

Energy Expenditure Prediction Using Raw Accelerometer Data in Simulated Free Living

ALEXANDER H. K. MONTOYE¹, LANAY M. MUDD², SUBIR BISWAS³, and KARIN A. PFEIFFER²

¹Human Performance Laboratory, Ball State University, Muncie, IN; ²Department of Kinesiology, Michigan State University, East Lansing, MI; and ³Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI

ABSTRACT

MONTOYE, A. H. K., L. M. MUDD, S. BISWAS, and K. A. PFEIFFER. Energy Expenditure Prediction Using Raw Accelerometer Data in Simulated Free Living. *Med. Sci. Sports Exerc.*, Vol. 47, No. 8, pp. 1735–1746, 2015. **Purpose:** The purpose of this study was to develop, validate, and compare energy expenditure (EE) prediction models for accelerometers placed on the hip, thigh, and wrists using simple accelerometer features as input variables in EE prediction models. **Methods:** Forty-four healthy adults participated in a 90-min, semistructured, simulated free-living activity protocol. During the protocol, participants engaged in 14 different sedentary, ambulatory, lifestyle, and exercise activities for 3–10 min each. Participants chose the order, duration, and intensity of activities. Four accelerometers were worn (right hip, right thigh, as well as right and left wrists) to predict EE compared with that measured by the criterion measure (portable metabolic analyzer). Artificial neural networks (ANNs) were created to predict EE from each accelerometer using a leave-one-out cross-validation approach. Accuracy of the ANN was evaluated using Pearson correlations, root mean square error, and bias. Several ANNs were developed using different input features to determine those most relevant for use in the models. **Results:** The ANNs for all four accelerometers achieved high measurement accuracy, with correlations of $r > 0.80$ for predicting EE. The thigh accelerometer provided the highest overall accuracy ($r = 0.90$) and lowest root mean square error (1.04 METs), and the differences between the thigh and the other monitors were more pronounced when fewer input variables were used in the predictive models. None of the predictive models had an overall bias for prediction of EE. **Conclusions:** A single accelerometer placed on the thigh provided the highest accuracy for EE prediction, although monitors worn on the wrists or hip can also be used with high measurement accuracy. **Key Words:** ACCELEROMETRY, MACHINE LEARNING, ARTIFICIAL NEURAL NETWORK, PHYSICAL ACTIVITY

Physical activity (PA) has long been recognized for its beneficial effects on many aspects of health. Because of these known health benefits, the most recent PA guidelines advocate that adults obtain a minimum of 150 min·wk⁻¹ of moderate- or vigorous-intensity PA (MVPA), which is defined as any activity that elicits an energy expenditure (EE) of at least 3.0 times the resting level or 3.0 METs (40). Accurate measurement of EE is vital for understanding the prevalence of meeting PA recommendations, identifying populations who may benefit from interventions aimed at increasing PA, and understanding the relation between PA and health.

Objective PA measurement tools such as activity monitors have shown considerable promise because of their relative ease of use and accurate measurement of PA for days or weeks at a time (41). Accelerometer-based activity monitors

in particular have seen dramatically increased use for measurement of free-living PA. Accelerometers are generally worn on the hip and record accelerations of the trunk as a person moves. These accelerations have traditionally been used as an independent variable in linear regression equations to predict EE. Linear regression approaches to prediction of EE are appealing because of their simplicity and their high accuracy in initial validation studies, which focused on measuring the EE of ambulatory activities (i.e., walking and running) in controlled settings (12). However, the linear relation between accelerations and EE does not seem to hold when applied to nonambulatory activities or free-living environments, resulting in much poorer prediction accuracy in such situations (14,36). To overcome these limitations, researchers have explored several avenues to improve PA measurement. One approach involves the use of more than one monitoring device to measure accelerations and/or other physiologic variables (i.e., HR) to improve EE measurement. The use of multimonitor systems has shown promise for improving EE measurement in several studies (3,7,42), but the use of multiple monitors dramatically increases participant and researcher burden, preventing these methods from being feasible for use in large surveillance, intervention, or epidemiologic studies.

Another approach to improving EE prediction has involved using techniques other than linear regression for modeling the

Address for correspondence: Alexander H. K. Montoye, Ph.D., Human Performance Laboratory 232, Ball State University, Muncie IN 47306; E-mail: ahmontoye@bsu.edu.

Submitted for publication September 2014.

Accepted for publication December 2014.

0195-9131/15/4708-1735/0

MEDICINE & SCIENCE IN SPORTS & EXERCISE®

Copyright © 2015 by the American College of Sports Medicine

DOI: 10.1249/MSS.0000000000000597

relation between acceleration data and EE. Machine learning, a branch of artificial intelligence, has become a popular modeling technique and has been shown to improve EE measurement in both laboratory-based and simulated free-living settings (11,30,35). However, there are still many unresolved questions regarding the use of machine learning for predicting EE. First, machine learning modeling may allow for accurate prediction of EE using accelerometers placed on body locations other than the hip (i.e., wrist, ankle, and thigh) but it is unclear whether accelerometers placed on alternate body locations can achieve the same measurement accuracy as that of a hip-mounted accelerometer. The wrist is a promising accelerometer placement location because of its utility in measuring sleep and activity type as well as ease of wear (15,18,22,44). In addition, accelerometers worn on the thigh have shown high accuracy for measuring ambulatory activity and sedentary behavior (13,31). Despite the potential for the wrist and thigh as appealing measurement sites, there is very limited evidence regarding their utility for measuring EE.

Second, a current limitation of using machine learning to model accelerometer data is that machine learning models are much more complex than traditional linear regression approaches both in the extraction of useful information (features) from accelerometer data to use as inputs into the models and the model creation itself. This complexity currently limits the use of machine learning and keeps it from being used on a wider scale. However, there is some evidence that the process of developing and using machine learning can be simplified without compromising measurement accuracy. In 2009, Staudenmayer et al. (35) took a large step toward simplifying the use of machine learning modeling. They used the R statistical software (a freely available, open-source software package) to develop a specific type of machine learning model (an artificial neural network (ANN)) to predict EE and activity type. In addition, they used simple time domain features (percentiles of the acceleration signal and autocorrelation) as input variables and achieved dramatically improved EE predictions over linear regression approaches. However, it is unknown whether the features they used as input variables in their models represent an optimal set of input variables for maximizing EE prediction accuracy.

Third, most validation studies are carried out in laboratory-based settings, which allows for good control of type, duration, and intensity of activities performed. However, there is considerable evidence that laboratory-based validation techniques have considerably lower accuracy when applied to free-living situations (6,19,36). Recent studies by Calabro et al. (4) and Mackintosh et al. (20) have used semistructured settings for validation of cut points and comparison of different types of accelerometers, but to our knowledge, no studies have used this type of environment for creation or validation of machine learning models for predicting EE.

The purposes of this study were threefold: 1) to validate and compare ANNs developed from accelerometers worn on the hip, thigh, and both wrists for prediction of EE in a semistructured, simulated, free-living setting, 2) to compare

accuracies achieved by the left and right wrist ANNs, and 3) to identify simple input features that maximize predictive accuracy while minimizing complexity of the ANNs.

METHODS

Participants. A total of 44 adults (22 male and 22 female participants) were recruited from the area of East Lansing, MI, via e-mail, flyers, and word of mouth for participation in this study. Exclusion criteria included the following: 1) participants who had known health conditions that prevented them from being able to perform MVPA safely, 2) participants who were bound to wheelchair or had orthopedic limitations that invalidated the use of accelerometry for activity measurement, or 3) those who fell outside the age range of 18–44 yr.

This study was approved by the Michigan State University institutional review board before participant recruitment. Details of the study were described to each participant immediately upon arriving at the Human Energy Research Laboratory, and a written informed consent was obtained before proceeding with the protocol.

Instrumentation. The instruments used in this study were ActiGraph GT3X+ accelerometers, GENEActiv accelerometers, and an Oxycon Mobile portable metabolic analyzer. The Oxycon portable metabolic analyzer provided a criterion measure of EE. The accelerometers and portable metabolic analyzer were synchronized to an external clock before each test; descriptions of the accelerometers and metabolic analyzer follow.

ActiGraph accelerometers. The ActiGraph (ActiGraph LLC, Pensacola, FL) is the most commonly used accelerometer on the market for PA research, and there is an abundance of literature regarding its reliability and validity for measurement of PA (12,23). Two GT3X+ models were placed on each participant during the study. One accelerometer was placed on the midline of the right thigh, one-third of the way between the inguinal crease and patella and adhered to the leg with hypoallergenic sticky tape. The other ActiGraph was mounted on the right hip at the anterior axillary line with an elastic belt. The ActiGraph GT3X+ records raw accelerations of up to ± 6 times gravitational force (6g) in three axes of movement. For the current protocol, the ActiGraph accelerometers recorded at a rate of 40 samples per second (40 Hz).

GENEA accelerometers. The GENEActiv (Activinsights Ltd., Kimbolton, Cambridgeshire, United Kingdom) is a newer accelerometer that has recently been validated for PA measurement (9). Like the ActiGraph, the GENEActiv records raw data of up to ± 6 g in three axes of movement. The GENEActiv devices were set to record acceleration data at a rate of 20 Hz for the current study. The GENEActiv is shaped like a watch and comes with a standard wrist strap, allowing for easy attachment to the wrist. Participants wore two GENEActiv accelerometers (one on each wrist) for this study. Each GENEActiv was fastened securely to the dorsal side of the wrist between the styloid processes of the radius and ulna (9).

The acceleration data for all four accelerometers were time stamped and stored within the monitors and were later downloaded to a computer for analysis.

Oxycon portable metabolic analyzer. The Oxycon Mobile (Cardinal Health, Yorba Linda, CA) portable metabolic analyzer was used to measure oxygen consumption ($\dot{V}O_2$) and carbon dioxide production ($\dot{V}CO_2$) during 13 of the 14 activities performed in the protocol (EE was recorded but not analyzed for the nonwear activity). The Oxycon is lightweight (950 g) and was worn on the back using a shoulder harness. Participants were fitted with a breathing mask (held in place by a mesh cap), which was attached to a digital turbine flow meter and gas sampling tube. The flow meter and sampling tube collected data for measurement of inspired and expired air volume so that $\dot{V}O_2$ and $\dot{V}CO_2$ could be calculated on a breath-by-breath basis. $\dot{V}O_2$ data were expressed in milliliters per kilogram per minute ($\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$) and converted to METs for analysis. Before each test, the Oxycon was calibrated according to manufacturer's specifications to ensure accurate measurements for flow rate and gas concentration. The Oxycon has been shown to provide valid $\dot{V}O_2$ measures over a range of exercise intensities (2,28) and was used as the criterion measure for EE in this study.

Procedure. Each participant reported to the Human Energy Research Laboratory for one visit. Participants were asked to refrain from eating for 3 h before visiting the laboratory to minimize the risk of discomfort while performing the activities and because food ingestion can affect EE values. Details of the study were discussed with each participant. A written informed consent was obtained, and a PA readiness questionnaire was administered to ensure that the participant was healthy and had no contraindications to engaging in MVPA. If participants had answered "yes" to any question on the questionnaire, they would have been asked to obtain physician approval before being able to participate in the study; however, this did not occur. Next, participant weight and height were taken by trained research assistants according to standardized methods (21). Weight was measured to the nearest 0.1 kg using a Seca digital scale (Seca, Hanover, Germany) with shoes off and weight

balanced on the center of the scale. Height was measured to the nearest 0.1 cm using a Harpenden stadiometer (Holtain Ltd., Crymych, United Kingdom). Two measurements were taken and averaged for both weight and height. If the two weights differed by more than 0.3 kg or if the two heights differed by more than 0.4 cm, a third measurement was taken and the closest two values were averaged. Body mass index (BMI) was calculated by dividing body weight by the square of height ($\text{kg}\cdot\text{m}^{-2}$). Age was assessed by asking participants to state their age in years, and handedness was assessed by asking participants which hand they prefer to use for the majority of their everyday activities.

Each participant wore the Oxycon metabolic analyzer, one ActiGraph on the hip, another ActiGraph on the thigh, one GENEa on the left wrist, and one GENEa on the right wrist while performing 14 activities (activity descriptions are provided in Table 1). These activities comprised a range of intensities from sedentary to vigorous and represented a mixture of sedentary, ambulatory, exercise, and lifestyle activities. Ambulatory activities (walking and running) are common in accelerometer validation literature; however, we added the sedentary, exercise, and lifestyle activities to determine the potential of the four accelerometers to measure a range of activity types and intensities often seen in free-living settings. In addition, we included a nonwear activity but this activity was not included in our analysis of EE prediction.

The 14 activities were performed in semistructured, simulated free-living setting, which took approximately 90 min to complete and took place in a laboratory room inside the Human Energy Research Laboratory and a hallway and stairwell outside the laboratory. A list of the activities was written on a whiteboard for participants at the beginning of the visit, and a description of each activity was given. The order of activities on the whiteboard was altered every four to five participants to avoid ordering effects during the visit. Participants completed each of the 14 activities for 3–10 min, with the order, intensity, and timing of the activities left up to each participant. A research assistant observed and recorded each activity on a handheld computer while it was being performed and periodically updated participants on which activities they still needed to complete. The nonwear

TABLE 1. Activities performed during the simulated free-living protocol.

Activity	Activity Intensity	Description of Activity ^a
Lying down (T ₁)	Sedentary	Lying on a mat on the floor
Reading (T ₂)	Sedentary	Reading a magazine article while sitting at a table
Computer (T ₃)	Sedentary	Clicking and typing while playing a computer game
Standing (T ₄)	Light	Standing still with arms at sides
Laundry (T ₅)	Light	Folding towels and putting them in a laundry basket
Sweeping (T ₆)	Light	Sweeping confetti into piles
Walking slowly (T ₇)	Light	Walking at a self-selected "leisure" pace in a hallway
Walking fast (T ₈)	Moderate	Walking at a self-selected "brisk" pace in a hallway
Jogging (T ₉)	Vigorous	Jogging at a self-selected pace in a hallway
Cycling (T ₁₀)	Moderate	Cycling on an ergometer at a self-selected cadence of 50–100 rpm with 1.0-kg resistance
Stair use (T ₁₁)	Vigorous	Walking up and down the stairs at a self-selected pace
Biceps curls (T ₁₂)	Light	Biceps curls with a 1.4-kg weight in each hand
Squats (T ₁₃)	Moderate	With feet shoulder-width apart, flexing at the knees (to a 90° angle) and then straightening legs
Nonwear (T ₁₄)	N/A	Not wearing the accelerometer

^aActivity order, intensity, and duration (3–10 min) were chosen by participants.

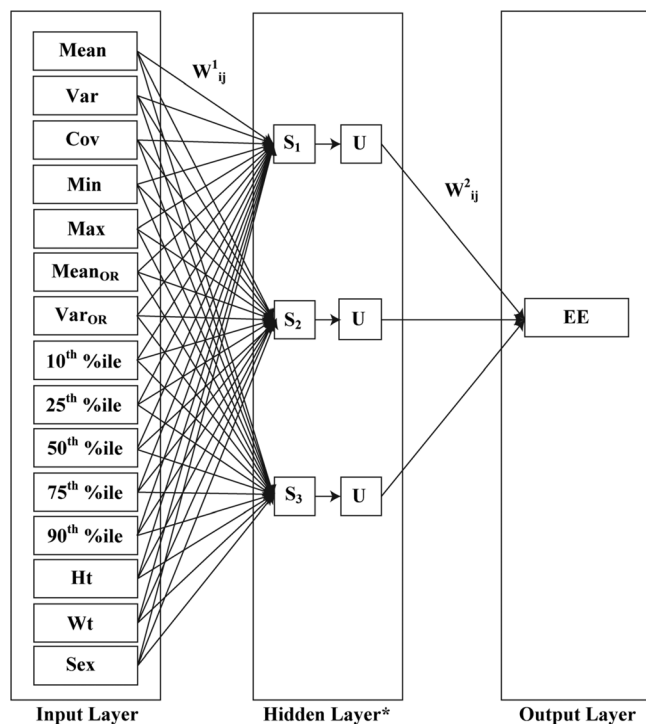


FIGURE 1—ANN for predicting EE. This figure is a graphical depiction of the ANN developed and tested in this study. The *input layer* contains the features used as input variables, whereas the *hidden layer* is a computation layer within the ANN. * Signifies that the *hidden layer* contains 15 hidden units, although only three are shown in the figure for simplicity. There were 36 accelerometer features and three participant characteristics features tested in this study, which are listed below. In addition, *S* represents summations of the *input layer* in the hidden units. *U* is the activation function for the *hidden layer*, and W^1 and W^2 are weight vectors for the inputs and summations, respectively. The accelerometer signal features (one of each per axis and three of each per accelerometer) are as follows: 1) mean, mean; 2) var, variance; 3) cov, covariance; 4) min, minimum; 5) max, maximum; 6) mean_{OR}, mean accelerometer orientation; 7) var_{OR}, variance of accelerometer orientation; 8) 10th %ile, 10th percentile; 9) 25th %ile, 25th percentile; 10) 50th %ile, 50th percentile; 11) 75th %ile, 75th percentile; 12) 90th %ile, 90th percentile. The participant characteristics features are as follows: 1) ht, participant's height; 2) wt, participant's weight; 3) sex, participant's sex (0 for females and 1 for males).

activity was saved until the end of the 90-min protocol. Upon completion of the protocol, participants were given a \$35 Target® gift card.

Data reduction and modeling. ANNs are nonlinear models, which take a set of inputs x_1, \dots, x_k and use them to predict a certain output variable y (e.g., EE or activity type),

where k is the number of features used to predict y . An ANN designed to predict EE (in METs) was developed for each accelerometer (example shown in Fig. 1).

The ANNs were created and tested using a leave-one-participant-out approach. In this approach, the ANN was first created from a “training” data set, where the input features and the outcome variable (EE in METs) were used to estimate the weights for each input feature. This training set consisted of the data from all but one participant in the study. Then, the ANN was tested on the data from the participant left out of the training phase. This testing was conducted by supplying the input features and comparing the predicted EE from the ANN to the measured EE from the criterion measure (Oxycon metabolic analyzer). This process was conducted with each participant's data used as the testing data once, therefore obtaining an ANN for each participant in the study. This process was conducted separately for each accelerometer placement site.

Several studies provide evidence that time domain features can be used to achieve high activity classification (70%–90% from a single accelerometer) and EE prediction accuracy without use of frequency domain features (8,24,35). Therefore, we constructed a set of 36 time domain features and three participant characteristics (Fig. 1) to use in constructing our ANNs. After constructing an ANN for each accelerometer placement site with all 39 features (set 1), we tested several subsets of this total feature set. First, we chose a feature set (set 2) using only mean and variance because members of our research group have achieved high measurement accuracy with these features in previous work (7,24). Next, we used a stepwise approach to select features that had correlations of less than $r = 0.70$ with features already included in the set. Using this approach, we created a feature set (set 3) consisting of mean, variance, minimum, and maximum of the acceleration signal. In our final feature set (set 4), we used the 10th, 25th, 50th, 75th, and 90th percentiles and covariance of the acceleration signal because similar feature sets have been used successfully in a previous work (35). These four feature sets can be seen in Table 2. We initially included lag-1 autocorrelation in this feature set because it can yield valuable information on the temporal nature of activities by assessing the correlation of two adjacent epochs of acceleration data. However, the calculation of autocorrelation involves dividing by the variance of the acceleration data within the adjacent windows, which is

TABLE 2. Feature sets used for creation and testing of ANNs.

Feature Set	Features Used	Total No. of Features Used
Set 1	Mean, variance, covariance, minimum, maximum, mean orientation, variance of orientation, and 10th, 25th, 50th, 75th, and 90th percentiles of acceleration signal, weight ^a , height ^a , and sex ^a	36 or 39 (12 accelerometer features per axis × three axes and either zero or three participant characteristic features)
Set 2	Mean and variance of acceleration signal, weight ^a , height ^a , and sex ^a	Six or nine (two accelerometer features per axis × three axes, and either zero or three participant characteristic features)
Set 3	Mean, variance, minimum, and maximum of acceleration signal, weight ^a , height ^a , and sex ^a	12 or 15 (four accelerometer features per axis × three axes, and either zero or three participant characteristic features)
Set 4	10th, 25th, 50th, 75th, and 90th percentiles and covariance of acceleration signal, weight ^a , height ^a , and sex ^a	18 or 21 (six accelerometer features per axis × three axes, and either zero or three participant characteristic features)

^aIndicates that models were created once with this variable and once without to determine whether addition of the variable improved predictive accuracy of the ANN.

0 for many sedentary activities and results in an invalid calculation. Thus, we used covariance as a substitute for lag-1 autocorrelation because covariance is simpler to calculate, is defined even when the variance is 0, and can provide information regarding the similarity of the accelerometer signals of adjacent data windows (similar to autocorrelation). These feature sets were also initially tested both including and excluding the three participant characteristics (weight, height, and sex) to determine whether demographic characteristics would improve accuracy of the ANNs. In addition, we chose to use 15 hidden units in the hidden layer on the basis of the number of activities in the study and the number of features being used (26). Skip-layer connections were not allowed, and a Broyden–Fletcher–Goldfarb–Shanno optimization algorithm was used, as is the standard in the *nnet* package in R (27). The R code used in this study is very similar to the R code used and available in the work of Staudenmayer et al. (35).

Oxycon data. In a previous study by members of our research group, we reintegrated breath-by-breath Oxycon portable metabolic analyzer data into 10-s and then 15-s epochs for analysis. However, with both epochs, we found that data loss occurred in participants with slower breathing rates (especially during sedentary activities), resulting in our reintegrating the data into 30-s epochs for our final analysis (24). Correspondingly, breath-by-breath Oxycon data from the simulated free-living protocol were reintegrated into 30-s epochs for measurement of EE in the current study. Absolute oxygen consumption values measured by the Oxycon were first converted to relative terms (by dividing by participant weight) and then converted into METs (by dividing relative oxygen consumption by 3.5) for analysis. The 30-s epochs of accelerometer data were used for training the ANN to predict EE (as described earlier). In addition, when testing the ANN, 30-s epochs were used for computing predicted EE for comparison with Oxycon-measured EE. Because the Oxycon recorded continuously and was not dependent on correctly identifying an activity type, all data, including transitions, were included for training and testing of the ANN. This means that non-steady-state data were used in training and testing of the ANNs.

Statistical analyses. After downloading the accelerometer and Oxycon data, all data processing was conducted in Microsoft Excel (Microsoft Corporation, Redmond, WA) and ANN creation was performed using the R statistical software package (R-project, Vienna, Austria) (33). R is an open-source software, which is freely available for download and has a special ANN library, which can be used for development and testing of ANNs (16,20). Thus, development and application of ANNs in R are less costly and much less complicated than machine learning algorithms developed using other software programs and R has been used successfully for creation of ANNs for predicting EE and activity type (19,35). Three summary statistics were calculated to test the accuracy of each ANN for predicting EE, as follows: Pearson correlations, root mean square error (RMSE), and bias.

A minimum correlation of $r = 0.60$ has been defined as having moderately high validity in the literature; therefore, we desired to obtain a correlation of $r \geq 0.60$ between predicted EE and Oxycon-measured EE (32). A sample size of 25 is needed to detect a correlation of $r = 0.60$ with 90% power (5); thus, our sample of 44 was more than sufficient for addressing this question. For RMSE, smaller values represent better prediction of the ANNs; our goal was to minimize RMSE to maximize accuracy of the ANNs. We used bias to evaluate systematic under- or overprediction of EE by the ANNs.

Correlations, RMSE, and biases were calculated separately for each of the four accelerometer placements and each of the four feature sets. In addition, MET predictions were derived from the vertical axis of the hip accelerometer data using the popularly used, cut point-based Freedson MET prediction equation (12) to compare the accuracy of the ANNs with this widely tested equation and the corresponding cut points for determining activity intensity.

Differences among correlations, RMSE, and biases among the four accelerometers were assessed using repeated-measures ANOVA (RMANOVA). In addition, differences among feature sets were evaluated using RMANOVA. Because correlations tend to be negatively skewed, we first performed a Fisher *Z* transformation to normalize the correlations before performing the RMANOVA. When the RMANOVA revealed statistically significant differences for any of the three analyses, *post hoc* dependent *t*-tests were conducted to determine differences among monitor placements or feature sets. The *a priori* alpha level was set at $P < 0.05$ for determining statistical significance. Statistical analyses were performed using SPSS version 22 (IBM Corporation, Armonk, NY).

RESULTS

Malfunction of the Oxycon metabolic analyzer (because of a bad battery) occurred in three participants, and accelerometer malfunction occurred in another two participants. These participants were excluded from further analyses, resulting in 39 of the 44 participants' data included in model creation and validation. Means and SD for participant characteristics (both those included in and excluded from analysis) are shown in Table 3. Although weight and BMI seemed higher in the females excluded in the final analysis, these differences were not statistically significant. Of the 39 participants included in the analysis, 13 were either overweight or obese according to BMI ($\geq 25 \text{ kg} \cdot \text{m}^{-2}$). In addition, four of the 39 participants included in the final analysis were left hand dominant, with the remaining 35 being right hand dominant.

The average length of time participants spent performing activities in the protocol was 80.6 min, with approximately 11.8 min spent in transitions (total time, 92.4 min). The average lengths of time that participants spent in each activity ranged from 4.5 min with squats to 7.6 min with stair

TABLE 3. Demographic characteristics of participants enrolled in study.

	Included in Analysis			Excluded from Analysis		
	All (n = 39)	Males (n = 19)	Females (n = 20)	All (n = 5)	Males (n = 3)	Females (n = 2)
Age (yr)	22.1 (4.3)	23.7 (5.0)	20.5 (2.7)	21.2 (2.9)	21.3 (4.0)	21.0 (1.4)
Weight (kg)	72.4 (16.2)	84.5 (13.1)	60.8 (8.9)	78.2 (21.2)	75.9 (4.9)	81.6 (41.4)
Height (cm)	171.4 (10.1)	179.1 (7.7)	164.1 (5.7)	167.8 (9.8)	175.9 (0.4)	157.1 (2.4)
BMI (kg·m ⁻²)	24.4 (3.6)	26.3 (3.4)	22.5 (2.6)	28.0 (9.1)	24.8 (1.7)	32.8 (15.8)

Data in the table are displayed as mean (SD).

use. Squats were the only activity performed for an average of less than 5 min per visit. Standing, sweeping, biceps curls, and jogging were performed for 5–6 min, whereas lying, walking slowly, walking fast, and cycling were performed for 6–7 min and reading, computer, laundry, and stair use were performed for 7–8 min, on average, per visit.

In initial testing of the four feature sets, it was found that the addition of weight, height, and sex yielded no gains in predictive accuracy of the ANNs. Therefore, these features were removed when training and testing the ANNs. Correlations, RMSE, and bias for predicted EE are shown in Table 4.

With correlations ranging from $r = 0.83$ to 0.90 for the four accelerometers across the four sets of features, all four monitor placements achieved correlations well above the $r = 0.60$ desired to indicate moderately high validity. The RMANOVA test among accelerometer placement sites revealed a test statistic of $F = 4.36$, indicating significant differences among the four placement sites. *Post hoc* tests revealed that the thigh-mounted ActiGraph accelerometer ANNs had significantly higher correlations with measured EE ($r = 0.89$ – 0.90) than the hip ($r = 0.83$ – 0.88) or wrist ($r = 0.84$ – 0.87) accelerometer ANNs for each of the four feature sets. The hip-mounted ActiGraph accelerometer ANN had significantly higher correlations with measured EE than the wrist-mounted GENE accelerometer ANNs

for feature set 3 only. In addition, correlations achieved by the left and right wrist accelerometer ANNs were similar for each of the four feature sets.

When comparing correlations achieved among the four sets of features, the thigh ANN accuracy was not affected by choice of feature set, meaning that even the simplest ANN (created from feature set 2) achieved similar accuracy to the ANNs which used more features. Conversely, for the hip accelerometer ANN, feature set 2 resulted in a significantly lower correlation than feature set 1; similarly, correlations dropped for the wrist accelerometer ANNs for feature sets 2 and 3 but not for set 4. Overall, the ANN created using feature set 4 resulted in similar correlations with measured EE to correlations achieved with feature set 1 for all four accelerometers.

For RMSE, the RMANOVA test revealed a test statistic of $F = 3.64$, indicating significant differences in RMSE among placement sites (Table 4). *Post hoc* tests revealed that the thigh accelerometer ANNs again outperformed the hip and wrist ANNs, with significantly lower RMSE values for all four feature sets. Differences in RMSE between the thigh and hip ANNs ranged from 8.7% to 22.2%, and differences between the thigh and wrists ranged from 10.6% to 22.1%. In comparison of the four feature sets, RMSE was highest for the hip ANN using feature set 2, but

TABLE 4. Correlations of measured vs predicted EE.

	ActiGraph Hip	ActiGraph Thigh	GENEA Left Wrist	GENEA Right Wrist
Correlations				
Feature set 1	0.88 (0.05)*	0.90 (0.06)	0.86 (0.06)*	0.87 (0.05)*
Feature set 2	0.83 (0.04)***	0.89 (0.05)	0.84 (0.05)***	0.84 (0.05)***
Feature set 3	0.87 (0.04)*	0.90 (0.05)	0.84 (0.05)***,***	0.84 (0.05)***,***
Feature set 4	0.86 (0.06)****	0.90 (0.07)	0.87 (0.04)***,***,***	0.87 (0.05)***,***,***
Freedson MET equation	0.79 (0.05)*****	N/A	N/A	N/A
RMSE (METs)				
Feature set 1	1.13 (0.21)*	1.04 (0.21)	1.18 (0.32)*	1.18 (0.27)*
Feature set 2	1.32 (0.21)***	1.08 (0.25)	1.26 (0.29)*	1.25 (0.26)*
Feature set 3	1.17 (0.22)*	1.04 (0.24)	1.26 (0.26)*	1.27 (0.29)*
Feature set 4	1.20 (0.26)****	1.04 (0.31)	1.15 (0.21)****	1.21 (0.35)*
Freedson MET equation	1.50 (0.28)*****	N/A	N/A	N/A
Bias (METs)				
Feature set 1	0.01 (0.35)	−0.01 (0.34)	−0.02 (0.32)	−0.01 (0.35)
Feature set 2	0.02 (0.59)	0.03 (0.42)	0.01 (0.41)	−0.03 (0.49)
Feature set 3	−0.03 (0.46)	0.01 (0.32)	−0.03 (0.35)	0.00 (0.47)
Feature set 4	−0.02 (0.36)	−0.01 (0.38)	0.01 (0.32)	−0.01 (0.38)
Freedson MET equation	−0.53 (0.33)*****	N/A	N/A	N/A

Correlations, RMSE, and bias are shown as mean (SD).

*Indicates significant differences ($P < 0.05$) from thigh accelerometer ANN.

**Indicates significant differences from feature set 1.

***Indicates significant differences from hip accelerometer ANN.

****Indicates significant differences from feature set 2.

*****Indicates significant differences from feature set 3.

*****Indicates significant differences from all ANNs and all feature sets.

sets 1, 3, and 4 yielded similar RMSE values. Conversely, RMSE values achieved by the thigh accelerometer ANN were similar across all four feature sets. For the left wrist accelerometer ANN, the ANN created with feature set 4 had significantly lower RMSE than the ANN created from feature set 2, but no other differences existed among feature sets. Despite a nonsignificant trend ($P < 0.10$) toward higher RMSE for feature sets 2 and 3 with the right wrist accelerometer ANNs, RMSE values were not significantly different among feature sets.

Average biases for each accelerometer are also shown in Table 4. The RMANOVA test statistic was $F = 0.062$, indicating no overall bias for any of the four monitor placements or for any of the four feature sets. This lack of bias indicates that none of the accelerometers had an overall overestimation or underestimation of EE in the total sample.

Table 4 also shows correlations, RMSE, and bias for predicted EE using the Freedson MET prediction equation. The overall correlation of $r = 0.79$ indicated moderately high validity of the Freedson equation, although this correlation is significantly lower than the correlations achieved by the ANNs for all four accelerometer placement sites and all four feature sets. Similarly, the average RSME of 1.50 METs achieved using the Freedson equation was 13.6%–32.7% higher than the RMSE achieved by the hip ANNs, 38.9%–44.2%,

19.0%–30.4%, and 18.1%–27.1% higher than the RMSE achieved by the thigh, left wrist, and right wrist ANNs, respectively. Moreover, the Freedson prediction equation underestimated the EE of activities performed in this visit by an average of 0.53 METs. To better determine bias at different activity intensities and for specific activities, plots were created for each feature set, comparing measured with predicted EE on an activity-by-activity basis. These plots can be seen in Figure 2. In the plot for feature set 1, the predicted EE by the Freedson MET equation is also included. The closer to the linear line each data point on the plots falls, the closer the prediction is to the measured value. Of note, the EE for sedentary activities was predicted accurately using the ANNs created for the hip- and thigh-mounted accelerometers but was slightly overpredicted using the ANNs for the wrist-mounted accelerometers. The EE for the slow walk was slightly overpredicted by all ANNs, whereas the EE for stair use was slightly underpredicted for all ANNs. In addition, the EE for cycling and squats tended to be underpredicted by all ANNs, with the exception of the ANN for the thigh-mounted accelerometer for cycling. Because of the large number of data points collected (>2000 for each activity), even slight differences between predicted and measured EE were statistically significant ($P < 0.05$). The Freedson MET equation dramatically underestimated the

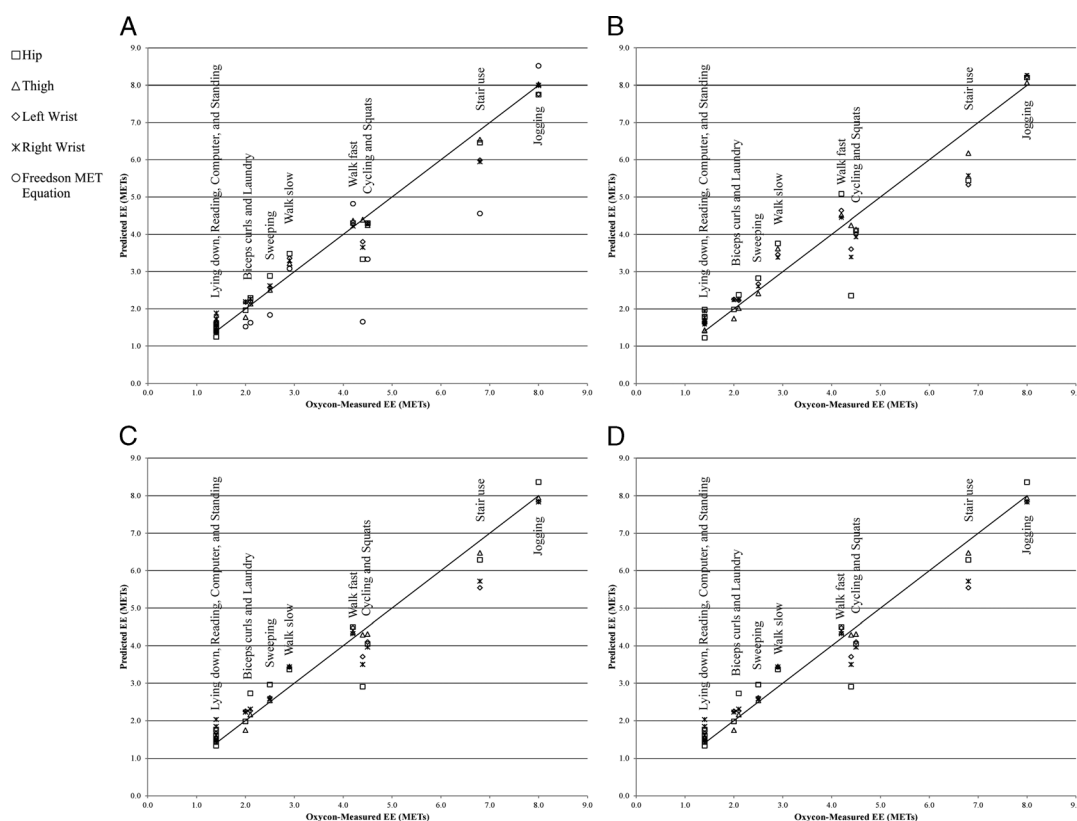


FIGURE 2—A comparison of measured and predicted EE among different accelerometer placements and feature sets. A, The difference between measured and predicted EE for feature set 1 across four accelerometer placement sites. The Freedson cut point-based MET prediction equation for the hip accelerometer is also shown. B, The difference between measured and predicted EE for feature set 2 across four accelerometer placement sites. C, The difference between measured and predicted EE for feature set 3 across four accelerometer placement sites. D, The difference between measured and predicted EE for feature set 4 across four accelerometer placement sites.

EE of biceps curls, laundry, sweeping, cycling, squats, and stair use but had fairly accurate overall estimations of the EE for the four sedentary activities, the two walking activities, and jogging.

DISCUSSION

The purposes of this study were 1) to validate and compare ANNs developed from accelerometers worn on the hip, thigh, and wrists for prediction of EE, 2) to compare accuracies of the left and right wrists, and 3) to use simple input features to maximize prediction accuracy while minimizing complexity of the developed ANNs.

Our results showed strong correlations between measured and predicted EE for ANNs developed for all four accelerometer placements and for all four feature sets. In addition, our results indicated no systematic bias by any of the ANNs for prediction of EE. Overall, the thigh-mounted accelerometer provided the highest correlations with measured EE and also the lowest RMSE of the placement sites for all four feature sets; in addition, the differences in performance were more apparent when comparing the ANNs developed with the simpler feature sets (i.e., sets 2 and 3). The thigh accelerometer ANN performance was not diminished for any of the four feature sets tested, meaning that even very simple inputs such as mean and variance of the acceleration signal were used to predict EE with a high degree of accuracy. Given previous work showing high accuracy for measuring sedentary behavior and ambulatory activities with thigh-mounted accelerometers (13,31,34), the results of this study further illustrate the utility of the thigh as a highly accurate placement site for activity and EE measurement.

Despite the superiority of the thigh-mounted accelerometer, it is worth emphasizing that the left and right wrist accelerometer ANNs provided only slightly lower accuracy than the thigh and comparable accuracy with the hip, resulting in high overall prediction accuracy of all four placement sites. Our finding of high prediction accuracy for the wrist accelerometer placement sites lies in contrast to studies that have used linear regression-based approaches for predicting EE. In the early days of using activity monitors, Montoye et al. (25) found significantly higher correlations for predicting EE using a hip-mounted motion sensor ($r = 0.71$) compared with a wrist-mounted accelerometer ($r = 0.40$) during ambulatory and exercise activities. Similarly, Swartz et al. (36) found that in a simulated free-living setting, the hip-mounted accelerometer predicted EE with a moderate correlation of $r = 0.56$ whereas the wrist-mounted accelerometer had very poor correlation ($r = 0.18$) with measured EE. Finally, as recently as 2013, Rosenberger et al. (29) found higher correlations ($r = 0.72$ vs $r = 0.36$) and lower error (0.55 vs 0.85 METs) when predicting EE from a hip-mounted accelerometer compared with that from a wrist-mounted accelerometer. It is important to note that these studies all used linear regression for their modeling technique; the consistent superiority

of the hip to the wrist when linear regression is used is not surprising, given that a hip monitor records movement of the trunk whereas wrist monitors record arm movement that may or may not be coupled with movement of the rest of the body, resulting in poor correlations of activity counts and EE.

A significant advantage of machine learning is its ability to recognize patterns in an acceleration signal rather than simply using magnitude of acceleration for prediction. Recent studies by Mannini et al. (22) and Zhang et al. (44) show very high activity classification accuracies (85%–97%) using a wrist accelerometer coupled with machine learning models, giving strong reason to believe that machine learning would also allow for high accuracy of EE prediction. The results of the current study support the utility of ANN modeling as a viable approach to analyzing wrist-mounted accelerometer data and provide further evidence of the superiority of machine learning to linear regression for modeling of accelerometer data.

Although the strengths and limitations of the Freedson MET prediction equation are already well-known from previous validation literature in a variety of settings, we used this equation in the current study as a benchmark for evaluating the performance of the ANNs. We found that the Freedson prediction equation explained approximately 61% of the variance (0.79^2) in measured EE, whereas the ANNs were able to account for 69%–81% of the variance in measured EE. In addition, the RMSE values achieved with the ANNs were 13.6%–44.2% lower than those achieved with the Freedson prediction equation, again showing dramatic improvements in EE prediction with the ANNs compared with the cut point approach. Finally, the Freedson prediction equation dramatically underpredicted the EE for non-ambulatory and nonsedentary activities in this protocol.

The high accuracy of wrist-mounted accelerometers for EE prediction found in this study is especially encouraging, given the utility of the wrist location for measuring sleep quality and its current use in large surveillance studies such as National Health and Nutrition Examination Survey (38). Moreover, wrist-mounted accelerometers are comfortable to wear and can be designed/disguised to look like watches, both of which may lead to improved compliance. With the ability to accurately measure sleep as well as activity type and EE, the wrist may represent an ideal blend of practicality and measurement accuracy for monitoring lifestyle behaviors and patterns.

In addition, the left and right wrist accelerometer ANNs achieved equally high accuracies for prediction of EE, which is informative, given that the current convention for wrist accelerometer wear is for an accelerometer to be worn on the nondominant wrist. Our results are in accordance with a 2012 study by Zhang et al. (44), which found that the classification accuracy rates for identifying four types of activities were 97% and 96% for the left and right wrist accelerometers, respectively. These studies point to the utility of either wrist for measurement of EE and classification of activity type, providing preliminary evidence that the

convention for accelerometer placement on the nondominant wrist may be unnecessary. It is worth noting that the ANNs developed for the wrist-mounted GENEa accelerometers may not be directly applicable to data collected from wrist-mounted ActiGraph accelerometers because of small differences in the capture of raw data from these two brands of accelerometers, as was recently reported by John et al. (16). However, the findings of this study and those of John et al. (16) provide strong evidence that ANNs developed for wrist-mounted ActiGraph accelerometers will be able to predict EE with high accuracy. We are currently starting data collection with wrist-mounted ActiGraph accelerometers to confirm this hypothesis. In addition, although the GENEa and ActiGraph monitors collected data at different sampling rates (20 and 40 Hz), another study by Zhang et al. (43) showed no difference in predictive accuracy among monitors collecting data at sampling rates from 10 to 80 Hz and we do not believe that the difference in sampling rates contributed to the different predictive accuracies seen in our study among the wrists, hip, and thigh placements.

The hip-mounted accelerometer achieved correlations of $r = 0.83$ – 0.88 and RMSE values of 1.13–1.32 METs with the different feature sets, and these statistics compare favorably with those achieved in previous studies. In a study conducted in a laboratory-based setting, Staudenmayer et al. (35) found that an ANN developed using data from a hip-mounted accelerometer predicted EE with an RMSE of 1.22 METs. This RMSE represented an improvement of 32%–71% over previously developed linear regression approaches tested in the study (35). In addition, their input features were very similar to our feature set 4 (the 10th, 25th, 50th, 75th, and 90th percentiles of the acceleration signal and lag-1 autocorrelation), lending additional support that feature sets using percentiles of the accelerometer signal and a measure of temporal change in the signal (i.e., autocorrelation and covariance) are viable for use in different settings and populations. Similarly, Lyden et al. (19) achieved intraclass correlation coefficients above 0.90 and RMSE values of 1.00 MET for predicting EE using an ANN developed from hip-mounted accelerometer data in a true free-living setting, again achieving superior accuracy when compared with that achieved using linear regression approaches. In another study, Rothney et al. (30) achieved a correlation of $r = 0.92$ and RMSE of 0.50 METs when predicting EE using an ANN developed from a hip-mounted accelerometer in a simulated free-living setting. Their slightly better accuracy is likely because of study design, especially given that their use of a linear regression approach to EE prediction yielded a correlation of $r = 0.89$ and an RMSE of 1.00 METs, both of which are considerably better than accuracy achieved in other studies (14,35,36). Despite the slightly higher RMSE values achieved by the ANN in our study, our results are encouraging, given that participants averaged an intensity of 3.3 METs across the duration of the protocol, which is higher than that achieved in many other studies and likely contributes to higher RMSE, as seen in previous work by our research group (24). Taken together,

these studies reinforce the high accuracy for EE prediction achievable using machine learning techniques on data from a single hip-mounted accelerometer, both in laboratory-based and free-living settings.

Our final objective in the current study was to use relatively simple methods for feature extraction and ANN creation and compare sets of input features to identify relevant feature sets that allow for high measurement accuracy while minimizing the complexity of the ANN, both in its structure and in its creation. To achieve the first part of this objective, all data cleaning and feature extraction were conducted using Microsoft Excel. Features were calculated and extracted using simple functions already built into Excel. Although this method is somewhat labor intensive, the key strength of this approach is that it is a viable method for feature extraction without knowledge of or access to powerful, complicated software packages. Use of macros in Excel requires additional knowledge of the software package but can also streamline the process of feature extraction. In addition, ANN creation was conducted using the R statistical software, which is a freely available open-source software package. Writing programs in R is complex and requires skill, but implementing programs that have already been written is relatively simple and can be accomplished with knowledge of only a few commands in R. Use of the *nnet* package for creating ANNs has been successfully accomplished by Staudenmayer et al. (35) and Lyden et al. (19), and considerable detail of the approach, including some of the codes for creating and testing the ANNs, can be found in their manuscripts.

To address the second part of our objective to simplify the use of ANNs, we sought to define an optimal subset of features that can be used without sacrificing measurement accuracy. For the thigh accelerometer, we found that choice of features had minimal effect on measurement accuracy, even in the simplest feature set (set 2) consisting of only mean and variance of the acceleration signal. A very similar set of features was used in a study by members of our research group, in which they were able to classify 14 activities with accuracy greater than 78% with a thigh accelerometer (8). Therefore, this minimal feature set seems to provide strong accuracy for both activity type classification and EE prediction for an ANN created from a thigh-mounted accelerometer.

For the ANNs developed from the hip and both wrist-mounted accelerometers, the ANN created using feature set 2 provided slightly lower correlations and higher RMSE than those provided by set 1; in contrast, the ANNs created using set 4 (the 10th, 25th, 50th, 75th, and 90th percentiles and covariance) yielded similar accuracy to that of set 1 for all four accelerometer placements. A similar set of features to that of set 4 has been used successfully to predict EE and classify activity type in several other studies, albeit these studies used accelerometer counts instead of raw data as inputs into the ANN (19,35,39). Thus, our findings further demonstrate the utility of using the 10th, 25th, 50th, 75th, and 90th percentiles and a measure of similarity in adjacent data windows for predicting EE.

Taken together, the findings of this study support the use of simple-to-compute acceleration features for achieving highly accurate estimates of free-living EE using machine learning. Moreover, choice of the number and type of features seems to alter EE prediction accuracy slightly but the practical significance of these small differences is likely minimal, indicating that researchers may be able to achieve high measurement accuracy for predicting EE using ANN with only a few simple-to-compute accelerometer features.

Study limitations and strengths. The current study had several limitations that should be addressed. First, study participants represented a fairly homogeneous group of college-age adults. Thus, our findings are not necessarily applicable to older populations or children/adolescents and require further validation before use in these populations. Second, the use of a semistructured, simulated free-living setting rather than a true free-living setting could be viewed as a limitation because some studies have used a true free-living setting for ANN creation and validation (19). Third, we did not measure resting $\dot{V}O_2$, which is known to vary across individuals (10). However, like creation of individual HR curves for improving the accuracy of EE prediction using HR, taking individual resting EE into account results in dramatically increased burden on researchers and participants; more importantly, individual resting EE measurement would limit the generalizability of our findings because it is not often possible to measure resting EE in intervention or epidemiologic studies, where accelerometers are often used. Instead of measuring resting $\dot{V}O_2$, it may be useful to include variables such as age and fat-free mass into prediction models because these variables account for the majority of variation in resting $\dot{V}O_2$ (17). However, our study did not find the inclusion of demographic variables such as weight, height, and sex to improve EE prediction when added as input features, so adding variables such as age or fat-free mass may be of limited utility. One final limitation is that we experienced some difficulties in keeping thigh-mounted accelerometers in their proper location during the protocol. Taping monitors on the thigh worked well initially but was less reliable once participants started to sweat. We attempted to secure the monitor using an elastic strap, but this often slipped throughout the session and was less comfortable for participants. There have been several studies that have successfully used thigh-mounted accelerometers for PA and SB measurement, and in future work, we hope to communicate with other researchers regarding optimal strategies for mounting accelerometers on the thigh because of their high measurement accuracy found in this study and their capacity to be worn inconspicuously (i.e., under clothing) to enhance compliance.

There are also several notable strengths of this study. First and foremost, we believe that the semistructured, simulated free-living setting represents a good blend of exerting some control over participant activities while still allowing considerable freedom for the order, intensity, and duration of activities chosen by participants. Troiano et al. (37) identified that PA tends to be performed in short bouts, meaning that a

steady state is rarely achieved during PA in free-living settings; their study provides rationale for the inclusion of transitions and non-steady-state activities in our study because our protocol is more similar to true free-living settings than a typical laboratory-based validation.

A true free-living setting may theoretically have the most real-world generalizability, but a major issue in true free-living settings is lack of a good criterion measure. Doubly labeled water provides an accurate estimate of total EE but cannot measure activity EE or minute-to-minute EE. As another approach, Lyden et al. (19) used a true free-living setting for their ANN creation and validation and direct observation as their criterion measure. Trained observers recorded activities being performed and later used activity classification to predict EE using the compendium of physical activities (1). This approach, although providing detailed information on the activities being performed by participants, is limited because the compendium is an estimate of activity EE and is not necessarily suitable for individual EE prediction. In addition, without imposing some structure in which participants must perform certain activities for a minimum time, it is likely that participants will spend the majority of their time in activities such as sitting and walking and minimal or no time performing other activities, limiting the generalizability of ANNs created from these data. By using a variety of activities across a wide range of intensities and including all transition data during the visit in our analysis, we incorporated many advantages of a free-living setting while also exerting enough control to ensure that a variety of activities were performed. In addition, in the simulated free-living setting, we were able to use a portable metabolic analyzer as our criterion measure, which is widely used as a criterion measure for EE measurement. Thus, the use of a simulated free-living setting provides a step closer to using ANN models for prediction of EE in a true free-living environment and provides additional evidence of their superiority to count-based regression approaches for predicting EE.

Another strength of the study was the use of Microsoft Excel and R statistical software for all stages of data cleaning, feature computation and extraction, as well as ANN creation and validation. These software programs are widely available, and they can be used to create and test machine learning algorithms with minimal experience in computational programming. Finally, it can sometimes be difficult to compare results across studies because of differences in protocol, number and types of activities performed, population used, and modeling approach(es) tested. By simultaneously using four accelerometer placement sites, our study allows for direct comparisons of monitors worn on different places on the body for accuracy in EE prediction.

CONCLUSIONS

In summary, our study provides strong preliminary evidence that ANNs developed from data collected from single accelerometers mounted on the thigh, hip, or wrists provide

highly accurate estimates of EE in a simulated free-living setting. Thigh-mounted accelerometers seem to perform with slightly better accuracy than hip- or wrist-mounted accelerometers, although this difference is fairly small. In addition, we have shown that choice of wrist (dominant vs nondominant) does not affect accuracy of EE prediction. Finally, our study builds off the work of others and highlights ways of reducing complexity of ANN model creation, hopefully allowing for this approach to be used by a wider group of researchers with skills in areas other than activity measurement (i.e., interventionists or epidemiologists). In future studies, we plan to extend our comparison of different placement sites for accuracy of activity classification and measurement of sedentary behavior and sleep across different populations. In addition,

we plan to experiment with using data from multiple monitors to further improve measurement accuracy over that achieved with a single monitor. Finally, we intend to cross-validate the algorithms developed in the study in a true free-living setting to provide support for their future use for EE prediction in epidemiologic or surveillance research.

The authors would like to thank Todd Buckingham, Ryan Hulteen, Brandon Krinock, Elizabeth Lenz, and Stefan Rowland for their assistance in data collection and participant recruitment. In addition, this research was supported by the Michigan State University College of Education and by a Student Award Program grant from the Blue Cross Blue Shield of Michigan Foundation.

The authors attest that they have no conflicts of interest to report.

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

REFERENCES

1. Ainsworth BE, Haskell WL, Herrmann SD, et al. 2011 Compendium of physical activities: a second update of codes and MET values. *Med Sci Sports Exerc.* 2011;43(8):1575–81.
2. Akkermans MA, Sillen MJ, Wouters EF, Spruit MA. Validation of the oxycon mobile metabolic system in healthy subjects. *J Sports Sci Med.* 2012;11(1):182–3.
3. Albinali F, Intille S, Haskell W, Rosenberger M. Using wearable activity type detection to improve physical activity energy expenditure estimation. In: *ACM Conference on Ubiquitous Computing*; Copenhagen (Denmark). 2010 Sep 26–29. pp. 311–20.
4. Calabro M, Lee JM, Saint-Maurice PF, Yoo H, Welk GJ. Validity of physical activity monitors for assessing lower intensity activity in adults. *Int J Behav Nutr Phys Act.* 2014;11(1):119.
5. Cohen S. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale (NJ): Lawrence Erlbaum Associates; 1988. pp. 407–67.
6. Crouter SE, Bassett DR Jr. A new 2-regression model for the Actical accelerometer. *Br J Sports Med.* 2008;42(3):217–24.
7. Dong B, Biswas S, Montoye A, Pfeiffer K. Comparing metabolic energy expenditure estimation using wearable multi-sensor network and single accelerometer. In: *Conference Proceedings - IEEE Engineering in Medicine and Biology Society*; Osaka (Japan). 2013 Jul 3–7. pp. 2866–9.
8. Dong B, Montoye A, Moore R, Pfeiffer K, Biswas S. Energy-aware activity classification using wearable sensor networks. In: *Proceedings of SPIE 8723, Sensing Technologies for Global Health, Military Medicine, and Environmental Monitoring III*. Baltimore (MD). 2013 Apr 29:87230Y1–7.
9. Eslinger DW, Rowlands AV, Hurst TL, Catt M, Murray P, Eston RG. Validation of the GENE Accelerometer. *Med Sci Sports Exerc.* 2011;43(6):1085–93.
10. Ferro-Luzzi A. Inter- and intra-individual variability of the human energy expenditure in the rest position [in Italian]. *Boll Soc Ital Biol Sper.* 1968;44(7):633–7.
11. Freedson PS, Lyden K, Kozey-Keadle S, Staudenmayer J. Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: validation on an independent sample. *J Appl Physiol (1985).* 2011;111(6):1804–12.
12. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc.* 1998;30(5):777–81.
13. Grant PM, Ryan CG, Tigbe WW, Granat MH. The validation of a novel activity monitor in the measurement of posture and motion during everyday activities. *Br J Sports Med.* 2006;40(12):992–7.
14. Hendelman D, Miller K, Baggett C, Debold E, Freedson P. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Med Sci Sports Exerc.* 2000;32(9 Suppl):S442–9.
15. Jean-Louis G, Kripke DF, Cole RJ, Assmus JD, Langer RD. Sleep detection with an accelerometer actigraph: comparisons with polysomnography. *Physiol Behav.* 2001;72(1–2):21–8.
16. John D, Sasaki J, Staudenmayer J, Mavilia M, Freedson PS. Comparison of raw acceleration from the GENE and ActiGraph GT3X+ activity monitors. *Sensors (Basel).* 2013;13(11):14754–63.
17. Johnstone AM, Murison SD, Duncan JS, Rance KA, Speakman JR. Factors influencing variation in basal metabolic rate include fat-free mass, fat mass, age, and circulating thyroxine but not sex, circulating leptin, or triiodothyronine. *Am J Clin Nutr.* 2005;82(5):941–8.
18. Kripke DF, Mullaney DJ, Messin S, Wyborne VG. Wrist actigraphic measures of sleep and rhythms. *Electroencephalogr Clin Neurophysiol.* 1978;44(5):674–6.
19. Lyden K, Keadle SK, Staudenmayer J, Freedson PS. A method to estimate free-living active and sedentary behavior from an accelerometer. *Med Sci Sports Exerc.* 2014;46(2):386–97.
20. Mackintosh KA, Fairclough SJ, Stratton G, Ridgers ND. A calibration protocol for population-specific accelerometer cut-points in children. *PLoS One.* 2012;7(5):e36919.
21. Malina R. Anthropometry. In: Maud C, Foster C, editors. *Physiological Assessment of Human Fitness*. Champaign (IL): Human Kinetics, Inc; 1995. pp. 205–19.
22. Mannini A, Intille SS, Rosenberger M, Sabatini AM, Haskell W. Activity recognition using a single accelerometer placed at the wrist or ankle. *Med Sci Sports Exerc.* 2013;45(11):2193–203.
23. Matthews CE. Calibration of accelerometer output for adults. *Med Sci Sports Exerc.* 2005;37(11 Suppl):S512–22.
24. Montoye A, Dong B, Biswas S, Pfeiffer K. Use of a wireless network of accelerometers for improved measurement of human energy expenditure. *Electronics.* 2014;3(2):205–20.
25. Montoye HJ, Washburn R, Servais S, Ertl A, Webster JG, Nagle FJ. Estimation of energy expenditure by a portable accelerometer. *Med Sci Sports Exerc.* 1983;15(5):403–7.
26. Preece SJ, Goulemas JY, Kenney LP, Howard D, Meijer K, Crompton R. Activity identification using body-mounted sensors—a review of classification techniques. *Physiol Meas.* 2009;30(4):R1–33.
27. R Core Development Team. R: A Language and Environment for Statistical Computing version 2.12.1. Available from: <http://www.r-project.org/>.
28. Rosdahl H, Gullstrand L, Salier-Eriksson J, Johansson P, Schantz P. Evaluation of the Oxycon Mobile metabolic system against the Douglas bag method. *Eur J Appl Physiol.* 2010;109(2):159–71.
29. Rosenberger ME, Haskell WL, Albinali F, Mota S, Nawyn J, Intille S. Estimating activity and sedentary behavior from an accelerometer on the hip or wrist. *Med Sci Sports Exerc.* 2013;45(5):964–75.

30. Rothney MP, Neumann M, Beziat A, Chen KY. An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *J Appl Physiol* (1985). 2007;103(4):1419–27.
31. Ryan CG, Grant PM, Tigbe WW, Granat MH. The validity and reliability of a novel activity monitor as a measure of walking. *Br J Sports Med*. 2006;40(9):779–84.
32. Safrit M, Wood T. *Introduction to Measurement in Physical Education and Exercise Science*. 3rd ed. St. Louis (MO): Mosby; 1995. p. 71.
33. Sirard J, Trost S, Pfeiffer K, Dowda M, Pate R. Calibration and evaluation of an objective measure of physical activity in pre-school children. *J Phys Act Health*. 2005;2(3):345–57.
34. Skotte J, Korshoj M, Kristiansen J, Hanisch C, Holtermann A. Detection of physical activity types using triaxial accelerometers. *J Phys Act Health*. 2014;11(1):76–84.
35. 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* (1985). 2009;107(4):1300–7.
36. Swartz AM, Strath SJ, Bassett 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(9 Suppl):S450–6.
37. 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):181–8.
38. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. *Br J Sports Med*. 2014;48(13):1019–23.
39. Trost SG, Wong WK, Pfeiffer KA, Zheng Y. Artificial neural networks to predict activity type and energy expenditure in youth. *Med Sci Sports Exerc*. 2012;44(9):1801–9.
40. US Department of Health and Human Services. *Physical Activity Guidelines Advisory Committee Report, 2008*. 2008. Washington (DC): US Department of Health and Human Services. Available from: <http://www.health.gov/paguidelines/guidelines>.
41. Welk GJ. Use of accelerometry-based activity monitors to assess physical activity. In: Welk GJ, ed. *Physical Activity Assessments for Health-Related Research*, Champaign (IL): Human Kinetics, Inc.; 2002. pp. 125–42.
42. Zhang K, Pi-Sunyer FX, Boozer CN. Improving energy expenditure estimation for physical activity. *Med Sci Sports Exerc*. 2004;36(5):883–9.
43. Zhang S, Murray P, Zillmer R, Eston RG, Catt M, Rowlands AV. Activity classification using the GENE: optimum sampling frequency and number of axes. *Med Sci Sports Exerc*. 2012;44(11):2228–34.
44. Zhang S, Rowlands AV, Murray P, Hurst TL. Physical activity classification using the GENEA wrist-worn accelerometer. *Med Sci Sports Exerc*. 2012;44(4):742–8.