EXAMINING VALIDITY OF CONSUMER-AVAILABLE ACTIVITY MONITORS IN MEASURING ENERGY EXPENDITURE, HEART RATE, AND STEPS

by

Veronika Pribyslavska

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy in Human Performance

Middle Tennessee State University
December 2017

Dissertation Committee:

Dr. Vaughn W. Barry, Chair

Dr. Jennifer L. Caputo

Dr. Dana K. Fuller

I dedicate this dissertation to my mom.

Tuto doktorskou práci věnuji své mamince. Moc tě miluju a děkuju za všechno.

ACKNOWLEDGEMENTS

I would like to express my gratitude to *all* the individuals who provided me with support and encouragement during the course of my doctoral studies, even to those who may not be mentioned on this acknowledgements page.

First and foremost, I would like to thank my fiancé, Eric M. Scudamore, who gave me the strength, through his continuous encouragement, reassurance, motivation, and confidence in my abilities, to “keep going”. Eric, I feel privileged to have gone through this chapter of my life along your side. You always reminded me that all the after-hour days spent in the office, all the sacrificed weekends, all the countless sleepless nights will in the end be worth it. As we both moved through the doctoral studies, we helped each other to grow professionally and we grew even closer together. I could not ask for a better partner in crime to share this part of my life journey.

I am also thankful for the support of my family. While I cherish all of them, the most influential individuals, specifically during my studies in the US, have been my mom Pavlína, my brother Petr, my dad Petr, my aunt Pavla, and my grandpa Jiří. I am very grateful for all that you have done and continue to do for me and for staying close, in spite of the thousands of miles separating us. I also want to extend my gratitude to Tammie and Mike Scudamore, my future family. I am humbled by your generosity, kindness, and support.

Lastly, writing this page would not be possible without my committee members. Dr. Barry, Dr. Caputo, and Dr. Fuller, I thank you all for the support and understanding when I decided to apply for and accept a job at Arkansas State University. I am beyond grateful for your willingness to commit your time in order for me to be able to start this new chapter in life. Dr. Barry and Dr. Caputo, thank you for always keeping your doors open for me, whether I needed a professional or personal advice. Finally, I want to express my appreciation to Dr. Farley, who was always there for me and went above and beyond to help me when complications presented themselves.

ABSTRACT

The first study of this dissertation evaluated the accuracy of the Fitbit Surge (FBS), Garmin Vívofit (GVF), and SenseWear armband (SWA) in measuring energy expenditure (EE), heart rate (HR), and steps during treadmill and cycling activities performed at two intensities in healthy, physically active individuals. In the second study, the monitors were evaluated in measuring EE and HR during a gym-based routine that included aerobic and resistance training activities performed by healthy, physically active participants.

In the first study, the activity monitors underestimated EE compared to the Oxycon Mobile (OM) metabolic analyzer across all bouts and their accuracy declined with vigorous intensity, signified by higher measurement error. The HR analyses revealed that the FBS and GVF yielded lower average HR (HR_{avg}), although the estimates were comparable to the Polar HR monitor (PM). The GVF had better HR accuracy over the FBS, however, the difference in accuracy was minimal during the moderate intensity treadmill bout. The same trend was observed for session maximal HR (HR_{max}). The step count analysis showed that all monitors accurately estimated steps during the vigorous intensity treadmill bout. During the moderate intensity bout, only the SWA had an equivalent step count with the video observation.

The results of the second study demonstrated that no monitor was equivalent to the OM in assessing EE. The FBS and GVF overestimated EE for all segments of the gym-based session. The SWA overestimated EE for the treadmill running bout but underestimated EE for the stationary cycling and resistance training bouts, which resulted

in reasonable whole-session EE estimates (450.9 ± 142.1 kcal) compared to the OM (470.6 ± 106.0 kcal), further supported by low measurement error. Equivalency testing for HR_{avg} data revealed that only the FBS did not agree with the PM during the stationary cycling bout. The GVF had superior accuracy in measuring HR indicated by lower measurement error across all segments of the session.

In conclusion, the activity monitors were the least accurate in measuring EE during common aerobic and resistance training activities. The monitors showed promising accuracy for measuring steps during treadmill walking and running. Lastly, the wrist-worn monitors demonstrated good potential in measuring HR_{avg} and HR_{max}, although the GVF appears to have a lower measurement error than the FBS.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER I

DISSERTATION INTRODUCTION

Physical activity (PA) and exercise are bodily movements produced by muscular contraction that result in energy expenditure (EE) above resting levels. The primary difference is that exercise is planned, structured, organized, and performed repeatedly with the goal of improving physical fitness. Engaging in PA and exercise provides a variety of health and fitness benefits (Blumenthal et al., 1989; Hollingworth, Harper, & Hamer, 2015; Kushi et al., 1997; Loprinzi, Lee, & Cardinal, 2013; Paffenbarger et al., 1993; Rockhill et al., 2001; Tanasescu et al., 2002; Williams, 2010).

In several large-scale investigations, in which self-report surveys of PA and exercise were employed, evidence of a dose-response relationship between PA and many disease outcomes including cardiovascular disease and premature mortality (Manson et al., 2002; Paffenbarger et al., 1993; Rockhill et al., 2001), hip fracture (Feskanich, Willett, & Colditz, 2002), and death from all-causes, cancer, and cardiovascular and respiratory diseases (Manson et al., 2002; Rockhill et al., 2001) has been documented. Accelerometry-based findings have also confirmed inverse dose-response relationships between PA and health complications, diseases, and mortality, respectively (Blumenthal et al., 1989; Dengel, Pratley, Hagberg, Rogus, & Goldberg, 1996; Healy et al., 2007; Loprinzi & Cardinal et al., 2013). Individuals who participate in at least light-intensity PA have a reduced risk of cardiovascular disease compared to individuals who participate in no PA. In addition, participation in higher amounts of light-intensity PA produces greater benefits (Loprinzi & Cardinal et al., 2013).

While there is substantial subjective and objective evidence supporting the relationship between PA and/or exercise with health and fitness benefits, there are methodological concerns in quantifying PA and/or exercise. Although self-report questionnaires are simple and accessible tools, inter-responder variability in interpreting PA terms may be problematic (Shepard, 2003). In addition, recalling past activities is difficult and may lead to incorrect subjective estimation of PA and/or exercise levels (Walsh, Hunter, Sirikul, & Gower, 2004). There are limitations associated with objective measures of PA and/or exercise as well, including the cost and training and experience requirements to operate the equipment, making some of these objective methods unfeasible outside of research settings. Motion sensors offer less obtrusive methods for objective assessment of health and fitness metrics and provide personalized feedback.

Continuous progression in technology has led to the development of wearable motion sensors available on the consumer market. These activity monitors can wirelessly connect to smartphones and social media applications for personal monitoring of fitness- and health-related parameters, such as time being active, step count, heart rate (HR), and an estimation of caloric expenditure and sleep patterns. These monitors are founded on accelerometry-based technology and may be combined with multiple sensors (such as skin and ambient temperatures, heat flux, galvanic skin response, or HR). Activity variables such as EE, steps, distance, or sleep are estimated using monitor-specific proprietary algorithms integrated with personal data (e.g. height, weight, age).

While commercially-available activity monitors are designed to target the general population, researchers are also including these monitors in research applications (Bai et al., 2016; Ellingson, Meyer, & Cook, 2016; Lee, Kim, & Welk, 2014; Price et al., 2016; Stahl, An, Dinkel, Noble, & Lee, 2016; Tucker, Bhammar, Sawyer, Buman, & Gaesser, 2015; Wang et al., 2015). The employment of activity monitors in research interventions has been shown to reduce prolonged bouts of sedentary behavior (Ellingson et al., 2016) and to increase walking activity in previously sedentary individuals (Kurti & Dallery, 2013). A valid activity measurement tool is fundamental to both epidemiological and intervention studies for establishment of appropriate PA and/or exercise recommendations. Therefore, it is important that the validity and the accuracy of these activity monitors be determined.

Numerous researchers have examined the validity of activity monitors in estimating EE (Bai et al., 2016; Lee et al., 2014; Nelson, Kaminsky, Dickin, & Montoye, 2016; Noah, Spierer, Gu, & Bronner, 2013), HR (Parak & Korhonen, 2014; Wallen, Gomersall, Keating, Wisløff, & Coombes, 2016) and steps (Hill, Wyatt, Reed, & Peters, 2003; Lindberg, 2000; Wilde, Sidman, & Corbin, 2001; Yamanouchi et al., 1995) in apparently healthy and active individuals. A comparison of previous findings is often challenging because there is large heterogeneity in research designs. In addition, some consumer-based monitors are represented in the literature more than other monitors.

Commonly assessed activity monitors for accuracy in estimating EE include the Fitbit, Jawbone, and Nike models, however, other monitors, such as the Garmin Vívofit (GVF), the Apple Watch (APW), and the BodyMedia Fit (BMF) have also been evaluated. In most studies, the accuracy of the monitors was assessed during laboratory conditions including treadmill walking/running or stationary cycling (Bai et al., 2016; Nelson et al., 2016; Noah et al., 2006; Stackpool, Porcari, Mikat, Gillette, & Foster, 2014). To evaluate the performance of the monitors, a few researchers also incorporated simulated daily-living activities, such as household chores, playing golf, tennis, or basketball, stair climbing, or elliptical (Lee et al., 2014; Tucker et al., 2015) and a resistance exercise session (Bai et al., 2016). Some studies have documented reasonable accuracy of certain Fitbit monitors (Bai et al., 2016, Lee at al., 2014), while others have not (Nelson et al., 2016; Noah et al., 2006; Stackpool et al., 2014). A wrist-worn Fitbit Surge (FBS) has only been included in two reports and shown to have poor accuracy in estimating EE during outdoor walking and running (Kirk, 2016) and laboratory-based treadmill ambulation and stationary cycling (Shcherbina et al., 2017). Findings for another wrist-worn monitor, the Fitbit Flex (FBF), indicated that, compared to a portable oxygen analyzer, the monitor produced lower EE estimates during stationary cycling (Nelson et al., 2016) and resistance training session (Bai et al., 2016), but higher EE during treadmill walking and jogging at self-selected intensities (Bai et al., 2016, Nelson et al., 2016).

Similarly, controversial results showing the Nike Fuelband (NFB) to be accurate in assessing whole-session EE (Tucker et al., 2014), but to overestimate total EE (Bai et al., 2016) have been reported. Contrasting findings have also been shown for the Jawbone Up (JBU) which underestimated EE during ambulatory activities on a treadmill across walking and running speeds (Price et al., 2016) and overestimated EE during treadmill walking (Bai et al., 2016). These discrepancies may be attributed to differences in study design, varying sample sizes, and/or varying statistical analyses used to evaluate the performance of the monitors. In addition, inaccurate measurement of HR via photoplethysmography (PPG) technology may increase the EE estimate error rate.

While commercially-available activity monitors have only been evaluated in a few studies for accuracy in assessing HR, there is also variability in these findings. For example, the Mio Alpha (MIA) has been shown to have reasonable accuracy for measuring HR for resting, treadmill walking and running (Stahl et al., 2016), and cycling (Parak & Korhonen, 2014; Wallen et al., 2016) activities, however, to underestimate HR during a weight lifting session (Spierer, Rosen, Litman, & Fujii, 2015). Stahl et al. (2016) and Wallen et al. (2016) concurrently evaluated the accuracy of the Fitbit Charge (FBC) and found that the monitor produced lower HR values than the criterion measure, but both authors concluded that the monitor performance was favorable. Lastly, Dooley, Golaszewski, and Bartholomew (2017) evaluated the validity of the FBC along with the APW in measuring HR during rest and walking and running on a treadmill. The monitors provided similar HR monitoring for some of the intensity bouts, but overestimated HR at low (APW) and high (FBC) intensities. A possible explanation for these discrepancies may be skin photosensitivity, as melanin concentration and skin pigmentation may impact the function of PPG technology (Fallow, Tarumi, & Tanaka, 2013), thereby altering the accuracy of the monitors.

Another commonly monitored activity variable is steps. Although commercially-available monitors tend to produce lower step counts than reference methods, these estimates have been considered favorable (An, Jones, Kang, Welk, & Lee, 2017; Case, Burwick, Volpp, & Patel 2015; Chen, Kuo, Pellegrini, & Hsu, 2016; Huang, Xu, Yu, & Shull, 2016). Hip-worn monitors provide more accurate estimates than wrist-worn monitors, however, the accuracy of the wrist-worn monitors is improved with increased speed (Chen et al., 2016; Diaz et al., 2015; Huang et al., 2016; Storm, Heller, & Mazza, 2015). For example, the GVF and JBU estimated steps with greater accuracy at a fast (5.0 mph) compared to a slow speed (3.0 mph; Chen et al., 2016). Similarly, Storm et al. (2015) showed that the JBU and the NFB produced lower error rates for ambulation at self-selected fast speed in comparison to a self-selected slow speed. The improved step count accuracy at higher speeds could be contributed to increased angular movement of the swinging arm (Thielemans, Meyns, & Bruijn, 2014), which in turn may enhance the ability of the monitors to detect the arm movement and yield favorable outcomes.

There is a need to examine the validity of consumer-based monitors in measuring activity variables to allow users to make educated decisions while purchasing monitors. In addition, establishing the accuracy and validity of these monitors will enable their utilization for research interventions. While many activity monitors have been represented in the literature, few studies have evaluated the accuracy of the FBS and GVF in measuring EE, HR, and steps in healthy individuals.

Overall Purpose

The investigations in the dissertation were intended to examine the validity of the FBS, GVF, and SWA activity monitors for assessing PA parameters. The primary purpose of the first study was to determine the validity during a laboratory-based protocol controlling for intensity and mode of exercise. The accuracy of the monitors in estimating EE compared to a portable metabolic analyzer during treadmill walking, treadmill running, and stationary leg cycling at two different intensities was evaluated in healthy, physically active individuals. The secondary objectives were to compare step count from the monitors to a video observation and to examine the accuracy of monitoring maximal and average HR compared to HR obtained via telemetry. The purpose of the second study was to validate the activity monitors against a portable metabolic analyzer for estimating EE during a simulated gym-based routine consisting of treadmill running, stationary cycling, and resistance training activities at self-selected intensities. The secondary purpose was to assess the accuracy of monitoring maximal and average HR compared to HR obtained by telemetry

CHAPTER II

REVIEW OF LITERATURE

The review of literature begins with a description of the terminology on physical activity (PA) and exercise, providing their definitions and listing commonalities and differences between the terms. The review then transitions into an overview of previous research examining health and fitness benefits associated with PA and exercise as examined by subjective and objective measurement methods. A brief paragraph is devoted to addressing limitations to using the mentioned instruments for the assessment of activity. The next section introduces wearable technology, a new appealing method for monitoring activity that targets the consumer market for personal use. While there is an array of wearable technology devices, the literature review will focus on activity monitors. The final section will focus on an overview of previous research examining the validity of activity monitors in measuring EE, HR, and step count.

Definitions of PA and Exercise

The terms PA and exercise are often interchanged. However, an important conceptual distinction exists between these terms. An understanding of this difference is necessary for appropriate interpretation and comparison of results across intervention and epidemiology studies (Caspersen, Powell, & Christenson, 1985).

Physical activity is bodily movement generated by contraction of the skeletal muscles that results in an increase in EE above resting level. The caloric contribution of

PA to the total EE occurs above the basal metabolic rate requirements and is determined by the nature of the activity. Physical activity can be divided into mutually-exclusive categories on the basis of will (voluntary or obligatory), intensity (light, moderate, or high), or segments of every-day life (sleep, work, leisure; Caspersen et al., 1985). In addition, leisure-time PA can be subcategorized into activities such as household tasks, sports, and conditioning exercises (Folsom et al., 1985). The amount of caloric expenditure is greatly influenced by the extent of participation in each category. Furthermore, the magnitude of EE depends on how much muscle mass is used to produce the bodily movements and on the frequency, intensity, and duration of muscular contractions. Energy expenditure will also vary across individuals based on fitness status, body composition, age, and sex.

Like PA, exercise also involves any bodily movement produced by skeletal muscles that results in increased EE. However, exercise is a more specific subset of PA in that it is planned, structured, repetitive, and goal-oriented. While the goal of exercise varies among individuals, the objective is either improvement or maintenance of one or more components of physical fitness that is either health- (body composition, cardiorespiratory fitness, flexibility, and muscular endurance and strength) or skill- (agility, balance, coordination, reaction time, speed, and power) related.

Because exercise is a subset of PA, it can constitute, in part or entirely, occupational and leisure-time PA. However, some categories of PA may be represented by exercise to a greater extent than others. For example, Caspersen and colleagues (1985) suggested that conditioning and many sports activities have the greatest probability to be deemed exercise. These two PA categories are almost always performed to maintain or improve physical fitness and, in such, are planned, structured, and repetitive. In contrast, occupational, household, and other daily activities are typically performed in the most time-efficient and energy-conserving manner with little or no concern for physical fitness. However, there may be instances in which occupational and household tasks are performed in such a manner to be considered exercise. For example, jobs requiring manual labor can be completed in a structured and planned way to improve muscular strength (e.g. a construction worker). Likewise, a person may plan to perform weekly household chores in a taxing manner to expend calories.

There are apparent similarities between PA and exercise. First, both involve bodily movement generated by skeletal muscles resulting in EE. Exercise, a subset of PA, is specifically intended to improve or maintain one or more components of physical fitness through structured, planned, and repetitive bodily movements. Both PA and exercise are positively related to physical fitness, resulting in a variety of health and performance benefits.

Benefits of PA and Exercise

Regular participation in PA and exercise is associated with a wide range of health and fitness benefits. To elicit these beneficial outcomes, all Americans should participate in 150 minutes of moderate intensity, 75 minutes of vigorous intensity, or any combination of the two intensities weekly (Physical activity guidelines advisory committee report, 2008). Additionally, based on the American College of Sports Medicine (ACSM), these amounts can be accumulated in 10-minute bouts (ACSM, 2014). Further improvements in health and fitness are obtained by engaging in higher intensities or durations of activity (Haskell et al., 2007). There is compelling subjective and objective evidence supporting decreased risk of chronic diseases, comorbidities, and premature mortality and improvements in physiological and physical variables from PA and exercise (Hollingworth et al., 2015; Kushi et al., 1997; Loprinzi & Lee et al., 2013; Paffenbarger et al., 1993; Rockhill et al., 2001; Williams, 2010).

Subjective measurement of PA and exercise and improvements in health

Participation in PA and exercise can be subjectively reported by having individuals self-report their activities, in either a recall fashion or an activity log. As with all self-report measures, there are concerns with lack of or incorrect recall and incorrect representation of intensity or duration of activities. These concerns notwithstanding, there exists a substantial body of literature documenting improvements in health relative to self-reported PA and/or exercise.

Participation in PA was associated with reduced risk of mortality in individuals who reported being more physically active compared to those who reported being less physically active (Kushi et al., 1997; Paffenbarger et al., 1993; Rockhill et al., 2001). Specifically, the risk of pre-mature all-cause mortality was inversely related to the total time of PA, suggesting that men who expended at least 3,500 kcal a week had half the risk compared to men who expended less than 500 kcal a week (Paffenbarger et al., 1993). Similarly, women who engaged in more hours of PA per week had lower risk for all-cause mortality and for death caused by cardiovascular diseases, cancer, and respiratory diseases (Rockhill et al., 2001). The positive effect of PA on health outcomes was also confirmed in ethnically-diverse postmenopausal women between the ages of 50 and 79 years in a large prospective cohort study (Manson et al., 2002). Baseline PA levels were assessed by a detailed self-report questionnaire and expressed as EE in metabolic equivalents of task (MET scores) per week. The 6-year follow-up revealed lower incidence of coronary heart disease (CHD) and total cardiovascular events for women who participated in regular PA, with decreased risk for greater quantities of PA.

Engaging in higher levels of PA has also been shown to reduce risk of hip fracture in postmenopausal women (Feskanich et al., 2002). In this prospective cohort study, postmenopausal women (40-77 years old) were followed for 12 years to assess the influence of PA on the incidence of hip fracture. Women who self-reported participation in higher amounts of PA (24 MET-hours a week) had 55% lower hip fracture risk than women who reported engaging in lower PA mounts (3 MET-hours a week). Additionally, a dose-response relationship between hip fracture risk and amount of PA indicated that the risk decreased by 6% for every 3 MET-hours increase in PA per week. It is apparent that PA is associated with many health benefits. In addition, improvements in cardiorespiratory function resulting from participation in PA lead to higher levels of habitual PA (Blair et al., 1995), which in turn results in many health benefits.

Several studies have also documented a positive influence of self-reported exercise on health and fitness (Hollingworth et al., 2015; Williams, 2013). Participation in aerobic exercise is associated with a decreased risk of all-cause mortality (Williams, 2013), cardiovascular diseases (Tanasescu et al., 2002; Williams, 2010; Williams, 2013), and cardiovascular risk factors (Hollingworth et al., 2015). A greater risk reduction in these maladies is achieved by individuals who reported engaging in higher amounts of exercise compared to individuals with lower amounts (Hollingworth et al., 2015; Tanasescu et al., 2002; Williams, 2010; Williams, 2013). A relationship between habitual cycling and cardiovascular risk markers (hypertension and hypercholesterolemia) and body mass index (BMI) was examined via an online survey completed by 6,949 male and female cyclists with a mean age of 48 years (Hollingworth et al., 2015). There was a dose-response relationship between self-reported cycling volume and risk of diagnosed hypertension and hypercholesterolemia. Furthermore, BMI was inversely related to cycling volume. This study suggests that higher cycling volumes are associated with a lower likelihood of cardiovascular risk factors in habitual male and female cyclists.

Similarly, the risk of angina and nonfatal and fatal CHD was lower in habitual male runners who reported running > 9 km/day compared to < 3 km/day (Williams, 2010). In addition, every additional km/day of running distance was associated with a 5% risk decrease in nonfatal CHD and revascularization procedures, a 7% risk reduction in nonfatal myocardial infarction, and a 10% decrease in angina. Tanasescu et al. (2002) also reported a dose-response relationship between self-reported exercise and the incidence of CHD. In addition, individuals who participated in greater amount of exercise, including activities such as running, cycling, swimming, or racquet sports, had lower BMI and prevalence of hypertension.

Collectively, these studies present sufficient evidence of the positive effects PA and exercise have on health, even when these activities are documented by self-report. In addition, there is a clear dose-response relationship between PA or exercise and the risk of disease. These conclusions are based on surveillance findings from either large cohort studies that require long-term follow-up of individuals or computer-, telephone-, or mail-mediated communication. Surveys call for an individual's ability to recall past PA or exercise participation to provide subjective quantification of activity. Objective measures of PA and exercise provide many benefits over subjective assessments.

Objective measurement of PA and exercise and improvements in health

Objective activity assessments may provide more precise measurements of intensity and duration of PA and exercise. Methods of objective assessment include direct and indirect calorimetry, doubly-labeled water, physiologic markers, heart-rate monitoring, and/or motion sensors.

Objective measurements to examine PA levels and the related benefits were employed in a population-based study from the 2003-2006 National health and Nutrition Examination Survey (NHANES). Accelerometry-assessed PA data exhibited a dose-response relationship between PA and having metabolic syndrome and its contributing risk factors (Loprinzi & Cardinal et al., 2013). A total of 5,538 individuals, 18 years of age or older, wore an accelerometer during all waking hours, except during water-based activities, for 7 days. Participation in light-intensity activities was associated with a decreased cardiovascular disease risk, although higher amounts of PA resulted in greater benefits. Furthermore, participants who engaged in a minimum of 71 minutes of moderate-to-vigorous PA a day were least likely to have metabolic syndrome and had the lowest triglyceride levels, waist circumference, and BMI, and the highest HDL levels.

Physical activity also has a positive influence on certain blood markers suggestive of chronic cardiovascular or metabolic diseases. For example, the association between PA and C-reactive protein, a marker of inflammation recognized to be related to cardiovascular disease risk, has been examined (Loprinzi & Lee et al., 2013). Physical activity was assessed by a hip-worn accelerometer over a 7-day period in a total sample of 4,555 participants from the 2003-2004 NHANES database. Physical activity was inversely related to C-reactive protein levels in adults, suggesting that PA has an anti-inflammatory effect and may therefore provide protection from cardiovascular disease. The relationship between PA and blood plasma glucose, a blood marker that is a precursor of type 2 diabetes, has also been studied (Healy et al., 2007). A total of 67 men and 106 women without diagnosed diabetes wore an accelerometer for 7 consecutive days and their PA levels were summarized into sedentary, light-intensity, and moderate-to-vigorous intensity. Individuals who were mostly sedentary had higher 2-h plasma glucose, while individuals who participated in higher moderate-to-vigorous intensity PA and increased light-intensity PA were associated with lower 2-h plasma glucose. Therefore, spending time in light- and moderate-to-vigorous PA helps to maintain blood plasma glucose within appropriate levels.

Evidence on the effects of exercise on health and fitness outcomes measured by objective assessments also exists. Aerobic exercise is associated with weight reduction (Dengel et al., 1996; Houmard et al., 2004), improved glucose metabolism (Dengel et al., 1996; Houmard et al., 2004) and improved plasma lipoproteins (Stefanick et al., 1998). Houmard et al. (2004) reported that 6-months of exercise training lead to improvements in insulin sensitivity in previously sedentary, overweight/obese individuals. The improvements were greater for a training program that incorporated 170 minutes compared to 115 minutes of exercise per week, supporting the evidence of a dose-response relationship between exercise and health benefits. Accordingly, a 10-month intervention including three weekly aerobic exercise sessions consisting of stationary cycling and walking and running on a treadmill for 30 minutes improved insulin sensitivity and reduced weight when paired with dietary modifications (Dengel et al., 1996). Another study examined the effect of 4-month aerobic exercise training program in older men and women (Blumenthal et al., 1989). The findings revealed that three, 45-minute sessions a week that consisted of leg and arm cycle ergometry and walking and running performed at 70% of maximum HR reserve improved cholesterol levels, diastolic blood pressure, and bone mineral content. In addition, improvements were also observed for cardiorespiratory fitness (Blumenthal et al., 1989; Dengel et al., 1996) and anaerobic threshold (Blumenthal et al., 1989).

Like aerobic training, regular participation in resistance training also elicits an array of positive changes in health status. In fact, resistance training is as effective as aerobic training in improving blood lipid profile in overweight/obese individuals (Schwingshackl, Missbach, Dias, König, & Hoffman, 2014) and for managing and treating type 2 diabetes (Yang, Scott, Mao, Tang, & Farmer, 2014). Shaibi et al. (2006) found that a 16-week resistance training program increased insulin sensitivity in overweight adolescents and this improvement was independent of changes in body composition. Insulin sensitivity was also improved in non-obese, young women (18-35 years) who participated in a 6-month resistance training program targeting enhancements in muscular strength and mass (Poehlman, Dvorak, DeNino, Brochu, & Ades, 2000). In addition, exercise that enhances muscular strength and mass increases regional bone mineral density and bone strength (Lohman et al., 1995), which may be especially valuable in individuals at risk for osteoporosis.

It is evident that PA and exercise offer a variety of positive health and fitness outcomes. Despite the well-known benefits, 31.1% of adults worldwide are inactive and approximately 50% of US adults meet aerobic activity guidelines, 30% meet muscle strengthening guidelines, and only 21% meet both guidelines (Centers for Disease Control and Prevention, 2013). The individual's ability to assess and monitor activity may induce behavioral changes and encourage PA and exercise (Bravata et al., 2007). The measurement tool must be valid, reliable, and practical (in terms of cost and utility; Laporte, Montoye, & Caspersen, 1985). It is, therefore, necessary that individuals have access to accurate and practical assessment tools. Direct and indirect calorimetry and doubly-labeled water are time consuming, require expensive equipment, and experience and training, confining these methods to research settings. Motion sensors may offer more convenient methods for assessing PA in free-living conditions.

Although accelerometers and PA questionnaires have been widely used as conventional methods to measure PA, there are limitations to each method as well. For instance, while questionnaires are relatively simple tools for PA assessment, participants must recall and quantify recent activity levels, which often results in an inaccurate estimation (Walsh et al., 2004). Furthermore, the reliability and validity of these subjective tools may be lacking (Shephard, 2003). Accelerometers were first introduced in 1961 (Cavagna, Saibene, & Margaria, 1961) to provide objective monitoring of PA and have been employed in many research settings since that time (Bouten, Koekkoek, Verduin, Kodde, & Janssen, 1997; Colley & Tremblay, 2011; Welk, Schaben, & Morrow, 2004). They provide a useful method of activity assessment, especially in large-scale studies. However, the relatively high cost and inconvenience of accessing PA data may be deemed as limitations. Continuous advancements in technology have led to recent developments of new activity tracking devices that are available for personal use and may offer more convenient and less expensive ways for individuals to monitor and track their own PA.

Activity Monitors

Wearable-technology products were introduced to the consumer marketplace a few years ago. Wearables, which is an abbreviated form of the term wearable technology, are small portable electronic devices that can either be worn on the body or connected to clothes and accessories. These gadgets include activity monitors, smart watches, global positioning system (GPS) tracking devices, HR monitors, and smart eye glasses. While there is an array of devices, the focus of this paper will be on wearable activity monitors and smart watches. Wearable activity monitors and smart watches will jointly be referred to as activity monitors in this paper.

Wearable technology has become ubiquitous, achieving a number-one fitness trend rank for the past two consecutive years (Thompson, 2015). The popularity of activity monitors has increased as they have become less obtrusive and more useful in their everyday utility (Evenson, Goto, & Furberg, 2015). According to the 2015 NPD Connected Intelligence Consumers and Wearables Report, one in 10 US adults reported owning an activity monitor to track activity and health patterns. The 2017 PA Council Report stated that ownership of a fitness tracking device among active individuals increased from 22.5% in 2014 to 28.0% in 2016. In addition, the interest in purchasing a wearable technology device to track PA between 2014 and 2016 increased from 3.0% to 10.7% in inactive individuals and from 9.3% to 25.7% in active individuals. Another survey reported that out of 1,000 Americans, 78% reported they would use an activity monitor and the reasons for using one included losing weight, achieving a fitness goal, or lower health insurance rates ("Where are Wearable Fitness Trackers Going for 2015?" 2014, October 30). Approximately 3.3 million monitors were sold between April 2013 to March 2014 and Fitbit (67 %), Jawbone (18%), and Nike (11%) represented 96% of the total trade (Danova, 2014). Stanley and colleagues (2014) forecasted a 154% annual growth rate in wearable shipments between 2013 and 2017. In addition, analysts predict wearable technology orders to increase from 9.7 million units in 2013 to 135 million in 2018, with wrist-worn devices accounting for 87% of the orders (Spann, 2015).

Activity monitors allow consumers to personally assess and track a variety of exercise modalities (walking, running, cycling, swimming, weight lifting, etc.) and other health- and fitness-related parameters, most commonly including HR, estimating caloric expenditure, counting steps, and monitoring sleep patterns. The convenience and practical utility of activity monitors enables consumers to track progress towards daily or long-term goals and to compare personal statistics with peers via specific platforms. The provision of interactive tools for goal-setting and feedback is an effective instrument to induce positive behavior and lifestyle modifications (Lyons, Lewis, Mayrsohn, & Rowland, 2014). Additionally, the application of consumer activity monitors has facilitated a growth of the Quantified Self movement, which is a self-quantification system for personal health information (Almalki, Gray, & Sanchez, 2015).

Activity monitors provide consumers with convenient and easy access to their health and fitness data. Generally, the collected data are transferred to smartphones, computers, or network storage clouds through wireless connectivity (for example Bluetooth). This wireless connectivity to external devices allows real-time activity data tracking utilizing device specific platforms. To track activity, the monitors utilize accelerometry-based technology that senses acceleration to detect bodily movements. Accelerometers can sense activity in one (uni-axial) to three (tri-axial) orthogonal planes (anteroposterior, mediolateral, and vertical; Chen & Bassett, 2005). Most modern activity monitors combine tri-axial accelerometry with other physiological sensors such as HR, skin and ambient temperatures, heat flux, or bioelectrical impedance (for respiration and galvanic skin response).

The inclusion of several sensors may improve monitors' accuracy in predicting EE or detecting activity mode. Some activity monitors interface with external accessories, for instance chest straps, while others implement technology called photoplethysmography (PPG) to measure HR. Photoplethysmography is a simple and low-cost optical technique that measures the amount of light absorbed by blood flow changes in the microvascular bed of tissues to detect HR (Allen, 2007). The PPG technology implements an optical emitter, typically light-emitting diode (LED) to shine light into the skin and a digital signal processor captures the light that is refracted and translates the signals into HR data. Activity variables such as EE, steps, distance, or sleep are estimated using monitor-specific proprietary algorithms integrating measured data along with information entered by the users.

Although commercially-available activity monitors have primarily targeted the consumer market, their use has also been implemented in research. Health and PA-related intervention studies have employed activity monitors for goal-setting, self-monitoring, measurement purposes, and behavior changes reinforcement (Cadmus-Bertram, Marcus, Patterson, Parker, & Morey, 2015; Ellingson et al., 2016; Kurti & Dallery, 2013; Wang et al., 2015). Evaluating the validity of monitors and establishing their accuracy should, however, precede their widespread application in intervention and measurement research. Although manufacturers conduct internal studies to evaluate their products, the documentation on the process and accuracy has a certain level of obscurity and the results reported are typically vague. Thus, external research has focused on examining the accuracy and validity of commercial activity monitors for various health and fitness metrics. In this review, the focus is on studies assessing EE, steps, and HR validity. Descriptive statistics of the most commonly represented activity monitors in the literature are displayed in Table 1.

Validity of Activity Monitors in Measuring Steps

Step count is a simple method of quantifying PA. It has been recommended that adults accumulate at least 10,000 steps per day to improve health (Hill et al., 2003;

Table 1

Consumer-available Activity Monitors Validated for Measuring Energy Expenditure, Steps, or Heart Rate

Monitor	Release date	Measurements	Ancillaries	Placement	Size (cm)	Mass (g)
Fitbit						
Tracker	2008 (discontinued)	Calories, steps		Bra, pocket, waist	5.5(h) × 1.9(w) × 1.4(d)	11
Ultra	2011 (discontinued)	Calories, steps	Altimeter	Bra, pocket, waist, wrist (requires a band)	5.5(h) × 1.9(w) × 1.4(d)	11
One	2012	Calories, steps	Altimeter	Bra, pocket, waist	4.8(h) × 1.9(w) × 1.0(d)	9
Zip	2013	Calories, steps		Bra, pocket, waist	3.6(h) × 2.9(w) × 1.0(d)	8
Flex	2013	Calories, steps		Wrist	S: 14.0-17.6(c) × 1.4(w) L: 16.1-20.9(c) × 1.4(w)	13 15
Charge	2014	Calories, steps	Altimeter	Wrist	S: 14.0–17.0(c) × 2.1(w) L: 16.1–20.0(c) × 2.1(w) XL: 19.8–23.0(c) × 2.1(w)	23
Surge	2015	Calories, HR, steps	Altimeter, GPS, LED	Wrist	S: 14.0–16.0(c) × 3.4(w) L: 16.0–19.8(c) × 3.4(w) XL: 19.8–22.6(c) × 3.4(w)	77
Charge HR	2015	Calories, HR, steps	Altimeter, LED	Wrist	L: 16.1–19.4(c) × 2.1(w) XL: 19.4–23.0(c) × 2.1(w)	23
Jawbone						
UP	2011 (discontinued)	Calories, steps		Wrist	S: 14.0–15.5(c) M: 15.5–18.0 21(c) L: 18.0–20.0(c)	19 21 23
UP24	2013 (discontinued)	Calories, steps		Wrist	S: 5.2(w) × 3.5(h) M: 6.3(w) × 4.0(h) L: 6.9(w) × 4.3(h)	19 21 23

Table 1

Consumer-available Activity Monitors Validated for Measuring Energy Expenditure, Steps, or Heart Rate (Continued)

Monitor	Release date	Measurements	Ancillaries	Placement	Size (cm)	Mass (g)
Nike						
Nike+ Fuelband	2013 (discontinued)	Calories, steps	LED	Wrist	S: 15.2(c) × 1.9(w) M: 17.2(c) × 1.9(w) L: 19.7(c) × 1.9(w)	27 - 35
BodyMedia						
Fit	2010 (discontinued)	Calories, steps		Upper arm	6.1(w) × 5.6(h)	45
Core	2012 (discontinued)	Calories, HR, steps		Upper arm	4.0(w) × 6.5(h)	30
Garmin						
Forerunner 225	2015	Calories, HR, steps	GPS, LED	Wrist	45(w) × 4.8(h) × 1.6(d)	54
Vívofit	2014	Calories, HR, steps	Heart rate strap	Wrist	2.1(w) × 1.1(h)	26
Apple						
Watch	2015	Calories, HR, steps	LED, Wi-Fi,	Wrist	S: 13.0-18.0(c) × 3.9(w) × 3.3(h) × 1.1(d) L: 14.5-21.5(c) × 4.2(w) × 3.6(h) × 1.1(d)	56 69
MIO						
Alpha	2013	HR	LED	Wrist	26.3(c) ×4.4(w)	56
Alpha 2	2015	Calories, HR		Wrist	25.2(c) × 4.3(w)	53

Note. c = Circumference; d = Depth; HR = Heart rate; h = Height; GPS = Global positioning system; L = Large; LED = Light-emitting diode; M = Medium; S = Small; w = Width; Wi-Fi = Wireless networking; XL = Extra-large.

Lindberg, 2000; Wilde et al., 2001; Yamanouchi et al., 1995). While commercially-available activity monitors provide consumers with an easy way to assess their daily step count, it is paramount to determine the accuracy of these values. Various wrist- and hip-worn activity monitors have been evaluated for step count accuracy in laboratory-based settings (Case et al., 2015; Chen et al., 2016; Diaz et al., 2015; Huang et al., 2016; Nelson et al., 2016; Noah et al., 2013; Takacs et al., 2014), during over-ground activities (Kirk, 2016; Storm et al., 2015), and in free-living conditions (An et al., 2017; Ferguson, Rowlands, Olds, & Maher, 2015). The step counts reported by the activity monitors have been compared to criterion measures that included pedometers, accelerometers, or manual step counting by a person or with video recording.

With respect to wrist-worn monitors, Chen et al. (2016) and Huang et al. (2016) reported reasonable accuracy for the GVF in measuring steps during short bouts of treadmill walking at slow, moderate, and fast speeds. The monitor produced low error rates, ranging between 1.5% and 5.6% across both studies. Another wrist-worn monitor, the NFB, yielded high error rates (up to 35%) while underestimating steps during slow walking, however, accuracy improved for walking at a faster speed (15% error rate; Storm et al., 2015). Two other wrist-worn monitors, the JBU and the FBF, were shown to underestimate steps during treadmill running (Chen et al., 2016; Nelson et al., 2016) and treadmill walking at slow (Case et al., 2015), moderate, and fast speeds (Chen et al., 2016, Huang et al., 2016). During slow walking on a treadmill, the FBF produced lower step count when worn on a dominant (461.8 ± 65.1) compared to a non-dominant (479.6 ± 35.6) wrist, however, the counts for both wear placements were significantly lower than the actual step count (513.7 ± 30.7; Chen et al., 2016). At the highest speed, the difference between the actual step count (815.3 ± 51.2) and the dominant (797.9 ± 46.3) and non-dominant (793.6 ± 43.4) sides, respectively was reduced. Diaz et al. (2015) also reported that the FBF underestimated steps by 16.3% during slow, 10.6% during moderate, and by 11.3% during brisk walking on a treadmill for 5 minutes at each speed. The JBU was suggested to underestimate steps during indoor and outdoor walking at slow, fast, and self-selected “natural” speeds on average by 36±178 steps (Storm et al., 2015), however, the error rate was smaller for the fast and self-selected (< 5%) speeds compared to the slow speed (10%).

The FBS provided similar step estimates to the actual step count during a 2-mile walk and run (Kirk, 2016). This was only a dissertation abstract and, unfortunately, no further details are disclosed. In contrast, the FBC, which is another wrist-worn monitor, underestimated steps (with 9.4% error) during 500 m of suburban walking, but had good inter-device reliability between dominant and non-dominant hands (de Man et al., 2016). Noah et al. (2013) reported the hip-worn FBU to be valid and reliable in estimating steps during level walking, but accuracy reduced when incline was incorporated. Although the step counts produced by the FBU (116.96 ± 6.85) compared to the research-grade monitor (117.22 ± 6.83) during incline walking were different statistically, the two monitors may be deemed equivalent in practical terms. In general, wrist-worn monitors were the least accurate at slower speed, but accuracy improved as speed increased.

Various hip-worn monitors have also been documented in the literature for their step-measuring accuracy. The Fitbit One (FBO) underestimated steps with lower accuracy for walking at a slow speed, but became more accurate during walking at a fast speed (Huang et al., 2016; Storm et al. 2015). The FBO, however, showed promising accuracy across all speeds with low error rates ranging between 1.0% (fast speed) and 3.8% (slow speed; Storm et al., 2016). Other authors reported concurring findings confirming the reasonable accuracy of the FBO in measuring steps during treadmill walking (Case et al., 2015; Diaz et al., 2015; Nelson et al., 2016; Takacs et al., 2014) and running (Diaz et al., 2015; Nelson et al., 2016). The error rate produced by the activity monitor was < 1.3% during 5-minute ambulation bouts on a treadmill at five speeds (54, 67, 80, 92, and 107 m•min^{-1}; Takacs et al., 2014), 1.5% for treadmill walking at 3 mph for 500 and 1,500 steps (Case et al., 2015), and < 3% at self-selected walking and running speeds (Nelson et al., 2016).

Other Fitbit models are also represented in the literature for step count validity. The FBZ is another hip-worn activity monitor that is considered a valid tool for measuring steps while ambulating on a treadmill at different speeds (An et al., 2017; Case et al., 2015). The monitor was consistently accurate across speeds from 2.0–5.0 mph and averaged 0.6% error across all speeds (An et al., 2017). In contrast, the monitor was less accurate in detecting steps for over-ground and treadmill ambulatory activities at slow to fast self-selected speeds and resulted in slight, although statistically significant, step count overestimation (2250 steps) compared to the observed step count (2206 steps; Nelson et al., 2015).

The step-counting accuracy of commercially-available activity monitors has also been assessed in healthy individuals during free-living conditions (An et al., 2017;

Ferguson et al., 2015). Various Fitbit activity monitors have been found to be accurate measures of steps during 24 hours (An et al., 2017) and 48 hours (Ferguson et al., 2015) of free-living activities. The Fitbit monitors represented in the investigations included the FBO, FBF, and FBZ and their accuracy was compared to research-grade monitors. The NFB is one of the least accurate monitors as it has been shown to undercount daily steps (An et al., 2017; Ferguson et al., 2015). The average underestimation of steps over a 48-hour period by the activity monitor was 2,529 (Ferguson et al., 2015).

Conclusions on the accuracy relative to steps

The general observation from the findings is that commercially-available activity monitors tend to underestimate actual step count. Overall, hip-worn monitors provide more accurate estimates compared to wrist-worn monitors. The wrist-worn monitors underestimate steps specifically at slower speeds with improved accuracy as speed increases. The greater accuracy at higher speeds may be explained by enhanced ability of the activity monitors to detect arm movement because the angular momentum of the swinging arm increases at higher speeds (Thielemans et al., 2014). One way to possibly improve accuracy of the monitors is to manually enter stride length if the option is available (Evenson et al., 2015).

Validity of Activity Monitors in Assessing EE

While counting steps may be an easy method to measure PA levels, monitoring EE may be more valuable for weight management purposes. Evaluating EE estimations may be achieved by comparing monitor estimates to EE measured by direct or indirect calorimetry. Most research studies use indirect calorimetry as the criterion measure (Bai et al., 2016; Lee et al., 2014; Nelson et al., 2016; Noah et al., 2014; Price et al., 2016), although the use of direct calorimetry has also been documented (Dannecker, Sazanova, Mlanson, Sazanov, & Browning., 2013). Certain activity monitors are represented in the literature to a greater extent than others and comparisons of findings across validation studies can be challenging due to the heterogeneity of protocols and statistical analyses used to evaluate the monitors' accuracy. In addition, only few studies have included an in-depth analysis to provide the reader with more precise insight on activity monitors' performance assessment.

The Fitbit is a commonly used monitor and various models (approximately 7) have been studied. Depending on the type of the activity monitor, it can be worn at the waist, pocket, bra, or wrist (see Table 1). The original Fitbit monitor, the Fitbit Tracker (FBT; also, referred to as the "original Fitbit"), tends to underestimate EE. For instance, Dannecker et al. (2013) showed that the monitor underestimated EE at different intensities while participants performed a series of activities including resting, stationary cycling, and walking at different grades and speeds for 3.5 hours in a room calorimeter. The monitor underestimated EE by 28.7%, however, the error was reduced to 12.9% when activities were manually classified via the monitor's web-based software. Underestimation of EE was also reported by Sasaki et al. (2015), who compared EE estimated by the FBT to a portable metabolic analyzer during sedentary, sport, and household activity routines. The monitor showed a systematic underestimation of EE across the routines with differences ranging from 104% (cycling) to 22% (treadmill walk at 4 mph and 5% grade). However, comparable EE estimation was found for running on a treadmill at 5.5 mph. In accord, the FBT underestimated EE during 6-minute bouts of level and incline (5%) walking at 3.5 mph and level running at 5.5 mph on a treadmill (Noah et al., 2013).

Other Fitbit monitors worn on the hip have been examined for EE accuracy during an assortment of activities. For example, the FBU, which was an upgraded model released few years after the FBT, underestimated EE during treadmill walking at a constant 3.5 mph speed (Noah et al., 2013) and at self-selected slow and brisk speeds (Gusmer, Bosch, Watkins, Ostrem, & Dengel, 2014), but slightly overestimated EE for running at 5.5 mph (Noah et al., 2013). The mean EE estimates during walking (6.7 ± 2.1 kcal) and running (14.3 ± 4.7 kcal) were statistically different from the metabolic analyzer (11.4 ± 2.9 kcal and 13.1 ± 2.5 kcal, respectively) with accuracy within 91-113% of EE and high correlations (r = .81 – .87; Noah et al., 2013). Slightly different outcomes for this monitor were found by Stackpool et al. (2014). They reported that the mean estimated EE, in comparison to the criterion measure, was lower during a 20-minute treadmill run (230.0 ± 50.5 kcal vs. 240.0 ± 47.3), and higher during a 20-minute treadmill walk (111.0 ± 22.8 kcal vs. 109.0 ± 19.6 kcal) at self-selected speeds. The mean EE values between the monitor and the criterion measure were not statistically different, however, they had a low correlation (r = .24) for the treadmill walk. Although both studies had similar sample sizes (N = 20, Stackpool et al., 2014; N = 23, Noah et al. 2013) of healthy male and female participants, the study protocols were different in that Stackpool et al. (2014) implemented walking and running at self-selected intensities as opposed to standardized speeds (Noah et al., 2013). This may have contributed to the discrepancies in the findings.

The FBT and FBU have been discontinued and replaced by two other hip-worn monitors, the FBO and Fitbit Zip (FBZ) that are still available on the market. Their accuracy in estimating EE has been documented in previous research. For example, in an investigation by Nelson et al. (2013), the FBZ and FBO underestimated stationary cycling at a self-selected intensity producing mean absolute percentage error (MAPE) up to 46% and overestimated EE during ambulatory activities on a treadmill at self-selected speed and grade, also with high error rates up to 68%. Although there is no set standard, ≤10% MAPE is considered reasonable and MAPE between 10% and 20% is considered moderately reasonable (Bai et al., 2016; Lee et al., 2014). Similar outcomes indicating overestimation of EE by the FBO were found during treadmill walking and running at various speeds on a level surface (Diaz et al., 2015) and on 1% incline (Price et al., 2016). Although the monitor demonstrated the ability to distinguish gross increases in EE by being highly correlated to the criterion measure across all speeds (1.6, 2.8, 4.0, 5.0, 6.2, and 7.5 mph), it overestimated EE with mean bias of 2.9 kcal per minute (Price et al., 2016).

In another investigation, the FBO underestimated EE with a 10.4% MAPE while the FBZ overestimated EE producing a MAPE of 10.1% during a continuous 69-minute session (Lee et al., 2014). The session consisted of simulated free-living activities including sedentary, ambulatory, household, and sport activities. For the first time in activity monitor validation research, the authors used a novel statistical analysis called equivalence testing to assess EE measurement agreement between monitors and the criterion measure. According to the equivalence testing analysis, the FBZ was considered equivalent to the criterion measure based on confidence intervals (CI) falling within an equivalence zone of ±10% from the measured mean EE. The authors concluded that the performance in estimating total EE by these two monitors was reasonable, particularly because they provided similar error rates relative to a well-established research monitor Actigraph GT3X (AG3X). The discrepancy in EE outcomes between the studies by Lee et al. (2014), Price et al. (2016), Diaz et al. (2015), and Nelson et al. (2016) may be explained by variations in the protocols. While the three latter studies reported the FBO accuracy for individual activities, Lee et al. (2014) analyzed total EE over the course of the whole session. It is possible that the monitor overestimated EE for some activities, but underestimated EE for most activities, thereby resulting in net EE underestimation of the whole session.

Wrist-worn models of the Fitbit, including the FBF, FBS, and FBC, have also been evaluated for assessing EE during a variety of activities. The monitors have been shown to vary in their accuracy depending on the activity mode as well as due to heterogeneity in study protocols. For example, the FBF has been shown to overestimate EE during treadmill walking and jogging at self-selected intensity (Bai et al., 2016; Diaz et al., 2015; Nelson et al., 2016), but to underestimate EE during stationary cycling (Nelson et al., 2016) and resting and resistance activities (Bai et al., 2016). The treadmill ambulation activities were similar in duration and the average EE overestimation was 40 kcal over 15 minutes (Nelson et al., 2016) and 52 kcal over 20 minutes (Bai et al., 2016).

The average EE underestimation was 5 kcal for the resting and 25 kcal for the resistance activities (Bai et al., 2016). The authors postulated that combining EE from all three activities (treadmill ambulation, resistance exercises, and resting tasks) resulted in a more reasonable EE estimation for the whole session due to the over- and under-estimates balancing out. It is also noteworthy that a MAPE provided by the FBF (16.8%) was comparable to a MAPE of the AGT3X (16.7%), which is the most commonly used research-grade monitor.

To date, there are only two reports on the FBS accuracy. An unpublished dissertation showed poor accuracy of the monitor when compared to a portable metabolic analyzer in estimating EE during a 2-mile walk and a 2-mile run on a treadmill (Kirk, 2016). Unfortunately, only the dissertation abstract is accessible and, therefore, no data statistics are obtainable. The other investigation also showed poor accuracy of the FBS indicated by a median error rate of 27.4% during a protocol consisting of slow and fast walking, slow and fast running, and exercising on a stationary cycle ergometer at a low and a high intensity (Shcherbina et al., 2017). However, it is important to note that only the last minute of each stage was used for the analysis, which may make generalizability of the results demonstrating the monitor's accuracy problematic.

The other wrist-worn Fitbit monitor, the FBC, overestimated EE compared to indirect calorimetry during 4-minute treadmill stages of light (2.5 mph), moderate (3.5 mph), and vigorous (5.5 mph) ambulation (Dooley et al., 2017). The EE estimates during the light and moderate bouts were considerably higher than the measured EE with large MAPE ranging from 45.8% (moderate) to 85.0% (light). In contrast, the FBC underestimated total EE during a routine consisting of 5-minute activities that included supine and seated rest, walking and running on a treadmill, and stationary cycling (Wallen et al., 2016). Although the literature is lacking in statistical evidence, it appears that the two Fitbit monitors are inaccurate in estimating EE during aerobic activities

Varying outcomes have been reported on the accuracy of the Fitbit. It appears that across all the Fitbit models, studies mostly indicate a bias toward EE underestimation for the hip-worn devices and EE overestimation for the wrist-worn devices. The hip-worn trackers seem to provide a more substantial agreement with a reference method in comparison to the wrist-worn monitors. This is specifically true for the FBC, which produced exceedingly large error rates ranging from 45.8% to 85%. In contrast, the FBF demonstrated reasonable accuracy signified by providing comparable MAPE to a well-established research-level monitor. The confounding findings on the accuracy of the monitor may be contributed to a diversity in study designs and sample sizes. In addition, the newer wrist-worn monitors (FBS and FBC) integrate PPG technology to measure HR. The proprietary algorithms include HR in the EE calculations for improved accuracy ("How does", 2016, May 16). However, it is possible that the EE estimate error would be augmented if the monitor inaccurately measured HR.

The validity of other consumer-based monitors has also been documented. Two commonly represented monitors in the literature include the JBU and NFB. In the study by Price et al. (2016), the JBU showed a bias to underestimate EE when participants walked at 1.5, 2.8, and 4.0 mph and to overestimate EE during running at 5.0, 6.2, and 7.5 mph. The EE estimates correlated strongly (*r* = .80) with the measured EE, which led the authors to conclude that the JBU is able to distinguish gross increases in EE across the speeds used in the study. Bai et al. (2016) reported that the JBU and the NFB yielded lower whole-session EE estimates than the criterion measure during semi-structured periods of sedentary, aerobic (walking and running on a treadmill), and resistance activities, with MAPE of 18.2% for the JBU and 17.1% for the NFB. Unlike the NFB, which consistently underestimated EE across all activities, the JBU underestimated EE during the 15-minute resting and 20-minute resistance activities, but overestimated EE for 20 minutes of treadmill walking and running. This resulted in a smaller difference in total EE between the criterion measure (316.8 ± 81.6 kcal) and the JBU (290.7 ± 99.0 kcal) in comparison to the NFB (274.5 ± 60.9 kcal). Findings presented by Stackpool (2014) indicated that both the JBU and the NFB overestimated EE while participants ran at self-selected speeds on a treadmill for 20 minutes. Furthermore, during a 20-minute self-paced treadmill walk, the JBU (123 ± 25.2 kcal) overestimated EE while the NFB (107 ± 24.2 kcal) had slightly lower mean EE compared to the criterion measure (109 ± 19.6 kcal).

It is not clear why the NFB performed differently for the treadmill walking and running in Bai et al. (2016) and Stackpool at al. (2014), considering the similar nature of the activities. One possible explanation could be in the way the EE data were obtained and analyzed. Unlike in the study by Stackpool et al. (2014) where the walking and running activities were performed and analyzed separately, Bai et al. (2016) analyzed the EE obtained during the 25-minute bout disregarding the activity modes and the proportion of participants who walked or ran. Therefore, the discrepancy in findings could be contributed to the different biomechanical and energy demands associated with walking and running. In addition, Bai et al. (2016) used JBU24, which is an updated version of the JBU. The monitor possibly operated on upgraded firmware integrating upgraded algorithms to compute EE.

The performances of the JBU and the NFB were also evaluated in the previously mentioned study by Lee at al. (2014). During the continuous 69 minutes of simulated daily activities routine performed by 60 participants, the monitors exhibited a systematic bias to underestimate EE, producing MAPE of 13% for the NFB and 12% for the JBU. Based on equivalence testing analysis, the mean EE by the JBU (338.8 ± 66.1 kcal) exceeded the lower limit of the ±10%, while NFB (350.2 ± 41.8 kcal) was accurate within a margin of ±10% error from the portable oxygen analyzer (356.9 ± 67.6 kcal), however, the correlation with the criterion measure was weak (r = .35). In a similar protocol, during which 24 participants performed a 60-minute activity routine, Tucker et al. (2015) reported the NFB to be accurate and reliable in estimating EE. The statistical results contrasted with those by Lee et al. (2014), demonstrating that the activity monitor yielded comparable EE estimates (246.0 ± 67.0 kcal) to the measured EE (243.0 ± 67.0 kcal) with no systematic bias and a moderate correlation (r = .77) with the criterion measure. Furthermore, the NFB produced a slightly higher MAPE of 16% with an upper CI limit exceeding the ±10% equivalence zone (although by only 4 kcal). While the discrepancy in findings is difficult to explain, it was proposed that adding resting EE to the monitor's activity EE estimates to assess total EE (Lee et al., 2014), as opposed to assessing activity EE only (Tucker et al., 2015), contributed to the differences. Another explanation could be that Lee at al. (2014) had a nearly three times larger sample size than Tucker at al. (2015). Lastly, it is possible that in right-handed participants, some activities (for example ironing, vacuuming, tennis, or basketball dribble) may not have been adequately detected by the monitor because the NFB was worn on the left wrist only in both studies.

The accuracy in measuring EE of a few other activity monitors have also been previously examined. For example, the GVF is another monitor that was concurrently assessed for validity in the above-mentioned study by Price et al. (2016). It was shown to be strongly correlated (r = .85) with the criterion measure across all ambulation speeds. Significant correlations were also reported for walking (r = .71) and running (r = .35) speeds separately. The GVF underestimated EE with a mean bias of 1.6 ± 2.4 kcal/min across all speeds, 1.7 ± 1.21 kcal/min for walking speeds with greater proportional bias at higher speeds, and 1.5 ± 3.1 kcal/min for running speeds. The largest mean difference, of approximately 40%, was reported for walking at 1.8 mph and the smallest mean difference of roughly 4% for running at 5.0 mph. Similar findings were reported by Alsubheen and colleagues (2016) showing that the GVF underestimated EE on average by 29.5% in comparison to indirect calorimetry. The differences between the monitor and indirect calorimetry decreased at higher intensities as indicated by lower measurement error. In another study, a different model from Garmin, the Forerunner 225 (GF225), overestimated EE and had a large magnitude of MAPE (31-155%) during a session that consisted of a resting measurement followed by 4-minute treadmill stages at 3 speeds and a 10-minute recovery period (Dooley et al., 2017). The possible source of discrepancy in the accuracy between the monitors may be grounded in that they utilize distinct firmware versions that integrate different regression equations to compute EE. In addition, the GVF can be paired with a HR strap and the GF225 has a build-in optical sensor to measure HR. Measuring activity HR may enhance EE estimation as HR data are incorporated in the proprietary calculation of active calories by the Garmin products (“Calorie Terminology”, n.d.). However, it appears that the HR monitor was not used in conjunction with the GVF, which could suggest why the activity monitor underestimated EE. Lastly, the differences in findings may be attributable to the sample size used in Dooley et al. (2017) as it was four times larger than the sample size of Price et al. (2016).

Another monitor, the Body Media Fit (BMF) is a commercially available monitor that has been derived from a research-grade SenseWear Armband monitor. The SWA is an accelerometer that integrates various heat- and galvanic-related variables and has been used for research. It provided valid and reliable measures of EE during a resistance training session targeting all major muscle groups (Reeve, Pumpa, & Ball, 2013) and during a level treadmill walking at two speeds (1.25 and 1.75 m•s^{-1}; Vernillo, Savoldelli, Pellegrini, & Schena, 2015). However, at high exercise intensities, the SWA underestimated EE during a circuit training session (Benito et al., 2012) and treadmill and outdoor running (Drenowatz & Eisenmann, 2011). It was suggested that addition of HR measurement may enhance the performance (Plasqui & Westeerterp, 2005). The upgraded version of the SWA, the BMF, integrates a HR-monitoring function among the other physiological sensors.

Studies have reported contradicting results for the BMF indicating both overestimation (Bai et al., 2016) and underestimation (Lee et al., 2014; Stackpool et al., 2014) of EE. The whole-session EE estimates by the monitor (351.0 ± 98.9 kcal) compared to a portable metabolic analyzer (316.8 ± 81.6 kcal) yielded a moderate MAPE of 15.3% (Bai et al., 2016), although in Lee et al. (2014), it produced a low MAPE of 9.3% for EE estimates (338.9 ± 59.4 kcal) compared to the reference method (356.9 ± 67.6 kcal). The EE estimates from the BMF fell within the equivalence zone and the monitor had the best overall results among all the monitors tested in the investigations (Bai et al., 2016; Lee et al., 2014). This opposes outcomes in Stackpool et al. (2014) who reported that the monitor's mean EE estimates (261.0 ± 52.4 kcal) significantly differed from the criterion measure EE (240.0 ± 47.3 kcal) during the 20-minute treadmill running bout. The main difference in the way the monitor accuracy was interpreted may be due to the statistical approaches used to evaluate its performance. Secondly, the varying sample size across the studies may again impact the results.

Lastly, the APW one of the most versatile activity monitors, has also been included in recent validation investigations. In the study by Dooley et al. (2017), the APW overestimated EE with MAPE 14.1%, 16.5%, and 20.0% for moderate-, vigorous-, and light-intensity treadmill activities, respectively. However, the monitor was highly inaccurate and overestimated EE with large error rates for the seated resting measurement (210%) and the recovery period (160%). Interesting findings showed an interaction between sex and device and BMI and device during the resting and recovery periods. The device overestimated EE for both sexes and all BMI categories, however, the magnitude of the effect was greater for males than females and greatest for participants who were overweight, followed by participants that were obese, and normal weight. Furthermore, an interaction during the moderate-intensity bout between BMI and the APW denoted that the monitor overestimated EE for participants who were overweight and obese, but not for participants who were of normal weight.

In contrast, Wallen et al. (2016) reported that the APW (162.6 ± 33.0 kcal) underestimated total EE compared to the criterion measure (285.7 ± 50.2 kcal) with a weak correlation (r = .16) across resting, treadmill, and cycling activities. The difference in the outcomes may be contributed to analyzing total EE estimates for the entire session (Wallen et al., 2016) as opposed to estimating EE for individual activities (Dooley et al., 2017). In addition, Dooley at al. (2017) included a large sample size (N = 62) of diverse participants in terms of sex, age, race, and fitness status, compared to a small (N = 22), more homogenous sample (Wallen et al., 2016), which possibly contributed to the differences in the findings.

Conclusions on the accuracy relative to EE

There is a large disparity in the literature regarding the accuracy of activity monitors in measuring EE. While some studies report that certain activity monitors are valid in estimating EE, other studies show contrasting results. The discrepancies are most likely attributable to differences in study design. For example, some studies included analyses of accuracy in estimating total EE over an entire session (Bai et al., 2016; Lee et al., 2014; Tucker et al., 2015; Wallen et al., 2016) while others analyze EE for individual activities (Diaz et al., 2015; Nelson et al., 2016; Price et al., 2016; Stackpool et al., 2014).

The sample size of different studies must also be taken into account. While some studies had larger sample sizes ($N \geq 50$; Bai et al., 2016; Dooley et al., 2017; Lee at al., 2014), others had sample sizes that was less than half in size (Noah et al., 2013; Price et al., 2016; Stackpool et al., 2014; Tucker et al., 2015; Wallen et al., 2016). The targeted population may also influence how accurately the monitors perform. Although most studies included relatively young, healthy male and female participants, Dooley et al. (2017) demonstrated that sex and BMI impacts monitor performance. To the author's knowledge, this was the only study to include an examination of the effects of demographic and anthropometric measures on estimation of EE.

Furthermore, the selected statistical analyses may influence how the accuracy of the monitors is interpreted. Using a traditional hypothesis approach with a focus on testing for a statistical difference between the criterion measure and the monitors may not offer the best insight on the accuracy of a monitor because this method does not imply equivalence between monitors (Hauck & Anderson, 1984). The study by Lee et al. (2014) was a pioneering investigation to introduce a novel statistical approach in activity monitors validation research called equivalence testing. Equivalence testing statistically examines measurement agreement between activity monitors and the criterion measure and, thus, makes it possible to determine if one method is "significantly equivalent" to another method. Whether a method is equivalent to another method is determined by specifying appropriate equivalence zone. While there is no standard, ±10% error zone has been used elsewhere (Bai et al., 2016, Lee et al., 2014; Stahl et al., 2016). Another common statistical approach used in validation studies is the Bland-Altman plot. It determines if existing systematic bias of the estimating method is present. Lastly, reporting MAPE, which provides an indicator of overall measurement error, or mean percent error rates has also been widely used. Standardized statistical analyses should be used to allow for comparison across studies and make interpretation of results practical.

Several other factors may impact the accuracy of a monitor. For instance, proper wear placement is not known for all devices studied in the literature and manufacturer instructions are not always clear. The manufacturer-specific instructions designate that the FBS should be positioned higher on the wrist and the BMF on the left upper arm, however, no concrete directions are provided for the remaining monitors. In addition, the EE calculation is determined by device-specific proprietary algorithms that are not disclosed. The manufacturers will only provide general information to the consumers about what variables are accounted for in the determination of EE. For example, the Garmin and Fitbit support pages list definitions related to caloric expenditure, disclosing that the type of activity, and the user's age, height, weight, and sex are included in the calculations. With the inclusion of HR monitoring function (either via ancillary HR strap or built-in PPG technology), the supporting monitors may also utilize HR to compute active calories. Therefore, errors in HR readings would consequently contribute to error in measuring EE.

Validity of Activity Monitors in Measuring HR

Heart rate monitoring is an easy and practical method for determining association between HR and health and fitness status (Blair, Goodyear, Gibbons, & Cooper, 1984; Wannamethee & Shaper, 1994) and assessing exercise intensity (Karvonen & Vuorimaa, 1988). With the integration of HR measurement via chest strap or PPG technology, activity monitors allow for personal monitoring of this health and fitness metric. While the inclusion of HR in proprietary algorithms calculating EE has enhanced EE predictions, the accuracy of PPG-based monitors in measuring HR has only been evaluated in a few studies. The validity of the monitors in measuring HR is achieved through comparison to electrocardiography (ECG; Parak & Korhonen, 2014; Wallen et al., 2016) or HR monitors (Dooley et al., 2017; Spierer et al., 2015; Stahl et al., 2016).

A common wrist-worn monitor for measuring HR is the MIA. Wallen et al. (2016) evaluated four wrist-worn monitors, including the MIA, in assessing HR during a 58-minute continuous routine that included supine, standing, and sitting rest periods, a treadmill routine consisting of the first three stages of the Bruce protocol, and a 25-watt step test on a stationary leg ergometer. The MIA exhibited a systematic bias and had a lower whole-session mean HR (97.7 ± 14.6 bmp) compared to the ECG (102.0 ± 13.4 bmp). The error across all the monitors ranged between 1-9%, however, the specific error for the MIA was not specified. The authors concluded that the monitors provided satisfactory measurement of HR, although the study included a relatively small sample size (*N* = 22) of young, healthy individuals. Similar procedures were used in another small study (*N* = 21) by Parak and Korhonen (2014), who compared the accuracy of the MIA to ECG during a 50-minute protocol that simulated activities such as resting in supine or seated positions, cycling (speed: 60 rpm and 90 rpm), walking (speed: 1.9 – 3.1 mph; grade: 0 – 10 %), and running (speed: 5.6 – 6.8 mph) on a treadmill. The monitor was within ±10 bpm from the ECG HR 86-87% of the time, but tended to underestimate

HR across all activities (mean error: 1.7%; MAPE: 5.2%). The monitor performed better during walking (mean error: 1.7%) and running (mean error: 1.9%) compared to cycling (mean error: 4.8%), however, it was the most accurate during the resting conditions (mean error: 0.5%).

The accuracy of the MIA was also evaluated in two larger investigations. Spierer et al. (2015) found the MIA to have different HR values while 50 participants performed 6-minute tasks including rest, walking, jogging, cycling, stair climbing, elliptical, and weight lifting. The monitor overestimated HR during the resting and walking tasks and underestimated HR for the other activities, but the difference was only significant during the weight lifting activity (error: 23.3 ± 31.9 bmp). Furthermore, when activity intensity increased, the measurement error also increased. This was specifically apparent during the running (5.3 mph) trial. In the other large study where the accuracy of the MIA to measure HR was assessed, 50 participants simultaneously wore six wrist-worn activity monitors during a routine consisting of resting, treadmill walking (speed: 1.9 – 4.0 mph), and treadmill running (5.0 – 6.0 mph) activities (Stahl et al., 2016). The mean HR measured by the MIA (110.5 ± 30.3 bpm) was shown to be equivalent to the Polar HR monitor (109.1 ± 29.3 bpm) determined by equivalence testing. The monitor's MAPE ranged between 0.8% to 16.0% across all activities, with the lowest error for the highest running speed (6.0 mph) and the highest error at the lowest walking speed (2.0 mph). These data contrast the findings by Spierer et al. (2015) who found that the performance of the MIA decreased with increasing speed. Collectively, the findings suggest varied accuracy of the MIA monitor in measuring HR. Some studies analyzed whole session HR data (Wallen et al., 2016) while others analyzed HR for individual activities (Parak & Korhonen, 2014; Stahl et al., 2016; Spierer et al., 2015) which could partially explain the discrepancies in the results.

As mentioned, other monitors with PPG technology have been examined for their accuracy. For example, Dooley et al. (2017), Wallen et al. (2016), and Stahl et al. (2016) examined the accuracy of two other monitors (the APW and the FBC) next to the previously mentioned MIA. In the study by Stahl et al. (2016), the FBC underestimated HR across all activities producing a whole-session HR (105.0 ± 30.6 bpm) that was lower than the ECG (109.1 ± 29.3 bmp) with an overall MAPE of 6.2%. However, the FBC was found to be equivalent to the criterion measure through the utilization of equivalence testing. Similarly, Wallen et al. (2016) reported that the FBC underestimated whole-session HR with mean difference of almost 10 bpm and produced the largest mean error of 9%. Slightly different findings reported by Dooley et al. (2017) showed that the FBC provided similar HR monitoring during moderate intensity (110.1 ± 16.7 bmp), but produced lower HR during vigorous intensity (144.0 ± 17.4 bpm) and higher HR values during light intensity (103.1 ± 17.5 bpm) compared to the Polar HR monitor (106.8 ± 16.4 bmp, 150.6 ± 21.3 bmp, 92.5 ± 13.7 bpm, respectively). The FBC produced a MAPE that ranged between 3.4% - 17.0%. In the same study, the APW measured comparable HR values to the reference method during rest and vigorous intensity work, but underestimated HR during light (89.2 ± 11.9 bpm vs. 92.5 ± 13.7 bpm) and moderate (101.0 ± 16.7 vs. 106.8 ± 16.4 bpm) intensities (Dooley et al., 2017). The APW had a low MAPE that ranged between 1.1 and 6.7% across activities. In addition, a small study including only 4 participants showed that the APW measured HR with 99.9% accuracy during walking for 200, 500, and 1,000 steps (El-Amrawy & Nounou, 2015).

Conclusions on the accuracy relative to HR

Similar to the large disparities in EE validity between activity monitors, discrepancies in HR validity across studies also exist. However, it is noteworthy, that monitors estimate HR more accurately than EE, as represented by a lower MAPE for HR measures. In spite of the accurate assessment of HR, there are a few proposed reasons why discrepancies exist in HR accuracy across studies. First, varying sample sizes may contribute to MAPE differences, as the number of participants fluctuated between 4 and 62. Secondly, the accuracy of the monitors with built-in PPG technology may be influenced by skin photosensitivity as melanin concentration and skin pigmentation can reduce the light wavelength emitted and thus attenuate the monitor's ability to detect pulse rate (Fallow et al., 2013). Indeed, Spierer et al. (2015) attributed the observed error to the variance in skin types. Although most studies did not assess skin photosensitivity, Dooley et al. (2017) included a diverse sample consisting of 47% non-White individuals. These data suggest that HR accuracy may have been impacted by the varying skin photosensitivity of the sample.

Additionally, the intensity of the activity has been proposed to also impact the accuracy of the HR measurement, as higher intensities, specifically during running, tend to disturb the skin to sensor interface introducing error (Spierer at al., 2015). While this is a valid concern, it conflicts with outcomes showing improved accuracy as intensity increased (Stahl et al., 2016; Parak & Korhonen, 2014). It is possible that, given complete and undisturbed contact between the monitor and skin, lower blood flow to the extremities during exercise at a lower intensity compared to a higher intensity may interfere with a monitor's ability to detect pulse. The improved perfusion at higher intensities could decrease the error rate.

Lastly, some discrepancies may be due to whole session HR data being analyzed in some studies (Wallen et al., 2016) while HR for individual activities being analyzed in other studies (Parak & Korhonen, 2014; Stahl et al., 2016; Spierer et al., 2015). As the validity of wrist-worn activity monitors in measuring HR was only examined in a few studies, more research in this area is warranted. Further, there is a scarcity of studies that have addressed certain factors (BMI, skin photosensitivity, or sex) that may influence the ability of PPG-based monitors to detect HR. Therefore, future research focusing on examining the impact of these factors is warranted.

CHAPTER III

VALIDATION OF ACTIVITY MONITORS IN ESTIMATING ENERGY EXPENDITURE, HEART RATE, AND STEPS IN LABORATORY CONTROLLED CONDITIONS

Introduction

Motion sensors, such as accelerometers and pedometers, are noninvasive, relatively inexpensive, and easy to use. These sensors have been applied in research to assess activity intensity, energy expenditure (EE), and steps (Bouten, Koekkoek, Verduin, Kodde, & Janssen, 1997; Colley & Tremblay, 2011; Welk, Schaben, & Morrow, 2004). Wearable technology has also become popular among market consumers, with one in 10adults in the United States owning an activity tracker to monitor PA and improve health ("The Demographics Divide", 2015, November 30).

Although the features vary, most commercially-available monitors can estimate EE, measure heart rate (HR), record steps, track active minutes, and/or monitor sleep patterns. Many of the monitors integrate multiple sensors (such as a tri-axial accelerometer, global positioning system (GPS), HR sensor, or skin temperature and galvanic responses gauge), which may improve estimation of EE (Brage et al., 2006; Evenson, Goto, & Furberg, 2015; Maddison & Ni-Muhurchu, 2009). In addition, some of the newer wrist-worn monitors include photoplethysmography (PPG) technology, an optical sensing technique that detects blood flow changes to determine HR (Allen, 2007).

Although these innovations may enhance the utility of activity monitors for personal and research purposes, their validity and accuracy remain to be determined. Most recently, commercial activity monitors were analyzed for step-measuring accuracy during treadmill ambulation under laboratory-controlled conditions (Chen, Kuo, Pellegrini, & Hsu, 2016; Diaz et al., 2015; Huang, Xu, Yu, & Shull, 2016) and during over-ground walking and running (Kirk, 2016; Storm, Heller, & Mazza, 2015). In general, hip-worn monitors provided more accurate estimates than wrist-worn monitors, however, the accuracy of the wrist-worn monitors improved as speed increased (Chen et al., 2016; Diaz et al., 2015; Huang et al., 2016). Additionally, the accuracy of the commercial monitors has been compared to established accelerometers, such as the SenseWear armband (SWA; Storm, Heller, & Mazza, 2015).

With respect to estimating EE, the validity of consumer-based monitors has been determined during free-living (Bai et al., 2016, Lee, Kim, & Welk, 2014; Tucker, Bhammar, Sawyer, Buman, & Gaesser, 2015) and laboratory-based (Dooley, Golaszewski, & Bartholomew, 2017; Nelson, Kaminsky, Dickin, & Montoye, 2016; Price et al., 2016) protocols. However, the validation studies disregard participants’ fitness status as the intensities investigated are usually based on a constant workload or an absolute intensity. This may be problematic because training status impacts HR response at a given intensity and as measurement of HR is often included in monitor-specific estimations of EE, disregarding training status can lead to error in the EE estimates. In addition, with the inclusion of PPG technology, users may rely on HR-monitoring to assess fitness status (Blair, Goodyear, Gibbons, & Cooper, 1984) and prescribe relative exercise intensity (Karvonen & Vuorimaa, 1988). Therefore, activity monitors with a HR function should be evaluated for accuracy in measuring HR and EE whilst participants exercise at relative intensities.

To date, there is a scarcity of validity research on the Fitbit Surge (FBS) and the Garmin Vívofit (GVF) and little is known about their accuracy in estimating EE, HR, and step count at specific exercise intensities. Therefore, the primary purpose of the current study was to examine the accuracy of three activity monitors (FBS, GVF, and SWA) in estimating EE in comparison to EE measured by a portable metabolic analyzer during treadmill walking, treadmill running, and stationary leg cycling at relative moderate and vigorous intensities in healthy, physically active individuals. The secondary objectives were to compare step count from the three activity monitors to a video observation and to assess HR-monitoring accuracy measured by the wrist-worn monitors (FBS and GVF).

Methods

Participants

Male (n = 23) and female (n = 11) participants completed three laboratory visits. All participants were physically active based on the Physical Activity Guidelines Advisory Committee Report (2008) and classified as either low or moderate risk based on American College of Sports Medicine (ACSM) cardiovascular risk assessment (ACSM, 2014). Moderate-risk participants obtained written medical clearance prior to participating in the study. All participants read and signed an informed consent form (see appendix A). The study was approved by the university Institutional Review Board (see Appendix B)

Instrumentation

Resting HR and blood pressure (BP) were obtained in a seated position after participants sat quietly for 5 minutes. To determine resting HR, participants were fitted with a Polar HR monitor (PM; Polar Electro, Kempele, Finland). Heart rate was recorded at the end of the 5 minutes of resting (Palatini et al., 2006). Next, resting BP was measured manually with a BP cuff and a stethoscope in a seated position with the participants' arm resting on a table at heart level (Pickering et al., 2005). Anthropometric and body composition measures were obtained in light clothing and with shoes removed. Standing height was determined with a wall-mounted stadiometer (Seca, Hamburg, Germany) to the nearest 0.1 cm. Body mass was measured with a digital scale (Sunbeam Products Inc., Health O Meter, Boca Ratoon, FL) to the nearest 0.1 kg. Participants' body composition was determined by a 3-site (males: chest, abdomen, and thigh; females: triceps, suprailiac, and thigh) skinfold procedure (Pollock, Schmidt, & Jackson, 1980) using a calibrated Harpenden skinfold caliper. Population-specific formulas (ACSM, 2014) were used to estimate percent body fat.

Activity monitoring instruments

Fitbit Surge (Fitbit Inc., San Francisco, CA). This wrist-worn activity monitor combines a 3-axis accelerometer and multiple sensors to track physical activity. The FBS measures pulse rate, estimates total and exercise EE, tracks daily steps, distance traveled, and floors climbed, and monitors sleep patterns. Continuous HR is detected by an optical sensor. When enabled, the monitor utilizes GPS to determine distance traveled, pace, and elevation. The FBS provides specific activity modes (e.g. treadmill running, free running, cycling, elliptical, spinning, hiking, weight training, or circuit training) to select from when tracking exercise. The rechargeable battery life of this device lasts up to 7 days, but when GPS is in use, the life can be reduced to 10 hours. Activity data can be viewed on a monitor-specific application by wirelessly syncing the device to a smartphone or computer.

Garmin Vívofit (Garmin Ltd, KS). The GVF is an accelerometer-based activity monitor worn on the wrist that estimates active and total EE, measures steps taken and distance traveled, and tracks sleep patterns. Heart rate-monitoring is available by pairing the GVF with an Ant$^+$ (Dynastream Innovations Inc., Canada) HR strap. The HR data are considered in the calculation of EE for fitness activities. The monitor measures active minutes and, when an individual has been inactive for an hour, a red bar appears on the display. The bar increases in length for each additional 15-minute block of inactivity. The monitor utilizes two replaceable coin cell batteries with a life of more than 1 year. Data recorded by the monitor can be either wirelessly synced with a smartphone or to a computer via a universal serial bus (USB) Ant$^+$ stick.

SenseWear Armband Mini (SWA; BodyMedia Inc., Pittsburgh, PA). The SWA is a multi-sensor activity monitor that measures steps and estimates EE by combining information from a 3-axis accelerometer with heat-related and galvanic skin response sensors. The monitor is worn on the upper left arm, midway between the olecranon and acromion processes. The SWA has a rechargeable battery that lasts approximately 7 days. The data are downloaded via a USB cord to a computer and accessed through monitor-specific software. Previously reported findings indicate the SWA to be valid in estimating

EE at rest and during level walking at two (1.25 and 1.75 $m \cdot s^{-1}$) speeds (Vernillo, Savoldelli, Pellegrini, & Schena, 2015).

Oxycon Mobile (OM; Carefusion Germany 234 GmbH, Hoechberg, Germany). The OM is a portable metabolic analyzer that was used as the criterion measure for EE measurements. The analyzer was calibrated before each session based on manufacturer's instructions. The OM includes two small portable units (a sensor and a receiver) that can be attached to a harness and strapped around a person's chest or back. Gas volumes are measured through a triple V™ turbine that connects to the sensor unit and a gas-collecting face mask. Gas is sampled via a sampling tube connected to the sensor unit and turbine. The OM can be paired with Polar technology to monitor HR. The measured parameters (such as the gas and air flow signals or HR) are telemetrically sent by the units to PC-based software (JLAB, Carefusion Germany 234 GmbH, Hoechberg, Germany). In a previous validation investigation, the OM measured comparable VO_2, carbon dioxide production (VCO_2), and ventilation to the Douglas bag method during different intensities (50 W, 100 W, 150 W, and 200 W) on a cycle ergometer (Rosdahl, Gullstrand, Salier-Eriksson, Johansson, & Schantz, 2010).

Digital video device (iPad, Apple Inc., Cupertino, CA). This video device was used as the reference method to compare actual steps and estimated step counts by the activity monitors. The device was positioned next to the treadmill to record participant's lower limbs during the 10-minute bouts of treadmill walking and running.

Procedures

Participants reported to the laboratory on three occasions, each separated by at least 48 hours. Participants were asked to abstain from exercise and alcohol and caffeine consumption 24 hours prior to each visit. During the first visit, participants completed a pre-participation health history questionnaire (ACSM, 2014) and a physical activity questionnaire (Global Physical Activity Questionnaire, World Health Organization, Geneva, Switzerland). Then, resting HR, BP, and anthropometric and body composition assessments were completed. The resting HR was used to calculate moderate (40% - < 60% of HR reserve) and vigorous (60-84% of HR reserve) intensity target HR zones (HRZ; Garber et al., 2011). Participants' demographic data (i.e. sex, age, and hand dominance) and smoking status were also collected and recorded during the visit.

Next, participants were fitted with a gas collection mask (Hans Rudolph Inc., V2 Mask, Shawnee, KS) and the OM for familiarization. After familiarization, participants performed an accommodation session of three, 10-minute ambulation bouts on a motorized treadmill (Fitnex, Dallas, TX) with 10-minute rest periods between each bout (Morgan, Martin, Krahenbuhl, & Baldini, 1990). For the initial 10-minute bout, the treadmill speed was maintained at 1.34 ms^{-1}. The second 10-minute bout was used to determine workloads for walking within the moderate HRZ and the third 10-minute bout was used to determine workloads for running within the vigorous HRZ. This was accomplished by having the participants walk at a starting speed of 1.34 ms^{-1} in the first two minutes of the second bout and by having the participants run at a starting speed of 1.8 ms^{-1} in the first two minutes of the third bout. Then, the workload was increased by asking participants if they wanted speed and/or grade increased until steady state HRs (defined as HR difference of less than 5 beats per minute; bpm, between the second and third minute at the specific speed and/or grade) at the low and high ends of each intensity HRZ were obtained. To monitor HR, participants wore the PM. The speeds and/or grades eliciting HR values in the appropriate intensity ranges were recorded to be used during visits 2 and 3.

After the treadmill accommodation, participants rested in a seated position until HR returned within 10 bpm of the baseline level. Participants then completed two, 10-minute familiarization periods (one for the moderate and one for the vigorous intensity) on a stationary leg ergometer (828E Monark Exercise AB, Verberg, Sweden). The pedaling frequency was kept constant at 70 rpm and the resistance (kilopond) was progressively increased based on participants' HR response until steady state HRs near the low and high ends of each intensity HRZ were achieved. The resistances that elicited the appropriate HR values were recorded and used during the next two visits.

Participants returned to the laboratory for visits 2 and 3 during the morning hours (6 am-10 am), after an overnight fast, to complete counterbalanced treadmill and cycling protocols. Upon arrival to the laboratory, participants' body mass was obtained. Next, all activity monitors were initialized using each participant's personal information (i.e. age, sex, body mass, height, smoking status, and hand dominance) and synchronized with a laptop computer. Then, the activity monitors, the Ant$^+$ HR monitor, the PM, and the OM were donned. The SWA was placed on the left triceps and the FBS and the GVF were worn on the right and left wrist in a counterbalanced order across participants and protocols. The PM monitor was placed around the chest at the level of the xyphoid process and the Ant^+ HR monitor was strapped directly below. Participants then completed the treadmill or cycling protocols.

The treadmill protocol consisted of counterbalanced 10-minute walking in the moderate HRZ and 10-minute running in the vigorous HRZ, each followed by a recovery period until participant's HR returned within 10 bpm of the baseline level. The treadmill was adjusted to the personalized workload based on the speed and/or grade recorded in the first visit. The investigator initiated all activity monitoring devices and then instructed the participant to step on the treadmill belt. Holding the treadmill handrails during ambulation was not permitted. Heart rate was continuously monitored and, if necessary, the investigator adjusted the speed and/or grade to ensure participants were exercising within the desired intensity range. At the end of the 10 minutes, participants were instructed to grasp the handrails and straddle the belt. Data recording by the activity monitors was immediately stopped.

For the cycling protocol, participants performed two, 10-minute bouts within the moderate and vigorous HRZs in a counterbalanced order. Each bout was also followed by a recovery period. To begin, participants pedaled at 70 rpm while the investigator adjusted the cycle ergometer to the desired resistance. Then, all activity monitors were activated. Participants were instructed to maintain pedaling frequency at 70 rpm throughout the 10 minutes. The PM constantly monitored HR and, if necessary, the resistance was adjusted to maintain HR within the desired intensity. At the end of the 10 minutes, participants stopped pedaling and the monitors were stopped.

Data processing

Criterion measures. Data acquired from the OM were processed using the manufacturer's software. The calculation of EE (kcal) from the OM is based on software specific algorithms considering urea nitrogen concentration and measured respiratory exchange ratio (RER) obtained by breath-by-breath VO_2 and VCO_2 analysis. Because urea nitrogen concentration was not attained in the study, a constant value of 15 g/day set by the program was used in the computation of EE. To allow for direct comparison between the activity monitors and the criterion measure, the OM minute-by-minute EE for each 10-minute bout was summed to obtain a total EE value. The HR measured by the PM was received and telemetrically transmitted by the OM receiver unit to the software and accessed at the end of the sessions. Minute-by-minute HR values were added and then divided by 10 to obtain average HR (HR_{avg}) for each activity bout. The highest HR reached during each activity bout (HR_{max}) was acquired by analyzing 5-second HR averages and recording the highest average value. The actual step count recorded by the digital video device was observed and recorded after completing the sessions. A clicker was used to count the observed steps. The running bouts were viewed in slow motion to avoid miscounting steps.

Activity monitors. To optimize activity recording by the FBS, the spinning, walking, and running settings were selected in respect to both cycling, moderate treadmill, and vigorous treadmill bouts. For the GVF, EE estimation was optimized by wearing the Ant^+ HR monitor and selecting the HR settings for all treadmill and cycling bouts. Upon completion of each activity bout, the GVF and FBS were wirelessly synced to a smartphone to access total EE estimates, step count, and average and maximal HR through monitor-specific mobile applications. However, the step count estimated by the GVF had to be determined by calculating the difference between step numbers noted from the monitor's display at the start and then again at the end of the treadmill bouts. The data obtained by the SWA were downloaded via a USB cable to a computer-based program and processed at the end of each session. The minute-by-minute EE and minute-by-minute step count data were combined to provide 10-minute estimates for each bout.

Statistical analyses

Descriptive characteristics were calculated for all study variables and reported as means ± standard deviations. Pearson correlations were calculated to assess overall group-level associations between the criterion measures and the activity monitors. Root mean square error (RMSE) was calculated to indicate overall measurement error. To examine measurement agreement for all variables between the criterion measures and the activity monitors, 95% equivalency testing analyses were performed. The estimates by the monitors were considered equivalent to the criterion-measured values if the 90% confidence intervals (CI) of the mean estimates fell within ±10% of the mean measured values. This approach has been used elsewhere (Bai et al., 2016; Lee, Kim, & Welk, 2014). Bland-Altman plots (Bland & Altman, 1999) with the corresponding 95% limits of agreement and a regression line were used to graphically represent individual variations and to examine systematic bias in the EE estimates. The estimated and measured means ± standard deviations and the 90% CI were obtained using the IBM SPSS statistical software package, version 24 (IBM Corp., Armonk, NY).

Results

Participants (*N* = 34, age = 25.8 ± 4.9 years, height = 171.7 ± 8.4 cm, body mass = 76.6 ± 18.8 kg, percent body fat = 15.5 ± 7.8%) completed two visits during which they performed moderate and vigorous intensity treadmill and cycling activities in a counterbalanced order. During the data collection process, some measuring device data were lost due to technical issues. Furthermore, one female participant did not complete the treadmill session due to illness, resulting in varying sample sizes for the measured variables across activity bouts. The varying sample size can be found in Tables 1, 2, and 3 for EE, HR_{avg} and HR_{max}, and step count, respectively.

As shown in Table 1, the activity monitors produced lower EE values compared to the OM across all activity bouts. The measurement error increased at higher intensities as signified by higher RMSE for vigorous intensity bouts compared to moderate intensity bouts. The activity monitors tended to underestimate HR_{avg} but yielded comparable HR_{max} values to the PM (Table 2). Finally, the activity monitors underestimated actual step count for both treadmill bouts (Table 3).

When assessing the EE correlation matrix between activity monitors and the OM, the monitors had the strongest relationship with the OM during the vigorous intensity (Table 4) treadmill bout, with the SWA having the strongest correlation, followed by the FBS, and the GVF. Strong correlation between the monitors and the OM was still found for the moderate intensity treadmill bout. When assessing the HR correlation matrix, the

Table 1

Descriptive Statistics and Root Mean Square Error for Energy Expenditure Obtained by the Oxycon Mobile and Activity Monitors

Mode / Intensity	*n*	EE (kcal)	RMSE
Cycling / Moderate			
OM	34	88.2 ± 16.5	-
FBS	34	63.0 ± 19.2	30.6
GVF	34	75.3 ± 17.4	19.1
SWA	32	44.6 ± 18.6	41.0
Cycling / Vigorous			
OM	33	114.7 ± 23.7	-
FBS	33	85.0 ± 26.2	43.1
GVF	32	113.2 ± 16.3	23.3
SWA	32	52.8 ± 22.6	67.3
Treadmill / Moderate			
OM	33	88.5 ± 20.9	-
FBS	33	82.7 ± 16.2	15.7
GVF	33	76.2 ± 17.2	21.2
SWA	32	65.1 ± 12.8	28.1
Treadmill / Vigorous			
OM	33	144.3 ± 31.6	-
FBS	33	123.8 ± 23.1	26.1
GVF	32	124.7 ± 20.6	27.7
SWA	32	123.0 ± 22.7	61.5

Note. EE = Energy expenditure; FBS = Fitbit Surge; GVF = Garmin Vívofit; OM = Oxycon Mobile; RMSE = Root mean square error; SWA = SenseWear Armband.

Table 2

Descriptive Statistics and Root Mean Square Error for Average and Maximal Heart Rate for the Polar and Activity Monitors

Mode / Intensity	*n*	Average HR (bpm)	RMSE	*n*	Maximal HR (bpm)	RMSE
Cycling / Moderate						
PM	34	122.3 ± 5.6	-	34	130.8 ± 6.5	-
FBS	34	109.3 ± 14.8	20.2	34	129.1 ± 8.3	5.8
GVF	32	120.1 ± 6.0	2.4	33	130.2 ± 6.9	1.2
Cycling / Vigorous						
PM	34	151.2 ± 7.0	-	34	160.0 ± 6.4	-
FBS	34	130.9 ± 19.2	28.8	34	159.5 ± 6.0	2.7
GVF	32	147.0 ± 7.2	5.0	33	159.8 ± 6.6	0.7
Treadmill / Moderate						
PM	33	123.1 ± 5.6	-	32	131.7 ± 6.4	-
FBS	33	111.9 ± 10.8	14.1	33	128.2 ± 7.9	6.7
GVF	30	117.0 ± 12.6	13.7	31	129.8 ± 11.1	8.9
Treadmill / Vigorous						
PM	32	159.3 ± 8.0	-	32	169.6 ± 7.6	-
FBS	33	148.3 ± 7.1	11.8	33	168.2 ± 9.0	5.4
GVF	30	152.0 ± 9.5	9.2	29	169.7 ± 8.0	0.9

Note. bpm = Beats per minute; FBS = Fitbit Surge; GVF = Garmin Vívofit; HR = Heart rate; PM = Polar monitor; RMSE = Root mean square error.

Table 3

Descriptive Statistics and Root Mean Square Error for Steps during the Treadmill Bouts

Intensity	*n*	Steps	RMSE
Moderate			
Video	32	1220.5 ± 53.6	-
FBS	33	1101.39 ± 123.9	175.1
GVF	32	1123.2 ± 148.0	185.1
SWA	32	1162.8 ± 61.8	71.6
Vigorous			
Video	30	1654.0 ± 110.1	-
FBS	33	1607.6 ± 106.3	71.7
GVF	32	1647.8 ± 109.8	25.1
SWA	32	1584.1 ± 70.5	102.9

Note. *N = 34.* FBS = Fitbit Surge; GVF = Garmin Vívofit; RMSE = Root mean square error; SWA = SenseWear Armband.

Table 4

Pearson's Correlation for Energy Expenditure between the Oxycon Mobile and Activity Monitors

Mode / Intensity	FBS	GVF	SWA
Cycling / Moderate			
OM	.09	.43*	.52**
FBS	1	.33	.17
GVF		1	.16
Cycling / Vigorous			
OM	.24	.34	.37*
FBS	1	.34	.18
GVF		1	.32
Treadmill/ Moderate			
OM	.71**	.59**	.63**
FBS	1	.71**	.60**
GVF		1	.37*
Treadmill / Vigorous			
OM	.74**	.66**	.88**
FBS	1	.77**	.85**
GVF		1	.78**

Note. FBS = Fitbit Surge; GVF = Garmin Vívofit; OM = Oxycon Mobile; SWA = SenseWear Armband. *p < .05, **p < .01.

GVF produced the strongest relationship with the PM. Strong correlations for HR_{avg} between the GVF and the PM were found for the moderate and vigorous intensity cycling bouts. Similarly, the GVF was strongly correlated with the PM for session HR_{max} during the moderate intensity cycling, vigorous intensity cycling, and vigorous intensity treadmill bouts (Table 5). Pearson's correlation analysis for step count revealed that steps estimated by the GVF strongly correlated with the video observation for the vigorous intensity bout, but had a weak correlation with the criterion measure for the moderate intensity bout (Table 6.). A similar pattern was found for the FBS, which was strongly correlated with the video observation for the vigorous intensity bout, but had a weak relationship with the criterion measure for the moderate intensity bout. The SWA had a strong correlation with the video observation for both intensity bouts.

The equivalency testing analyses for agreement in EE (Figure 1) indicated that the activity monitors were not equivalent to the OM for any activity bout except for the GVF during the vigorous intensity cycling bout, indicated by the monitor's 90% CI (lower limit: 108.3; upper limit: 118.1 kcal) within the ±10% of the OM measured mean EE (103.7; 126.7 kcal). During the moderate intensity treadmill bout, the FBS 90% CI interval was nearly entirely contained within the equivalency zone. The lower limit (77.9 kcal) missed the lower bound (79.7 kcal) of the equivalency zone by only 1.8 kcal (Figure 1c). Similar pattern was seen for the GVF during the moderate intensity cycling bout, where the lower limit of the 90% CI (70.0 kcal) missed the lower bound (74.0 kcal) of the equivalency zone by 4 kcal (Figure 1a). For the HR equivalency analysis, estimates from the FBS and GVF were considered equivalent to the PM, with the exception of

Table 5

Pearson's Correlation for Average and Maximal Heart Rate between the Polar and Activity Monitors

	Average HR		Maximal HR	
Mode / Intensity	FBS	GVF	FBS	GVF
Cycling / Moderate				
PM	.20	.97**	.73**	.99**
FBS	1	.20	1	.84**
GVF		1		1
Cycling / Vigorous				
PM	-.30	.96**	.91**	.99**
FBS	1	-.22	1	.91**
GVF		1		1
Treadmill/ Moderate				
PM	.58**	.28	.69**	.59**
FBS	1	.08	1	.29
GVF		1		1
Treadmill / Vigorous				
PM	.73**	.83**	.54**	.99**
FBS	1	.62**	1	.72**
GVF		1		1

Note. FBS = Fitbit Surge; GVF = Garmin Vívofit; HR = Heart rate; PM = Polar monitor. $^{*}p < .05$, $^{**}p < .01$.

Table 6

Pearson's Correlation for Steps between the Activity Monitors and Video Observation during the Treadmill Bouts

Intensity	FBS	GVF	SWA
Moderate			
Video	.08	-.05	.73**
FBS	1	.49**	.13
GVF		1	-.06
SWA			1
Vigorous			
Video	.84**	.98**	.64**
FBS	1	.83**	.56**
GVF		1	.61**
SWA			1

Note. FBS = Fitbit Surge; GVF = Garmin Vívofit; SWA = SenseWear Armband. ***p* < .01.

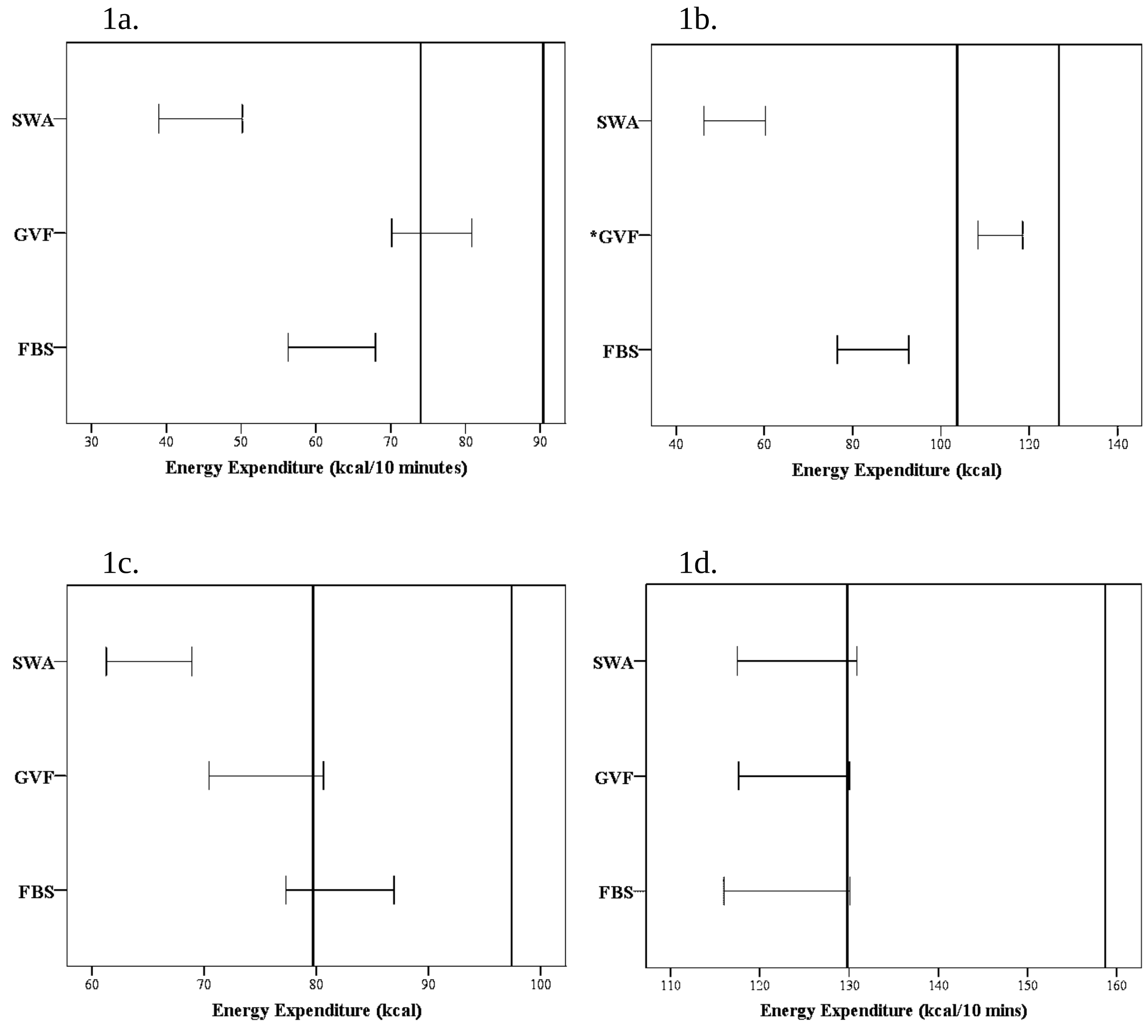


Figure 1. Equivalency testing for energy expenditure agreement between the Oxycon Mobile and activity monitors. 1a. The moderate intensity cycling bout. 1b. The vigorous intensity cycling bout. 1c. The moderate intensity treadmill bout.1d. The vigorous intensity treadmill bout. Vertical lines represent the ±10% equivalency zone of the mean measured energy expenditure by the Oxycon Mobile. *Within the equivalency zone. FBS = Fitbit Surge; GVF = Garmin Vívofit; OM = Oxycon mobile; SWA = SenseWear Armband.

HR_{avg} obtained by the FBS during the moderate intensity treadmill bout (Table 7). The monitor’s lower level of the 90% CI missed the lower bound of the equivalency zone bound by 2 bmp. The step count equivalency analysis showed that only the SWA produced step count estimates equivalent to the video observation during the moderate intensity treadmill bout. All three activity monitors were equivalent to the video observation during the vigorous intensity treadmill bout (Table 8).

The distribution of individual error and systematic bias in the EE estimates were evaluated by Bland-Altman plots (Figures 2, 3, 4, and 5). The narrowest 95% limits of agreement during the moderate intensity cycling bout (Figure 2) were found for the SWA (difference = 68.3 kcal; Figure 2c), followed by the GVF (difference = 70.9 kcal; Figure 2b), and the FBS (difference = 94.9 kcal; Figure 2a). The narrowest limits of agreement during the vigorous intensity cycling bout (Figure 3) were produced by the GVF (difference = 91.8 kcal; Figure 3b), followed by the SWA (difference = 104.2 kcal; Figure 3c), and the FBS (difference = 122.7 kcal; Figure 3a). During this bout, a significant slope for the regression analysis was found for the GVF. For the moderate intensity treadmill bout (Figure 4), the FBS had the narrowest limits of agreement (difference = 58.0 kcal; Figure 4a), followed by the SWA (difference = 64.3 kcal; Figure 4c), and the GVF (difference = 68.6 kcal; Figure 4b). Significant slope was found for the SWA. Lastly, during the vigorous intensity treadmill bout (Figure 5), the narrowest limits of agreement were found for the SWA (difference = 62.7 kcal; Figure 5c), followed by the FBS (difference = 83.5 kcal; Figure 5a), and the GVF (difference = 93.1 kcal; Figure 5b). Significant slope from the regression analysis was found for the FBS and SWA.

Table 7

Equivalency Testing for Average and Maximal Heart Rate between the Polar and Activity Monitors

			90% CI (bpm)		
Variable	Mode / Intensity	-10% (bpm)	Lower	Upper	+10% (bpm)
HR_{avg}	Cycling / Moderate				
	FBS*	101.1	105.0	113.6	134.5
	GVF*	101.1	118.3	121.9	134.5
	Cycling / Vigorous				
	FBS*	136.1	125.3	136.5	166.4
	GVF*	136.1	144.8	149.1	166.4
	Treadmill / Moderate				
	FBS	110.8	108.8	115.1	135.4
	GVF*	110.8	113.1	120.9	135.4
	Treadmill / Vigorous				
	FBS*	143.4	146.2	150.4	175.2
	GVF*	143.4	149.1	155.0	175.2
HR_{max}	Cycling / Moderate				
	FBS*	117.7	126.7	131.5	143.9
	GVF*	117.7	128.1	132.2	143.9
	Cycling / Vigorous				
	FBS*	144.0	157.7	161.2	176.0
	GVF*	144.0	157.9	161.8	176.0
	Treadmill / Moderate				
	FBS*	118.5	125.9	130.5	144.9
	GVF*	118.5	126.4	133.2	144.9
	Treadmill / Vigorous				
	FBS*	152.6	165.5	170.8	186.5
	GVF*	152.6	167.1	172.2	186.5

Note. bpm = Beats per minute; CI = Confidence interval; FBS = Fitbit Surge; GVF = Garmin Vívofit; HR_{avg} = Average heart rate; HRmax= Maximal heart rate. *Within the equivalency zone.

Table 8

Equivalency Testing for Steps between the Activity Monitors and Video for the Treadmill Bouts

		90% CI		
Intensity	-10%	Lower	Upper	+10%
Moderate				
FBS	1098.5	1064.9	1137.9	1342.6
GVF	1098.5	1078.8	1167.6	1342.6
SWA*	1098.5	1144.3	1181.4	1342.6
Vigorous				
FBS*	1488.6	1576.2	1638.9	1819.4
GVF*	1488.6	1614.9	1680.7	1819.4
SWA*	1488.6	1563.0	1605.2	1819.4

Note. CI = Confidence interval; FBS = Fitbit Surge; GVF = Garmin Vívofit; SWA = SenseWear armband. *Within the equivalency zone.

2a. 2b.

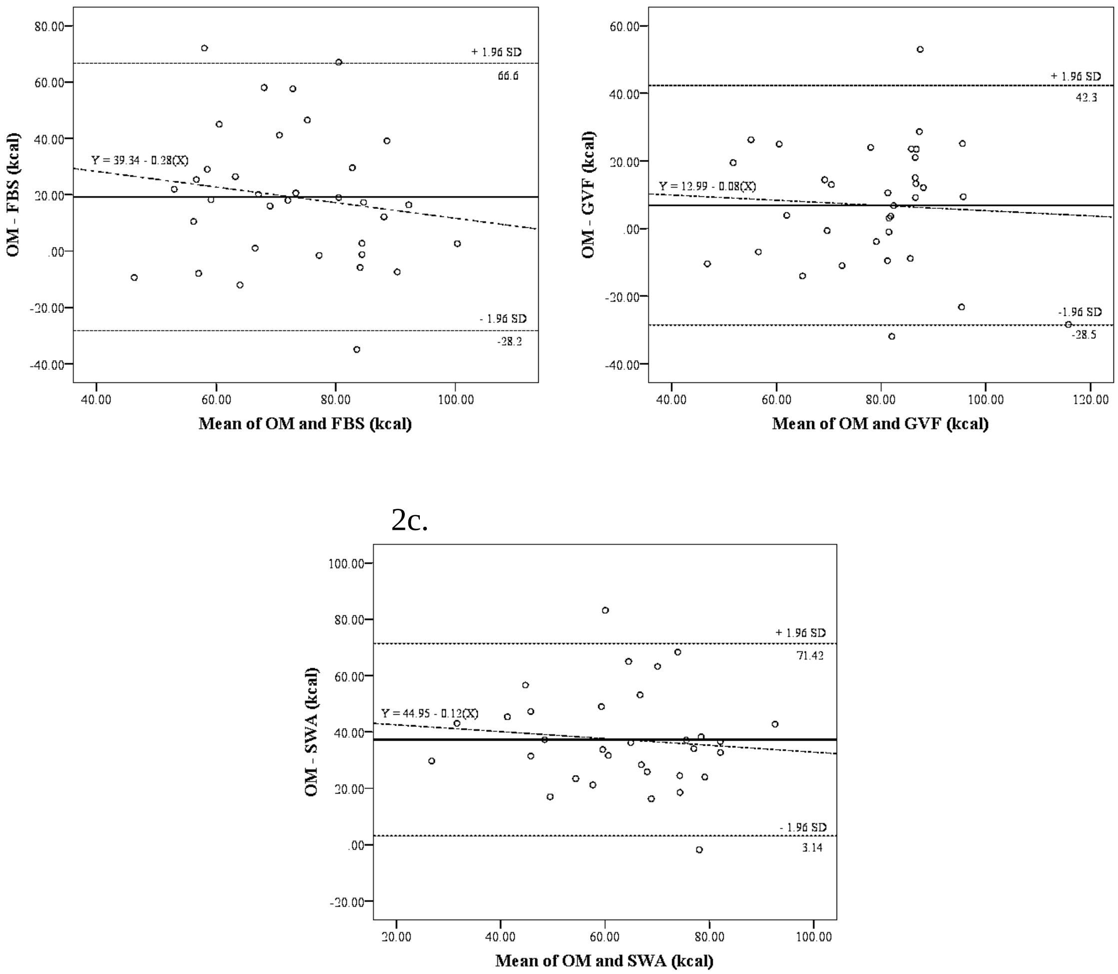
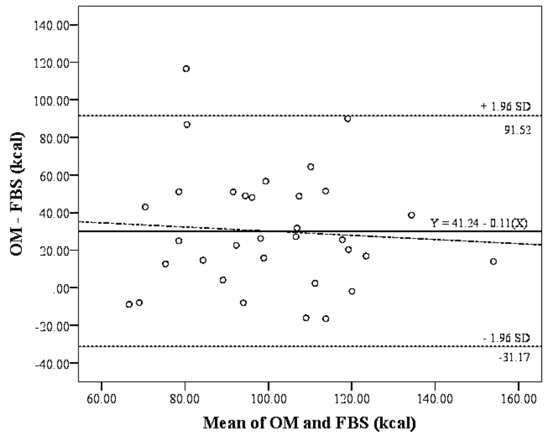
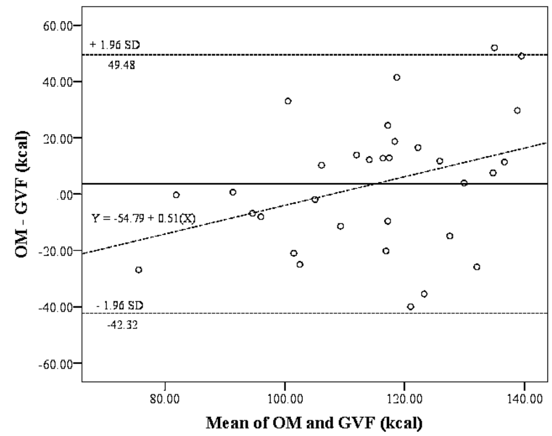


Figure 2. Bland-Altman plots for differences in energy expenditure between the Oxycon Mobile and activity monitors for the moderate intensity cycling bout. 2a. Individual errors between the OM and the FBS. 2b. Individual errors between the OM and the GVF. 2c. Individual errors between the OM and the SWA. FBS = Fitbit Surge; GVF = Garmin Vívofit; OM = Oxycon Mobile; SWA = SenseWear Armband.

3a.



3b.



3c.

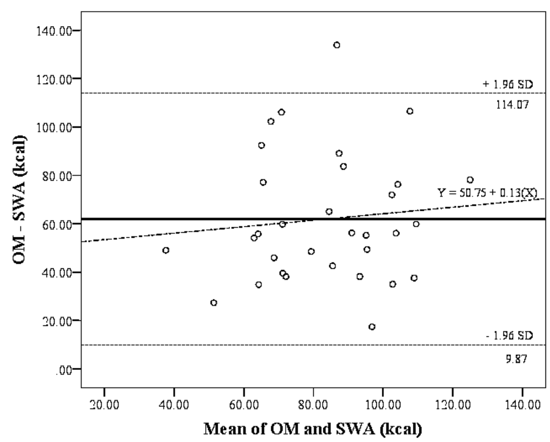
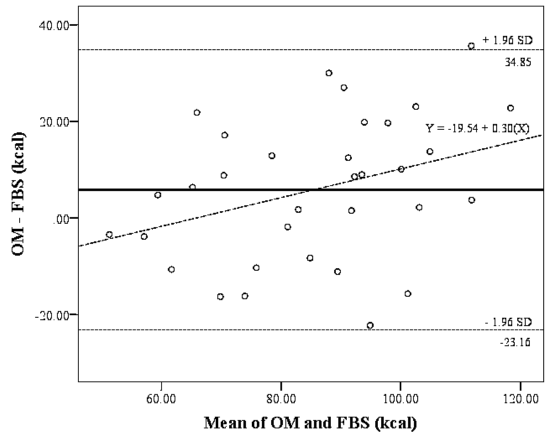
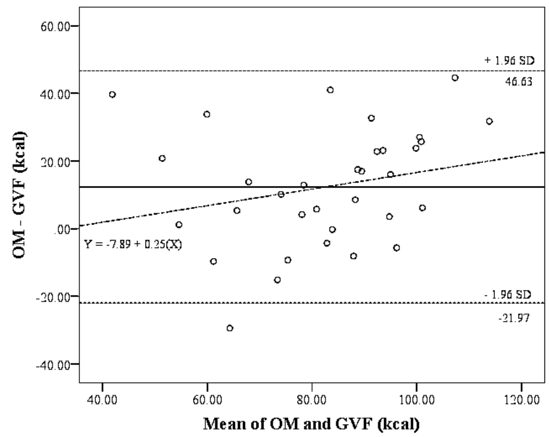


Figure 3. Bland-Altman plots for differences in energy expenditure between the Oxycon Mobile and activity monitors for the vigorous intensity cycling bout. 3a. Individual errors between the OM and the FBS. 3b. Individual errors between the OM and the GVF. 3c. Individual errors between the OM and the SWA. FBS = Fitbit Surge; GVF = Garmin Vívofit; OM = Oxycon Mobile; SWA = SenseWear Armband.

4a.



4b.



4c.

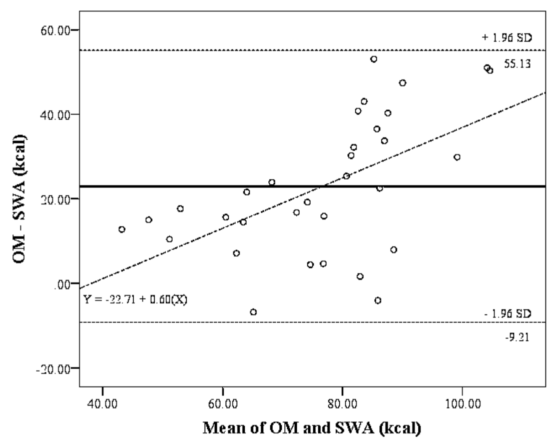


Figure 4. Bland-Altman plots for differences in energy expenditure between the Oxycon Mobile and activity monitors for the moderate intensity treadmill bout. 4a. Individual errors between the OM and the FBS. 4b. Individual errors between the OM and the GVF. 4c. Individual errors between the OM and the SWA. FBS = Fitbit Surge; GVF = Garmin Vívofit; OM = Oxycon Mobile; SWA = SenseWear Armband.

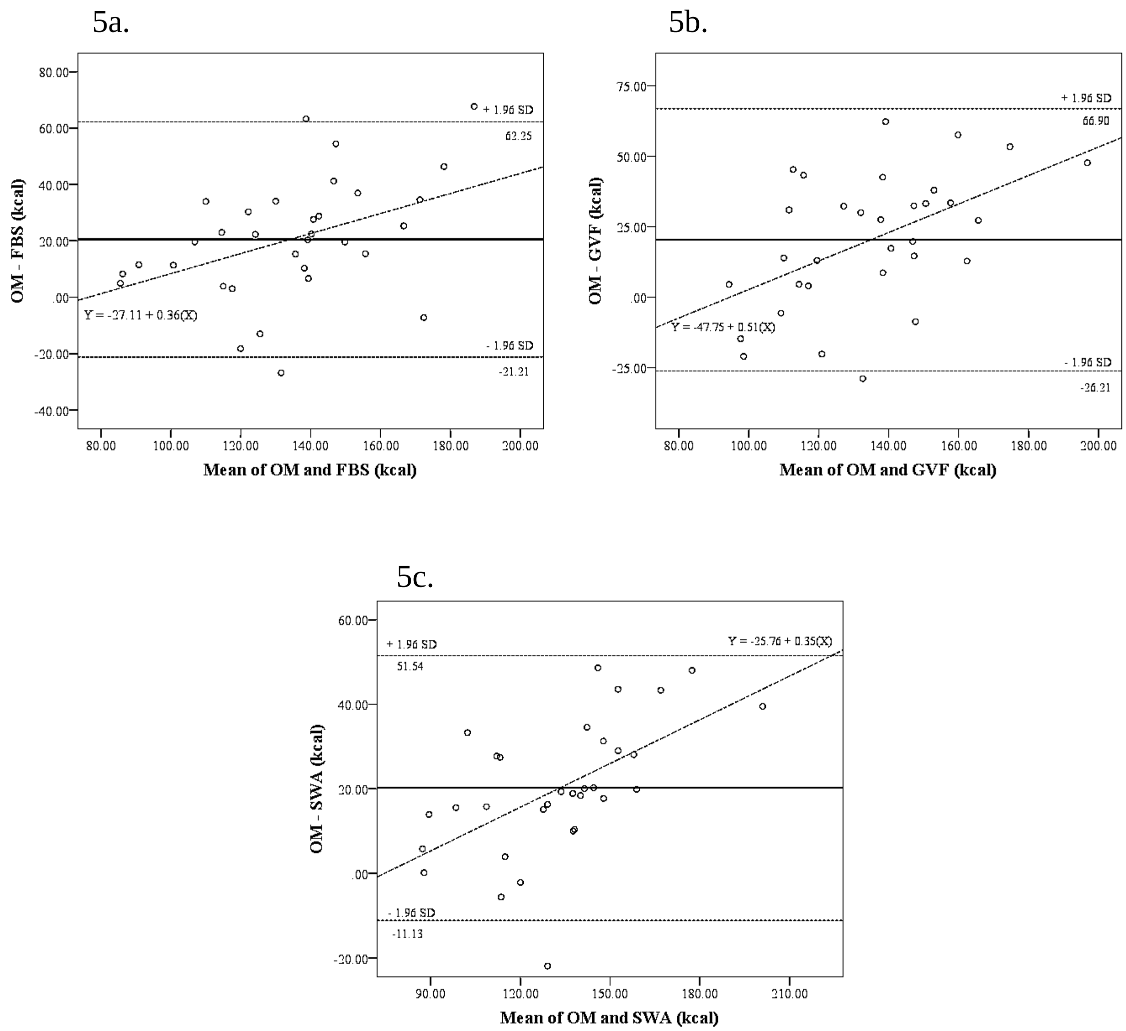


Figure 5. Bland-Altman plots for differences in energy expenditure between the Oxycon Mobile and activity monitors for the vigorous intensity treadmill bout. 5a. Individual errors between the OM and the FBS. 5b. Individual errors between the OM and the GVF. 5c. Individual errors between the OM and the SWA. FBS = Fitbit Surge; GVF = Garmin Vívofit; OM = Oxycon Mobile; SWA = SenseWear Armband.

Discussion

The present study evaluated the accuracy of activity monitors including the FBS, GVF, and SWA in estimating EE, HR_{avg}, HR_{max}, and steps during stationary cycling and treadmill activities at two relative intensities in healthy, physically active adults. The unique feature of the current study was that the accuracy of the activity monitors was assessed relative to each individual's HRZs. In addition, evaluating accuracy by relative HR intensity may offer a more practical application of the results. In general, the activity monitors yielded lower EE compared to EE measured by the OM. The FBS and the GVF performed favorably in measuring HR, producing comparable HR_{avg} and HR_{max} to the PM. The step analysis also revealed promising accuracy for the activity monitors when compared to the actual step count obtained via video observation.

The EE analyses demonstrate that the activity monitors underestimate EE across all activity bouts with increased error at higher intensities (Table 1). Some monitors however, had more promising accuracy compared with the OM. For instance, the FBS was highly correlated (r = .77) with the OM, produced the lowest RMSE (15.7), and showed no systematic bias of individual errors with the narrowest limits of agreement (Figure 4a) during the moderate intensity treadmill bout. While the results from equivalency testing indicate that the FBS did not yield significantly equivalent EE estimates to the criterion measure (Figure 1c), it should be noted that the monitor's 90% CI missed the lower bound of the equivalency zone by only 1.8 kcal. The GVF produced favorable results compared to the OM during the moderate intensity cycling bout. The GVF was moderately correlated with the OM (r = .43), yielded relatively low error

(RMSE = 19.1), and showed no systematic bias of individual errors (Figure 2b). Similar to the FBS, the equivalency testing suggests that GVF did not yield equivalent estimates to the measured EE (Figure 1b), with the monitor's 90% CI missed the lower bound of the equivalency zone by 4 kcal.

The heterogeneity of protocols in previous research poses challenges for comparison of findings across studies. In addition, some activity monitors have been utilized in validation research more than others. Only a few studies have examined the validity of the GVF in estimating EE, concluding the monitor underestimates EE during treadmill ambulation (Alsubheen, George, Baker, Rohr, & Basset, 2016; Price et al., 2016). In the study by Price and colleagues (2016), the GVF underestimated EE with a systematic bias of 1.67 ± 1.21 kcal·min^{-1} across treadmill walking (0.70-1.80 ms^{-1}) and with a systematic bias of 1.45 ± 3.10 kcal·min^{-1} across treadmill running (2.22-3.33 ms^{-1}) speeds. A similar pattern for the GVF to systematically underestimate EE during the vigorous intensity (running) treadmill bout was observed in the present study (Figure 5b). Likewise, Alsubheen et al. (2016) reported that the GVF underestimated EE on average by 29.5% compared to indirect calorimetry. However, the differences between the monitor and indirect calorimetry decreased at higher intensities, which is in contrast to the results of the present study. In general, the present investigation and the study by Alsubheen et al. (2016) and Price et al. (2016) collectively found a trend for the GVF to underestimate EE during treadmill ambulation. In addition, the current study included evaluation of the GVF in estimating EE during stationary cycling. To the author's knowledge, no other study has done so. Interpretation of the monitor's results is challenging due to apparently contradicting findings. For instance, while during the vigorous intensity cycling bout, the GVF yielded equivalent EE based on the monitor's 90% CI within the equivalency zone, it did not correlate with the OM (r = .34), and systematically underestimated EE based on Bland-Altman analysis (Figure 3b). It is difficult to resolve how the monitor can produce equivalent group-level estimates and yet, not be correlated with the OM.

Evaluation of the FBS accuracy revealed that the monitor performed poorly for the cycling bouts with increased error during higher intensity bouts (Table 1). While there was no apparent systematic bias for the FBS during the cycling bouts, the monitor did not correlate with the OM (Table 4) and the equivalency testing showed the monitor not to be equivalent to the OM (Figures 1a-1b). The monitor performed better during the moderate intensity treadmill bout, but during the vigorous intensity treadmill bout, it did not produce equivalent EE estimates to the OM (Figure 1c), with systematic EE underestimation (Figure 5a) and relatively high RMSE (26.1). These findings are similar to results of another study, where the FBS underestimated EE with a mean bias of 25.0 ± 16.0 kcal and produced 90% CI outside of the equivalency zone during structured periods of walking and running (0.9-2.7 ms^{-1}) on a treadmill (Massey, Funk, Thiebaud, & Patton, 2017). It is interesting to note that older versions of wrist-worn Fitbit monitors, such as the Fitibit Flex or Fitbit Charge, have been reported to overestimate treadmill walking and running EE (Bai et al., 2016; Diaz et al., 2015; Dooley et al, 2017; Nelson et al., 2016). Although EE calculation is based on proprietary algorithms, the inclusion of PPG with the FBS may contribute to the discrepancy in the accuracy across models.

The accuracy of the SWA was also concurrently examined in this study. Based on the results from equivalency testing, the SWA was not equivalent to the OM for any of the activity bouts (Figures 1a-1d), yielding lower EE estimates compared to the OM with relatively high RMSE. As seen in Table 1, the measurement error increased at higher intensities. These findings are in accordance with previous research demonstrating that the SWA underestimated EE during treadmill running (Drenowatz & Eisenmann, 2011; Koehler, deMarees, Braun, & Schaenzer, 2010) and stationary cycling at two different (150 W and 450 W) workloads (Jakicic et al., 2004). Drenowatz and Eisenmann (2011) found that the measurement error increased across intensity levels (65, 70, and 85% of maximal VO_2) and concluded that the larger measurement error at higher intensities was due to a ceiling effect in EE measurement by the monitor. In contrast to these and current findings, an overestimation of EE by the SWA at exercise intensities up to 3.6 ms^{-1} (8 mph) has been reported (King, Torres, Potter, Brooks, & Coleman, 2004). The difference in the results may be attributed to the outdated monitor algorithm used in the King et al. (2004) study.

The monitors seem to be more accurate in measuring HR_{avg} and HR_{max} compared to their accuracy in estimating EE. The results of equivalency testing show that the FBS and GVF measured HR_{avg} and HR_{max} that were equivalent to the PM across all activity bouts with the exception of the FBS during the moderate intensity treadmill bout (Table 5). However, the monitor's 90% CI missed the lower bound of the equivalency zone bound only by 2 bmp. The RMSE (Table 1) suggests that the monitors performed better in detecting HR_{max} than HR_{avg} indicated by lower measurement errors. For example, during the vigorous intensity cycling bout, the FBS produced relatively high RMSE for HR_{avg} (28.8), but yielded considerably lower RMSE for HR_{max} (2.7). The same trend was seen for both wrist-worn monitors across all activity bouts.

In general, the GVF had lower measurement error for both HR variables compared to the FBS. The implementation of a HR strap (GVF) as opposed to PPG technology (FBS) to obtain HR measurements may contribute to the improved accuracy of the GVF. Existing literature evaluating the accuracy of various consumer-based monitors utilizing PPG technology demonstrates poor accuracy (Dooley et al., 2017; Wallen, Gomersall, Keating, Wisløff, & Coombes, 2016) and the tendency to underestimate HR during a variety of activities (Stahl, An, Dinkel, Noble, & Lee, 2016; Wallen et al., 2016). The Fitbit Charge was found to underestimated HR (105.0 ± 30.6 bpm) compared to the criterion measure (109.1 ± 29.3 bmp) across treadmill walking (0.8-1.8 ms^{-1}) and running (2.2- 2.7 ms^{-1}) speeds. However, the monitor was found to be equivalent to the criterion measure based on equivalency testing. Although a different model of the Fitbit monitor was used, these findings are similar to those of the current study where the FBS produced lower HR_{avg} across activities, but was found to be equivalent to the criterion measure with the exception of the moderate intensity treadmill bout.

Caution should, however, be taken when interpreting the results for HR accuracy. While the GVF and FBS produced relatively accurate HR measurements based on equivalency testing, the monitors did not correlate with the PM (Table 5) and produced relatively high RMSE (Table 2) for some activity bouts. This is specifically true for the

FBS during the cycling bouts. The reason for this occurrence can be potentially explained by sampling error of the monitor during the cycling activity. Although this was not a part of the study purpose, an intermittent HR detection by the monitor was observed during data collection. While it is not clear what caused the HR signal to be lost for some participants, it may be that flexion of the wrist interrupted the HR-measuring ability of the monitor. Another possible explanation is that compared to treadmill exercise, lower blood flow to the upper extremities during leg ergometry may interfere with the monitor's ability to detect pulse. This may explain the improved measurement error and correlation with the PM during the treadmill bouts. Interestingly, during the moderate intensity treadmill bout, the GVF did not correlate with the PM (r = .28), producing the highest RMSE (13.7). The second highest RMSE (11.8) for this monitor occurred during the vigorous intensity treadmill bout. In previous research, decreased correlation between HR monitors and ECG during treadmill activities was attributed to upper body movement (Lee & Gorelick, 2011; Montgomery et al., 2009), which may elucidate the lower correlation and accuracy for the treadmill bouts in the present study.

The results of the step count analysis indicated that the monitors tended to underestimate actual step count as indicated by lower mean values compared to the video observation (Table 3). However, some of the activity monitors displayed a promising ability to measure step count. The SWA showed good measurement agreement with the video observation based on equivalency analysis (Table 8) and high correlation with the actual step count (Table 5) for both intensity bouts. Improved accuracy of the monitor was detected for the moderate intensity (walking) based on smaller RMSE in comparison to vigorous intensity (running). In previous work, the SWA was found to underestimate steps during indoor and outdoor walking, with higher underestimation at slow compared to fast speeds (Storm, Heller, & Mazzà, 2015). The ambulation activities and statistical analyses differed between the current study and the study by Storm et al. (2015), therefore, careful consideration should be given to the interpretation of findings.

As for the wrist-worn monitors, the GVF and FBS produced equivalent step count with strong correlation with the criterion measure only during the vigorous intensity bout. The measurement errors for the GVF and FBS were also lower at the higher intensity treadmill bout. The findings for the accuracy of the GVF in measuring steps is in agreement with previous work reporting that the monitor tended to underestimate actual steps, but the accuracy increased at the fastest treadmill speed (2.2 ms^{-1}) compared to slower treadmill speeds (Chen et al., 2015). Alsubheen et al. (2016) found that the GVF underestimated steps during treadmill walking at 0% incline, but no significant differences were observed at 5% and 10% inclines. Collectively, it appears that the monitor has a tendency to underestimate actual step count with improved accuracy as intensity of the activity increases. Similar findings have been reported for the wrist-worn Fitbit monitor, although an older Fitbit model (Fitbit Flex) has been implemented in step validation research (Chen et al., 2015; Diaz et al., 2015).

There are some limitations of the current study. First, the sample population included only young (18-45 years), healthy, and physically active individuals. Thus, the generalization of the study findings should not be implemented with other groups of different ages, fitness levels, and/or health status. Secondly, application of the results is limited to the activities during which the monitors were examined, including stationary cycling and treadmill ambulation. Lastly, the study included validation of only three activity monitors. In addition, the SWA has since been discontinued and is no longer available in the market. Future research should focus on measurement accuracy of activity monitors during a variety of physical activities, and include a more diverse sample to address how the accuracy of these monitors is impacted under a variety of conditions.

In conclusion, the current study demonstrates that the FBS, GVF, and SWA tend to underestimate exercise variables during treadmill ambulatory activities and stationary cycling when compared to the criterion measures, specifically at higher intensities. However, some of these estimates were promising. The activity monitors appear to be a valuable tool in measuring HR and step count while the accuracy of EE estimates should be interpreted with caution. While the monitors included in the current study underestimated EE, it may be more beneficial for weight loss purposes for activity monitors to underestimate rather than overestimate EE. The current findings provide consumers and researchers with insight on the functionality of these activity monitors during common aerobic activities including treadmill walking, running, and stationary cycling performed at relative HR intensity. The current results indicate acceptable validity of the activity monitors in measuring HR_{avg}, HR_{max}, and step variables. The utility of the findings may be valuable for personal measurement of these physical activity variables during exercise. It is important for individuals to obtain accurate real-time physical activity feedback to help maintain or increase their physical activity levels.

In addition, the monitors used in the current study may also be utilized in research applications for behavioral and measurement purposes in place of expensive and complex laboratory equipment.

CHAPTER III REFERENCES

Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, *28*(3), R1-R39.

Alsubheen, S. A. A., George, A. M., Baker, A., Rohr, L. E., & Basset, F. A. (2016). Accuracy of the Vivofit activity tracker. *Journal of Medical Engineering & Technology*, *40*(6), 298-306.

American College of Sports Medicine. (2014). *ACSM's guidelines for exercise testing and prescription,* (9th ed.). Baltimore: Lippincott, Williams & Wilkins.

Bai, Y., Welk, G. J., Nam, Y. H., Lee, J. A., Lee, J. M., Kim, Y., . . . Dixon, P. M. (2016). Comparison of consumer and research monitors under semistructured settings. *Medicine and Science in Sports and Exercise, 48*(1), 151-158.

Benito, P. J., Neiva, C., González-Quijano, P. S., Cupeiro, R., Morencos, E., & Peinado, A. B. (2012). Validation of the SenseWear armband in circuit resistance training with different loads. *European Journal of Applied Physiology*, *112*(8), 3155-3159.

Blair, S. N., Goodyear, N. N., Gibbons, L. W., & Cooper, K. H. (1984). Physical fitness and incidence of hypertension in healthy normotensive men and women. *Journal of the American Medical Association*, *252*(4), 487-490.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research, 8*(2), 135-160.

Brage, S., Brage, N., Ekelund, U., Luan, J. A., Franks, P. W., Froberg, K., & Wareham, N. J. (2006). Effect of combined movement and heart rate monitor placement on physical activity estimates during treadmill locomotion and free-living. *European Journal of Applied Physiology*, *96*(5), 517-524.

Bouten, C. V., Koekkoek, K. T., Verduin, M., Kodde, R., & Janssen, J. D. (1997). A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Transactions on Biomedical Engineering*, *44*(3), 136-147.

Chen, M. D., Kuo, C. C., Pellegrini, C. A., & Hsu, M. J. (2016). Accuracy of wristband activity monitors during ambulation and activities. *Medicine and Science in Sports and Exercise*, *48*(10), 1942-1949.

Colley, R. C., & Tremblay, M. S. (2011). Moderate and vigorous physical activity intensity cut-points for the Actical accelerometer. *Journal of Sports Sciences*, *29*(8), 783-789.

Diaz, K. M., Krupka, D. J., Chang, M. J., Peacock, J., Ma, Y., Goldsmith, J., . . . Davidson, K. W. (2015). Fitbit®: An accurate and reliable device for wireless physical activity tracking. *International Journal of Cardiology*, *185*, 138-140.

Dooley, E. E., Golaszewski, N. M., & Bartholomew, J. B. (2017). Estimating accuracy at exercise intensities: A comparative study of self-monitoring heart rate and physical activity wearable devices. *Journal of Medical Internet Research, mHealth and uHealth*, *5*(3), e34.

Drenowatz, C., & Eisenmann, J. C. (2011). Validation of the SenseWear Armband at high intensity exercise. *European Journal of Applied Physiology, 111*(5), 883-887.

Evenson, K. R., Goto, M. M., & Furberg, R. D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, *12*(1), 1-22.

Fitbit. (2016, May 16). How does Fitbit estimate how many calories I've burned? Retrieved from http://help.fitbit.com/articles/en_US/Help_article/1381.

Garber, C. E., Blissmer, B., Deschenes, M. R., Franklin, B. A., Lamonte, M. J., Lee, I. M., . . . Swain, D. P. (2011). American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: Guidance for prescribing exercise. *Medicine and Science in Sports and Exercise*, *43*(7), 1334-1359.

Garmin. (n.d.). Calorie terminology. Retrieved from https://support.garmin.com/faqSearch/enUS/faq/content/lkl4cwCLlK7ox362uGQEV7

Huang, Y., Xu, J., Yu, B., & Shull, P. B. (2016). Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. *Gait and Posture*, *48*, 36-41.

Jakicic, J. M., Marcus, M., Gallagher, K. I., Randall, C., Thomas, E., Goss, F. L., & Robertson, R. J. (2004). Evaluation of the SenseWear Pro Armband to assess energy expenditure during exercise. *Medicine and Science in Sports and Exercise, 36*(5), 897-904.

Karvonen, J., & Vuorimaa, T. (1988). Heart rate and exercise intensity during sports activities. *Sports Medicine*, *5*(5), 303-311.

King, G. A., Torres, N., Potter, C., Brooks, T. J., & Coleman, K. J. (2004). Comparison of activity monitors to estimate energy cost of treadmill exercise. *Medicine and Science in Sports and Exercise, 36*(7), 1244–1251.

Kirk, S. E. (2016). *Comparison of the Apple Watch, Fitbit Surge, and Actigraph GT9X Link in Measuring Energy Expenditure, Steps, Distance, and Heart Rate* (Unpublished doctoral dissertation). Cleveland State University, Cleveland, OH.

Koehler, K., deMarees, M., Braun H., & Schaenzer, W. (2010). Comparison of two portable devices for assessing energy expenditure during high-intensity running. *Medicine and Science in Sports and Exercise, 42*(5), S304–S305.

Lee, C. M., & Gorelick, M. (2011). Validity of the Smarthealth watch to measure heart rate during rest and exercise. *Measurement in Physical Education and Exercise Science, 15*(1), 18-25.

Lee, J. M., Kim, Y., & Welk, G. J. (2014). Validity of consumer-based physical activity monitors. *Medicine and Science in Sports and Exercise*, *46*(9), 1840-1848.

Maddison, R., & Ni Muhurchu, C. (2009). Global positioning system: A new opportunity in physical activity measurement. *International Journal of Behavioral Nutrition and Physical Activity, 23*, 1489-1495.

Massey, B. L., Funk, M. D.; Thiebaud, R., & Patton, J. C. (2017). Accuracy of wrist-worn physical activity monitors to measure energy expenditure, *International Journal of Exercise Science: Conference Proceedings*, *9*(2), Article 90.

Montgomery, P. G., Green, D. J., Etxebarria, N., Pyne, D. B., Saunders, P. U., & Minahan, C. L. (2009). Validation of heart rate monitor-based predictions of oxygen uptake and energy expenditure. *The Journal of Strength and Conditioning Research*, *23*(5), 1489-1495.

Morgan, D. W., Martin, P. E., Krahenbuhl, G. S., & Baldini, F. D. (1990). Variability in running economy and mechanics among trained male runners. *Medicine and Science in Sports and Exercise*, *23*(3), 378-383.

Nelson, M. B., Kaminsky, L. A., Dickin, D. C., & Montoye, A. H. (2016). Validity of consumer-based physical activity monitors for specific activity types. *Medicine and Science in Sports and Exercise*, *48*(8), 1619-1628.

Palatini, P., Benetos, A., Grassi, G., Julius, S., Kjeldsen, S. E., Mancia, G., . . . Zanchetti, A. (2006). Identification and management of the hypertensive patient with elevated heart rate: Statement of a European Society of Hypertension consensus meeting. *Journal of Hypertension*, *24*(4), 603-610.

Physical Activity Guidelines Advisory Committee. (2008). Physical activity guidelines advisory committee report, 2008. *Washington, DC: US Department of Health and Human Services*, A1-H14.

Pickering, T. G., Hall, J. E., Appel, L. J., Falkner, B. E., Graves, J., Hill, M. N., . . . Roccella, E. J. (2005). Recommendations for blood pressure measurement in humans and experimental animals. *Circulation*, *111*(5), 697-716.

Pollock, M. L., Schmidt, D. H., & Jackson, A. S. (1980). Measurement of cardiorespiratory fitness and body composition in the clinical setting. *Comprehensive Therapy, 6*, 12-27.

Price, K., Bird, S. R., Lythgo, N., Raj, I. S., Wong, J. Y., & Lynch, C. (2016). Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. *Journal of Medical Engineering and Technology*, *41*, 208-215.

Rosdahl, H., Gullstrand, L., Salier-Eriksson, J., Johansson, P., & Schantz, P. (2010). Evaluation of the Oxycon mobile metabolic system against the Douglas bag method. *European Journal of Applied Physiology*, *109*(2), 159-171.

Stahl, S. E., An, H. S., Dinkel, D. M., Noble, J. M., & Lee, J. M. (2016). How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? *BMJ Open Sport and Exercise Medicine*, *2*(1), e000106.

Storm, F. A., Heller, B. W., & Mazzà, C. (2015). Step detection and activity recognition accuracy of seven physical activity monitors. *PLoS One*, *10*(3), e0118723.

Tucker, W. J., Bhammar, D. M., Sawyer, B. J., Buman, M. P., & Gaesser, G. A. (2015). Validity and reliability of Nike+ Fuelband for estimating physical activity energy expenditure. *BMC Sports Science, Medicine and Rehabilitation*, *7*(1), 14-20.

Vernillo, G., Savoldelli, A., Pellegrini, B., & Schena, F. (2015). Validity of the SenseWear armband to assess energy expenditure in graded walking. *Journal of Physical Activity and Health*, *12*(2), 178-183.

Wallen, M. P., Gomersall, S. R., Keating, S. E., Wisløff, U., & Coombes, J. S. (2016). Accuracy of heart rate watches: Implications for weight management. *PloS one*, *11*(5), e0154420.

Welk, G. J., Schaben, J. A., & Morrow Jr, J. R. (2004). Reliability of accelerometry-based activity monitors: A generalizability study. *Medicine and Science in Sports and Exercise*, *36*(9), 1637-1645.

APPENDICES FOR STUDY I

APPENDIX A

Informed Consent Form

Principal Investigator: Veronika Pribyslavska
Study Title: Comparison of Consumer and Research Activity Monitors in Estimating Energy Expenditure
Institution: Middle Tennessee State University

Name of participant: ______________________________ Age: __________

The following information is provided to inform you about the research project and your participation in it. Please read this form carefully and feel free to ask any questions you may have about this study and the information given below. You will be given an opportunity to ask questions, and your questions will be answered. Also, you will be given a copy of this consent form.

Your participation in this research study is voluntary. You are also free to withdraw from this study at any time. In the event new information becomes available that may affect the risks or benefits associated with this research study or your willingness to participate in it, you will be notified so that you can make an informed decision whether or not to continue your participation in this study.

For additional information about giving consent or your rights as a participant in this study, please feel free to contact the MTSU Office of Compliance at (615) 494-8918.

1. **Purpose of the study:**
   The purpose of this study is to validate several activity trackers including the Fitbit Surge™, Garmin Vívofit®, and SenseWear Armband™ during a treadmill walk, treadmill run, and stationary leg cycling.

2. **Description of procedures to be followed and approximate duration of the study:**
   The study will consist of three visits. Before the first visit, you will fill out a brief questionnaire sent to you via email to determine your eligibility in my study. If you qualify, we will schedule your first visit. Today is the first visit. After you have read this document (the informed consent) and your questions have been answered, you will be asked to sign it.

   After you have signed this form, you will undergo initial screening consisting of completing a cardiovascular risk factor classification form, resting blood pressure assessment, and Global Physical Activity questionnaire. If you are classified as moderate risk, you will be asked to obtain a written medical clearance from your physician before continuing in the study. After you complete the forms and blood pressure measurement, you will be asked to rest in a seated position for 5 minutes. Your resting heart rate will be measured afterwards. Next, your standing height, body mass, and body fat percentage assessments will be completed. Afterwards, you will be familiarized with the equipment to be used in this study by completing three 10-minute treadmill bouts while wearing the equipment, each separated by a 10-minute resting period. During this time, you will be asked to walk and run on the treadmill to help us determine your personalized speed and grade for the exercise trial. After the treadmill session, you will rest for 10 minutes and then complete two 10-minute cycling bouts. You will be given 10-minute resting periods between each exercise bout. During the total 20 minutes of cycling, we will determine the resistance that you will be cycling at for the experimental trial.

   Before the second and third visits, you will be asked to abstain from exercise and consumption of alcohol and caffeine 24-hours prior to reporting to the lab. You will also be asked to report to the lab in the morning after a 12-hour fast. During the second and third visits, you will perform either treadmill or stationary cycle activities. The order of the activities will be randomly chosen for you. Upon your arrival, we will measure your resting metabolic rate using a portable metabolic analyzer (Oxycon Mobile). The mobile metabolic analyzer will be worn on your chest or back using a light weight backpack and all the expired gases will be collected via a mask that will be placed on your face around the nose and the mouth. You will be breathing room air. The mask will only measure how much oxygen your body is using. You will be asked to rest in a supine position for 10 minutes in a quiet room. After the 10 minutes of rest, we will continue to measure your resting metabolic rate for another 15 minutes in the same position. Thereafter, you will be fitted with 3 activity monitors (Fitbit Surge™, Garmin Vívofit®, SenseWear Armband) and 2 heart rate monitors (ANT^{+}, Dynastream Innovations; Polar Electro). The Fitbit Surge and Garmin Vívofit will be worn on your wrists and the SenseWear armband will be located on the left triceps.

All devices will be personalized using your age, gender, height, weight, and hand dominance. At no time during the study will the GPS function on any of the devices be in use.

Once all devices are initialized and properly positioned, you will be asked to complete the exercise sessions. For the treadmill protocol, you will warm-up on a motorized treadmill for 5 minutes at your self-selected speed and grade. Following the warm-up, you will complete a 10-minute walk at moderate intensity and a 10-minute run at vigorous intensity. The order of execution will be randomly selected for you. A video recording device will be positioned next to the treadmill to record your lower limbs to observe steps completed during the walk and run. After each 10-minute exercise bout, we will measure your excess post-exercise oxygen consumption (EPOC). This will measure how much oxygen your body is using after exercise. For this measurement, you will be asked to stand on the treadmill for 5 minutes and then lay down for the remaining time. We will measure your EPOC until it returns back to resting levels. For the cycling protocol, you will also perform a 5-minute self-selected warm up on a stationary cycle ergometer. After the warm-up, you will complete 10 minutes of cycling at moderate and 10 minutes of cycling at vigorous intensity levels. The order of the intensities will be randomly selected for you. We will also measure your EPOC after each 10-minute bout. You will be asked to remain seated on the bike for the first 5 minutes of the measurement and to lay down for the remaining time.

Video recording

Your lower limbs will be recorded during the treadmill walk and treadmill run. A video recording device will be positioned next to the treadmill and capturing your lower limbs only (from just above the knees down). This will allow us to analyze steps you complete during the treadmill walk and run after the exercise session and compare to the step count recorded by the activity trackers.

3. **Expected costs:** There is no cost associated with participating in this study.

4. **Description of the discomforts, inconveniences, and/or risks that can be reasonably expected as a result of participation in this study:**
The potential risks include experiencing fatigue, shortness of breath, and/or dizziness. The possibility of experiencing soreness 24 to 48 hours post the exercise sessions might also occur. Risk of injury also exists. This injury might be to the muscular or skeletal system. This risk of injury is inherent with any type of exercise. In an unlikely event, irritation in the upper arm from wearing the armband or irritation of the wrist or ankle from wearing the activity trackers might occur. You might also feel discomfort from the mask while performing the exercises. If you are uncomfortable at any time during the study, please let me know immediately. You will also be allowed to stop at any time during the study if you feel any discomfort. You may also stop at any time for no particular reason. Please be reminded that your participation is completely voluntary and you have the right to stop or withdraw at any time with no questions asked.

5. **Compensation in case of study-related injury:**
MTSU will not provide compensation in the case of study related injury.

6. **Anticipated benefits from this study:**

a) The potential benefits to science and humankind that may result from this study include understanding the validity of different methods in measuring energy expenditure and step count during exercise. Raising awareness about activity trackers and their purpose in reducing sedentary behavior might promote behavioral changes to enhance lifestyle and lead to healthier living.

b) The potential benefits to you from this study include familiarization with the functions and operation of recent consumer-based activity trackers. The findings of the study might assist you with decision-making about obtaining commercially-available activity trackers for personal purposes.

7. **Alternative treatments available:** N/A

8. **Compensation for participation:** Participants will not be compensated for participating in this study.

9. **Circumstances under which the Principal Investigator may withdraw you from study participation:**
You might be withdrawn from the study if your, another person's, or the equipment's safety is endangered by your participation. If you fail to complete or miss any of the sessions according to the study protocol, you might also be withdrawn from the study. Your participation may also be withdrawn if you do not meet the health screening and physical activity levels criteria or if found out that the questionnaires were inaccurately filled out.

10. **What happens if you choose to withdraw from study participation:**
Participation in this study is completely voluntary and, therefore, you have the right to withdraw from the study at any time for any reason without any penalty or negative consequences to you. Additionally, you have the right to remain anonymous and your data to remain confidential.

11. **Contact Information.**
If you should have any questions about this research study, please feel free to contact Veronika Pribyslavska at 615-987-2248 or my faculty advisors Dr. Vaughn Barry at 615-898-5535, Dr. Jennifer Caputo at 615-898-5547, or Dr. John Coons at 615-494-7973.

12. **Confidentiality.**
All efforts, within reason, will be made to keep the personal information in your research record private but total privacy cannot be promised. Your information may be shared with MTSU or the government, such as the Middle Tennessee State University Institutional Review Board, Federal Government Office for Human Research Protections, if you or someone else is in danger, or if we are required to do so by law.

13. **STATEMENT BY PERSON AGREEING TO PARTICIPATE IN THIS STUDY**
I have read this informed consent document and the material contained in it has been explained to me verbally. I understand each part of the document, all my questions have been answered, and I freely and voluntarily choose to participate in this study.

Date _______________ Signature of patient/volunteer _______________

Consent obtained by:

Date _______________ Signature _______________

Printed Name and Title _______________

APPENDIX B

IRB Approval Letter

IRB
INSTITUTIONAL REVIEW BOARD
Office of Research Compliance,
010A Sam Ingram Building,
2269 Middle Tennessee Blvd
Murfreesboro, TN 37129

MIDDLE TENNESSEE STATE UNIVERSITY

IRBN001 - EXPEDITED PROTOCOL APPROVAL NOTICE

Wednesday, May 25, 2016

Investigator(s): Veronika Pribyslavska (PI), Jenn Caputo, John Coons, Kala Young and Vaughn Barry (FA)
Investigator(s') Email(s): *vp7u@mtmail.mtsu.edu; vaughn.barry@mtsu.edu*
Department: Human Health and Performance

Study Title: ***Comparison of consumer and resarch activity monitors in estimating energy expenditure***
Protocol ID: **16-2274**

Dear Investigator(s),

The above identified research proposal has been reviewed by the MTSU Institutional Review Board (IRB) through the **EXPEDITED** mechanism under 45 CFR 46.110 and 21 CFR 56.110 within the category *(4) Collection of data through noninvasive procedures* A summary of the IRB action and other particulars in regard to this protocol application is tabulated as shown below:

IRB Action	APPROVED for one year from the date of this notification	
Date of expiration	5/25/2017	
Sample Size	60 (SIXTY)	
Participant Pool	Individuals who are between the ages of 18 and 45 years. Additional inclusion criteria include: 1) falling in low or moderate risk classification category according to ACSM risk classification, and 2) engaging in physical activity (as defined by the Physical activity guidelines advisory committee report, 2008) assessed by the Global Physical Activity Questionnaire (GPAQ, World Health Organization, Geneva, Switzerland).	
Exceptions	Allow collection of identifiable information to enable the researchers to make follow up sessions. However, the identifiable information MUST be destroyed once the data were transcribed.	
Restrictions	Participant prescreening and signed informed consent	
Comments	NONE	
Amendments	**Date**	**Post-approval Amendments**
		NONE

This protocol can be continued for up to THREE years (5/25/2019) by obtaining a continuation approval prior to 5/25/2017. Refer to the following schedule to plan your annual project reports and be aware that you may not receive a separate reminder to complete your continuing reviews. Failure in obtaining an approval for continuation will automatically result in cancellation of this protocol. Moreover, the completion of this study MUST be notified to the Office of Compliance by filing a final report in order to close-out the protocol.

Continuing Review Schedule:

Reporting Period	Requisition Deadline	IRB Comments
First year report	4/25/2017	INCOMPLETE
Second year report	4/25/2018	INCOMPLETE
Final report	4/25/2019	INCOMPLETE

The investigator(s) indicated in this notification should read and abide by all of the post-approval conditions imposed with this approval. Refer to the post-approval guidelines posted in the MTSU IRB's website. Any unanticipated harms to participants or adverse events must be reported to the Office of Compliance at (615) 494-8918 within 48 hours of the incident. Amendments to this protocol must be approved by the IRB. Inclusion of new researchers must also be approved by the Office of Compliance before they begin to work on the project.

All of the research-related records, which include signed consent forms, investigator information and other documents related to the study, must be retained by the PI or the faculty advisor (if the PI is a student) at the secure location mentioned in the protocol application. The data storage must be maintained for at least three (3) years after study completion. Subsequently, the researcher may destroy the data in a manner that maintains confidentiality and anonymity. IRB reserves the right to modify, change or cancel the terms of this letter without prior notice. Be advised that IRB also reserves the right to inspect or audit your records if needed.

Sincerely,

Institutional Review Board
Middle Tennessee State University
Email: irb_information@mtsu.edu (for questions)
irb_submissions@mtsu.edu (for documents)

Quick Links:

Click here for a detailed list of the post-approval responsibilities.
More information on expedited procedures can be found here.

CHAPTER IV

ACCURACY ASSESSMENT OF ACTIVITY MONITORS IN MEASURING ENERGY EXPENDITURE AND HEART RATE DURING A GYM-BASED ROUTINE

Introduction

Commercially-available activity monitors have become popular among consumers to maximize health and fitness benefits associated with aerobic and resistance training. These monitors allow tracking of personal activity patterns and monitoring of exercise intensity and energy expenditure (EE) to assist with outcomes such as weight control, improved health, and increased physical fitness (“Where are Wearable”, 2014, October 30). In addition, intervention research on wearable technology demonstrates the utility of consumer-based monitors in goal-setting, self-monitoring, and behavior change reinforcement (Ellingson, Meyer, & Cook, 2016; Kurti & Dallery, 2013; Wang et al., 2015).

Many activity monitors combine mechanical and physiological measures in a single device in an attempt to improve accuracy of EE estimates. For example, the SenseWear Armband (SWA; BodyMedia Inc., Pittsburgh, PA) is an accelerometer that integrates heat- and galvanic-related variables. The SWA was found to provide valid and reliable measures of EE (Reeve, Pumpa, & Ball, 2013; Vernillo, Savoldelli, Pellegrini, & Schena, 2015), but to underestimate EE at high exercise intensities (Benito et al., 2012; Drenowatz & Eisenmann, 2011).

It has been suggested that the addition of heart rate (HR) measurement to accelerometry-based monitors may enhance the performance of activity monitors in estimating EE (Plasqui & Westeerterp, 2005). Some monitors, such as the Garmin Vívofit (GVF), interface with chest straps to measure HR while other monitors, including the Fitbit Surge (FBS) or Apple Watch (APW), implement optical blood flow sensing technology called photoplethysmography (PPG) for HR detection.

The accuracy of activity monitors is often examined during structured bouts of aerobic activities (Dooley, Golaszewski, & Bartholomew, 2017; Price et al., 2016; Stahl, An, Dinkel, Noble, & Lee, 2014; Wallen, Gomersall, Keating, Wisløff, & Coombes, 2016). Due to the popularity of these monitors with the general public, it is important to evaluate the performance of commercial activity monitors during activities that simulate real-life conditions to establish their validity for personal measurement and intervention purposes. Therefore, the primary purpose of the current study was to examine the accuracy of EE estimates obtained by the FBS, GVF, and SWA against a portable metabolic analyzer during a gym-based session consisting of treadmill, cycling, and resistance activities in healthy, physically active individuals. The secondary purpose was to compare HR readings obtained by the FBS and GVF to a reference HR monitor.

Methods

Participants

The sample consisted of 37 male (n = 21) and female (n = 16) participants. Participants reported no major illnesses, medical complications, or musculoskeletal injuries. Physical activity level was assessed with a Global Physical Activity

Questionnaire (World Health Organization, Geneva, Switzerland) and all participants met the minimum physical activity recommendation criteria (Physical Activity Guidelines Advisory Committee report, 2008). Additionally, all participants had at least 1-year experience of resistance training and had engaged in resistance training exercises for at least 2 days a week over the past 6 months prior to study participation. All participants provided written informed consent (see Appendix A). The university Institutional Review Board approved the study (see Appendix B)

Activity monitoring instruments

Fitbit Surge (Fitbit Inc., San Francisco, CA). The FBS is a light-weight and water-resistant activity monitor that is worn on the wrist. The tracker is powered by a rechargeable battery that lasts up to 7 days. The monitor has a monochrome liquid crystal display (LCD) touch-screen display. It continuously monitors pulse rate, estimates EE, records steps, tracks distance covered and stairs climbed, and monitors sleep patterns. When enabled, a built-in global positioning system (GPS) receiver measures maximal and average pace, and tracks elevation and distance covered. A maximum of 35 hours of GPS data can be stored on the device. According to the manufacturer, estimated exercise EE can be enhanced by selecting an activity from a variety of exercise-specific modes. The FBS allows the consumer to set daily goals and tracks progress. Activity data are displayed on the monitor’s screen or can be later viewed and extracted after wirelessly synchronizing with a computer or a smartphone.

Garmin Vívofit (Garmin Ltd, KS). The GVF is a small wrist-worn activity tracker that estimates EE, records steps, tracks distance traveled, and monitors sleep activity. It features a move bar that indicates a period of inactivity that can be reset by short-term continuous bodily movement. The tracker has an option to pair with a HR transmitter to enable HR-monitoring during fitness activities. According to the manufacturer, this improves EE estimation. Replaceable coin cell batteries power the monitor for more than 1 year. Data tracked are shown on a segmented LCD display. The GVF provides a personalized daily step goal based on prior activity levels. The monitor is water resistant up to 50 meters. Data recorded by the GVF can be wirelessly synchronized and viewed on a computer or smartphone application.

SenseWear Armband Mini (BodyMedia Inc., Pittsburgh, PA). The SWA is a small armband device that estimates EE by incorporating accelerometry, heat flux, galvanic skin response, and skin temperature sensors. The SWA also monitors steps and provides time spent and intensity level of minute-by-minute activities. The SWA is worn on the back of the left arm, half-way between the Olecranon and Acromion processes. Data recorded by the armband are accessed by connecting the SWA to a computer via a universal serial bus (USB) cable.

Oxycon Mobile (OM; Carefusion Germany 234 GmbH, Hoechberg, Germany). The OM is a portable metabolic analyzer including a sensor and a receiver unit with identical dimensions of 126 x 96 x 41 mm (length x width x height). The device weighs 950 g including strap, battery, and mask. Data acquired by the measuring units are telemetrically transmitted to a base unit, which provides real-time data projected on a laptop screen via manufacturer-specific software (JLAB, Carefusion Germany 234 GmbH, Hoechberg, Germany). Calculated EE in kilocalories (kcal) is determined based on software-specific algorithms that take into consideration urea nitrogen concentration (a constant value of 15 g/day set by the program) and respiratory exchange ratio (RER) obtained by breath-by-breath oxygen consumption (VO_2) and carbon dioxide production (VCO_2). The OM was calibrated before every session based on the manufacturer's instruction. The HR data obtained by the PM (Polar Electro, Kempele, Finland) was received by the OM receiver unit and telemetrically transmitted to the software and accessed at the end of the sessions

Procedures

Visit one. On the first laboratory visit, the study procedures were explained, and participants signed the consent form. Then, participants completed the PA questionnaire and the health screening using pre-participation health history questionnaires (American College of Sports and Medicine; ACSM, 2017). After completing the paperwork, participants rested 5 minutes in a seated position prior to a resting blood pressure assessment. Participants' sex, age, smoking status, and hand dominance were documented for activity monitor device configuration. Participants were asked to empty their pockets and remove heavy clothes and shoes for body mass (to the nearest 0.1 kg; Seca, Hamburg, Germany) and standing height (to the nearest 0.1 cm; Sunbeam Products Inc., Health O Meter, Boca Ratoon, FL) assessments. Body composition was then assessed with a Harpenden skinfold caliper (Baty International, England) using a 3-site skinfold procedure (chest, abdomen, and thigh for males; triceps, suprailiac, and thigh for females; Pollock, Schmidt, & Jackson, 1980). Percent body fat (BF%) was estimated from calculation of body density and population specific BF% formulas.

Following the resting assessments, participants were fitted with a gas collection mask (Hans Rudolph Inc., V2 Mask, Shawnee, KS) and the OM for familiarization. Participants were allowed to perform any activity (e.g. walking on a treadmill, performing body weight exercises) while wearing the equipment. When participants felt accustomed to the OM, the device and the mask were removed. Following familiarization, participants received this list of resistance exercises: back squat, step forward lunge, box step ups, leg press, dead lift, hamstring curls, dumbbell bench press, machine chest fly, seated cable rows, lat pulldown, reverse fly, shoulder press, dumbbell bicep curls, cable bicep curls, cable triceps extensions, dips, crunches, Russian twists, Superman, planks, and table tops. From this list, participants created their own resistance training routine to perform during the second visit. Participants were instructed to self-select three upper body exercises, three lower body exercises, and two core exercises and weights enabling them to perform 8-12 repetitions for 2-4 sets (ACSM, 2017). At the end of the first visit, participants were shown the weight room and were allowed to familiarize themselves with any of the resistance training equipment and given the opportunity to ask questions about any of the exercises.

Visit two. Participants reported to the laboratory in a 2-hour postprandial state and after refraining from exercise and consumption of alcohol and caffeine for the prior 24 hours. Upon arrival, participants' current body mass was assessed. Then, they were fitted with the activity measuring devices. The OM was attached to a harness that was fastened over the shoulders and the upper torso. The SWA was placed on the left triceps. The FBS and the GVF were worn on the right and left wrists, as per the manufacturer's recommendations, in a counterbalanced order across participants. Additionally, the PM was strapped around the participant's chest right below the pectoralis muscles and an Ant+ (Dynastream Innovations Inc., Alberta, Canada) HR transmitter was placed directly below the PM. The Ant+ HR transmitter was used in conjunction with the GVF. Before data collection started, all devices were personalized to each participant. The wrist-worn monitors were synchronized with a smart phone (Apple Inc., Cupertino, CA) and the SWA was synchronized with a computer through the monitor specific program.

After properly donning the measuring devices, participants completed a 75-minute session comprised of 3 exercise bouts performed in a counterbalanced order across participants: 1) 15 minutes of level-grade treadmill running (Fitnex, Dallas, TX) at self-selected speeds. Participants were allowed to increase or decrease the speed at any time, but were instructed to sustain running for the entire 15 minutes; 2) 15 minutes of stationary cycling on a semi-recumbent bike (Biodex, Shirley, NY) or leg ergometer (828E Monark Exercise AB, Verberg, Sweden) at a self-selected intensity. Participants were allowed to adjust the pedaling frequency and/or resistance at any point during the activity; and 3) 35 minutes of resistance training routine, during which participants performed the resistance exercises selected on the first visit. Five-minute breaks following the first and second exercise bouts were included to facilitate transition to the next station, data recording, and resetting all devices for the following activity. Participants were instructed to perform the gym-based activities in the same manner in which they would perform a personal workout routine in their free-time.

Data processing

Estimated EE from the activity monitors and the OM was recorded in kcals. Exercise EE, average HR (HR_{avg}), and maximal HR (HR_{max}) data from the FBS and GVF were synchronized with a smart phone following each exercise bout and accessed from the monitor-specific mobile applications after the session. The EE estimates from the SWA were extracted using specific software (version 8.1) following completion of each exercise session. The EE obtained by the SWA was calculated based on proprietary algorithms that considered participant's height, weight, age, sex, and hand dominance. The software reports the data in minute-by-minute values, which were summed for each activity bout to provide total EE estimates. The EE data measured by the OM were processed after the exercise sessions through manufacturer-specific software. Proprietary algorithms calculated EE estimates as minute-by-minute averages, therefore, the estimates were summed to represent total EE for each session bout. Because the PM technology can be paired with the OM, exercise HR data were also obtained through the OM software and reported as minute-by-minute readings. These HR minute-by-minute averages were averaged out for each exercise bout to provide HR_{avg} estimates. The HR_{max} was determined by analyzing 5-second averages and reporting the highest average value for each segment of the session.

Statistical analyses

Descriptive statistics for participant characteristics, EE, HR_{max}, and HR_{avg} are reported as means and standard deviations. The validity of EE, HR_{max}, and HR_{avg} estimates from the activity monitors were analyzed for each bout separately. The EE estimates were also combined and analyzed to represent whole-session estimates. Pearson's correlations for EE, HR_{max}, and HR_{avg} were computed to determine overall association with the criterion measures.

Next, equivalency testing analyses were implemented to examine agreement between the activity monitors and the criterion measures for EE, HR_{avg}, and HR_{max}. To determine if measurements were equivalent, 90% confidence intervals (CI) of the estimates and ± 10% equivalence zone for the mean of the criterion measures were pre-determined (Bai et al., 2016; Lee, Kim, & Welk, 2014; Stahl et al., 2014). Mean absolute percentage error (MAPE) was computed to represent overall measurement error between the estimates and the criterion methods. The MAPE was calculated as [(|criterion measure – monitor|) / criterion measure] x 100. The monitors were considered to have a reasonable error rate if MAPE was ≤ 20% for EE (Bai et al., 2016; Lee et al., 2014) and ≤ 10% for HR_{avg} and HR_{max} (Dooley et al., 2017; Stahl et al., 2014). Lastly, Bland-Altman plots with 95% limits of agreements and a regression line were used to determine systematic bias of the EE estimates (Bland & Altman, 1999). The means and standard deviations, 90% CI, MAPE, Bland-Altman plots, and regression lines were obtained from the IBM SPSS statistical software package version 24 (IBM Corp., Armonk, NY).

Results

The 37 participants completed a gym-based session consisting of aerobic activities and resistance training performed at self-selected intensities. Descriptive statistics of the participants are reported in Table 1. There were no differences between male and female participants for age, however, male participants were significantly taller, had a higher body mass, and lower BF% compared to female participants.

During testing, some data were lost from each activity monitor. An initialization problem with the FBS for one participant caused that data from this monitor not to be recorded for the session. For another participant, there was a synchronization problem between the GVF and the smartphone during the stationary cycling and resistance training bouts. Lastly, the SWA-specific software encountered an error during data download and the data were lost for another participant's session. In addition, the first five participants performed the stationary cycling bout on a semi-recumbent bike, but due to equipment malfunction, the remaining 32 participants completed the cycling bout on a stationary leg ergometer.

In comparison to the OM, the wrist-worn activity monitors produced higher EE estimates for all segments of the exercise session (Table 2). The SWA overestimated EE during the treadmill running bout and underestimated EE for the rest. Analysis of overall measurement error revealed acceptable values (i.e. $\leq 20\%$) with the smallest MAPE across monitors being produced by the SWA during the whole-session EE estimates. During the treadmill running bouts, the SWA and GVF were found to have an acceptable measurement error. During this bout, the FBS produced its smallest measurement error,

Table 1

Participants' Descriptive Statistics

Variable	Males (n = 21)	Females (n = 16)	Full sample (N = 37)
Age (years)	26.6 ± 5.0	26.7 ± 5.0	26.6 ± 5.0
Height (cm)	177.0 ± 5.9	*164.4 ± 5.4	171.5 ± 8.5
Body mass (kg)	85.2 ± 9.4	*65.6 ± 9.6	76.7 ± 13.6
Body fat (%)	10.8 ± 4.0	*21.5 ± 5.0	15.4 ± 7.0

Note. * = Significantly different from male participants ($p < .01$).

Table 2

Descriptive Statistics and Mean Absolute Percentage Error for Energy Expenditure Obtained by the Oxycon Mobile and Activity Monitors

Exercise bout	*n*	EE (kcal)	MAPE (%)
Stationary cycling			
OM	37	109.2 ± 25.8	-
FBS	36	119.1 ± 42.5	33.0 ± 25.6
GVF	36	127.8 ± 38.2	31.1 ± 24.4
SWA	36	83.6 ± 30.7	30.4 ± 19.5
Treadmill running			
OM	37	163.3 ± 39.3	-
FBS	36	186.0 ± 38.5	21.9 ± 15.1
GVF	37	178.9 ± 35.1	18.0 ± 16.0
SWA	36	184.5 ± 36.9	16.6 ± 11.3
Resistance training			
OM	37	198.1 ± 56.0	-
FBS	36	243.9 ± 82.6	29.8 ± 25.1
GVF	36	279.3 ± 112.8	52.3 ± 30.2
SWA	36	182.8 ± 106.1	22.9 ± 28.5
Whole session			
OM	37	470.6 ± 106.0	-
FBS	36	549.0 ± 136.9	23.2 ± 18.1
GVF	37	575.0 ± 165.1	30.6 ± 23.8
SWA	36	450.9 ± 142.1	11.9 ± 10.5

Note. EE = Energy expenditure; MAPE = Mean absolute percentage error; OM = Oxycon Mobile; FBS = Fitbit Surge; GVF = Garmin Vívofit; SWA = SenseWear Armband.

however, this error did not meet the acceptance criteria. During the resistance training bout, the GVF had the highest MAPE across monitors.

Both monitors tended to underestimate HR_{avg} compared to the PM for all exercise bouts, with the GVF yielding lower estimates than the FBS (Table 3). During these bouts, the wrist-worn monitors measured HR_{max} and HR_{avg} with acceptable error rates (i.e. ≤ 10%), however, the GVF had an improved accuracy over the FBS as indicated by lower error rates. The magnitude of measurement error was lower for HR_{max} than for HR_{avg} in both monitors. When assessing HR values during the resistance training bouts, the activity monitors provided comparable measurements of HR_{max} to the PM, with the exception of the underestimation of HR_{max} by the FBS.

When looking at the correlational statistics, EE estimates for all monitors correlated with the OM for each exercise bouts and the whole session (Table 4). Strong relationships between the activity monitors and the OM were found for the treadmill running, resistance training, and whole session, while lower correlation coefficients were reported for the stationary cycling bout. The strongest correlation was found between the SWA and the OM for the treadmill running bout while the weakest correlation with the OM was the SWA during the stationary cycling bout. When assessing the HR data, the wrist-worn activity monitors were strongly correlated with the PM in measuring HR_{avg} and HR_{max} for all conditions (Table 5).

Through the utilization of equivalency testing analysis, it was determined that no monitor was equivalent to the OM in the estimation of EE (Figure 1). During the stationary cycling (Figure 1a), the SWA produced the narrowest 90% CI (74.9-82.2 kcal)

Table 3

Descriptive Statistics and Mean Absolute Percentage Error for Average and Maximal Heart Rate Obtained by the Polar and Activity Monitors

Exercise bout	*n*	Average HR (bpm)	MAPE (%)	Maximal HR (bpm)	MAPE (%)
Stationary cycling					
PM	37	132.1 ± 20.4	-	146.3 ± 20.3	-
FBS	36	123.9 ± 17.4	5.9 ± 8.1	146.7 ± 19.8	1.4 ± 2.0
GVF	36	130.8 ± 19.5	2.0 ± 1.5	146.6 ± 19.5	0.8 ± 1.8
Treadmill running					
PM	37	153.3 ± 15.1	-	165.7 ± 16.5	-
FBS	36	148.1 ± 12.9	3.1 ± 2.7	164.3 ± 15.1	1.5 ± 1.9
GVF	37	150.0 ± 16.6	2.4 ± 4.5	165.8 ± 17.2	0.9 ± 1.5
Resistance training					
PM	37	128.8 ± 18.2	-	162.6 ± 16.0	-
FBS	36	117.0 ± 15.2	9.5 ± 6.4	155.3 ± 15.2	7.0 ± 8.0
GVF	36	127.2 ± 18.9	2.4 ± 6.1	163.3 ± 15.2	0.9 ± 2.2

Note. bpm = Beats per minute; HR = Heart rate; MAPE = Mean absolute percentage error; PM = Polar monitor; FBS = Fitbit Surge; GVF = Garmin Vívofit.

Table 4

Pearson's Correlation for Energy Expenditure between the Oxycon Mobile and Activity Monitors

Exercise bout	FBS	GVF	SWA
Stationary cycling			
OM	$.43^{**}$	$.50^{**}$	$.39^{*}$
FBS	1	$.71^{**}$	.25
GVF		1	$.42^{*}$
Treadmill running			
OM	$.69^{**}$	$.66^{**}$	$.85^{**}$
FBS	1	$.74^{**}$	$.72^{**}$
GVF		1	$.62^{**}$
Resistance training			
OM	$.78^{**}$	$.76^{**}$	$.64^{**}$
FBS	1	$.84^{**}$	$.71^{**}$
GVF		1	$.54^{**}$
Whole session			
OM	$.73^{**}$	$.60^{**}$	$.83^{**}$
FBS	1	$.86^{**}$	$.69^{**}$
GVF		1	$.51^{**}$

Note. FBS = Fitbit Surge; GVF = Garmin Vívofit; OM = Oxycon Mobile; SWA = SenseWear Armband. $*p < .05$, $**p < .01$.

Table 5

Pearson's Correlation for Average and Maximal Heart Rate between the Polar and Activity Monitors

	Average HR		Maximal HR	
Exercise bout	FBS	GVF	FBS	GVF
Stationary cycling				
PM	$.77^{**}$	$.99^{**}$	$.99^{**}$	$.99^{**}$
FBS	1	$.74^{**}$	1	$.99^{**}$
GVF		1		1
Treadmill running				
PM	$.95^{**}$	$.90^{**}$	$.97^{**}$	$.99^{**}$
FBS	1	$.86^{**}$	1	$.97^{**}$
GVF		1		1
Resistance training				
PM	$.82^{**}$	$.90^{**}$	$.63^{**}$	$.97^{**}$
FBS	1	$.83^{**}$	1	$.61^{**}$
GVF		1		1

Note. FBS = Fitbit Surge; GVF = Garmin Vívofit; HR = Heart rate; PM = Polar monitor. $^{**}p < .01$.

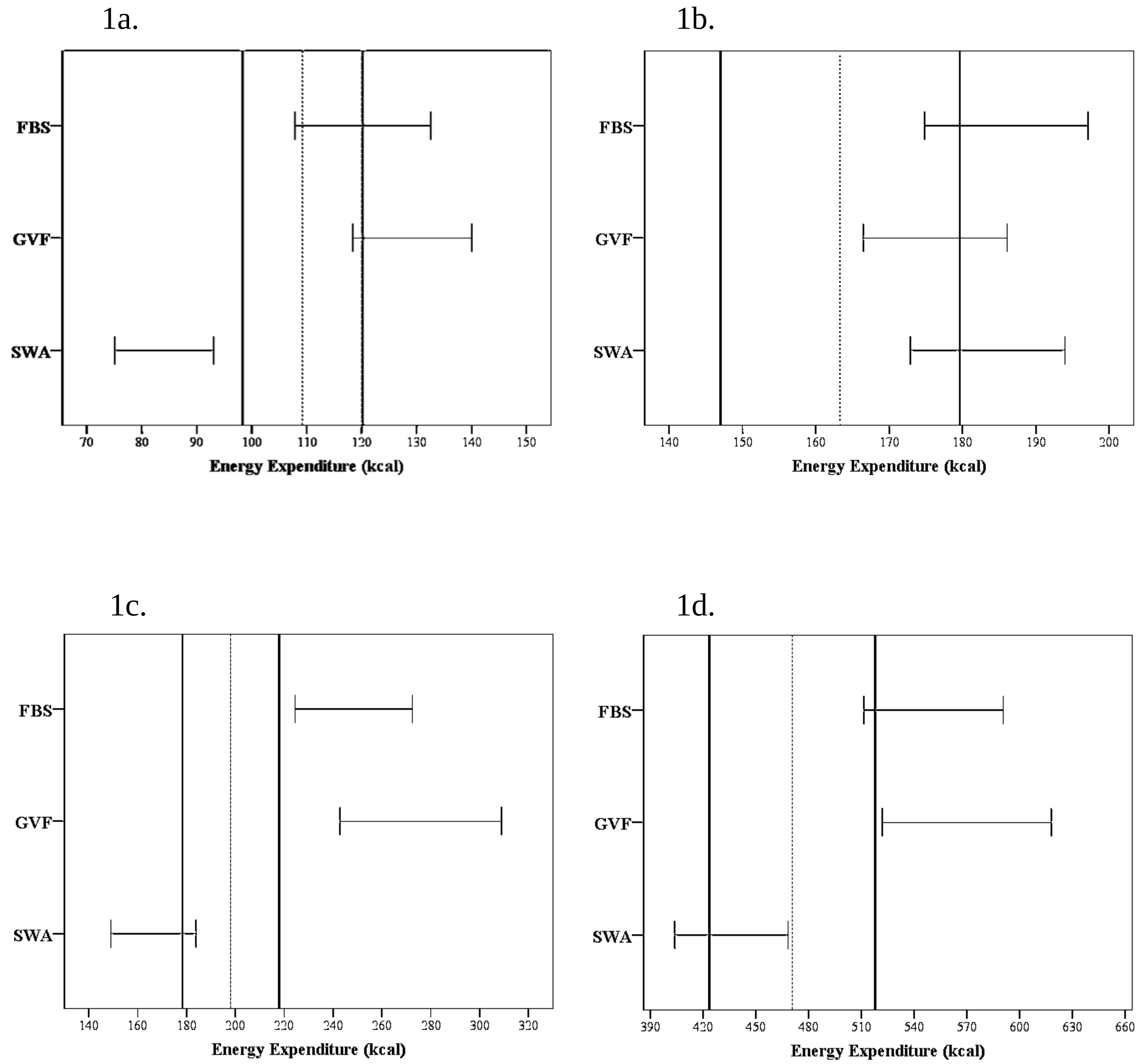


Figure 1. Equivalency testing for energy expenditure agreement between the Oxycon Mobile and activity monitors. 1a. The stationary cycling bout. 1b. The treadmill running bout. 1c. The resistance training bout. 1d. The whole session. Dashed vertical line represents mean EE measured by the OM and solid vertical lines represent the ±10% equivalency zone of the mean measured EE by the OM. EE = Energy expenditure; FBS = Fitbit Surge; GVF = Garmin Vívofit; OM = Oxycon Mobile; SWA = SenseWear Armband.

and still it fell entirely outside equivalency zone (98.3-120.2 kcal). The SWA yielded the most favorable agreement for the whole session as the monitor's 90% CI (410.9-500.1 kcal) overlapped the higher end of the equivalency zone (423.6-517.7 kcal) by only 12.7 kcal. The widest 90% CI was found for the GVF (529.2-620.8 kcal) for whole-session EE (Figure 1d). Only the FBS was not considered equivalent to the PM during the resistance training bout (Figure 2) as the 90% CI (112.7-121.2 bmp) was not entirely contained within the equivalency zone (115.9-141.7 bmp). More favorable equivalency testing results were reported for HR_{avg} (Figure 2) and HR_{max} (Figure 3), indicating nearly perfect measurement agreement between the wrist-worn monitors and the PM.

Bland-Altman plot analyses examined the distribution of error and evaluated the systematic bias in EE estimates (Figures 4, 5, 6, and 7). All monitors produced their narrowest 95% limits of agreement during the treadmill running bout (Figure 5). The SWA had the narrowest limits of agreement (difference = 83.6 kcal; Figure 5c), followed by the FBS (difference = 115.3 kcal; Figure 5a), and the GVF (difference = 121.6 kcal; Figure 5b). The widest limits of agreement were reported for the GVF for whole-session EE (difference = 517.1 kcal; Figure 7b). Significant bias was observed for the FBS during the stationary cycling (10.1 ± 39.2 kcal; Figure 4a), the resistance training (47.4 ± 52.7 kcal; Figure 6a), and for the whole session (83.2 ± 93.7 kcal; Figure 7a). Similarly, significant bias was also found for the GVF during the stationary cycling bout (18.6 ± 33.7 kcal; Figure 4b), the resistance training bout (82.0 ± 79.2 kcal; Figure 6b), and whole-session EE (104.4 ± 131.9 kcal; Figure 7b).

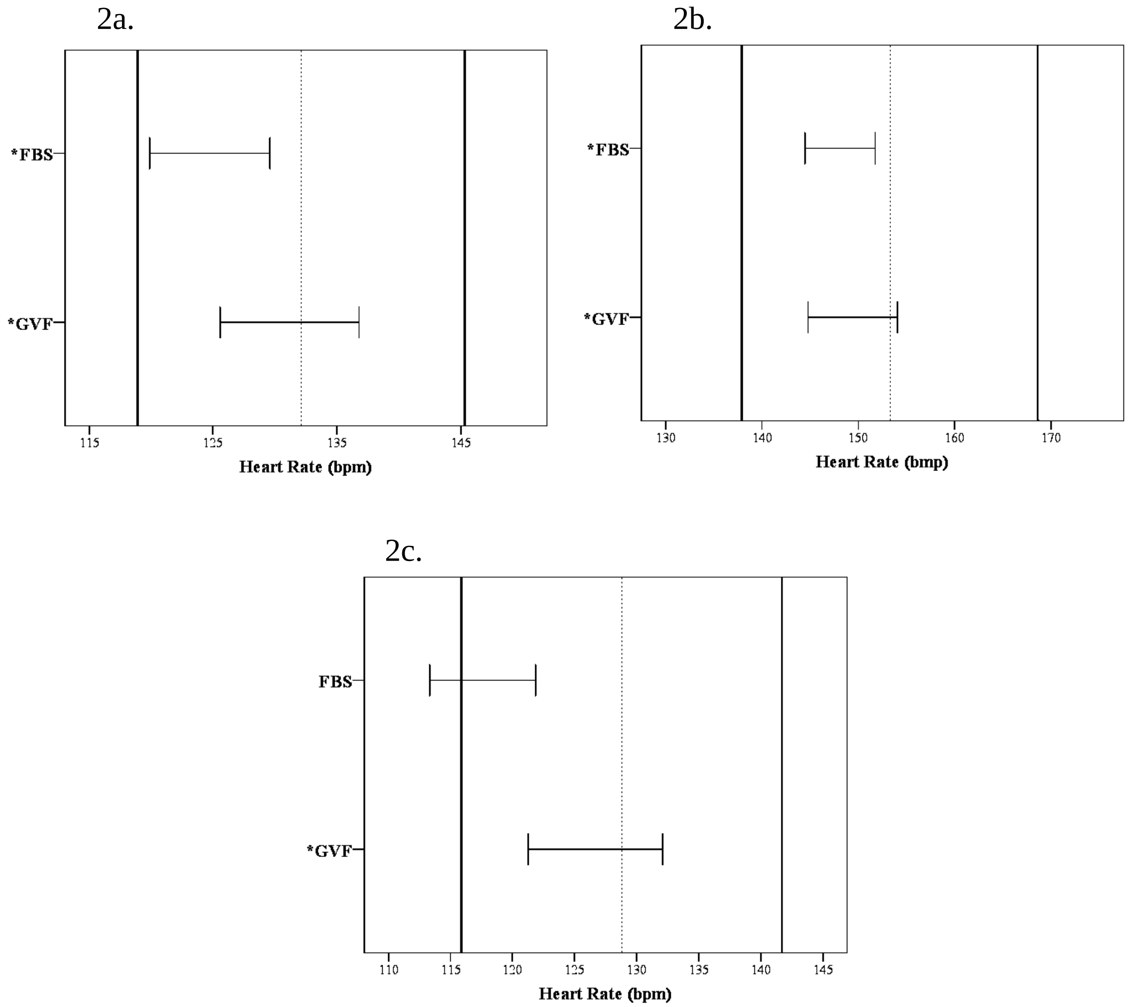


Figure 2. Equivalency testing for average heart rate agreement between the Polar and activity monitors. 2a. The stationary cycling bout. 2b. The treadmill running bout. 2c. The resistance training bout. Dashed vertical line represents mean HR_{avg} measured by the PM and solid vertical lines represent the ±10% equivalency zone of the mean HR_{avg} measured by the PM. Bpm = Beats per minute; FBS = Fitbit Surge; GVF = Garmin Vívofit; HR_{avg} = Average heart rate; PM = Polar monitor. *Within the equivalency zone.

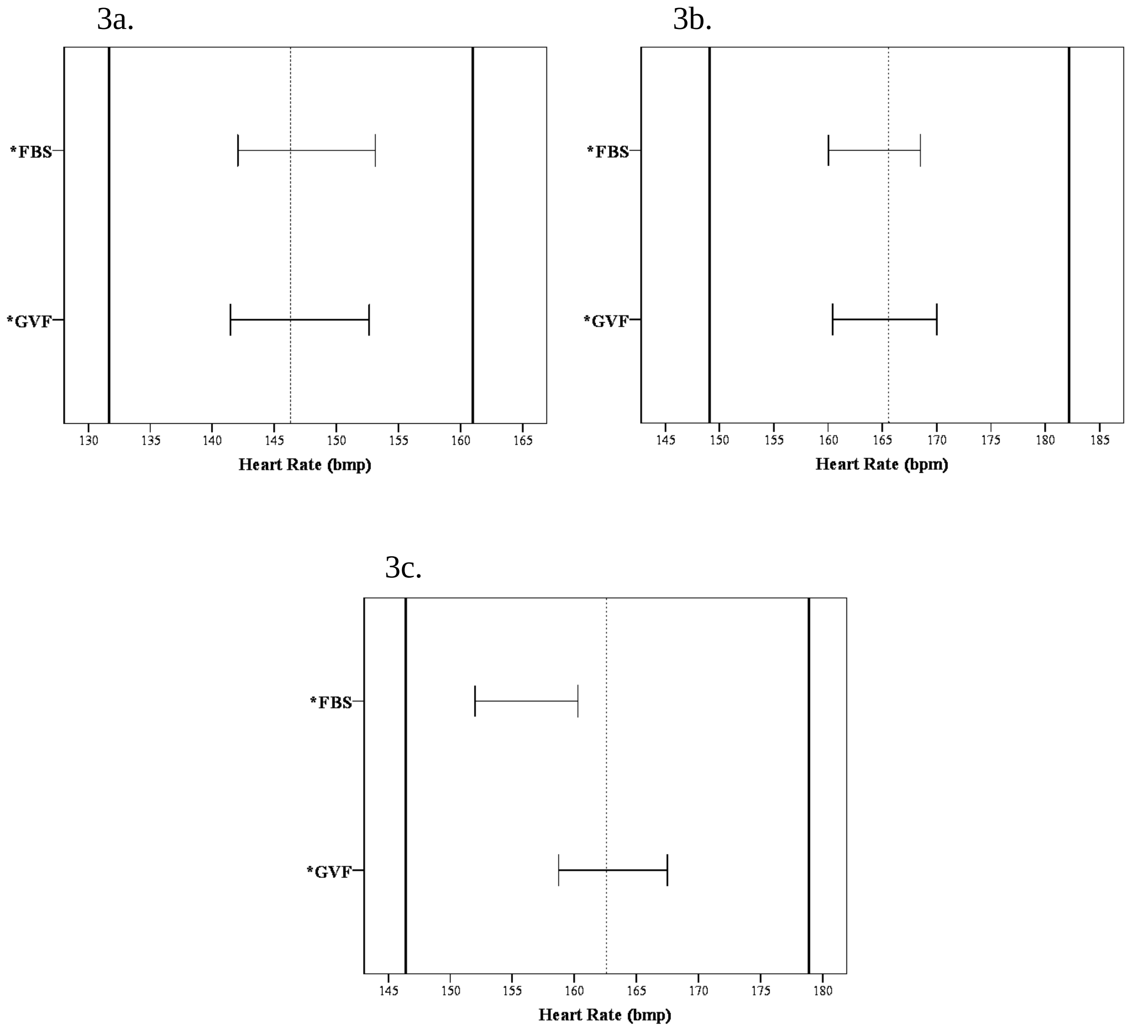
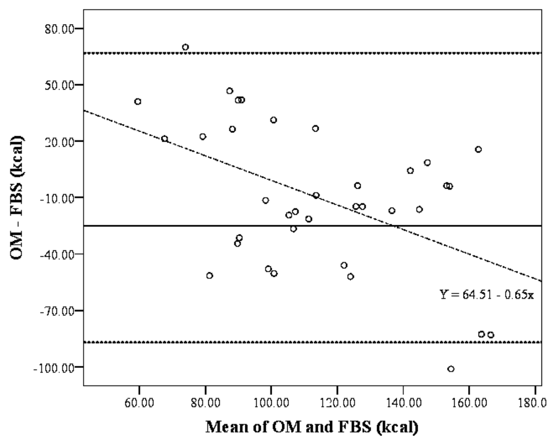
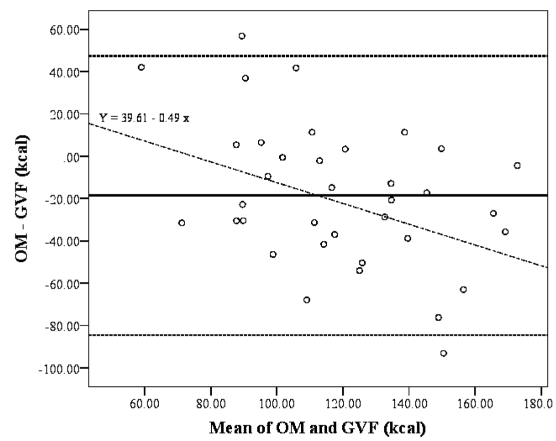


Figure 3. Equivalency testing for agreement in maximal heart rate between the Polar and activity monitors. 3a. The stationary cycling bout. 3b. The treadmill running bout. 3c. The resistance training bout. Dashed vertical line represents mean HR_{max} measured by the PM and solid vertical lines represent the ±10% equivalency zone of the mean HR_{max} measured by the PM. Bpm = beats per minute; FBS = Fitbit Surge; GVF = Garmin Vívofit; HR_{max} = Maximal heart rate; PM = Polar monitor. *Within the equivalency zone.

4a.



4b.



4c.

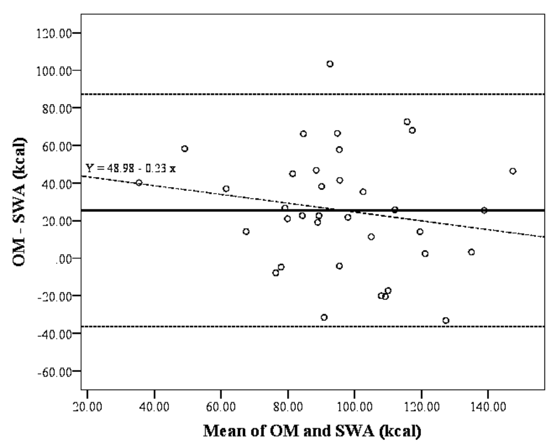
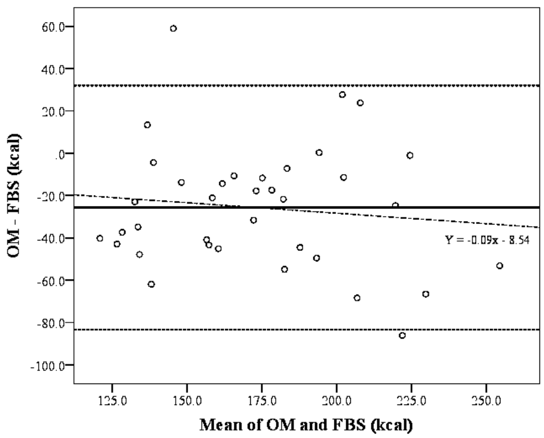
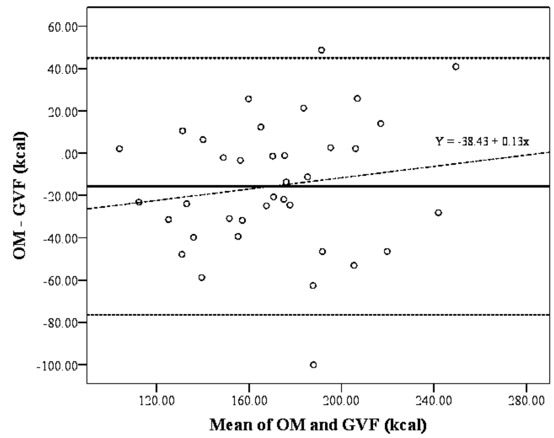


Figure 4. Bland-Altman plots for differences in energy expenditure between the Oxycon Mobile and activity monitors for the stationary cycling bout. 4a. Individual errors between the OM and the FBS. 4b. Individual errors between the OM and the GVF. 4c. Individual errors between the OM and the SWA. FBS = Fitbit Surge; GVF = Garmin Vívofit; OM = Oxycon Mobile; SWA = SenseWear Armband.

5a.



5b.



5c.

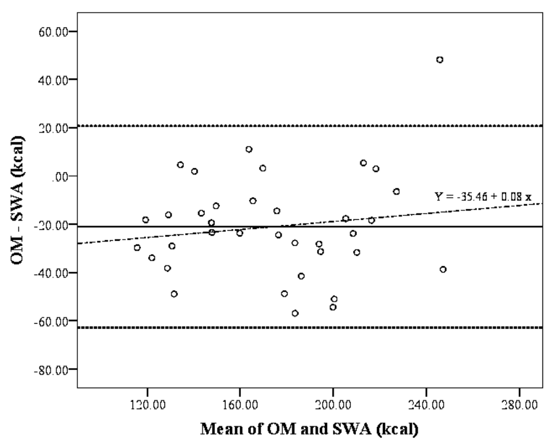
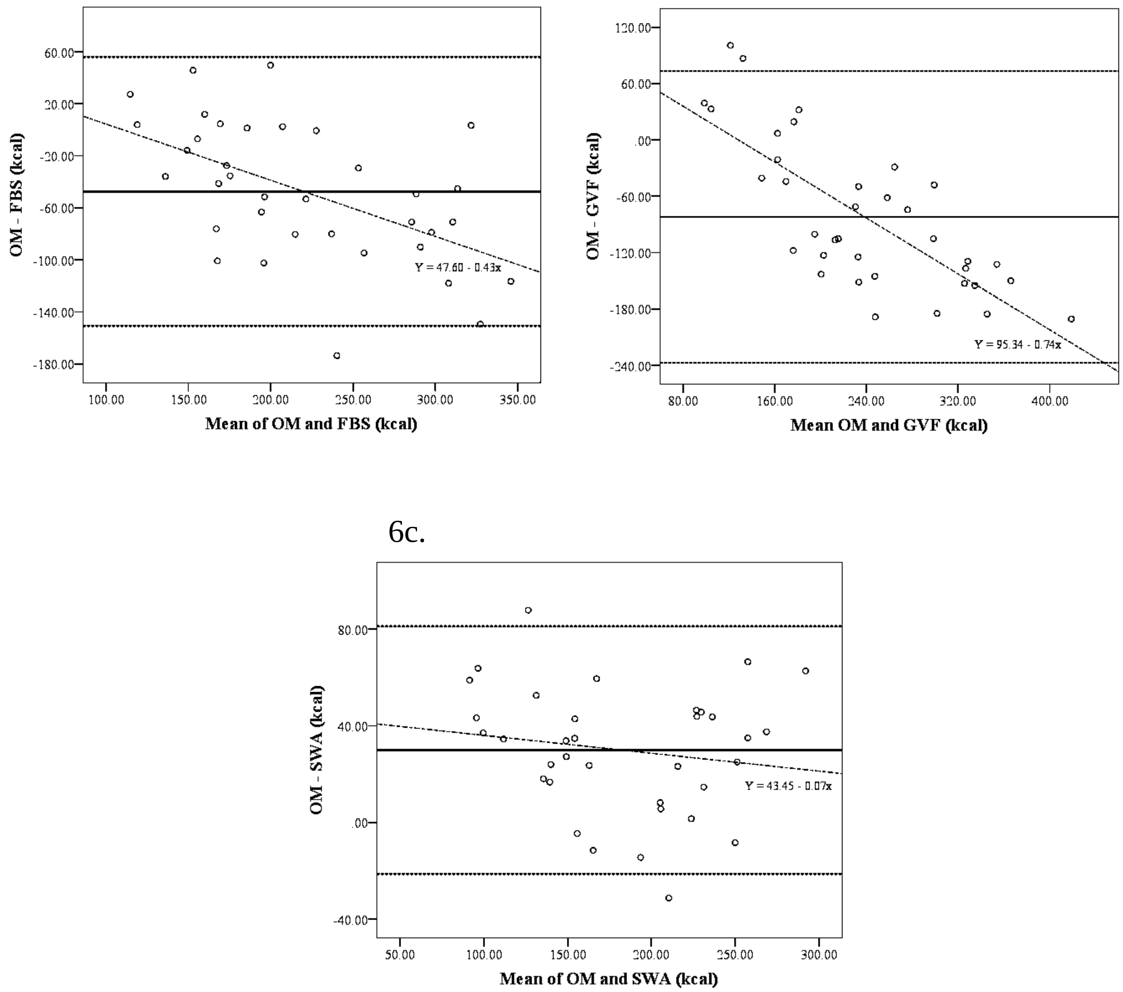


Figure 5. Bland-Altman plots for differences in energy expenditure between the Oxycon Mobile and activity monitors for the treadmill running bout.5a. Individual errors between the OM and the FBS. 5b. Individual errors between the OM and the GVF. 5c. Individual errors between the OM and the SWA. FBS = Fitbit Surge; GVF = Garmin Vívofit; OM = Oxycon Mobile; SWA = SenseWear Armband.

6a.

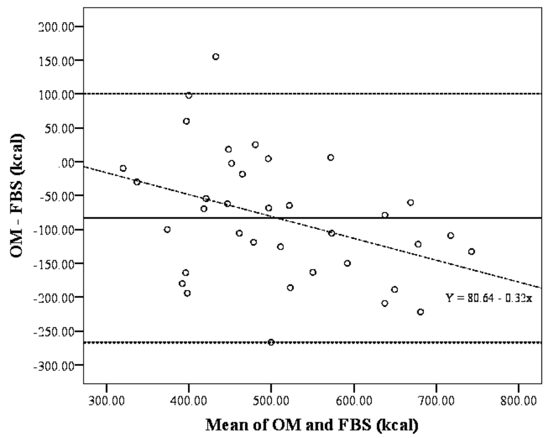
6b.



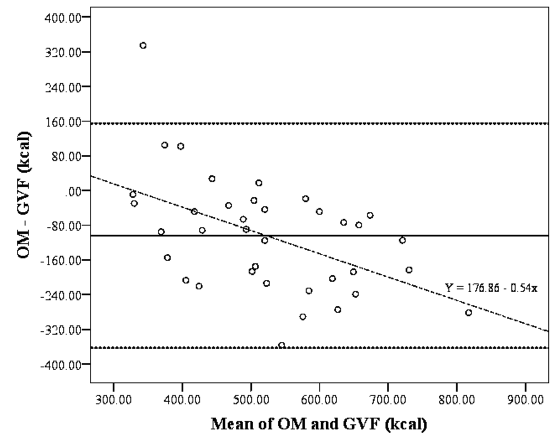
6c.

Figure 6. Bland-Altman plots for differences in energy expenditure between the Oxycon Mobile and activity monitors for the resistance training bout. 6a. Individual errors between the OM and the FBS. 6b. Individual errors between the OM and the GVF. 6c. Individual errors between the OM and the SWA. FBS = Fitbit Surge; GVF = Garmin Vívofit; OM = Oxycon Mobile; SWA = SenseWear Armband.

7a.



7b.



7c.

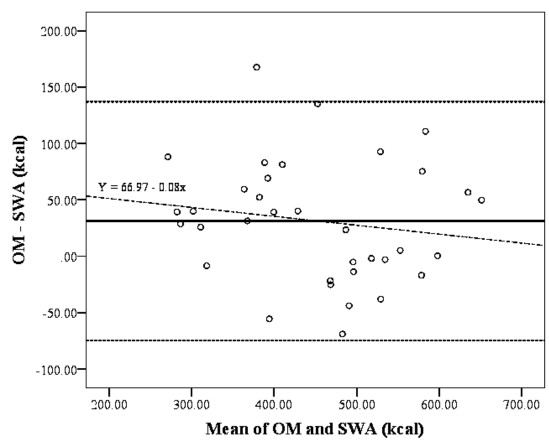


Figure 7. Bland-Altman plots for differences in energy expenditure between the Oxycon Mobile and activity monitors for the whole session. 7a. Individual errors between the OM and the FBS. 7b. Individual errors between the OM and the GVF. 7c. Individual errors between the OM and the SWA. FBS = Fitbit Surge; GVF = Garmin Vívofit; OM = Oxycon Mobile; SWA = SenseWear Armband.

Discussion

The present study examined the accuracy of the FBS, GVF, and SWA for estimating EE and HR during a gym-based routine that consisted of treadmill running, stationary cycling, and resistance training activities in healthy, physically active males and females. The study protocol was designed to assess the performance of the monitors under real-life conditions to replicate actual usage by consumers. The wrist-worn monitors consistently yielded higher EE estimates while the SWA both over- and under-estimated EE across the gym-based routine compared to the criterion measure (OM). Evaluation of HR data revealed that the FBS and GVF tend to measure lower HR_{avg} compared to the PM, but these estimates showed promising accuracy. The FBS and GVF were the most accurate in the measurement of HR_{max}, as the wrist-worn monitors had equivalent values to the PM, with low overall measurement error.

Existing validation research demonstrates large variability in the accuracy of various activity monitors in estimating EE during diverse activities. In the current study, the FBS and the GVF overestimated EE for all exercise bouts, which subsequently resulted in overall overestimation of whole-session EE. The presence of a proportional bias in the errors of EE estimates for stationary cycling (Figure 4), resistance training (Figure 6), and the whole-session (Figure 7) suggests that the monitors had a systematic tendency to yield higher EE compared to the OM during the specific periods. Contrasting outcomes are presented in previous research, indicating that the FBS (Massey, Funk, Thiebaud, & Patton, 2017) and the GVF (Alsubheen, George, Baker, Rohr, & Basset, 2016; Price et al., 2016) yielded lower EE estimates compared to indirect calorimetry during treadmill activities. In addition, the GVF underestimated EE with a proportional bias (Price et al., 2016). In the current study, the 5-minute breaks between exercise bouts may not have provide enough time for participants to recover for the next exercise bout. Consequently, participants' HR may have stayed elevated for the beginning of the subsequent exercise bout, which may have resulted in overestimation of EE by the activity monitors. This may potentially explain the discrepancy in the findings by Alsubheen et al. (2016) and Price et al. (2016) and the current study. In addition, the sample size in the current study was approximately twice as large compared to the others (Alsubheen et al., 2016, Massey et al., 2017; Price et al., 2016), which could also contribute to the disparity in the outcomes.

Comparison of present results obtained by the GVF and the FBS from stationary cycling and resistance training is difficult because, to the author's knowledge, no other study been conducted to examine the accuracy of these monitors under these conditions. However, in contrast to the current findings, a different model of the Fitbit (Fitbit Flex) along with five other wrist-worn activity monitors were found to underestimate EE during 25 minutes of a resistance exercise routine performed on a training machine (Bai et al., 2016). The varying accuracy could be attributed to differences in the resistance training protocols in the two investigations. In the study by Bai et al. (2016), the resistance exercises were self-selected and it is possible that participants completed more leg and/or core exercises than upper body exercises, while in the current study, participants completed equal amounts of leg and arm exercises (i.e. 3 sets of each). Because leg and core exercises require little to no hand and arm movements, the wrist-worn monitors would then consequently underestimate the work performed and yield lower EE.

In regard to the estimates obtained by the SWA, the armband overestimated treadmill running EE, but underestimated EE for the stationary cycling and resistance training bouts, with no evident proportional bias (Figures 4c, 5c, 6c, and 7c). Contradictory to the present findings, the SWA was previously found to underestimate EE during 10-minutes of high intensity (65, 75, and 85% of maximal VO_2) treadmill running (Drenowatz & Eisenmann, 2011) and during 5-minute running bouts above 6.3 mph (Koehler, deMarees, Braun, & Schaenzer, 2010). One possible reason for the difference in the findings could be that the two latter studies used an older version of the armband's firmware (version 6.1), while the current SWA was using updated algorithms (version 8.1).

The armband yielded lower EE estimates compared to the OM during the resistance training routine, which is consistent with past research. Benito et al. (2012) reported the SWA underestimated EE with increased error at higher intensities during circuit-style resistance training performed on an exercise machine at three different intensities (30, 50, and 70% of participants' 15 reputation maximum). Similar findings by Reeve et al. (2013) showed the SWA underestimated EE by 23.7% compared to a portable metabolic analyzer while participants performed 10 repetitions of 9 resistance exercises. It was postulated that the SWA fails to recognize the load being moved during resistance activities because the armband's algorithms depend on measurement of body acceleration (Benito et al., 2012). This would consequently lead to the same EE estimates for a movement regardless of the weight lifted, resulting in inaccurate EE estimation.

The inclusion of equivalency testing made it possible to evaluate group-level agreement between the monitors and the OM. No monitor yielded EE estimates that were equivalent to the OM (Figure 1). Similar findings were presented by Bai et al. (2016), who examined various consumer-based monitors during a protocol that resembled that of the current study. Although different monitors were used, the authors also reported that the whole-session EE estimates were not in agreement with the criterion measure (the OM). In addition, the Fitbit Flex used by Bai et al. (2016) had 90% CI that overlapped the upper level of the equivalency zone. These outcomes correspond to those produced by the FBS, which 90% CI also overlapped the upper bound of the equivalency zone for whole-session EE (Figure 1d).

The activity monitors produced overall measurement error of EE ranging between 11.9% (SWA) and 52.3% (GVF) across the exercise session (Table 2). Previous validation investigations have also shown a wide range of MAPE, although different models and monitors have been used. For example, Dooley and colleagues (2017) found that three wrist-worn monitors (Apple Watch, Fitbit Charge, and Garmin Forerunner) estimated EE during light to vigorous intensity treadmill activities with error rates as low as 14.1% to as high as 85.0%. This range is much higher in comparison to the activity-specific (treadmill) MAPE in the current study (16.6% and 21.9%). In the study by Bai et al. (2016), the MAPE range was 15.3-52.6%, and monitors generally produced higher error when evaluated separately for each exercise bout as opposed to the whole session.

This pattern was produced by the SWA in the current study. The reasonable measurement error may be attributable to the over- and under-estimation of EE across the exercise bouts balancing out.

The SWA consistently produced lower MAPE compared to the FBS and GVF for all conditions and estimated EE with an acceptable measurement error (i.e. < 20%) during the treadmill running and for the whole session. In accordance, a newer version of the SWA, the BodyMedia Core, was also found to have the lowest MAPE values among six other activity monitors and to produce MAPE of < 20% for treadmill exercise and whole session EE (Bai et al., 2016). From the wrist-worn monitors, only the GVF met the MAPE acceptance criteria for treadmill running, however, the monitor also had the highest MAPE (52.3%) when estimating resistance training EE. This conflicting accuracy should be interpreted with caution. The GVF may not be able to accurately estimate the energy cost associated with resistance activities due to inability to detect the increased energy cost associated with increased loads. In contrast, the FBS had a lower MAPE (29.8%) than the GVF for the resistance training bout. The improved accuracy of the FBS is likely grounded in the integration of activity mode in the proprietary EE estimation, but the error rate is still considerably high.

In addition to the evaluation of EE estimates, the wrist-worn monitors were also assessed for accuracy in measuring HR_{avg} and HR_{max} over the three exercise periods. In general, the FBS and GVF measured exercise HR more accurately in comparison to estimating exercise EE. Both wrist-worn monitors were strongly correlated with the PM for HR_{avg} and HR_{max} values. The FBS and the GVF measured slightly lower mean HR_{avg}

than the PM (Table 3), however, the measurement error was within the acceptable range (i.e. < 10%) for both monitors across all three exercise bouts. The GVF demonstrated superior ability to detect HR_{avg} and HR_{max} as indicated by lower MAPE. The use of traditional HR strap in conjunction with the GVF is most likely associated with the increased accuracy (Terbizan, Dolezal, & Albano, 2002).

Scarcity of HR validation research including the GVF and the FBS poses challenges for interpretation and comparison of current findings. Other models of these wrist-worn monitors have, however, been evaluated. In a study by Stahl and colleagues (2016), the Fitbit Charge HR demonstrated comparable ability to measure HR_{avg} during treadmill running as the FBS. Both monitors strongly correlated with the PM, produced low MAPEs (Fitbit Charge HR = 1.7-2.5%; FBS = 3.1%), and had good agreement with the criterion measures indicated by the 90% CIs within the equivalency zones. Slightly higher MAPE (5.8%) for HR_{avg} obtained by the Fitbit Charge HR during treadmill running was reported by Dooley et al. (2017). Collectively, the findings suggest that the Fitbit monitor may serve as a valid tool for optical detection of HR during treadmill running.

Previous research evaluating the accuracy of optically-sensing HR illustrates inconsistency in the relative validity of various activity monitors to obtain HR data (Dooley et al., 2017; Parak & Karhonen, 2014; Stahl et al., 2016; Wallen et al., 2016). The accuracy of monitors differs not only across studies, but also across activities and intensities within a single protocol. Stahl et al. (2016) concurrently assessed six wrist-worn monitors during treadmill walking and running and found the HR_{avg} estimates to be equivalent to the PM, with measurement error rate generally < 10%. Conversely, Dooley et al. (2017) reported that MAPE for HR_{avg} varied between 3.3% and 24.4% among three monitors and across light, moderate, and vigorous intensities of treadmill ambulation. In the present study, the HR_{avg} obtained by the GVF had a consistently low MAPE (2.0-2.4%), but the measurement error produced by the FBS fluctuated across exercise periods (3.1-9.5%). It is noteworthy that the FBS measured HR with the lowest accuracy during the resistance training activities, which coincides with a study by Spierer, Rosen, Litman, & Fujii (2015). It is possible that movements of the wrist and the forearm, occurring particularly during upper body resistance exercises, disrupted the interface between skin and the monitor, resulting in sampling error by the FBS.

Some limitations of the study include that the application of study findings can only be generalized to samples with similar characteristics and the same modes of activities and monitors. Another limitation regarding the activity monitors is that the GVF does not have the ability to segment activities based on mode or duration (i.e. activity files are only created for activities ≥ 10 minutes in duration). The inability to recognize the activity mode may lead to elevated EE estimates by the monitor. Furthermore, according to the manufacturer, the Ant^+ HR monitor for the GVF should be worn around the chest. In the study, the Ant^+ HR monitor was always worn below the PM. This wear placement could have impacted its accuracy. Additionally, the EE and HR from the wrist-worn monitors are reported as averages over the activity-specific time periods. Evaluating the accuracy of the monitors for shorter-time averages (minute-by-minute, 30-s-by-30-s) would provide further insight on measuring ability. However, obtaining these shorter averages requires specialized and costly software. Lastly, due to equipment malfunction, 5 participants completed their study session on a semi-recumbent bike, while the rest used a stationary cycle ergometer. Past investigations have reported no difference in physiological variables between the two methods (Eckstrom, 2000; Griffiths, 1989).

In conclusion, the current study protocol simulated real-world conditions in order to facilitate naturalistic application of the findings. During the gym-based routine, no monitor accurately estimated EE, however, the SWA had the most favorable estimates. The wrist-worn monitors demonstrated comparable performance for both EE and HR estimates. The GVF and the FBS both systematically overestimated EE and were equally accurate in measuring HR_{avg} and HR_{max} across the aerobic and resistance exercises. The GVF produced slightly lower HR measurement errors, which was most likely due to utilization of HR strap. The utility of activity monitors in promoting behavioral changes and lifestyle modifications relies on their ability to accurately assess health- and fitness-related parameters (e.g. EE and HR). However, monitors' features including appearance, cost, functions, data accessibility, and data sharing on interacting platforms will also influence their utility. Considering the comparable inaccuracy between the wrist-worn monitors, the FBS facility to continuously and effortlessly monitor HR and to track activity-specific variables (potentially providing improved accuracy as seen during the resistance exercise bout) may make this monitor more attractive to consumers for personal usage.

CHAPTER IV REFERENCES

Alsubheen, S. A., George, A. M., Baker, A., Rohr, L. E., & Basset, F. A. (2016). Accuracy of the Vívofit activity tracker. *Journal of Medical Engineering and Technology*, *40*(6), 298-306.

American College of Sports Medicine. (2017). *ACSM's guidelines for exercise testing and prescription,* (10th ed.). Baltimore: Lippincott, Williams & Wilkins.

Bai, Y., Welk, G. J., Nam, Y. H., Lee, J. A., Lee, J. M., Kim, Y., . . . Dixon, P. M. (2016). Comparison of consumer and research monitors under semistructured settings. *Medicine and Science in Sports and Exercise, 48*(1), 151-158.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research, 8*(2), 135-160.

Benito, P. J., Neiva, C., González-Quijano, P. S., Cupeiro, R., Morencos, E., & Peinado, A. B. (2012). Validation of the SenseWear armband in circuit resistance training with different loads. *European Journal of Applied Physiology*, *112*(8), 3155-3159.

Dooley, E. E., Golaszewski, N. M., & Bartholomew, J. B. (2017). Estimating accuracy at exercise intensities: A comparative study of self-monitoring heart rate and physical activity wearable devices. *Journal of Medical Internet Research, mHealth and uHealth*, *5*(3), e34.

Drenowatz, C., & Eisenmann, J. C. (2011). Validation of the SenseWear Armband at high intensity exercise. *European Journal of Applied Physiology, 111*(5), 883-887.

Eckstrom, N. A. (2000). *A comparison of submaximal exercise responses to upright and semi-recumbent cycle ergometers with habituation*. (Master's thesis, University of Wisconsin, La Crosse).

Ellingson, L. D., Meyer, J. D., & Cook, D. B. (2016). Wearable technology reduces prolonged bouts of sedentary behavior. *Translational Journal of the American College of Sports Medicine, 1*(2), 10-17.

Griffiths, H. I. Comparison of energy expenditure among three different cycling positions (1989). *Theses, Dissertations, Professional Papers.* Paper 1862.

Kurti, A. N., & Dallery, J. (2013). Internet-based contingency management increases walking in sedentary adults. *Journal of Applied Behavior Analysis*, *46*(3), 568-581.

Kelton. (2014, October 30). Where are Wearable Fitness Trackers Going for 2015? Retrieved from http://keltonglobal.com/in-the-media/where-are-wearable-fitness-trackers-going-for-2015/

Koehler, K., de Marees, M., Braun, H., & Schaenzer, W. (2010). Comparison of two portable devices for assessing energy expenditure during high-intensity running. *Medicine and Science in Sports and Exercise*, *42*(5), 304-305.

Lee, J. M., Kim, Y., & Welk, G. J. (2014). Validity of consumer-based physical activity monitors. *Medicine and Science in Sports and Exercise*, *46*(9), 1840-1848.

Massey, B. L., Funk, M. D., Thiebaud, R. P., & Patton, J. C. (2017). Accuracy of wrist-worn physical activity monitors to measure energy expenditure. *International Journal of Exercise Science: Conference Proceedings*, *2*(9), article 90.

Parak, J., & Korhonen, I. (2014). Evaluation of wearable consumer heart rate monitors based on photopletysmography. *Engineering in Medicine and Biology Society, 2014*, 3670-3673.

Physical Activity Guidelines Advisory Committee. (2008). *Physical activity guidelines advisory committee report, 2008.* Washington, DC: US Department of Health and Human Services, A1-H14.

Plasqui G., & Westerterp, K. R. (2005). Accelerometry and heart rate as a measure of physical fitness: Proof of concept. *Medicine and Science in Sports and Exercise, 37*(5), 872-876.

Pollock, M. L., Schmidt, D. H., & Jackson, A. S. (1980). Measurement of cardiorespiratory fitness and body composition in the clinical setting. *Comprehensive Therapy, 6*, 12-27.

Price, K., Bird, S. R., Lythgo, N., Raj, I. S., Wong, J. Y., & Lynch, C. (2016). Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. *Journal of Medical Engineering and Technology, 41*, 208-215.

Reeve, M. D., Pumpa, K. L., & Ball, N. (2013). Accuracy of the SenseWear Armband Mini and the BodyMedia FIT in resistance training. *Journal of Science and Medicine in Sport, 17*(6), 630-634.

Spierer, D. K., Rosen, Z., Litman, L. L., & Fujii, K. (2015). Validation of photoplethysmography as a method to detect heart rate during rest and exercise. *Journal of Medical Engineering and Technology, 39*(5), 264-271.

Stahl, S. E., An, H. S., Dinkel, D. M., Noble, J. M., & Lee, J. M. (2016). How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? *BMJ Open Sport and Exercise Medicine*, *2*(1), e000106.

Terbizan, D. J., Dolezal, B. A., & Albano, C. (2002). Validity of seven commercially available heart rate monitors. *Measurement in Physical Education and Exercise Science*, *6*(4), 243-247.

Vernillo, G., Savoldelli, A., Pellegrini, B., & Schena, F. (2015). Validity of the SenseWear armband to assess energy expenditure in graded walking. *Journal of Physical Activity and Health*, *12*(2), 178-183.

Wallen, M. P., Gomersall, S. R., Keating, S. E., Wisløff, U., & Coombes, J. S. (2016). Accuracy of heart rate watches: Implications for weight management. *PloS one*, *11*(5), e0154420.

Wang, J. B., Cadmus-Bertram, L. A., Natarajan, L., White, M. M., Madanat, H., Nichols, J. F., . . . Pierce, J. P. (2015). Wearable sensor/device (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: A randomized controlled trial. *Telemedicine and e-Health*, *21*(10), 782-792.

APPENDICES FOR STUDY II

APPENDIX A

Informed Consent Form

Middle Tennessee State University Institutional Review Board
Informed Consent Document for Research

Principal Investigator: Veronika Pribyslavska
Study Title: Accuracy Assessment of Activity Monitors in Measuring Energy Expenditure and Heart Rate during a Gym-Based Routine
Institution: Middle Tennessee State University

Name of participant: ____ Age: __________

The following information is provided to inform you about the research project and your participation in it. Please read this form carefully and feel free to ask any questions you may have about this study and the information given below. You will be given an opportunity to ask questions, and your questions will be answered. Also, you will be given a copy of this consent form.

Your participation in this research study is voluntary. You are also free to withdraw from this study at any time. In the event new information becomes available that may affect the risks or benefits associated with this research study or your willingness to participate in it, you will be notified so that you can make an informed decision whether or not to continue your participation in this study.

For additional information about giving consent or your rights as a participant in this study, please feel free to contact the MTSU Office of Compliance at (615) 494-8918.

1. **Purpose of the study:**
   The purpose of this study is to validate activity monitors including the Fitbit Surge™, Garmin Vívofit®, and SenseWear™ Armband during a gym-based routine consisting of treadmill running, stationary cycling, and resistance exercises.

2. **Description of procedures to be followed and approximate duration of the study:**
   The study will consist of two visits. The first visit will last approximately 1 hour and will consist of baseline measurements. The second visit will last approximately 80 minutes with you performing a gym-based exercise session. Today is the first visit. After reading and signing this document, you will be asked to complete a physical activity questionnaire and a health screening process.

   To complete today's session, your standing height, body mass, and body fat percentage assessments will be completed. Your sex, age, smoking status, and hand dominance will be also collected for later configuration of the activity monitors. Afterwards, you will be familiarized with the equipment to be used in this study. You will be fitted with a portable metabolic analyzer and allowed to perform any activities while wearing it. After you are comfortable with the equipment, you will take it off and will receive a list of resistance exercises. From this list, you will be asked to create a resistance training routine with the assistance of the investigator. This routine will be used when you come back for the second visit. Finally, you will be shown the weight room and will be allowed to use any of the equipment and perform any of the exercises from your routine.

   Before the second visit, you will be asked to abstain from exercise and consumption of alcohol and caffeine 24-hours prior to reporting to the lab. You will also be asked to report to the lab after not eating for at least 2 hours. Upon your arrival, your current body weight will be measured. Then, you will be fitted with all the activity monitors (Fitbit Surge™, Garmin Vívofit®, and SenseWear™ Armband), and 2 heart rate monitors (ANT+, Dynastream Innovations; Polar Electro). The Fitbit Surge and Garmin Vívofit will be worn on your wrists and the SenseWear armband will be located on the left triceps. All devices will be personalized using your age, sex, height, weight, smoking status, and hand dominance. At no time during the study will the GPS function on any of the devices be in use. You will be also fitted with the portable metabolic analyzer you tried on during the first visit. The analyzer's mask will be placed over your nose and mouth. You will be breathing room air throughout the trial. The mask will only measure how much oxygen your body is using.

Once all devices are initialized and properly positioned, you will be asked to complete the gym-based exercise session. You will complete the session that will consist of a 15-minute run on a treadmill, a 15-minute cycling on a stationary bike, and a 35-minute resistance training routine. The order of the activities will be randomly selected for you. During the treadmill and cycling bout, you will be exercising at self-selected intensities. You will be allowed to change the treadmill speed at any time but you will run at a level grade. During the resistance training routine, you will perform the exercises you selected in visit one.

3. **Expected costs:**
   There is no cost associated with participating in this study.

4. **Description of the discomforts, inconveniences, and/or risks that can be reasonably expected as a result of participation in this study:**
   The potential risks include experiencing fatigue, shortness of breath, and/or dizziness. The possibility of experiencing soreness 24 to 48 hours post the exercise sessions might also occur. Risk of injury also exists. This injury might be to the muscular or skeletal system. This risk of injury is inherent with any type of exercise. In an unlikely event, irritation in the upper arm from wearing the armband or irritation of the wrist or from wearing the activity trackers might occur. You might also feel discomfort from the mask while performing the exercises. If you are uncomfortable at any time during the study, please let me know immediately.

5. **Compensation in case of study-related injury:**
   MTSU will not provide compensation in the case of study related injury.

6. **Anticipated benefits from this study:**

a) The potential benefits to science and humankind that may result from this study include understanding the validity of different methods in measuring energy expenditure and heart rate during exercise. Raising awareness about activity monitors and their purpose in reducing sedentary behavior might promote behavioral changes to enhance lifestyles and lead to healthier living.

b) The potential benefits to you from this study include familiarization with the functions and operation of recent consumer-based activity monitors. The findings of the study might assist you with decision-making about obtaining commercially-available activity monitors for personal use.

7. **Alternative treatments available:**
   N/A

8. **Compensation for participation:**
   Participants will not be compensated for participating in this study.

9. **Circumstances under which the Principal Investigator may withdraw you from study participation:**
   You may be withdrawn from the study if your, another person's, or the equipment's safety is endangered by your participation. If you fail to complete or miss any of the sessions according to the study protocol, you might also be withdrawn from the study. Your participation may also be withdrawn if you do not meet the health screening and physical activity levels criteria or if found out that the questionnaires were inaccurately filled out.

10. **What happens if you choose to withdraw from study participation:**
    Participation in this study is completely voluntary and, therefore, you have the right to withdraw from the study at any time for any reason without any penalty or negative consequences to you. Additionally, you have the right to remain anonymous and your data to remain confidential.

Participation in this study is completely voluntary and, therefore, you have the right to withdraw from the study at any time for any reason without any penalty or negative consequences to you. Additionally, you have the right to remain anonymous and your data to remain confidential.

11. Contact Information.
If you should have any questions about this research study, please feel free to contact Veronika Pribyslavska at 615-987-2248 or my faculty advisors Dr. Vaughn Barry at 615-898-5535, or Dr. Jennifer Caputo at 615-898-5547.

12. **Confidentiality.**
All efforts, within reason, will be made to keep the personal information in your research record private but total privacy cannot be promised. Your information may be shared with MTSU or the government, such as the Middle Tennessee State University Institutional Review Board, Federal Government Office for Human Research Protections, if you or someone else is in danger, or if we are required to do so by law.

13. STATEMENT BY PERSON AGREEING TO PARTICIPATE IN THIS STUDY
I have read this informed consent document and the material contained in it has been explained to me verbally. I understand each part of the document, all my questions have been answered, and I freely and voluntarily choose to participate in this study.

____________ ________________________

Date Signature of patient/volunteer

Consent obtained by:

____________ ________________________

Date Signature

____

Printed Name and Title

APPENDIX B

IRB Approval Letter

IRB
INSTITUTIONAL REVIEW BOARD
Office of Research Compliance,
010A Sam Ingram Building,
2269 Middle Tennessee Blvd
Murfreesboro, TN 37129

MIDDLE TENNESSEE STATE UNIVERSITY

IRBN001 - EXPEDITED PROTOCOL APPROVAL NOTICE

Tuesday, May 23, 2017

Principal Investigator	**Veronika Pribyslavska** (Student)
Faculty Advisor	Vaughn W. Barry
Co-Investigators	Jenn L. Caputo and Dana K. Fuller
Investigator Email(s)	*vp2u@mtmail.mtsu.edu; vaughn.barry@mtsu.edu*
Department	Health and Human performance
Protocol Title	***Accuracy assessment of activity monitors in measuing energy expenditure and heart rate during a gym-based routine***
Protocol ID	**17-2250**

Dear Investigator(s),

The above identified research proposal has been reviewed by the MTSU Institutional Review Board (IRB) through the **EXPEDITED** mechanism under 45 CFR 46.110 and 21 CFR 56.110 within the category *(4) Collection of data through noninvasive procedures* A summary of the IRB action and other particulars in regard to this protocol application is tabulated as shown below:

IRB Action	APPROVED for one year from the date of this notification
Date of expiration	**5/31/2018**
Participant Size	60 (SIXTY)
Participant Pool	Adult (age group 18 to 45 years)
Exceptions	1. Permitted to recruit participants through word of mouth. 2. Collection of identifiable information to facilitate the project is permitted and identifiable data must be destroyed. 3. Full committee review waived - ACSM prescreening
Restrictions	**1. Mandatory signed informed consent.** **2. Research team must complete CITI HIPS training.** **3. Prescreening participants is mandatory.**
Comments	NONE

This protocol can be continued for up to THREE years (**5/31/2020**) by obtaining a continuation approval prior to **5/31/2018**. Refer to the following schedule to plan your annual project reports and be aware that you may not receive a separate reminder to complete your continuing reviews. Failure in obtaining an approval for continuation will automatically result in cancellation of this protocol. Moreover, the completion of this study MUST be notified to the Office of Compliance by filing a final report in order to close-out the protocol.

Continuing Review Schedule:

Reporting Period	Requisition Deadline	IRB Comments
First year report	4/30/2018	TO BE COMPLETED
Second year report	4/30/2019	TO BE COMPLETED
Final report	4/30/2020	TO BE COMPLETED

Post-approval Protocol Amendments:

Date	**Amendment(s)**	**IRB Comments**
NONE	NONE	NONE

The investigator(s) indicated in this notification should read and abide by all of the post-approval conditions imposed with this approval. Refer to the post-approval guidelines posted in the MTSU IRB's website. Any unanticipated harms to participants or adverse events must be reported to the Office of Compliance at (615) 494-8918 within 48 hours of the incident. Amendments to this protocol must be approved by the IRB. Inclusion of new researchers must also be approved by the Office of Compliance before they begin to work on the project.

All of the research-related records, which include signed consent forms, investigator information and other documents related to the study, must be retained by the PI or the faculty advisor (if the PI is a student) at the secure location mentioned in the protocol application. The data storage must be maintained for at least three (3) years after study completion. Subsequently, the researcher may destroy the data in a manner that maintains confidentiality and anonymity. IRB reserves the right to modify, change or cancel the terms of this letter without prior notice. Be advised that IRB also reserves the right to inspect or audit your records if needed.

Sincerely,

Institutional Review Board
Middle Tennessee State University

Quick Links:

Click here for a detailed list of the post-approval responsibilities.
More information on expedited procedures can be found here.

CHAPTER V

OVERALL CONCLUSIONS

The purpose of this dissertation was to evaluate the accuracy of three activity monitors (FBS, GVF, SWA) in measuring EE, HR, and steps under controlled and free-living conditions. The FBS and GVF are wrist-worn monitors that are available on the consumer market and the SWA (now discontinued) is a multi-sensor armband commonly used for research purposes (Benito et al., 2012; Drenowatz & Eisenmann, 2011; King, Torres, Potter, Brooks, & Coleman, 2004; Koehler, deMarees, Braun, & Schaenzer, 2010; Vernillo et al., 2015). These monitors offer relatively inexpensive and simple objective assessments of physical activity. To evaluate the accuracy of the monitors, their estimates were compared to criterion measures including a portable metabolic analyzer (OM) for EE, a Polar HR chest strap (PM) for HR, and a video observation for step count. Various statistical analyses were implemented to assess individual and group agreements with the criterion measures.

In study 1, healthy, physically active male and female participants (N = 34) completed a laboratory-controlled protocol that consisted of treadmill and stationary cycling activities performed at two intensities. The treadmill and stationary cycling protocols each included a 10-minute bout performed at moderate intensity and a 10-minute bout performed at vigorous intensity HRZs. Each activity bout was analyzed separately. In study 2, healthy and physically active male and female participants (N =

37) completed a gym-based routine comprised of 15 minutes of stationary cycling, 15 minutes of treadmill running, and 35 minutes of resistance training, all at self-selected intensities. The exercise bouts were analyzed separately and combined to represent whole-session estimates.

The findings of the first study indicated that the monitors yielded lower EE in comparison to the OM for all activity bouts, but the measurement error of the estimates was higher at the vigorous intensity. This error was specifically high for the SWA. Some of the estimates were, however, favorable. For example, the FBS was highly correlated ($r = .77$) with the OM, yielding the lowest measurement error (15.7 kcal), and producing the lowest limits of agreement with no systematic bias of error distribution during the moderate treadmill bout. Through the utilization of equivalency testing, no monitor was equivalent to the OM except for the GVF during the vigorous cycling bout. This was an interesting finding, considering that the monitor did not correlate with the OM ($r = .34$) and showed systematic underestimation of EE based on Bland-Altman plots.

The GVF and FBS showed improved accuracy in measuring HR_{avg} and HR_{max} compared to their accuracy in estimating EE. The monitors tended to measure slightly lower HR_{avg} compared to the PM, but these estimates had good agreement with the criterion measure. The GVF had a lower measurement error than the FBS for both HR variables, however, the difference between the errors of the two monitors for HR_{avg} was minimal during the treadmill bouts. The equivalency testing revealed that HR_{avg} and HR_{max} obtained by the monitors were equivalent to the PM for all activity bouts, with the exception of HR_{avg} measured by the FBS during the moderate intensity treadmill bout.

Some equivocal outcomes were observed when the HR_{avg} data were further examined through the application of other analyses. For example, the FBS did not correlate with the PM and produced relatively large measurement error for both cycling bouts. Similarly, the GVF did not correlate with the PM and produced its highest error rate during the moderate intensity treadmill protocol.

Lastly, the step count analysis in study 1 revealed that the monitors tended to underestimate the actual steps, which is in agreement with previous investigations (Alsubheen et al., 2016; Price et al., 2016; Storm et al., 2015,). However, some of the estimates had a promising agreement with the video observation. For example, based on the equivalency testing results, during the vigorous treadmill bout, all three monitors were considered equivalent to the video observation. The activity monitors also strongly correlated with the video observation. In addition, the SWA yielded equivalent estimates to the video observation with strong correlation (r = .73) for the moderate treadmill bout.

Findings of the second study revealed that the FBS and GVF consistently overestimated EE for all exercise bouts of the gym-based session. The SWA overestimated EE for the treadmill running bout but underestimated EE for the stationary cycling and resistance training bouts. The SWA had smaller MAPE for all segments of the session compared to the wrist-worn monitors. In addition, the armband had the smallest and acceptable (i.e. <20%) measurement error for whole-session EE. The SWA and GVF also had an acceptable MAPE during the treadmill running. In contrast, the GVF had the highest MAPE across monitors for the resistance training bout. Lastly, the results of equivalency testing analysis revealed that no monitor produced equivalent estimates to the EE measured by the OM.

Results of the HR_{avg} and HR_{max} analyses again showed improved accuracy of the wrist-worn monitors in measuring HR compared to estimating EE. The GVF appeared to be more accurate based on a lower measurement error for both HR variables. Both HR_{avg} and HR_{max} estimates obtained by the GVF and FBS were highly correlated with the PM, had acceptable MAPE (i.e. <10%), and were found equivalent (with the exception of HR_{avg} obtained by the FBS for the resistance training bout) to the criterion measure through implementation of the equivalency testing analysis.

Comparison of the current findings with existing literature is difficult due to differences in study protocols and activity monitors implemented in these investigations. Only a few validation studies have assessed accuracy of the GVF and FBS. For example, the GVF was found to underestimate EE during treadmill ambulation (Alsubheen et al., 2016; Price et al., 2016). This is in agreement with the results of the first study, indicating the GVF underestimated EE during the treadmill bouts, however, findings of the second study show contrasting outcomes. Similar conflicting outcomes were found for the FBS. Results from a previous study showed that the monitor underestimated EE and was not considered equivalent to the criterion measure during structured periods of treadmill walking and running (Massey, Funk, Thiebaud, & Patton, 2017), which is in accordance to the findings of the first study, however, opposes the findings of the second study. It is possible that the varying EE accuracy in studies 1 and 2 is due to differences in rest durations between activity bouts. Unlike in study 1, where participants rested in a supine position until reaching baseline VO_2 levels, allowing sufficient time for HR recovery, the resting periods in study 2 were only 5 minutes, which may not be enough time for adequate recovery. This possibly resulted in having elevated HR as participants started the subsequent exercise bout. The overestimation of EE by the GVF and FBS in study 2 may be attributed to the elevated HR. Inconsistent findings were also reported for the SWA. Results from the first study correspond with previous findings showing the armband underestimates EE during high intensity treadmill running (Drenowatz & Eisenmann, 2011; Koehler et al., 2010), which is controversial to the findings of the second study and to those reported by King et al. (2004).

Comparison of the resistance training EE findings from study 2 agrees with previous research that demonstrates that the SWA underestimated EE (Benito et al., 2012, Reeve et al., 2013). This can be contributed to the monitor's inability to detect the higher energy cost associated with increased loads. The same can be concluded for the wrist-worn monitors, specifically for the GVF that produced the highest measurement error during the resistance training bout. The FBS also produced relatively high error rate for this bout, but it appears to have improved accuracy via the ability to track selected activities.

Similarly, lack of validation research examining HR measurement accuracy for the GVF and FBS presents challenges for comparison of current findings. In studies 1 and 2 of the dissertation, the wrist-worn monitors demonstrated good accuracy in assessing HR_{avg} and HR_{max} during various activities. Previous studies including other models of the Fitbit monitor, suggest reasonable HR measurements compared to the criterion measures for treadmill activities (Dooley et al., 2017; Stahl et al., 2016). Therefore, it could be concluded that the Fitbit monitor provides a valid method for HR measurement during treadmill ambulation. However, it is noteworthy that the FBS measured HR with the lowest accuracy during the cycling and resistance training activities, possibly due to the monitor experiencing sampling error. This could be attributed to a loss of skin contact caused by movements of the wrist and the forearm during upper body resistance exercises or due to an inability to detect pulse resulting from decreased blood flow to the arms during leg ergometry.

In conclusion, the results from studies 1 and 2 demonstrate that the activity monitors produce accurate estimates for some activities, but not for others, and provide preliminary support of the less established GVF and FBS in the literature. Specifically, EE estimates had promising accuracy for the treadmill activities, suggesting that the monitors appear to have an improved performance during treadmill compared to stationary cycling and resistance training activities. In addition, the utility of the FBS and GVF in monitoring exercise HR is supported by the dissertation study findings. In regard to the comparable performance between the FBS and the GVF, the FBS may provide additional value to users due to its advanced technological features. Taken collectively, the results demonstrate good potential for the monitors in terms of estimating PA variables, specifically HR and steps. Overall, the dissertation findings suggest that the activity monitors could be used as alternative methods for objective assessment of PA or exercise for personal purposes or intervention research applications.

DISSERTATION REFERENCES

Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, *28*(3), R1-R39.

Alsubheen, S. A. A., George, A. M., Baker, A., Rohr, L. E., & Basset, F. A. (2016). Accuracy of the Vivofit activity tracker. *Journal of Medical Engineering and Technology*, *40*(6), 298-306.

American College of Sports Medicine. (2014). *ACSM's guidelines for exercise testing and prescription,* (9th ed.). Baltimore: Lippincott, Williams & Wilkins.

An, H. S., Jones, G. C., Kang, S. K., Welk, G. J., & Lee, J. M. (2017). How valid are wearable physical activity trackers for measuring steps? *European Journal of Sport Science*, *17*(3), 360-368.

Almalki, M., Gray, K., & Sanchez, F. M. (2015). The use of self-quantification systems for personal health information: Big data management activities and prospects. *Health Information Science and Systems, 3*(1), S1-S11.

Bai, Y., Welk, G. J., Nam, Y. H., Lee, J. A., Lee, J. M., Kim, Y., . . . Dixon, P. M. (2016). Comparison of consumer and research monitors under semistructured settings. *Medicine and Science in Sports and Exercise, 48*(1), 151-158.

Benito, P. J., Neiva, C., González-Quijano, P. S., Cupeiro, R., Morencos, E., & Peinado, A. B. (2012). Validation of the SenseWear armband in circuit resistance training with different loads. *European Journal of Applied Physiology*, *112*(8), 3155-3159.

Blair, S. N., Goodyear, N. N., Gibbons, L. W., & Cooper, K. H. (1984). Physical fitness and incidence of hypertension in healthy normotensive men and women. *Journal of the American Medical Association*, *252*(4), 487-490.

Blair, S. N., Kohl, H. W., Barlow, C. E., Paffenbarger, R. S., Gibbons, L. W., & Macera, C. A. (1995). Changes in physical fitness and all-cause mortality. A prospective study of healthy and unhealthy men. *Journal of the American Medical Association*, *273*(14), 1093-1098.

Blumenthal, J. A., Emery, C. F., Madden, D. J., George, L. K., Coleman, R. E., Riddle, M. W., ... & Williams, R. S. (1989). Cardiovascular and behavioral effects of aerobic exercise training in healthy older men and women. *Journal of Gerontology*, *44*(5), M147-M157.

Bouten, C. V., Koekkoek, K. T., Verduin, M., Kodde, R., & Janssen, J. D. (1997). A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Transactions on Biomedical Engineering*, *44*(3), 136-147.

Bravata, D. M., Smith-Spangler, C., Sundaram, V., Gienger, A. L., Lin, N., Lewis, R., . . . Sirard, J. R. (2007). Using pedometers to increase physical activity and improve health: A systematic review. *Journal of the American Medical Association*, *298*(19), 2296-2304.

Cadmus-Bertram, L. A., Marcus, B. H., Patterson, R. E., Parker, B. A., & Morey, B. L. (2015). Randomized trial of a Fitbit-based physical activity intervention for women. *American Journal of Preventive Medicine*, *49*(3), 414-418.

Case, M. A., Burwick, H. A., Volpp, K. G., & Patel, M. S. (2015). Accuracy of smartphone applications and wearable devices for tracking physical activity data. *Journal of the American Medical Association*, *313*(6), 625-626.

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. *Public Health Reports*, *100*(2), 126-131.

Cavagna, G., Saibene, F., & Margaria, R. (1961). A three-directional accelerometer for analyzing body movements. *Journal of Applied Physiology*, *16*(1), 191.

Chen, K. Y., & Bassett, D. R. (2005). The technology of accelerometry-based activity monitors: Current and future. *Medicine and Science in Sports and Exercise*, *37*(11), S490-S500.

Chen, M. D., Kuo, C. C., Pellegrini, C. A., & Hsu, M. J. (2016). Accuracy of wristband activity monitors during ambulation and activities. *Medicine and Science in Sports and Exercise*, *48*(10), 1942-1949.

Centers for Disease Control and Prevention. (2013). Adult participation in aerobic and muscle-strengthening physical activities--United States, 2011. *Morbidity and Mortality Weekly Report*, *62*(17), 326-330.

Colley, R. C., & Tremblay, M. S. (2011). Moderate and vigorous physical activity intensity cut-points for the Actical accelerometer. *Journal of Sports Sciences*, *29*(8), 783-789.

Dannecker, K. L., Sazonova, N. A., Melanson, E. L., Sazonov, E. S., & Browning, R. C. (2013). A comparison of energy expenditure estimation of several physical activity monitors. *Medicine and Science in Sports and Exercise*, *45*(11), 2105-2112.

Danova, T. (2014, May 5). Just 3.3 million fitness trackers were sold in the US in the past year. Retrieved from http://www.businessinsider.com/33-million-fitness-trackers-were-sold-in-the-us-in-the-past-year-2014-5

Drenowatz, C., & Eisenmann, J. C. (2011). Validation of the SenseWear Armband at high intensity exercise. *European Journal of Applied Physiology, 111*(5), 883-887.

de Man, M., Vanderploeg, E., Aimers, N., MacMahon, C., Wise, L., & Parrington, L. (2016). Validity and inter-device reliability of dominant and non-dominant wrist worn activity trackers in suburban walking. *Sensoria: A Journal of Mind, Brain & Culture, 12*(2), 40-45.

Dengel, D. R., Pratley, R. E., Hagberg, J. M., Rogus, E. M., & Goldberg, A. P. (1996). Distinct effects of aerobic exercise training and weight loss on glucose homeostasis in obese sedentary men. *Journal of Applied Physiology*, *81*(1), 318-325.

Diaz, K. M., Krupka, D. J., Chang, M. J., Peacock, J., Ma, Y., Goldsmith, J, . . . , & Davidson, K. W. (2015). Fitbit®: An accurate and reliable device for wireless physical activity tracking. *International Journal of Cardiology*, *185*, 138-140.

Dooley, E. E., Golaszewski, N. M., & Bartholomew, J. B. (2017). Estimating accuracy at exercise intensities: A comparative study of self-monitoring heart rate and physical activity wearable devices. *Journal of Medical Internet Research, mHealth and uHealth*, *5*(3), e34.

El-Amrawy, F., & Nounou, M. I. (2015). Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? *Healthcare Informatics Research*, *21*(4), 315-320.

Ellingson, L. D., Meyer, J. D., & Cook, D. B. (2016). Wearable technology reduces prolonged bouts of sedentary behavior. *Translational Journal of the American College of Sports Medicine*, *1*(2), 10-17.

Evenson, K. R., Goto, M. M., & Furberg, R. D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, *12*(1), 1-22.

Fallow, B. A., Tarumi, T., & Tanaka, H. (2013). Influence of skin type and wavelength on light wave reflectance. *Journal of Clinical Monitoring and Computing*, *27*(3), 313-317.

Ferguson, T., Rowlands, A. V., Olds, T., & Maher, C. (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: A cross-sectional study. *International Journal of Behavioral Nutrition and Physical Activity*, *12*(1), 42-50.

Feskanich, D., Willett, W., & Colditz, G. (2002). Walking and leisure-time activity and risk of hip fracture in postmenopausal women. *Journal of the American Medical Association, 288*(18), 2300-2306.

Fitbit. (2016, May 16). How does Fitbit estimate how many calories I've burned? Retrieved from http://help.fitbit.com/articles/en_US/Help_article/1381

Folsom, A. R., Caspersen, C. J., Taylor, H. L., Jacobs, D. R., Luepker, R. V., Gomez-Marin, O., . . . Blackburn, H. (1985). Leisure time physical activity and its relationship to coronary risk factors in a population-based sample The Minnesota Heart Survey. *American Journal of Epidemiology*, *121*(4), 570-579.

Garmin. (n.d.). Calorie Terminology. Retrieved from https://support.garmin.com/faqSearch/enUS/faq/content/lkl4cwCLlK7ox362uGQEV7

Gusmer, R. J., Bosch, T. A., Watkins, A. N., Ostrem, J. D., & Dengel, D. R. (2014). Comparison of FitBit® Ultra to ActiGraph™ GT1M for assessment of physical activity in young adults during treadmill walking. *The Open Sports Medicine Journal*, *8*(1), 11-15.

Haskell, W. L., Lee, I. M., Pate, R. R., Powell, K. E., Blair, S. N., Franklin, B. A., . . . Bauman, A. (2007). Physical activity and public health. Updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Circulation, 116*, 1081-1093.

Hauck, W. W., & Anderson, S. (1984). A new statistical procedure for testing equivalence in two-group comparative bioavailability trials. *Journal of Pharmacokinetics and Pharmacodynamics*, *12*(1), 83-91.

Healy, G. N., Dunstan, D. W., Salmon, J., Cerin, E., Shaw, J. E., Zimmet, P. Z., & Owen, N. (2007). Objectively measured light-intensity physical activity is independently associated with 2-h plasma glucose. *Diabetes Care*, *30*(6), 1384-1389.

Hill, J. O., Wyatt, H. R., Reed, G. W., & Peters, J. C. (2003). Obesity and the environment: Where do we go from here? *Science*, *299*(5608), 853-855.

Hollingworth, M., Harper, A., & Hamer, M. (2015). Dose–response associations between cycling activity and risk of hypertension in regular cyclists: The UK Cycling for Health study. *Journal of Human Hypertension*, *29*(4), 219-223.

Houmard, J. A., Tanner, C. J., Slentz, C. A., Duscha, B. D., McCartney, J. S., & Kraus, W. E. (2004). Effect of the volume and intensity of exercise training on insulin sensitivity. *Journal of Applied Physiology*, *96*(1), 101-106.

Huang, Y., Xu, J., Yu, B., & Shull, P. B. (2016). Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. *Gait and Posture*, *48*, 36-41.

Karvonen, J., & Vuorimaa, T. (1988). Heart rate and exercise intensity during sports activities. *Sports Medicine*, *5*(5), 303-311.

Kelton. (2014, October 30). Where are Wearable Fitness Trackers Going for 2015? Retrieved from http://keltonglobal.com/in-the-media/where-are-wearable-fitness-trackers-going-for-2015/

King, G. A., Torres, N., Potter, C., Brooks, T. J., & Coleman, K. J. (2004). Comparison of activity monitors to estimate energy cost of treadmill exercise. *Medicine and Science in Sports and Exercise*, *36*(7), 1244–1251.

Kirk, S. E. (2016). *Comparison of the Apple Watch, Fitbit Surge, and Actigraph GT9X Link in Measuring Energy Expenditure, Steps, Distance, and Heart Rate* (Unpublished doctoral dissertation). Cleveland State University, Cleveland, OH.

Koehler, K., deMarees, M., Braun H., & Schaenzer, W. (2010). Comparison of two portable devices for assessing energy expenditure during high-intensity running. *Medicine and Science in Sports and Exercise, 42*(5), S304–S305.

Kurti, A. N., & Dallery, J. (2013). Internet-based contingency management increases walking in sedentary adults. *Journal of Applied Behavior Analysis*, *46*(3), 568-581.

Kushi, L. H., Fee, R. M., Folsom, A. R., Mink, P. J., Anderson, K. E., & Sellers, T. A. (1997). Physical activity and mortality in postmenopausal women. *Journal of the American Medical Association*, *277*(16), 1287-1292.

Laporte, R. E., Montoye, H. J., & Caspersen, C. J. (1985). Assessment of physical activity in epidemiologic research: Problems and prospects. *Public Health Reports*, *100*(2), 131-146.

Lee, J. M., Kim, Y., & Welk, G. J. (2014). Validity of consumer-based physical activity monitors. *Medicine and Science in Sports and Exercise, 46*(9), 1840-1848.

Lindberg, R. (2000). Active living: On the road with the 10,000 Stepssm Program. *Journal of the American Dietetic Association*, *100*(8), 878-879.

Lohman, T., Going, S., Hall, M., Ritenbaugh, C., Bare, L., Hill, A., . . . Pamenter, R. (1995). Effects of resistance training on regional and total bone mineral density in premenopausal women: A randomized prospective study. *Journal of Bone and Mineral Research*, *10*(7), 1015-1024.

Loprinzi, P. D., Cardinal, B., Crespo, C., Brodowicz, G., Andersen, R., Sullivan, E., & Smit, E. (2013). Objectively measured physical activity and C-reactive protein: National Health and Nutrition Examination Survey 2003–2004. *Scandinavian Journal of Medicine and Science in Sports*, *23*(2), 164-170.

Loprinzi, P. D., Lee, H., & Cardinal, B. J. (2013). Dose response association between physical activity and biological, demographic, and perceptions of health variables. *Obesity Facts*, *6*(4), 380-392.

Lyons, E. J., Lewis, Z. H., Mayrsohn, B. G., & Rowland, J. L. (2014). Behavior change techniques implemented in electronic lifestyle activity monitors: A systematic content analysis. *Journal of Medical Internet Research*, *16*(8), e192-e206.

Manson, J. E., Greenland, P., LaCroix, A. Z., Stefanick, M. L., Mouton, C. P., Oberman, A., . . . Siscovick, D. S. (2002). Walking compared with vigorous exercise for the prevention of cardiovascular events in women. *New England Journal of Medicine*, *347*(10), 716-725.

Massey, B. L., Funk, M. D., Thiebaud, R. P., & Patton, J. C. (2017). Accuracy of wrist-worn physical activity monitors to measure energy expenditure. *International Journal of Exercise Science: Conference Proceedings*, *2*(9), article 90.

Nelson, M. B., Kaminsky, L. A., Dickin, D. C., & Montoye, A. H. (2016). Validity of consumer-based physical activity monitors for specific activity types. *Medicine and Science in Sports and Exercise, 48*(8), 1619-1628.

Noah, A. J., Spierer, D. K., Gu, J., & Bronner, S. (2013). Comparison of steps and energy expenditure assessment in adults of Fitbit Tracker and Ultra to the Actical and indirect calorimetry. *Journal of Medical Engineering and Technology, 37*(7), 456-462.

Paffenbarger Jr, R. S., Hyde, R. T., Wing, A. L., Lee, I. M., Jung, D. L., & Kampert, J. B. (1993). The association of changes in physical-activity level and other lifestyle characteristics with mortality among men. *New England Journal of Medicine*, *328*(8), 538-545.

Parak, J., & Korhonen, I. (2014). Evaluation of wearable consumer heart rate monitors based on photopletysmography. *Engineering in Medicine and Biology Society, 2014,* 3670-3673.

Physical Activity Guidelines Advisory Committee. (2008). *Physical activity guidelines advisory committee report, 2008.* Washington, DC: US Department of Health and Human Services, A1-H14.

Plasqui G., & Westerterp, K. R. (2005). Accelerometry and heart rate as a measure of physical fitness: Proof of concept. *Medicine and Science in Sports and Exercise*, *37*(5), 872-876.

Poehlman, E. T., Dvorak, R. V., DeNino, W. F., Brochu, M., & Ades, P. A. (2000). Effects of resistance training and endurance training on insulin sensitivity in nonobese, young women: A controlled randomized trial 1. *The Journal of Clinical Endocrinology and Metabolism*, *85*(7), 2463-2468.

Price, K., Bird, S. R., Lythgo, N., Raj, I. S., Wong, J. Y., & Lynch, C. (2016). Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. *Journal of Medical Engineering and Technology*, *41*, 208-215.

Reeve, M. D., Pumpa, K. L., & Ball, N. (2013). Accuracy of the SenseWear Armband Mini and the BodyMedia FIT in resistance training. *Journal of Science and Medicine in Sport*, *17*(6), 630-634.

Rivera, G., & van der Meulen, G. (2014). Gartner says in 2015, 50 percent of people considering buying a smart wristband will choose a smartwatch instead. Retrieved from http://www.gartner.com/newsroom/id/2913318

Rockhill, B., Willett, W. C., Manson, J. E., Leitzmann, M. F., Stampfer, M. J., Hunter, D. J., & Colditz, G. A. (2001). Physical activity and mortality: A prospective study among women. *American Journal of Public Health*, *91*(4), 578-583.

Sasaki, J. E., Hickey, A., Mavilia, M., Tedesco, J., John, D., Keadle, S. K., & Freedson, P. S. (2015). Validation of the Fitbit wireless activity tracker for prediction of energy expenditure. *Journal of Physical Activity and Health*, *12*(2), 149-154.

Schwingshackl, L., Missbach, B., Dias, S., König, J., & Hoffmann, G. (2014). Impact of different training modalities on glycaemic control and blood lipids in patients with type 2 diabetes: A systematic review and network meta-analysis. *Diabetologia*, *57*, 1789-1797.

Shaibi, G. Q., Cruz, M. L., Ball, G. D., Weigensberg, M. J., Salem, G. J., Crespo, N. C., & Goran, M. I. (2006). Effects of resistance training on insulin sensitivity in overweight Latino adolescent males. *Medicine and Science in Sports and Exercise*, *38*(7), 1208-1215.

Shcherbina, A., Mattsson, C. M., Waggott, D., Salisbury, H., Christle, J. W., Hastie, T., . . . & Ashley, E. A. (2017). Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *Journal of Personalized Medicine*, *7*(3), 1-12.

Shephard, R. J. (2003). Limits to the measurement of habitual physical activity by questionnaires. *British Journal of Sports Medicine*, *37*(3), 197-206.

Spann, S. (2015). Wearable fitness devices: Personal health data privacy in Washington State. *Seattle University Law Review*, *39*, 1411-1432.

Spierer, D. K., Rosen, Z., Litman, L. L., & Fujii, K. (2015). Validation of photoplethysmography as a method to detect heart rate during rest and exercise. *Journal of Medical Engineering and Technology*, *39*(5), 264-271.

Stackpool, C. M., Porcari, J. P., Mikat, R. P., Gillette, C., & Foster, C. (2014). The accuracy of various activity trackers in estimating steps taken and energy expenditure. *Journal of Fitness Research*, *3*(3), 32-48.

Stahl, S. E., An, H. S., Dinkel, D. M., Noble, J. M., & Lee, J. M. (2016). How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? *BMJ Open Sport and Exercise Medicine*, *2*(1), e000106.

Stefanick, M. L., Mackey, S., Sheehan, M., Ellsworth, N., Haskell, W. L., & Wood, P. D. (1998). Effects of diet and exercise in men and postmenopausal women with low levels of HDL cholesterol and high levels of LDL cholesterol. *The New England Journal of Medicine*, *339*(1), 12-20.

Storm, F. A., Heller, B. W., & Mazzà, C. (2015). Step detection and activity recognition accuracy of seven physical activity monitors. *PLoS One*, *10*(3), e0118723.

Takacs, J., Pollock, C. L., Guenther, J. R., Bahar, M., Napier, C., & Hunt, M. A. (2014). Validation of the Fitbit One activity monitor device during treadmill walking. *Journal of Science and Medicine in Sport*, *17*(5), 496-500.

Tanasescu, M., Leitzmann, M. F., Rimm, E. B., Willett, W. C., Stampfer, M. J., & Hu, F. B. (2002). Exercise type and intensity in relation to coronary heart disease in men. *Journal of the American Medical Association*, *288*(16), 1994-2000.

Thielemans, V., Meyns, P., & Bruijn, S. M. (2014). Is angular momentum in the horizontal plane during gait a controlled variable? *Human Movement Science*, *34*, 205-216.

Thompson, W. R. (2015). Worldwide survey of fitness trends for 2016: 10th anniversary edition. *ACSM'S Health and Fitness Journal*, *19*(6), 9-18.

Tucker, W. J., Bhammar, D. M., Sawyer, B. J., Buman, M. P., & Gaesser, G. A. (2015). Validity and reliability of Nike+ Fuelband for estimating physical activity energy expenditure. *BMC Sports Science, Medicine and Rehabilitation*, *7*(1), 14-20.

Vernillo, G., Savoldelli, A., Pellegrini, B., & Schena, F. (2015). Validity of the SenseWear armband to assess energy expenditure in graded walking. *Journal of Physical Activity and Health*, *12*(2), 178-183.

Wallen, M. P., Gomersall, S. R., Keating, S. E., Wisløff, U., & Coombes, J. S. (2016). Accuracy of heart rate watches: Implications for weight management. *PloS one*, *11*(5), e0154420.

Walsh, M. C., Hunter, G. R., Sirikul, B., & Gower, B. A. (2004). Comparison of self-reported with objectively assessed energy expenditure in black and white women before and after weight loss. *The American Journal of Clinical Nutrition*, *79*(6), 1013-1019.

Wang, J. B., Cadmus-Bertram, L. A., Natarajan, L., White, M. M., Madanat, H., Nichols, J. F., . . . Pierce, J. P. (2015). Wearable sensor/device (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: A randomized controlled trial. *Telemedicine and e-Health*, *21*(10), 782-792.

Wannamethee, G., & Shaper, A. G. (1994). The association between heart rate and blood pressure, blood lipids and other cardiovascular risk factors. *Journal of Cardiovascular Risk*, *1*(3), 223-230.

Wilde, B. E., Sidman, C. L., & Corbin, C. B. (2001). A 10,000-step count as a physical activity target for sedentary women. *Research Quarterly for Exercise and Sport*, *72*(4), 411-414.

Williams, P. T. (2010). Reductions in incident coronary heart disease risk above guideline physical activity levels in men. *Atherosclerosis*, *209*(2), 524-527.

Williams, P. T. (2013). Dose-response relationship of physical activity to premature and total all-cause and cardiovascular disease mortality in walkers. *PLoS One*, *8*(11), e78777.

Welk, G. J., Schaben, J. A., & Morrow Jr, J. R. (2004). Reliability of accelerometry-based activity monitors: A generalizability study. *Medicine and Science in Sports and Exercise*, *36*(9), 1637-1645.

Yamanouchi, K., Shinozaki, T., Chikada, K., Nishikawa, T., Ito, K., Shimizu, S., . . . Oshida, Y. (1995). Daily walking combined with diet therapy is a useful means for obese NIDDM patients not only to reduce body weight but also to improve insulin sensitivity. *Diabetes Care*, *18*(6), 775-778.

Yang, Z., Scott, C. A., Mao, C., Tang, J., & Farmer, A. J. (2014). Resistance exercise versus aerobic exercise for type 2 diabetes: A systematic review and meta-analysis. *Sports Medicine*, *44*(4), 487-499.