ESTABLISHING THE RELIABILITY OF SEVERAL CONSUMER-BASED PHYSICAL ACTIVITY MONITORS

A THESIS

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF SCIENCE

BY

JOSHUA MICHAEL BOCK

DR. ALEXANDER MONTOYE – ADVISOR

BALL STATE UNIVERSITY

MUNCIE, INDIANA

JULY 2016

Acknowledgements

First, I would like to thank my family for their continued support. Mom, Dad, Joe, Dawn, Nick, and Lyla – you guys have made these past six years so much easier on me; thank you for always being there for me.

Dr. Montoye, working with you these past two years has been an incredible learning experience. You've taught me a skill set that I can carry for the rest of my life; saying I've learned a lot from you is a strong understatement ($p < 0.01$). Thank you for everything!

Drs. Matt Harber, Leonard Kaminsky, Todd Trappe, and Scott Trappe; thank you for all you've taught me inside and outside the classroom. The knowledge and lessons I've gained from you all while in the HPL are beyond valuable.

Dr. Mazzetti, I wouldn't be writing this if it weren't for you. Thank you for introducing me to research as an undergraduate, your help in getting me to Ball State, and for all of your continued support as I transition to the next level.

Jordan, Beard, Hidde, Kelsie, Zoe: where did the past two years go? It seems not too long ago we were eating Jimmy John's during orientation week; couldn't think of a better group to go through this program with. Thanks for all the great times and lifelong memories! We did it!

Ben, Kristine, Kristen, Amanda, Dan, and Megan: thank you for your mentorship during the first year of the program and showing me how to have a good time in Muncie, Indiana. I'm incredibly grateful to have spent time with you all.

Rachel, Allison, Nagle, Alexis, Chad, and Izzy: you guys rock! Had a lot of great times during our year together; best of luck with the rest of your time in Muncie and careers!

Steve, Greg, Tuttle, Kevin(s), Ryan, Kaleen, Tim, Voss, Jen, and Brandon: thanks for all of your help with essays, applications, beverage selection, and for always being down to talk science. I wouldn't be nearly as prepared for a doc program if it weren't for you guys. Best of luck with your career endeavors!

Abby, it's blown me away how much of a support system you have become in such a short amount of time. I'm so lucky to have met you at this point in my life. Thanks for believing in me, tolerating my rants about science, putting up with the ridiculous things I say/do, and for keeping me on track when I'm the furthest thing from focused. Here's to Iowa and the adventures that will follow!

Table of Contents

List of Figures

List of Tables

Chapter I

Introduction

Background.

Regular physical activity (PA) has long been known to have positive effects on health-related variables (Morris 1954). Defined as "any bodily movement produced by the contraction of skeletal muscles that results in a substantial increase in caloric requirements over resting energy expenditure," PA has been shown to yield favorable health benefits such as lowering body weight, blood pressure, cholesterol, and reducing the risk of cardiovascular or coronary artery disease (CVD and CAD, respectively) (Garber 2011, PAGAC 2008). Obesity, diabetes, and CVD cost the U.S. almost $800 billion each year; regularly participating in PA improves these chronic conditions and has the potential to reduce their associated costs. Despite the known benefits of regularly engaging in PA, 3-51% of adults living in the U.S. do not meet the 2008 Physical Activity Guidelines for Americans; depending upon the instrument used for PA measurement (Troiano 2008, CDC 2010).

To better understand the role of PA in reducing the societal burden of these diseases, it is important to be able to measure PA accurately and reliably. Wareham et al. describes five reasons for improving PA measurement techniques. These reasons include being able to identify which activity intensity provides the best health benefits, quantify the dose-response relationship of PA in studies, examine PA without cultural bias, monitor intermittent bouts of PA, and more accurately measure the efficacy of interventions (Wareham 1998).

Subjective and objective methods are available to quantify PA. The most common subjective tool used for estimating PA is a questionnaire (e.g. the International Physical Activity Questionnaire) (Albanes 1990). Subjective PA measurements are easy and inexpensive to use but often yield overestimations of PA due to recall bias. As such, questionnaires are used predominately when objective measuring instruments are impractical or when other aspects of PA (e.g. recalling past PA or location in which PA takes place) are of interest (Ferguson 2015). Objective measurements of PA are commonly made using a pedometer or accelerometer (Tudor-Locke 2014). A pedometer is a step-counting device commonly worn on the hip. Accelerometers are more advanced than pedometers due to their ability to measure and record time-stamped accelerations as a person moves. Their ability to measure the magnitude of motions and time stamp data allow accelerometers to estimate PA intensity in addition to steps and energy expenditure. Most research-grade accelerometers express their collected data in units known as activity counts. Once counts are assigned to a period of time (called an epoch), cut-points are used to classify that epoch's activity intensity. Cut-points are pre-determined thresholds used to classify a series of counts into sedentary, light, moderate, vigorous, or very vigorous PA. There are many series of cut-points in the literature; most notably, the series published by Freedson et al. in 1998 for use with the research-grade ActiGraph GT3X+ accelerometer (Freedson 1998). Despite their accuracy and precision, research-grade accelerometers are rarely used by the common consumer due to their complex units of measurement, lack of real-time feedback, high cost, and dependence on software for data analysis.

Within the last decade, companies such as Fitbit and Jawbone have begun producing consumer-based PA monitors (CPAMs) that use technology similar to accelerometers but are offered at a lower price. CPAMs also produce PA estimations in units that are commonly understood such as steps, Calories (kcals), and active minutes. CPAMs have become increasingly popular in recent years; over 3.3 million were sold in 2013 alone with 87% coming from Fitbit and Jawbone (Danova 2014). With the overwhelming presence these devices have in today's market, it is important to understand their accuracy and reliability.

Despite the popularity of CPAMs, limited research exists regarding their validity and even less research exists to support their reliability to estimate steps, kcals, and active minutes. Knowing the reliability of CPAMs would allow consumers to have a better understanding of how their PA habits compare to others and how their PA changes over time. There are two variations of reliability known as intra and inter-monitor reliability. Intra-monitor reliability (also known as test-retest reliability) is the analysis of data collected from a single monitor, assessed against itself, when performing an identical activity multiple times. Inter-monitor reliability is a comparison of two identical monitors against each other during a single activity. Both forms of reliability are important to measure in order to best understand the performance of CPAMs. Consumers, who would likely purchase only one CPAM for PA tracking, would be interested in the intra-monitor reliability of CPAMs to be confident in PA estimations on a day-to-day basis. Establishing the inter-monitor reliability provides benefits to both researchers and consumers. Researchers using CPAMs in their studies need to know if using multiple models of the same monitor will provide consistent results; if not, the results of their studies could be invalid. Additionally, CPAM users often compete against one another to see who can accumulate the most PA; knowing the inter-monitor reliability will ensure that consumers with the same model of CPAM have a fair comparison of PA levels.

There are only a few previous studies assessing CPAM reliability. Kooiman et al. assessed the test-retest reliability (intra-monitor reliability) of the Fitbit Flex (FF), Jawbone Up (JU), and Fitbit Zip (FZ) to estimate steps using a laboratory-based protocol. The FF and JU were worn on the non-dominant arm while the FZ was carried in a front pants pocket. During a structured laboratory protocol, participants walked on a treadmill for 30 minutes at approximately 3.0 mph. Intra-class correlations for the FF, JU, and FZ in the laboratory setting were 0.81, 0.83, and 0.90, respectively (Kooiman 2015). A limitation to Kooiman's study includes only using a single speed on the treadmill during the study.

Diaz et al. examined both the validity and inter-monitor reliability of the Fitbit One (FO) and FF monitors. Three FOs were worn (two on the right hip, one on the left) and two FFs were used; one on each wrist. During the study, participants (n=23) walked/jogged on a treadmill at various speeds (1.9, 3.0, 4.0, and 5.2 mph) for 6 minutes. Results from this study showed inter-monitor correlations of 0.96 and 0.99 for FO estimations of steps and kcals, respectively. Diaz also showed correlations of 0.90 and 0.95 for steps and kcals, respectively, between the left and right wrist FF monitors (Diaz 2015). Limitations of Diaz's study include only using a treadmill for activities and no investigation of intra-monitor reliability.

The studies by Kooiman and Diaz provide some evidence that CPAMs have high inter- and intra-monitor reliability; however, neither study assessed both intra- and inter-monitor reliability, nor did either study examine the reliability of CPAMs for non-ambulatory activities such as sedentary or household activities.

Purpose

The purpose of this study was to determine the intra- and inter-monitor reliability of several popular CPAMS for estimations of steps, kcals, and active minutes in a structured laboratory setting as well as inter-monitor reliability steps, kcals, and active minutes in a free-living setting.

Aims and Hypotheses.

Aim 1. To establish the intra-monitor reliability of the Fitbit One (FO), Fitbit Zip (FZ), Fitbit Flex (FF), and Jawbone Up24 (JU) when estimating steps, kcals, and active minutes during sedentary, household, and ambulatory activities within a laboratory setting.

Hypothesis 1a. All CPAMs will have higher intra-monitor reliability (ICC > 0.8) for steps, kcals, and active minutes (11).

Hypothesis 1b. Hip-mounted CPAMs will be significantly more reliable than wrist-mounted CPAMs for steps, kcals, and active minutes.

Hypothesis 1c. The FF will be significantly more reliable than the JU for steps, kcals, and active minutes.

Aim 2. To assess the inter-monitor reliability of the FO, FZ, FF, and JU to estimate steps, kcals, and active minutes during sedentary, household, and ambulatory activities within a laboratory setting.

Hypothesis 2a. All CPAMs will have high inter-monitor reliability ($r > 0.8$) for steps, kcals, and active minutes (11).

Hypothesis 2b. The FF will be significantly more reliable than JU for steps, kcals, and active minutes.

Aim 3. To assess the inter-monitor reliability of the FO, FZ, FF, and JU to estimate steps, kcals, and active minutes in a free-living setting.

Hypothesis 3a. All CPAMs will be significantly less reliable for steps, kcals, and active minutes in the free-living setting compared to the structured laboratory setting.

Delimitations

Adult participants were recruited from the East Central Indiana region and were able to perform all activities within the protocol. Participants were excluded on the basis of illness, unstable chronic conditions, and pregnancy due to an increased risk of adverse responses to exercise. Participants completed a structured protocol during their first visit to the Human Performance Laboratory (HPL); the protocol was identical during all visits for all subjects. By standardizing the protocol amongst all participants, we were able to compare CPAMs against each other for specific activities.

While it is not possible to perform all activities an individual may complete during daily living within a 120 minute period, the protocol used in this study included 7 sedentary, household, and ambulatory activities. Following the initial visit to the laboratory, participants were assigned a set of wrist- or hip-mounted CPAMs to wear in their everyday environment for the remainder of the day and the following morning in order to assess the inter-monitor reliability in a free-living setting.

Definitions.

Kcals – An abbreviation for kilocalorie; a unit used to express the metabolic demand of an activity

Inter-Monitor Reliability – The degree of agreement between different units of the same monitor assessing the same diagnostic activity

Intra-Monitor Reliability – The degree of agreement among repeated administrations of a diagnostic activity by a single monitor

Manipulation check – Using a criterion measure to assess inadvertent changes of a variable during the replication of an activity

Metabolic Equivalent (MET) – A unit used to represent the average individual's rate of oxygen consumption at rest (3.5 ml of oxygen per kilogram of body weight per minute); multiples of this numeric are used to express activity intensity (ex. 2.0 METs)

Physical Activity – Any bodily movement resulting in an increased metabolic rate from rest

Portable Metabolic Analyzer – A breath-by-breath gas analyzer that uses oxygen consumption and carbon dioxide production to determine the metabolic demand of an activity

Reliability – The ability to produce an estimated value with equal variance over a series of trials

Step – The picking up of the heel and toe of a foot and replacing it on the ground

Chapter II

Review of the Literature

Introduction

Understanding the practical implications of the inter- and intra-monitor reliability of consumer-based physical activity monitors (CPAMs) begins with knowing the comprehensive benefits of physical activity (PA) itself. The tools used to quantify PA are diverse in complexity, validity, and practicality each with their own unique sets of positive and negative characteristics. The following sections will review the health benefits of PA, the methods used to quantify PA, the strengths and limitations of each instrument, as well as a review of their validity and reliability.

Part 1: Physical Activity

Regularly engaging in PA has been shown to yield favorable short and long-term health benefits since the 1950s (Morris 1954). Some of these benefits include lower body weight, blood pressure, LDL-cholesterol, and a reduced risk of cardiovascular or coronary artery disease (CVD and CAD, respectively) (Garber 2011, PAGA 2008). Blair et al. in 1996 conducted a meta-analysis of the literature present and found an inverse, dose-response relationship between PA and mortality due to coronary artery disease. Despite the extensive review of the literature, Blair et al. failed to identify the volume of PA required to minimize risk of mortality (Blair 1996). Regular PA has also been shown to increase cardiorespiratory fitness which has been shown to be the greatest predictor of all-cause mortality (Meyers 2002). Organizations such as the American College of Sports Medicine, Center for Disease Control, and the World Health Organization have detailed the positive impact PA can have on health outcomes, the systemic damage sedentary time can cause, and strategies individuals can use to incorporate PA into their daily regimen. The minimal amount of PA required to obtain these health benefits has been quantified using three variables: steps, kcals, and moderate to vigorous physical activity (MVPA).

10,000 steps

Butcher et al. conducted a study to examine the minimal amount of steps which needed to be taken in a day in order to obtain minimal benefits in health related outcomes (Butcher 2008). They found that taking 10,000 steps per day helped lower LDLs and improve coronary vascular health. Taking 10,000 steps per day has also been shown to stimulate weight loss in overweight and sedentary adults (Schneider 2006). A study conducted by Tudor-Locke et al. in 2011 confirmed these findings and produced a classification system for PA based upon the amount of steps taken in a day. Nomenclature for this system is as follows: a day can be classified as sedentary if <5,000 steps are taken; low-active if 5,001-7,499 steps are taken; somewhat active if 7,500-9,999 steps are taken; active if 10,000-12,455 steps are taken; and very active if >12,500 steps are taken in a given day (Tudor-Locke 2011). While steps per day is an unrefined measure of PA, these studies show that taking 10,000 steps per day promotes cardiovascular health and in unhealthy, sedentary populations it can stimulate modest amounts of weight loss (Cocate 2013).

1,000 kcals

Energy expenditure, in the form of kcals, is a popular variable used to quantify PA. Sesso et al. conducted a study investigating the relationship of kcals expended and the risk of developing coronary heart disease. They found that middle-aged, older men who expended ≥4,200 kJ/week (1,000 kcals) were less likely to develop coronary heart disease than those who expended <4,200 kJ/week. Their results also showed no significant decrease in relative risk when energy expenditure was increased beyond 4,200 kJ/week (Sesso 2000). Jakicic et al. published a report on the appropriate caloric balance for weight loss. They found that a caloric expenditure of 1,000 kcals/day would elicit a weight loss of 1-2 pounds per week (Jakicic 2001). These findings seem to conflict topically; however, the aims of each paper differ slightly and consumers should adjust their energy expenditure goals to fit their personalized health objectives.

<u>150/75 minutes</u>

Shiroma et al. investigated the impact of different proportions of vigorous PA in a regimen and its impact on all-cause and cardiovascular mortality. This study found that 150 minutes of MVPA and 75 minutes of vigorous PA had similar improvements on mortality outcomes (Shiroma 2014). Nybo et al. found that higher intensities of PA can improve some health parameters beyond lower intensity training; specifically maximal oxygen uptake (VO_{2max}) which has been shown to reduce all-cause mortality by 11% per MET increase (Meyers 2002, Nybo 2010). The 2008 Physical Activity Guidelines for Americans suggests that the average American should partake in 150 minutes of MVPA, 75 minutes of vigorous PA, or a combination of both each week. These recommendations incorporate the duration and intensity components present in the literature (PAGAC 2008). In doing so, the 2008 PAGA allows everyday activities (e.g. walking) to be counted towards meeting guidelines. These Guidelines also highlight a quality component of PA by differentiating moderate and vigorous PA.

Part 2: Physical Activity Assessment

Wareham et al. discusses the significance of PA assessment using subjective and objective instruments concluding that there is an inverse relationship between practical use and cost of PA assessment instruments at the epidemiological level (Wareham 1998). PA assessment has proven to be a difficult task; scientists and researchers have developed instruments such as logs, pedometers, accelerometers, and most recently: consumer-based PA monitors (CPAMs). Each instrument has its own set of positive and negative characteristics as well as practical uses.

Logs & Questionnaires

Since the 1970's, logs and questionnaires (hereafter called logs) have been used as a cost-effective way to assess PA habits within a large population (**Taylor 1978**). Logs are commonly used to assess PA after it has already been completed. Albanes et al. conducted a study which compared eight (8) questionnaires both self- and interviewer-administered. The results show error ranges of 14 to 39% for energy expenditure and correlations ranging between 0.05 and 0.54 amongst each other (Albanes 1990). Taylor et al. found that individuals overestimate PA intensity on a treadmill using the Minnesota LTPA log. They also found that occupationally light PA, which has been shown to have positive effects on health, was underestimated using the Minnesota LTPA scale (Taylor 1978). Taylor et al. has shown that individuals perceive PA differently and that logs, while cost effective, do not produce valid PA measurements. These results show that individuals recall PA differently whether due to memory loss, wording of the log, types of activities being assessed, or recall bias.

Using logs to estimate PA on an epidemiological level certainly has practical justification (e.g. cost and time effective); however, this practicality comes with compromised accuracy. It is also important to note that not all logs estimate PA equally and caution should be used when selecting a log or questionnaire for research purposes.

Pedometers

Pedometers are commonly worn on the user's right hip to detect vertical motion of the hip when a person steps. The movements are counted and displayed in real-time for the wearer to see. Pedometers are of particular appeal to consumers due to their ease of use and the easy-to-understand units. Corporations have used pedometers to improve the PA habits of their employees. Thomas et al. conducted a study which examined the PA habits of employees whom took part in a program using pedometers. Their study found that after the completion of the program, 43% of responders purchased a pedometer, 65% reported changing their PA levels, 70% included family members in their PA changes, and 54% of responders reported an increased knowledge of PA and its health benefits (Thomas 2006). Pedometers have also been shown to accurately estimate PA completed by children as well as adults. A study done by Clemes et al. showed that children above the age of five (5) can have their day to day ambulation accurately monitored with a pedometer (Clemes 2013). It is important to note the wear time of the pedometer when interpreting the results. In 2015, Laurson et al. showed that when wear time (in days per week) exceeded four (4) days, there was a significant increase in the amount of boys and girls who met the recommended steps. They also suggested that the wear time should be ten (10) hours at minimum further stating that anything less would not accurately capture the day's PA (Laurson 2015).

Pedometers have been shown to improve measurements of PA habits across a wide range of ages; however, pedometers have difficulty estimating PA intensity thus creating a need for a monitor that can adequately quantify various degrees of PA intensity.

Accelerometers

Accelerometers measure time-stamped accelerations of the body segment on which they are worn; accelerations of the body are related to intensity of the movement, allowing accelerations to differentiate PA into an array of intensities. Robson et al. investigated the relationship between sporadic PA and regimented PA; intensity being the most prominent contrast between the two. They found that 98% of MVPA accumulated in the average person's day is homogenously mixed with lower intensity PA making the ability to distinguish between these intensities significant (Robson 2015). Dowd et al. showed that accelerometers are valuable tools to measure sedentary time in addition to PA, further supporting their ability to differentiate between PA intensities (Dowd 2012). Another major difference between pedometers and accelerometers is the anatomical location where they are worn; pedometers are most commonly worn on the hip whereas accelerometers have been designed to be worn on the hip, wrist, or thigh. Accelerometers are more advanced than pedometers in that they produce intensity-specific PA estimations and can be worn on different anatomical locations; however, research-grade devices are not ideal for the day-to-day consumer. Data collected by an accelerometer are often labeled as "counts" and require further analysis to interpret. Accelerometers can also cost significantly more than pedometers further decreasing their popularity amongst the average consumer. Finally, most accelerometers do not provide real-time feedback and are, therefore, less useful to individuals looking for tracking progress toward daily or weekly PA goals.

Consumer-Based Physical Activity Monitors

CPAMs are the most recent development in the PA assessment spectrum. Using technology similar to accelerometers, CPAMs estimate PA into variables understood by the everyday consumer (e.g. steps, kcals, and active minutes). They do so by quantifying accelerations, processing these accelerations through company-specific propriety algorithms, and labeling them as PA. Various companies have produced models of CPAMs; most commonly are the Fitbit and Jawbone brands; over 3.3 million were sold in 2013 alone with 87% coming from Fitbit and Jawbone (Danova 2014).

CPAMs are alternative PA assessment instruments to the pedometer and research grade accelerometer. As previously discussed, pedometers are commonly worn on the hip and measure PA in the form of steps. Pedometers are cost effective alternatives to research-grade accelerometers as well as CPAMs; however, their estimations of PA are severely limited. Due to their measurements being made exclusively in the frontal plane, pedometers are not able to detect the intensity of PA. CPAMs offer the ability to quantify PA intensity but are also capable of synthesizing the accelerations into several variables whereas pedometers are restricted to a single PA variable.

One of the biggest advantages CPAMs have over research-grade accelerometers is the ability to provide consumers with real-time feedback of their PA.

Research-grade accelerometers are often produced without a display screen forcing users to download the data, process the accelerations, and use a series of cut-points to quantify their PA. CPAMs are also sold and at a fraction of the cost research-grade accelerometers are sold display PA data without any data analysis decisions that need to be made by the wearer, further increasing their appeal to the everyday consumer.

Part 3: Protocol Selection

Three styles of protocols are used to assess CPAMs in humans; they can be broadly grouped as structured laboratory, simulated free-living and free-living protocols. Each style of protocol exhibits a balance between structure, set by the researchers, and freedom allotted to the participant. Previous studies have used all three styles; their similarities and differences will be discussed in the following subsections.

<u>Structured Laboratory Protocol</u>

Structured laboratory protocols are one of the most popular protocol styles used to validate PA monitors in humans. Characteristically, these protocols have strong controls placed on the activity protocol as well as the participant and allow for use of high-quality criterion measures (e.g. indirect calorimetry). A limitation to this model lies in its applicability; consumers rarely participate in a series of activities for strict intervals of time (e.g. sweeping for 5 minutes then walking for 5 minutes), there is some question about the generalizability of findings from this type of study design (Bastian 2015). Despite this limitation, researchers often choose this model due to its prevalence in the literature allowing studies to be compared amongst each other and opportunity to study many types of activities (Nelson 2016).

Simulated Free-Living Protocol

Simulated free-living protocols are a hybrid between structured laboratory protocols and free-living protocols. This model commonly consists of an activity list organized by the research team which the participant chooses from to construct their own protocol. By structuring the protocol this way, researchers maintain a level of control over the participant and can use high-quality criterion measures; however, by giving the participant more freedom over the activities, the monitors' results are more applicable than structured laboratory protocols to performance in daily living. A limitation to this model of protocol is it's scarceness in the literature; of the three protocol models, simulated free-living protocols are the least common making their results difficult to compare to other studies. Also, they typically occur in a laboratory which still limits applicability to free-living settings.

Free-Living Protocol

A free-living protocol is the most applicable protocol type to activities of daily living. In this style of protocol, participants are equipped with the monitors being investigated along with an instrument which generates a criterion PA measurement. Participants wear this equipment for a set period of time after which, the monitors are returned and data can be analyzed. A significant limitation to this model of protocol is the inability to generate activity-specific data since the participants are performing an array of activity types, at different intensities, and at unspecified times of the day. Additionally, free-living protocols characteristically use criterion instruments which lack quality to measure PA. Each protocol type has its individual set of advantages and disadvantages; however, little research has been done to evaluate a CPAMs' performance using two or more of these protocol models.

Part 4: Validity of Consumer-Based Physical Activity Monitors

CPAMs offer an affordably balanced product to the everyday consumer by combining the accelerometer technology of a research-grade accelerometer with the low cost and usability of a pedometer. With these products becoming increasingly popular, it is important to understand their accuracy and reliability. The following subsections will review validation studies conducted on CPAM estimations of steps, kcals, and active minutes.

Steps

Taking 10,000 steps per day is a widely recognized daily PA goal set by many consumers and has been shown to have positive health benefits since 2000 (Kajioka 2008). In 2015, Storm et al. investigated the validity of the JU, worn on the right wrist, and FO, worn on the left hip, to estimate steps using a structured laboratory protocol as well as a simulated free-living protocol. The structured laboratory portion of Storm's study was 11 minutes in length and consisted of a 20 meter walk, descending 24steps, outdoor walk, ascending 24 steps, and an indoor walk. Simulated free-living activities included standing, taking a lift, working at a computer, lying down, and ascending/descending steps. Both the JU and FO showed higher mean absolute percent errors during slower activity speeds when compared to faster activity speeds. Mean absolute percent errors for both monitors were below 5% during all slow, self-selected, and fast activity speeds with the exception of the JU during slow speeds which had a mean absolute percent error of 11% (Storm 2015). Case et al. investigated the validity of the FO, FZ, Fitbit Flex (FF), and JU to estimate steps using a structured laboratory protocol.

Case had 14 participants walk on a treadmill at 3.0 mph for 500 and 1,500 steps with both distances completed twice; steps were tallied via direct observation. At 500 steps the accuracy of the FO, FZ, FF, and JU were 99%, 99%, 93%, and 96, respectively. The accuracy of these CPAMs did not change significantly at the 1,500 steps mark with 99%, 99%, 92%, and 98% accuracy for the FO, FZ, FF, and JU, respectively (Case 2015). Stackpool et al. assessed the validity of the JU to estimate steps using a structured laboratory protocol. Activities used were walking and running on a treadmill, using an elliptical, and while completing agility drills using 20 participants. Each participant partook in 20 minutes of each activity for a total of 80 minutes; steps were recorded via direct observation. Pearson correlations between the JU and criterion step count were 0.98, 0.98, 0.99, and 0.34 for the treadmill walk, treadmill run, elliptical, and agility drills, respectively (Stackpool 2014). From these studies, it is clear that CPAMs can estimate steps with a high degree of accuracy in a structured laboratory setting. However, once activities become increasingly subjective to the movement patterns of the consumer, their validity suffers.

<u>Kcals</u>

Losing weight by expending energy through PA is a common goal of consumers who plan to improve their PA habits. This common purpose places great significance on the accuracy of CPAMs to estimate the expenditure of kcals. Lee et al. studied the validity of the FZ, FO, and JU to estimate kcals using 30 men and 30 women in a structured laboratory setting. The 69 minute protocol used 13 activities which were organized into four categories: sedentary, walking, running, and moderate-to-vigorous. All activities lasted for five minutes with the exception of treadmill activities which lasted for three minutes. Mean absolute percent errors were calculated between the CPAMs and a criterion measurement; the results were 10.1%, 10.4%, and 12.2% for the FO, FZ, and JU, respectively across all activities (Lee 2014).

Ferguson and colleagues used 21 adult participants to investigate the FO, FZ, and JU using a free-living protocol. During the protocol, participants wore one of the aforementioned monitors along with a BodyMedia SenseWear Model MF (BodyMedia Inc., Pittsburgh, PA) for energy expenditure criterion measurements. The results showed Pearson correlations of 0.79, 0.81, and 0.79 with mean absolute differences of 349, 484, and 866 kcals for the FO, FZ, and JU, respectively, compared to the BodyMedia SenseWear (Ferguson 2015). In 2015, Bai et al. investigated the validity of the FF and JU to estimate energy expenditure using a laboratory-based protocol that differed from Lee's. The FF was worn on the left wrist while the JU was worn on the right wrist. Bai's protocol consisted of 20 minutes of sedentary time, 25 minutes at a self-selected treadmill speed, and 25 minutes of resistance training. Energy expenditure was measured continuously throughout the trial via indirect calorimetry with total energy expenditure from the trial used for analysis. Bai found the FF and JU overestimated kcals by 20.4 and 23.1kcals, respectively; the mean absolute percent errors were 16.8% for the FF and 18.2% for the JU (Bai 2015). These studies show a range of error exists with CPAM estimations of energy expenditure which could be attributable to the variability of activities between protocols, the use of different criterion measurement instruments, as well as biographical differences between samples.

Active Minutes

150 minutes of moderate intensity PA, 75 minutes of vigorous PA, or a combination of both became a benchmark for PA with the publication of the 2008 PA Guidelines for Americans. Despite the popularity of CPAMs and the notoriety of the 2008 PA Guidelines, little research has been done to validate CPAMs' ability to estimate these active minutes. Ferguson et al. investigated the FO, FZ, and JU in a free-living setting using a BodyMedia SenseWear Model MF worn on the left upper arm. The 52 participants went about their daily routine wearing the CPAMs and SenseWear for 48 hours before while data was collected. Pearson correlations between the CPAMs and criterion for active minutes were 0.91, 0.88, and 0.81 for the FO, FZ, and JU, respectively. Ferguson also calculated the mean absolute differences between CPAMs and the criterion to be 58.6, 89.8, and 18.0 active minutes over the protocol's 48 hours (Ferguson 2015). To better understand CPAM's performance to estimate active minutes, further research should investigate these CPAMs using structured laboratory and simulated free-living protocols.

Part 5: Reliability of Consumer-Based Physical Activity Monitors

Many of the aforementioned studies have investigated the validity, or accuracy of CPAMs. However, few are constructed in such a way that the consistency, or reliability, of CPAMs can be assessed. There are two variations of reliability; intra- and inter-monitor reliability. Both variations are unique in the characteristics of their analysis and provide separate points of view on the performance of CPAMs.

Intra-Monitor Reliability

More commonly known as test-retest reliability, intra-monitor reliability is the quantitative ability of a single CPAM to measure an activity when the same activity is completed several times. Methodologically, this is done using a structured laboratory protocol which includes activities done for the same period of time, at the same intensity, under the same conditions, multiple times. Kooiman et al. investigated the intra-monitor reliability of the FF, JU, and FZ to estimate steps using a structured laboratory protocol. Kooiman had 33 participants walk at 4.8 km/h (approximately 3.0 mph) for 30 minutes in consecutive bouts with data collected before and after bouts. Intra-class correlation coefficients were 0.81, 0.83, and 0.90 for the FF, JU, and FZ, respectively (Kooiman 2015). Limitations to this study include treadmill walking being the only activity performed. Kooiman's study is the only study in the literature present at the time of this review indicating need for further research evaluating the intra-monitor reliability of CPAMs.

Inter-Monitor Reliability

This type of reliability examines the performance of a CPAM, against a second unit of the same monitor, during the same activity. Logistically, two models of the same CPAM must be worn simultaneously by the participant, ideally in the same anatomical location while completing an activity. By orienting the monitors in this fashion, variables external to the CPAM itself are minimized. Diaz et al. utilized this protocol model to investigate the ability of the FO and FF to reliably estimate steps and kcals using a structured laboratory protocol. Three FOs were used, two worn on the left hip and one worn on the right; one FF was worn on each wrist. Participants in Diaz's study completed four bouts on the treadmill with each bout lasting six minutes; treadmill speeds were 1.9, 3.0, 4.0, and 5.2 miles per hour. Inter-device correlations for the FO on the right hip were 0.99 (steps) and 0.96 (kcals) for steps and kcals. Correlations for FOs worn on the right and left hips were 0.99 and 0.97 for steps and kcals, respectively. The inter-device correlations for the FFs were 0.90 (steps) and 0.95 (kcals) for steps and kcals (Diaz 2015).

Dontje et al. used a slightly different approach to investigate the inter-monitor reliability of the Fitbit Ultra to predict steps. In his study, Dontje used a single subject who wore 10 Fitbit Ultras for eight days except while sleeping and during water-oriented activities (e.g. showering). The Ultras were equally worn on the right and left hip during the eight day trial. The results showed no significant differences amongst all 45 potential pairs of monitors when estimating steps in a free-living protocol (Dontje 2015). Takacs et al. also investigated the inter-monitor reliability of the FO to estimate steps. Utilizing three FOs on different mounting sites (e.g. one on each hip and another in a front pants pocket), Takacs had 30 participants walk/jog on a treadmill at 5 speeds: 2.0, 2.5, 3.0, 3.4, and 4.0 miles per hour for 5 minutes each. Correlations ranged between 0.97 and 1.00 for all three monitors. (Takacs 2014). Together, the limited available research suggests high reliability of a limited variety of CPAMs for a limited number of tests activities.

Both variants of reliability carry their respective significance in quantifying habitual PA changes over time. Intra-monitor reliability provides insight to the consistency which a CPAM will quantify the PA completed during an activity. This is of particular concern when obtaining an amount of PA is the goal during repetitive exercise bouts. Consumers need to know if their CPAM is quantifying their PA consistently when they complete the same exercise. Inter-monitor reliability is more significant to researchers than consumers. Studies with a large sample size or use multiple models of the same CPAM (e.g. epidemiological studies) require consistent measurements to produce refined results and compare individuals wearing different CPAMs. If a particular model of CPAM has a low inter-monitor reliability, comparability of participants' data will be difficult. Furthermore, conclusions drawn from studies that have used CPAMs with low inter-monitor reliability should be interpreted with caution.

Part 6: Summary of Current Literature

Americans are becoming increasingly aware of the health benefits obtained from improved PA habits. Individuals who purchase CPAMs rely on their outputs to improve overall health. The validity of these CPAMs is somewhat well established; researchers have been diligent in assessing the ability of these monitors to estimate steps, kcals, and active minutes using an array of activities in all three protocol settings. However, there has been less research conducted which investigates the reliability of CPAMs; in particular, their intra-monitor reliability. More research needs to be done which examines a greater amount of CPAMs, the reliability of CPAMs during different activity intensities, as well as the reliability of the active minutes output in different protocol settings.

CHAPTER III

METHODOLOGY

Sample

Thirty participants between the ages of 18 and 80 years were recruited for this study. All participants were recruited from the East Central region of Indiana and were recruited via email and by word-of-mouth. Participants were excluded from this study if they possessed one or more of the following attributes: acute illness, unstable chronic conditions (e.g. severe hypertension), or if they were/are pregnant. In order to participate in this study, the participants must have also been able to perform all activities within the protocol (e.g. jogging). Participants were compensated with $20 for their involvement with the study.

Equipment

During the first of two visits to the HPL (Ball State University, Muncie, IN) the participants were outfitted with a total of nine PA monitors and a Cosmed $K4b^2$ portable metabolic analyzer (Cosmed Srl, Rome, Italy). These PA monitors consisted of two FFs, two FOs, two FZs (Fitbit Inc., San Francisco, CA), two JUs (AliphCom dba Jawbone, San Francisco, CA), and one activPAL (PAL technologies, Glasgow, Scotland). Steps, kcals, and active minutes from the FFs, FOs, FZs, and JUs were recorded by the CPAMs and collected using wireless internet, four iPod Touches, two iPads (Apple Inc., Cupertino, California), and the Fitbit and Jawbone mobile applications. A handheld tally counter was used by research assistants to count steps taken during each activity (serving as criterion measure of steps taken during the laboratory protocol).

The Cosmed was used to generate the criterion measure for kcals in this study. The $K4b^2$ system was worn via harness; the battery was placed on the participant's mid-back with the analyzer worn against the sternum (total weight <1kg). A mask was worn in conjunction with the $K4b^2$ and was used to collect expired gas samples from the mouth via a sampling line. $K4b^2$ data were downloaded from the analyzer into Microsoft Excel (Microsoft Co., Redmond, Washington) following the first visit. The $K4b^2$ calculated kcals based upon oxygen consumption and carbon dioxide production. The equation used by the $K4b^2$ is $3.781*VO_2 + 1.237*VCO_2$ and is based upon a study done by Elia et al. in 1992 (Elia 1992). Initial training for the $K4b^2$ was provided by a Cosmed company representative. Kcal data collected using the $K4b^2$ was not used to assess CPAM accuracy; its use was to determine if significant differences in energy expenditure occurred between bouts (e.g. manipulation check) via a dependent samples t-test.

The activPAL was used as a criterion measurement for steps during the free-living portion of the study but was not used for the current analyses. The activPAL was worn on the middle of the anterior surface of the right thigh and was securely held to the thigh via Tegaderm (3M Health Care, St. Paul, MN). ActivPAL data were downloaded from the monitor after visit two via the PAL Analysis v7.2.32 software package. Both the activPAL and the $K4b^2$ have been shown to generate valid and reliable measurements and served as criterion measures in this study (Dowd 2015, Storm 2015).

Procedures

Participants completed all required paperwork and the laboratory protocol during the first visit to the HPL; the second visit was used only to return the PA monitors. Prior to the initiation of the first visit, all participants read and signed the informed consent document. Height and weight measurements were taken thereafter; all measurements were recorded to the nearest centimeter and 0.1 kg, respectively. The participants' height, weight, and date of birth were then entered into all CPAMs and the $K4b^2$. Research assistants synchronized the accelerometer computer and Cosmed computer to a common clock prior to the initialization of all monitors. Participants were then outfitted with all monitors and the $K4b^2$. Once all of the CPAMs and analyzers were outfitted, initial readings of steps and kcals were collected from all CPAMs while the participant was in the seated position.

Following baseline data collection, participants carried out the laboratory-based activity protocol. Each participant underwent in an identical protocol where all activities lasted for five minutes, excluding transition time between activities. The only exception was the ascending and descending stairs activity which consisted of ascending and descending a flight of a stairs five times. All activities were performed twice in succession with CPAM data collected before and after each activity bout (e.g. typing was done for five minutes, data were collected, a second bout of typing was done for five minutes, and data were collected again). The activity protocol was structured in the following order: typing, reading, sweeping, treadmill walk at 2.0mph, treadmill walk at a moderate pace (3.0-3.5mph), treadmill jog (4.0-8.0mph), and ascending/descending stairs. The moderate treadmill walk, treadmill jog, and stair climb occurred at a self selected pace. The duration of the first visit averaged 120 minutes. $K4b^2$ data were downloaded following completion of the first visit. Start and end times of each activity were also recorded to match $K4b^2$ data to specific activities.

Participants also completed a free-living protocol after their first laboratory visit. During this protocol, the participants continued to wear either the wrist-mounted (FFs and JUs) or hip-mounted (FOs and FZs) CPAMs and the activPAL. These monitors were worn for the rest of the day and returned the following morning. The monitors assigned to the participants were counter-balanced among participants so each placement site (hip or wrist) was used during 15 free-living sessions.

Data Analysis

Data analysis for the first aim, intra-monitor reliability, was assessed via intra-class correlations independently for each outcome variable of interest (steps, kcals, and active minutes). Data used in this analysis came from only one monitor of each brand during both bouts of each activity. For wrist monitors, data from the distal monitor were used; for hip monitors, data from the anterior monitor were used. Data were excluded if the monitor reported a negative value, if the monitor did not update (e.g. no values were recorded), if the kcals value was not updated from the previous activity bout, or if a data point from one monitor was paired with data from its respective pair for any of the aforementioned data points. Once correlations were calculated per monitor and per participant, a repeated-measures analysis of variance was used to determine if significant differences existed among the CPAMs' intra-class correlations. Medians of the absolute differences were also calculated by determining the absolute difference between bout one and bout two for each variable of interest, then finding the median value for each monitor and variable.

The second aim, inter-monitor reliability, was assessed by comparing data from each monitor brand with its pair (e.g. one FF against the other FF) for one activity bout. For this analysis, data from only the first bout of each activity were used. Data were excluded if the monitor reported a negative value, if the monitor did not update (e.g. no values were recorded), if the kcals value was not updated from the previous activity bout, or if a data point was paired with any of the aforementioned data points. Pearson correlations, calculated per monitor and per participant, were used to define the inter-monitor reliability for each CPAM; a repeated-measures analysis of variance was used to determine if significant differences existed for correlations among CPAMs. Median absolute differences were also calculated by calculating the absolute difference between CPAM models for each variable of interest and finding the median of each respective data series.

A third aim identified the inter-monitor reliability of these CPAMs during a free-living protocol. Similar to the analysis for Aim 2, Pearson correlations were used to determine the inter-monitor reliability for these CPAMs to estimate steps, kcals, and active minutes.

Dependent samples t-tests using data collected from the $K4b^2$ (kcals) and direct observation (steps) were used to check for manipulation between bout one and bout two for all variables of interest.

Chapter IV

Journal Manuscript

This manuscript conforms to the format specifications for the authors of the journal:

Measurement in Physical Education and Exercise Science

Abstract: Establishing the Reliability of Several Consumer-Based Physical Activity Monitors.

BACKGROUND: Consumer-based physical activity monitors (CPAMs) have become increasingly popular in recent years; however, little research exists investigating their reliability. **PURPOSE:** The purpose of this study was to determine the intra- and inter-monitor reliability of several popular CPAMS for estimations of steps and Calories (kcals) in a structured laboratory setting as well as inter-monitor reliability for steps, kcals, and active minutes in a free-living setting. **METHODS:** Thirty participants (8 male, mean age 23.1±2.1 years) were recruited for this study. During the first of two visits, participants were outfitted with two each of four CPAM models (Fitbit One [FO], Jawbone Up24 [JU], Fitbit Flex [FF], and the Fitbit Zip [FZ]). Participants then completed a protocol consisting of seven activities: typing, reading, sweeping, slow treadmill walk (2.0 mph), brisk treadmill walk (3.0-3.5 mph), treadmill jog (4.0-8.0 mph), and ascending/descending a flight of stairs five times. Each activity was completed twice in succession and lasted five minutes per bout (except ascending/descending stairs). Participants were then given two pairs of CPAMs (four wrist- or hip-mounted) to wear in a free-living setting for one day. Intra- and inter-monitor reliability was assessed with intra-class and Pearson correlations, respectively. A repeated measures analysis of variance was used to identify if differences existed among monitor models. **RESULTS:** Intra-monitor reliability using raw, step data was 0.96±0.14, 0.92±0.28, 0.87±0.18, and 0.65±0.49 for the FO, FZ, JU, and FF, respectively. Once filtered, the same respective values were 0.97±0.13, 0.99±0.04, 0.88±0.20, and 0.89±0.20. The intra-monitor reliability for kcals using raw data was 0.75±0.14, 0.71±0.28, 0.68±0.18, and 0.54±0.49 for the FO, FZ, JU, and FF, respectively. After filtering, the correlations were 0.94±0.12, 0.90±0.17, 0.79±0.30, and 0.77±0.37 for the same monitors, respectively. Inter-monitor reliability in the laboratory for steps was 0.99±0.03, 0.96±0.13, 0.93±0.11, and 0.77±0.28 for the FO, FZ, JU, and FF, respectively using raw data. Once filtered, the correlations for steps were 0.99±0.03, 0.98±0.05, 0.93±0.09, and 0.81±0.28 for the same monitors, respectively. Inter-monitor reliability using raw kcal data was 0.86±0.31, 0.81±0.35, 0.75±0.34, and 0.43±0.53 for the FO, FZ, JU, and FF, respectively. After the data were filtered, kcal correlations were 0.91±0.22, 0.93±0.15, 0.78±0.31, and 0.64±0.38 for the same monitors, respectively. In most, but not all cases, hip-mounted CPAMs (FO and FZ) had higher correlations than wrist-mounted CPAMs (FF and JU) for intra- and inter-monitor reliability ($p < 0.05$). Free-living, raw inter-monitor reliability for steps was 0.99, 0.93, 0.69, and 0.50; kcal correlations were 0.35, 0.96, 0.69, and 0.77; and active minute correlations were 0.83, 0.52, 0.92, and 0.43 for the FO, JU, FF, and FZ, respectively. After the data were adjusted, correlations were 0.99, 0.94, 0.94, and 0.61; kcal correlations remained the same; and active minute correlations were 0.92, 0.48, 0.92, and 0.98 for the same monitors, respectively. **CONCLUSIONS:** Once adjusted, the intra- and inter-monitor reliability for all hip-mounted CPAMs was strong (≥ 0.80) for steps and kcals in the laboratory setting; free-living correlations were more sporadic yet strong for most monitors and variables of interest. The reliability of CPAMs appears to be strong which suggests that CPAMs provide reliable estimations of PA in laboratory and free-living settings.

Key Words: physical activity, accelerometry, steps, energy expenditure, activity tracker

INTRODUCTION

Regular physical activity (PA) has long been known to have positive effects on health-related variables such as body weight, cardiorespiratory fitness, blood pressure, cholesterol, and risk of cardiovascular or coronary artery disease (Garber, 2011, PAGAC, 2008, Morris, 1954). Despite the known benefits of regularly engaging in PA, 3-51% of adults living in the U.S. do not meet the 2008 Physical Activity Guidelines for Americans; depending upon the instrument used for PA measurement (Troiano, 2008, CDC 2010). To better understand the role of PA in reducing the societal burden of these non-communicable diseases, it is important to be able to measure PA accurately and reliably. Higher quality measurement techniques allow for the identification of which activity intensity provides optimal health benefits, monitoring intermittent bouts of PA, and more accurately measuring the effectiveness of interventions for promoting behavior change (Wareham, 1998).

Within the last decade, companies such as Fitbit and Jawbone have begun producing consumer-based PA monitors (CPAMs) that use technology similar to research-grade accelerometers but are offered at a lower price. CPAMs also estimate PA using variables commonly understood such as steps, Calories (kcals), and active minutes. CPAMs have become increasingly popular in recent years; over 3.3 million were sold in 2013 alone with 87% coming from Fitbit and Jawbone brands (Danova, 2014). With the overwhelming presence these devices have in today's market, it is important to understand their accuracy and reliability.

Knowing the reliability of CPAMs would allow consumers to have a better understanding of how their PA habits compare to others and how their PA changes over time. Two variations of reliability, intra and inter-monitor, must be considered when assessing the reliability of CPAMs. Intra-monitor reliability (also known as test-retest reliability) is the analysis of data collected from a single monitor, assessed against itself, when performing an identical activity multiple times. Inter-monitor reliability is a comparison of two identical monitors against each other during a single activity. Both forms of reliability are important to measure in order to best understand the performance of CPAMs.

While several studies have assessed CPAM validity, fewer studies have examined their reliability (Ferguson 2015, Kaewkannate 2016, Nelson, 2016, Storm 2015). Kooiman et al. assessed the intra-monitor reliability of the Fitbit Flex (FF), Jawbone Up (JU), and Fitbit Zip (FZ) to estimate steps using a laboratory-based treadmill walk protocol at 3.0 mph for 30 minutes and repeated this activity twice. The FF and JU were worn on the non-dominant arm while the FZ was carried in a front pants pocket. Intra-class correlations (ICCs) for the FF, JU, and FZ were 0.81, 0.83, and 0.90, respectively (Kooiman, 2015).

Takacs et al. also investigated the inter-monitor reliability of the Fitbit One (FO) in a laboratory setting at various walking and jogging speeds. Utilizing three FOs on different mounting sites (e.g. one on each hip and another in a front pants pocket), Takacs had 30 participants walk/jog on a treadmill at 5 speeds: 2.0, 2.5, 3.0, 3.4, and 4.0 miles per hour for 5 minutes each. ICCs ranged between 0.97 and 1.00 for all three monitors. The validity of these FOs was also strong; all monitors had < 1.3% error when compared to a directly observed step count (Takacs, 2014). Dontje et al. and Mammen et al. used case-studies to evaluate the intra-monitor reliability of Fitbit Ultra

(FU) to estimate steps. Their results showed strong ICCs, 0.995-1.000, and mean differences between 0 and 5% (Dontje, 2013, Mammen, 2012).

Diaz et al. examined both the validity and inter-monitor reliability of the FO and FF monitors for laboratory based walking and jogging. Results from this study showed inter-monitor reliability of 0.96 and 0.99 for FO estimations of steps and kcals, respectively. Diaz also showed correlations of 0.90 and 0.95 for steps and kcals, respectively, between the left and right wrist FF monitors (Diaz, 2015). Dontje et al. used a slightly different approach to investigate the inter-monitor reliability of the Fitbit Ultra to predict steps. In his study, Dontje used a single subject who wore 10 Fitbit Ultras for eight days except while sleeping and during water-oriented activities (e.g. showering). The Ultras were equally worn on the right and left hip during the eight day trial. The results showed no significant differences amongst all 45 potential pairs of monitors when estimating steps in a free-living protocol (Dontje, 2015).

These studies show strong reliability of CPAMs; however, the available literature on CPAM reliability is very limited regarding the diversity of activity protocols and the evaluation of only one type of reliability. To the authors' knowledge, there have been no studies which simultaneously examined both intra- and inter-monitor reliability. Additionally, little work has been done to assess CPAM reliability in a free-living setting with Dontje's work being the only study present in the literature. Therefore, the purpose of this study was to assess the intra- and inter-monitor reliability of several CPAMs to estimate steps and kcals in doing a variety of activities as well as the inter-monitor reliability of the same CPAMs to estimate steps, kcals, and active minutes in a free-living setting.

METHODS

Subjects

Participants in this study were 30 healthy adults recruited from the East-Central region of Indiana through word-of-mouth. All participants were right handed, Caucasian, and predominately female (8 male). To be eligible for this study, participants must have been free of gait abnormalities, free of acute illness, and have not been pregnant; they must also have been able to complete the protocol without risking injury beyond that of regular exercise. Prior to participating in the study, all participants read an informed consent form approved by Ball State University's Institutional Review Board which was voluntarily signed prior to the initiation of data collection.

Equipment

During the first visit, participants wore 9 PA monitors; 8 of which were CPAMs. The ninth monitor, an activPAL, was not used during data analysis. In addition to the PA monitors, participants wore a portable metabolic analyzer. Descriptions of the equipment used are provided below.

Fitbit One (Fitbit Inc., San Francisco, CA). The FO, a hip-mounted CPAM, was used to estimate steps, kcals, and active minutes. This CPAM has an internal, rechargeable battery and provides real-time feedback to its user. The FO has the capability to synchronize with the Fitbit Mobile Application via a Wi-Fi internet connection allowing the user to track PA over time.

Fitbit Zip (Fitbit Inc., San Francisco, CA). The FZ, a hip-mounted CPAM, was used to estimate steps, kcals, and active minutes. The FZ uses a CR-2032 watch battery and has the capability to synchronize with the Fitbit Mobile Application via a Wi-Fi internet connection.

Jawbone Up24 (AliphCom dba Jawbone, San Francisco, CA). The JU, a wrist-mounted CPAM, was used to estimate steps, kcals, and active minutes. This CPAM utilizes an internal, rechargeable battery and can provide real-time feedback to its user via Wi-Fi connection and the UP Mobile Application.

Fitbit Flex (Fitbit Inc., San Francisco, CA). The FF is a wrist-mounted CPAM which was used to estimate steps, kcals, and active minutes. This monitor utilizes an internalized, rechargeable battery and requires the Fitbit Mobile Application and a Wi-Fi connection to track PA.

Cosmed $K4b^2$ (Cosmed Srl, Rome, Italy). The $K4b^2$ is a portable metabolic analyzer which samples expired gas from the mouth and calculates oxygen consumption (VO_2) and carbon dioxide production (VCO_2). The $K4b^2$ was used as a criterion measurement for kcals and has been validated for of both steady state and interval-oriented activities. The equation used by the $K4b^2$ to calculate kcals is $3.781*VO_2 + 1.237*VCO_2$ and is based upon a study done by Elia et al. in 1992 (Elia, 1992).

Protocol

Participants completed all required paperwork and the laboratory protocol during the first visit to the HPL; the second visit was used only to return the PA monitors. Height and weight measurements were taken following the signing of the informed consent; all measurements were recorded to the nearest centimeter and 0.1 kg, respectively. The participants' height, weight, and date of birth were then entered into all CPAMs and the $K4b^2$. Research assistants synchronized the accelerometer computer and $K4b^2$ computer to a common clock prior to the initialization of all monitors. Participants were then outfitted with all monitors and the $K4b^2$. Once all of the CPAMs and analyzers were outfitted, initial readings of steps, kcals, and active minutes were collected from all CPAMs while the participant was in a seated position.

Following baseline data collection, participants carried out the laboratory-based activity protocol. Each participant underwent in an identical protocol where all activities lasted for 5 minutes, excluding transition time between activities. The only exception was the ascending and descending stairs activity which consisted of ascending and descending a flight of a stairs five times. All activities were performed twice in succession with CPAM data collected before and after each activity bout (e.g. typing was done for 5 minutes, data were collected, a second bout of typing was done for five minutes, and data were collected again). The activity protocol was structured in the following order: typing, reading, sweeping, slow treadmill walk at 2.0mph, brisk treadmill walk (3.0-3.5mph), treadmill jog (4.0-8.0mph), and ascending/descending stairs. The brisk treadmill walk, treadmill jog, and stairs activities occurred at a pace chosen by the participant. A list of activities, their descriptions, and average PA accumulated per bout is located in Table 2. The duration of the first visit averaged 120 minutes. $K4b^2$ data were downloaded following completion of the first visit and were used as a criterion measurement of kcals and active minutes during the laboratory protocol. Start and end times of each activity were also recorded to match $K4b^2$ data to specific activities. A trained research assistant, using a hand-held tally counter, generated the criterion measurements for steps during the laboratory activity protocol. The assistants were trained using a video produced by the principal investigator and lead author of this project. A step was defined as the foot completely lifting off the ground and was counted upon its return to the ground; shuffle and pivot steps were not counted.

Participants also completed a free-living protocol after their first laboratory visit. During this protocol, the participants continued to wear either the wrist-mounted (FFs and JUs) or hip-mounted (FOs and FZs) CPAMs and the activPAL. These monitors were worn for the rest of the day and returned to the lab the following morning. The monitors assigned to the participants were counter-balanced among participants so that each placement site (hip or wrist) was used during 15 free-living sessions.

Data Analysis

Intra-monitor reliability was assessed via intra-class correlations (ICCs) independently for each CPAM model (FO, FZ, JU, and FF) and outcome variable of interest (steps and kcals). Data used in this analysis came from only one monitor of each brand during both bouts of each activity. For wrist monitors, data from the distal monitor were used; for hip monitors, data from the anterior monitor were used. From the seven activities, an ICC was calculated for each participant's data. Once correlations were calculated for each participant, a repeated-measures analysis of variance (RM-ANOVA) with Tukey post-hoc analysis was used to determine if significant differences existed among the CPAMs' ICCs. Median absolute differences (MAD) were also calculated by calculating the absolute difference between each pair of monitors (intra-) and between bouts for the distal or anterior monitors (inter-). Once these values were determined, the median value was found per monitor and variable.

Inter-monitor reliability was assessed by comparing data from each monitor brand with its pair (e.g. one FF against the other FF) for one activity bout. For this analysis, data from only the first bout of each activity were used. Pearson correlations, calculated for each CPAM model and outcome variable of interest, were used to define the inter-monitor reliability for each CPAM; an RM-ANOVA with a Tukey post-hoc analysis was used to determine if significant differences existed for correlations among CPAMs. MAD and MPD were calculated for each CPAM model and variable of interest.

Dependent samples t-tests were used on criterion data collected during both activity bouts conducted using the laboratory protocol to check for manipulation ensuring that each activity in both bouts, and among participants, elicited similar PA.

Free-living data were analyzed using Pearson correlations in a similar fashion the laboratory inter-monitor reliability analysis. MAD and MPD were also calculated to describe the free-living data; this was conducted in the same manner as the laboratory inter-monitor MAD and MPD analyses.

All data were analyzed twice; once in its raw form and again with a series of exclusion criteria developed to exclude data likely influenced by monitor malfunctions. The exclusion criteria for the lab were as follow: 1) if any PA data were negative or if the kcals variable was not updated for a given CPAM, all data from that monitor were excluded for the respective activity. 2) if data from the CPAM's pair were also excluded (e.g. if one FO had bad data, data from the other FO were also excluded). Exclusion criteria for the free-living portion were if any PA variable was negative or if steps taken over the course of the setting were ≤ 150. If a data point within a CPAM met either of the aforementioned criteria, all data from that monitor was excluded from the participant. All analyses were conducted in SPSS version 23.0 (IBM, Armonk, NY) and Microsoft Excel (Microsoft, Redmond, WA). A p-value of < 0.05 was used to determine statistical significance. Nomenclature for correlation strength was assigned using thresholds published by Safrit et al. in 1995 (Safrit 1995).

Results

Participant characteristics are shown in Table 1 as a full sample and separately for each analysis. Dependent samples t-tests on the criterion measurements from the laboratory protocol revealed no significant differences between bouts 1 and 2 for any PA variable during any activity.

Intra-Monitor Reliability

Two participants were excluded from the intra-monitor reliability analysis, during both raw and filtered analyses. The two subjects removed from raw analysis had incomplete data sheets from data collection. Two subjects were removed from the filtered analysis due to errors encountered during data collection (e.g. poor synchronization of the mobile application resulting in a loss steps or kcals from an activity). The ICCs for steps and kcals are shown in Figures 1 and 2, respectively. All step ICCs were strong (≥ 0.80) using both raw and filtered data with the exception of the FF when using raw data which was moderately strong (≥ 0.60). Significant differences in the raw data set for steps were observed between the FF and all other CPAMs. In the filtered data set, significantly different ICCs were observed between the FZ and the

JU as well as between the FZ and FF for steps. All raw ICCs for kcals were moderately strong; when the data were reanalyzed after being filtered, the FO and FZ ICCs were strong. No significant differences in the raw correlations were observed; however, once filtered, there was a significant difference between the FO and the JU as well as the FO and FF for kcals.

Median Absolute Differences: Intra-Monitor

Data for MAD are shown in Table 3. Raw MAD values for steps were 3.0, 5.5, 18.5, and 0.0 for the FO, FZ, JU, and FF, respectively. When filtered, the MAD step values changed to 3.0, 4.0, 16.0, and 17.0 for the same monitors, respectively. Kcal MAD values were 3.0 for all monitors except the FF whose MAD was 0.1 kcals. After the kcal data were filtered, MAD values for all CPAMs were 3.0 kcals.

Inter-Monitor Reliability: Laboratory Setting

All 30 participants' data were included in the inter-monitor reliability analysis when raw and filtered data were used. Correlations for steps and kcals are shown in Figures 3 and 4, respectively. All step correlations were strong (≥ 0.80) with the exception of the FF when raw data was used which was moderately strong (≥ 0.60). Significant differences in step correlations were observed between the FO and JU, FO and FF, FZ and FF, and between the JU and FF when raw data were used. After the filtering process, significant differences were seen between the FO and JU, FO and FF, as well as between the FZ and FF. Kcal correlations calculated using raw data were all of moderate strength (≥ 0.60) with the exception of the FF (r=0.43). Once filtered, the FO and FZ correlations were strong (≥ 0.80); the JU and FF correlations were of moderate strength. Significant differences in raw kcal correlations were observed between the FO and JU, FO and FF, FZ and FF, and between the JU and FF. After the data were filtered, significant differences were observed between the FO and FF, FZ and JU, FZ, and FF, as well as between the JU and FF.

Median Absolute Differences: Inter-Monitor

Results from MAD are shown in Table 3. Raw MAD values for steps were 12.5, 29.7, 51.8, and 287.0 for the FO, FZ, JU, and FF, respectively. When filtered, the MAD step values changed to 11.5, 21.2, 50.1, and 245.0 for the same monitors, respectively. Kcal MAD values were 10.1, 64.3, 12.2, and 111.3 for the FO, FZ, JU, and FF, respectively. After the kcal data were filtered, MAD values decreased to 3.6, 62.8, 10.5, and 15.9 for the same monitors, respectively.

Inter-Monitor Reliability: Free-Living Setting

Each pair of CPAMs (wrist- or hip-mounted) was worn by 15 participants; minimum wear time was not mandated; however, the mean wear time in hours was 5.66 ± 3.77 hours. Characteristics of the participants who wore the hip- and wrist-mounted CPAMs can be viewed independently in Table 1. Raw, step correlations in this setting were 0.99, 0.93, 0.69, and 0.50 for the FO, FZ, JU, and FF, respectively. Once filtered, all correlations met moderate strength criteria (≥ 0.60) and were 0.99, 0.94, 0.94, and 0.61 for the same respective monitors. Raw, kcal correlations were of moderate strength with the exception of the FO (r=0.35). None of the free-living kcal data met the exclusion criteria therefore; the raw and filtered correlations were equivalent at 0.35, 0.96, 0.84, and 0.77 for the FO, FZ, JU, and FF, respectively. Raw, active minute correlations were 0.91, 0.47, 0.92, and 0.98 for the FO, FZ, JU, and FF, respectively.

After data were filtered, all correlations were classified as strong (≥ 0.80) with the exception of the FZ which decreased to r=0.48; the other correlations were 0.92, 0.92, and 0.98 for the FO, JU, and FF, respectively.

Median Absolute Differences: Free-Living Setting

MAD was also calculated using the free-living data and are shown in Table 4. MAD values for steps using raw data were 35, 134, 763, and 1,824 for the FO, FZ, JU, and FF, respectively. When filtered, the MAD values for steps were 35, 128, 731, and 154 for the same monitors, respectively. Raw, kcal MAD values were 30, 60, 26, and 88 for the FO, FZ, JU, and FF, respectively. None of the kcal data met exclusion criteria; therefore, filtered MAD values were identical to the raw kcal MAD values. Raw active minute MAD values were 0, 0, 6, and 6 for the FO, FZ, JU, and FF, respectively. When filtered, only the FF's MAD value changed (filtered MAD value = 0).

DISCUSSION

This study's purposes were to determine and compare the intra- and inter-monitor reliability of several CPAMs to estimate steps and kcals in a laboratory setting, and to determine the inter-monitor reliability of several CPAMs to estimate steps, kcals, and active minutes in a free-living setting.

The intra-monitor reliability findings from the raw data did not support the hypotheses set forth by the authors; the FF ICC failed to meet "strong" criteria for steps with an ICC of 0.65 (all other CPAM ICCs were ≥ 0.80) and all monitors' raw kcal data failed to meet this hypothesis. After the data were filtered, all monitors met the "strong" criteria for steps (Safrit, 1995). However, only the hip-mounted monitors (FO and FZ) met the “strong” criteria for kcals; the wrist-mounted monitor ICCs were classified as having moderate strength (≥ 0.60). These correlations are slightly lower than those found by Kooiman et al. who determined intra-monitor reliability for steps using the FF, JU, and FZ monitors; their ICCs were 0.81, 0.83, and 0.90, respectively (Kooiman, 2015). Participants in Kooiman’s study had an identical average age to the present study (23.1 years). Discrepancies between the studies could be attributable to a difference in activity protocol. Kooiman only used a single treadmill walking activity whereas the present study used seven different activities including both ambulatory (e.g. walking) and non-ambulatory (e.g. typing, sweeping). The larger number and greater variety of activities may give our results a more realistic indication of these monitors’ reliability. To our knowledge, there are no other studies that have investigated the intra-monitor reliability of the FO for steps or any CPAMs to estimate kcals; therefore, these findings extend the knowledge of past studies, showing consistently strong intra-monitor reliability, especially for step estimates, with a variety of CPAMs and activities.

Raw, step data from the inter-monitor reliability analyses did not fully support the researchers’ hypothesis that all CPAMs would have a strong Pearson correlation ≥ 0.80; the FF had a correlation of 0.77 for steps (Safrit, 1995). Once filtered, the FF’s correlation rose to 0.81 and all CPAM step correlations were strong. Raw, kcal correlations for the wrist-mounted CPAMs (JU and FF) failed to meet the researchers’ hypotheses of being ≥ 0.80; the hip-mounted monitors both had raw kcal correlations ≥ 0.80. Following the filtering process, all correlations rose however the FF was the only monitor whose classified strength increased to “moderate,” r=0.64 (Safrit, 1995). A study by Diaz et al. investigated the inter-monitor reliability of the FO and FF to estimate steps and kcals. Their results for steps were 0.99 (FO) and 0.90 (FF); the kcal correlations were slightly lower: 0.96 and 0.95 for the FO and FF, respectively (Diaz, 2015). The results from this analysis showed lower correlations than Diaz's findings for steps and kcals. The differences in results are likely due to the inclusion of various activities (present study) whereas Diaz used only a treadmill at speeds ranging from 1.9 to 5.2 miles per hour. As with the intra-monitor analysis, the greater variability in activities likely gives the present study better generalizability than previous works.

Takacs et al. also investigated the inter-monitor reliability of the FO to estimate steps. Takacs' showed that for each of the three mounting sites used for the FO, the inter-monitor reliability was > 0.97 (Takacs, 2014). The present study revealed the inter-monitor reliability of the FO in a laboratory setting to be 0.99 when estimating steps. The similarities in findings between the results of these studies is likely attributable to the similarities in protocol structure, participant characteristics, and mounting sites used for the FO. Together, these studies provide strong evidence to support the high inter-monitor reliability for the FO during ambulatory and non-ambulatory activities in a laboratory setting.

While all CPAMs in the present study yielded moderate to strong intra- and inter-monitor reliability in the laboratory, the hip-mounted CPAMs (FO and FZ) had statistically higher reliability than the wrist-mounted CPAMs (JU and FF) for both steps and kcals. Given the greater variability and higher accelerations of arm movement compared to hip movement while performing basic tasks, these results were expected. However, wrist-mounted CPAMs have better compliance than hip-mounted CPAMs

(Kamada, 2016, Fairclough, 2016, Troiano, 2014). Additionally, there are a greater number of wrist-mounted CPAMs than hip-mounted CPAMs on the market suggesting that they may be the more popular models. Accordingly, the choice of CPAM placement (wrist vs. hip) may depend on the importance of optimal reliability vs. optimal compliance and comfort.

Additionally, a review article by Evenson et al. reported findings from over 20 validity and reliability studies (Evenson, 2016). Evenson's findings found the inter-monitor reliability and validity for steps to be high when estimating steps using a treadmill-oriented protocol. They also noted that hip-mounted CPAMs outperformed monitors worn elsewhere on the body. Energy expenditure validity was found to be low despite the inter-monitor reliability being high. Evenson noted that the bias in energy expenditure was most often an underestimation. Results from Evenson's review are in partial agreement with findings from a recent study by Nelson et al. Nelson's findings showed <10% mean absolute percent error for steps during ambulatory activities but an overestimation in energy expenditure of 16-40% during similar activities. These studies show CPAM validity is variable and greatly influenced by the activity performed; however, their estimations appear highly reliable. This study adds evidence that while CPAMs may not provide precise estimations of PA, they are reliable instruments for detecting changes in PA habits over time.

A limitation of this study design was the length of the activity bouts. Fitbit algorithms minimally require 10 minutes of active time in order for their monitors to produce any quantification of PA for the active minutes variable. Since activity bouts in the laboratory's activity protocol lasted 5 minutes, the active minutes variable was not able to be assessed as a part of the intra- or inter-monitor reliability of these CPAMs in the laboratory setting. Further limitations include the small proportion of male participants in the study (27%) and a narrow age range (20-28 years). Also, sweeping was the only household-oriented activity in the present study. Further research should include the evaluation of intra- and inter-monitor reliability of CPAMs using more activities of daily living (e.g. folding laundry, doing dishes, gardening).

Strengths of the present study are the inclusion of the active minutes variable into the free-living analysis, as well as the simultaneous analysis of intra- and inter-monitor reliability. This study included both laboratory and free-living aspects, which provides better generalizability of results than studies without a free-living component. Additionally, the laboratory activity protocol included a variety of activities not previously assessed.

In conclusion, these CPAMs provide moderate to strongly reliable estimations of PA in the laboratory. Correlations calculated from the free-living session were also moderate to strong for most CPAMs and variables. These favorable findings suggest these CPAMs provide reliable estimates of PA in both laboratory and free-living settings.

References

1. CDC. Physical Activity Data and Statistics. 2010. http://www.cdc.gov/physicalactivity/data/
2. Danova T. Just 3.3 million fitness trackers were sold in the US in the past year. Businessinsider.com. 2014.
3. Diaz KM, Krupka DJ, Chang MJ, Peacock J, Ma Y, Goldsmith J, Schwartz JE, Davidson KW. Fitbit: an accurate and reliable device for wireless physical activity tracking. *Int J Cardiol.* 2015; 185: 138-140.
4. Dontje ML, de Groot M, Lengton RR, van der Schans CP, Krijnen WP. Measuring steps with the Fitbit activity tracker: An inter-device reliability study. J Med Eng Tech. 2015;39:286–90
5. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Inter J Behav Nutri Phys Act. 2015; 12:159.
6. Fairclough SJ, Noonan R, Rowlands AV, van Hees V, Knowles Z, Boddy LM. Wear compliance and activity in children wearing wrist- and hip-mounted accelerometers. Med Sci Sports Exerc. 2016 48: 245-53.
7. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, Nieman DC, Swain DP. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Med Sci Sports Exerc*. 2011: 1334-1339.
8. Kaewkannate, Kanitthika, and Soochan Kim. "A Comparison of Wearable Fitness Devices." BMC Public Health 16 (2016): 433.
9. Kamada M, Shiroma EJ, Harris TB, Lee IM. Comparison of physical activity assessed husing hip- and wrist-worn accelerometers. Gait Posture. 2016; 44: 23-28.
10. Kooiman T.J.M., Dontje M.L., Sprenger S.R., Krijnen W.P., van der Schans C.P., de Groot M. Reliability and validity of ten consumer activity monitors. BMC Sport Sci Med Rehab. 2015; 24.
11. Mammen G, Gardiner S, Senthinathan A, McClemont L, Stone M, Faulkner G. Is this bit fit? Measuring the quality of the FitBit step-counter. Health Fit J Can. 2012; 5:30–9.
12. Morris JN, Heady JA, Raffle PAB, Roberts CG, Parks JW. Coronary heart-disease and physical activity of work. Lancet 1953; 1053-1057.
13. Nelson MB, Kaminsky LA, Dickin DC, Montoye AHK. Validity of consumer-based physical activity monitors. Med Sci Sport Exer. 2016. Epub ahead of print.
14. PAGAC. Physical Activity Guidlines Advisory Committee Report, 2008. Washington, DC: US Department of Health and Human Services.
15. Safrit MJ and Wood TM. *Introduction to measurement in physical education and exercise science.* 1995; 3rd ed., p. 71. St. Louis: Mosby.
16. 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. 2014; 17:496–500.

17. 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. 2014;17:496–500.
18. 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: 181-188.
19. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. Br J Sports Med. 2014; 48: 1019-23.
20. Wareham NJ and Rennie KL. The assessment of physical activity in individuals and populations: why try to be more precise about how physical activity is assessed? *Inter J Obesity*. 1998: 22; S30-S38.

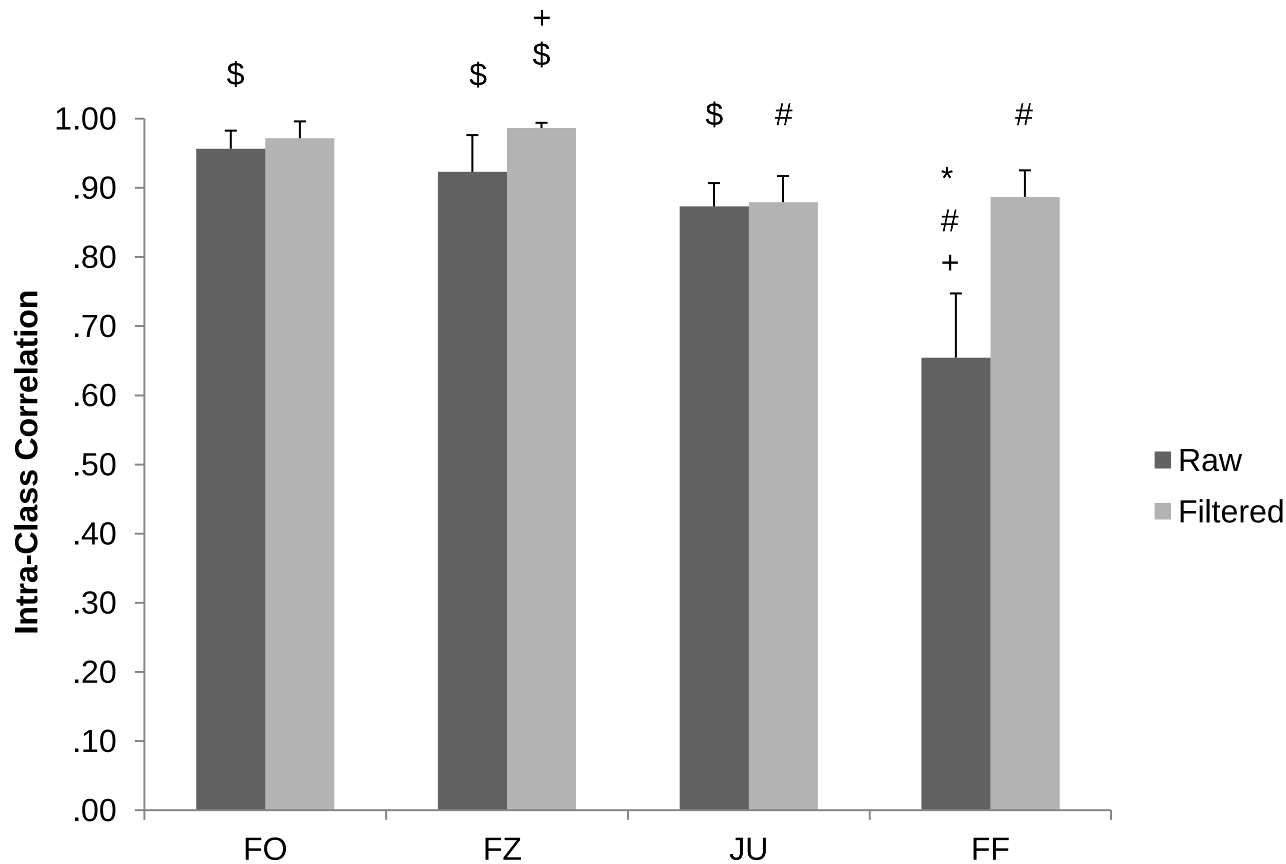

Figure 1: Intra-monitor reliability correlations for steps from the laboratory setting. All values are represented as intra-class correlations (ICCs) with standard error (SE) bars. * Indicates significant difference from FO. # Indicates significant difference from FZ. + Indicates significant difference from JU. $ Indicates significant difference from FF.

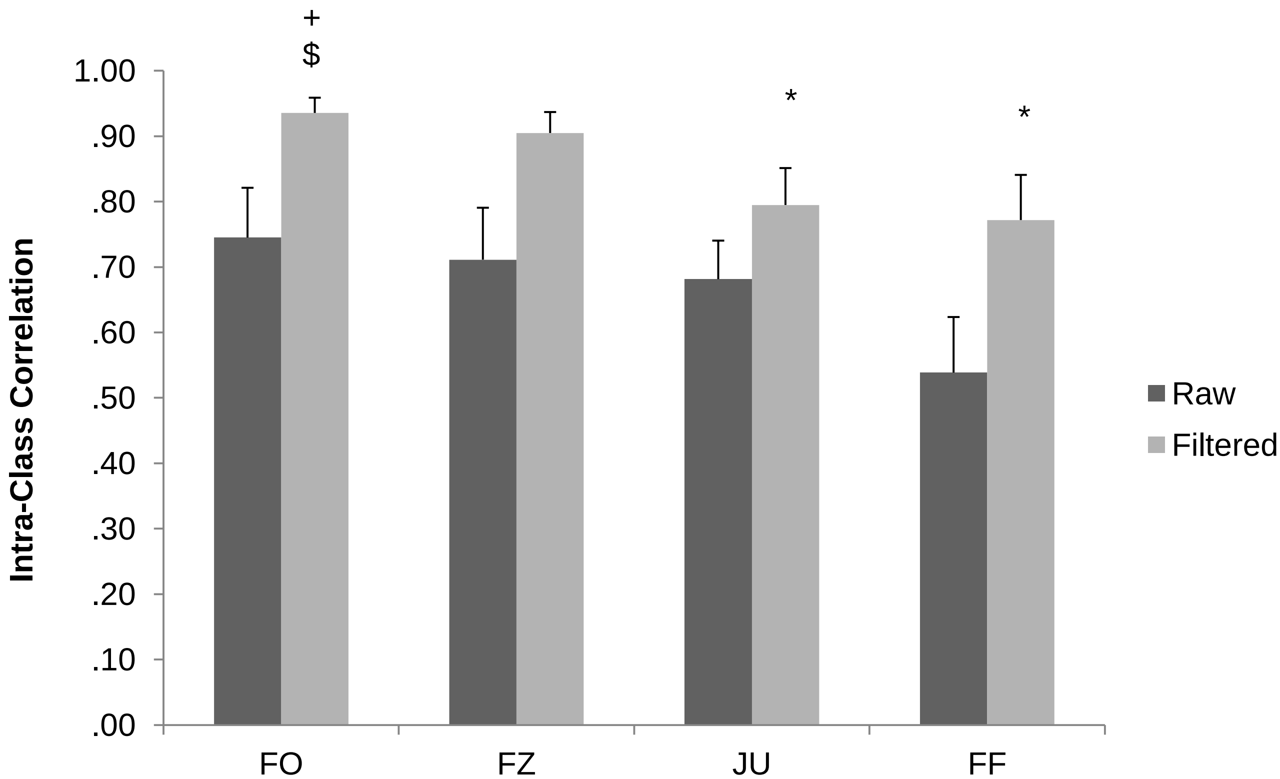

Figure 2: Intra-monitor reliability correlations for kcals from the laboratory setting. All values are represented as intra-class correlations (ICCs) with standard error (SE) bars. * Indicates significant difference from FO. # Indicates significant difference from FZ. + Indicates significant difference from JU. $ Indicates significant difference from FF.

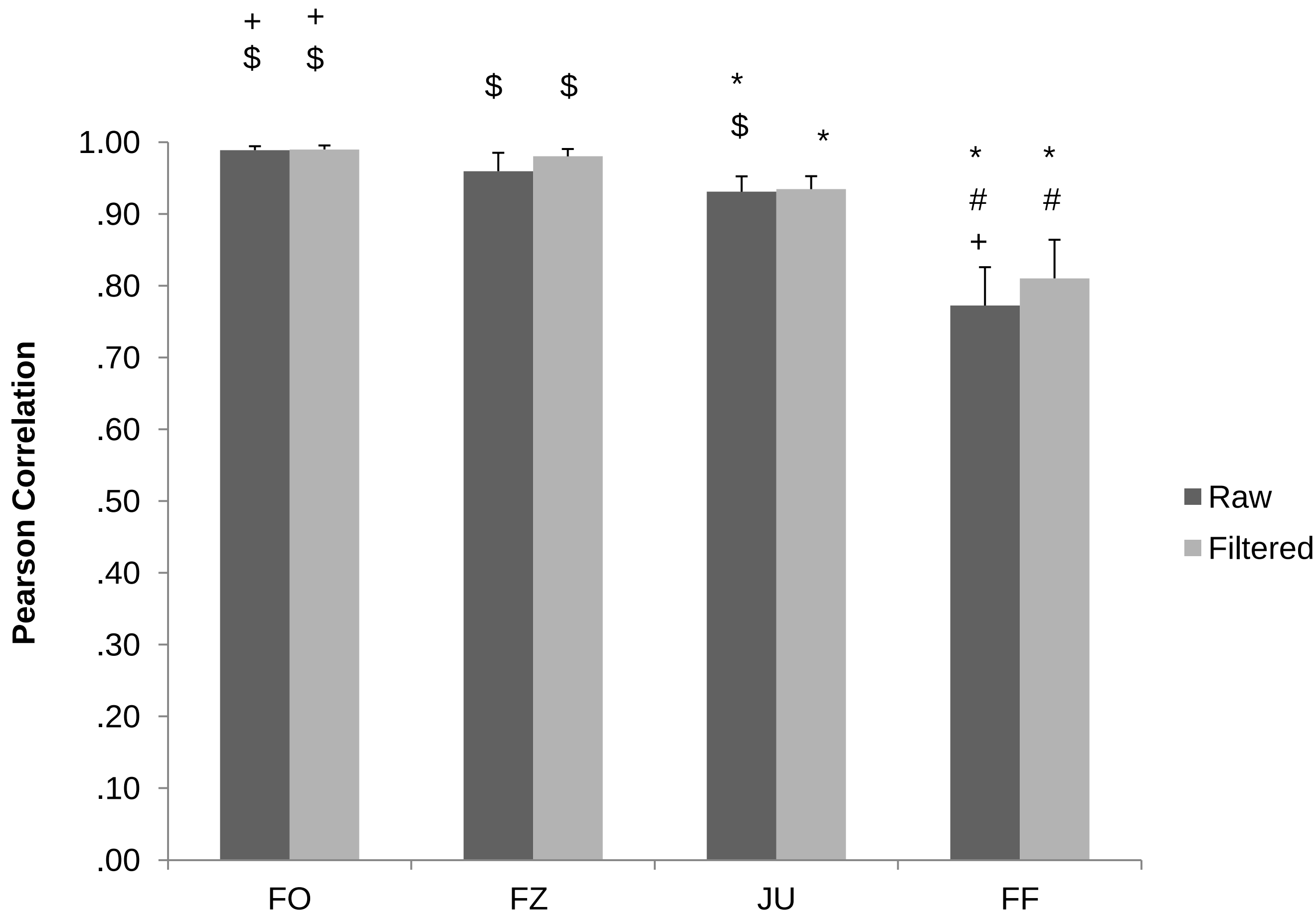

Figure 3: Inter-monitor reliability correlations for steps from the laboratory setting. All values are represented as correlations with SE bars. * Indicates significant difference from FO. # Indicates significant difference from FZ. + Indicates significant difference from JU. $ Indicates significant difference from FF.

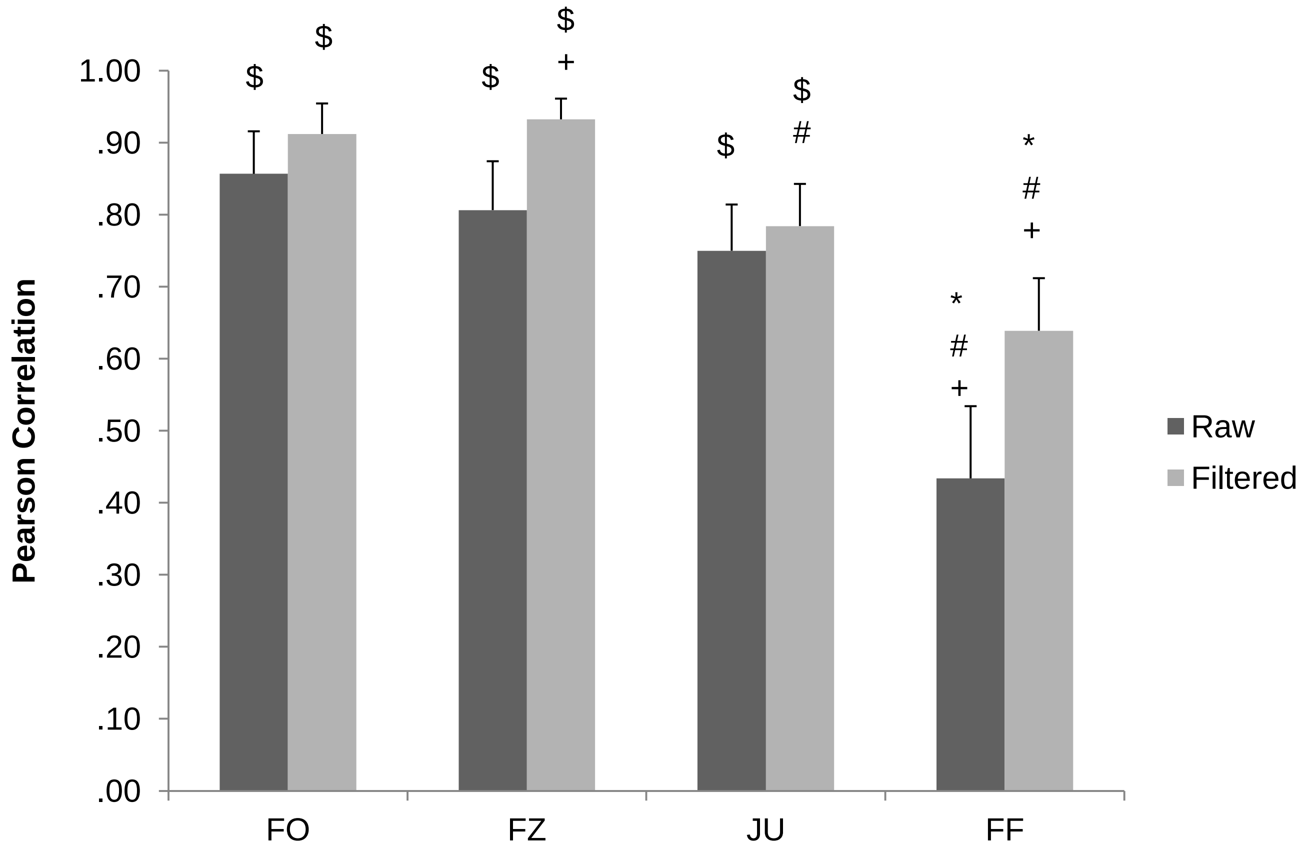

Figure 4: Inter-monitor reliability correlations for kcals from the laboratory setting. All values are represented as correlations with SE bars. * Indicates significant difference from FO. # Indicates significant difference from FZ. + Indicates significant difference from JU. $ Indicates significant difference from FF.

Table 1: Descriptive characteristics of participant groups and activities**.**

	All Participants	ICCs (n=28)	Pearson (n=30)	FL Hip (n=15)	FL Wrist (n=15)
Age	23.1 ± 2.1	23.0 ± 2.1	23.1 ± 2.1	23.8 ± 2.4	22.4 ± 1.7
BMI	23.3 ± 3.4	23.4 ± 3.5	23.3 ± 3.4	23.3 ± 2.7	23.3 ± 4.0
Treadmill Brisk	3.3 ± 0.2	3.3 ± 0.2	3.3 ± 0.2	3.3 ± 0.2	3.4 ± 0.2
Treadmill Jog	5.4 ± 1.3	5.5 ± 1.3	5.4 ± 1.3	5.2 ± 1.4	5.6 ± 1.2
Steps	4,267 ± 325.1	4,304 ± 314.3	4,267 ± 325.1	4,234 ± 346.1	4,300 ± 311.3
Kcals	261 ± 56.4	264 ± 57.5	261 ± 56.4	250 ± 53.0	273 ± 59.3

Age in years. BMI in kg*m-^{2}. Treadmill activities in miles per hour. Step and kcal statistics calculated using criterion data from the laboratory protocol. FL = free-living setting.

Table 2: Description of activities and average physical activity per activity.

Activity	Description	Steps	Kcals
Typing	Seated at desk, typing	0.0 ± 0.0	8.2 ± 2.4
Reading	Seated at desk, reading a magazine	0.0 ± 0.0	8.2 ± 1.8
Sweeping	Swept confetti within an 15" x 6" area, hand placement was chosen by participants	174.8 ± 62.1	17.0 ± 4.1
Walk	Slow walk on a treadmill at 2.0 mph	471.0 ± 41.1	17.9 ± 3.7
Brisk Walk	Brisk walk on a treadmill, participants selected speed between 3.0-3.5 mph	581.5 ± 56.5	24.9 ± 5.1
Jog	Jogging on a treadmill, participants selected speed between 4.0-8.0 mph	748.9 ± 102.4	46.3 ± 14.4
Stairs*	Ascended/descended a flight of stairs 5 times at a self-selected pace	126.4 ± 29.0	8.3 ± 2.6

***** Indicates activity not completed within a structured amount of time. Data presented as Mean ± standard deviation (SD) from criterion data.

Table 3: Median of the absolute differences between bouts (intra) and between monitors (inter).

	FO		FZ		JU		FF	
	Steps	Kcals	Steps	Kcals	Steps	Kcals	Steps	Kcals
Intra-	-	-	-	-	-	-	-	-
Raw	3.0	3.0	5.5	3.0	18.5	3.0	0.0	0.1
Refined	3.0	3.0	4.0	3.0	16.0	3.0	17.0	3.0
Inter-	-	-	-	-	-	-	-	-
Raw	12.7	10.1	29.7	64.3	51.8	12.2	287.0	111.3
Refined	11.5	3.6	21.2	62.8	50.1	10.5	245.0	15.9

FO = Fitbit One. FZ = Fitbit Zip. JU = Jawbone Up24. FF = Fitbit Flex.

Table 5: Median of the absolute percent differences between bouts (intra) and between monitors (inter).

	FO		FZ		JU		FF	
	Steps	Kcals	Steps	Kcals	Steps	Kcals	Steps	Kcals
Intra-	-	-	-	-	-	-	-	-
Raw	0.7%	15.4%	1.0%	16.1%	4.6%	25.0%	5.6%	18.2%
Filtered	0.7%	16.7%	0.8%	16.1%	4.5%	22.2%	5.1%	15.4%
Inter-	-	-	-	-	-	-	-	-
Raw	0.1%	9.8%	0.2%	11.8%	1.0%	18.9%	2.3%	16.0%
Filtered	0.1%	10.5%	0.2%	11.8%	1.0%	18.2%	2.1%	15.4%

FO = Fitbit One. FZ = Fitbit Zip. JU = Jawbone Up24. FF = Fitbit Flex.

Table 4: Median of the mean absolute differences between monitors in the free-living setting.

		Raw	Refined
FO	Steps	35	35
	Kcals	34	34
	Amins	0	0
FZ	Steps	134	128
	Kcals	60	60
	Amins	0	0
JU	Steps	763	731
	Kcals	26	26
	Amins	6	6
FF	Steps	1824	154
	Kcals	88	88
	Amins	6	0

FO = Fitbit One. FZ = Fitbit Zip. JU = Jawbone Up24. FF = Fitbit Flex. Amins = Active Minutes

Table 5: Percentage of data removed per monitor, variable, and analysis.

	FO	FZ	JU	FF
Intra-Raw	-	-	-	-
Steps	0.0%	1.4%	7.1%	1.0%
Kcals	0.0%	1.4%	7.1%	1.0%
Intra-Filtered	-	-	-	-
Steps	4.8%	8.6%	15.7%	18.1%
Kcals	2.1%	5.2%	11.9%	13.8%
Inter-Raw	-	-	-	-
Steps	0.0%	1.4%	6.2%	4.3%
Kcals	0.0%	1.4%	6.2%	4.3%
Inter-Filtered	-	-	-	-
Steps	7.1%	6.7%	16.2%	16.2%
Kcals	4.0%	4.8%	10.2%	13.8%
Free-Living Raw	-	-	-	-
Steps	0.0%	0.0%	2.2%	6.7%
Kcals	0.0%	0.0%	0.0%	0.0%
Amins	2.2%	2.2%	2.2%	11.1%
Free-Living Filtered	-	-	-	-
Steps	0.0%	1.1%	4.4%	8.9%
Kcals	0.0%	0.0%	0.0%	0.0%
Amins	4.4%	2.2%	2.2%	11.1%

FO = Fitbit One. FZ = Fitbit Zip. JU = Jawbone Up24. FF = Fitbit Flex. Amins = Active Minutes

Chapter V

Summary and Conclusions

The present study assessed the intra- and inter-monitor reliability of four consumer-based physical activity monitors (CPAMs), the Fitbit One (FO), Jawbone Up24 (JU), Fitbit Flex (FF), and the Fitbit Zip (FZ), to estimate steps and Calories (kcals) using a highly controlled activity protocol. Additionally, the inter-monitor reliability of the aforementioned monitors was assessed in a free-living setting for steps, kcals, and active minutes. Another purpose of this study was to compare the reliability of these CPAMs against one another to determine if any monitor significantly outperformed the others or if there were any monitors which performed significantly worse than the others.

The authors hypothesized that all CPAMs would have strong intra- and inter-monitor reliability (≥ 0.80) during the laboratory protocol but that the hip-mounted CPAMs (FO and FZ) would have significantly higher intra- and inter-monitor reliability than the wrist-mounted CPAMs (JU and FF). The authors also hypothesized than correlations in the free-living setting would be weaker than those measured in the laboratory setting for each respective monitor.

Findings from the intra-monitor reliability analysis did not support the authors' hypotheses. Most raw step ICCs were strong with the exception of the FF; once filtered for technical errors, all monitors had strong reliability. All of the raw kcal ICCs were moderately strong but following the filtering process, the hip-mounted CPAMs met "strong" criteria. Results from the inter-monitor reliability analysis were similar; all

CPAMs had raw correlations ≥ 0.80 for steps with the exception of the FF; but once filtered, it met "strong" criteria. Raw, kcal correlations for the hip-mounted CPAMs were strong, the JU was moderately strong, and the FF's correlation was weak. After being filtered, the hip-mounted CPAMs were strong and the wrist-mounted CPAMs were moderately strong. Correlations calculated using raw step data from the free-living portion of this study were strong for the FO, FZ, and JU but was moderately strong for the FF; the filtering process did not improve any of the classifications. Kcal correlations from the free-living portion were strong for the FZ and JU, moderate for the FF, and weak for the FO. None of the free-living data was affected by the filtering process therefore, none of the classifications were changed. Active minute correlations were strong for the FO and JU but weak for the FZ and FF; after being filtered, the FF correlation improved to "strong" but the FZ did not change classifications. In conclusion, these data support the use of CPAMs as instruments to track changes in PA habits over time via the steps variable.

Recommendations for Future Research

Future research should include activity bouts which last a minimum of 10 minutes; Fitbit algorithms require a minimum of 10 minutes in a given bout in order to accumulate active minutes and therefore longer bouts must be used to assess reliability of the active minutes variable. The activity protocol could also include more activities of daily living (e.g. folding laundry, making a bed). Including such activities would improve the generalizability of each CPAM's intra- and inter-monitor reliability.

Additionally, the present study investigated monitors which have since been succeeded by other models. These newer models produce the same PA variables as those researched in this study as well as heart rate, ambulatory pace, and sedentary time. Future research should assess the intra- and inter-monitor reliability of these more recent models and all the PA variables which they produce.

References

1. Albanes D, Conway JM, Taylor PR. Validation and comparison of eight physical activity questionnaires. *Epidemiology.* 1990; 1:65-71.
2. Bai Y, Welk GJ, Nam YH, Lee JA, Lee JM, Kim Y et al. Comparison of consumer and research monitors under semistructured settings. Med Sci Sports Exercise. 2015, in press.
3. Bastian T, Maire A, Dugas J, Ataya A, Villars C, Gris F, Perrin E, Caritu Y, Doron M, Blanc S, Ballon Pierre, Simon C. J Appl Physiol. 2015; 118: 716-722.
4. Blair SN and Connelly JC. How much physical activity should we do? The case for moderate amounts and intensities of physical activities. *Research Quarterly.* 1996: 193-205.
5. Butcher LR, Thomas A, Blackx K, Roberts A, Webb R, Morris K. Low-intensity exercise exerts beneficial effects on plasma lipids via PPARγ. Med Sports Sci Exer. 2008; 1263-70.
6. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. JAMA. 2015; 313:625–6.
7. CDC. Physical Activity Data and Statistics. 2010. http://www.cdc.gov/physicalactivity/data/
8. Clemes SA and Biddle SJH. The use of pedometers for monitoring physical activity in children and adolescents: measurement and considerations. J Physical Act Health. 2013; 249-262.
9. Cocate PG, de Oliveira A, Hermsdorff HHM, Alfenas RC, Amorim PRS, Longo GZ, Peluzio MCG, Faria FR, Natali AJ. Benefits and relationship of steps walked per day to cardiometabolic risk factor in Brazilian middle-aged men. J Sci Med Sport. 2014; 283-287.
10. 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; 45:2105–12.
11. Danova T. Just 3.3 million fitness trackers were sold in the US in the past year. Businessinsider.com. 2014.
12. Diaz KM, Krupka DJ, Chang MJ, Peacock J, Ma Y, Goldsmith J, Schwartz JE, Davidson KW. Fitbit: an accurate and reliable device for wireless physical activity tracking. *Int J Cardiol.* 2015; 185: 138-140.
13. Dontje ML, de Groot M, Lengton RR, van der Schans CP, Krijnen WP. Measuring steps with the Fitbit activity tracker: An inter-device reliability study. J Med Eng Tech. 2015;39:286–90
14. Dowd KP, Harrington DM, Donnelly AE. Criterion and concurrent validity of the activPAL professional physical activity monitor in adolescent females. *PLOS One.* 7: 1-9.
15. Elia M and Livesey G. Energy expenditure and fuel selection in biological systems: the theory and practice of calculations based on indirect calorimetry and tracer methods. World Rev Nutr Diet. 1992; 70: 68-131.
16. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Inter J Behav Nutri Phys Act. 2015; 12:159.

17. Fairclough SJ, Noonan R, Rowlands AV, van Hees V, Knowles Z, Boddy LM. Wear compliance and activity in children wearing wrist- and hip-mounted accelerometers. Med Sci Sports Exerc. 2016 48: 245-53.
18. 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. *IntJ Behav Nutr Phys Act.* 2015; 12:42.
19. Freedson PS, Melanson E, Sirard J. Calibration of the computer science and applications, inc. accelerometer. *Med Sci Sports Exerc.* 1998; 30: 777-81.
20. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, Nieman DC, Swain DP. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Med Sci Sports Exerc*. 2011: 1334-1339.
21. Jakicic JM, Clark K, Coleman E, Donnelly JE, Foreyt J, Melanson E, Volek J, Volpe SL. Appropriate intervention strategies for weight loss and prevention of weight regain in adults. Med Sci Sport Exer. 2001; 2145-2156.
22. Kaewkannate, Kanitthika, and Soochan Kim. "A Comparison of Wearable Fitness Devices." BMC Public Health 16 (2016): 433.
23. Kajioka T, Shimokata H, Sato Y. The effect of daily walking on body fat distribution. Enviro Health and Prev Med. 2000; 85-89.
24. Kamada M, Shiroma EJ, Harris TB, Lee IM. Comparison of physical activity assessed husing hip- and wrist-worn accelerometers. Gait Posture. 2016; 44: 23-28.
25. Kooiman T.J.M., Dontje M.L., Sprenger S.R., Krijnen W.P., van der Schans C.P., de Groot M. Reliability and validity of ten consumer activity monitors. BMC Sport Sci Med Rehab. 2015; 24.
26. Laurson KR, Welk GJ, Eisenmann JC. Estimating physical activity in children: impact of pedometer wear time and metric. J Phys Act Health. 2015; 12: 124-131.
27. Lee JM, Kim Y, Welk GJ. Validity of consumer-based physical activity monitors. Med Sci Sports Exerc. 2014; 46:1840–8.
28. Mammen G, Gardiner S, Senthinathan A, McClemont L, Stone M, Faulkner G. Is this bit fit? Measuring the quality of the FitBit step-counter. Health Fit J Can. 2012; 5:30–9.
29. Meyers J, Prakash M, Froelicher V, Partington S, Atwood JE. Exercise capacity and mortality among men referred for exercise testing. NEJM 2002; 793-801.
30. Morris JN, Heady JA, Raffle PAB, Roberts CG, Parks JW. Coronary heart-disease and physical activity of work. Lancet 1953; 1053-1057.
31. Nelson MB, Kaminsky LA, Dickin DC, Montoye AHK. Validity of consumer-based physical activity monitors. Med Sci Sport Exer. 2016.
32. Nybo L, Sundstrup E, Jakobson MD, Mohr M, Hornstrup T, Simonsen L, Bulow J, Randers MB, Nielsen JJ, Aagaard P, Krustrup P. High-intensity training versus traditional exercise interventions for promoting hearlth. Med Sci Sport Exer. 2010; 1951-1968.
33. PAGAC. Physical Activity Guidlines Advisory Committee Report, 2008. Washington, DC: US Department of Health and Human Services.

34. Robson J and Janssen I. A description of the volume and intensity of sporadic physical activity among adults. BMC Sport Sci Med rehab. 2015; 2.
35. Safrit MJ and Wood TM. *Introduction to measurement in physical education and exercise science.* 1995; 3rd ed., p. 71. St. Louis: Mosby.
36. Schneider PL, Bassett DR, Thompson DL, Pronk NP,Bielak KM. Amer J Health Prom. 2006; 2: 85-89.
37. Sesso HD, Paffenbarger RS, Lee I. Physical activity and heart disease in men: the Harvard alumni health study. Circulation. 2000; 102: 975-980.
38. Stackpool CM, Porcari JP, Mikat RP, Gillette C, Foster C. The accuracy of various activity trackers in estimating steps taken and energy expenditure. JFit Res. 2014; 3:32–48.
39. Storm FA, Heller BW, Mazza C. Step detection and activity recognition accuracy of seven physical activity monitors. *Plos One.* 2015; 10: 1-13.
40. 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. 2014; 17:496–500.
41. Taylor HL, Jacobs DR, Schucker B, Knudsen J, Leon AS, Debacker G. A questionnaire for the assessment of leisure time physical activities. J Chron Dis. 1978; 1978: 741-755.
42. Thomas L and Williams M. Promoting physical activity in the workplace: using pedometers to increase daily activity levels. Health Prom J Aust. 2006; 17: 97-102.
43. 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: 181-188.
44. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. Br J Sports Med. 2014; 48: 1019-23.
45. Tudor-Locke C, Craig CL, Brown WJ, Clemes SA, Cocker KD, Giles-Corti B, Hatano Y, Inoue S, Matsudo SM, Mutrie N, Oppert JM, Rowe DA, Schmidt MD, Schofield GM, Spence JC, Teixeira PJ, Tully MA, Blair SB. How many steps/day are enough? For adults. Inter J Behav Nutri Phys Act. 2011; 8(79).
46. Tudor-Locke C, Barreira TV, Schuna JM. Comparison of step outputs for waist and wrist accelerometer attachment sites. *Med Sports Sci Exer.* 2014; 839-842.
47. Wareham NJ and Rennie KL. The assessment of physical activity in individuals and populations: why try to be more precise about how physical activity is assessed? *Inter J Obesity*. 1998: 22; S30-S38.