8-2015

Accuracy of Energy Expenditure Predictions and Activity Identification in Consumer-Based Activity Monitors

James Andrew Woodman
University of Tennessee - Knoxville, jwoodma1@vols.utk.edu

To the Graduate Council:

I am submitting herewith a thesis written by James Andrew Woodman entitled "Accuracy of Energy Expenditure Predictions and Activity Identification in Consumer-Based Activity Monitors." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Kinesiology.

Scott E. Crouter, Major Professor

We have read this thesis and recommend its acceptance:

Eugene Fitzhugh, David R. Bassett

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Accuracy of Energy Expenditure Predictions and Activity Identification in Consumer-Based Activity Monitors

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

James Andrew Woodman
August 2015

DEDICATION

This thesis is dedicated to my family. With your continuous love and support I am able to achieve any goal I set for myself. Thank you Mom and Dad, and thank you Jennifer, for helping me succeed in all of my academic endeavors.

ACKNOWLEDGEMENTS

I would like to acknowledge and thank each of the faculty members serving on my thesis committee. Dr. David Bassett, who opened the door to this education and sparked my interest in research. Dr. Eugene Fitzhugh, who was the first to teach me what research is, how to interpret it, and what it could mean. Most importantly, I would like to express the utmost sincere appreciation of my advising professor, Dr. Scott Crouter. He taught me how to conduct, write, and complete a research study, he carefully guided the trajectory of my education at this university, and his advice allowed me to realize my true potential. For these reasons I am forever grateful.

I also wish to acknowledge my professors and colleagues in the Applied Physiology program. I would not have gotten this far without the help of faculty members and other graduate students within this department. Dr. Dawn Coe and William Boyer have been instrumental in the completion of this research project, and deserve special thanks for their tremendous contributions.

ABSTRACT

INTRODUCTION: Consumer-based physical activity (PA) monitors are increasingly common, and must be validated against criterion measures to determine which models are accurate. Such studies will aid PA intervention research and individual consumers purchasing these devices. **METHODS:** Thirty participants (mean ±[plus or minus] SD; age, 25.5 ± 3.7 years; BMI, 24.9 ± 2.6 kg/m^2[meters squared]) completed a structured PA routine including 11 activities ranging from sedentary behaviors to vigorous intensities. During the routine participants wore an Oxycon portable calorimeter (criterion measure of energy expenditure (EE)), a Basis Peak and Garmin Vivofit on the non-dominant wrist, and three Withings Pulse devices (right hip, shirt collar, dominant wrist). Two repeated measures ANOVAs were used to examine differences between the Oxycon and predicted EE from each monitor, and also examine differences between three Withings placements. Intraclass correlation coefficients (ICC) was calculated to determine reliability of EE predictions between Withings placement sites. Paired samples T-tests were used to determine mean differences between directly observed minutes of structured walking, running, and cycling compared to Basis Band predictions. **RESULTS:** The Basis Peak was the only device not significantly different from measured gross EE for the entire PA routine (P>[is greater than]0.05), however it had large individual error (95% prediction interval, -290.4 to +233.1 kilocalories (kcals)). All devices were significantly different from measured EE for at least eight individual activities (P<[is less than]0.05); Basis (mean error range: 0.4-24.9 kcals, 3.1%-92.8%), Garmin (0.65-32.5 kcals, 4.3%-78.4%), Withings wrist (0.8-34.7 kcals, 5.4%-69.8%), Withings collar (0.6-29.0 kcals, 4.6%-69.9%), and Withings hip (0.9-29.0 kcals, 6.5%-69.9%). Withings ICC ranged from 0.085-0.558. The Basis Peak correctly identified ≥[is greater than or equal to] 92% of directly observed minutes during treadmill walking, over-ground walking, and over-ground running (P>0.05). However, only 40.4% of over-ground cycling minutes were correctly identified and no stationary cycling minutes were identified (P<0.001). **CONCLUSION:** The Basis Peak had the most accurate EE predictions, on average, for the entire PA routine. The Withings Pulse hip and shirt collar predictions are most similar, but inaccurate compared to the criterion. The Basis Peak activity identification function accurately predicts minutes spent walking and running, but not cycling.

Key Words: Accelerometer; Physical Activity; Objective Monitoring; Indirect Calorimetry; Validation, Activity Recognition

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

Physical activity (PA) is an important component of health. It is well documented that PA decreases risk of developing cardiovascular and metabolic disease (1, 4), and increasing daily PA reduces the risk of chronic disease (38). The 2003-2004 National Health and Nutrition Examination Survey indicates that less than 5% of adults in the United States are obtaining the recommended amount of PA (as assessed by accelerometry) (64). When this was reported in 2007, the CDC and ACSM PA public health recommendations were to accumulate at least 30 minutes of moderate intensity PA on at least five days a week (43).

The study of PA in large populations has historically used subjective self-report measures (18). Such studies have provided valuable information to public health researchers, assessed compliance with PA recommendations, and informed the development of current PA guidelines (25, 60). Investigators have shown less than acceptable validity and reliability (<0.50 correlation with objective measures) across many different self-report measures (33, 54). Advances in technology have allowed researchers to rely less on self-report in favor of objective monitoring (e.g. accelerometers) to assess PA. Prediction equations associated with research accelerometers are extensively validated in the literature, and are used to track time spent in light, moderate, and vigorous PA intensities, as well as total daily energy expenditure (EE) (13, 20, 28, 35, 48, 65). Validation studies often include structured activity and predominantly use gas exchange analysis as the criterion measure of EE. Validation of prediction formulas have been conducted with the ActiGraph (13, 48), GENEActiv (20), Actical (28), and Tritrac RT3 (65) accelerometers.

When combining goals and PA log, consumer-based monitors can help produce increases in PA for previously sedentary adults. The device itself does not cause an increase in PA, but could be used as a tool for PA-based prevention of chronic disease (39). These devices have been included in recent PA research (2, 8, 12, 21, 30, 31, 58, 63), and have become common in the market. Given their increasing use, validation of these devices is essential to ensure researchers are obtaining accurate and reliable PA data.

Various methods have been used to validate PA monitors. Some research protocols use structured activity bouts with standardized work rates and intensities for all participants. These methods accommodate laboratory-based criterion measures of EE such as direct calorimetry (whole-room calorimetry) and indirect calorimetry (gas exchange analysis), but are less representative of normal daily activity. Other validation methods use unstructured activity bouts where participants self-select work rates and intensities, which resembles normal daily activity more than structured protocols. Validation using free-living activity involves the least amount of researcher intervention, and most closely represents normal daily activity. Common criterion measures of EE include portable indirect calorimetry, doubly labeled water, and direct observation. While doubly labeled water is very accurate, it cannot measure intensity or duration of activity, and data must be collected for weeks at a time. Portable indirect calorimetry is a validated measure of EE (19, 37, 44, 46) that can measure PA intensity. These devices allow participants to move about with minimal restriction, and data can be collected over a short (<1 day) period of time.

Noah and colleagues (41) conducted a validation study assessing predicted EE of the Fitbit Tracker, Fitbit Ultra, and Actical activity monitors, using portable calorimetry as a criterion measure. This study had five structured activity bouts, each six minutes in length, consisting of one seated resting measure, three treadmill activity bouts (3.5 mph, 0% incline; 3.5 mph, 5% incline; 5.5 mph, 0% incline), and a bout of stair stepping. The findings indicate that all devices had significant differences in mean estimated kilocalories (kcals) ($P<0.006$); mean error ranging from 8.9% - 48.6% across all devices (41). Dannecker and colleagues (17) conducted a validation study of several devices using whole-room calorimetry as an EE criterion measure, and found that during both structured and various unstructured activities the Fitbit Tracker underestimated EE, on average, by 136.2 kcals (27.3%) compared to whole room calorimetry ($P<0.001$). These study results suggest that some consumer-based PA monitors may underestimate EE as compared to criterion measures.

The Basis Band is a consumer-based activity monitor that has been included in only one validation study. Lee and colleagues (32) found the first generation of this device (Basis B1) underestimated EE, on average, by 24% when compared to portable indirect calorimetry. During a 69 minute PA routine the B1 estimated 271.1 kcals for total EE, while the portable indirect calorimeter measured 356.9 kcals total EE, a mean difference of 85.5 kcals. The BodyMedia FIT, Philips DirectLife, Fitbit One, Jawbone UP, NikeFuel Band, and ActiGraph GT3X+ underestimated EE, on average (1.9% - 10.2%). The Fitbit Zip was the only device that overestimated EE (3.7%). In this study, the Basis B1 was placed on the right wrist for all participants, while the two other wrist-worn monitors (DirectLife and Jawbone UP) were placed on the left wrist. The Basis B1

manufacturers state that their monitor should be worn on the non-dominant wrist; since the majority of individuals are right-handed, the placement of this device was likely inconsistent with manufacturer specifications for a majority of participants. This may have been a source of error in the Basis Band estimates obtained by Lee and colleagues. Additionally this device has a novel activity identification function that was not validated by Lee and colleagues; it can identify minutes spent walking, minutes spent running, and minutes spent cycling.

The Garmin VivoFit and Withings Pulse are new activity monitors that have been included in few, if any, validation studies. The VivoFit is a wrist-worn device, and the Pulse can be worn on a hip, a wrist, or clipped to a shirt collar. The VivoFit and Pulse both display a cumulative EE value and do not report minute-by-minute data, in contrast to the Basis Band and research accelerometers. As many consumer-based devices are unable to report minute-by-minute data, validation studies must be designed to accommodate this limitation. Given these limitations, a validation study of the VivoFit and Pulse is not feasible with traditional methods that use minute-by-minute data to calculate point estimates of steady state EE. However, a feasible method is to subtract cumulative EE values displayed at the end of an activity bout from values reported at the beginning of a bout.

In the time since Lee et al. (32) examined the Basis B1, a second generation model has been released (Basis Peak) and the previous model is no longer manufactured. No published study to date has included the second generation Basis Peak. Further studies are needed to establish the accuracy of the Basis Band EE estimation function, and to examine the Basis Band's activity identification function that

was not included in the previous validation study (32). The Withings Pulse has been included in only one validation study to date (21), and the Garmin VivoFit has not been included in any studies to date. Therefore, a study validating EE estimates provided by the Basis, VivoFit, and Pulse could provide useful information to both researchers and consumers.

Statement of the Problem

Validation studies of research accelerometers have traditionally used structured activity bouts and point estimates of steady state EE data to validate prediction formulas associated with the devices. Although many consumer-based PA monitors update EE estimates every minute, most do not provide access to minute-by-minute data, and thus a validation study must be designed to accommodate this limitation. Previous validation studies have found that most consumer-based monitors underestimate EE, and these inaccuracies can mislead consumers who wish to use such devices for EE estimation. Currently there is very little research on the validity of EE estimates provided by consumer-based PA monitors, meaning there is a lack of information available for those choosing a PA monitor for research studies or individual consumer use. Therefore, both researchers and consumers can benefit from a validation of the newest consumer-based monitors.

Statement of Purpose

The first purpose of this study is to examine the accuracy of EE estimates provided by consumer-based activity monitors during a structured PA routine. Devices will include the wrist-worn Basis Band and Garmin VivoFit, as well as three Withings Pulses worn on the wrist, shirt collar, and right hip. The second purpose is to investigate

the relationship of Withings Pulse EE estimates between three placement sites. The third purpose is to determine whether the Basis Band correctly identifies the number of minutes spent in structured walking activity, running activity, and cycling activity.

Research Questions

Question 1: During a structured activity routine, are EE estimates from the Basis Band, Garmin VivoFit, and Withings Pulse valid compared to a criterion measure of EE?

Hypothesis 1: Compared to portable indirect calorimetry, all consumer-based monitors will provide statistically significant mean differences in EE estimates during a structured activity routine.

Question 2: Does the Withings Pulse provide different EE predictions among three placement sites?

Hypothesis 2: The Withings Pulse will provide statistically significant differences among three placement sites.

Question 3: Does the Basis Band correctly identify minutes spent in structured walking activity, structured running activity, and structured cycling activity during the PA routine?

Hypothesis 3a: The Basis Band will not correctly identify all minutes spent in walking activity.

Hypothesis 3b**:** The Basis Band will not correctly identify all minutes spent in running activity.

Hypothesis 3c**:** The Basis Band will not correctly identify all minutes spent in cycling activity**.**

Delimitations

1. Participants shall be between 18-65 years
2. Participants must be able to answer “No” to all questions on a PAR-Q.
3. Participants will be excluded if they are obese, pregnant, or have orthopedic or musculoskeletal issues that would limit activity.
4. Participants must be able to run at five miles per hour on a treadmill for five minutes.
5. Participants will be asked to abstain from alcohol and vigorous exercise for 24 hours prior to lab visits, and will be asked to abstain from eating and caffeine consumption in the four hours preceding lab visits.

Limitations

1. It is assumed that participants will follow guidelines for alcohol, exercise, eating, and caffeine consumption guidelines, though some participants may not follow all guidelines and could therefore affect EE measurement.
2. Participants will be exposed to some risk inherent to vigorous intensity PA, and are expected to answer the PAR-Q truthfully.
3. Weather and campus events may interfere with outdoor activities.
4. Reasonable time commitment for participants will limit the total duration of data collection; data should be collected within one hour and thirty minutes.

CHAPTER II: REVIEW OF LITERATURE

Introduction

PA is any body movement that requires skeletal muscles to contract, therefore increasing EE above resting value (3). Physiologists and epidemiologists measure PA and EE in research, and find important relationships between low levels of PA and chronic disease. Sesso et al. (53) found that total PA and vigorous PA offer a significant reduction in risk of coronary heart disease in middle-aged and older men. The importance of PA is well documented, however decreases in total PA have been observed, driven by decreases in work-related PA, transportation-related PA, and PA in the home (7). At the same time, obesity rates have risen in all segments of the American population (42). Troiano and colleagues examined PA in a probability sample of 6,329 Americans, and found less than 5% of adults accumulate 30 minutes of moderate intensity PA on at least five days a week (64).

Troiano and other researchers use both subjective and objective methods to estimate PA and EE. These methods are validated against criterion measures such as doubly labeled water and calorimetry. The most common objective estimates of PA and EE use accelerometer based PA monitors. Most research-grade activity monitors are not suitable for the average individual; they are marketed to research professionals, provide data that is not as useful to an individual consumer, and require software that is often prohibitively expensive for consumers. Consumer-based devices are marketed to the public for personal PA tracking, and have recently appeared in PA studies (39), though there is limited research available examining the validity of such devices. The purpose of this literature review is to examine various methods of PA and EE

measurement, validation of these methods, and the role of consumer-based devices in research and consumer use.

Measurement of Energy Expenditure

Doubly Labeled Water

The doubly labeled water method is the gold standard for measurement of free-living EE. Doubly labeled water was originally developed for use in animals, and in the past 50 years has been adapted for use in humans. In this method, the $^{2}H^{18}O$ form of water is administered in a loading dose that comes to equilibrium with normal water in the body. Fluid, most commonly urine, is periodically sampled to determine the excretion rates of each isotope of the $^{2}H^{18}O$. The ^{2}H isotope is excreted only in water, while the ^{18}O isotope is excreted in water and carbon dioxide (CO_2). The difference in excretion rates between the two isotopes can be used to measure the amount of CO_2 produced, which is directly related to EE.

The doubly labeled water measure is typically used over a period of one to three weeks, and is used to measure EE during free-living activity. The accuracy of this measure is within 2%-8% of actual EE (kcals), and depends on dose, number of samples, and length of data collection (52). Doubly labeled water cannot provide a measure over intervals of a few days or hours or be used to measure intensity, and only total daily EE can be reported. In order to measure EE over a shorter period of time, other methods such as direct and indirect calorimetry must be used.

Room Calorimetry

Room calorimetry measures a participant's EE in a closed system. This method uses an insulated room equipped with an isothermal, heat sink, convection, or indirect

gas measurement system. Data from these systems are used in prediction equations to calculate EE from a measure of heat energy produced, or oxygen (O_2) consumption and CO_2 production. This method is very accurate but can be expensive, costing upwards of $1,000,000, and they require a full-time technician to maintain and operate the system (34). Sampling periods vary, less than five minutes is needed with isothermal systems, and ten to twenty minutes is required for heat sink and convection systems (34). Room calorimetry has been used as a criterion measure of EE in validation studies of accelerometer prediction equations, but is limited by physical confinement of participants to a small room.

Indirect Calorimetry

This method measures inspired and expired O_2 and CO_2, giving a value for oxygen consumption and CO_2 production. From the ratio of O_2 consumption to CO_2 production (41), stoichiometric calculations are applied to the oxidation reactions of each substrate, from which rate of energy production can be calculated (22). Metabolic carts are common in exercise testing and EE measurement of PA, but are not portable thus limiting their use to the laboratory. Some portable devices that use indirect calorimetry, such as the Cosmed K4b^2 and Oxycon Mobile, have been validated as a method of EE measurement (19, 37, 44, 46). Portable devices offer the advantage of moving about with minimal restriction. These devices use a facemask with tubes c$_o$nnected to a small gas analyzer strapped to a participant's torso, and can be used to collect EE data in free-living situations.

Oxycon Mobile

The Oxycon Mobile (CareFusion Corp, San Diego, CA) is a portable indirect calorimeter that provides measures of oxygen consumption (VO_2) and carbon dioxide production (VCO_2). The device has two units measuring 126 x 96 x 41 mm each, and a total weight of 950 grams (including backpack, battery, and mask). Gas measurement from the unit has ≤3% error for VO_2 and VCO_2, with a sensor range of 0-7 L/min. It uses a compact volume sensor that is insensitive to humidity, is guaranteed to be precise between -10 and +50 degrees C, and has ≤2% error. Breath-by-breath data are collected, and can be summed to intervals ranging from five seconds to one minute. VO_2 and VCO_2 are used to calculate respiratory exchange ratio (RER), from which EE (kcals) can be determined. EE is calculated by multiplying VO_2 (L/min), the RER caloric equivalent value, and time. Referent RER caloric equivalence tables are widely available in published textbooks. This device has been shown to be valid compared to the Douglas Bag method (44) and a lab-based metabolic cart (42).

Measurement of Physical Activity

Self-Reporting Methods

Self-report measures have historically been used in PA studies, though these measures are subjective. Relying on personal recall can result in reporting error, which could affect study outcomes. Lee and colleagues (33) conducted a systematic review of validation studies on a widely used self-reporting measure, the International Physical Activity Questionnaire short form (IPAQ-SF). The researchers sought studies comparing the IPAQ-SF to objective measures such as accelerometers, doubly labeled water, and absolute fitness measures. They found varying but similar results across studies, with

correlations to objective measures of total PA ranging from 0.09 up to 0.39. The IPAQ-SF mostly overestimated PA level, in the range of 36% to 173%. These results indicate the IPAQ-SF is a weak indicator of both absolute and relative PA (33).

R. J. Shephard (54) conducted a thorough review examining the limits of many questionnaires that are used to assess PA. Questionnaires were qualitatively analyzed based on primary elements of assessment such as activity type, intensity, frequency, duration, and aerobic vs. resistance activities. Other elements examined include level of detail in questionnaires, length of assessed period, respondent classification or categorization, and respondent supervision during surveys. Reliability is reported as test-retest correlation, ranging from 0.3 to 0.88 across various studies, and trended towards lower correlation as time between test and retest increases. This research also reports validity of many questionnaires in a variety of populations; Pearson correlations (R) with doubly labeled water ranged from R=0.57 to R=0.79 and correlations with accelerometers or pedometers ranged from R=0.22 to R=0.78. The author conclusively notes that quantitatively interpreting data from questionnaires is not advisable (54).

Accelerometer Methods

Accelerometer-based activity monitors are now a common objective measure of PA in research. Raw acceleration data collected by an accelerometer is interpreted using prediction equations, and provides researchers with data related to EE and time spent in sedentary behaviors and light, moderate, and vigorous PA. These devices are prevalent in PA studies and there is much research supporting the validity of associated prediction equations for PA and EE.

Accelerometer-based activity monitors consist of a micro electro mechanical system (MEMS) that senses the energy of motion and converts it to another measure (voltage) which is then stored as a raw acceleration data in the device's memory (10). Raw data is filtered with frequency parameters, then prediction equations are applied to provide meaningful data on PA and EE. The newest devices use piezoresistive MEMS that exhibit an electrical charge differential in response to mechanical deformation from acceleration, which is detected by a differential capacitance sensor, then recorded as raw data (10). Most current activity monitors incorporate three accelerometer MEMS orthogonally oriented to provide data in three exes. Post-processed data is almost universally reported as counts per epoch (e.g. counts per minute), though differences in proprietary frequency filtering can lead to different counts between devices given the same raw acceleration signal (10). These differences prevent researchers from comparing counts from studies using different devices.

Many early studies of activity monitors yielded EE prediction equations (14, 23, 61). Linear approaches have used the correlation coefficient between counts per epoch and EE measured by calorimetry, whereas nonlinear approaches have been developed using power parameters or logarithmic function. Nonlinear regression could possibly be more accurate for activities that have been shown to have nonlinear correlation between EE and intensity (9, 11).

Machine learning algorithms have also been developed to predict EE from activity monitors. Computer-based pattern recognition uses computers in statistical, syntactic, or neural approaches to classify data (5). Machine learning algorithms can classify continuous outcome variables (e.g. EE) with regression analysis, as well as

categorical classification using statistical clustering (e.g. time spent in sedentary, light, moderate, and vigorous PA) (49, 57).

Validation of Activity Monitors

Validation methods vary greatly in criterion measures and activity protocols used. Indirect calorimetry is a commonly used criterion method in validation studies as it can provide minute-by-minute measures of EE as well as time spent in sedentary behavior and light, moderate, and vigorous PA intensity. For longer duration measurements of one to three weeks, the gold standard of doubly labeled water is used for a precise criterion measurement of total EE.

In a review paper on research-based devices, Plasqui and colleagues (45) focused on free-living studies that used doubly labeled water as a criterion measure of EE. The researchers limited to studies conducted since 2007, 25 articles were chosen for final inclusion in their review. Eighteen accelerometers were identified: BioTel 3dNX, Accusplit AX-120, ActiGraph models 7164 and GT1M, Dynastream AMP-331, Actiheart, ActiReg, BodyMedia SenseWear Pro and Mini, ActivPAL, GENEActiv, Suzuken Lifecorder, Minisun IDEEA, New Lifestyles NL-2000, Tritrac and Tritrac RT3, Tracmor and $Tracmor_D$. Diverse population samples were represented in these studies and included healthy adults of various nationalities, normal and overweight youth, clinical patients, pregnant and non-pregnant women, monozygotic twins, and critically ill children.

Plasqui et al. found wide variability in correlation of EE and PA level derived from doubly labeled water with corresponding estimates from activity monitors. R-values ranged from 0.17 to 0.91. The Actiheart total EE estimates demonstrated high

correlation to doubly labeled water in normal and overweight youth age five to eighteen years, with R=0.86–0.91. The ActivPAL also demonstrated activity EE estimates that were highly correlated to doubly labeled water in cancer patients and healthy controls, with R=0.72–0.89. The Sensewear Pro3 and Mini provided total EE and activity EE estimates that were moderately to highly correlated with doubly labeled water in 30 healthy adults, with R=0.71–0.82 and 0.69–0.84, respectively. The Tracmor activity EE estimates also demonstrated high correlation to doubly labeled water in seven critically ill children, with R=0.85. The authors acknowledge that several factors that could affect the validity of a device, including research goals, population to be studied, outcome variables desired, budget concerns, and activity patterns of participant populations. This diversity of factors affecting validity helps explain why activity monitors are often validated for specific populations, and in certain PA situations such as structured, unstructured, or free-living activity.

ActiGraph

The most current ActiGraph device is a small, lightweight (14 grams), and water resistant model called the ActiGraph GT9X Link. It can record and store raw acceleration data at a frequency ranging from 30-100 Hz, and has gigabytes of memory. To initialize or download data from an ActiGraph, the manufacturer provided Actilife software is required. The ActiGraph and its associated cut-points or prediction formulas are the most heavily validated of research-based activity monitors, with application in a variety of populations (45).

Crouter, Churilla, Bassett (13) conducted a study on validity of select published EE prediction equations related to three research-based accelerometers (ActiGraph,

Actical, and AMP-331). This study had 48 participants, included three structured activity routines for data collection, and used a criterion measure of portable indirect calorimetry (Cosmed K4b^2). Activity routines consisted of six bouts, each ten minutes long, with a one or two minute break between bouts. Most participants completed only one routine, and each routine was completed by at least 20 participants. Multivariate analyses were conducted to compare measured EE to predictions given by each equation. Comparisons were conducted on EE predictions from each activity bout, as well as predicted EE over the entire activity routine. No prediction equation was found to accurately predict EE in all activities; furthermore all equations provided significant underestimations of time spent doing vigorous intensity activities. ($P<0.05$). Some prediction equations provided close estimates for certain activities, with over- or underestimates for other activities. For example, the Freedson 1998 equation (23) was developed using walking and jogging activities, and it predicts those activities well, however it tends to underestimate other activities. The authors note that prediction equations are valid during activities they were developed with (13).

Rothney and colleagues (48) validated the 2006 Crouter 2-regression hip model (C2RM) (13) for predicting EE with the ActiGraph GT1M; criterion measures included whole room indirect calorimetry and doubly labeled water. This study collected data on 34 healthy adults aged 20-67 years for the room calorimetry analysis, and a subset of 22 participants were used for the doubly labeled water analysis. Data was collected during free-living activities lasting about 24 hours in the room calorimeter, with 14 days of data collection for the doubly labeled water protocol during which participants also wore the ActiGraph. Results indicated that during waking hours spent in the room

calorimeter C2RM significantly overestimated (P<0.001) total metabolic equivalents (METs; 1 MET=3.5 ml/kg/min VO_2), but applying a low pass five point median filter altered C2RM estimates to a non-significant difference (P=0.419). Results showed that applying the low pass filter to C2RM predictions provided significantly lower mean differences from doubly labeled water values (P<0.001) compared to unfiltered predictions (48).

Accelerometers are used to assess sedentary behavior, an important outcome measure in PA research. Certain aspects of sedentary behavior can be difficult to assess with self-reporting; short (i.e. less than five minutes) breaks in sedentary time are more easily tracked with activity monitors. Kozey-Keadle and others (29) compared the ActiGraph GT3X predictions of sedentary time, using the low frequency extension filter and one second epochs, to a criterion of direct observation. ActiGraph sedentary cut-points of less than 50, 100, 150, 200, and 250 vertical axis counts per minute were examined in order to determine which had the highest validity for assessing time spent in sedentary behavior. Researchers recruited twenty (five male and fifteen female) overweight and obese (more than 25 kg/m^2) participants who were at least 25 years of age. Two conditions (one week each) were assessed: first a baseline measure during which participants were asked not to change current PA level, then a measure during which participants were advised of daily sitting time reduction strategies. During one day of each condition, trained research assistants conducted direct observation at the participants' place of work, each measure lasting six continuous hours. Results indicated the 150 counts per minute cut-point was most valid for predicting sedentary

minutes when compared to direct observation, with a 1.8% bias (-0.9 minutes) and a 95% confidence interval ranging from 14.1 to 15.9 minutes (29).

In a recent validation study, Lyden and colleagues (35) evaluated nine published EE prediction equations for the ActiGraph and Actical, as well as two proprietary equations for the Tritrac RT3. The researchers recruited 277 participants with a mean ± standard deviation (SD) age of 38.3 ± 12.4 years and a BMI of 24.8 ± 4.2 kg/m^2. Each participant completed a two-part activity protocol consisting of activity bouts lasting seven minutes, with four minutes of rest in between. The first part of the protocol was treadmill activities at 1.34, 1.56, and 2.23 meters per second; participants completed seven minutes of each speed at both 0% and 3% grade, in a randomized order. The second part included activities of daily living conducted at a self-selected pace; all participants ascended and descended stairs, and moved a six kilogram box, while two other activities were selected at random from a list of 14 activities that included cleaning a room, dusting, laundry, mopping, sweeping, vacuuming, washing dishes, painting gardening, mowing, raking, trimming, basketball, and tennis (35).

The findings of Lyden et al. (35) agreed with their previously conducted research, which indicated the ActiGraph, Actical, and Tritrac RT3 cannot provide accurate EE predictions across a wide array of activities (35). It was shown that no prediction equation could accurately classify activities across the entire intensity spectrum (light, moderate, and vigorous). ActiGraph equations used in this analysis included the Freedson 1998 MET equation (23) and kcal equation, the Swartz 2000 equation (61), and the Crouter 2006 C2RM equation (13). The Freedson MET equation provided underestimations of EE for all activities of daily living, with a bias of -2.0 METs and a

95% confidence interval ranging from -0.8 to -0.7. The Swartz equation provides overestimates of EE for light intensity activities and underestimates of vigorous activities, and the C2RM most accurately predicted EE for activities in the 2.5-8.3 MET range. Findings from this study indicated linear regression modeling is most appropriate for predicting EE during activities they were developed with, and accuracy is lost when equations are used to predict EE during dissimilar activities. Importantly, the authors note the nonlinear C2RM has promising potential for discriminating locomotion from activities of daily life, and predicting EE accurately across the spectrum of PA intensity (35).

Sasaki and fellows (51) conducted a study to compare the older ActiGraph GT1M model to the newer GT3X model; 50 participants (28 men, 22 women; mean ± SD age 26.9 ± 7.7 years; BMI) in good health were recruited to complete walking and running activities on a treadmill. An Oxycon Mobile portable calorimeter was used as the criterion measure. Participants were asked to complete four stages; a 4.8 and 6.4 kilometer per hour walk, as well as a 9.7 and 12 kilometer per hour run, with five minutes of rest between these stages. Data was excluded if the participant was unable to complete at least one minute of a treadmill stage, and stages with more than one whole minute of data were averaged for a mean counts per minute during the stage. Between the two ActiGraph models, researchers compared counts from the vertical axis, antero-posterior axis (62), and the vector magnitude of these axes (51).

The findings of Sasaki et al. (51) revealed significant differences in vector magnitude and vertical axis counts per minute between the two ActiGraph models ($P<0.0125$). The GT1M vector magnitude counts were significantly higher than GT3X

vertical axis counts, with a mean percent difference of 21% at 4.8 kilometers per hour, 38% at 9.7 kilometers per hour, and 45% at 12 kilometers per hour. The authors mention that differences between GT1M and GT3X firmware could have contributed to differences in vector magnitude counts. Consistent with previous research, there were no significant differences found for vertical axis counts (51). This study illustrates the challenges one might face when trying to compare data between studies that use different ActiGraph models, and the authors suggest such comparisons be made only with vertical axis counts.

Santos-Lozano and colleagues (50) conducted a validation study of existing ActiGraph prediction formulas and PA intensity cut-points, using a lab-based indirect calorimeter (Oxycon Pro metabolic cart) as the criterion measure. This study used a sample of 97 participants divided into three groups: youth, adults, and older adults. Each participant performed a structured PA routine during data collection. The PA routine began with ten minutes of rest, followed by four treadmill bouts (ten minutes each; three, five, seven, and nine kilometers per hour) with five minutes of rest in between, then ten minutes of a sit-stand-sit activity. The researchers examined accelerometer predictions of MET level during each activity as determined by the Sasaki equation (51), the work-energy theorem provided by Actilife software, and a model that combined the Freedson equation with the work-energy model. A repeated measures ANOVA was used to find within group differences in calorimeter and accelerometer derived METs during each activity, and BIAS was calculated as measured METs minus predicted METs, ± standard deviation. The results of the study indicated that the least accurate prediction formula was the work-energy theorem

applied to adults, with a BIAS of -1.856 ± 2.848. Results further indicated the most accurate prediction formula was the combined Freedson/work-energy theorem applied to children, with a BIAS of -0.053 ± 1.776.

Swartz et al. (61) conducted a free-living validation of the CSA accelerometer using activity domains such as yardwork, family care, housework, recreation, occupation, and conditioning. Seventy participants were recruited, mean ± SD age 41 ± 15 years, BMI 26.0 ± 5.4 kg/m^2. A wrist-worn and a hip-worn CSA accelerometer were worn by each participant as they completed at least one, and up to six activities within each domain. Each domain had at least three individual activities, and researchers collected data for five to 12 participants per individual activity. The Cosmed K4b^2 was used as a criterion measure of EE, and accelerometers were calibrated at the start, middle, and end of data collection. The purpose of the study was to calculate prediction algorithms and examine whether combining data from wrist and hip placements into a bivariate regression increases accuracy of EE predictions for the CSA accelerometer. Researchers found statistically significant improvement when analyzing data combined from both hip and wrist placements (hip, R=0.563, P<0.001; wrist R=0.181, P=0.003; both, R=0.586, P<0.001). However, it was concluded that the improvement was not great enough to warrant the additional time required for analysis and the added cost of one additional accelerometer. The authors also noted limitations to accelerometer use; they cannot identify different walking surfaces and cannot account for additional EE from walking with a load, pushing a weighted object, or ascending stairs.

Strath and colleagues (59) examined the accuracy of five published prediction equations in a study of EE during free-living activity. Ten participants were recruited

(mean ± SD age 26 ± 3 years, BMI 24.4 ± 5.0 kg/m^2) to complete a variety of free-living activities at work or home, with no intervention from the researchers present during data collection. Five to six hours of data were collected for each participant, and breaks were taken every two hours to change the Cosmed battery and allow participants to drink water. EE was converted to METs and each minute was classified by PA intensity; resting/light activity (<3 METs), moderate activity (3-6 METs), or hard activity (>6 METs). Equations examined in this study include the Freedson 1998 equation (23), the Hendelman 2000 equations (walking only, all activities) (27), the Swartz 2000 equation (61), and the Nichols 2000 equation (40).

Strath et al. (59) found the Freedson, Hendelman walking, and Nichols equations overestimated resting/light intensity (13%, 14%, 12%, respectively), and underestimated moderate intensity (60%, 60%, 55%, respectively). The Hendelman equation (all activities) underestimated resting/light intensity (29%), and overestimated moderate (120%). The Swartz equation had no significant mean differences for any PA intensity, but large individual error. These results indicate that no single prediction equation accurately predicted time spent in PA across all levels of intensity, and the authors noted large individual error for each equation examined.

Another validation study of free-living EE (15) was conducted using the ActiGraph GT1M and the Cosmed K4b^2 portable calorimeter. Twelve male and 17 female participants (group mean ± SD; age 25 ± 4.6 years, BMI 25.0 ± 4.6 kg/m^2) were asked to wear the ActiGraph and the Cosmed as a researcher followed them for five to six hours of either work and/or leisure-time apart from work. Activities included sedentary behavior, activities of daily living, recreational activities, and manual labor.

Breaks in data collection were required every two hours in order to change the battery on the Cosmed unit and to allow participants to drink water. The purpose of this study was to compare the Crouter 2006 C2RM equation (13), the Crouter 2010 equation (16) and the NHANES (64) and Matthews (36) cut-points with a gold-standard criterion measure of EE and time spent in sedentary behaviors, and light (1.5-2.9 MET), moderate (3-5.9 METs), and vigorous (≥6 METs) PA.

Crouter et al. (15) found no significant differences for mean EE (METs) between the Cosmed and the Crouter 2010 equation during the entire six hours of measurement ($P>0.05$). The Cosmed measured a mean EE of 1.90 ± 0.68 METs and the Crouter 2010 equation predicted 2.08 ± 0.77 METs. The Crouter 2006 equation predicted 2.32 ± 0.84 METs, a 22% overestimation compared to the Cosmed. Predictions from the 2006 equation were significantly different from both the 2010 equation and the criterion measure ($P<0.05$). The 2010 equation significantly underestimated time spent in sedentary behaviors by 20.8% ($P<0.05$), while it increasingly overestimated time spend in light, moderate, and vigorous PA by 9.5%-62.4% ($P>0.05$). The 2006 equation did not have significantly different estimates of time spent in sedentary behaviors and vigorous activity ($P>0.05$), but significantly underestimated light PA and overestimated moderate PA (34.4%, 76.5%, respectively; $P<0.05$). There were also significant differences observed between the two Crouter equations for light and moderate PA ($P<0.05$). Both NHANES and Matthews cut-points overestimated time spend in sedentary behaviors (9.9%, $P>0.05$; both) and underestimated time spent in vigorous PA (56.7%, $P>0.05$; both). These cut-points were also both significantly different from the Crouter 2010 equation for time spent in sedentary behavior ($P<0.05$). A finding of practical importance

is the large underestimation of moderate PA by the NHANES cut-points. This means that previous work (64) may have underestimated the percent of Americans who meet the 2007 guidelines for moderate PA. Additionally, it was noted that the Crouter 2010 equation may not have been a significant improvement of the 2006 equation, as the newer equation overestimated moderate PA by 44%.

Consumer-Based Activity Monitors

With an increasing number of consumer devices becoming commercially available (31), there is a need for researchers to provide information about their validity. Examples of these activity monitors will be listed and described below, concluding in a review of current validation studies examining selected devices.

Basis Band

The Basis Band is a wrist-worn personal activity monitor with a digital readout. It weighs 44 g, measures 3.6 x 2.7 cm (W x H), and is 27.3 cm long. It has a triaxial accelerometer, an optical blood-flow sensor, two thermometers, and a galvanic skin response sensor. Data from these sensors are used to estimate non-activity heart rate, count steps per day, and predict EE. This device has a touchscreen surface that displays current time by default. Users can swipe the touchscreen in each direction to access watch functions and view data on heart rate and EE. Additionally this device is able to identify three activities: walking, running, and cycling. However, the manufacturer does not explain how these activities are identified. The original model is called the Basis B1 Band, and is no longer in production. The most current model is named Basis Peak, and has not yet been included in any published research.

After this device is charged for the first time, it must be initialized with a smartphone via Bluetooth in order to update firmware on the device. Users must create a profile within the application, and data is stored on the manufacturer's server. Gender, age, height, and weight must be input when creating a profile, and each can be changed at a later time. Data syncing is possible with a computer via the USB docking station, or with a smartphone via freely available iPhone or android applications. The application provides minute-by-minute data on measures such as EE, heart rate, and steps. When using the device, the EE value displayed reflects the total daily EE value up to the current point in time. The touchscreen does not display minute-by-minute data, so a researcher recorded EE values on a data sheet before and after each activity.

Garmin VivoFit

The VivoFit is a novel accelerometer based wrist-worn activity monitor that has not yet been included in any published validation research. It has a digital readout, is water resistant, and uses replaceable coin cell batteries for a one year battery life. It weighs 25.5 g, measures 2.1 cm x 1.05 cm, and comes with two options for wristband size (small, fits 12 to 17.5 cm wrist; large, fits 15.2 to 21 cm wrist). It estimates steps per day, combined distance walking and running per day, and EE. It is unclear if the device predicts distance separate from other activities, and the manufacturer does not release any information explaining how steps, distance, or EE are predicted.

The software interface provided for use with the device is called Garmin Connect, a free application that can be downloaded to a computer or compatible smartphone (android, iPhone). An account must be created within the Garmin Connect application

and a user must input gender, age, height, and weight (each can be changed at a later time). The VivoFit has one physical button that scrolls through each measure (steps, distance, EE) and establishes a Bluetooth connection. To establish this connection, the Garmin Connect application must be installed on the phone or computer, the device must be within one foot of the phone or USB stick, and the button on the wristband must be held down for three seconds. When this device is charged for the first time, it must be initialized via Bluetooth to either a smartphone or a computer that has the provided Bluetooth USB stick plugged in. When using the device, the EE value displayed reflects the total daily EE value up to the current point in time.

Withings Pulse

This accelerometer based activity tracker has been included in one validation study to date (21). It can be worn in a clip on either the hip or the collar area of clothing, on the upper arm, wrist, or placed in a pocket. The Pulse is small (4.3 x 2.2 x 0.8 cm), lightweight (8 g), uses an accelerometer to track EE, and has a fingertip heart rate and SpO_2 sensor. This Withings Pulse estimates steps per day, elevation or distance ascended during hiking and walking up stairs, combined walking and running distance per day, net EE and heart rate. The software interface provided by the manufacturer is called Withings Health Mate, and it is freely available for android and iPhone.

The Withings pulse has one physical button that scrolls through each measure (steps, elevation, distance, EE) and establishes a Bluetooth connection. To establish this connection, the Garmin Connect application must be installed on an iPhone or android smartphone and the device must be within a few inches of the phone. The user holds down the button for three seconds, and the word “sync” appears as the device

attempts a connection. When this device is charged for the first time it must be initialized via Bluetooth so the device can update its software. When using the device, the EE value displayed reflects the activity EE (total EE minus resting EE) value up to the current point in time.

Validation of Consumer-Based Activity Monitors

In a validation study that included many consumer-based activity monitors, Lee and colleagues (32) examined EE estimates over a 69 minute period of activity, using portable indirect calorimetry as a criterion measure. The activity protocol in this study was labeled as “free-living”, however length of each activity and order of activities were both structured. PA data were collected on 60 participants with a mean ± standard deviation (SD) age, 28.6 ± 6.4 years; BMI, 24.3 kg/m^2 ± 2.6; body fat, 17.7% ± 6.2. All participants concurrently wore the Basis B1 Band, BodyMedia FIT, Philips DirectLife, Fitbit One, Fitbit Zip, Jawbone UP Band, NikeFuel Band, and ActiGraph GT3X+ activity monitors along with an Oxycon Mobile 5.0 portable calorimeter. The outcome measure of gross EE was expressed in kcals. The statistical analysis incorporated equivalence testing to determine whether each activity monitor was significantly equivalent to the Oxycon.

Mean absolute percentage error for the Basis B1 was the highest at 23.5%, compared to 9.3% (BodyMedia FIT), 10.1% (Fitbit Zip), 10.4% (Fitbit One), 12.2% (Jawbone UP), 12.6% (ActiGraph GT3X+), 12.8% (Philips DirectLife), and 13% (NikeFuel Band) (32). A 90% confidence interval for the Basis Band estimates of EE fell outside of the proposed equivalence interval (criterion mean +/- 10%) (32). These results indicate that, on average, the Basis B1 EE predictions are not statistically

equivalent to the Oxycon. One important factor than may have influenced this finding was that participants were instructed to wear the Basis Band on the right wrist. Basis states their device should be worn on the non-dominant wrist, so the device placement was likely incorrect for a majority of people.

Stahl and Insana (56) compared the Fitbit to a self-report assessment called Community Health Activities Model Program for Seniors (CHAMPS). Results demonstrated a significant ($P<0.05$) correlation ($R=0.61$) when examining total daily EE estimates between the Fitbit device and the CHAMPS. The researchers concluded the Fitbit is able to make predictions that are acceptably correlative to a self-report PA measure (56), demonstrating the concurrently validity of the Fitbit compared to subjective methods currently in use.

Takacs and colleagues (62) validated the Fitbit One step and distance estimates during treadmill walking at different speeds. Three devices were worn simultaneously in three alternate placements sites: left hip, right hip, and in the front pocket of the dominant leg. Thirty participants volunteered for the study, and each conducted one session of data collection. Participants walked for five consecutive minutes at each of five different speeds (0.90, 1.12, 1.33, 1.54, and 1.78 meters per second). Direct observation was used as a criterion measure for step count; two researchers counted steps during each participant's session. The distance output from the treadmill served as the criterion for distance. Fitbit step counts were not significantly different from either direct observation counts ($P>0.05$) with concordance correlation between 0.97 and 1.00. Percent relative error for distance estimations was below 1.3% at each treadmill speed.

The three Fitbits in different placement sites had a high interclass correlation coefficients (ICC) for estimates of both speed and distance (ICC≥0.94).

Sieverdes and colleagues (55) recruited twenty five participants (twelve male and thirteen female) with a mean ± SD age of 27.6 ± 4.5 years and BMI of 22.4 ± 2.5 kg/m^2 to validate the Mywellness Key during two identical treadmill protocols. Participants were asked to warm-up with a two minute walk at 2.25 kilometers per hour, then five minute stages of walking at 3.22, 5.96, and 6.44 kilometers per hour, followed by a five minute stage of jogging at 7.24 kilometers per hour. All trials were performed at a 0% grade. A ParvoMedics Cart calorimeter was used as a criterion measure of EE. Pearson's correlation was used to analyze inter-device reliability; researchers found R=0.95–0.96 for stages one, two, and three, and R=0.79–0.84 for stage four (P<0.0001). ICC were used for inter-device reliability, for stages one through four ICC=0.993 (95% confidence interval =0.991–0.995). Pearson's correlation was used for validity analysis; VO_2 from the last two minutes of each stage was compared to a computed VO_2 based on Mywellness Key's average counts per minute over the last two minutes. Researchers found R=0.895–0.902 between these two variables (P<0.0001), indicating acceptable validity (55).

Ferguson et al. (21) recently conducted a cross-validation study comparing seven consumer-based activity monitors with two validated research-grade monitors for measures of step count (ActiGraph GT3X+), moderate-to-vigorous PA (MVPA) (ActiGraph GT3X+), and EE (BodyMedia SenseWear). Twenty-one healthy adults (10 male, 11 female) participated in this study, wearing each activity monitor over a 48-hour period. For step count, strong correlations were observed when compared to the

Actigraph GT3X+ (R≥0.94; Nike FuelBand, Striiv, Misfit Shine, Jawbone UP, Withings Pulse, Fitbit Zip, and Fitbit One). For EE predictions compared to the SenseWear, correlations ranged from R=0.74-0.81. These results indicate some consumer-based monitors can make EE predictions similar to research-grade devices, however true accuracy of the consumer devices could not be established. This study did not use a gold-standard criterion such as portable indirect calorimetry, therefore the accuracy of consumer devices could not be examined.

Consumer-Based Activity Monitors in Intervention Studies

Consumer-based activity monitors can be a useful tool for tracking PA and EE in intervention studies. Many consumer-based models are cheaper than research-based monitors, have user-friendly digital displays, and smartphone connectivity that allows users to easily sync data via Bluetooth. Meyer and Hein (39) expect them to play a role in reducing prevalence of cardiovascular disease by facilitating behavior changes that lead to increases in PA, resulting in lower risk of cardiovascular disease. With the continual advancement in wireless data integration, these devices are becoming integrated in social media, user experience is improving (31), and more individuals are using them.

Meyer and Hein (39) recently examined the potential role of consumer-based activity monitors for PA tracking, gathering qualitative data on user experience, and comparison to subjective methods. The authors of this study used two devices concurrently, the Fitbit Ultra and the Garmin Forerunner 110, to collect three weeks of PA data on ten participants in northwestern Germany. The Fitbit was a clip-based model worn in a pocket or attached to clothing, and the Garmin Forerunner 110 is a wrist-worn

device with an integrated GPS unit that comes with a heart rate sensor mounted on a chest strap. The researchers conducted meetings before and after the three weeks, with phone and email correspondence held during data collection. Participants were given simple heart-healthy guidelines such as completing three weekly exercise sessions each at least 30 minutes in duration, obtain six to eight hours of sleep each night, and monitor these habits on a regular basis. After data collection, participants were administered a questionnaire with items such as a system usability scale for the PA monitors, a section on user experience and perceived effect on behaviors, questions regarding future potential for use, and a section asking for self-assessment of how well guidelines were followed (39).

Objectively assessed PA from Meyer and Hein's study was reported as group mean ± SD or % of data within a given range. Results indicated steps per day averaged 10,045 ± 3,243, and active minutes per day averaged 41.4 ± 22.2. Active minutes comprised 22%-31% of endurance training sessions. Subjective self-report of PA was recorded at the end of data collection using a four-point scale that asked participants to classify various PA measures (e.g. steps per day, daily activity minutes, and duration of sleep) into a range of four quartiles. The researchers found results agreed with previous research in that vigorous intensity PA had strong correlation between subjective and objective measures, and fair to moderate correlation between subjective and objective measures of moderate PA (39). The authors' qualitative assessment of participants' experience indicated that there was excitement about using the Fitbit and an appreciation for daily feedback on PA. Overall, participants reported positive feedback; the system helped track their activity and provided motivation for healthier behavior.

CHAPTER III: MANUSCRIPT

INTRODUCTION

Physical inactivity is detrimental to health, and in the U.S. it is considered the largest public health issue of the 21st century (5). Based upon objective measurement of physical activity (PA), it is estimated that less than 5% of U.S. adults meet the guideline of 30 minutes of moderate intensity PA on at least five days a week (64). Regular PA has been shown to reduce the risk of type II diabetes, heart disease, certain types of cancer, and even delay age-related decline in cognition and functional physical capacity (5).

Objective monitoring can be a useful method to track PA and has been shown to help motivate sedentary individuals to increase their PA (63). Consumer-based PA monitors are a widely available form of objective monitoring; use of these devices is trending upwards, and they have been employed in a variety of research applications (2, 39). Different models of the Fitbit, one of the popular consumer-based monitors, have been included in numerous PA monitor validation studies (7, 16, 20, 23, 29, 31, 41, 56, 58, 62). Other consumer-based devices represented in validation studies are the Nike FuelBand (7, 20, 23, 29, 58), Jawbone Up (7, 20, 29, 58), and Misfit Shine (20).

Previous research finds some consumer-based activity monitors can accurately estimate energy expenditure (EE), with Pearson correlations (R) ranging from R=0.74-0.81 observed between consumer monitors (Misfit Shine, Jawbone UP, Withings Pulse, Fitbit Zip and One) and a research-grade device (BodyMedia SenseWear) (20). EE is a common measure of PA provided by many consumer-based PA monitors, and there is a wide range of validity in EE predictions from these devices (16). For example, the Fitbit

and Fitbit Ultra have been shown to have high correlations to measured EE during walking and jogging; intraclass correlation coefficients (ICC) ranged from 0.56-0.72 and 0.81-0.87, respectively (41). In contrast, other consumer-based monitors, such as the Basis B1 (first generation), have compared poorly to measured EE during 69 minutes of structured activities (R=0.136; mean absolute percentage error (MAPE), 23.5%) (31).

With an increasing number of consumer-oriented monitors being released (30), it is important for researchers to validate these devices. To our knowledge the Basis Peak (second-generation) and the Garmin Vivofit have not been included in previous validation studies, and the Withings Pulse has only been included in one validation study (20). The Basis Peak has a function to identify time spent in walking, running, and cycling activities but this function was not examined in previous research. Additionally, the Withings Pulse manufacturer says their device can be worn anywhere, and previous research has not examined the effect of placement on EE predictions. Therefore, the primary purpose of this study is to examine the accuracy EE predictions from these consumer-based PA monitors, compared to a criterion measure of portable calorimetry, during structured PA. The second purpose of this study is to investigate the relationship of EE predictions among three placement sites for the Withings Pulse. The third purpose of this study is to validate the Basis Peak's activity identification function, which estimates time spent walking, running, and cycling.

METHODS

Participants

Twenty-eight participants (mean ± SD; age, 25.5 ± 3.7 years; BMI, 24.9 ± 2.6 kg/m^2) were recruited via word of mouth, flyers, email, and social media from The

University of Tennessee, Knoxville and surrounding areas. Exclusion criteria included pregnancy, obesity (BMI ≥ 30 kg/m^2), orthopedic or musculoskeletal issues that would limit activity, or not being able to run on a treadmill for five minutes at 134.1 $m \cdot min^{-1}$ and 0% incline. Upon visiting the lab, participants were given a verbal explanation of the study, and screened for exclusion criteria using the Physical Activity Readiness Questionnaire (PAR-Q). Prior to participation, participants signed an informed consent form. This study was conducted with approval from the University of Tennessee Institutional Review Board.

Procedures

Participants were asked to abstain from alcohol and vigorous exercise for 24 hours prior to data collection, and abstain from eating and caffeine consumption for four hours prior. Weight and height were measured in light clothing and no shoes, using a physician's scale and stadiometer, respectively. Participants were fitted with a heart rate monitor, A Basis Peak and Garmin VivoFit on the non-dominant wrist, three Withings Pulse (dominant wrist, shirt collar, and right hip) and an Oxycon portable calorimeter. Participants were then asked to complete a structured PA routine consisting of 11 activities that lasted a total of approximately 90 minutes. EE values from all activity monitors were recorded immediately before and after each activity. Participants were asked to complete ten minutes of supine lying rest and five minutes of the other 10 activities, with a minimum of two minutes of transition time between activities. Activities were completed in the following order:

1) Supine rest
2) computer usage in a seated position

3) folding clothes in a seated position

4) sweeping a floor

5) treadmill walking at 80.5 m/min and 7% incline

6) continuously ascending and descending stairs

7) over-ground walking at a self-selected pace on a sidewalk, track, or in a gym

8) over-ground running at a self-selected pace on a sidewalk, track, or in a gym

9) seated rest

10) over-ground cycling outside on a standard bicycle at a self-selected pace

11) cycling on a Lode ergometer at 100 watts

Devices

Oxycon: The Oxycon Mobile (CareFusion Corp, San Diego, CA) is a portable indirect calorimeter that provides measures of oxygen consumption (VO_2) and carbon dioxide production (VCO_2). The device has two units measuring 126 x 96 x 41 mm each, and a total weight of 950 grams (including backpack, battery, and mask). Breath-by-breath data are collected, and can be summed to intervals ranging from five seconds to one minute. This device has been shown to be valid compared to the Douglas Bag method (46) and a lab-based metabolic cart (44). The device was calibrated before each test; procedures consisted of ambient air sampling, volumetric calibration with a 3 liter syringe, and gas calibration using a mixture of 16% O_2 and 4% CO_2.

Basis Peak: The Basis Peak (Basis Science, Inc., San Francisco, CA) wrist-worn activity monitor is lightweight (44 grams), measures 3.6 x 2.7 cm, has a 27.3 cm wristband, and is water resistant up to 5 ATM. It has a battery life of 2-3 days, depending on use, and is charged through a docking station connected to a computer.

Sensors within this device include a triaxial accelerometer, two thermometers, an optical blood-flow sensor, and a galvanic skin response sensor. Data from these sensors are used to estimate heart rate, steps taken, and predict gross EE that are displayed on a touchscreen. Additionally, this device uses its sensors to identify how many minutes are spent in three activities: walking, running, and cycling. A profile was created with the MyBasis application using the researcher's smartphone. The same smartphone was used to edit the profile for each participant; gender, age, height, and weight were modified, and then synced to the Basis device. All data is owned and stored on company servers, which was accessed via smartphone and computer-based web browser.

Garmin VivoFit: The Garmin VivoFit (Garmin Ltd., Schaffhausen, Switzerland) is a water resistant wrist-worn activity monitor that weighs 25.5 grams and measures 2.1 cm x 1.05 cm. It includes two band sizes to accommodate wrist circumferences ranging from 12 to 21 cm, and utilizes a coin cell battery that provides up to a year of battery life. A digital readout displays estimates of steps taken, walking and running distance, and gross EE. A profile was created with the Garmin Connect application using the researcher's smartphone. The same smartphone was used to edit the profile for each participant; gender, age, height, and weight were modified, and then synced to the Garmin device.

Withings Pulse: The Withings Pulse (Withings, Issy les Moulineaux, France) is a small (4.3 x 2.2 x 0.8 cm), lightweight (8 grams) device that is not water resistant. Variables estimated include steps, walking and running distance, and net EE. This device does

not require participant data to be entered prior to use, and can be worn on the hip, shirt collar, or either wrist.

Data Processing

Breath-by-breath VO_2 and VCO_2 from the Oxycon were used to compute EE, which was then averaged over a 15-sec period, and used as the criterion variable that the activity monitors would be compared against. EE data were analyzed for the entire PA routine, as well as for individual activities. Oxycon EE values were obtained for the entire routine (including transitions) by summing all 15-sec values. To examine EE during individual activities, rate of EE was calculated and presented as kilocalories (kcals) per minute. To calculate Oxycon kcals per minute for analysis of individual activities, steady-state EE data were required. Therefore, Oxycon data between minute 2:30 to 4:30 were averaged to obtain criterion values for kcals per minute. To calculate consumer monitor kcals per minute, total gross EE predictions for each activity were divided by the activity duration.

The activity monitor EE predictions for the entire routine were computed as the difference of values from the beginning of the PA routine to the end of the routine. Predictions for individual activities were computed by subtracting the EE value at the end of each activity from the EE value at the start of the activity. Since the Withings Pulse provides estimates of net EE while the Garmin VivoFit and Basis Peak estimate gross EE, we chose to convert all values to gross EE so a direct comparison could be made. Thus, basal metabolic rate (BMR) for each participant was calculated using the Harris-Benedict equation (25), which was added to the net EE value from the Withings Pulse for an estimate of gross EE.

Data for the Basis Peak identification of time spent in structured walking, running, and cycling minutes were obtained via the MyBasis app. The activity routine commenced on the minute according to the internal clock in the Basis, such that the Basis measures of time spent in structured activities could be compared to direct observation of these behaviors. All structured activity bouts were started on the minute, however not all bouts ended on the minute. To ensure only valid data were included, the first and last whole minute of each activity bout were excluded from this analysis.

Statistical Analysis

All analyses were conducted using IBM SPSS statistics software version 22 (IBM, Armonk, NY). For all analyses, an alpha of 0.05 was used to denote statistical significance and data are presented as mean ± standard deviation. Repeated measures ANOVAs were used to examine differences between measured EE (Oxycon) and predicted EE from each consumer-based monitor. This was conducted for the entire PA routine, as well as for each structured activity. When necessary, within-subjects contrasts were used to determine where significant differences existed between measured and predicted EE values. Bland-Altman plots were created to show the range of each monitor's individual error, using dashed lines to represent a 95% prediction interval (95% PI) and a solid line to represent the mean error score. Accurate devices will have a narrow 95% PI a mean error score close to zero.

A separate repeated measures ANOVA was used to test for mean differences between the different Withings Pulse placement sites. This was completed for analyses of the entire activity routine. When needed, Bonferroni post-hoc testing was used to find which placements were significantly different. ICC was calculated to examine reliability

among the three Withings Pulse placements over the entire PA routine. Since any systematic differences between placement sites is important, all ICC tests were performed for absolute agreement. Excellent reliability is defined as ICC≥0.75; fair to good, 0.4≤ICC<0.75; and poor, ICC<0.4 (47).

Paired samples T-tests were used to determine mean differences between directly observed minutes and Basis Peak identified minutes of treadmill walking, over-ground walking, over-ground running, over-ground cycling, and stationary cycling.

RESULTS

Physical characteristics of the participants are presented in Table 1. During analysis of individual activities, one participant's Oxycon data was not retrievable due to downloading error. On some occasions, head movement during testing caused temporary occlusions in the Oxycon sampling line resulting in the exclusion of eight out of 297 individual activity bouts: over-ground running (3), seated rest (2), over-ground cycling (1), and stationary cycling (2) bouts. Data from the occlusions were also removed for analysis of the entire PA routine.

Table 1. Physical characteristics of participants. Values are mean ± SD (range).

	Male (n = 20)	Female (n = 8)	All Participants (N= 28)
Age (years)	26.3 ± 4.9 (21.5-33.8)	23.4 ± 1.9 (21.5-26.3)	25.5 ± 3.7 (21.5-33.8)
Height (cm)	179.7 ± 5.2 (171.5-191.0)	165.9 ± 5.0 (157.5-173.0)	175.7 ± 8.1 (157.5-191.0)
Weight (kg)	83.3 ± 8.4 (65.6-96.2)	62.3 ± 4.5 (56.2-68.8)	77.3 ± 12.2 (56.2-96.2)
BMI (kg/m^2)	25.8 ± 2.2 (21.2-29.9)	22.7 ± 2.5 (19.6-26.1)	24.9 ± 2.6 (19.6-29.9)

BMI: body mass index.

For the entire PA routine, there were significant differences between the measured EE and predicted EE from all activity monitors (P<0.05), except for the Basis

Peak (P=0.257; Table 2, Figure 1). On average, the Basis Peak predicted EE was 7% higher than measured EE. The Garmin VivoFit significantly underestimated measured EE by 44.8% (P<0.001) and all three Withings placement sites underestimated measured EE by 41.6%-64.4% (P<0.001).

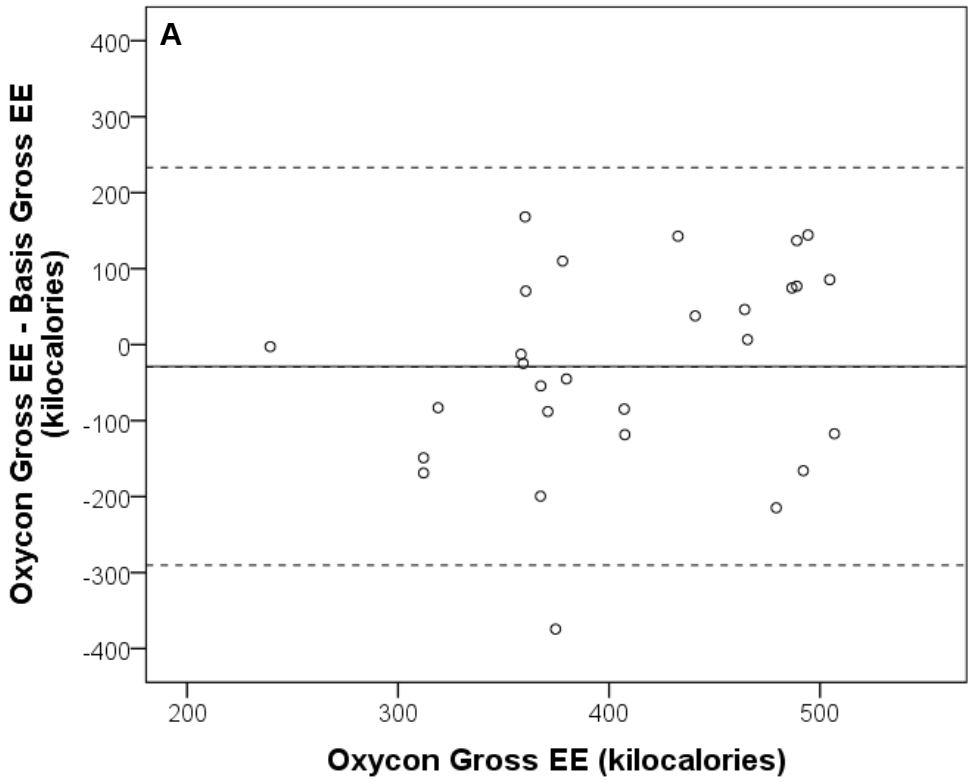
Table 2. Gross energy expenditure (kilocalories) and mean difference (device minus Oxycon) for entire physical activity routine.

Device	Mean ± SD	Mean Difference ± SD	P value
Oxycon	407.8 ± 71.4	—	—
Basis Peak	436.5 ± 132.6	-28.7 ± 131.1	0.257
Garmin VivoFit	225.0 ± 43.1	182.8 ± 44.6	< 0.001
Withings Wrist	145.2 ± 17.4	262.7 ± 60.1	< 0.001
Withings Shirt Collar	234.9 ± 31.6	172.0 ± 56.7	< 0.001
Withings Hip	238.1 ± 31.3	169.8 ± 56.6	< 0.001

Figure 1 shows the Bland-Altman plots for the gross EE during the entire PA routine. The Basis Peak had the lowest mean error for predicting EE (28.7 kcals), however it had large individual error; 95% PI, -290.4 to +233.1 kcals. Other devices had greater mean error (169.8-262.7 kcals), with less individual error; 95% PI, +93.8 to +271.8 kcals (Garmin VivoFit), +142.7 to +382.6 kcals (Withings wrist), +59.8 to +286.2 kcals (Withings shirt collar), and +56.7 to +282.8 kcals (Withings hip).

Table 3 shows the mean measured and predicted gross EE for all 11 individual activities. The Basis Peak significantly over- or under-estimated eight activities (P<0.05), with mean differences ranging from 0.3 to 24.9 kcals/min (38.1%-84.2%). The Garmin VivoFit significantly underestimated all individual activities except seated computer use (P>0.05), with mean differences ranging from 0.1 to 2.8 kcals/min (7%-81%). (P<0.05). The Withings wrist placement significantly underestimated 10 individual

Figure 1. Bland-Altman plots of gross energy expenditure over the entire activity routine for A) Basis Peak, B) Gamin VivoFit, C) Withings Pulse wrist, D) Withings Pulse shirt collar, E) Withings Pulse hip.



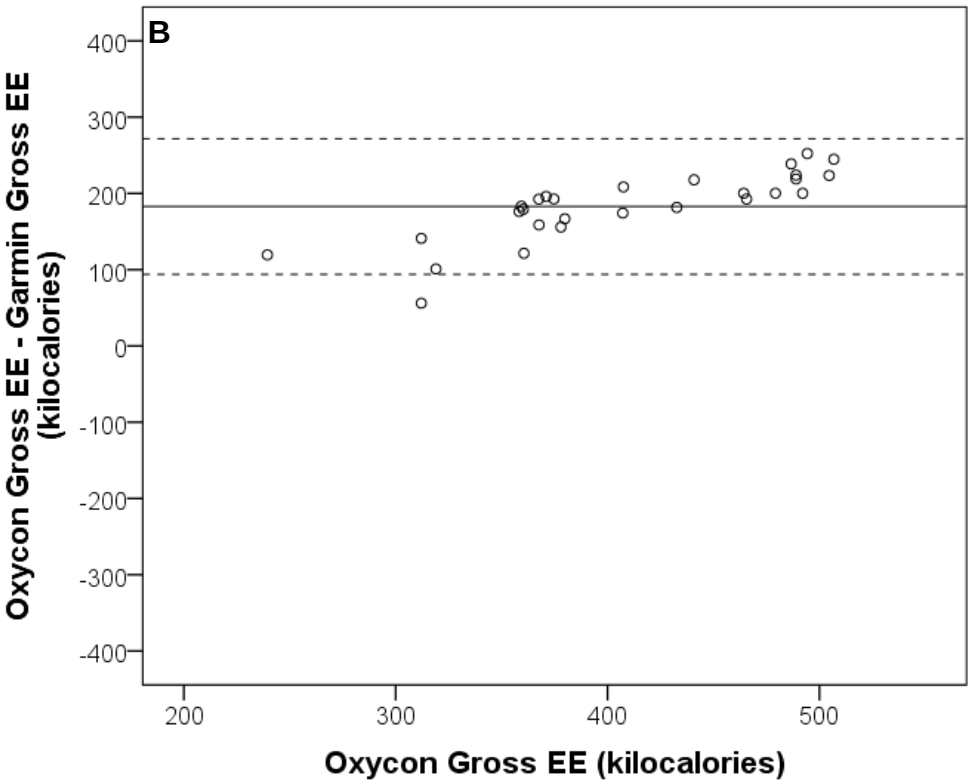
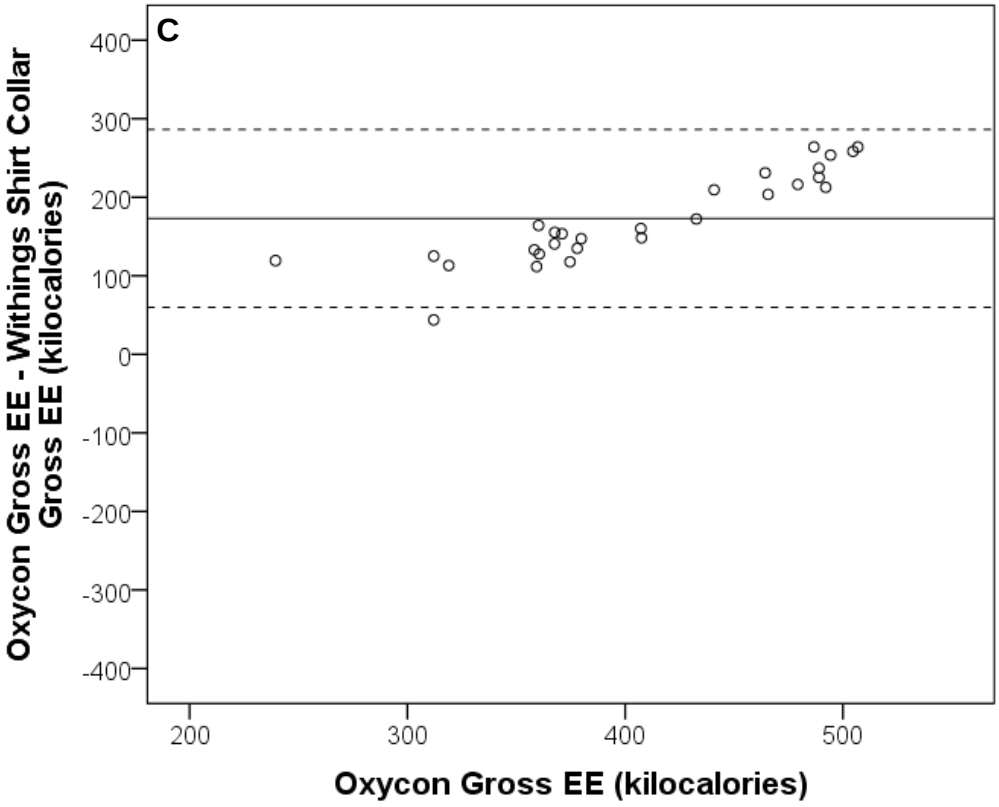


Figure 1 continued



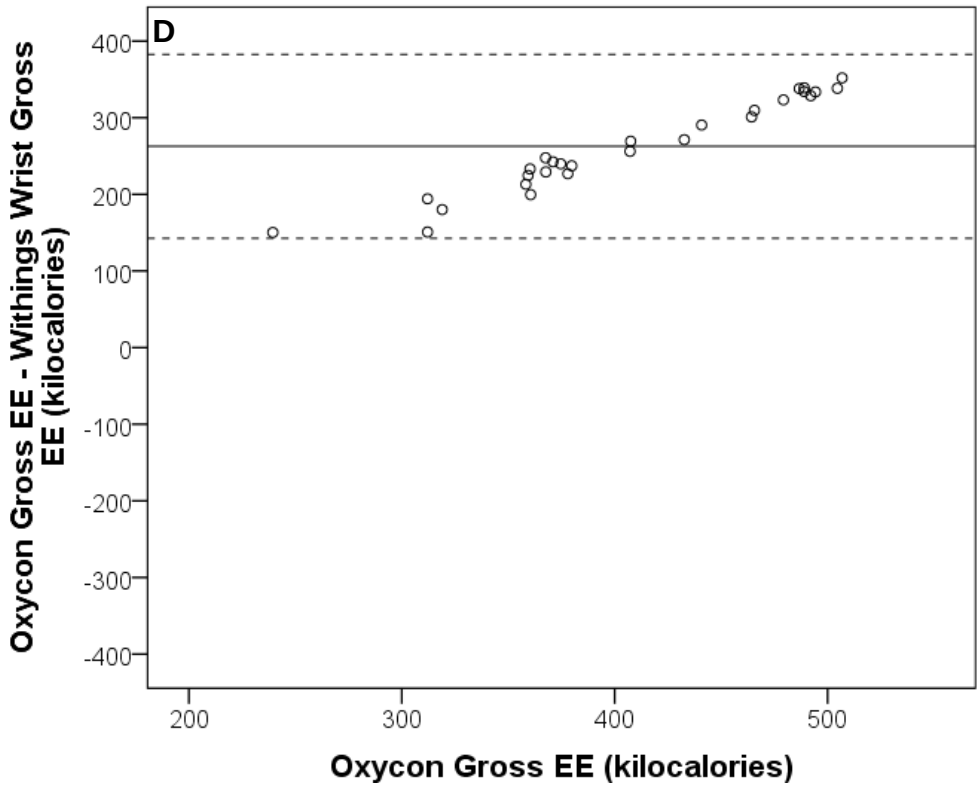


Figure 1 continued

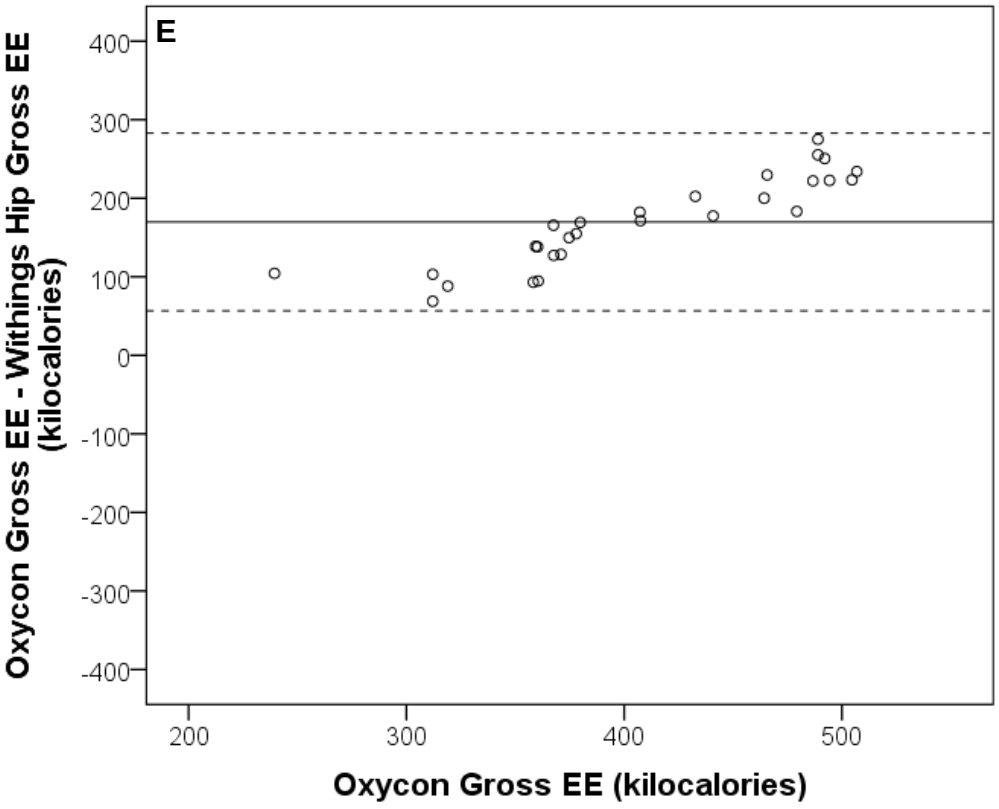


Figure 1 continued

Table 3. Mean ± SD measured and predicted gross energy expenditure (kcal/min) for each individual activity.

Activity	N	Oxycon	Basis Peak	Garmin VivoFit	Withings wrist	Withings shirt collar	Withings hip
Supine lying rest	27	1.5 ± 0.3	1.2 ± 0.3*	1.4 ± 0.3*	1.2 ± 0.2*	1.2 ± 0.2*	1.2 ± 0.2*
Seated computer use	27	1.8 ± 0.4	1.6 ± 0.6	1.7 ± 0.3	2.6 ± 0.4*	2.5 ± 0.3*	2.5 ± 0.3*
Seated folding towels	27	3.1 ± 0.5	2.7 ± 0.5*	2.0 ± 0.4*	2.7 ± 0.3*	2.5 ± 0.3*	2.5 ± 0.3*
Sweeping a floor	27	4.2 ± 0.8	2.8 ± 0.7*	2.5 ± 0.7*	2.8 ± 0.4*	2.5 ± 0.4*	2.6 ± 0.3*
Treadmill walking (80.5 m/min, 7% incline)	27	8.5 ± 1.6	10.1 ± 2.4*	4.2 ± 0.7*	3.4 ± 0.4*	5.6 ± 0.4*	5.7 ± 0.5*
Up and down stairs	27	9.1 ± 1.3	10.3 ± 2.7*	4.1 ± 0.8*	3.9 ± 0.3*	7.2 ± 0.8*	7.3 ± 0.8*
Over-ground walking (avg. speed 79 m/min)	27	5.6 ± 1.1	10.3 ± 2.30*	4.2 ± 0.8*	3.5 ± 0.5*	5.7 ± 0.8	5.7 ± 0.8
Over-ground running (avg. speed 150 m/min)	24	13.7 ± 3.2	14.3 ± 2.6	10.8 ± 2.4*	5.2 ± 0.9*	13.5 ± 2.1	13.6 ± 2.0
Seated rest	25	2.5 ± 0.7	2.6 ± 3.3	1.7 ± 0.3	2.4 ± 0.2	2.5 ± 0.5	2.4 ± 0.5
Over-ground cycling (avg. speed 207 m/min)	26	8.7 ± 2.6	5.4 ± 3.5*	3.2 ± 7.1*	1.2 ± 0.2*	2.8 ± 0.7*	2.8 ± 0.8*
Stationary cycling (100 watts)	25	9.2 ± 1.2	7.0 ± 5.1*	1.8 ± 0.6*	1.2 ± 0.2*	2.5 ± 0.4*	2.5 ± 0.4*

* Significantly different from the Oxycon, $P < 0.05$

activities, with mean differences ranging from 0.8 to 8.5 kcals/min (43.2%-73.1%) (P<0.05). The Withings shirt collar and hip placements significantly over- or under-estimated the same nine activities from 0.7 to 6.7 kcals/min (38%-73%). (P<0.05).

Figure 2 shows predicted mean gross EE for each Withings Pulse placement site for the entire activity routine. For the entire routine, the shirt collar and hip placements were both significantly higher than the wrist placement (P<0.001) but not significantly different from each other (P>0.05). The shirt collar and hip placements had fair to good reliability (ICC=0.558, P<0.05). Pairs containing the wrist placement had poor reliability; wrist and hip ICC=0.085 and wrist and collar ICC=0.094 (P<0.05). For seated computer use, seated rest, and stationary cycling there were no significant differences between the three Withings Pulse placement locations (P>0.05). For all other individual activities, the shirt collar and hip placements were both significantly different from the wrist placement (P<0.001) but not significantly different from each other (P>0.05).

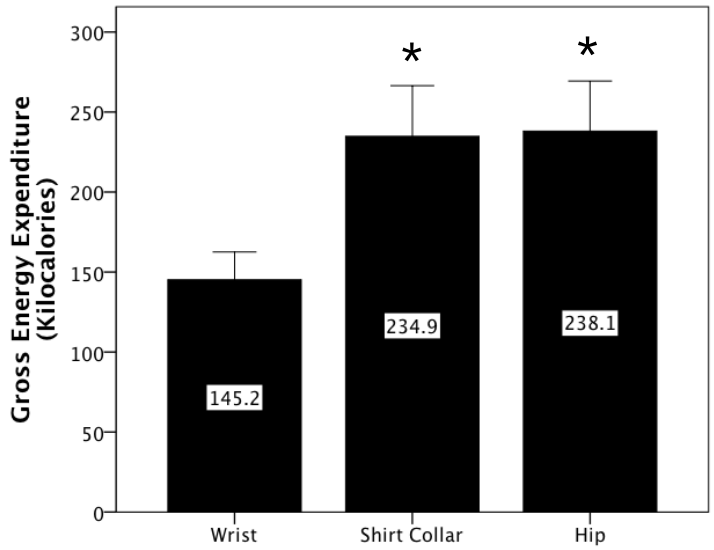


Figure 2. Withings gross energy expenditure for the entire physical activity routine. Error bars represent standard deviation. * Denotes significant different from wrist location (P<0.05).

Figure 3 shows the percent of minutes correctly classified by the Basis Peak during treadmill walking, over-ground walking, over-ground running, over-ground cycling, and stationary cycling. For treadmill walking, over-ground walking and over-ground running, ≥ 92% of minutes were correctly classified (P>0.05). For over-ground cycling only 40.4% of minutes were correctly classified (P<0.001). Compared to direct observation, zero stationary cycling minutes were correctly identified (P<0.001).

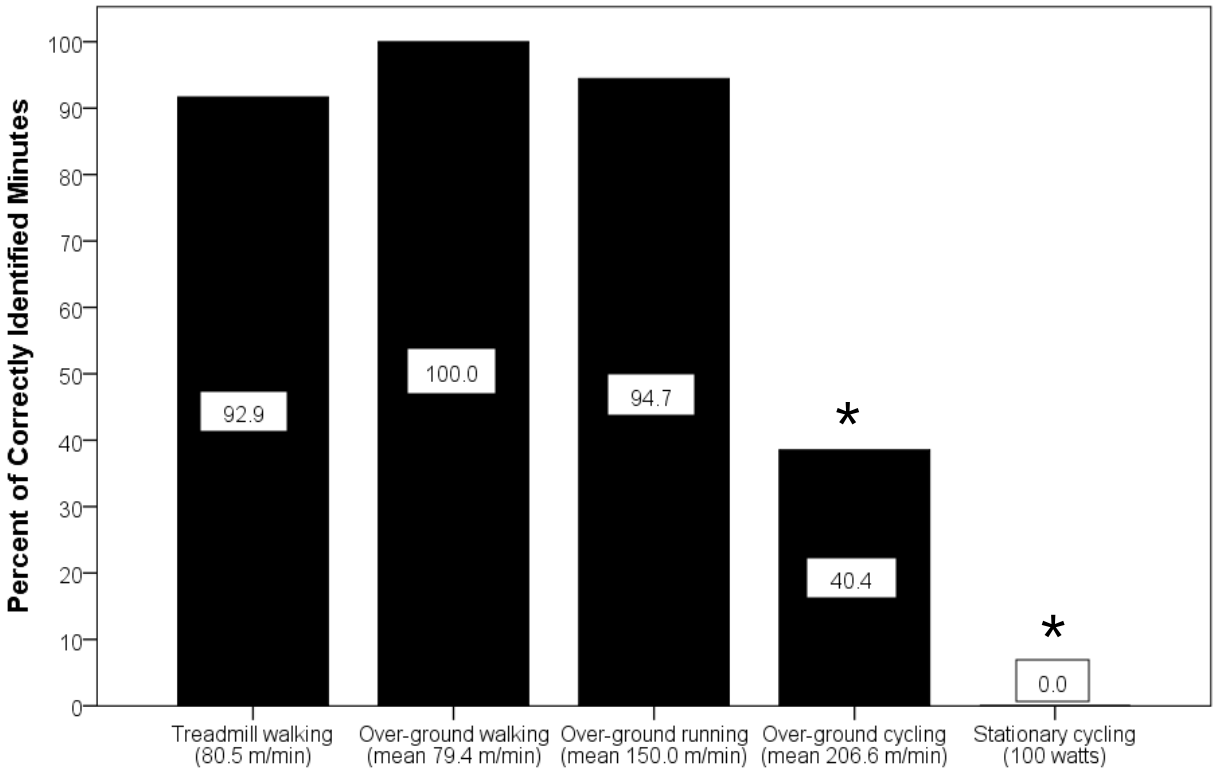


Figure 3. Basis Peak activity identification during structured bouts of walking, running, and cycling. ***Denotes significant difference from measured time.**

DISCUSSION

The primary findings from this study are that: 1) The Basis Peak was the only monitor not significantly different from measured EE for the entire physical activity routine; however it had the largest 95% PI. 2) All devices performed poorly for estimating gross EE of individual activities. 3) For the entire routine and eight of 11 individual activities EE predictions from the Withings Pulse wrist placement site were significantly different from the shirt collar and hip placement sites, while the shirt collar and hip placements were not different from each other. 4) The Basis Peak activity identification function correctly identified more than 90% of walking and running minutes, but could not accurately predict cycling, with zero stationary cycling minutes identified.

The second-generation Basis Peak predictions of gross EE were, on average, similar to measured EE over the entire structured PA routine. This is a significant improvement compared to the first-generation Basis B1 (31). Previous findings showed the Basis B1 underestimated measured EE during a 69 minute PA routine by 85.8 kcals (MAPE=24%) and performed the worst compared to seven other devices. The best performers were the Bodymedia Fit (MAPE=9.3%) Fitbit Zip, (MAPE=10.1%) and Fitbit One (MAPE=10.4%) (31). Although improvements were observed in the current study, the Basis Peak EE predictions were significantly different from the Oxycon during eight individual activities, indicating this device does not accurately predict EE of individual activities. For individuals that wish to measure EE for specific activity bouts, this device is likely to provide over- or under-estimates.

The current study found that all three Withings Pulse placements and the Garmin VivoFit performed poorly and significantly underestimated measured EE for the entire activity routine, and predictions were significantly different for at least eight individual activities. These results indicate users of the Withings Pulse and Garmin VivoFit could be receiving incorrect information about their daily EE. The Basis Peak did not significantly under- or over-estimate EE, however it had a wide range of individual error, so users of this device may also be receiving incorrect information. Accurate daily EE estimates are required for individuals seeking weight loss through a caloric intake deficit. Weight loss is a common goal of many consumers who purchase PA monitors, and those who purchase the Basis Peak, Withings Pulse or Garmin VivoFit for this goal are likely recieving inaccurate daily EE estimates. Consumers should be cautious of using these devices for the purpose of estimating daily EE.

The Withings Pulse device provided more consistent predictions between two of the three different placement sites (shirt collar, hip). The wrist placement was significantly different from either shirt collar or hip placements. If the same EE prediction algorithm is used for multiple placements of a single device on one individual, predictions will be different for each placement site. This effect was observed in the current study. The device literature did not explicitly state guidelines for placement; sites were chosen based on the device accessories (wristband, clip), as well as images from the manufacturer's instruction booklet and website. No input of placement site was required for use, so the Withings Pulse could not have used a different algorithm for wrist, shirt collar, and hip predictions.

In a prior study of concurrent validity, the Withings Pulse EE predictions demonstrated strong correlation with a research-grade PA monitor, (20). In that study, the validated Bodymedia SenseWear was used as criterion to investigate EE predictions from five consumer-based monitors. This study found moderate to strong correlations with the criterion (Withings Pulse, MisFit Shine, Jawbone UP, Fitbit Zip, Fitbit One; R=0.74-0.81. (20). A study of concurrent validity between consumer-based and research-grade devices is limited because the criterion provides an estimate and not a measure of EE, therefore true accuracy cannot be determined for the consumer devices. Results from the current study indicate the Withings Pulse cannot accurately predict EE during many individual physical activities, when compared to indirect calorimetry.

The Basis Peak accurately identified more than 92% of walking and running time, and could be used to help individuals estimate total weekly walking and running time. Such information helps individuals determine whether they meet PA guidelines. Accuracy of this function is limited to walking and running; the device cannot identify stationary cycling, and only 40% of over-ground cycling minutes were identified. This device could still be used to encourage individuals that enjoy walking and running to increase weekly activity.

Strengths of this study include the criterion measure of EE (Oxycon) and use of direct observation in comparison to the Basis Peak activity identification function. There were also some limitations to this study. There was some data loss, and participants were a homogenous group of highly fit college-aged adults that consisted of mostly males. Future studies should include free-living activity to improve the generalizability of

results, providing more information about how these devices are performing in daily use. Such information could help researchers choose the most accurate monitors for PA intervention studies.

In summary, findings suggest that the consumer-based monitors examined provided poor estimates for EE of individual activities and the entire physical activity routing. The Basis Peak was the only device to not be significantly different from measured EE during the entire routine; however like the other devices it had large individual error. Caution should be used with devices that suggest multiple placement sites but do not provide site specific prediction algorithms as with the Withings Pulse. Lastly, while the Basis Peak worked well for estimating time spent walking and running, caution should be used when predicting time spent cycling. Future research should examine a wider range of activities, as well as, free-living activity to further evaluate how the devices are actually used in a real-world setting.

LIST OF REFERENCES

1. Ahmed HM, Blaha MJ, Nasir K, Rivera JJ, Blumenthal RS. Effects of physical activity on cardiovascular disease. *Am J Cardiol*. 2012;109(2):288-95.

2. Albert MV, Deeny S, McCarthy C, Valentin J, Jayaraman A. Monitoring Daily Function in Persons With Transfemoral Amputations Using a Commercial Activity Monitor: A Feasibility Study. *PM&R*. 2014;6(12):1120-7.

3. American College of Sports Medicine. ACSM's guidelines for exercise testing and prescription. 9th ed: Lippincott Williams & Wilkins; 2013.

4. Archer E, Blair SN. Physical activity and the prevention of cardiovascular disease: from evolution to epidemiology. *Prog Cardiovasc Dis*. 2011;53(6):387-96.

5. Bassett Jr DR, Rowlands AV, Trost SG. Calibration and validation of wearable monitors. *Med Sci Sports Exerc*. 2012;44(1):S32.

6. Blair SN. Physical inactivity: the biggest public health problem of the 21st century. *Br J Sports Med*. 2009;43(1):1-2.

7. Brownson RC, Boehmer TK, Luke DA. Declining rates of physical activity in the United States: what are the contributors? *Annu Rev Public Health*. 2005;26:421-43.

8. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. *JAMA*. 2015;313(6):625-6.

9. Chen KY, Acra SA, Donahue CL, Sun M, Buchowski MS. Efficiency of walking and stepping: relationship to body fatness. *Obes Res*. 2004;12(6):982-9.

10. Chen KY, Janz KF, Zhu W, Brychta RJ. Re-defining the roles of sensors in objective physical activity monitoring. *Med Sci Sports Exerc*. 2012;44(1):S13.

11. Chen KY, Sun M. Improving energy expenditure estimation by using a triaxial accelerometer. *J Appl Physiol*. 1997;83(6):2112-22.

12. Crespo NC, Sallis JF, Conway TL, Saelens BE, Frank LD. Worksite physical activity policies and environments in relation to employee physical activity. *Am J Health Promot*. 2011;25(4):264-71.

13. Crouter SE, Churilla JR, Bassett DR, Jr. Estimating energy expenditure using accelerometers. *Eur J Appl Physiol*. 2006;98(6):601-12.

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

15. Crouter SE, DellaValle DM, Haas JD, Frongillo EA, Bassett DR. Validity of actiGraph 2-regression model and Matthews and NHANES and cut-points for assessing free-Living physical activity. *J Phys Act Health*. 2013;10(4):504.

16. Crouter SE, Kuffel E, Haas JD, Frongillo EA, Bassett Jr DR. A refined 2-regression model for the ActiGraph accelerometer. *Med Sci Sports Exerc*. 2010;42(5):1029.

17. Dannecker KL, Sazonova NA, Melanson EL, Sazonov ES, Browning RC. A comparison of energy expenditure estimation of several physical activity monitors. *Med Sci Sports Exerc*. 2013;45(11):2105-12.

18. Dishman R, Heath G, Lee I-M. Physical Activity Epidemiology: Human Kinetics; 2004.

19. Duffield R, Dawson B, Pinnington H, Wong P. Accuracy and reliability of a Cosmed K4b^2 portable gas analysis system. *J Sci Med Sport*. 2004;7(1):11-22.

20. Esliger DW, Rowlands AV, Hurst TL, Catt M, Murray P, Eston RG. Validation of the GENEA Accelerometer. *Med Sci Sports Exerc*. 2011;43(6):1085-93.

21. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *Int J Behav Nutr Phys Act*. 2015;12(1):42.

22. Ferrannini E. The theoretical bases of indirect calorimetry: a review. *Metabolism*. 1988;37(3):287-301.

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

24. Fulk GD, Combs SA, Danks KA, Nirider CD, Raja B, Reisman DS. Accuracy of 2 activity monitors in detecting steps in people with stroke and traumatic brain injury. *Phys Ther*. 2014;94(2):222-9.

25. Hansen BH, Kolle E, Dyrstad SM, Holme I, Anderssen SA. Accelerometer-determined physical activity in adults and older people. *Med Sci Sports Exerc*. 2012;44(2):266-72.

26. Harris JA, Benedict FG. A biometric study of human basal metabolism. *Proc Natl Acad Sci U S A*. 1918;4(12):370.

27. Hendelman D, Miller K, Baggett C, Debold E, Freedson P. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Med Sci Sports Exerc*. 2000;32(9):S442-9.

28. Hooker SP, Feeney A, Hutto B, Pfeiffer KA, McIver K, Heil DP, et al. Validation of the actical activity monitor in middle-aged and older adults. *J Phys Act Health*. 2011;8(3):372-81.

29. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson PS. Validation of wearable monitors for assessing sedentary behavior. *Med Sci Sports Exerc*. 2011;43(8):1561-7.

30. Kramer CL. Accuracy of Commercially-Available Accelerometers For Measuring Steps: Arizona State University; 2014.

31. Lee J-M, Kim Y-W, Welk GJ. TRACK IT: Validity and Utility of Consumer-Based Physical Activity Monitors. *ACSMs Health Fit J*. 2014;18(4):16-21.

32. Lee J-M, Kim Y, Welk GJ. Validity of Consumer-Based Physical Activity Monitors. *Med Sci Sports Exerc*. 2014;46(9):1840-8.

33. Lee PH, Macfarlane DJ, Lam T, Stewart SM. Validity of the international physical activity questionnaire short form (IPAQ-SF): A systematic. *Int J Behav Nutr Phys Act*. 2011;8(115):1-11.

34. Levine JA. Measurement of energy expenditure. *Public Health Nutr*. 2005;8(7a):1123-32.

35. Lyden K, Kozey SL, Staudenmeyer JW, Freedson PS. A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. *Eur J Appl Physiol*. 2011;111(2):187-201.

36. Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR, et al. Amount of time spent in sedentary behaviors in the United States, 2003–2004. *Am J Epidemiol*. 2008;167(7):875-81.

37. McLaughlin JE, King GA, Howley ET, Bassett DR, Jr., Ainsworth BE. Validation of the COSMED K4 b2 portable metabolic system. *Int J Sports Med*. 2001;22(4):280-4.

38. Metkus TS, Jr., Baughman KL, Thompson PD. Exercise prescription and primary prevention of cardiovascular disease. *Circulation*. 2010;121(23):2601-4.

39. Meyer J, Hein A. Live Long and Prosper: Potentials of Low-Cost Consumer Devices for the Prevention of Cardiovascular Diseases. *Medicine 20*. 2013;2(2):e7.

40. Nichols JF, Morgan CG, Chabot LE, Sallis JF, Calfas KJ. Assessment of physical activity with the Computer Science and Applications, Inc., accelerometer: laboratory versus field validation. *Res Q Exerc Sport*. 2000;71(1):36-43.

41. Noah JA, Spierer DK, Gu J, Bronner S. Comparison of steps and energy expenditure assessment in adults of Fitbit Tracker and Ultra to the Actical and indirect calorimetry. *J Med Eng Technol*. 2013;37(7):456-62.

42. Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999-2004. *JAMA*. 2006;295(13):1549-55.

43. Pate RR, Pratt M, Blair SN, Haskell WL, Macera CA, Bouchard C, et al. Physical activity and public health: a recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *JAMA*. 1995;273(5):402-7.

44. Perret C, Mueller G. Validation of a new portable ergospirometric device (Oxycon Mobile) during exercise. *Int J Sports Med*. 2006;27(5):363-7.

45. Plasqui G, Bonomi A, Westerterp K. Daily physical activity assessment with accelerometers: new insights and validation studies. *Obes Rev*. 2013;14(6):451-62.

46. 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.

47. Rosner B. Fundamentals of biostatistics: Cengage Learning; 2010.

48. Rothney MP, Brychta RJ, Meade NN, Chen KY, Buchowski MS. Validation of the ActiGraph two-regression model for predicting energy expenditure. *Med Sci Sports Exerc*. 2010;42(9):1785-92.

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

50. Santos-Lozano A, Santín-Medeiros F, Cardon G, Torres-Luque G, Bailón R, Bergmeir C, et al. Actigraph GT3X: Validation and determination of physical activity intensity cut points. *Int J Sports Med*. 2013;34(11):975-82.

51. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. *J Sci Med Sport*. 2011;14(5):411-6.

52. Schoeller DA. Measurement of energy expenditure in free-living humans by using doubly labeled water. *J Nutr*. 1988;118(11):1278-89.

53. Sesso HD, Paffenbarger RS, Lee I-M. Physical activity and coronary heart disease in men the Harvard Alumni Health Study. *Circulation*. 2000;102(9):975-80.

54. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. *Br J Sports Med*. 2003;37(3):197-206.

55. Sieverdes JC, Wickel EE, Hand GA, Bergamin M, Moran RR, Blair SN. Reliability and validity of the Mywellness Key physical activity monitor. *Clin Epidemiol*. 2013;5:13.

56. Stahl ST, Insana SP. Caloric expenditure assessment among older adults: Criterion validity of a novel accelerometry device. *J Health Psychol*. 2013;19(11):1382-7.

57. Staudenmayer J, Pober D, Crouter S, Bassett D, Freedson P. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *J Appl Physiol*. 2009;107(4):1300-7.

58. Storm FA, Heller BW, Mazzà C. Step Detection and Activity Recognition Accuracy of Seven Physical Activity Monitors. *PloS One*. 2015;10(3):e0118723.

59. Strath S, Bassett Jr D, Swartz A. Comparison of MTI accelerometer cut-points for predicting time spent in physical activity. *Int J Sports Med*. 2003;24(4):298-303.

60. Strath SJ, Kaminsky LA, Ainsworth BE, Ekelund U, Freedson PS, Gary RA, et al. Guide to the assessment of physical activity: Clinical and research applications: a scientific statement from the American Heart Association. *Circulation*. 2013;128(20):2259-79.

61. Swartz AM, Strath SJ, Bassett DR, O Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Med Sci Sports Exerc*. 2000;32(9):S450-S6.

62. Takacs J, Pollock CL, Guenther JR, Bahar M, Napier C, Hunt MA. Validation of the Fitbit One activity monitor device during treadmill walking. *J Sci Med Sport*. 2013;17(5):496-500.

63. Thorndike AN, Mills S, Sonnenberg L, Palakshappa D, Gao T, Pau CT, et al. Activity monitor intervention to promote physical activity of physicians-in-training: randomized controlled trial. *PloS One*. 2014;9(6):e100251.

64. 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.

65. Welk GJ, Blair SN, Wood K, Jones S, Thompson RW. A comparative evaluation of three accelerometry-based physical activity monitors. *Med Sci Sports Exerc*. 2000;32(9):S489-S97.

APPENDIX

INFORMED CONSENT FORM

Study Title: Validation of energy expenditure predictions from consumer-based activity monitors.

Principal Investigators: Andrew Woodman and Scott E. Crouter, Ph.D.

Institution: The University of Tennessee, Knoxville

This information is provided to tell you about the research project. Please read this form carefully and ask any questions you may have about this study. Your questions will be answered before we ask you to sign it. Also, you will be given a copy of this consent form to take home.

INTRODUCTION
The purpose of this study is to assess the validity and accuracy of various commercially available physical activity monitors.

ELIGIBILITY
To be in this study, you must be between 18 and 65 years of age, complete a Physical Activity Readiness Questionnaire (PAR-Q), and have no contraindications to moderate and vigorous physical activity.

GENERAL TESTING SCHEDULE
1. You will be given an informed consent form, the study will be described, and you will have time to ask questions. If you decide to continue, we require your initials, signature, and date on an informed consent.

2. You will be asked to complete a Physical Activity Readiness Questionnaire (PAR-Q). This will assess your readiness for physical activity.

3. We will measure and record on a data sheet your age, height, and weight.

4. If you are female you will be asked to verbally confirm that you are not pregnant.

5. If you have contraindication to exercise or do not meet eligibility requirements for BMI (based on height and weight), pregnancy status, age, or running ability, participation in this study will stop here.

6. A researcher will help you put on a portable metabolic system (a backpack and facemask)

7. You will put on two wrist-worn devices on one arm, another wrist-worn device on the other arm, one hip worn device attached to your waist, and another device clipped to the collar of your shirt.

8. You will be asked to complete ten minutes of rest, lying on your back.

9. You will be asked to complete a structured activity routine, 5 minutes each activity, with a 1-3 minutes rest in between. A researcher will be recording notes during this time.

- quiet study or computer usage in a seated position
- folding clothes at a table in a seated position
- sweeping a floor
- walking on a treadmill at three miles per hour with a 7% grade
- continuously ascending and descending flights of stairs
- walking at a self-selected pace in the HPER gym or on the track
- running at a self-selected pace in the HPER gym or on the track
- five minute rest period
- cycling overground on a standard bicycle, at a self-selected pace
- cycling on a Monark ergometer at 60 revolutions per minute with 2 kiloponds of resistance

_____ Participant's Initials

RISKS
The American College of Sports Medicine (ACSM) states that absolute contraindications to exercise testing include, but are not limited to, recent (within 2 days) history of acute cardiac events, arrhythmias that are not controlled by medication, symptoms of heart failure, or presence of acute infection (28). There is a very small risk of heart attack in healthy individuals performing moderate intensity activity, though risk increases in those with cardiovascular disease performing vigorous intensity activity (28). Possible risk of moderate and vigorous intensity physical activity include musculoskeletal injury, headaches, dizziness, abnormally high blood pressure, and a relative risk of cardiac events (28, 29).

BENEFITS
Participation in this study will help assess the validity and accuracy of consumer-based physical activity monitors, and contribute to future research in the field of physical activity assessment.

CONFIDENTIALITY
The information obtained from this study will be treated as confidential. Confidentiality will be maintained in the analysis and presentation of the data through the use of an ID number that we will assign to you. Your name and ID number will be recorded at the beginning of the study and this information will be placed in a file cabinet that will be locked and only accessible to study investigators.

COMPENSATION
There will be no compensation for participating in this research study.

EMERGENCY MEDICAL TREATMENT
The University of Tennessee does not "automatically" reimburse participants for medical claims or other compensation. If physical injury is suffered in the course of research, or for more information, please notify the student researcher Andrew Woodman at (815) 621-1730, or the faculty supervisor of this study Dr. Scott E. Crouter at (865) 974-1272.

CONTACT INFORMATION
If you have questions at any time about the study or the procedures, (or you experience adverse effects as a result of participating in this study,) you may contact the researcher Andrew Woodman by phone at (815) 621-1730 and by email at jwoodma1@vols.utk.edu or the faculty supervisor of the project, Dr. Scott E. Crouter, at 1914 Andy Holt Ave., 334 HPER Bldg., Knoxville, TN, (865) 974-1272. If you have questions about your rights as a participant, contact the Office of Research Compliance Officer at (865) 974-3466.

PARTICIPATION
Your participation in this study is voluntary; you may decline to participate. If you decide to participate, you may withdraw from the study at any time without penalty and without loss of benefits to which you would otherwise be entitled. If you withdraw from the study before data collection is completed, your data will be returned to you or destroyed.

CONSENT
I have read the above information, and I have received a copy of this form. I agree to participate in this study.

Participant's signature: _____ Date: _____

Investigator's signature: _____ Date: _____

Physical Activity Readiness Questionnaire (PARQ)

Regular physical activity is fun and healthy, and increasingly more people are starting to become more active every day. Being more active is very safe for most people. However, people should check with their doctor before they start becoming much more physically active.

If you are planning to become much more physically active than you are now, start by answering the seven questions in the box below. If you are between the ages of 15 and 69, the PAR-Q will tell you if you should check with your doctor before you start. If you are over 69 years of age, and you are not used to being very active, check with your doctor.

Common sense is your best guide when you answer these questions. Please read the questions carefully and answer each one honestly: check YES or NO.

YES NO 1. Has your doctor ever said that you have a heart condition and that you should only do physical activity recommended by a doctor?

YES NO 2. Do you feel pain in your chest when you do physical activity?

YES NO 3. In the past month, have you had chest pain when you were not doing physical activity?

YES NO 4. Do you lose your balance because of dizziness or do you ever lose consciousness?

YES NO 5. Do you have a bone or joint problem that could be made worse by a change in your physical activity?

YES NO 6. Do you know of any other reason why you should not be doing physical activity?

I have read, understood and completed this questionnaire. Any questions I had were answered to my full satisfaction.

Name (Print): _____

Signature: _____ Date: _____

VITA

James Andrew Woodman was born December 3rd, 1988 to mother Mari K. Woodman, JD and father James L. Woodman, JD, MD. Andrew graduated with honors from Northern Illinois University in the summer of 2012, earning his Bachelor's of Science Degree. He majored in Kinesiology, with a concentration in Exercise Science and a minor in Fine Arts. He continued work with the YMCA of Rock River Valley for one year after earning his undergraduate degree, and then continued on to graduate school in the Fall of 2013. Andrew graduated from the University of Tennessee, Knoxville in the Summer of 2015, earning his Master's Degree in Kinesiology with a specialization in Exercise Physiology. He was accepted to the University of Vermont post-baccalaureate pre-medicine program, and will begin his curriculum at UVM in the Fall of 2015. Andrew will prepare for a career as a primary care physician, becoming an advocate for the philosophy of "Exercise is Medicine".