, 0.1)	0.9	0.7 (0.5, 0.9)	1.6
Walking 1.56 msec^{-1} 0% gr	0.0* (−0.1, 0.1)	0.9	0.7 (0.6, 0.8)	1.1	−0.3 (−0.4, −0.2)	1.0	0.6 (0.5, 0.8)	1.6
Running 2.23 msec^{-1} 0% gr	−1.1 (−1.3, −1.0)	1.8	−0.8 (−0.9, −0.6)	1.5	−1.4 (−1.7, −1.2)	2.2	−1.2 (−1.4, −0.9)	2.0
Walking 1.34 msec^{-1} 3% gr	−0.8 (−0.9, −0.7)	1.1	0.1 (0.0, 0.1)	0.7	−0.9 (−1.0, −0.8)	1.3	−0.2 (−1.0, −0.8)	1.4
Walking 1.56 msec^{-1} 3% gr	−1.0 (−1.1, −0.9)	1.3	−0.2 (−03, −0.1)	0.8	−1.3 (−1.4, −1.2)	1.6	−0.6 (−0.7, −0.4)	1.4
Running 2.23 msec^{-1} 3% gr	−2.4 (−2.5, −2.2)	2.7	−2.0 (−2.2, −1.8)	2.3	−2.7 (−2.9, −2.4)	3.1	−2.7 (−2.9, −2.5)	3.1
Ascend Stairs	−5.9 (−6.2, −5.7)	6.2	−5.1 (−5.3, −4.9)	5.3	−3.9 (−4.1, −3.6)	4.3	−7.1 (−7.4, −6.8)	7.4
Basketball	−3.2 (−3.6, −2.7)	3.5	−2.5 (−3.0, −2.0)	3.0	−0.8 (−1.2, −0.3)	1.7	−3.3 (−4.0, −2.7)	4.0
Move 6kg Box	−1.4 (−1.5, −1.3)	1.6	−0.4 (−0.5, −0.3)	1.0	0.8 (0.7, 0.9)	1.2	−0.9 (−1.1, −0.7)	1.9
Descend Stairs	0.0* (−0.2, 0.1)	1.1	0.8 (0.7, 0.9)	1.3	2.1 (1.9, 2.2)	2.5	0.6 (0.4, 0.8)	1.8
Dishes	−0.4 (−0.5, −0.3)	0.5	0.7 (0.6, 0.8)	0.8	−0.8 (−0.9, −0.7)	0.9	0.4* (−0.1, 0.9)	1.7
Dust	−0.8 (−1.0, −0.7)	0.9	0.3 (0.2, 0.4)	0.5	0.3 (0.2, 0.5)	0.5	−0.4* (−0.8, 0.1)	1.5

	Freedson (MET)		Swartz (MET)		Crouter (MET)		Freedson (kcal)	
	Bias (95% CI)	RMSE	Bias (95% CI)	RMSE	Bias (95% CI)	RMSE	Bias (95% CI)	RMSE
Garden	−1.2 (−1.5, −0.9)	1.5	−0.2* (−0.4, 0.1)	0.9	0.5 (0.2, 0.8)	1.0	−0.6 (−1.2, −0.1)	1.8
Laundry	−0.7 (−0.8, −0.6)	0.8	0.4 (0.3, 0.5)	0.6	0.0 (−0.3, 0.2)	0.8	0.1* (−0.4, 0.6)	1.6
Mop	−1.5 (−1.8, −1.3)	1.7	−0.5 (−0.7, −0.2)	0.8	−0.2 (−0.4, 0.1)	0.9	−1.0 (−1.6, −0.4)	2.1
Mow	−2.3 (−2.5, −2.0)	2.4	−1.3 (−1.6, −1.0)	1.6	−0.2 (−0.5, 0.1)	1.0	−2.1 (−2.6, −1.7)	2.6
Paint	−0.9 (−1.1, −0.7)	1.1	0.1 (0.0, 0.3)	0.6	0.5 (0.3, 0.7)	0.9	−0.4* (−1.0, 0.2)	1.9
Rake	−2.2 (−2.5, −1.9)	2.4	−1.1 (−1.4, −0.8)	1.4	−0.8 (−1.1, −0.5)	1.2	−1.8 (−2.4, −1.2)	2.6
Clean Room	−0.7 (−1.1, −0.2)	1.5	0.1* (−0.3, 0.5)	1.3	1.7 (1.3, 2.2)	2.2	−0.1* (−0.8, 0.6)	2.2
Sweep	−1.2 (−1.4, −1.0)	1.4	−0.1* (−0.3, 0.1)	0.6	0.0 (−0.2, 0.3)	0.8	−0.6 (−1.0, −0.1)	1.6
Tennis	−4.9 (−5.3, −4.4)	5.1	−4.1 (−4.5, −3.6)	4.3	−2.4 (−2.9, −2.0)	2.8	−5.5 (−6.0, −5.0)	5.8
Trim	−1.5 (−1.7, −1.3)	1.6	−0.4 (−0.5, −0.2)	0.7	−0.7 (−1.0, −0.5)	1.1	−1.2 (−1.7, −0.7)	2.0
Vacuum	−1.3 (−1.5, −1.2)	1.4	−0.2 (−0.4, −0.1)	0.6	0.0 (−0.2, 0.2)	0.6	−1.1 (−1.5, −0.6)	1.7

CI = Confidence Interval; RMSE = Root Mean Squared Error; m = meters; sec = seconds; gr = grade;

*
Predicted EE not significantly different than measured EE

Table 4

Actical Bias (95% CI); RMSE

	Klippel & Heil 1R (MET)		Klippel & Heil 2R (MET)		Crouter AC (MET)		Heil 1R (kcals)		Heil 2R (kcals)	
	Bias (95% CI)	RMSE	Bias (95% CI)	RMSE	Bias (95% CI)	RMSE	Bias (95% CI)	RMSE	Bias (95% CI)	RMSE
Across All Activities	−0.8 (−0.9, −0.7)	2.2	−0.8 (−0.8, −0.7)	2.1	−0.1* (−0.2, 0.0)	2.3	−1.3 (−1.4, −1.2)	2.9	−1.3 (−1.4, −1.2)	2.8
Treadmill Activities	0.1 (0.1, 0.2)	1.1	0.0* (−0.1, 0.1)	1.1	0.2 (0.0, 0.3)	2.1	−0.2 (−0.3, −0.1)	1.3	−0.4 (−0.5, −0.3)	1.3
Activities of Daily Living	−1.9 (−2.0, −1.7)	3.0	−1.6 (−1.8, −1.5)	2.9	−0.5 (−0.6, −0.3)	2.5	−2.7 (−2.8, −2.5)	4.0	−2.4 (−2.6, −2.2)	3.9
Walking 1.34 msec⁻¹ 0% gr	0.4 (0.4, 0.5)	0.7	0.4 (0.3, 0.5)	0.7	0.0* (−0.2, 0.1)	1.1	0.2 (0.2, 0.3)	0.7	0.2 (0.1, 0.2)	0.7
Walking 1.56 msec⁻¹ 0% gr	0.4 (0.3, 0.5)	0.8	0.3 (0.2, 0.4)	0.8	−0.2 (−0.4, −0.1)	1.2	0.2 (0.1, 0.3)	0.9	0.0* (−0.1, 0.1)	0.8
Running 2.23 msec⁻¹ 0% gr	1.1 (0.9, 1.2)	1.8	0.9 (0.7, 1.0)	1.6	2.6 (2.3, 2.9)	3.6	0.9 (0.6, 1.1)	2.0	0.5 (0.3, 0.8)	1.8
Walking 1.34 msec⁻¹ 3% gr	−0.5 (−0.5, −0.4)	0.7	−0.5 (−0.6, −0.5)	0.8	−1.0 (−1.1, −0.9)	1.4	−0.9 (−1.0, −0.8)	1.1	−1.0 (−1.1, −0.9)	1.2
Walking 1.56 msec⁻¹ 3% gr	−0.6 (−0.7, −0.5)	0.9	−0.7 (−0.8, −0.6)	1.0	−1.3 (−1.4, −1.1)	1.7	−1.1 (−1.2, −1.0)	1.3	−1.2 (−1.3, −1.1)	1.5
Running 2.23 msec⁻¹ 3% gr	−0.1* (−0.3, 0.2)	1.4	−0.2* (−0.5, 0.0)	1.4	1.6 (1.3, 2.0)	3.0	−0.5 (−0.7, −0.2)	1.8	−0.8 (−1.1, −0.6)	1.9
Ascend Stairs	−5.7 (−5.9, −5.5)	5.9	−5.4 (−5.6, −5.2)	5.7	−4.0 (−4.2, −3.7)	4.4	−7.6 (−7.9, −7.3)	8.0	−7.3 (−7.6, −7.0)	7.7
Basketball	−2.9 (−3.3, −2.4)	3.2	−2.9 (−3.4, −2.5)	3.3	−0.7 (−1.3, −0.1)	1.9	−4.1 (−4.7, −3.6)	4.5	−4.2 (−4.8, −3.7)	4.6
Move 6kg Box	−1.2 (−1.3, −1.1)	1.4	−0.5 (−0.6, −0.4)	1.2	0.1 (0.0, 0.3)	1.2	−1.7 (−1.9, −1.6)	2.0	−0.9 (−1.0, −0.7)	1.6
Descend Stairs	1.0 (0.8, 1.1)	1.3	0.9 (0.8, 1.0)	1.3	2.7 (2.5, 3.0)	3.2	0.8 (0.7, 1.0)	1.4	0.7 (0.6, 0.8)	1.3
Dishes	−0.8 (−0.9, −0.7)	0.9	−0.8 (−0.9, −0.7)	0.9	−0.8 (−0.9, −0.7)	0.8	−1.2 (−1.3, −1.0)	1.2	−1.2 (−1.3, −1.0)	1.2
Dust	−1.1 (−1.3, −1.0)	1.2	−1.1 (−1.3, −1.0)	1.2	−0.1* (−0.3, 0.0)	0.5	−1.7 (−1.9, −1.5)	1.8	−1.7 (−1.9, −1.5)	1.8

	Klippel & Heil 1R (MET)		Klippel & Heil 2R (MET)		Crouter AC (MET)		Heil 1R (kcals)		Heil 2R (kcals)	
	Bias (95% CI)	RMSE	Bias (95% CI)	RMSE	Bias (95% CI)	RMSE	Bias (95% CI)	RMSE	Bias (95% CI)	RMSE
Garden	−1.8 (−2.1, −1.5)	2.0	−1.7 (−2.0, −1.5)	1.9	−0.7 (−0.9, −0.4)	1.1	−2.7 (−3.1, −2.3)	2.9	−2.6 (−3.0, −2.3)	2.8
Laundry	−1.1* (−1.2, 1.0)	1.2	−1.1 (−1.2, −1.0)	1.2	−0.5 (−0.7, −0.3)	0.8	−1.5 (−1.7, −1.4)	1.6	−1.5 (−1.7, −1.4)	1.6
Mop	−1.8 (−2.1, −1.5)	2.0	−1.8 (−2.1, −1.5)	2.0	−0.9 (−1.2, −0.6)	1.3	−2.5 (−2.8, −2.2)	2.7	−2.5 (−2.8, −2.2)	2.7
Mow	−1.8 (−2.1, −1.5)	2.0	−1.5 (−1.8, −1.1)	1.8	−0.5 (−1.0, −0.1)	1.4	−2.5 (−2.8, −2.1)	2.7	−2.1 (−2.5, −1.7)	2.4
Paint	−1.6 (−1.8, −1.4)	1.7	−1.6 (−1.8, −1.3)	1.7	−0.7 (−0.9, −0.4)	1.0	−2.3 (−2.6, −1.9)	2.5	−2.2 (−2.6, −1.8)	2.5
Rake	−2.2 (−2.5, −1.9)	2.4	−2.1 (−2.5, −1.8)	2.4	−1.0 (−1.4, −0.7)	1.4	−3.3 (−3.8, −2.9)	3.7	−3.3 (−3.8, −2.8)	3.6
Clean Room	−1.6 (−2.0, −1.3)	1.9	−1.1 (−1.4, −0.7)	1.5	−0.2* (−0.5, 0.1)	1.0	−2.3 (−2.7, −1.9)	2.6	−1.6 (−2.1, −1.2)	2.0
Sweep	−1.6 (−1.8, −1.4)	1.7	−1.6 (−1.8, −1.4)	1.7	−0.7 (−1.0, −0.4)	1.1	−2.4 (−2.7, −2.1)	2.5	−2.4 (−2.7, −2.1)	2.5
Tennis	−4.4 (−4.9, −3.9)	4.7	−4.5 (−5.0, −4.0)	4.8	−2.1 (−2.7, −1.5)	2.7	−5.6 (−6.2, −5.1)	5.9	−5.8 (−6.3, −5.2)	6.0
Trim	−1.7 (−1.9, −1.5)	1.8	−1.7 (−1.9, −1.5)	1.8	−0.8 (−1.1, −0.6)	1.1	−2.4 (−2.6, −2.1)	2.5	−2.3 (−2.6, −2.1)	2.5
Vacuum	−1.5 (−1.7, −1.3)	1.6	−1.5 (−1.7, −1.3)	1.6	−0.6 (−0.8, −0.3)	0.9	−2.2 (−2.4, −2.0)	2.3	−2.2 (−2.4, −2.0)	2.3

CI = Confidence Interval; RMSE = Root Mean Squared Error; m = meters; sec = seconds; gr = grade;

*
Predicted EE *not* significantly different than measured EE

Table 5

RT3 Bias (95% CI); RMSE

	RT3 Gross EE (kcals)		RT3 Activity EE (kcals)	
	Bias (95% CI)	RMSE	Bias (95% CI)	RMSE
Across All Activities	−0.5 (−0.6, −0.3)	2.9	−0.5 (−0.6, −0.4)	2.9
Treadmill Activities	0.5 (0.4, 0.6)	1.8	0.5 (0.4, 0.6)	1.8
Activities of Daily Living	−1.6 (−1.8, −1.4)	3.8	−1.7 (−1.9, −1.5)	3.8
Walking 1.34 m·sec^{-1} 0% gr	0.9 (0.8, 1.0)	1.3	0.8 (0.7, 1.0)	1.3
Walking 1.56 m·sec^{-1} 0% gr	1.1 (1.0, 1.3)	1.6	1.1 (0.9, 1.2)	1.6
Running 2.23 m·sec^{-1} 0% gr	1.8 (1.5, 2.1)	2.9	1.8 (1.4, 2.1)	2.9
Walking 1.34 m·sec^{-1} 3% gr	−0.4 (−0.5, −0.3)	1.9	−0.4 (−0.6, −0.4)	1.3
Walking 1.56 m·sec^{-1} 3% gr	−0.4 (−0.5, −0.2)	1.3	−0.5 (−0.6, −0.3)	1.3
Running 2.23 m·sec^{-1} 3% gr	0.2[*] (−0.2, 0.5)	2.4	0.1[*] (−0.3, 0.5)	2.4
Ascend Stairs	−7.6 (−7.9, −7.2)	7.9	−7.6 (−7.9, −7.3)	7.9
Basketball	−2.2 (−2.8, −1.7)	2.7	−2.3 (−2.9, −1.8)	2.8
Move 6kg Box	−0.7 (−0.9, −0.5)	1.7	−0.8 (−1.0, −0.6)	1.7
Descend Stairs	2.1 (1.8, 2.3)	2.6	2.0 (1.8, 2.2)	2.5
Dishes	−0.6 (−0.8, −0.3)	1.0	−0.7 (−0.9, −0.4)	1.0
Dust	−1.1 (−1.2, −0.9)	1.2	−1.1 (−1.3, −0.9)	1.2
Garden	−0.9 (−1.5, −0.3)	2.0	−0.9 (−1.6, −0.3)	2.0
Laundry	−1.0 (−1.1, −0.8)	1.1	−1.0 (−1.1, −0.9)	1.1
Mop	−1.6 (−1.9, −1.3)	1.8	−1.7 (−1.9, −1.4)	1.9
Mow	0.2[*] (−0.6, 0.9)	2.2	0.1[*] (−0.7, 0.8)	2.2
Paint	−1.7 (−2.0, −1.3)	1.9	−1.7 (−2.1, −1.4)	2.0
Rake	−0.4[*] (−1.2, 0.4)	2.5	−0.5[*] (−1.3, 0.3)	2.5
Clean Room	−1.0 (−1.7, −0.2)	2.2	−1.0 (−1.7, −0.3)	2.2

	RT3 Gross EE (kcals)		RT3 Activity EE (kcals)	
	Bias (95% CI)	RMSE	Bias (95% CI)	RMSE
Sweep	−1.4 (−1.7, −1.2)	1.6	−1.5 (−1.7, −1.2)	1.7
Tennis	−4.0 (−4.6, −3.4)	4.4	−4.1 (−4.6, −3.5)	4.4
Trim	−0.1 [*] (−0.7, 0.5)	1.8	−0.2 [*] (−0.8, 0.4)	1.8
Vacuum	−1.4 (−1.6, −1.2)	1.5	−1.5 (−1.6, −1.3)	1.6

CI = Confidence Interval; RMSE = Root Mean Squared Error; m = meters; sec = seconds; gr = grade;

[*] Predicted EE *not* significantly different than measured EE