
Routledge
Taylor & Francis Group

Measurement in Physical Education and Exercise Science

ISSN: 1091-367X (Print) 1532-7841 (Online) Journal homepage: https://www.tandfonline.com/loi/hmpe20

# Validation of Accelerometer-Based Energy Expenditure Prediction Models in Structured and Simulated Free-Living Settings

Alexander H. K. Montoye, Scott A. Conger, Christopher P. Connolly, Mary T. Imboden, M. Benjamin Nelson, Josh M. Bock & Leonard A. Kaminsky

**To cite this article:** Alexander H. K. Montoye, Scott A. Conger, Christopher P. Connolly, Mary T. Imboden, M. Benjamin Nelson, Josh M. Bock & Leonard A. Kaminsky (2017) Validation of Accelerometer-Based Energy Expenditure Prediction Models in Structured and Simulated Free-Living Settings, Measurement in Physical Education and Exercise Science, 21:4, 223-234, DOI: 10.1080/1091367X.2017.1337638

**To link to this article:** https://doi.org/10.1080/1091367X.2017.1337638

Published online: 22 Jun 2017.

MEASUREMENT IN PHYSICAL EDUCATION AND EXERCISE SCIENCE
2017, VOL. 21, NO. 4, 223–234
https://doi.org/10.1080/1091367X.2017.1337638

# Validation of Accelerometer-Based Energy Expenditure Prediction Models in Structured and Simulated Free-Living Settings

Alexander H. K. Montoye[a,b], Scott A. Conger[c], Christopher P. Connolly[d], Mary T. Imboden[a], M. Benjamin Nelson[a], Josh M. Bock[a], and Leonard A. Kaminsky[e]

[a]Clinical Exercise Physiology Program, Ball State University, Muncie, Indiana; [b]Department of Integrative Physiology and Health Science, Alma College, Alma, Michigan; [c]Department of Kinesiology, Boise State University, Boise, Idaho; [d]Department of Educational Leadership, Sports Studies, & Educational/Counseling Psychology, Washington State University, Pullman, Washington; [e]Fisher Institute for Health and Well-Being, Ball State University, Muncie, Indiana

**ABSTRACT**

This study compared accuracy of energy expenditure (EE) prediction models from accelerometer data collected in structured and simulated free-living settings. Twenty-four adults (mean age 45.8 years, 50% female) performed two sessions of 11 to 21 activities, wearing four ActiGraph GT9X Link activity monitors (right hip, ankle, both wrists) and a metabolic analyzer (EE criterion). Visit 1 (V1) involved structured, 5-min activities dictated by researchers; Visit 2 (V2) allowed participants activity choice and duration (simulated free-living). EE prediction models were developed incorporating data from one setting (V1/V2; V2/V2) or both settings (V1V2/V2). The V1V2/V2 method had the lowest root mean square error (RMSE) for EE prediction (1.04–1.23 vs. 1.10–1.34 METs for V1/V2, V2/V2), and the ankle-worn accelerometer had the lowest RMSE of all accelerometers (1.04–1.18 vs. 1.17–1.34 METs for other placements). The ankle-worn accelerometer and associated EE prediction models developed using data from both structured and simulated free-living settings should be considered for optimal EE prediction accuracy.

**KEYWORDS**

ActiGraph; artificial neural network; machine learning; physical activity; validity

## Introduction

Physical activity (PA) and sedentary behavior (SB) both have important independent effects on health. While there is a long history of research on the health benefits of PA (Morris & Heady, 1953; PAGAC, 2008), recent research has focused on improving health by reducing SB (Wilmot et al., 2012). Thus, interventions to improve health should address increasing PA and reducing SB (Department of Health, 2011). Accelerometer-based PA monitors provide an objective measure of both PA (Freedson, Melanson, & Sirard, 1998; Troiano, McClain, Brychta, & Chen, 2014) and SB (Byrom, Stratton, Mc Carthy, & Muehlhausen, 2016; Kozey-Keadle, Libertine, Lyden, Staudenmayer, & Freedson, 2011). While the use of accelerometers is common, there is a lack of consensus for the optimal body location for accelerometer placement and the analytic methods used to translate raw acceleration data into meaningful PA outcome measures. One such measure is energy expenditure (EE), which can be used to calculate number of kilocalories burned but is more often used to characterize activity intensity (i.e., sedentary, light, moderate, or vigorous; Lyden, Keadle, Staudenmayer, & Freedson, 2014; Montoye, Mudd, Biswas, & Pfeiffer, 2015; Staudenmayer, He, Hickey, Sasaki, & Freedson, 2015). Information regarding time spent in each activity intensity is valuable as it can be used to assess adherence to PA guidelines, which in the United States advocates that adults achieve ≥ 150 min/week of moderate-intensity PA, ≥ 75 min/week of vigorous-intensity PA, or an equivalent combination (Physical Activity Guidelines Advisory Committee, 2008).

Early studies utilized accelerometers worn on several body locations, including the hip, lower back, wrist, and ankle (LaPorte et al., 1979; Montoye et al., 1983; Wong, Webster, Montoye, & Washburn, 1981). However, hip-worn accelerometers emerged as a preferred placement site due to the hip's proximity to the center of mass, thus providing a good estimate of whole body movement and higher EE measurement accuracy than accelerometers worn on other body locations (Freedson et al., 1998; Montoye et al., 1983; Swartz et al., 2000). In recent years, accelerometer placement on the wrist has regained popularity to improve wear-time compliance (Fairclough et al., 2016; Troiano et al., 2014; van Hees et al., 2011), and new

**CONTACT** Alexander H. K. Montoye montoyeah@alma.edu Integrative Physiology and Health Science, 614 W. Superior, Alma, Michigan 48801.
Current affiliation for Mr. Nelson is Wake Forest School of Medicine, Winston-Salem, North Carolina; current affiliation for Mr. Bock is University of Iowa, Iowa City, Iowa.

data collection and analytic methods have evolved to allow for markedly improved EE prediction accuracy (Montoye et al., 2015; Montoye, Pivarnik, Mudd, Biswas, & Pfeiffer, 2016c; Staudenmayer et al., 2015). For similar reasons, the ankle may also be an appealing accelerometer location. One study by Karabulut et al. found the ankle to be superior to the hip for recognition of steps (Karabulut, Crouter, & Bassett, 2005), and a study by Dong et al. found superior activity type classification accuracy of an ankle-worn accelerometer to a wrist-worn accelerometer (Dong, Montoye, Moore, Pfeiffer, & Biswas, 2013). With walking being a commonly reported PA and ambulation being an important part of many other lifestyle activities (Physical Activity Guidelines Advisory Committee, 2008), accelerometer data from the ankle may provide information about PA that is not detected by the wrist or hip. Additionally, while we are unaware of ankle-worn accelerometers being used previously for assessment of SB, thigh-worn accelerometers have been shown to be highly accurate for SB assessment (Grant, Ryan, Tigbe, & Granat, 2006; Kozey-Keadle et al., 2011). Given that both thigh- and ankle-worn accelerometers measure primarily lower-body movement, an ankle-worn accelerometer may also have utility for SB assessment.

Recent studies indicate that accelerometers, coupled with advanced machine learning modeling, have higher accuracy for EE and activity intensity prediction than traditional cut-point-based methods and allow for monitor placement on alternate body locations (Montoye et al., 2015; Montoye, Pivarnik, Mudd, Biswas, & Pfeiffer, 2016b; Staudenmayer et al., 2015). Many of the machine learning and cut-point based models used to develop EE predictions algorithms have utilized structured activities in a laboratory setting (Freedson et al., 1998; Puyau, Adolph, Vohra, & Butte, 2002; Sasaki, John, & Freedson, 2011; Trost, Wong, Pfeiffer, & Zheng, 2012), yet, some evidence suggests that predictive models developed using data collected in a strictly controlled setting perform poorly for free-living EE or activity intensity prediction (Lyden, Keadle, Staudenmayer, & Freedson, 2014). Prediction models may be improved by using free-living data or a combination of structured laboratory and free-living data in the model development.

Finally, the accelerometer placement location that is optimal for EE prediction accuracy has not been established (Ellis, Kerr, Godbole, Staudenmayer, & Lanckriet, 2016; Mannini, Intille, Rosenberger, Sabatini, & Haskell, 2013; Trost, Zheng, & Wong, 2014). Accelerometers worn on various locations (i.e., hip, wrist, thigh, and ankle), coupled with machine learning prediction models, have been utilized to assess the type of PA (Montoye et al., 2015; Staudenmayer, Pober, Crouter, Bassett, & Freedson, 2009) and to predict EE during various exercise and free-living activities (Lyden Keadle, et al., 2014; Staudenmayer et al., 2009). While these studies have demonstrated that several accelerometer locations may be acceptable for PA and SB assessment, a lack of a consensus remains as to which location is superior. In addition, it is unclear if incorporating less structured data in the development of a predictive model will improve its accuracy. Therefore, the purpose of this study was twofold: (a) to compare the accuracy of machine learning EE prediction models developed with a combination of structured and simulated free-living data for predicting EE and activity intensity (i.e., sedentary, light, moderate, and vigorous) in a simulated free-living setting, and (b) to compare the accuracy of the prediction models developed with accelerometers worn on different body locations.

## Methods

### Participants

For this study, 30 apparently healthy adults aged 18 to 80 years were recruited via flier, e-mail, and word of mouth. For equal age and sex distribution, 10 participants (five male, five female) were recruited from each of the following age ranges: 18 to 40, 41 to 60, and 61 to 80 years. This study was approved by Ball State University's Institutional Review Board prior to study initiation, and all participants provided written informed consent before beginning the study.

### Procedure

Participants were asked to complete two separate laboratory sessions, each following a period of 2 to 3 hr of fasting, no tobacco or caffeine use, and no exercise. For both visits, participants were fitted with four ActiGraph GT9X Link (ActiGraph LLC, Pensacola, FL, USA) accelerometers and a COSMED K4B$^2$ (COSMED, Rome, Italy) portable metabolic analyzer. The ActiGraph Link is the newest accelerometer produced by ActiGraph, weighing 14 grams and sampling raw data at 30 to 100 Hz with a dynamic range of ± 8 g. The Link's accelerometer sensor is the same as ActiGraph's previous accelerometer, the GT3X+. However, the Link is smaller than the GT3X+ (3.5 × 3.5 × 1 cm vs. 4.6 × 3.3 × 1.5 cm) and has a display that can be activated to provide real-time feedback to the wearer. For the purpose of this study, the display screen was disabled. The accelerometers were worn on the lateral aspect of participants' right ankle, the anterior axillary line of the right hip, and the dorsal side of the left and right wrists. These were set to record raw, triaxial data at a sampling rate of 60 Hz. The COSMED analyzer was worn on the chest via a shoulder harness and connected via sampling lines to a facemask, which was individually fitted over participants' nose and mouth to capture all expired gases. The COSMED has been shown to provide accurate measures of oxygen consumption ($VO_2$) across a range of activities and was used as the criterion measure of $VO_2$ (subsequently converted to EE in METs) for this study (McLaughlin, King, Howley, Bassett, & Ainsworth, 2001; Pinnington, Wong, Tay, Green, & Dawson, 2001).

### Visit 1

The Visit 1 (V1) setting was designed to be highly structured, similar in design to how most accelerometer validation studies are conducted (Freedson et al., 1998; Sasaki et al., 2011), and required ~ 2 to 2.5 hr to complete. During V1, height and weight were taken according to standardized procedures (Malina, 1995). After being fitted with the accelerometers and metabolic analyzer, participants performed 11 activities from a list of 21 possible activities. Each participant started V1 by lying in a supine position for 10 min. The remaining 10 activities and activity order were chosen by the research staff such that each participant performed two additional sedentary behaviors, four household/chore activities, and four ambulatory/exercise activities for 5 min each, with activities performed in order of generally increasing intensity. V1 activities were selected by the research staff so that specific activities within a given category were performed by approximately the same number of participants. A few select activities deemed as common day-to-day activities (e.g., using computer, watching television) were sometimes assigned more often than activities deemed less common (e.g., playing cards). For the sedentary behaviors and household/chore activities, participants were instructed to perform these activities as they normally would outside of the laboratory. For the ambulatory/exercise activities, participants self-selected the speed/intensity of the activity. Following each activity, participants were given 1 to 2 min of rest before starting the next activity. Table 1 shows the 21 categorized activities (sedentary, household/chore, or ambulatory/exercise), the number of participants who performed each activity, and the intensity of the ambulatory/exercise activities in V1.

### Visit 2

Visit 2 (V2) was designed as a "simulated free-living" setting, where participants had a considerable amount of freedom to perform activities as they would during a normal day. Simulated free-living settings have been used successfully for accelerometer validations in several studies in the belief that they provide improved generalizability to free-living settings (Montoye et al., 2015; Staudenmayer et al., 2015). For V2, participants were fitted with the same four accelerometers and COSMED as used during V1. Participants were then given an 80-min block of time, during which they were instructed to perform ≥ 12 activities from Table 1 for 2 to 15 min each, with at least four activities performed in each of the three categories (sedentary, household/chore, ambulatory/exercise). Additionally, it was requested that participants spend ≥ 40 min performing activities in the sedentary category in an attempt to provide a simulation of a typical day, where the average adult spends ≥ 50% of waking hours in sedentary behaviors (Donaldson, Montoye, Tuttle, & Kaminsky, 2016; Matthews et al., 2008). For V2, the order of activities, amount of time spent in each activity (limited to 2–15 min), and the activity intensity were left to the discretion

**Table 1.** Activities performed in the current study.

| Activity | Category | Number of participants performing activity—V1 | Average speed/ workload—V1 | Number of participants performing activity—V2 | Average speed/ workload—V2 |
|---|---|---|---|---|---|
| Lying down | SB | 24 | N/A | 17 | N/A |
| Using computer | SB | 13 | N/A | 17 | N/A |
| Watching television | SB | 11 | N/A | 20 | N/A |
| Writing | SB | 6 | N/A | 4 | N/A |
| Playing cards | SB | 9 | N/A | 18 | N/A |
| Reading | SB | 8 | N/A | 22 | N/A |
| Standing | HC | 14 | N/A | 11 | N/A |
| Dusting | HC | 10 | N/A | 13 | N/A |
| Making bed | HC | 13 | N/A | 14 | N/A |
| Folding laundry | HC | 11 | N/A | 18 | N/A |
| Sweeping | HC | 13 | N/A | 15 | N/A |
| Vacuuming | HC | 12 | N/A | 10 | N/A |
| Gardening; scooping dirt with hand shovel | HC | 14 | N/A | 7 | N/A |
| Picking up items (< 1 kg) off of floor | HC | 10 | N/A | 8 | N/A |
| Slow over-ground walking (miles/hr) | AE | 15 | 2.5 (0.5) | 17 | 2.4 (0.5) |
| Brisk over-ground walking (miles/hr) | AE | 19 | 3.5 (0.5) | 17 | 3.2 (0.6) |
| Treadmill walking (miles/hr) | AE | 16 | 2.9 (0.5) | 22 | 2.6 (1.1) |
| Stationary cycling (Watts) | AE | 12 | 64.0 (36.3) | 16 | 51.3 (16.6) |
| Stair climbing/descending (11-step flights/min) | AE | 12 | 3.6 (1.2) | 14 | 3.2 (1.3) |
| Overground jogging (miles/hr) | AE | 11 | 4.4 (0.7) | 5 | 4.3 (0.7) |
| Treadmill jogging (miles/hr) | AE | 11 | 5.1 (0.8) | 7 | 5.2 (0.9) |

*Note.* V1: Visit 1, structured laboratory setting; V2: Visit 2, simulated free-living setting; SB: sedentary activity category; HC: household/chore activity category; AE: Ambulatory/exercise activity category; Intensities for AE activities are shown as mean (standard deviation).

of the participants. Table 1 shows the number of participants who chose to perform each activity, as well as the average intensity of the ambulatory/exercise activities in V2.

## Data processing

Following data collection, breath-by-breath COSMED data were reintegrated to 30-sec windows. Relative $VO_2$ ($ml\cdot kg^{-1}\cdot min^{-1}$) measured by the COSMED was converted to METs by dividing by 3.5 $ml\cdot kg^{-1}\cdot min^{-1}$. EE during the study was expressed as METs from the COSMED data. The intensity of each 30-sec window was then determined using absolute MET thresholds: ≤ 1.5 = SB, 1.6 to 2.9 = light, 3.0 to 5.9 = moderate, ≥ 6.0 = vigorous, and ≥ 3.0 = moderate-to-vigorous (MVPA; Ainsworth et al., 2011) and summed to determine the total time spent in each activity intensity.

Raw, unfiltered data from the accelerometers were downloaded following each visit. From the raw data, six features (10th, 25th, 50th, 75th, and 90th percentiles of signal and covariance between adjacent windows) were extracted from each of the three accelerometer axes in non-overlapping, 30-sec windows, resulting in 18 total features used. This feature set has been used previously for EE prediction models (Montoye et al., 2015). For V1, only data collected during the activities were included for model training and testing, with the data from the transitions between activities removed. For V2, all data collected during the activities and in the transitions between activities were included in model training and testing.

## Model training for prediction of energy expenditure

Artificial neural networks (ANNs), which are machine learning models capable of being trained to predict continuous or categorical variables, were developed from the data collected during the current study and were trained to predict EE as a continuous variable (in METs) from each accelerometer. After EE was predicted in each 30-sec window, the absolute intensity of each 30-sec window was determined using the same absolute MET thresholds as used for the criterion measure (≤ 1.5 = SB, 1.6–2.9 = light, 3.0–5.9 = moderate, ≥ 6.0 = vigorous, and ≥ 3.0 = MVPA). More detailed description on the ANN framework can be found in a review by Preece et al. (2009). The *nnet* package in the R software (version 2.12.1; R Core Development Team, Auckland, NZ) was used for ANN training and testing. A feed-forward ANN was developed with one hidden layer and five nodes in the hidden layer, and skip-layer connections were not allowed (R Core Development Team). A leave-one-out cross-validation approach was used to maximize the efficiency of the data for ANN training and testing.

As mentioned previously, accelerometers are used to assess free-living activity, but analytic models to predict PA from accelerometer data are most often developed in strictly controlled laboratory settings (similar to V1 of the current study) which bear little resemblance to free-living. Accordingly, analytic models developed using data from strictly controlled settings tend to have high accuracy in the strictly controlled environment but predict with lower accuracy when detecting free-living PA (Bastian et al., 2015; Gyllensten & Bonomi, 2011; Lyden, Keadle, et al., 2014; Montoye et al., 2015). Therefore, when developing modeling techniques, it is important to validate or cross-validate these techniques in a free-living or simulated free-living setting. Due to lack of high-quality criterion measures available to measure free-living activity intensity and EE, simulated free-living settings which take place in a laboratory are popular because they allow for use of high-quality criterion measures while also incorporating some elements of free-living (i.e., choice of activities and method of performing activities). Because V2 (simulated free-living setting) is assumed to be similar to a true free-living setting, accuracy of the ANNs in this study was evaluated by their predictive accuracy when applied to V2 data. ANNs were trained using three different methods (Table 2) to determine if different data used to train the ANNs would impact their accuracy for prediction of EE. The ANNs developed for this study can be accessed at the following link, along with example code and data for their use and implementation: https://drive.google.com/file/d/0B-BgdTzyd2OxUDhwRWR6OTJwZmM.

## Data analysis

Accuracy of the ANN models for EE prediction was evaluated using the variance ($R^2$) accounted for between predicted versus measured EE and root mean square error (RMSE), a measure of individual predictive error. Additionally, time in each activity intensity measured by the COSMED was compared to time in each intensity predicted by the ANNs.

Two primary comparisons were evaluated during this study. The first was to determine if the data used to train the ANNs would impact their accuracy for EE prediction.

**Table 2.** Methods used to train and test ANNs.

| Method | Data used for training/developing ANNs | Data used for testing ANNs |
|---|---|---|
| V1/V2 | V1 | V2 |
| V2/V2 | V2 | V2 |
| V1V2/V2 | V1 and V2 | V2 |

*Note.* ANN: Artificial neural network; V1: Visit 1, structured laboratory setting; V2: Visit 2, simulated free-living setting.

Repeated-measures analysis of variance (RMANOVA) statistics were performed separately for each accelerometer to determine if significant differences in accuracy existed among training methods listed in Table 2. Also, RMANOVAs were performed to determine if significant differences existed between COSMED-measured and ANN-predicted times in each activity intensity.

The second comparison was to determine the accelerometer placement (right hip, right ankle, and right and left wrists) that had the highest accuracy for EE prediction. Separately for each training method, RMANOVA tests were performed to detect differences in accuracy among accelerometer placement locations. When the RMANOVA revealed statistically significant differences, post hoc pairwise comparisons (with a least significant difference correction) were performed. A $p$-value of $< .05$ was used to denote statistical significance. All statistical analyses were conducted using SPSS version 23.0 (IBM Corp., Armonk, North Castle, NY).

To be included in the analysis, complete participant data from all accelerometers and the COSMED for both V1 and V2 was required. The COSMED malfunctioned during testing of two participants, and accelerometer issues (malfunction, initialization, or placement issues) occurred with four separate participants, resulting in 24 participants for the final analyses. Demographics of those included in the final analysis are located in Table 3. The GPower 3.0.10 software (GPower, Dusseldorf, Germany) was used to calculate required effect sizes. In order to detect a significant difference among the four accelerometer placements using RMANOVA, we desired 80% power at an alpha level of .05 to detect an effect size of .50; with these parameters, a sample size of 12 was required. Using the same desired power, alpha, and effect size, a sample size of 12 was required to detect a significant difference among the three training settings (R Core Development Team, 2011). Thus, even after removal of participants for which there were data collection issues, our sample size was more than adequate to detect differences of moderate effect size (Cohen, 1977).

Upon calculation of ANN performance, correlation and RMSE data from two participants were determined to be outliers as they were > 1.5 times the interquartile range outside of the first or third quartiles. Data are presented both with and without the outliers included.

## Results

Due to the large number of statistical analyses performed, $F$ statistics and $p$-values for all RMANOVA tests are presented together in Table 4. $R^2$ data are shown in Table 5. For comparison of accelerometer placements, the ankle-worn accelerometer had $R^2$ significantly higher than all other accelerometer placements for all training methods, with the range for the V1/V2 training method being 9.7% to 17.2% higher than other placements with outliers included and 9.0% to 15.8% with outliers excluded, 9.0% to 27.3% higher than other placements for the V2/V2 training method with outliers included and 10.3% to 29.3% with outliers excluded, and 5.9% to 20.0% higher than other placements for the V1V2/V2 training method with outliers included and 6.8% to 20.0% with outliers excluded. Additionally, the hip-worn accelerometer had significantly higher $R^2$ than both wrist-worn accelerometers for V2/V2 with outliers included (16.4% higher) and excluded (15.3% to 17.2% higher) and significantly higher $R^2$ than the right wrist-worn accelerometer for V1V2/V2 with outliers included (13.3%) and outliers excluded (12.3%). There were no differences in $R^2$ between wrist accelerometer placements.

In comparison of training methods, the V2/V2 method resulted in significantly lower $R^2$ than the other two methods for the left wrist-worn accelerometer with outliers excluded (4.6% to 15.3%). Additionally, there was a nonsignificant trend ($.05 < p < .10$) for higher $R^2$ for the left wrist-worn accelerometer with the V1V2/V2 training method compared to the V1/V2 method with outliers included and for the right wrist-worn accelerometer

**Table 3.** Participant demographics.

| | All ($n$ = 24) | Males ($n$ = 12) | Females ($n$ = 12) |
|---|---|---|---|
| Outliers included | | | |
| Age (years) | 45.8 (19.4) | 49.2 (19.6) | 44.0 (19.5) |
| Height (cm) | 173.9 (8.7) | 179.8 (7.1) | 168.1 (5.8) |
| Weight (kg) | 79.8 (15.5) | 88.8 (13.1) | 70.7 (12.4) |
| Body mass index (kg·m$^{-2}$) | 26.1 (3.5) | 27.4 (3.2) | 24.9 (3.6) |
| Number of participants who are right-hand dominant | 1 | 0 | 1 |
| Outliers excluded | All ($n$ = 22) | Males ($n$ = 12) | Females ($n$ = 10) |
| Age (years) | 46.8 (19.3) | 49.2 (19.6) | 46.0 (19.5) |
| Height (cm) | 174.7 (8.8) | 179.8 (7.1) | 167.9 (6.0) |
| Weight (kg) | 80.2 (15.5) | 88.8 (13.1) | 69.9 (12.7) |
| Body mass index (kg·m$^{-2}$) | 26.1 (3.6) | 27.4 (3.2) | 24.7 (3.8) |
| Number of participants who are right-hand dominant | 0 | 0 | 0 |

*Note.* Data are displayed as mean (standard deviation).

**Table 4.** *F*-statistics and *p*-values for RMANOVA statistical tests.

| Training method/Accelerometer placement | Measure | Outliers included | | Outliers excluded | |
|---|---|---|---|---|---|
| | | *F*-statistic | *P*-value | *F*-statistic | *P*-value |
| *Energy expenditure—Differences among placements* | | | | | |
| V1/V2 | Correlation | 3.376 | *.023* | 10.616 | < *.001* |
| V2/V2 | Correlation | 6.442 | *.001* | 8.322 | < *.001* |
| V1V2/V2 | Correlation | 7.859 | < *.001* | 9.159 | < *.001* |
| V1/V2 | RMSE | 0.331 | .803 | 4.663 | *.005* |
| V2/V2 | RMSE | 5.319 | *.002* | 6.330 | *.001* |
| V1V2/V2 | RMSE | 4.341 | *.007* | 4.351 | *.008* |
| *Energy expenditure—Differences among training methods* | | | | | |
| Ankle | Correlation | 1.067 | .353 | 1.088 | .347 |
| Hip | Correlation | 1.923 | .158 | 1.862 | .168 |
| LW | Correlation | 2.906 | .065 | 3.265 | *.048* |
| RW | Correlation | 1.632 | .207 | 2.694 | .079 |
| Ankle | RMSE | 0.450 | .640 | 1.764 | .184 |
| Hip | RMSE | 0.213 | .809 | 1.147 | .327 |
| LW | RMSE | 1.024 | .367 | 1.555 | .223 |
| RW | RMSE | 1.350 | .269 | 2.013 | .146 |
| *Time (min) in each activity intensity—Difference from criterion* | | | | | |
| V1/V2 | SB | 15.254 | < *.001* | 12.467 | < *.001* |
| V1/V2 | LPA | 18.885 | < *.001* | 15.424 | < *.001* |
| V1/V2 | MPA | 1.246 | .297 | 0.956 | .436 |
| V1/V2 | VPA | 8.057 | *.009* | 2.917 | *.026* |
| V1/V2 | MVPA | 2.335 | .061 | 1.851 | .127 |
| V2/V2 | SB | 15.970 | < *.001* | 13.255 | < *.001* |
| V2/V2 | LPA | 18.311 | < *.001* | 18.331 | < *.001* |
| V2/V2 | MPA | 0.281 | .890 | 0.205 | .935 |
| V2/V2 | VPA | 6.300 | < *.001* | 4.670 | *.002* |
| V2/V2 | MVPA | 1.088 | .367 | 1.095 | .364 |
| V1V2/V2 | SB | 9.120 | < *.001* | 7.003 | < *.001* |
| V1V2/V2 | LPA | 15.245 | < *.001* | 12.398 | < *.001* |
| V1V2/V2 | MPA | 0.718 | .582 | 0.196 | .940 |
| V1V2/V2 | VPA | 7.565 | < *.001* | 6.610 | < *.001* |
| V1V2/V2 | MVPA | 2.110 | .086 | 1.385 | .246 |

*Note*. RMANOVA: repeated-measures analysis of variance; RMSE: root mean square error; V1: Visit 1, structured laboratory setting; V2: Visit 2, simulated free-living setting; SB: sedentary activity category; LPA: light-intensity physical activity category; MPA: moderate-intensity physical activity category; VPA: vigorous-intensity physical activity category; MVPA: moderate- or vigorous-intensity physical activity category.

**Table 5.** $R^2$ values for predicted versus measured EE.

| Training/Testing | Ankle | Hip | Wrist—Left | Wrist—Right |
|---|---|---|---|---|
| Outliers included | | | | |
| V1/V2 | .68 (.05)$^{2}$ | .62 (.04)$^{1}$ | .61 (.04)$^{1}$ | .58 (.04)$^{1}$ |
| V2/V2 | .70 (.05)$^{2}$ | .64 (.04)$^{1}$ | .55 (.05)$^{1,2}$ | .55 (.05)$^{1,2}$ |
| V1V2/V2 | .72 (.05)$^{2}$ | .68 (.04)$^{1}$ | .63 (.04)$^{1}$ | .60 (.04)$^{1,2}$ |
| Outliers excluded | | | | |
| V1/V2 | .73 (.04)$^{2}$ | .67 (.03)$^{1}$ | .65 (.03)$^{1}$ | .63 (.03)$^{1}$ |
| V2/V2 | .75 (.03)$^{2}$ | .68 (.03)$^{1}$ | .59 (.04)$^{1,2}$ | .58 (.04)$^{1,2}$ |
| V1V2/V2 | .78 (.02)$^{2}$ | .73 (.03)$^{1}$ | .68 (.03)$^{1,3,4}$ | .65 (.03)$^{1,2}$ |

*Note*. [1]Indicates significant difference from ankle; [2]Indicates significant difference from hip; [3]Indicates significant difference from V2/V2; [4]Indicates significant difference from V1/V2; EE: energy expenditure; V1/V2: Artificial neural networks (ANNs) were trained using data from V1 and tested using data from V2; V2/V2: ANNs were trained using data from V2 and tested using data from V2; V1V2/V2: ANNs were trained using data from both V1 and V2 and tested using data from V2.

with the V1V2/V2 method compared to the V2/V2 method (outliers excluded).

Table 6 presents RMSE values for each accelerometer location and all of the different training methods. In comparison of accelerometer placements, the ankle-worn accelerometer had significantly lower RMSE than all other accelerometers for V2/V2 (11.6% to 17.9% lower) and V1V2/V2 (11.9% to 15.4% lower) with outliers included and all three training methods with outliers excluded (11.4% to 16.5% lower for V1/V2, 13.1% to 23.1% lower for V2/V2, and 12.7% to 18.3% lower for V1V2/V2). Moreover, the hip-worn accelerometer had significantly lower RMSE than the right wrist-worn accelerometer for V2/V2, with inclusion and exclusion of outliers (8.2% and 11.6% lower, respectively). In comparison of training methods, there were no significant differences for any of the accelerometer placements for any training method, although point estimates trended non-significantly toward lower RMSE in V1V2/V2 for all four accelerometer placements.

While point estimates of time spent in each activity intensity changed slightly with outliers excluded, it did not change statistical significance of differences between measured and estimated time spent in each activity intensity; therefore, data are presented only with outliers included (Figure 1). The ANNs developed for all accelerometer locations and with all training methods significantly underestimated time (in min) spent in SB (measured: 29.1; estimated range: 15.3–22.5 min) and correspondingly overestimated time spent in light-intensity PA (measured: 19.9; estimated range: 28.1–40.7 min). For moderate-intensity PA, the only significant difference between measured and estimated time was from the ankle-worn accelerometer

**Table 6.** RMSE (in METs) for predicted versus measured EE.

| Training/Testing | Ankle | Hip | Wrist—Left | Wrist—Right |
|---|---|---|---|---|
| Outliers included | | | | |
| V1/V2 | 1.18 (0.15) | 1.17 (0.08) | 1.21 (0.07) | 1.28 (0.08) |
| V2/V2 | 1.10 (0.12) | 1.23 (0.13)[1] | 1.33 (0.13)[1] | 1.34 (0.12)[1,2] |
| V1V2/V2 | 1.04 (0.11) | 1.18 (0.13)[1] | 1.23 (0.13)[1] | 1.22 (0.11)[1] |
| Outliers excluded | | | | |
| V1/V2 | 1.01 (0.07) | 1.14 (0.08)[1] | 1.17 (0.06)[1] | 1.21 (0.07)[1] |
| V2/V2 | 0.93 (0.04)[2] | 1.07 (0.07)[1] | 1.19 (0.09)[1] | 1.21 (0.07)[1,2] |
| V1V2/V2 | 0.89 (0.04)[2] | 1.02 (0.07)[1] | 1.07 (0.06)[1] | 1.09 (0.06)[1] |

*Note*. [1]Indicates significant difference from ankle; [2]Indicates significant difference from hip; EE: energy expenditure; RMSE: root mean square error; METs: metabolic equivalents; V1/V2: Artificial neural networks (ANNs) were trained using data from V1 and tested using data from V2; V2/V2: ANNs were trained using data from V2 and tested using data from V2; V1V2/V2: ANNs were trained using data from both V1 and V2 and tested using data from V2.

with the V1/V2 training method, with a mean difference of 1.3 min. Conversely, for vigorous-intensity PA, all accelerometer locations and training methods yielded significant underestimations compared to the criterion (measured: 4.8; estimated range: 1.4–3.4 min) with the exception of the ankle-worn accelerometer with the V1/V2 training method (estimated: 4.0 min). For MVPA (Figure 1e), the ankle- and hip-worn accelerometer estimates were not significantly different from the criterion measure for any training method used. However, the left wrist-worn accelerometer significantly underestimated time spent in MVPA with training methods V1/V2 and V2/V2, and the right wrist-worn accelerometer significantly underestimated time spent in MVPA with training methods V1/V2 and V1V2/V2.

## Discussion

The major findings of this study suggest that, in general, ANN models developed for ankle-worn accelerometers were superior, compared to ANN models developed for hip- or wrist-worn accelerometers, for predicting EE. We are unaware of previous studies that have assessed the accuracy of EE prediction from an ankle-worn accelerometer; however, a study by Mannini et al. (2013) found superior classification accuracy of an ankle-worn accelerometer (95.0%) compared to a wrist-worn accelerometer (84.7%) for

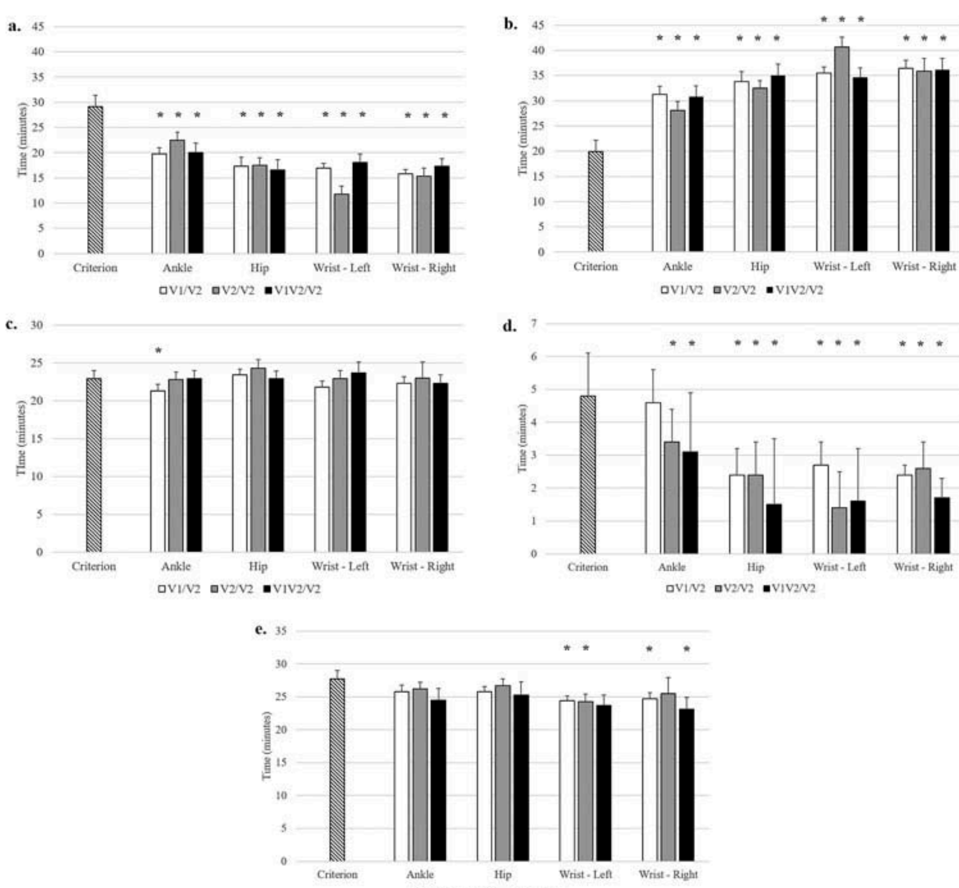


**Figure 1.** Comparison of measured and estimated time spent in physical activity intensities among accelerometer locations and training methods.

*Note*. (a) Sedentary behavior; (b) light-intensity physical activity; (c) moderate-intensity physical activity; (d) vigorous-intensity physical activity; (e) moderate- to vigorous-intensity physical activity. *Indicates significant differences from criterion measure ($p < .05$).

correctly classifying activities into four general categories: sedentary (sitting, computer use, reading, typing), cycling (indoor and outdoor), ambulation (overground/treadmill walking, up/down stairs), and other activities (sweeping, painting with a roller). Therefore, limited available evidence supports the use of an ankle-worn accelerometer for assessment of multiple PA constructs.

In addition, the accuracy of our models developed with data collected from the hip-worn accelerometer was also higher than the accuracy achieved by either the right or left wrist-worn accelerometers. These findings are similar to a previous study by members of our research group, finding that a thigh-worn accelerometer had higher accuracy than hip- or wrist-worn accelerometers for EE prediction, with the hip-worn accelerometer also outperforming the wrist-worn accelerometers (Montoye et al., 2015). It appears that, for the measurement of EE, an accelerometer placed somewhere on the leg is superior to capture movements which translate to EE compared to accelerometers placed on the arms or on the hip when using ANN prediction models. That a leg-worn accelerometer is superior to wrist-worn accelerometers is not surprising given that movement of the arms may not correspond to full body movement, whereas movement of the ankle or thigh is much more likely to reflect walking or other full body movements. However, the superior performance of the ankle and thigh to the hip is informative. Early accelerometer validations found that the highest accuracy of EE measurement came from accelerometers located near the center of the body (Montoye et al., 1983; Swartz et al., 2000). It seems that the thigh/ankle body placement is able to capture ambulatory movements well, even for slow movement speeds, which translates to improved prediction of EE.

Another important finding from this study was that choice of wrist placement had minimal effect on overall EE prediction accuracy. This finding is in agreement with previous work. Members of our research group, using a different population and different brands of accelerometers, found no difference in accuracy of EE prediction models or activity recognition accuracy between wrists using several types of machine learning and linear modeling techniques (Montoye et al., 2016c; Montoye, Begum, Henning, & Pfeiffer, 2017; Montoye, Pivarnik, Mudd, Biswas, & Pfeiffer, 2016a). However, other previous studies indicate that wrist placement (right vs. left) may affect MVPA estimates, which coincides with our finding of differences in estimates of time spent in MVPA compared to the criterion measure (left wrist-worn accelerometer more accurate with V1V2/V2 training method, right wrist-worn accelerometer more accurate with V2/V2 training method) (Esliger et al., 2011; Montoye et al., 2016b). Therefore, available evidence suggests that some outcome variables may be more sensitive to the specific wrist placement than others. This has implications for comparability of data from studies like the National Health and Nutrition Examination study (uses accelerometer worn on non-dominant wrist) and UK Biobank study (uses accelerometer worn on the dominant wrist).

This study also found that the inclusion of training data from both a laboratory and semi-structured, simulated free-living setting (V1V2/V2) offered a small improvement in the accuracy of the prediction models compared to using training data strictly from a controlled laboratory setting (V1/V2) or from only a simulated free-living setting (V2/V2). While the point estimates for $R^2$ and RMSE appeared to indicate improvement in accuracy for all four accelerometer placements, only a few reached statistical significance. There were no differences in the $R^2$ or the RMSE among any of the models developed with data collected from the ankle-, hip-, or right-wrist worn accelerometers. Although the $R^2$ values for the left wrist-worn accelerometer were higher using the V1V2/V2 method compared to V1/V2 or V2/V2 methods with outliers excluded, RMSE was not significantly different. This study was adequately powered to detect an effect size as low as ~ .3; although a larger sample size may have allowed adequate power to detect statistically significant differences among training methods, the effect sizes would be small (Cohen, 1977). Thus, this finding suggests that inclusion of data from structured and simulated free-living settings when developing ANN models for EE prediction offers a small benefit over using data from only one setting.

Our finding that both laboratory and free-living/simulated free-living data be used to optimize EE measurement differs from previous research evaluating PA assessment in the laboratory versus in a free-living setting. Several previous studies demonstrate that activity recognition from accelerometers (i.e., correctly classifying the activity being performed) is much poorer in a free-living setting if the prediction models were developed in a structured laboratory setting compared to a free-living setting (Bastian et al., 2015; Gyllensten & Bonomi, 2011; Sasaki et al., 2016). It may be that activity movements are considerably more variable in the free-living environment compared to the laboratory for different tasks, which would make free-living activity recognition much harder without free-living training data. On the other hand, EE for given tasks (i.e., sedentary activities) may not be as variable and, therefore, would be better assessed even in the free-living using laboratory-trained prediction models. Another possibility is that activity-type prediction accuracy, which has seen classification accuracies > 90% in many laboratory-based studies (Cleland et al., 2013; Dong et al., 2013; Skotte, Korshoj, Kristiansen, Hanisch, & Holtermann, 2014), is better overall than EE prediction. Thus, activity recognition may be easy to perform in the laboratory but difficult to perform outside the laboratory, whereas EE assessment is difficult to perform accurately in both settings. It should be noted that a study by Lyden, Keadle, et al., (2014), which used a hip-worn accelerometer coupled with a machine learning model developed in a laboratory, had much lower EE measurement accuracy in a free-living setting than in the laboratory. However, their study used only the hip-worn accelerometer and used activity counts rather than raw data, which renders their findings difficult to compare directly. It may be that using raw data and/or choosing accelerometer placement locations other than the hip allow for higher accuracy for EE assessment to be achieved in the laboratory and also in a free-living setting.

The present study also evaluated the ability of the models to predict the time spent in each of the activity intensity categories, which is often used to determine if PA guidelines are being met and assess time spent in SB. Each of the accelerometer locations and prediction models underestimated the time spent in SB and overestimated the time spent in light-intensity PA. During SB, people are rarely completely motionless. The models tended to classify these extraneous movements which did not lead to increases in metabolic cost as light-intensity PA. This was consistent across all prediction model methods used.

Conversely, most of the prediction models did well at classifying the time spent in moderate-intensity PA while most tended to underestimate the time spent in vigorous-intensity PA. It should be noted that the criterion-measured time spent in vigorous-intensity PA was considerably lower than the criterion-measured time spent in sedentary time, light- or moderate-intensity PA (4.8 min vs. 20–30 min). Also, cycling was performed by all but nine participants in the study and involves minimal hip or wrist movement. This may partly explain the greater underestimation of vigorous-intensity PA by the hip- and wrist-worn accelerometers compared to the ankle-worn accelerometer. That said, all accelerometers and associated prediction models underestimated time spent at the extreme intensity ends (SB and vigorous-intensity PA) indicated a "bias toward the mean," a common issue observed in other EE prediction studies (Mackintosh, Montoye, Pfeiffer, & McNarry, 2016; Staudenmayer et al., 2015). Analysis of time spent in MVPA revealed that all models developed for the hip- and ankle-worn accelerometers performed well at the group level, whereas the models developed for both wrist-worn accelerometers tended to underestimate the time spent in MVPA. This indicates that the models developed for hip- and ankle-worn accelerometers may have use for MVPA prediction in free-living applications as long as differentiating between moderate- and vigorous-intensity PA is not of interest. However, the same cannot be said for SB, which was underestimated in our study by all models and accelerometer placements.

The accuracy of the ANN models seen in this study are similar to or slightly lower than ANN models developed in previous work but higher than accuracy of cut-points/regression equations developed for count-based, hip-worn accelerometer data. In a previous study, Montoye et al. were able to achieve higher $R^2$ (as high as $R^2$ = .81 vs. .72 in the current study [outliers included]) but similar RMSE (as low as 1.03 METs vs. 1.04 METs in the current study [outliers included]) for EE prediction with an ANN developed for a thigh-worn accelerometer (Montoye et al., 2015). By comparison, Montoye et al. found that a hip-worn accelerometer using the count-based Freedson 1998 equation (Freedson et al., 1998) had poorer accuracy, as indicated by lower $R^2$ (.64) and higher RMSE (1.50 METs), than our study's best-performing ANNs. In another study, Staudenmayer et al. achieved RMSE as low as 1.22 METs for EE prediction using a random forest machine learning model and a wrist-worn accelerometer (dominant wrist) but RMSE of 1.67 METs for EE prediction with a hip-worn accelerometer and the Freedson 1998 count-based equation (Freedson et al., 1998; Staudenmayer et al., 2015); the random forest accuracy is similar to the performance of our ANNs trained in V1V2/V2 for both wrist-worn accelerometers (1.23 METs for left wrist, 1.22 METs for right wrist), but our models far outperformed the hip-worn accelerometer with a count-based regression equation. The slightly lower accuracy for our ANN models compared to the ANNs published by Montoye et al. could be partly attributed to the variety of activities and freedom given to participants in the performance of the activities in this study, both of which increase the variability in the data. Additionally, the previous study by Montoye et al. used a thigh-worn accelerometer, which is closer to the center of mass while still located on the leg, indicating that the thigh may be preferred over the ankle for optimal EE prediction accuracy. However, the similar RMSE in our study to Montoye et al. and Staudenmayer et al. is encouraging, especially since our sample is more diverse in fitness level (data not shown), age, and body mass index (Table 3) than these two studies (Montoye et al., 2015, age 22.1 ± 4.3 years; Staudenmayer et al., 2015, 24.1 ± 4.5 years). While models developed in a similar group of people tend to have high accuracy when applied to data collected in similar individuals, it is likely that those models will have lower accuracy when applied to individuals with physical characteristics different from the training group. Conversely, we purposely sampled 10 participants in each of three age ranges (18–40, 41–60, 61–80) to increase the variability in our sample, thereby improving generalizability to other populations. If the highest possible accuracy is needed for EE assessment, it may be advisable to use a model developed for the specific population being measured. However, a model developed from a more diverse training sample, such as in our study, would be likely preferred when assessing a more varied population.

This study had several strengths. The gold-standard of measured $VO_2$ was used to determine EE and activity intensity. The population included participants from a wide age range with each completing a variety of different activities, which may give these models better generalizability to varied populations. The use of ANNs is also a strength as they have been frequently used in validation studies, offering comparability of these models to those previously developed in different populations and different activities. Finally, the variety of activities used and the two distinct settings (structured laboratory and simulated free-living) may give these results better generalizability for free-living EE measurement because they increase variability in the dataset. There were, however, several limitations to the study design. The sample size was relatively small, and there were no true free-living activities in this study. The difficulty in using a high-quality criterion measure of temporal EE measurement in free-living has precluded free-living EE validations. However, it may be possible to use other PA outcome measures (e.g., activity type, posture, or intensity), which may be more feasible to assess in a free-living setting using direct observation or body cameras (Doherty et al., 2013; Kerr et al., 2013; Lyden, Petruski, Mix, Staudenmayer, & Freedson, 2014). These possibilities should be explored further in future research. Another potential limitation is that no previous studies have compared GT9X Link data to data collected by previous generations of ActiGraph, so the present study cannot confirm that the machine learning models created in this study would work with previous generations of ActiGraph accelerometers. However, the accelerometer sensor in the Link is reportedly the same as in the GT3X+ (the most recent previous ActiGraph accelerometer), so good data comparability would be expected but should be evaluated in future research.

## Conclusion

This study found that optimal EE prediction accuracy was obtained using an accelerometer mounted on the right ankle, although accuracy of hip- and wrist-worn accelerometers was also acceptable. Choice of accelerometer placement should take into account the higher accuracy achieved by accelerometers worn on the leg, but practical considerations such as compliance may also play a role in choice of accelerometer placement. Additionally, ANN models developed using both laboratory and simulated free-living data performed minimally better than models developed using only laboratory or only simulated free-living data, suggesting that EE prediction models should be developed using both structured laboratory and simulated free-living/free-living data.

## Funding

This work was supported by a Ball State University ASPiRE grant and the Ball State University College of Applied Sciences and Technology.

## References

Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett, Jr., D. R., Tudor-Locke, C., … Leon, A. S. (2011). 2011 Compendium of physical activities: A second update of codes and MET values. *Medicine and Science in Sports and Exercise*, *43*(8), 1575–1581. doi:10.1249/MSS.0b013e31821ece12

Bastian, T., Maire, A., Dugas, J., Ataya, A., Villars, C., Gris, F., … Simon, C. (2015). Automatic identification of physical activity types and sedentary behaviors from 3-axial accelerometer: Lab-based calibrations are not enough. *Journal of Applied Physiology*, *118*(6), 716–722. doi:10.1152/japplphysiol.01189.2013

Byrom, B., Stratton, G., McCarthy, M., & Muehlhausen, W. (2016). Objective measurement of sedentary behaviour using accelerometers. *International Journal of Obesity*, *40*(11), 1809–1812. doi:10.1038/ijo.2016.136

Cohen J. Statistical power analysis for the behavioral sciences. New York (NY): Academic Press; 1977.

Cleland, I., Kikhia, B., Nugent, C., Boytsov, A., Hallberg, J., Synnes, K., … Finlay, D. (2013). Optimal placement of accelerometers for the detection of everyday activities. *Sensors*, *13*(7), 9183–9200. doi:10.3390/s130709183

Department of Health, Physical Activity, Health Improvement and Protection. (2011). *Start active, stay active: A report on physical activity for health from the four home countries' Chief Medical Officers*. Retrieved from https://www.sportengland.org/media/2928/dh_128210.pdf

Doherty, A. R., Kelly, P., Kerr, J., Marshall, S., Oliver, M., Badland, H., … Foster, C. (2013). Using wearable cameras to categorise type and context of accelerometer-identified episodes of physical activity. *The International Journal of Behavioral Nutrition and Physical Activity*, *10*, 22. doi:10.1186/1479-5868-10-22

Donaldson, S. C., Montoye, A. H., Tuttle, M. S., & Kaminsky, L. A. (2016). Variability of objectively measured sedentary behavior. *Medicine and Science in Sports and Exercise*, *48*(4), 755–761. doi:10.1249/MSS.0000000000000828

Dong, B., Montoye, A., Moore, R., Pfeiffer, K., & Biswas, S. (2013). *Energy-aware activity classification using wearable sensor networks*. Conference Proceedings of the International Society for Optical Engineering. 87230Y. doi:10.1117/12.2018134

Ellis, K., Kerr, J., Godbole, S., Staudenmayer, J., & Lanckriet, G. (2016). Hip and wrist accelerometer algorithms for free-living behavior classification. *Medicine and Science in Sports and Exercise*, *48*(5), 933–940. doi:10.1249/MSS.0000000000000840

Esliger, D. W., Rowlands, A. V., Hurst, T. L., Catt, M., Murray, P., & Eston, R. G. (2011). Validation of the GENEA accelerometer. *Medicine and Science in Sports and Exercise*, *43*(6), 1085–1093. doi:10.1249/MSS.0b013e31820513be

Fairclough, S. J., Noonan, R., Rowlands, A. V., Van Hees, V., Knowles, Z., & Boddy, L. M. (2016). Wear compliance and activity in children wearing wrist- and hip-mounted accelerometers. *Medicine and Science in Sports and Exercise*, *48* (2), 245–253. doi:10.1249/MSS.0000000000000771

Freedson, P. S., Melanson, E., & Sirard, J. (1998). Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine and Science in Sports and Exercise*, *30*(5), 777–781. doi:10.1097/00005768-199805000-00021

Grant, P. M., Ryan, C. G., Tigbe, W. W., & Granat, M. H. (2006). The validation of a novel activity monitor in the measurement of posture and motion during everyday activities. *British Journal of Sports Medicine*, *40*(12), 992–997. doi:10.1136/bjsm.2006.030262

Gyllensten, I. C., & Bonomi, A. G. (2011). Identifying types of physical activity with a single accelerometer: Evaluating laboratory-trained algorithms in daily life. *IEEE Transactions on Biomedical Engineering*, *58*(9), 2656–2663. doi:10.1109/TBME.2011.2160723

Karabulut, M., Crouter, S. E., & Bassett, Jr., D. R. (2005). Comparison of two waist-mounted and two ankle-mounted electronic pedometers. *European Journal of Applied Physiology*, *95*(4), 335–343. doi:10.1007/s00421-005-0018-3

Kerr, J., Marshall, S. J., Godbole, S., Chen, J., Legge, A., Doherty, A. R., … Foster, C. (2013). Using the SenseCam to improve classifications of sedentary behavior in free-living settings. *American Journal of Preventive Medicine*, *44* (3), 290–296. doi:10.1016/j.amepre.2012.11.004

Kozey-Keadle, S., Libertine, A., Lyden, K., Staudenmayer, J., & Freedson, P. S. (2011). Validation of wearable monitors for assessing sedentary behavior. *Medicine and Science in Sports and Exercise*, *43*(8), 1561–1567. doi:10.1249/MSS.0b013e31820ce174

LaPorte, R. E., Kuller, L. H., Kupfer, D. J., McPartland, R. J., Matthews, G., & Caspersen, C. (1979). An objective measure of physical activity for epidemiologic research. *American Journal of Epidemiology*, *109*(2), 158–168. doi:10.1093/oxfordjournals.aje.a112671

Lyden, K., Keadle, S. K., Staudenmayer, J., & Freedson, P. S. (2014). A method to estimate free-living active and sedentary behavior from an accelerometer. *Medicine and Science in Sports and Exercise*, *46*(2), 386–397. doi:10.1249/MSS.0b013e3182a42a2d

Lyden, K., Petruski, N., Mix, S., Staudenmayer, J., & Freedson, P. (2014). Direct observation is a valid criterion for estimating physical activity and sedentary behavior. *Journal of Physical Activity & Health 11*, (4), 860–863. doi:10.1123/jpah.2012-0290

Mackintosh, K. A., Montoye, A. H. K., Pfeiffer, K. A., & McNarry, M. (2016). Investigating optimal accelerometer placement for energy expenditure prediction in children using a machine learning approach. *Physiological Measurement*, *37*(10), 1728–1740. doi:10.1088/0967-3334/37/10/1728

Malina, R. (1995). Anthropometry. In Peter J. Maud and Carl Foster (Eds.) *Physiological assessment of human fitness* (pp. 205–219). Champaign, IL: Human Kinetics, Inc.

Mannini, A., Intille, S. S., Rosenberger, M., Sabatini, A. M., & Haskell, W. (2013). Activity recognition using a single accelerometer placed at the wrist or ankle. *Medicine and Science in Sports and Exercise*, *45*(11), 2193–2203. doi:10.1249/MSS.0b013e31829736d6

Matthews, C. E., Chen, K. Y., Freedson, P. S., Buchowski, M. S., Beech, B. M., Pate, R. R., & Troiano, R. P. (2008). Amount of time spent in sedentary behaviors in the United States, 2003–2004. *American Journal of Epidemiology*, *167*(7), 875–881. doi:10.1093/aje/kwm390

McLaughlin, J. E., King, G. A., Howley, E. T., Bassett, Jr., D. R., & Ainsworth, B. E. (2001). Validation of the COSMED K4b2 portable metabolic system. *International Journal of Sports Medicine*, *22*(4), 280–284. doi:10.1055/s-2001-13816

Montoye, A. H., Mudd, L. M., Biswas, S., & Pfeiffer, K. A. (2015). Energy expenditure prediction using raw accelerometer data in simulated free living. *Medicine and Science in Sports and Exercise*, *47*(8), 1735–1746. doi:10.1249/MSS.0000000000000597

Montoye, A. H. K., Begum, M., Henning, Z., & Pfeiffer, K. A. (2017). Comparision of linear and non-linear models for predicting energy expenditure from raw accelerometer data. *Physiological Measurement*, *38*(2), 343–357. doi:10.1088/1361-6579/38/2/343

Montoye, A. H. K., Pivarnik, J. M., Mudd, L. M., Biswas, S., & Pfeiffer, K. A. (2016a). Comparison of activity type classification accuracy from accelerometers worn on the wrists, hip, and thigh. *Measurement in Physical Education and Exercise Science*, *20*(3), 173–183. doi:10.1080/1091367X.2016.1192038

Montoye, A. H. K., Pivarnik, J. M., Mudd, L. M., Biswas, S., & Pfeiffer, K. A. (2016b). Validation and comparison of accelerometers worn on the hip, thigh, and wrists for measuring physical activity and sedentary behavior. *AIMS Public Health*, *3*(2), 298–312. doi:10.3934/publichealth.2016.2.298

Montoye, A. H. K., Pivarnik, J. M., Mudd, L. M., Biswas, S., & Pfeiffer, K. A. (2016c). Wrist-independent energy expenditure prediction models from raw accelerometer data. *Physiological Measurement*, *37*(10), 1770–1784. doi:10.1088/0967-3334/37/10/1770

Montoye, H. J., Washburn, R., Servais, S., Ertl, A., Webster, J. G., & Nagle, F. J. (1983). Estimation of energy expenditure by a portable accelerometer. *Medicine and Science in Sports and Exercise*, *15*(5), 403–407. doi:10.1249/00005768-198315050-00010

Morris, J. N., & Heady, J. A. (1953). Mortality in relation to the physical activity of work: A preliminary note on experience in middle age. *British Journal of Industrial Medicine*, *10*(4), 245–254.

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

Pinnington, H. C., Wong, P., Tay, J., Green, D., & Dawson, B. (2001). The level of accuracy and agreement in measures of

FEO2, FECO2 and VE between the COSMED K4b2 portable, respiratory gas analysis system and a metabolic cart. *Journal of Science and Medicine in Sport*, *4*(3), 324–335. doi:10.1016/S1440-2440(01)80041-4

Preece, S. J., Goulermas, J. Y., Kenney, L. P., Howard, D., Meijer, K., & Crompton, R. (2009). Activity identification using body-mounted sensors: A review of classification techniques. *Physiological Measurement*, *30*(4), R1–R33. doi:10.1088/0967-3334/30/4/R01

Puyau, M. R., Adolph, A. L., Vohra, F. A., & Butte, N. F. (2002). Validation and calibration of physical activity monitors in children. *Obesity Research*, *10*(3), 150–157. doi:10.1038/oby.2002.24

R Core Development Team. (2011) *R: A Language and Environment for Statistical Computing*. Version 2.12.1. Retrieved from http://www.r-project.org/

Sasaki, J. E., Hickey, A. M., Staudenmayer, J. W., John, D., Kent, J. A., & Freedson, P. S. (2016). Performance of activity classification algorithms in free-living older adults. *Medicine and Science in Sports and Exercise*, *48*(5), 941–950. doi:10.1249/MSS.0000000000000844

Sasaki, J. E., John, D., & Freedson, P. S. (2011). Validation and comparison of ActiGraph activity monitors. *Journal of Science and Medicine in Sport*, *14*(5), 411–416. doi:10.1016/j.jsams.2011.04.003

Skotte, J., Korshoj, M., Kristiansen, J., Hanisch, C., & Holtermann, A. (2014). Detection of physical activity types using triaxial accelerometers. *Journal of Physical Activity & Health*, *11*(1), 76–84. doi:10.1123/jpah.2011-0347

Staudenmayer, J., He, S., Hickey, A., Sasaki, J., & Freedson, P. (2015). Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. *Journal of Applied Physiology*, *119*(4), 396–403. doi:10.1152/japplphysiol.00026.2015

Staudenmayer, J., Pober, D., Crouter, S., Bassett, D., & Freedson, P. (2009). An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *Journal of Applied Physiology*, *107*(4), 1300–1307. doi:10.1152/japplphysiol.00465.2009

Swartz, A. M., Strath, S. J., Bassett, Jr., D. R., O'Brien, W. L., King, G. A., & Ainsworth, B. E. (2000). Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Medicine and Science in Sports and Exercise*, *32*(9 Suppl), S450–S456. doi:10.1097/00005768-200009001-00003

Troiano, R. P., McClain, J. J., Brychta, R. J., & Chen, K. Y. (2014). Evolution of accelerometer methods for physical activity research. *British Journal of Sports Medicine*, *48*(13), 1019–1023. doi:10.1136/bjsports-2014-093546

Trost, S. G., Wong, W. K., Pfeiffer, K. A., & Zheng, Y. (2012). Artificial neural networks to predict activity type and energy expenditure in youth. *Medicine and Science in Sports and Exercise*, *44*(9), 1801–1809. doi:10.1249/MSS.0b013e318258ac11

Trost, S. G., Zheng, Y., & Wong, W. K. (2014). Machine learning for activity recognition: Hip versus wrist data. *Physiological Measurement*, *35*(11), 2183–2189. doi:10.1088/0967-3334/35/11/2183

van Hees, V. T., Renstrom, F., Wright, A., Gradmark, A., Catt, M., Chen, K. Y., … Franks, P. W. (2011). Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. *Plos One*, *6*(7), e22922. doi:10.1371/journal.pone.0022922

Wilmot, E. G., Edwardson, C. L., Achana, F. A., Davies, M. J., Gorely, T., Gray, L. J., … Biddle, S. J. (2012). Sedentary time in adults and the association with diabetes, cardiovascular disease and death: Systematic review and meta-analysis. *Diabetologia*, *55*(11), 2895–2905. doi:10.1007/s00125-012-2677-z

Wong, T. C., Webster, J. G., Montoye, H. J., & Washburn, R. (1981). Portable accelerometer device for measuring human energy expenditure. *IEEE Transactions on Biomedical Engineering*, *28*(6), 467–471. doi:10.1109/TBME.1981.324820