2020

Accuracy of predicted energy expenditure from a Fitbit Inspire HR activity monitor during short- and long-duration exercise

Cailyn Van Camp

Accuracy of Predicted Energy Expenditure from a Fitbit Inspire HR Activity Monitor During Short- and Long-Duration Exercise

by

Cailyn Van Camp

Thesis

Submitted to the School of Health Promotion and Human Performance

Eastern Michigan University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Exercise Physiology

Thesis Committee:

Rebecca Moore, Ph.D., Chair

Catherine Gammon, Ph.D.

Anthony Moreno, Ph.D.

June 23, 2020

Ypsilanti, Michigan

Dedication

To my Mom and Dad and for supporting me through all chapters of my life. Thank you for pushing me to be the best version of myself. To my sisters, for being my biggest fans and best friends no matter what. Thank you for your endless support and for reminding me that life is not always meant to be stressful, but fun.

Acknowledgements

To Becca for being a great advisor and seeing the potential in me that I didn’t see for myself. Thank you for your advice and guidance throughout the last two years and for pushing me to pursue a doctorate degree. To Catherine for allowing me the opportunities in research and for treating me as an equal. Thank you for helping me grow in the field and for your mentorship through my graduate studies. To Andrea, for the countless laughs and heart-to-heart conversations. Thank you for your endless support, I am forever thankful for your friendship. To Andy, for navigating the craziness of graduate school with me. Thank you for always being willing to help and for the infinite number of coffee runs. To my friends, for believing in me when I didn’t believe in myself. Thank you for the unconditional support, I am eternally grateful for you all. Finally, to the undergraduates that assisted in this project, thank you. Without you all, this project could not have been completed.

Abstract

PURPOSE: To determine the accuracy of predicted Energy Expenditure (EE) reported by a wrist-worn activity monitor compared to measured EE during both a long- and short-duration exercise. METHODS: In addition to a VO_{2max} treadmill test, a running speed at approximately 70 - 75% of that VO_{2max} was found during the first visit. The second and third visit was comprised of either a 30-minute or 10-minute run at the speed previously determined. A wrist activity monitor was worn and VO_2 and EE were recorded by a metabolic cart. Pearson correlation, paired samples t-test, and repeated measures ANOVAs compared predicted and measured EE. An independent samples t-test determined significant differences in characteristics between fitness groups ($p < 0.05$). RESULTS: *N = 25 (60% male).* A significant correlation was found between predicted EE and measured EE for both short and long duration ($p < 0.001$). The repeated measures ANOVA determined the interactive effect of measurement mode and fitness level was significant. CONCLUSIONS: Overall, there is a strong correlation between criterion and predictive measurement, however, consumers should exercise caution when using predicted measures.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Obesity was recognized by the American Medical Association as a disease in 2013 (Sljivic and Gusenoff, 2019). Obesity is defined as an increase in both the amount and size of fat cells in the body that may negatively affect health (Kopelman, 2000). The American College of Sports Medicine (ACSM, 2018a) classifies someone who has a body mass index (BMI) $\geq 30.0\ kg/m^2$ as obese. Despite recent efforts to control the obesity epidemic, in 2015-2016, approximately 39.8% or 93.3 million U.S adults were considered obese (Hales, Carrol, Fryar, and Ogden, 2017). Obesity has been shown to increase an individual's risk of all-cause mortality; incidence or mortality for chronic diseases, such as certain cancers and cardiovascular disease (CVD); and incidence of type 2 diabetes mellitus (T2DM) (ACSM, 2018a). Mokdad et al. (1999) reported that those with a BMI greater than or equal to 40 kg/m^2 have seven times greater the risk of being diagnosed with diabetes, six times the risk for hypertension, and two times the risk of high cholesterol. Additionally, Thompson, Edelsberg, Colditz, Bird, and Oster (1999) found that risk for hypertension is two times greater and risk for T2DM is three times greater for moderately obese males than non-obese males, and life expectancy is reduced by one year. When looking at mortality records (1986-2006), it was determined that approximately 18.2% of deaths of U.S adults were associated with increased BMI (Masters et al., 2013). Being obese led to 111,909 excess deaths compared to normal weight individuals in 2000 (Flegal, Graubard, and Williamson, 2005). In not only high resource, but low resource countries, the number of individuals considered obese is continually increasing. This rise in obesity is partially caused by a positive energy balance in which people are consuming more calories daily than they are expending. It is crucial that research is continued to not only better understand the causes of obesity, but also to determine the best practices for treatment and prevention.

Energy Imbalance

One of the determinants of obesity includes an individual having a greater caloric input than caloric output, otherwise known as a positive energy balance (Wright and Aronne, 2012). Energy balance is reached when an individual consumes and expends an equal number of calories. The total amount of energy that is expended by the body within a 24-hour time period is referred to as total energy expenditure (TEE; Ndahimana and Kim, 2017). TEE for an individual is comprised of their resting energy expenditure (REE), thermic effect of food (TEF), and energy expenditure from activity (AEE). REE refers to the amount of energy necessary to maintain metabolic functions at rest. These functions include maintenance of body temperature and functioning organs (Ndahimana and Kim, 2017). An individual's REE can be affected by gender, body temperature, age, decreased energy intake, genetics, body composition, and hormones. The TEF is the energy needed to digest, absorb, transport and metabolize food, store nutrients, and eliminate wastes (Ndahimana and Kim, 2017). Finally, AEE refers to energy expended during activity. This can be highly variable from person to person and is affected by the intensity, duration, and frequency of the activity (Ndahimana and Kim, 2017). AEE relies on the assumption that because physical activity requires the contraction of skeletal muscle, that the larger amount of muscle that is used, the higher the EE is (Vanhees et al., 2005).

Physical Activity to Promote Energy Expenditure

Physical activity (PA) is defined by ACSM (2018b) as "any bodily movement produced by the contraction of skeletal muscles that results in a substantial increase in caloric requirements over resting energy expenditure". According to the *Physical Activity Guidelines for Americans*, adults should engage in at least 150 to 300 minutes of moderate-intensity or 75 to 150 minutes of vigorous-intensity, aerobic physical activity a week (U.S. Department of Health and Human Services, 2018). Although there is substantial supportive evidence by ACSM, the Centers for Disease Control (CDC), the National Institute of Health (NIH), and the U.S. Surgeon General in the benefit of physical activity, it is estimated that 23% of adults and 81% of adolescents do not meet the physical activity guidelines (World Health Organization, 2019). Many studies have found a strong association between the prevalence of obesity and a sedentary lifestyle (Shields and Tremblay, 2008; Heinonen et al., 2013). A sedentary lifestyle refers to participation in activities that promote minimal movement and low energy expenditure (EE), such as TV viewing (Reilly, Penpraze, Hislop, Grant, and Paton, 2008). Shields and Tremblay (2008) found that 25% of men that had more sedentary lifestyles (≥ 21 hrs/week of TV viewing) were categorized as obese, opposed to only 14% of those who were more active (≤ 5 hrs/week of TV viewing).

Self-Monitoring to Promote Physical Activity Adherence

Overweight and obese individuals have reported weight maintenance and weight loss to be difficult tasks. Many who have had success with these tasks have credited their achievements to behavioral modification (Montesi et al., 2016). Included in this modification is an increase in physical activity levels. Self-monitoring has been shown to significantly increase physical activity in several populations (Cadmus-Bertram, Marcus, Patterson, Parker, and Morey, 2015; Ashe et al., 2015). Self-monitoring allows individuals to track their daily activity and receive real time feedback. Activity energy expenditure measurement is important to analyze the dose relationships between disease and physical activity levels and to determine the benefits of different intensities and types of activities. Along with its three components, PA (and EE) can be assessed using both objective and subjective methods. Each method of assessing physical activity levels has strengths and limitations that make it practical for differing populations, activities, and settings.

Methods of EE Measurement

The criterion measures of EE include direct and indirect calorimetry and doubly labelled water (DLW). Subjective measures include self-report measures such as surveys or questionnaires, food and activity diaries, and direct observation. Objective measures include heart rate monitoring, pedometry, accelerometry, and wearable devices. Wearable devices refer to devices that provide real time feedback on physical activity variables such as steps, calories burned, and heart rate (Montoye, Mitrzyk and Molesky, 2017). Approximately 33% of the global population use a fitness tracking device to track their health, including calories in and calories out (Weinswig, 2017). It has been proposed that by 2021, the usage of wearable activity monitors will increase to one in every five people (Maslakovic, Johnson, and Jovin, 2017). Due to this increase in usage, it is important that the variables being estimated from these devices are accurate.

Validity of Wearable Devices

Validation studies have been done to determine the accuracy of the estimations provided by wearable activity monitors. Previous research has found that EE reported by wearable activity monitors is accurate during short duration exercise (Diaz et al., 2015; Reddy et al., 2018; Kendall, Bellovary, and Gothe, 2019; Dondzilla and Garner, 2016). However, when utilizing discontinuous, long-duration protocols, estimations were inaccurate (Shcherbina et al., 2017; Chowdhury, Western, Nightingale, Peacock, and Thompson, 2017).

There is currently no research validating these wearable devices during a continuous, long-duration bout of exercise.

Significance

Although there are studies aimed at validating the EE estimated by these devices, new technology is always developing. Currently, there is no research on the validity of the Fitbit® Inspire HR activity monitor. Additionally, there is no research on the EE estimate of wrist-worn activity monitors during long-duration exercise or comparing the accuracy of the EE estimate of both a short- and long- duration protocol. Novel to this study is the comparison of EE measurement between moderate and high fit individuals at a relative intensity. Validation of new activity monitoring technology is important to inform consumers of any inconsistency in measurement. This research project will fill the gap in literature of accuracy of EE depending on duration of exercise and individual fitness level.

Purpose of Study

1. To examine the accuracy of the estimated energy expenditure using a Fitbit Inspire HR activity monitor during short- and long-duration exercise.
2. To determine if the Fitbit Inspire HR activity monitor will accurately predict energy expenditure during short- and long-duration exercise for individuals with a moderate cardiorespiratory fitness level (VO_{2max} in lower half of median split; ≤ 65^{th} percentile for VO_{2max}).
3. To determine if the Fitbit Inspire HR activity monitor will accurately predict energy expenditure during short- and long-duration exercise for individuals with a high cardiorespiratory fitness level (VO_{2max} in upper half of median split; > 65^{th} percentile for VO_{2max}).

Hypotheses

1. The Fitbit Inspire HR will not accurately predict energy expenditure during short- and long-duration exercise
2. The Fitbit Inspire HR will accurately predict energy expenditure during short- and long-duration exercise for individuals with a moderate cardiorespiratory fitness level (low end of median split; ≤ 65^{th} percentile for VO_{2max})
3. The Fitbit Inspire HR will not accurately predict energy expenditure during short- and long-duration exercise for individuals with a high cardiorespiratory fitness level (high end of median split; > 65^{th} percentile for VO_{2max})

Definitions

- Energy Expenditure: The amount of energy (kJ or kcal) needed to perform a specific task (i.e, digestion, resting body function, and physical activity; Vanhees et al., 2005).
- Cardiorespiratory Fitness: The ability of the circulatory and respiratory system to supply oxygen during sustained physical activity (ACSM, 2018b)
- Wearable Activity Monitor (WFT): Devices that provide real time feedback on physical activity variables such as steps, calories burned, and heart rate (Montoye et al., 2017).

Limitations of the study

A limitation of the study is a similar demographic of participants. Recruiting on a college campus made it difficult to recruit a diverse sample. Another limitation is that only one wrist-worn activity monitor was used. This doesn't allow for comparison to other wearable devices. Lastly, participants will be uncontrolled before visits. This means that there will be no control for food intake, caffeine ingestion, sleep, etc. before exercise. These are all factors that have been found to affect VO_2 values.

Strengths of the study

One strength of this study was that all participants completed the short- and long-duration protocol at an intensity of 70-75% of their measured $VO_{2max.}$ This allowed for all participants to exercise at a relative intensity. A second strength was the use of the Fitbit® Inspire HR activity monitor. This is a new device that has not yet been validated. Lastly, a long-duration exercise protocol was used. There is no previous research examining the validity of activity monitors during a continuous, long-duration exercise protocol.

Chapter 2: Review of Literature

Criterion Measurement of Energy Expenditure

In all chemical reactions, there is heat exchanged between the object and the environment. Initial studies of these heat reactions aimed to measure this heat by observing the change in temperature. This heat exchange became known as "heat energy" and defined heat as the energy exchanged per unit of time between two systems (Kenny, Notley, and Gagnon, 2017). The measurement of heat energy is calorimetry and is measured by a calorimeter (Kenny et al, 2017). Over time, calorimetry has been divided into a direct and indirect measurement. The measurement of heat energy has been used to determine caloric output.

Direct and Indirect Calorimetry. The earliest known study referenced for calorimetry was a study done by Robert Boyle in 1660. Boyle placed both mice and burning flames in separate sealed jars. He observed that as he removed air from the jars, the flame began to die out and the mice began to be inactive and dreary. When he returned air to the jar, the flame began to burn again, and the mouse became lively. This led to the finding that both fire and life were processes of combustion and that both life and fire were supported by air (Boyle, 1660).

About eight years later, Mayrow, expanded on Boyle's discovery and placed the sealed jars of mice in water (Ainslie, Reilly, and Westerterp, 2003). He observed that as the mice breathed, the jars changed placement in the water. He noticed that once $1/14^{th}$ of the air in the jar was depleted, the mice died. From this study, it became known that air was composed of different parts and that only some parts of that air were actually used for respiration. Around this same time (1669), Richard Lower, cut open the thorax of a dog and attached a device to control respiration. He observed that that blood returning to the heart was bright red, while the blood entering the lungs was of a purple color. He concluded this was due to the blood being mixed with inspired air (Lower, 1669; Karamanou, Tsoucalas, and Androutsos, 2013).

In the late 1700's, crucial advancements in the field of metabolism were made by French scientist Lavoisier and his assistant Laplace. In the winter of 1780, Lavoisier and Laplace engineered the calorimeter (Lavoisier and Laplace, 1780). Their calorimeter was made of sheet metal and was composed of an inner layer of ice and an outer layer of snow. In theory, they suspected they could place a hot object in the calorimeter, and it would cool, releasing heat and melting the ice. Lavoisier and Laplace used this instrument to test a theory of respiration. They placed guinea pigs in the calorimeter and observed that the heat from the guinea pig's respiration, melted the inner layer of ice. They recorded the weight of the water, and by using the latent heat of ice, the heat absorbed, hence produced, was calculated (Underwood, 1943). They also estimated carbon dioxide (CO_2) production by measuring the heat and CO_2 produced by burning charcoal. This allowed them to estimate the amount of heat lost per unit of CO_2 produced. This experiment resulted in the realization that heat was produced by the combustion of carbon but left the question of where heat was produced within the animal.

Four years later, Lavoisier and his new assistant Seguin, continued Lavoisier's research on respiration (Underwood, 1943). They used both animals and humans and placed them in chambers to measure oxygen consumption and carbon dioxide production. They placed animals in jars and used silk bags that were secured around the mouth and nose for humans (Lavoiser and Seguin, 1793). Their initial study consisted of placing a guinea pig in a jar over water. He observed that "noxious" gas had to be removed and that vital air consumed was always the same. They also observed that guinea pigs consumed more oxygen after eating and during movement. Following animal experiments, they began studies on humans. (Karamanou et al., 2013). They used a mask attached to the face with a tube extending to a trough of a known amount of oxygen. This allowed Lavoisier to measure the amount of oxygen consumed by Seguin (Karamanou et al., 2013). By using this mask for various studies, they concluded that a person consumes more oxygen in lower temperatures, those at rest consume less oxygen than those standing or active, and more oxygen is consumed after a meal. Lavoisier was the first to suggest that an increase in heart rate is proportional to the work done and that oxygen consumption is affected by personal factors (Lavoiser and Seguin, 1793). From this collection of experiments, it was concluded how the physiological process of metabolism worked with respiration and dissipating heat (Karamanou et al., 2013). The process of analyzing oxygen consumption and carbon dioxide production led to the ability to quantify energy expenditure. This process later became known as indirect calorimetry.

Over the next century, there were many advancements in indirect calorimetry. Cesav Mansuete Despretz and Pierre Louis Dulong (1824) designed a water bath calorimeter that determined temperature of a known water mass. Henrik Scharling designed a water-cooled calorimeter in 1849 that was used for large animals and humans. The same year, Henry Regnault and Jules Reiset designed and perfected a closed-circuit system that measured O_2 consumption and CO_2 production (Poncet and Dahlberg, 2011; Mtaweh, et al., 2018). His calorimeter had a chamber in which a living animal was placed. There was a tube going to the chamber that supplied oxygen and a tube leaving the chamber that took expired air past materials that would absorb CO2. From their studies, they concluded differences in oxygen transformation regarding food consumption (Poncet and Dahlberg, 2011).

Around this same time, Carl von Voit and Max von Pettenkofer modified the calorimeter to allow for better quantification of O_2 consumption and CO_2 production (Kenny et al., 2017). They were able to determine amounts of carbon, nitrogen and O_2 that were metabolized based on diet. They determined that in metabolism one of the three substrates, carbohydrates, fats or proteins were oxidized (Mtaweh et al., 2018). Shortly after in 1894, Rubner validated the use of indirect calorimetry against direct calorimetry. He placed dogs into a chamber and measured the thermal gradient to measure heat loss. The calorimeter had a Pettenkofer respirator attached that was able to measure the gas ratios. He found that between his direct measure of heat and the gas exchange there was only a 0.2% difference in energy expenditure (Kenny et al., 2017; Mtaweh et al., 2018).

In 1901, Atwater, with the assistance of Rosa, Langworthy, and Benedict, designed a human calorimeter with a respiration chamber (Kenny et al., 2017). They called their design a respiration calorimeter and used it to further validate the relationship between direct and indirect calorimetry. The Atwater-Rosa calorimeter was used to measure the non-protein carbon in respiration and calculate the O_2 absorbed for metabolism of fats and carbs (Mtaweh et al., 2018). These findings allowed for energy expenditure to be estimated by the respiratory quotient (RQ) rather than using nitrogen and carbon. It also allowed for the creation of standard formulas to be created on the caloric equivalents of O_2 and CO_2. It was Atwater who eventually coined this process as indirect calorimetry. Following this research, Benedict went on to create a smaller instrument for measuring oxygen called the "Benedict apparatus" in 1907. This device went on to be used in many hospitals for use in the clinical setting and in studying metabolism in newborns (Benedict, 1919; Benedict and Talbot, 1915), athletes (Benedict and Smith, 1915), vegetarians (Benedict and Roth, 1915), children (Benedict, 1919), men and women, and starving people (Benedict and Roth, 1918).

Following the production of the Benedict apparatus, calorimetry began to be used more for metabolism. In 1910, Dubois and Veeder studied metabolism in those with diabetes using the Sage calorimeter, which allowed them to measure energy expenditure and heat loss over an extended period of time. While using the Sage calorimeter, they found an average error of 0.9% for heat loss, 0.6% for CO_2 production, and 1.6% for O_2 consumption (DuBois and Veeder, 1910). It was Dubois and colleagues who took the idea of indirect calorimetry and were able to standardize values at rest in both healthy and sick adult populations (Peabody, Meyer, and Dubois, 1916) and later in infants and children (Dubois, 1916). This became known as basal metabolic rate (BMR) or the energy used in a post absorptive state without movement (Peabody et al., 1916), In 1919, Benedict and his team validated a set of equations used to predict BMR in a group of 239 men and women. This equation became known has the Harris-Benedict equation and relies on the height, weight, and sex to predict BEE (Haugen, Chan, and Li, 2007).

In 1906, Douglas and Haldane, developed a semi-portable gas analyzer for field experiments. The analyzer was later improved by Douglas in 1911, making it more readily portable, and was able to be used during activities such as cycling, canoeing, and hiking. However, many areas of error have been found for the bag apparatus more recently, when compared to direct measures. Carter and Jeukendrup (2002) found that the Douglas bag resulted in significantly lower VCO_2 values than more recent methods of gas analyzation. The Douglas bag was used to expand on maximal oxygen intake and the term oxygen debt.

In 1922, Archibald Hill and Hartley Lupton used the Douglas bag with a discontinuous incremental speed protocol to record values of O_2 consumption and CO_2 production. This was the foundation for maximal oxygen consumption (VO_{2max}) and it was the first instance in which it was determined that those of higher fitness had considerably higher O_2 consumption during intense exercise (Hale, 2008). Additional methods of EE measurement have been created and validated using direct or indirect calorimetry as the criterion measurement.

Doubly Labelled Water. The method of doubly labelled water (DLW) was developed in the mid-1900's by Lifson and colleagues, but it wasn't until the 1980's that the technique was applied to humans. This method is non-invasive and causes little stress onto subjects. Because it is non-invasive, it has become the gold standard of EE measurement in free-living conditions (Westerterp, 2017). DLW eliminates the change in behavior seen with other methods of EE measurement such as wearable devices (Westerterp, 2017). In the DLW technique, a standardized amount of $^2H_2{}^{18}O$ and H_2O are ingested. Energy expenditure (CO_2 production) is measured by calculating the difference in elimination rates of the two isotopes in urine, blood, or saliva.

Schoeller and VanSanten (1982) published the first study validating DLW in humans. In their research, they compared the dietary intake plus the change in body stores to DLW in four adults (three males, one female). They found a difference in EE of 2 ± 6% between the two methods, successfully validating the technique in humans. In 1988, DLW was validated again against indirect calorimetry, using a respiratory chamber and differing doses of labelled water. Nine healthy, adult males remained in the chamber for 4 days and the difference in EE ranged from 8 ± 9% for the low dose and 4 ± 5% for the high dose, hence further validating the method (Schoeller et al., 1986)

Later research validated the use of DLW as a means for determining EE with an average error rate of 0-10%. Studies were done in different populations including adults (Seale, Rumpler, Conway, and Miles, 1990), infants (Roberts, Coward, Schlingenseipen, Nohria, and Lucas, 1986; Roberts et al, 1988; Jones et al., 1987), and lean and obese subjects (Ravussin, Harper, Rising, and Bogardus, 1991). In addition to validating REE, DLW had been validated in measuring EE during physical activity. These activities include biking (Westerterp, Saris, van Es, and ten Hoor, 1986), activities of high- and low-intensity (Westerterp, Brouns, and Saris, 1988), and occupational work (Singh et al., 1989).

Due to the accuracy of the technique, DLW is frequently used as the method of EE measurement in field studies. DLW does have disadvantages that make in inapplicable for large population studies. Limitations of DLW include the expense of the isotopes that are used, the complexity of analysis, and that it is an estimation made from the measure of CO_2 production (Goran, Poehlman, and Johnson, 1995).

Subjective Measures of Physical Activity

Due to the ability of subjective measures of physical activity to be cost-efficient and useable for large population studies, they are a very common report of activity levels (Vanhees et al., 2005). Subjective measurements include any type of survey or questionnaire including: self-report, interview assisted, and diaries. However, these types of measurements can lead to highly inaccurate estimations of physical activity due to over reporting or inability to recall activity (Sallis and Saelens, 2000). Surveys and questionnaires rely on one's ability to recall past activity and can include measures of mode, duration, and frequency (Sylvia, Bernstein, Hubbard, and Keating, 2014). They can range in time frame and administration.

Self-Report Questionnaires. One method of subjective measure is self-report questionnaires. In 2007, Maddison et al. compared results from the International Physical Activity Questionnaire (IPAQ) and the New Zealand Physical Activity Questionnaire (NZPAQ) to doubly labelled water. In 36 adults, they found an underestimation of physical activity of 27% with the IPAQ and 59% with the NZPAQ. However, Bonnefoy et al. (2001) looked at 10 different physical activity questionnaires and compared the results to DLW. They concluded that a few of the studied questionnaires showed validity in measuring PA levels, but that many did not.

Food and Activity Diaries. Another common subjective measure of physical activity includes food and activity diaries (Sylvia et al., 2014). In diaries, individuals can recall activity as it occurs, rather than recalling it making diaries slightly more accurate (Sylvia et al., 2014). Bratteby, Sandhagen, Fan, and Samuelson (1997) compared a 7-day activity diary to doubly labelled water in adolescents. They found a mean difference of 1.2% between the diary and DLW and concluded that the activity diary method was a valid method of EE. Similarly, Koebnick et al. (2005) found a good correlation between energy intake recorded via a food log and EE by DLW.

Direct Observation. The third most common subjective measurement type is direct observation. In direct observation, an individual observes and records physical activity (Sylvia et al.,2014; Vanhees et al., 2005). Direct observation is beneficial for children due to their lack of being able to recall activity and allows for direct details of the PA. Direct observation has been found to have a high level of agreement of PA levels between observers

(Sylvia et al., 2014). Puhl, Greaves, Hoyt, and Baranowski (1989) used a Children's Activity Rating Scale (CARS) to classify energy expenditure in children during field observation. VO_2s were taken to classify the activity as relative intensities. It was reported that the CARS observation was reliable in evaluating physical activity levels in children and that there was an 84% agreement among observers. Additionally, Bailey et al. (1995) used the modified fargo activity time sampling survey to record observations of children every three seconds for four-hour time blocks and then use VO_2 to classify intensity. This led to a reliable method of PA classification and a strong agreement of 91% between observers coding of PA.

Objective Measures of Physical Activity

Due to the reporting bias of subjective measures of physical activity, a more accurate measure of physical activity was needed (Troiana et al., 2008). It has been determined that objective measures of physical activity are more precise than subjective measures, due to them being able to recall all physical activity, not just the activity that an individual can remember. Objective measures are able to report on different areas of PA including frequency, intensity, time, and in some instances, even type of activity (Silfee et al., 2018). Heart rate (HR) monitoring, pedometers and accelerometers are three common forms of PA monitoring.

Heart Rate Monitoring. Heart rate monitoring relies on the assumption that an individual's heart rate increases linearly with intensity (O_2 consumption; Andre and Wolf, 2007). However, research has found that a calibration calculation is needed for heart rate to accurately report energy expenditure (Morio, Ritz, Verdier, and Montaurier, 1997; Racette, Schoeller, and Kushner, 1995; Schulz, Westerterp, and Bruck, 1989). HR measurements have been found to overestimate high intensity EE and underestimate low intensity EE (Lof, Hannestad, Forsum, 2002). The lack of accuracy at low intensity is due to variation in heart rate due to individual factors such as caffeine intake, age, fitness level, temperature, nutrition, sleep, stress or illness (Andre and Wolf, 2007; Hills, Mokhtar, Byrne, 2014; Hettiarachchi, Hanoun, Nahavandi, and Nahavandi, 2019). HR monitors can be worn on the wrist, arm, and chest (Hettiarachchi et al., 2019).

Polar® HR monitors are among the most popular brands of available devices. In a study by Hettiarachchi et al. (2019), the Polar® OH1 was validated against ECG during moderate- and high-intensity exercise. It was reported that the device had a mean bias of 0.27-0.33 bpm, hence making it a valid measurement in both the lab and field setting for HR. In an additional study by Engstrom, Ottosson, Wohlfart, Grundstrom and Wisen (2012), the Polar® RS400 was validated using cycle ergometry. There was a significant correlation found between HR measured by the Polar® device and the ECG (mean difference = 0.7-4.3 bpm).

Chest strap HR monitors rely on electrocardiac sensors and have been validated against electrocardiograms (ECGs) for accurately reporting HR (Terbizan, Dolezal, and Albano, 2002; Engstrom et al., 2012). Wrist-worn devices use photoplethysmography (PPG) to monitor heart rate. This technology has been deemed valid to ECG (Temko, 2017) but has inconclusive findings based on how photosensitive an individual's skin is and how much artifact there is during activity (Spierer, Rosen, Litman, and Fujii, 2015). HR monitors have then been validated for measuring EE against DLW in free-living conditions in healthy adults (Rafamantanantsoa et al., 2002; Schulz et al., 1989; Davidson, McNeill, Haggarty, Smith, and Franklin, 1997), those of different body masses (Racette et al., 1995; Lof et al., 2003), and geriatric populations (Morio et al., 1997; Rothenberg, Bosaeus, Lernfelt, Landahl, Steen, 1998).

Limitations of chest strap HR monitors include comfortability and conductivity. While wrist worn monitors are smaller and more convenient, their accuracy is not as high as a chest worn device. In a free-living or non-lab setting, compliance of actually wearing the device is highly differential among subjects (Hettiarachchi et al., 2019).

Pedometers. Pedometers are small devices that use a spring to measure movement in the vertical plane. More often than not, pedometers are used to measure step counts and are usually worn on the hip (Vanhees et al., 2005). After having a measured step count, it can be converted to distance using stride length. Pedometers are best for measuring activity in the vertical plane such as walking and running (Sylvia et al., 2014; Tudor-Locke, Williams, Reis and Pluto, 2012) and are much more affordable than other monitoring devices (Tudor-Locke et al., 2012).

Pedometer measurements are most commonly validated in the lab setting against direct observation. This means that an investigator will manually count steps and compare to the output of the device. This type of validation has been done in children (Nishikido et al. 1982; Kilanowski, Consali, and Epstein, 1999) and adults (Hoodless, Stainer, Savic, Batin, Hawkins, and Cowley, 1994; Selin, Winkel, and Stockholm-MUSIC I study group, 1994; Bassett et al., 1996) in the laboratory/field setting.

When looking at pedometers output compared to self-reported physical activity the findings are much more inconsistent. A study by Nishikido et al. (1982) reported that there was no significant correlation between mother’s and teacher’s report of PA levels of kindergarteners and the measurement from the pedometers. Similarly, Zahiri, Schmalzried, Szuszczewicz, and Armstutz (1998) concluded that in a study of joint replacement patients, there was no significant correlation between the patient's report of their own activity and the measurement from the pedometers. However, Edelman and Smits (1984) had 84 subjects wear a pedometer and keep an activity diary for 5 days. They reported a strong correlation between the diary and pedometer, finding the two to be useful to use together.

Lastly, pedometers have been used to predict energy expenditure. This is done using regression equations with input of steps taken. When compared to the doubly labelled water method, multiple studies have reported no significant correlation with pedometers (Leenders, Sherman, Nagaraja, and Lawrence, 2001; Fogelholm et al., 1998). Studies have also compared EE estimated by pedometers to indirect calorimetry in the field setting, which have inconclusive results. Bassett et al. (2000) reported a nonsignificant correlation between indirect calorimetry and pedometry, while Eston, Rowlands, and Ingledew (1998) concluded a significant correlation, even suggesting pedometry be used for large population studies. Irimagawa and Imamiya (1993), looked at the relationship between HR EE and pedometry and found the two to be significantly correlated. Also looking at HR, Kashiwazaki, Inaoka, Tsuguyoshi, and Kondo (1986) found significant correlations in factory works while commuting, but no correlation when at home, signifying the use of pedometers to access activities such as walking, but not sedentary behavior.

The ability of pedometers to report levels of physical activity is high, but results of EE aren't as clear. Pedometers are an accepted device for measuring activity levels during movement, but are not suggested for EE for those that spend extended times in sedentary behavior. Additionally, pedometers only measure up and down, vertical activity, so it would not be suggested for measurement for an activity such as swimming or cycling. Pedometers are also unable to measure characteristics of activity (duration, frequency, or intensity), they can cause behavioral shifts in subjects, and are unable to store large amount of data, limiting their ability to be used for all studies (Sylvia et al., 2014).

Accelerometers. Within the last few decades, accelerometers have become increasingly more popular. Accelerometers are similar to pedometers, except they can be multiaxial, meaning they are able to measure movement in multiple planes (Vanhees et al., 2005). Accelerometers do not use the spring mechanism that pedometers use, but a piezoelectric transducer and microprocessors. These mechanisms allow for magnitude and direction of acceleration to be measured and for those measurements to be converted into EE, activity type, intensity, and duration (Sylvia et al., 2014; Vanhees et al., 2005). Commonly, they are worn on the hip, around the waist, on the back, or more recently, around the wrist (Vanhelst et al., 2012). Accelerometers have become very popular due to their small size, their ability to store large amounts of data, and their ability to measure multiple characteristics of activity (i.e. frequency, intensity, time, type; Freedson and Miller, 2000; Sylvia et al., 2014; Vanhees et al., 2005).

A review of 47 studies conducted in 2008 determined that the most commonly used accelerometer from 2006-2016 was the Actigraph® (Silfee et al., 20018). Accelerometry has been validated for predicting EE from physical activity using the golden standard of DLW. Assah et al. (2009) conducted a study comparing physical activity energy expenditure from a hip worn Actigraph® accelerometer to energy expenditure from DLW in 33 adults in a free-living environment. Their results concluded that the physical activity energy expenditure (PAEE) from the accelerometer was significantly correlated to the PAEE from the DLW technique over 7-days. Similarly, Johansson, Rossander-Hulthen, Slinde, and Ekblom, (2006) compared the MTI Actigraph® EE output to doubly labelled water. Twenty-seven subjects wore an accelerometer for 14 days. They concluded that there was no significant difference between the EE measured by the Actigraph® and the DLW.

In the lab setting, accelerometers have been validated using indirect calorimetry. Kumahara et al. (2004) found the 24-hr TEE of 79 Japanese subjects using a room respiratory chamber. Additionally, subjects were wearing a Lifecorder® accelerometer. Subjects performed two 30-minute walking exercises on a treadmill. In an additional part to the study, 10 men ran at three speeds and walked and six speeds for four minutes each. They concluded that the accelerometer was accurate when measuring energy expenditure during the activities, but significantly underestimated the EE when the subjects were sedentary. Similarly, Vanhelst et al. (2012) aimed to validate the Vivago® wrist-worn accelerometer against indirect calorimetry. Twenty-one subjects performed six, 10-minute periods of activity at differing intensities (sedentary, light, moderate, and vigorous). During activity, subjects wore a wrist accelerometer and were hooked to a metabolic cart. They reported a significant correlation between the accelerometer EE output and the oxygen consumption output.

Many accelerometers have been validated against gold standards of EE measurement making them common devices for measuring PA. Despite their accuracy, accelerometers can be highly expensive, causing them to be impractical for large epidemiological studies (Sylvia et al., 2014). Additionally, they require researchers to be well-versed with technology and can cause subject reactivity in studies (Vanhees et al., 2005). Furthermore, accelerometers were found to be more accurate in determining PAL and EE when paired with an additional method of measurement (i.e. heart rate; Johansson et al., 2006; Brage et al., 2015; Chang, Lin, Ho, and Huang, 2010).

Wearable Fitness Technology. Following the realization that multiple methods of measurement together was most accurate for quantifying PA and the advancements in technology, multi-sensor devices were created and commercialized. Worldwide, the revenue from wearable fitness trackers (WTFs) was 2.57 billion dollars in 2018 and is projected to increase to 3.33 billion by 2022 (Liu, 2019b). Three of the highest selling WFTs are Fitbit®, Apple®, and Garmin® (Liu, 2019a). The most recent of these devices are able to record and report to users their steps, HR, and active minutes. Using these measures, these devices are pre-programmed with factory developed algorithms that can then estimate other measures such as energy expenditure, sleep stages, and intensity levels. These devices are then able to sync to smartphone applications or computers for users to access information across devices. (Liu, 2019a).

The use of wearable fitness technology has been used as a behavioral modification for weight loss and weight maintenance. Several findings have concluded that the use of these devices in interventions has proved to be successful. For example, Cadmus-Bertram et al. (2015) recruited 51 inactive women and split them into a control group that received a basic pedometer and an experimental group that received a Fitbit® tracking device. All women were asked to participate in 150-minutes of moderate/vigorous physical activity (MVPA) per week. Upon completion of the intervention, the group that received the Fitbit® had significant increases in MVPA and steps taken, where the pedometer group did not. In a similar study done by Ashe et al. (2015), 25 participants were split into an intervention and control group. The control received group-based education while the intervention group received education, support, PA prescription, and a Fitbit® activity monitor. After six months of the intervention, the intervention group had significantly higher step counts/day, greater weight loss, and reduced diastolic blood pressure than the control group. Due to the use of WFTs in behavior/lifestyle modification, being highly accurate is imperative.

In the literature, validation studies on wearable devices can be divided into several categories based on protocol. Protocols include discontinuous-short protocols, discontinuous-long protocols, and continuous-short duration protocols. Discontinuous refers to a protocol that is broken up into several bouts of exercise, while continuous is just one bout. Long-duration refers to a protocol of 30 minutes or more, while short duration is anything less than 30 minutes (Schmidt, Biwer, and Kalscheuer, 2001; Daley and Welch, 2004). Protocols also differ based on intensity. Some protocols consist of a set intensity or speed, while others allow participants to choose their own intensity.

Discontinuous Protocol-Short Duration. A study by Diaz et al. (2015) compared the energy expenditure from two Fitbit™ One devices placed at the hip and torso and a wrist-worn Fitbit™ Flex to indirect calorimetry. The protocol consisted of a discontinuous, four-phase treadmill test of increasing intensity. Each phase was six minutes and a three-minute rest period was taken between each one. They found a moderate agreement between the EE of the devices and indirect calorimetry. The torso placed Fitbit™ One, hip placed Fitbit™ One, and Fitbit™ Flex had a 9.7%-19.9%, 3.4-12.9%, and 24.5-83.4% error respectively. Reddy et al. (2018) used a similar short duration, discontinuous protocol. A Fitbit™ Charge 2 and Garmin™ Vivosmart HR+ were compared to a portable COSMED analyzer. During the first visit, subjects completed either a maximal treadmill or cycling test followed by a free weight circuit composed of six different exercises for two set and eight repetitions. The second visit was composed of 28-minutes of ADLs broken into 3-minutes, with 5-minute breaks between each activity. After the ADLs, subjects participated in a 27-minute interval training session that was 2 minutes of high intensity activity followed by 2 minutes of low intensity. EE from both the Fitbit™ and the Garmin™ were found to be significantly different from indirect calorimetry for all activities.

Discontinuous Protocol: Long-Duration. Chowdhury et al. (2017) also used a discontinuous protocol to compare EE of the Microsoft™ Band, Apple™ Watch, Fitbit™ Charge HR, and Jawbone™ UP24 to indirect calorimetry. In the first 24-minute block of activity, four activities of daily living (ADL) were done for five minutes of duration each. Subjects then took a 10-15-minute break before continuing on with a 64-minute block of activity. The second protocol consisted of four ten-minute bouts of walking, jogging and cycling. Each activity was separated by a five-minute standing break. Following the visit, each subject took the device home and was asked to wear it for at least 36 hours. They found that the Apple™ Watch reported the most accurate EE. However, the Fitbit™ Charge HR was more accurate in the free-living protocol. Ultimately, none of the devices were strongly correlated to the research devices. Another study by Shcherbina et al. (2017) used a discontinuous protocol of long-duration. EE from an Apple™ Watch, Basis Peak™, Fitbit™ Surge, Microsoft™ Band, Mio™ Alpha 2, Pulse On™, and Samsung™ Gear S2 was compared to indirect calorimetry. During the study, subjects performed a 40-minute block of activity consisting of sitting for five minutes, walking for ten minutes, then transitioning to running for 10-minutes, resting for 1-minute, then transitioning to cycling for ten minutes. At the conclusion of the study, it was reported that no device had an error < 20% and the errors ranged from 27.4% (Fitbit™) to 92.6% (Pulse On™). All devices underestimated EE for both walking and running.

Continuous Protocol: Short-Duration. Kendall et al. (2019) looked at EE estimated by the Basis™ Watch, Fitbit™ Flex, Polar™ FT7, Jawbone™, Omron™ Pedometer, and Actigraph™ compared to indirect calorimetry. For the study, participants completed a maximal graded treadmill test consisted of a continuous protocol at a self-selected pace with gradually increasing grade of 2% every 2-minutes until exertion. They found that during maximal treadmill testing, all devices were significantly correlated with indirect calorimetry, but that correlations were stronger in lower fit individuals (lower intensities). Alike, Dondzilla and Garner (2016) also used a short duration protocol while comparing EE estimations of the Fitbit™ Charge and the Jabra™ Sport Plus wireless earbuds to indirect calorimetry. Subjects performed four 5-minute treadmill stages at varying intensities. Both devices were found to significantly underestimate EE, but when looking at individual stages, both devices were moderately correlated at moderate speeds. One last short duration study by Dooley, Golaszewski, and Bartholomew (2017) compared EE of the Apple™ Watch, Fitbit™ Charge HR, and the Garmin™ Forerunner 225 to gas analysis of the metabolic cart. Subjects performed 4-minute stages of light, moderate, and vigorous activity on a treadmill. Both the Apple Watch and the Garmin overestimated EE at all stages and had error ranges of 14.07-210.84% and 30.77-155.05% respectively. The Fitbit™ Charge HR overestimated at all stages except vigorous and had an error range of (16.85-84.98%).

Self-Selected Intensity. Stackpool, Porcari, Mikat, Gillette, and Foster (2015) allowed their subjects to self-select the intensity of the activities. When comparing the Nike™ FuelBand, Fitbit™ Ultra, Jawbone™ UP, BodyMedia™ Fit Core, and Adidas™ MiCoach to indirect calorimetry, a two-day protocol was used. The first visit was comprised of a 20-minute walk, followed by a ten minute rest, and then a 20-minute run at self-selected paces.

During the second visit, participants performed 20 minutes at a self-selected intensity on an elliptical and then a series of agility drills. It was reported that none of the wearable devices were accurate across all of the activities, however the Fitbit™ Ultra only estimated a significantly different EE for the agility drills. Similarly, Montoye et al. (2017) compared the EE estimated by the Fitbit™ Charge HR and the Hexoskin™ smart shirt against the Parvo™ medabolic cart. A protocol consisting of 14 exercises was used including 11 in the lab and three on a track. Participants were allowed to self-select paces within a range for each walking, jogging, and cycling activity as long as their pace remained constant throughout the activity. Each activity was performed for five minutes and participants were able to take a break after each activity. An overestimation of EE was reported for both the Fitbit™ (43.7%) and the Hexoskin™ (27.9%). Bai, Hibbing, Mantis, and Welk (2018) looked at the correlation between Apple™ Watch 1 and Fitbit™ Charge HR and the Oxycon Mobile. The study consisted of three different stage with a 5-minute break between each stage. The first stage was 25 minutes of sedentary behavior where participants were required to sit at a desk but were allowed to choose an activity such as reading or using a computer. Next, subjects were asked to walk or jog at a self-selected pace for 25 minutes on a treadmill. Subjects were able to change pace throughout the stage. The final stage was comprised of 25 minutes of ADLs. Subjects were given little direction and were able to choose from a number of activities. From this study, the mean error in the Charge was 32.9% and the Apple™ Watch was 15.2%.

Missing from these validation studies is a methodology utilizing a continuous, long-duration exercise protocol. Additionally, to our knowledge, there is no research comparing the accuracy of wearable devices between a short- and long-duration protocol. It is important to validate using this long-duration protocol because the recommended physical activity is 30 minutes a day, 5 days per week, or 150 minutes of moderate-to-vigorous PA (ACSM, 2018b). Individuals are likely to engage in 30-minute bouts of exercise to meet this requirement. Ensuring that estimates of EE reported by these monitors are accurate is crucial, so consumers are able to adjust caloric intake appropriately. This will assist in weight management and potentially increase weight loss or weight maintenance among users.

Chapter 3: Methods

Participants

Thirty-one adults from Eastern Michigan University and surrounding areas were recruited as a convenience, non-probability sample. Participants had to be between the ages of 18 and 35 and had to be able to run for at least 30 minutes. An informed consent was given prior to testing. Only those not considered at risk by the Physical Activity Readiness Questionnaire (American College of Sports Medicine, 2019) were able to participate in the study. Additionally, following Visit 1, subjects were assigned to a fitness group based on their cardiorespiratory fitness (CRF) level. To determine these groups, relative maximal oxygen consumption (VO_{2max} (mL/kg/min)) was categorized based on percentile according to the ACSM guidelines. These percentiles are a method of standardization based on gender and age. A median split of this population was done, splitting participants at the 65th percentile. For this study those in the 65th percentile or below were considered moderately fit, while those above the 65th percentile were considered high fit. Individuals below the 35th percentile were excluded from the study. The rationale behind splitting subjects based on fitness level was to determine if the accuracy of energy expenditure varies based on CRF.

Procedure

Equipment for each visit included a FitBit™ Inspire HR activity monitor, True™ treadmill, Parvo™ metabolic cart, Polar™ heart rate monitor, Tanita™ BWB-800 scale, and Detecto™ stadiometer. Each participant completed three visits to the Eastern Michigan University Running Science Laboratory. Each visit was held at least 72 hours apart, but all three visits were completed within a two-week time period. Visits 2 and 3 consisted of either a short or long-duration running protocol. Whether the participant completed the long or short duration protocol following the incremental maximal treadmill test was randomized and counterbalanced using a random number generator.

A Fitbit™ Inspire HR activity monitor was worn by the participant in Visits 2 and 3. The Fitbit™ Inspire HR records steps taken, daily calorie expenditure, exercise calorie expenditure, heart rate, sleep stages, active minutes, pace, and distance traveled (Fitbit, 2019). The Fitbit™ devices predicted EE with an algorithm that combines BMR using factors such as age, height, weight and sex and your daily activity with HR (Fitbit, 2019). Predicted calorie expenditure by the activity monitor during the exercise portion of each visit was taken and accuracy was compared to the energy expenditure from the Parvo™ metabolic cart.

Visit 1: Incremental Graded Maximal Treadmill Test. During the first visit, a continuous, incremental maximal treadmill test protocol, adopted from Kendall (2019), was completed on a treadmill to determine VO_{2max}. Oxygen consumption and carbon dioxide expiration was measured via indirect calorimetry using a Parvo™ metabolic cart. A mouthpiece and nose clip were worn to ensure accurate measurement of gas exchange. The metabolic analysis was measured breath by breath and aggregated to 15 seconds and the subjects VO_2 (L/min and mL/kg/min), METS, and energy expenditure (kcals) were recorded. For the participant to reach maximal oxygen consumption, two of the three following criteria were met: (a) oxygen consumption leveled off despite increasing work rate (b) the respiratory exchange ratio (RER) was ≥ 1.1 3 and (c) heart rate was no less than 15 beats below age-predicated max (HR_{max} = 220 - age; Kline et al., 1987).

The treadmill test protocol was comprised of two-minute stages. Participants performed a warm-up on the treadmill for 3 minutes at 3.0 mph and 0% grade. Following the warm-up, the treadmill was set to a self-selected running pace, but remained at 0% grade for the first stage. The remainder of the test was conducted at this speed and every two minutes, the grade of the treadmill was increased by 2%. Participants were verbally encouraged to continue the test for as long as possible. Once the participant reached exhaustion, the test was concluded. The mouthpiece and nose clip were removed, and the metabolic analysis was completed by the Parvo™. The treadmill was returned to 0% grade and 2.5 mph and participants performed a cool down for five minutes or until HR fell below 120 bpm.

The participant was given a ten-minute break following the cool down. During this time, the VO_{2max} report from the metabolic cart was used to determine if a true max was reached. To meet the first criteria, the VO2 of the last two minutes was averaged and there was no more than a 2 ml/kg/min difference. For the second criteria to be met, the RER column was looked at and it was determined if the RER was over 1.1. For the last criteria to be met, the HR column was used to determine if the highest heart rate was less than 15 beats from the age-predicted max. If two of the three criteria are met, it was determined that the participant reached a true VO_{2max}. Seventy to seventy-five percent of the participants VO_{2max} was calculated using the following equations: (VO_{2max}) x (0.70) and (VO_{2max}) x (0.75). This value was used to determine the running speed to be used in the remaining visits.

Following the 10-minute break, the running speed was found for the remaining two visits (70-75% of VO_{2max}). Three minute stages were used to ensure steady state is reached. The participant completed a three minute warm up at 3.0 mph and then the treadmill speed was increased to 6.0mph for an additional three minute stage. The speed of the treadmill was increased or decreased by 0.5 mph for each three minute stage, until a steady state of seventy to seventy-five percent of the participants VO_{2max} was reached. This speed was recorded for use at future visits. The participant then completed a 3-minute cool down at 2.0 mph or until heart rate fell below 120bpm. Randomization was used to determine if the participant was to complete the long-duration or short duration protocol during visit two using a random number generator.

Visit 2 or 3: Long-Duration. Participants returned to the lab at least 72 hours, but no longer than 2 weeks, after the previous visit. The long-duration protocol was completed at a moderate-to-vigorous intensity, determined by ACSM's exercise guidelines, to be 70% to 75%. This intensity (speed) was calculated from the previous visit's maximal treadmill test. The participant was connected to the Parvo™ metabolic cart for the duration of the visit. Prior to the test, each subject's characteristics (birthday, weight, height, and gender) were inputted into the Fitbit™ mobile app and the device was synced to the Fitbit™ mobile app using IOS 13 software. Participants then placed the Fitbit™ on their non-dominant arm, one to two fingers above their wrist (Fitbit, 2019). The treadmill was set to the speed found during the second half of the initial visit. Simultaneously, the activity monitor was set to the treadmill exercise setting, the Parvo™ metabolic cart test was started, and the subject began running. The participant continued to exercise at this intensity for thirty minutes. The Fitbit™ activity monitor and Parvo™ metabolic cart were both paused at the thirty-minute time stamp and the participant stepped to the sides of the treadmill. Following the exercise stage, the mouthpiece and Fitbit™ activity monitor were removed. The participant completed a three minute cooldown at 2.0 mph or until heart rate fell below 120 bpm. Energy expenditure was recorded from the metabolic cart text report and was taken from the Fitbit™ exercise report after the device was synced to the mobile app.

Visit 2 or 3: Short Duration. Participants returned to the lab at least 72 hours after the previous visit, but no longer than two weeks from the initial visit. The same protocol as the long-duration visit was used except the participant only continued to exercise at this intensity for ten minutes. Following the exercise stage, the mouth piece and Fitbit™ activity monitor were removed. The participant completed a 3-minute cooldown at 2.0 mph or until heart rate fell below 120 bpm. Energy expenditure was recorded from the metabolic cart text report and was taken from the Fitbit™ exercise report after the device was synced to the mobile app.

Statistical Analysis

Descriptive statistics were used to describe participants' characteristics (gender, age, height, weight, VO_{2max}, VO_{2max} percentile rank, speed, and EE measurements) and were reported as mean (SD). Pearson correlations for short- and long-duration were used to assess group-level associations between measured EE (criterion measure) and predicted EE from the activity monitor. A paired samples t-test was performed to determine group differences in descriptive characteristics such as age, height, weight, VO_{2max}, and running speed and between measured EE (criterion measure) and predicted EE from the activity monitor A repeated measures ANOVA was used to determine the interactive effect of measurement mode and duration on measured EE (criterion measure) and predicted EE from the activity monitor. A mixed-measures ANOVA was used to determine the interactive effect of measurement mode and fitness level on the of energy expenditure estimates from both the metabolic cart and Fitbit™ activity monitor. Bland-Altman plots were created to display levels of agreement between measurement modes. Statistical significance was determined at a p-value of $p < 0.05$. All statistical analyses were performed using IBM SPSS statistical software version 26.

Chapter 4: Results

Descriptive Characteristics

Of the 31 participants that were recruited to participate in the study, 25 completed all three visits (81%). Four participants that did not return for all three visits and two participants who did not reach the criteria for VO_{2max} were excluded from the study. Of these 25 participants, the majority were male (60%) and the mean age was 23(5.0) years old. The average VO_{2max} of the participants was 50.4(7.0) mL/kg/min. The average speed in which participants were running at 70-75% of their VO_{2max} during both the short- and long-duration protocols was 6.8(0.8) mph. Characteristics of the sample can be found in Table 1. Descriptives are represented as mean(SD) or percentage. To create moderate- and high-fitness groups, each subject was assigned a percentile rank, according to ACSM. The sample was then split at the median VO_{2max} percentile (65th percentile); those in the 65th percentile or below were considered moderately fit, while those above the 65th percentile were considered high fit.

Table 1.
Descriptive Characteristics

	Total Sample (*N* = 25)	Moderate Fitness (*N* = 13)	High Fitness (*N* = 12)
Age (yrs)	23.0(5.0)	23.3(4.2)	22.2(5.4)
Gender (% male)	60%	77%	42%
Weight (kg)	74.2(13.9)	80.7(10.7)	67.2(13.9)*
Height (cm)	173.5(9.0)	173.9(8.0)	172.9(10.7)
VO_{2max} (mL/kg/min)	50.4(7.0)	46.4(5.1)	54.8(6.2)**
VO2 max percentile (%)	70th	55th	85th
Speed (mph)	6.8(0.8)	6.4(0.5)	7.2(0.8)**

Note. Mean (SD) or % and differences between groups (significant at p < 0.05) (** significant at p < 0.001).*

Moderate v. High-Fit Group Differences

The moderate fit group was comprised of 10 males and three females with an age of 23.3(4.2) years, weight of 80.7(10.7) kg, height 173.9(8.0) cm, VO_{2max} of 46.4(5.1) mL/kg/min, and running speed of 6.4(0.5) mph. The high fit group was comprised of five males and seven females with an age of 22.2(5.4) years, weight of 67.2(13.9) kg, height 172.9(10.7) cm, VO_{2max} of 54.8(6.2) mL/kg/min, and running speed of 7.2(0.8) mph. Group characteristics and differences are displayed in Table 1. An independent samples t-test determined group differences in these descriptive characteristics between the moderately-fit and high-fit groups. The independent samples t-test determined that there were significant differences in weight $t(24) = 2.71$, $p = 0.013$, VO_{2max} $t(24) = -3.75$, $p = 0.001$, and speed $t(24) = -3.45$, $p = 0.002$. There was not a significant difference between groups for age ($p = 0.590$) or height ($p = 0.793$).

Measured v. Estimated Energy Expenditure

Table 2 displays the average energy expenditure outputs for both measurement methods during both the short- and long-duration protocols for the entire sample. The mean difference in energy expenditure between the metabolic cart and the Fitbit activity monitor during the short duration exercise was 12.4(12.3) kcals. The mean difference in energy expenditure between the metabolic cart and the Fitbit activity monitor during the long-duration exercise was 20.7(55.1) kcals. For the whole sample, the correlation between measured and estimated energy expenditure for the short duration run was $R = .860$ ($p < 0.001$) and for the long-duration run was $R = .785$ ($p < 0.001$). A paired-samples t-test determined that during the short duration run, the average energy expenditure from the cart (126.8(23.6) kcals) was significantly greater than the average energy expenditure from the

Fitbit (114.4(23.0) kcals), $t(24) = -5.041$, $p < 0.001$. However, for the whole sample during the long-duration run, the average energy expenditure from the cart (398.8(75.7) kcals) was not significantly different from the average energy expenditure from the Fitbit (378.2(88.2) kcals), $t(24) = -1.877$, $p = 0.073$.

Table 2.
Average Energy Expenditure Estimates from Metabolic Cart and Fitbit

Short Duration Run (10-minute)		Long-duration Run (30-minute)	
Metabolic Cart (measured kcals)	Fitbit Inspire HR (estimated kcals)	Metabolic Cart (measured kcals)	Fitbit Inspire HR (estimated kcals)
126.8(23.6)	114.4(23.0)	398.8(75.7)	378.2(88.2)

Bland-Altman plots indicated the differences between the metabolic cart (criterion) and Fitbit activity monitor (y-axis) against average energy expenditure (x-axis). Figures 1 and 2 present these differences with limits of agreement for the 10-minute run duration and 30-minute run duration, respectively. A positive value of mean difference indicates an underestimation of Fitbit activity monitor compared to the criterion measurement. A negative value of mean difference, therefore, indicates and overestimation by the activity monitor compared to the criterion measurement. These differences and the range between the upper and lower limits of agreement are important in determining the validity of the consumer grade device. The larger the limits of agreement, the less accurate the WFT is. The mean difference in energy expenditure between the metabolic cart and Fitbit activity monitor during the short duration run was 12.4 kcals and the mean difference in energy expenditure between the metabolic cart and Fitbit activity monitor during the long-duration run was 20.7 kcals. The 10-minute run displayed more narrow limits of agreement (48.3 kcals or 40% of AEE) compared to the 30-minute run (216 kcals or 56% of AEE). Figure 1. indicates that individuals who expended more kcals during the 10-minute run tended to display larger measurement error.

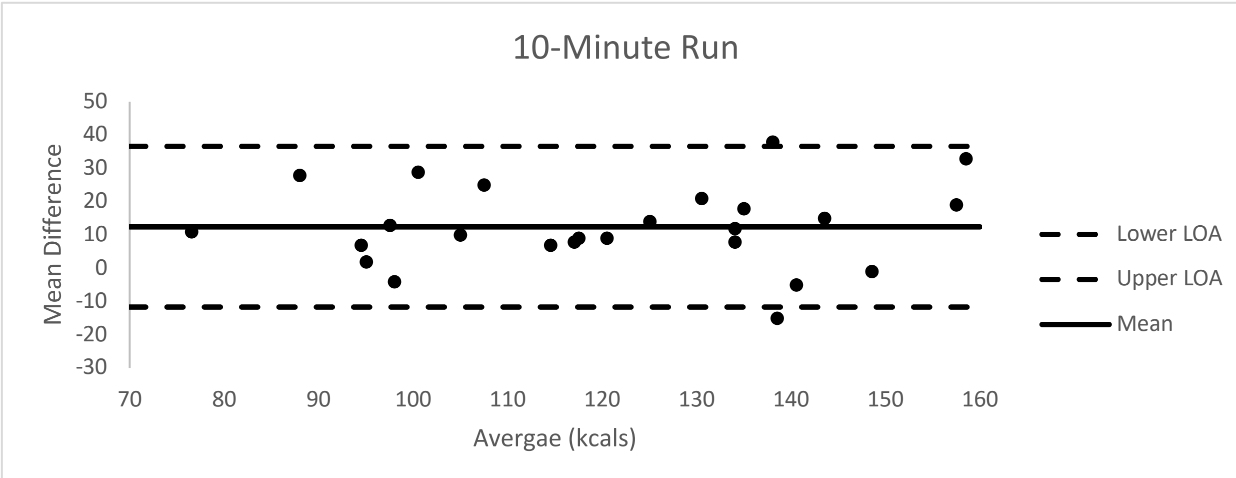

Figure 1. 10-minute run measurement agreement.

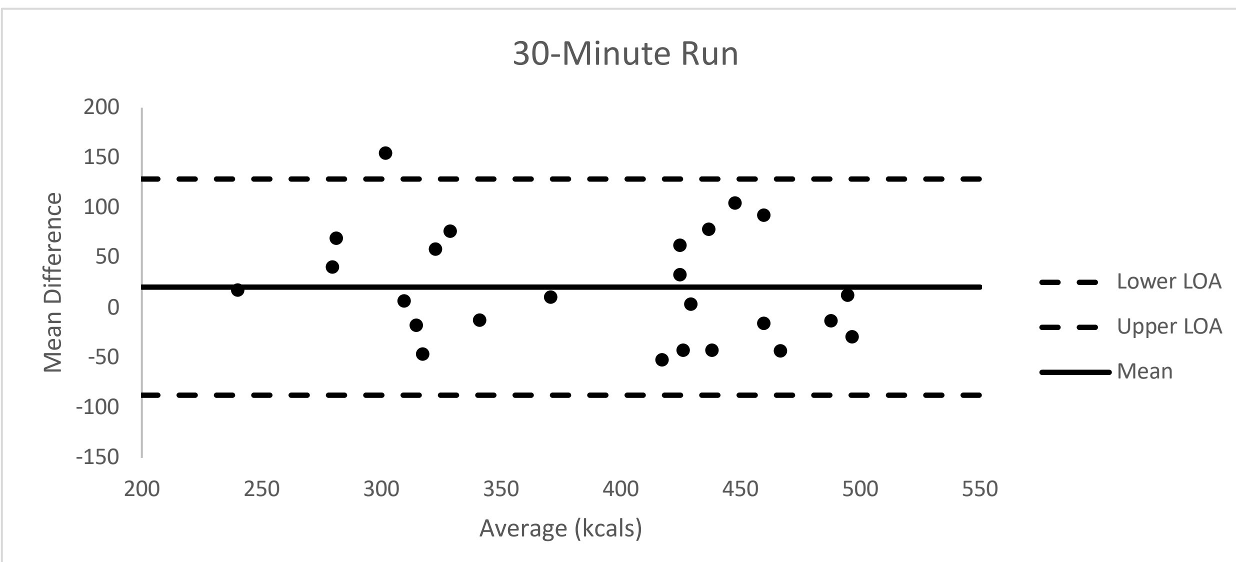

Figure 2. 30-minute run measurement agreement.

Short v. Long-Duration

A 2 (method) x 2 (duration) repeated-measures ANOVA revealed that for the whole sample, the main effects of measurement mode and exercise duration on energy expenditure were significant (p = 0.018; p < 0.001). The interactive effect of exercise duration and measurement mode on estimated energy expenditure was not significant (p = 0.385). Thus, differences in energy expenditure estimates from the two measurement modes were not influenced by run duration for the whole sample. Figure 3 represents the differences in mean between measurement mode for both run durations.

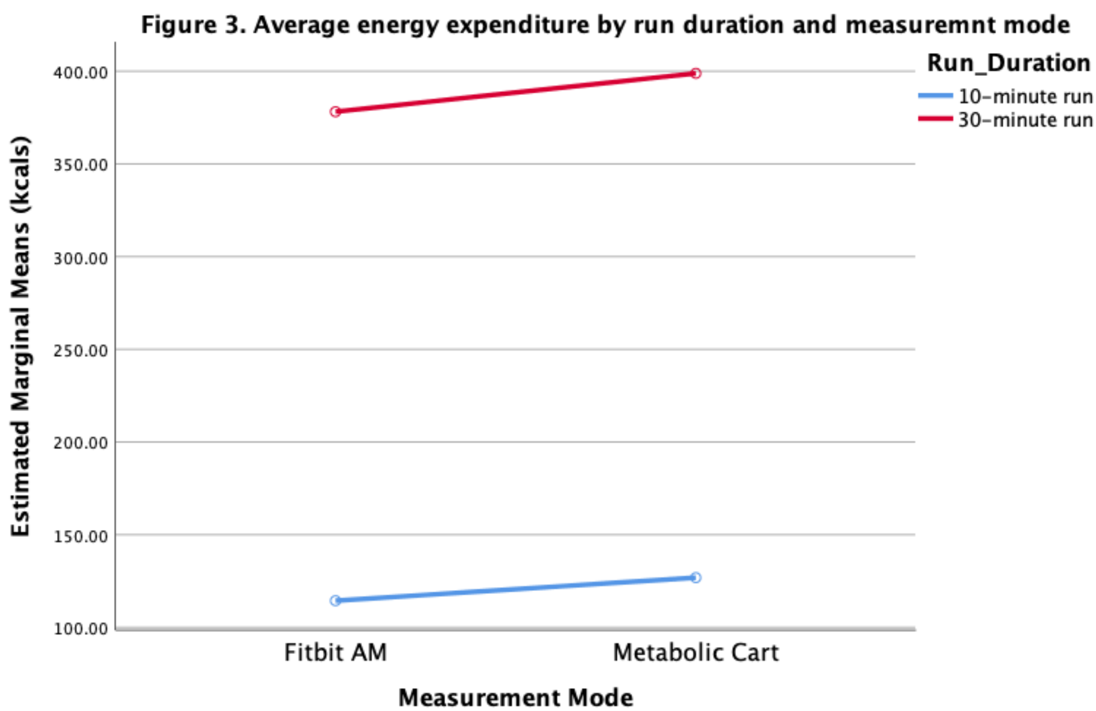

Figure 3. Short duration (1) v. long-duration (2) measurement error.

Moderate v. High Fitness

For the short duration run, a paired samples t-tests revealed that during the short duration protocol there were no significant differences in energy expenditure in the moderately fit group (p = 0.054), but that there were significant differences in energy expenditure in the high fit group (p < 0.001). A 2 (method) x 2 (fitness level) mixed-measure ANOVA revealed that the main effect of measurement mode was significant (p < 0.001) while the main effect of fitness level was not significant (p = 0.078). The interactive effect of fitness level and measurement mode on estimated energy expenditure was significant (p = 0.006). Thus, differences in energy expenditure estimates from the two measurement modes were influenced by fitness level. Figure 4 represents the differences in means between measurement mode for both the moderate and high fit groups.

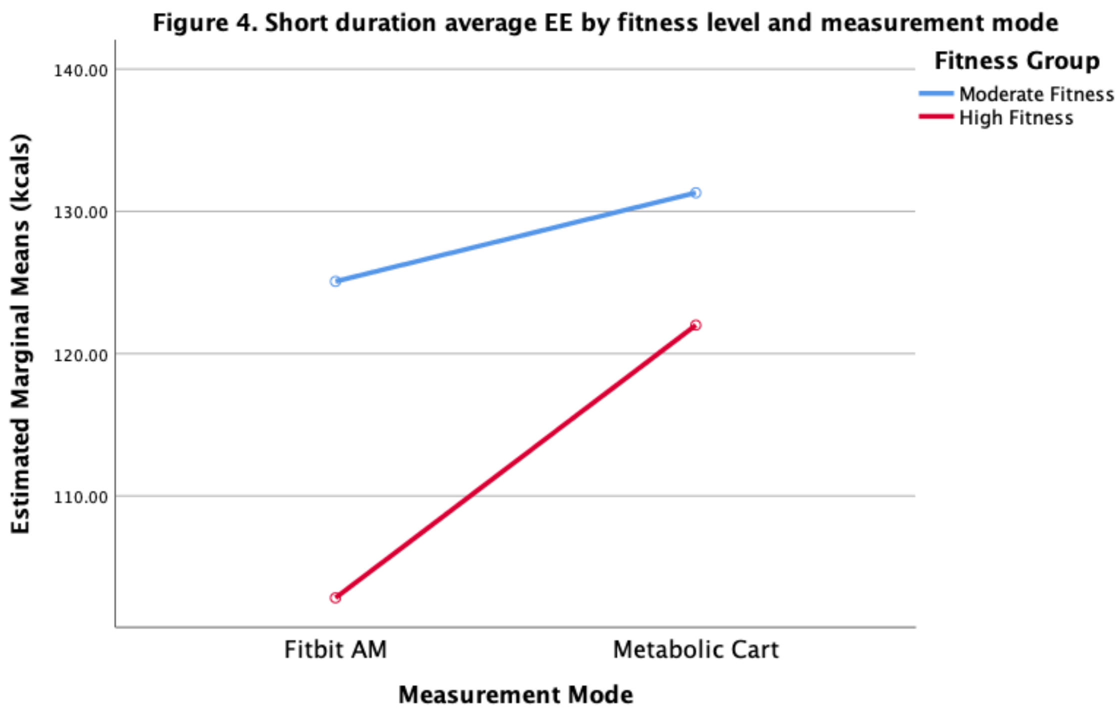

Figure 4. Short duration, moderate v. high fitness measurement error.

For the long-duration run, a 2 (method) x 2 (fitness level) mixed-measure ANOVA revealed that the main effects of measurement mode and fitness level were significant (p = 0.021; p = 0.020 respectively). The interactive effect of fitness level and measurement mode on estimated energy expenditure was significant (p = 0.001). Thus, differences in energy expenditure estimates from the two measurement modes were influenced by fitness level. A paired samples t-tests revealed that during the long-duration protocol there were no significant differences in energy expenditure in the moderately fit group (p = 0.310), but that there were significant differences in energy expenditure in the high fit group (p = 0.003).

Figure 5 represents the differences in means between measurement mode for both the moderate and high fit groups.

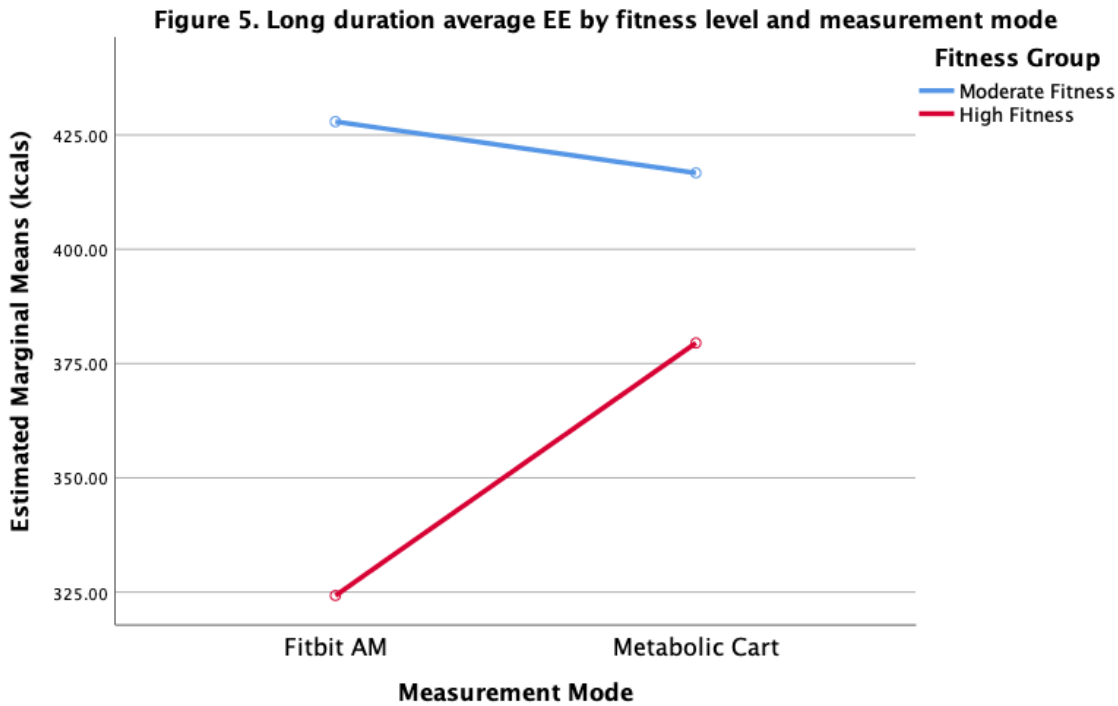

Figure 5. Long-duration, moderate v. high fitness measurement error.

Table 3 represents the mean difference, lower area of agreement, and upper area of agreement for each of the four groups (moderate fitness-short duration, high fitness-short duration, moderate fitness-long-duration, and high fitness-long-duration).

Table 3.
Energy Expenditure Estimation Error Based on Run Duration and Fitness Level

	Mean Difference	Lower Area of Agreement	Upper Area of Agreement
Short Duration-Moderate Fitness	6.2	-13.6	26
Short Duration – High Fitness	19.2	-1.1	39.4
Long-duration-Moderate Fitness	-11.2	-83.2	60.8
Long-duration – High Fitness	55.3	-38.9	149.4

Chapter 5: Discussion

The current study aimed to determine if the Fitbit™ Inspire HR is valid in estimating energy expenditure during both long and short duration exercise at a moderate to vigorous intensity (70-75% of VO_{2max}) compared to a criterion measure (Parvo™ metabolic cart). The secondary aim of this study was to determine if accuracy of energy expenditure estimates differed based on an individual's cardiorespiratory fitness level. It was hypothesized that the activity monitor would not be accurate in predicting overall energy expenditure during both the short- and long-durations of exercise. It was also hypothesized that the activity monitor would be accurate in predicting energy expenditure during both exercise durations for moderate fit individuals, but not for high fit individuals. To our knowledge, this is the first study to compare energy expenditure prediction accuracy during short- and long-duration exercise using the Fitbit™ Inspire HR activity monitor.

Estimated v. Criterion Energy Expenditure

The results of this study revealed strong correlations between the Fitbit activity monitor and criterion measurement for the short duration protocol ($R = 0.860$, $p < 0.001$). Despite this strong correlation, the t-test reported significant differences between the two methods of energy expenditure measurement ($p < 0.001$) during the short duration run. Energy expenditure during this protocol was underestimated by the Fitbit activity monitor. A study conducted by Kendall et al. (2019), reported similar results. They determined that during an exercise bout (~10 minutes) of increasing intensity, the Fitbit activity monitor had a strong correlation with indirect calorimetry ($R = 0.807$). Additionally, like the current study, Kendall et al. (2019) reported an overall underestimation of energy expenditure by the activity monitor. Dondzilla and Garner (2016) also reported an underestimation of energy expenditure, but moderate correlations for the Fitbit activity monitor. The Bland-Altman plots visually displayed the larger margin of error during the 10-minute exercise bout (Figure 1). As the average energy expenditure increased for an individual during the 10-minute bout, so did the measurement error. Thus, those individuals that expended higher amounts of energy showed larger discrepancies between the criterion and predicted EE measurements.

Similar to the short duration protocol, a moderate correlation was found in measurement method during the long-duration protocol (R = 0.785, $p < 0.001$). However, the t-test did not find any significant differences in the measurement methods ($p = 0.073$). These results are contradicted by the literature. In a previous study of long-duration, subjects performed 10-minute bouts of four different exercises of differing intensities (Chowdhury et al., 2017). Across intensities, a mean absolute percent error (MAPE) of 36 ± 22 was reported and the study concluded that the Fitbit activity monitor was not equivalent to research grade devices. Similarly, Shcherbina et al. (2017) utilized a 40-minute protocol transitioning between walking, running and cycling. Their results concluded a percent error of 27.4% for the Fitbit activity monitor and found significant underestimation of EE by the device compared to indirect calorimetry. The current study may have resulted in significant correlations due to the exercise bout being the same intensity throughout. Starting and stopping the devices during transitions between multiple forms (i.e., sedentary, running, cycling) of activity of differing intensities (i.e., light, moderate, vigorous) allows for more room for error in recording energy expenditure estimates.

Run Durations Effect on Energy Expenditure

Independently, the main effects of measurement mode and run duration on energy expenditure were significant ($p = 0.018$, $p < 0.001$). We expect run duration to have a significant impact on energy expenditure because as time progresses, energy expenditure continues to increase. The interactive effect of measurement method and run duration, however, was not significant ($p = 0.385$). This means that the error in energy expenditure was not different depending on run duration. To our knowledge, this is the first study to examine the differences in prediction accuracy between a short- and long-duration protocol. Seeing no difference in error between the two durations allows us to conduct shorter activity protocols and generalize it to a longer duration of activity. Being able to do this would allow more participants to be tested, resulting in larger studies that can ensure accurate results.

Fitness Level's Effect on Energy Expenditure

Independently, the main effects of measurement mode and fitness level on energy expenditure were significant ($p < 0.05$). The interactive effect of fitness level and measurement mode on energy expenditure was also found to be significant ($p = 0.001$). This means that the differences in energy expenditure error were dependent on fitness level. Figures 4 and 5 display theses errors. The high-fitness group saw much larger errors in energy expenditure between the Fitbit and the metabolic cart with the Fitbit significantly underestimating energy expenditure during both the short- and long-duration exercise for these individuals. These results are similar to a previous study by Kendall et al. (2019) that reported a much stronger correlation between a Fitbit activity monitor and indirect calorimetry in low fit individuals ($R = 0.934$) than high fit individuals ($R = 0.791$). Kendall et al. (2019) attributed this difference to activity monitors being unable to account for EE adjustments due to incline treadmill running or running intensities. However, in our study, participants all ran at the same intensity for the duration of the protocol. Thus, it can be assumed that differences in the estimation error may be due to the activity monitor's algorithm for energy expenditure.

Factors Included in Fitbit's EE Algorithm

According to the help section found on Fitbit.com, the energy expenditure algorithm combines factors such as BMR (calculated using height, weight, age and sex), activity data (step counts and distance), and HR to calculate calories burned daily and during physical activity. To improve accuracy, Fitbit™ suggests placing the activity monitor into the appropriate exercise mode, signifying the use of exercise mode in the algorithm as well. Table 4. displays the differences in descriptive characteristics between the high- and moderately- fit groups. When comparing these characteristics three factors were significantly different (weight, $VO_{2\ max}$, and speed of runs). We can assume that the differences in accuracy of energy expenditure estimation between the two groups could be due, in part, to these differences.

Basal Metabolic Rate. Most commonly BMR equations use variables such as height, weight, age, and sex to predict resting energy expenditure. In our sample, the high-fit group had a significantly lower weight than the moderately fit group. However, the height and age were not significantly different. Typically, trained individuals tend to have a higher percentage of fat-free mass (i.e., muscle and bone tissue and water). One major factor determining BMR has been shown to be fat-free mass (Sjodin et al., 1996; Haff and Weijs, 2014).

Some of the most popular prediction equations for BMR have been validated using healthy, adult individuals (Harris and Benedict, 1918; Mifflin et al., 1990; Owen et al., 1987; Owen et al., 1986; Food and Agricultural Organization, World Health Organization, United

Nations University, 1985). When using these equations to predict BMR in different population, they have been shown to be inaccurate. A study by Sjodin et al. (1996) researched differences in BMR between athletes and non-athletes and determined that athletes had a significantly higher BMR than estimated calculations. More recently, Haaf and Weijs (2014) compared the BMR of recreational athletes determined by indirect calorimetry to 12 equations for predicting energy expenditure. They determined that some of the most widely used equations for estimating EE showed less than 50% accuracy, but their developed equations based on FFM were much more accurate.

In this study, the high-fit group not only had a significantly lower weight than the moderately fit group but had a significantly higher VO2max (54.8 mL/kg/min). This value classified them as more fit than the average population (ACSM, 2018d). Knowing this, it can be assumed that the general prediction equations used in the general public may not be as accurate for their fitness levels and weight status. This would cause the high-fit group to have a lower estimated BMR to include in the calorie expenditure estimation from the Fitbit™ activity monitor, causing an underestimation of physical activity caloric expenditure.

Physical Activity Data. The Fitbit™ Inspire HR activity monitor collects physical activity data such as distance travelled using step count and stride length. According to Fitbit (2019) stride length is predetermined by the device using an individual's height and sex. However, if a user does multiple outdoor runs using a GPS to track distance, a stride length is calculated using those distances and step counts. Although step counts and distance travelled were not factors examined in our work, previous studies have been done to validate these variables.

A study by Wahl, Duking, Droszez, Wahl, and Mester (2017) aimed to validate eleven wearable devices for both step counts and distance travelled during treadmill exercise at varying intensities. They found that compared to direct observation, step counts by the devices were valid and displayed a MAPE of < 2%. However, the estimated distance travelled resulted in a MAPE of 1.3-29.9% and a significant underestimation of distance travelled at higher velocities (MAPE = -18.1-58.3%). Similarly, Haung, Xu, Yu, and Shull (2016) validated a number of devices during treadmill walking of increasing velocities. The reported that the majority of devices were valid in reporting step counts across all speeds. When reporting on distance travelled, they found a significant underestimation of distance travelled at faster walking speeds for most devices (including three Fitbit monitors).

It has been suggested that this inaccuracy in distance estimation could be due to the device using inaccurate stride lengths (Takacs et al., 2014). In our current work, significantly faster speeds were utilized by the high-fit group (Table 3). If an underestimation of distance travelled at faster speeds is utilized in the algorithm, this could result in an underestimation of caloric expenditure predicted by the device.

Heart Rate. The third variable used in the Fitbit™ algorithm for caloric expenditure is heart rate data. The Fitbit™ devices uses PurePulse™ technology to detect changes in blood volume and determine heart rate (bpm; Fitbit, 2020). Like step count and distance, heart rate data was not validated in this work, however previous research has found strong correlations between the HR from wrist worn devices and criterion measures such as Polar™ HR straps.

Bai et al. (2018) compared the HR from the Fitbit™ Charge HR to the Polar™ HR strap during sedentary, aerobic, and stimulated free-living activities. They reported a strong correlation between HR from the Fitbit™ and Polar™ monitor with an error range from -0.2 - 2.3% across the three activity types. Similarly, Shcherbina et al. (2018) reported strong correlation between seven wrist worn devices and a 12-lead electrocardiogram (ECG) during walking, running, and cycling. Across all devices and activities, error ranged from 0.9 - 9.0%.

Assuming our data was in line with the previous literature, HR would have been accurate for use in the algorithm. It is known that trained individuals have significantly lower heart rates at rest (Achten and Jeukendrup, 2003) and during submaximal exercise (Achten and Jeukendrup, 2003). When working at the same intensity, high-fit individuals will have a lower heart rate than those of a lower fitness level. In our study, all subjects completed the short- and long-duration running protocols at 70 - 75% of their determined VO_{2max}. Subjects in the high-fit group would not have had to work as hard as those in the moderately fit group to continue at this intensity. If the device recorded a lower heart rate value for the high-fit subjects, this would result in a lower energy expenditure estimation. However, this factor may not have contributed to the underestimation of EE by the Fitbit™ due to the linear relationship between HR and EE resulting in the assumption that a lower HR is associated with a lower caloric output (Keytel et al., 2005).

Strengths and Limitations

Strengths of this study included the use of continuous long-duration protocol, new technology that has not yet been reported on, and using a relative intensity of exercise (70 - 75% VO2max) for all participants. To our knowledge, in the literature, there are no studies validating the energy expenditure predicted from wearable activity monitors during an exercise bout of 30 minutes of exercise. Previous validations studies look at exercise ranging in duration from 3 to 20 minutes. Additionally, no other studies compare measurement error of a short duration protocol to a long-duration protocol. A wearable device that was released to the public in March 2019 was utilized in this study. Validation of new technology allow for us to determine improvement in factors estimated by wearable devices. Lastly, all subjects completed both protocols at the same intensity. Rather than choosing a set speed or allowing subjects to self-select a pace, each participant ran at 70-75% of their VO2max. This lets us generalize our findings to moderate-to-vigorous exercise as determined by ACSM (2018d). Additionally, in order to determine this relative intensity for each participant, we were able to determine VO2max. Finding VO2max allowed us to compare the measurement error between individuals of differing fitness levels. This allowed us to identify discrepancies in the measurement accuracy between moderate and high fit subjects.

Limitations of this study include a small sample size of relatively similar characteristics, the use of a single activity monitor, and the lack of control prior to laboratory visits. Subjects were recruited as a convenience sample from Eastern Michigan University's campus. We only accepted subjects between the ages of 18 - 35, which limited our sample. The fitness level of our subjects were similar, with only four participant's VO2max values falling below the 50^{th} percentile (fair) and only seven falling below the 60^{th} percentile (good) according to ACSM's guidelines. Only one consumer grade device was used in this study. Larger studies are able to validate multiple devices against a criterion method and make suggestions on which devices may be more accurate than others. Lastly, our participants were uncontrolled before visits. It is known that factors such as food intake, caffeine, and exercise can influence exercise performance and VO2 (Chowdury et al., 2017). Not asking participants to abstain from these factors could have an influence on performance measurements.

It is also important to note the complexity of energy expenditure. Energy expenditure varies greatly depending on the individual. Physiological factors such as fitness level, body composition, gender, economy, etc., along with environmental factors such as temperature, food intake, hydration, etc., all play a role in caloric output at rest and during physical activity. Not being able to control for all factors only allows us a glimpse into the validity of energy expenditure estimation by these wearable devices.

Future research

Future research should focus on increasing the number of devices validated to include other wearable consumer grade devices such as models manufactured by Apple™ and Garmin™. Given how rapidly technology is developed, newer Fitbit™ models should be included in these studies. Comparing multiple devices to the criterion measurement mode would produce results of energy expenditure estimation during long-duration exercise across devices. Doing this gives us a better understanding of which devices underestimate or overestimate energy expenditure and why that may be. Comparing estimations across devices from the same manufacturer would determine if energy expenditure algorithms are improving as technology develops. Additionally, future research should incorporate larger sample sizes with a larger range of fitness levels. Although, we were able to split our sample into a moderate and low fitness level, with only four subjects under the 50^{th} percentile, we were unable to fully determine measurement differences between a low-, moderate-, and high-fitness group. Having larger groups can give us more confidence that group differences exist. Lastly, other factors (i.e., heart rate, stride length, distance) should be recorded during the exercise protocols. This would allow us to determine which factors in the algorithms may be causing inaccurate estimates to be produced. Recording additional factors would also let us determine differences in fitness groups. By determining significant differences in these factors, it may allow for assumptions to be made about why these algorithms are more accurate for certain populations.

Conclusion

In conclusion, consumers should be cautious when using the Fitbit Inspire HR to determine caloric output. Significant differences were reported between measurement modes (metabolic cart v. Fitbit™) in all cases, except when looking at the full sample during the long-duration run. Those of higher fitness levels should take these estimations with a higher degree of caution due to the greater error seen when comparing these groups to those of a more moderate fitness level. Future studies should aim to validate additional devices in high fit individuals to determine which brands of activity monitors are more applicable to which subgroup of the population. The usage of wearable monitors should not be discredited, however, due to their ability to increase adherence to physical activity regimens (Cadmus-Bertram et al., 2015).

References

Achten, J., and Jeukendrup, A. E. (2003). Heart Rate Monitoring. *Sports Medicine*, *33*(7), 517–538.

Ainslie, P., Reilly, T., and Westerterp, K. (2003). Estimating human energy expenditure. *Journal of Sports Medicine, 33*(9), 683-698.

American College of Sports Medicine. (2018a). Overweight and obesity. In *ASCM's guidelines for exercise testing and prescription* (10th Edition ed., p. 287). Philadelphia: Wolters Kluwer.

American College of Sports Medicine. (2018b). Benefits and risks associated with physical activity. In *ACSM's guidelines for exercise testing and prescription* (10th Edition ed., p. 1-5). Philadelphia: Wolters Kluwer.

American College of Sports Medicine. (2018c). Sedentary behavior and health. In *ACSM's guidelines for exercise testing and prescription* (10th Edition ed., pp. 6-7). Philadelphia: Wolters Kluwer.

American College of Sports Medicine. (2018d). Interpretation of Results. In *ACSM's guidelines for exercise testing and prescription* (10th Edition ed., p. 92-94). Philadelphia: Wolters Kluwer.

Andre, D., and Wolf, D. (2007). Recent advances in free-living physical activity monitoring: A review. *Journal of Diabetes, Science and Technology, 1*(5), 760-767.

Ashe, M., Winters, M., Hoppmann, C., Dawes, M., Gardiner, P., Giangregorio, L., . . . McKay, H. (2015). "Not just another walking program": Everyday Activity Supports You (EASY) model-a randomized pilot study for a parallel randomized controlled trial. *Pilot and Feasibility Studies, 1*(4), 1-12.

Assah, F., Ekelund, U., Brage, S., Corder, K., Wright, A., Mbanya, J. C., and Wareham, N. (2009). Predicting physical activity energy expenditure using accelerometry in adults from sub-Sahara Africa. *Obesity (Silver Spring), 17*(8), 1588-1595.

Bai, Y., Hibbing, P., Mantis, C., and Welk, G. J. (2018, August). Comparative evaluation of heart rate-based monitors: Apple watch v. Fitbit Charge HR. *Journal of Sports Science, 36*(15), 1734-1741.

Bailey, R., Olson, J., Pepper, S., Porszasz, J., Barstow, T., and Cooper, D. (1995). The level of tempo of children's phsyical activities: an observational study. *Medicine and Science in Sports and Exercise, 27*(7), 1033-1041.

Bassett, D., Ainsworth, B., Leggett, S., Mathien, C., Main, J., Hunter, D., and Duncan, G. (1996). Accuracy of five electric pedometers for measuring distance walked. *Medical Science and Sports Exercise, 28*(8), 1071-1077.

Bassett, D., Ainsworth, B., Swartz, A., Strath, S., O'Brien, W., and King, G. (2000). Validity of four motion sensors in measuring moderate intensity physical activity. *Medicine and Science in Sports and Exercise, 32*(9), S471-S480.

Benedict, F. G. (1919). Energy requirements of children from birth to puberty. *The Boston Medical and Surgical Journal*, *181*(5), 107–139.

Benedict, F. G., and Roth, P. (1915). The basal caloric output of vegetarians as compared with that of non-vegetarians of like weight and height. *Proceedings of the National Academy of Sciences, 1*(2), 100–101.

Benedict F. G., Smith H. M. (1915) The metabolism of athletes as compared with normal individuals of similar height and weight. *Journal of Biological Chemistry, 20*(3), 243–252.

Benedict, F. G., and Talbot, F. B. (1915). The physiology of the new-born infant. *Proceedings of the National Academy of Sciences*, *1*(12), 600–602.

Benedict, F. G., and Roth, P. (1918). Effects of a prolonged reduction in diet on 25 Men: I. influence on basal metabolism and nitrogen excretion. *Proceedings of the National Academy of Sciences*, *4*(6), 149–152.

Bonnefoy, M., Normand, S., Pachiaudi, C., Lacour, J. R., Laville, M., and Kostka, T. (2001). Simultaneous validation of ten physical activity questionnaires in older men: A doubly labeled water study. *Journal of the American Geriatrics Society, 49*(1), 28–35.

Boyle, R. (1660). *New experiments physico-mechanical, touching the spring of air, and its effects, (made for the most part, in a new pneumatical engine) written by way of lettr to the right honorable Charles Lord Vicount of Dunfarvan, eldest son to the Earl of Corke.* Oxford: H. Hall.

Brage, S., Westgate, K., Franks, P. W., Stegle, O., Wright, A., Ekelund, U., and Wareham, N. (2015). Estimation of free-living energy expenditure by heart rate and movement sensing: A doubly-labelled water study. *PLoS ONE, 10*(9), 517-524.

Bratteby, L.-E., Sandhagen, B., Fan, H., and Samuelson, G. (1997). A 7-day activity diary for assessment of daily energy expenditure validated by the doubly labelled water method in adolescents. *European Journal of Clinical Nutrition, 51*(9), 585–591.

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

Carter, J. M., Jones, D. A., and Jeukendrup, A. E. (2002). Validity and reliability of three commercially available breath-by-breath respiratory systems. *Medicine and Science in Sports and Exercise, 34*(5), 435-441.

Chang, C.-H., Lin, K.-C., Ho, C.-S., and Huang, C.-C. (2010, May). Accuracy of the energy expenditure during uphill exercise measured by the waist-worn Actigraph. *Journal of Exercise Science and Fitness, 17*(2), 62-66.

Chowdhury, E. A., Western, M. J., Nightingale, T. E., Peacok, O. J., and Thompson, D. (2017, February). Assessment of laboratory and daily energy expenditure estimates from consumer multi-sensor physical activity monitors. *PLoS ONE, 12*(2), 1-15.

Daley, A. J., and Welch, A. (2004). The effects of 15 min and 30 min of exercise on affective responses both during and after exercise. *Journal of Sports Sciences, 22*(7), 621–628.

Davidson, L., McNeil, G., Haggarty, P., Smith, J., and Franklin, M. (1997). Free-living energy expenditure of adult men assessed by continous heart-rate monitoring and doubly-labelled water. *Britian Journal of Nutrition, 78*(5), 695-708.

Diaz, K., Krupka, D., Chang, M., Shaffer, J., Ma, Y., Goldsmith, J., . . . Davidson, K. (2016). Validation of the Fitbit One for physical activity measurement at an upper torso attachment site. *BMC Research Notes, 213*(9), 1-9.

Dondzilla, C., and Garner, D. (2016, June). Comparative accuracy of fitness tracking modalities in quantifying energy expenditure. *Journal of Medical Engineering and Technology, 40*(6), 325-329.

Dooley, E., Golaszewski, N., and Bartholomew, J. (2017). Estimating accuracy at execise intensities: A comparative study of self-monitoring heart and physical activity devices. *JMIR mHealth and uHealth, 5*(3), e34.

Dubois, E. F. (1916). The Archives of Internal Medication. In *Twelfth paper: The metabolism of boys 12 and 13 years old compared with the metabolism at other ages* (Vol. 17). California: American Medical Association.

DuBois, E. F., and Veeder, B. S. (1910). The total energy requirement in diabetes mellitus. *Archives of International Medicine, 1*, 37-46.

Edelman, B., and Smits, G. (1984). The pedometer: A reassessment of its usefulnes in the measurement of activity levels. *Perceptual and Motor Skills, 58*, 151-158.

Engstrom, E., Ottosson, E., Wohlfart, B., Drundstrom, N., and Wisen, A. (2012). Comparison of heart rate measured by Polar RS400 and ECG, validity and repeatability. *Advances in Physiotherapy, 14*(3), 115-122.

Eston, R. G., Rowlands, A. V., and Ingledew, D. K. (1998). Validity of heart rate, pedometry, and accelerometry for predicting the energy cost of children's activities. *Journal of Applied Physiology, 84*(1), 362–371.

Food and Agricultural Organization, World Health Organization, and United Nations University. (1985). Energy and protein requirements report of a joint expert consultation. *WHO Technical Report Series* (no. 724). Geneva: WHO.

Fitbit (2019). How does my Fitbit device calculate my daily activity? Retrieved from https://help.fitbit.com/articles/en_US/Help_article/1141/?q=calories+burnedandl=en_USandfs=Searchandpn=1#calories

Fitbit (2020). How do I track my heart rate with my Fitbit device? Retrieved from https://help.fitbit.com/articles/en_US/Help_article/1565/?q=heart+rateandl=en_USandfs=Searchandpn=1

Flegal, K. M., Graubard, B. I., and Williamson, D. F. (2005). Excess deaths associated wth underweight, overweight, and obesity. *JAMA, 293*(15), 1861-1867.

Fogelholm, M., Hiilloskorpi, H., Laukkanen, R., Oja, P., Van Marken, L., and Westerterp, K. (1998, August). Assessment of energy expenditure in overweight women. *Medicine and Science in Sports and Exercise, 30*(8), 1191-1197.

Freedson, P. S., and Miller, K. (2000). Objective monitoring of physical activity using motion sensors and heart rate. *Research Quarterly for Exercise and Sport, 71*(sup2), 21–29.

Goran, M. I., Poehlman, E. T., and Johnson, R. K. (1995). Energy requirements across the life span: New findings based on measurement of total energy expenditure with doubly labeled water. *Nutrition Research, 15*(1), 115-150.

Haaf, T. T., and Weijs, P. J. M. (2014). Resting energy expenditure prediction in recreational athletes of 18–35 years: Confirmation of Cunningham equation and an improved weight-based alternative. *PLoS ONE*, *9*(10), 1-8.

Hale, T. (2008). History of developmental in sport and exercise physiology: A. V. Hill, maximal oxygen uptake, and oxygen debt. *Journal of Sports Science, 26*(4), 365-400.

Hales, C., Carrol, M. D., Fryar, C. D., and Ogden, C. L. (2017). *Prevalence of obesity among adults and youth: United States, 2015-2016*. Hyattsville, MD: Centers for Disease Control and Prevention.

Harris, J. A., and Benedict, F. G. (1918). A biometric study of human basal metabolism. *Proceedings of the National Academy of Sciences*, *4*(12), 370–373.

Huang, Y., Xu, J., Yu, B., and 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.

Haugen, H., Chan, L.-N., and Li, F. (2007). Indirect calorimetry: A practical guide for clinicians. *Nutrition in Clinical Practice, 22*(4), 377-385.

Heinonen, I., Helajarvi, H., Pahkala, K., Heinonen, O. J., Hirvensalo, M., Palve, K., . . . Raitakari, O. T. (2013). Sedentary behaviours and obesity in adults: The cardiovascular risk in young finns study. *BMJ Open*, 1-11.

Hettiarachchi, I., Hanoun, S., Nahavandi, D., and Nahavandi, S. (2019). Validation of the Polar OH1 optical heart rate sensor for moderate and high intensity physical activities. *PLoS ONE, 14*(5), 1-13.

Hills, A. P., Mokhtar, N., and Byrne, N. M. (2014). Assessment of physical activity and energy expenditure: An overview of objective measures. *Frontiers in Nutrition, 1*, 1-16.

Hoodless, D., Stainer, K., Savic, N., Batin, P., Hawkins, M., and Cowley, A. (1994). Reduced customary activity in chronic heart failure: Assessment with a new shoe-mounted pedometer. *International Journal of Cardiology*, *43*(1), 39–42.

Irimagawa, S., and Imamiya, S. (1993, April 1). Industrial hygenic study on nursing activities comparison of energy expenditure between pedometer and Holter electrocardiograph. *Kitasato Archives of Experimental Medicine, 65*, 99-105.

Johansson, P., Rossander-Hulthen, L., Slinde, F., and Ekblom, B. (2006). Accelerometry combined with heart telemetry in the assessment of the energy expenditure. *British Journal of Nutrition, 95*, 631-639.

Jones, P., Winthrop, A., Schoeller, D., Swyer, P., Smith, J., Filler, R., and Heim, T. (1987). Validation of doubly labeled water for assessing energy expenditure in infants. *Journal of Pediatric Research, 21*(3), 242-246.

Karamanou, M., Tsoucalas, G., and Androutsos, G. (2013). Hallmarks in the study of respiratory physiology and the crucial role of Antoine-Laurent de Lavoisier (1743-1794). *American Journal of Physiology Lung Cellular and Molecular Physiology, 305*(9), L591-L594.

Kashiwazaki, H., Inaoka, T., Suzuki, T., and Kondo, Y. (1986, February). Correlations of pedometer readings with energy expenditure in workers during free-living daily activities. *European Journal of Applied Physiology and Occupational Physiology, 54*(6), 585-590.

Kendall, B., Bellovary, B., and Gothe, N. P. (2018). Validity of wearable activity monitors for tracking steps and estimating energy expenditure during a graded maximal treadmill test. *Journal of Sports Sciences, 37*(1), 42-49.

Kenny, G., Notley, S., and Gagnon, D. (2017). Direct calorimetry: A brief historical review of its use in the study of human metabolsm and thermoregulation. *European Journal of Applied Physiology*, 1765-1785.

Keytel, L., Goedecke, J., Noakes, T., Hiiloskorpi, H., Laukkanen, R., van der Merwe, L., Lambert, E. (2005). Prediction of energy expenditure from heart rate monitoring during submaximal exercise. *Journal of Sports Sciences, 23*(3), 289-297.

Kilanowski, C., Consalvi, A., and Epstein, L. (1999). Validation of an electric pedometer for measurment of physical actiivty in children. *Pediatric Exercise Science, 11*(1), 63-68.

Kline, G.M., Porcari, J.P., Hintermesiter, R., Freedson, P.S., Ward, A., McCarron, R.F., . . . Rippe, J.M. (1987). Estimation of VO2max from a one-mile track walk, gender, age, and body weight. Medicine and Science in Sports and Exercise, 19(3), 253-259.

Koebnick, C., Wagner, K., Thielecke, F., Dieter, G., Hohne, A., Franke, A., . . . Zunft, H. (2005). An easy-to-use semiquantitative food record validated for energy intake by using doubly labelled water technique. *European Journal of Clinical Nutrition, 59*, 989-995.

Kopelman, P. G. (2000). Obesity as a medical problem. *Nature, 404*, 636-643.

Kumahara, H., Schutz, Y., Ayabe, M., Yoshioka, M., Yoshitake, Y., Shindo, M., … Tanaka, H. (2004). The use of uniaxial accelerometry for the assessment of physical-activity-related energy expenditure: a validation study against whole-body indirect calorimetry. *British Journal of Nutrition, 91*(2), 235–243.

Lavoisier, A.-L., and Laplace, P. S. (1780). *Memoire sur la chaleur.*

Lavoiser, A.-L., and Seguin, A. (1793). *First memoir on the respiration of animals* .

Leenders, N., Sherman, M., Nagaraja, H., and Lawrence, K. (2001, July). Evaluations of methods to assess physical activity in free-living conditions. *Medicine and Science in Sports and Exercise, 33*(7), 1233-1240.

Liu, S. (2019a). *Fitness and activity tracker: Statistics and facts*. Retrieved September 2019, from Statista: https://www.statista.com/topics/4393/fitness-and-activity-tracker/?

Liu, S. (2019b). *Fitness tracker device revenue worldwide 2016-2022*. Retrieved September 2019, from Statista: https://www.statista.com/statistics/610433/wearable-healthcare-device-revenue-worldwide/

Lof, M., Hannestad, U., and Forsum, E. (2003). Comparison of commonly used procedures, including the doubly-labelled water technique, in the estimation of total energy expenditure of women with special reference to the significance of body fatness. *Britain Journal of Nutrition, 90*(5), 961-968.

Lower, R. (1669). *Tractatus de corde item de motu and colore sanguinis et chyli in eum transitu.* London: SCORE.

Maddison, R., Ni Mhurchu, C., Jiang, Y., Vander. Hoom, S., Rodgers, A., Lawes, C., and Rush, E. (2007). International Physical Activity Questionnarie (IPAQ) and New Zealand Physical Activity Questionnaire (NZPAQ): A doubly labelled water validation. *International Journal of Behavioral Nutrition and Physical Activity, 62*(4), 1-9.

Maslakovic, M., Johnson, D., and Jovin, I. (2017, February 13). Number of Americans using fitness trackers to double by 2021. Retrieved from https://gadgetsandwearables.com/2017/02/06/americans-fitness-trackers/.

Masters, R. K., Reither, E. N., Powers, D. A., Claire, Y. Y., Burger, A. E., and Link, B. G. (2013). The impact of obesity on US mortality levels: The importance of the age and cohort factors in population estimates. *American Journal of Public Health, 103*(10), 1895-1901.

Mifflin, M. D., Jeor, S. T. S., Hill, L. A., Scott, B. J., Daugherty, S. A., and Koh, Y. O. (1990).

A new predictive equation for resting energy expenditure in healthy individuals. *The American Journal of Clinical Nutrition*, *51*(2), 241–247.

Mokdad, A. H. (1999). The spread of the obesity epidemic in the United States, 1991-1998. *Journal of American Medical Association, 282*(16), 1519.

Montesi, L., El Ghoch, M., Brodosi, L., Calugi, S., Marchesini, G., and Dalle Grave, R. (2016). Long-term weight loss maintenance for obesity: A mulidisciplinary approach. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy, 9*, 37-44.

Montoye, A. H. K., Mitrzyk, J. R., and Molesky, M. J. (2017). Comparative accuracy of a wrist-worn activity tracker and a smart shirt for physical activity assessment. *Measurement in Physical Education and Exercise Science, 21*(4), 201–211.

Morio, B., Ritz, P., Verdier, E., and Montaurier, C. (1997). Critical evaluation of the factorial and heart-rate recording methods for the determination of energy expenditure of free-living elderly people. *Cambridge University Press, 78*(5), 709-722.

Mtaweh, H., Tuira, L., Floh, A. A., and Parshuram, C. S. (2018). Indirect calorimetry: History, technology, and application. *Frontiers in pediatrics*, 1-8.

Ndahimana, D., and Kim, E.-K. (2017). Measurement methods for physical activity and energy expenditure: A review. *Clinical Nutrition Research*, 68-75.

Nishikido, N., Kashiwazaki, H., and Suzuki, T. (1982). Preschool children's daily activities: Direct observation, pedometry or questionnaire. *Journal of Human Ergology, 11*(2), 214-218.

Owen, O. E., Kavle, E., Owen, R. S., Polansky, M., Caprio, S., Mozzoli, M. A., … Boden, G. (1986). A reappraisal of caloric requirements in healthy women. *The American Journal of Clinical Nutrition*, *44*(1), 1–19.

Owen, O. E., Holup, J. L., D'Alessio, D. A., Craig, E. S., Polansky, M., Smalley, K. J., … Mozzoli, M. A. (1987). A reappraisal of the caloric requirements of men. *The American Journal of Clinical Nutrition*, *46*(6), 875–885.

Peabody, F. W., Meyer, A. L., and Dubois, E. F. (1916). Sixteenth paper: The basal metabolism of patients with cardiac and renal disease. In *The Archives of Internal Medicine* (Vol. 17, pp. 980-982). California: American Medical Association.

Poncet, S., and Dahlberg, L. (2011). The legacy of Henri Victor Regnault in the arts and sciences. *International Journal of Arts and Sciences, 4*(13), 377-400.

Puhl, J., Greaves, K., Hoyt, M., and Baranowski, T. (1990). Children's Activity Rating Scale (CARS): Description and calibration. *Research Quarterly for Exercise and Sport, 61*(1), 26-36.

Racette, S., Schoeller, D., and Kushner, R. (1995). Comparison of heart rate and physical activity recall with doubly labeled water in obese women. *Medicine Science and Sports Exercise, 27*(1), 126-133.

Rafamantanantsoa, H., Ebine, N., Yoshioka, M., Higuchi, H., Yoshitake, Y., Tanaka, H., . . . Jones, P. (2002). Validation of the three alternative methods to measure total energy expenditure against doubly labeled water method for older japanese men. *Journal of Nutritional Science (Tokyo), 48*(6), 517-523.

Ravussin, E., Harper, I., Rising, R., and Bogardus, C. (1991). Energy expenditure by doubly labeled water: Validation in lean and obese subjects. *American Journal of Physiology, 261*(3), E402-E409.

Reddy, R. K., Pooni, R., Zaharieva, D., Senf, B., Youssef, J., Dassau, E., . . . Jacobs, P. G. (2018, December). Accuracy of wrist-worn activity monitors during common daily physical activities and types of structured exercise: Evaluation study. *JMIR mHealth and uHealth, 6*(12), e10338.

Reilly, J., Penpraze, V., Hislop, J., Grant, S., and Paton, J. (2008). Objective measurements of physical activity and sedentary behaviour: Review with new data. *Archives of Disease in Childhood, 93*(7), 614.

Roberts, S. B., Coward, W. A., Ewing, G., Savage, J., Colte, T. J., and Lucas, A. (1988). Efffects of weaning on accuracy of doubly labeled water method in infants. *Journal of Applied Physiology, 254*(4), R622-R627.

Roberts, S. B., Coward, W. A., Schlingenseipen, K.-H., Nohria, V., and Lucas, A. (1986). Comparison of the doubly labelled water (2H2180) method with indirect calorimetry and a nutrient-balance study for simultaneous determination of energy expenditure, water intake, and metabolize energy intake in preterm infants. *Journal of Clinical Nutrition, 44*, 315-322.

Rothenberg, E., Bosaeus, I., Lernfelt, B., Landahl, S., and Steen, B. (1998). Energy intake and expenditure: Validation of a diet history by heart rate monitoring, activity diary and doubly labeled water. *European Journal of Clinical Nutrition, 52*(11), 832-838.

Sallis, J. F., and Saelens, B. E. (2000). Assessment of Physical Activity by self-report: Status, limitations and future directions. *Research Quarterly for Exercise and Sport, 71*(2), 1-14.

Schoeller, D. A., and van Santen, E. (1982). Measurement of energy expenditure in humans by doubly labelled water method. *Journal of Applied Physiology, 53*(4), 955-959.

Schoeller, D. A., Ravussin, E., Schutz, Y., Acheson, K. J., Baertschi, P., and Jequier, E. (1986). Energy expenditure by doubly labelled water: validation in humans and proposed calculation. *American Journal of Physiology, 250*(5), 823-830.

Schmidt, W. D., Biwer, C. J., and Kalscheuer, L. K. (2001). Effects of long versus short bout exercise on fitness and weight loss in overweight females. *Journal of the American College of Nutrition*, *20*(5), 494–501.

Schulz, S., Westerterp, K., and Bruck, K. (1989). Comparison of energy expenditure by the doubly labeled water technique with energy intake, heart rate, and activity recording in man. *American Journal of Clinical Nutrition, 49*(6), 1146-1154.

Seale, J., Rumpler, W., Conway, J., and Miles, C. (1990). Comparison of doubly labeled water, intake-balance, and direct- and indirect-calorimetry methods for measuring energy expenditure in adult men. *American Journal of Clinical Nutrition, 52*(1), 66-71.

Selin, K., Winkel, J., and group, S.-M. I. (1994). Evaluation of two instruments for recornding sitting and standing postures and sumber of foot steps. *Applied Ergonomics, 25*(1), 41-46.

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*(2).

Shields, M., and Tremblay, M. S. (2008, June). Sedentary behavior and obesity. *Health Reoports, 19*(2), 19-30.

Silfee, V. J., Haughton, C. F., Jake-Schoffman, D. E., Lopez-Cepero, A., May, C., Sreedhara, M., . . . Lemon, S. C. (2018). Objective measurement of physical activity outcomes in lifestyle interventions among adults: A systematic review. *Preventive Medicine Reports, 11*, 74-80.

Singh, J., Prentice, A. M., Diaz, E., Coward, W. A., Ashford, J., Sawyer, M., and Whitehead, R. G. (1989). Energy expenditure of Gambian women during peak agricultural activity measured by the doubly-labelled water method. *Britian Journal of Nutrition, 62*(2), 315-329.

Sjodin, A. M., Forslund, A. H., Westerterp, K. R., Andersson, A. B., Forslund, J. M., and Hambraeus, L. M. (1996). The influence of physical activity on BMR. *Medicine and Science in Sports and Exercise*, *28*(1), 85–91.

Sljivic, S., and Gusenoff, J. A. (2019, January). The obesity epidemic and bariatic trends. *Clinics in Plastic Surgery, 46*(1), 1-7.

Spierer, D., Rosen, Z., Litman, L., and 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., Gillette, C., and Foster, C. (2015, January). Are activity trackers accurate? *ACE Prosource*, 1-3.

Sylvia, L., Bernstein, E., Hubbard, J., and Keating, L. A. (2014). A practical guide to measuring physical activity. *Academy of Nutrition and Dietetics, 114*(2), 199-208.

Takacs, J., Pollock, C. L., Guenther, J. R., Bahar, M., Napier, C., and 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.

Temko, A. (2017). Accurate heart rate monitoring during physical exercises using PPG. *IEEE Transactions on Biomedical Engineering, 64*(9), 2016-2024.

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

Thompson, D., Edelsberg, J., Colditz, G., Bird, A., and Oster, G. (1999). Lifetime health and economic consequences of obesity. *Arch Intern Med, 159*(18), 2177-2183.

Troiana, R., Berrigan, D., Dodd, K., Masse, L., Tilert, T., and McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise, 40*(1), 181-188.

Tudor-Locke, C., Williams, J. E., Reis, J. P., and Pluto, D. (2002). Utility of pedometers for assessing phsyical activity. *Journal of Sports Medicine, 32*(12), 795-808.

U.S. Department of Health and Human Services. (2018). *Physical activity guidelines for Americans* (2nd ed.). Washington D.C., U.S.

Underwood, E. A. (1943). Lavoisier and the history of respiration. In W. Langdon-Brown, *Section of the history of medicine* (Vol. 37, pp. 247-262).

Vanhees, L., Lefevre, J., Philippaerts, R., Martens, M., Huygens, W., Troosters, T., and Beunen, G. (2005). How to assess physical activity? How to assess physical fitness? *European Journal of Cardiovascular Prevention on Rehabilitation, 12*(1), 102-114.

Vanhelst, J., Hurdiel, R., Mikulovic, J., Bui-Xuan, G., Fardy, P., Theunynck, and Beghin, L. (2012, August). Validation of the Vivago wrist-worn accelerometer in the assessment of physical activity. *BMC Public Health, 12*(690), 1-7.

Wahl, Y., Düking, P., Droszez, A., Wahl, P., and Mester, J. (2017). Criterion-validity of commercially available physical activity tracker to estimate step count, covered distance and energy expenditure during sports conditions. *Frontiers in Physiology*, *8*, 725.

Weinswig, D. (2017, July 17). Deep dive: Wellness as a luxury: Part 1- living well. *FUNG Global Retail and Technology*. 1-17.

Westerterp, K. R. (2017). Doubly labelled water assessment of energy expenditure: Principle, practice, and promise. *European Journal of Applied Physiology, 117*, 1277-1285.

Westerterp, K. R., Saris, W. H., van Es, M., and ten Hoor, F. (1986). Use of the doubly labeled water technique in humans during sustained exercise. *Journal of Applied Physiology, 61*(6), 2162-2167.

Westerterp, K., Brouns, F., and Saris, W. H. (1988). Comparison of doubly labeled water with respirometry at low- and high-activity levels. *Journal of Applied Physiology, 65*(1), 53-56.

World Health Organization. (2019). *Physical inactivity: A global public health problem*. Retrieved from https://www.who.int/ncds/prevention/physical-activity/inactivity-global-health-problem/en/

Wright, S. M., and Aronne, L. J. (2012, October). Causes of obesity. *Abdominal Imaging, 37*(5), 730-732.

Zahiri, C., Schmalzried, T., Szuszczewicz, E., and Amstutz, H. (1998). Assessing activity in joint replacement patients. *Journal of Arthoplasty, 13,* 890-895