Wearable Technology and the Measurement of Physical Activity Intensity and Volume

by

Kyle Stephen Weber

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Science
in
Kinesiology

Waterloo, Ontario, Canada, 2020

Author's declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Physical activity has been linked to numerous health outcomes including a decreased risk of chronic disease and an increased quality of life; it is therefore an important component of rehabilitative and preventative programs and for tracking disease progression. Activity is often described using the Frequency-Intensity-Type-Time-Volume principle. Intensity and volume are of particular importance as they are used to relate activity to health outcomes. However, measuring intensity and volume outside laboratories poses many challenges. A new category of technology called wearables has improved the ability to objectively and continuously measure intensity and volume in free-living using accelerometers and portable electrocardiogram (ECG) sensors. Accelerometers can be worn on different body locations and have been used to estimate activity intensity and volume. However, several different analytical approaches, or models, have been used to date. Quantifying the differences in activity-related outcome measures from these different wearables models has important implications in guiding the clinical decision making involved in rehabilitation and tracking disease progression. This thesis aimed to describe the differences in total activity volume (time spent in sedentary, light, and moderate-to-vigorous activity) and moment-to-moment agreement in activity intensity measured by four wearables models, and to determine if model performance was consistent for those who were relatively active compared to inactive. These models included an existing wrist accelerometer model (Wrist), a novel ankle accelerometer model (Ankle) that used activity counts to predict gait speed, a heart rate model (HR), and a model that combined heart rate with the new ankle model (HRAcc). Data from the *ONDRI@Home* project's control cohort were used. Participants wore a chest-mounted ECG and accelerometers on the wrist and ankle for a period of 5-7 days. To develop the new Ankle model, a subset of participants also performed a treadmill protocol. Data were collapsed into 15-second epochs. Only epochs when participants were awake and all devices provided usable data were included in analyses. Participants that provided less than 30 hours of usable data were excluded from analyses. Due to the volume of lost data, a subset of analyses was conducted using data from epochs where pairs of models provided valid data. Activity volumes were reported as a percentage of usable data. Moment-to-moment agreement in activity intensity was assessed using Cohen's kappa.

Significant differences in activity volume between models were found at each activity intensity. Moment-to-moment agreement in intensity was in the fair-to-almost perfect range; agreement was highest between the HR and HRAcc models, and lowest between both the HR and Ankle and HR and Wrist models. Model performance was consistent across activity levels. However, model agreement was greater for those who were more active. The Ankle model demonstrated excellent performance; activity counts explained more than 98% of the variance in gait speed and prediction error was less than .04 m s^{-1}.

Given the clinically significant magnitude of differences in activity volume and the large range of moment-to-moment agreement in intensity, physical activity outcome measures from different models should not be considered equivalent. This thesis highlights the limitations of using wearables models related to different types of activities, how devices measure intensity, and physiological differences which may have affected model performance. Many of these limitations can be overcome by multi-device models that use individualized data and relative intensity measures. Multi-device models likely have the ability to better represent activity duration, timing, and intensity and should therefore be the focus of future research.

Acknowledgements

I would like to extend my thanks to my committee, Drs. Laura Middleton and Chris Vigna, and to my supervisor, Dr. Bill McIlroy. I've thoroughly enjoyed growing as a researcher during my graduate studies and your feedback and guidance have been invaluable.

Thank you to the Ontario Neurodegenerative Disease Research Initiative (ONDRI) for providing funding which allowed me to focus more directly on my research.

Thank you to all my colleagues, past and present, who have contributed to data collection and processing. Special thanks go out to Kit Beyer, David Ding, Michelle Lee, Adam Vert, Vanessa Thai, and Matthew Wong – all your work has saved me many headaches!

Lastly, thank you to my wife, Stephanie, for always pretending to understand what I'm talking about with a smile on your face, for being my proofreader-in-chief, and for being someone I can always bounce ideas off of. Thank you for all your support over the last couple years.

Table of contents

List of figures

List of tables

Chapter 1: Background and Literature Review

1.1 Physical Activity

1.1.1 The Importance of Activity and Exercise

There is an abundance of literature describing the relationship between physical activity and positive health outcomes. These benefits include reduced risk of chronic disease such as cardiovascular disease, osteoporosis, type 2 diabetes, cancers, and psychiatric conditions (World Health Organization [WHO], n.d.; Canadian Society for Exercise Physiology [CSEP], 2011). Additionally, increased activity has been associated with improved quality of life (Bize, Johnson, & Plotnikoff, 2007; Berger & Tobar, 2011) and to improvements in both true and one's perceived physical function (Berger & Tobar, 2011). These improvements can be attributed to movement – both exercise and overall physical activity – so it is crucial to differentiate the two. Caspersen and colleagues (1985) define physical activity as "any bodily movement produced by skeletal muscles that results in energy expenditure" (p. 126). Exercise, a subset of physical activity, is defined as "planned, structured, repetitive, and purposive in the sense that the improvement or maintenance of one or more components of physical health is an objective" (Caspersen et al., 1985, p. 128).

More broadly, movements that are classified as non-exercise physical activity are often unstructured and are typically only described by when they occur during the day (i.e. during sleep, leisure time, or work) (Caspersen, Powell, & Christenson, 1985).

Exercise and physical activity can both be described using the frequency-intensity-type-time-volume (FITT-V) principle. In this principle, frequency refers to how often activity is performed, intensity to how physically or metabolically demanding the activity is, type to the mode of activity, time to the duration of activity, and volume to describe the total quantity of activity as the product of time and intensity (American College of Sports Medicine [ACSM], 2014). Guidelines have been developed that state the recommended volume of activity that is associated with "substantial health benefits" (ACSM, 2014, p. 8). Both the ACSM and CSEP recommend a minimum of 150 minutes of moderate-to-vigorous intensity exercise every week;

these guidelines are widely used. However, they describe the minimum recommended activity volume and a dose-response relationship has been established between activity and multiple aspects of health in that more activity leads to greater health improvements (ACSM, 2014; Janssen & LeBlanc, 2010). Therefore, describing one's activity with the FITT-V principle has important implications for both proactive and reactive health interventions.

1.1.2 Measuring Physical Activity

While frequency, time, and type of activity are fairly easy to quantify, activity intensity (and therefore volume) are not as easy to measure. Subjectively, intensity can be described using different rating scales such as Borg's Scale of Perceived Exertion (Borg, 1982). Objectively, physical activity intensity can be quantified using several measures. Energy expenditure can be used as a measure of intensity when expressed as a rate (e.g. kcal kg^{-1} min^{-1}) since the generation of movement requires more energy to sustain. This value is often measured as the increase in energy expenditure above resting levels. When energy demand increases, O_2 consumption (VO_2) increases to meet this need. Increased VO_2 demand is met by increasing cardiac output which is partially accomplished by increasing heart rate (Plowman & Smith, 2007). The relationship between heart rate and VO_2 is linear and quite strong during activity, especially when expressed in terms relative to the individual's activity capacity (Strath et al., 2000), so both heart rate and VO_2 can be used as measures of energy expenditure and therefore physical activity intensity. These three measures all quantify intensity on a continuous scale, but intensity can also be categorized as sedentary, light, moderate, or vigorous. This is commonly done using metabolic equivalent of task (METs) thresholds of 1.5, 3.0, and 6.0 METs, respectively (Powell, Carson, Dowd, & Donnelly, 2017; Assah et al., 2011). Reporting activity intensity as METs or categorically provides a measure of intensity that is easier to interpret than energy expenditure when providing feedback to patients, study participants, or healthcare practitioners. Since being able to provide these individuals with meaningful data is one of the major goals for the project within which this thesis falls (see section 2.1.0 for more details about the ONDRI@Home project), this thesis reports activity intensity as categories as opposed to energy expenditure measured on a continuous scale.

Ideally, to measure physical activity intensity, a direct measure of energy expenditure would be taken. This is possible in the laboratory using direct calorimetry. By measuring changes in temperature of an isolated chamber caused by the subject's thermogenesis, energy expenditure can be calculated (Hills et al., 2014). However, these chambers are small and do not allow the subject to behave as they would on a day-to-day-basis (free-living). Due to these restrictions, *indirect* calorimetry is more commonly used to assess physical activity intensity. Outside calorimeters, measuring VO_2 is considered the gold standard but it requires expensive equipment. This has led to the development of alternative methods that estimate VO_2. Due to the relationships that VO_2 has with energy expenditure, heart rate, and activity intensity, VO_2 can be estimated using activity protocols while measuring outcomes such as heart rate or cycling power (ACSM, 2014). Heart rate is commonly measured but variability caused by non-activity factors such as caffeine intake, stress, environmental temperature, and hydration (Freedson & Miller, 2000; Villars et al., 2012; Brage et al., 2004) reduces the accuracy of activity intensity estimates if heart rate is used on its own (Villars et al., 2012). These factors have a more noticeable effect during low intensity activity (Warren et al., 2010; Hills et al., 2014).

While these relationships hold true outside the laboratory as well, VO_2 is much easier to measure or estimate in the laboratory due to the nature of the activity being performed. For example, if VO_2 is being estimated with validated equations in the laboratory, the physical activity being assessed typically has well-defined start and end times, consists of a single type of activity, and performance measures (e.g. pedalling cadence, treadmill speed) can be easily recorded. Taken together, these characteristics make measuring activity over short time periods in the laboratory relative straightforward.

1.1.3 Measuring Physical Activity in Free-Living

Although exercise is very commonly associated with health benefits, unstructured physical activity from daily life can also be of sufficient intensity to induce these benefits (Strath, Bassett, Swartz, & Thompson, 2001; Brooks et al., 2004) so it is important to be able to measure physical activity in free-living in addition to structured, single-type activity in the laboratory. Indirect calorimetry could be considered the gold standard in free-living as well since portable gas exchange systems are available and used in research, but it is not practical to wear an airtight facemask that measures gas exchange for any considerable length of time. In free-living, the gold standard for measuring energy expenditure is doubly labelled water because this technique is non-invasive and allows the subject to behave as they normally would (Hills et al., 2014; Freedson & Miller, 2000). Performing doubly labelled water is very expensive (Hills et al., 2014; Skender et al., 2016) and only provides information about *total* energy expenditure; it does not provide any information about energy expenditure with temporal resolution greater than the length of the study period (Hills et al., 2014) which is often seven or more days. While this is useful for measuring energy expenditure over an extended time period, doubly labelled water is not useful for measuring moment-to-moment changes in energy expenditure.

As of 2009, the most common way to measure activity in free-living (Westerterp, 2009) and in large-scale studies (Shepard & Aoyagi, 2012) was the use of self-report methods such as questionnaires, activity logs, and interviews. Although these methods are widely used, their reliability and validity have been questioned and they are not able to detect differences in activity on a day-to-day timescale or in total activity volume over a given time period (Hills et al., 2014). Subjective measures are prone to bias and frequently underestimate sedentary time while overestimating activity (Skender et al., 2016; Ryan et al., 2018), but also underestimate the lower intensity activities performed in daily living compared to lower intensity exercise (Ainsworth, 2009).

Despite several techniques that can be used to measure physical activity, the limitations of self-report methods, doubly labelled water, and VO_2 measurement mean that there is no true and practical gold standard for continuous monitoring of physical activity intensity in free-living that provides good temporal resolution (Hills et al., 2014). Recent advances in wearable sensors (wearables), such as accelerometers and portable electrocardiogram (ECG), systems have changed how physical activity is measured in free-living and provide opportunities for data acquisition that self-report methods do not. This technology allows data to be collected objectively and continuously over the course of several days to weeks, long enough to reveal a person's "normal" activity pattern (Sasaki, Hickey, Staudenmayer, Kent, & Freedson, 2017; Dillon et al., 2016). While some consider accelerometry the gold standard for continuous physical activity monitoring (van Blarigan et al., 2017; Rejeski et al., 2016; Sanders et al., 2019), there are significant limitations. Accelerometry is limited by a lack of standardization in device wear location, collection duration, and data post-processing (Bassett, Troiano, McClain, &

Wolff, 2015). For example, the location of the accelerometer on the body affects which outcome measures can be obtained and their accuracies. Although wrist- and hip-worn accelerometers are more common, ankle-worn accelerometers have been shown to be more accurate for measuring step counts at slower gait speeds (Simpson et al., 2015; Klassen et al., 2016; Storti et al., 2008) and can be used for gait (Lee, Cho, Lee, Lee, & Yang, 2007) and balance (Turcot, Allet, Golay, Hoffmeyer, & Armand, 2009) analyses. This makes ankle-worn accelerometry a powerful tool for assessing ambulatory physical activity, especially in those who move slower.

Previous work has shown that combining accelerometry and a measure of heart rate increases the accuracy of estimating energy expenditure (Haskell, Lee, Evans, & Irby, 1993; Strath et al., 2001; Strath, Bassett, Thompson, & Swartz, 2002; Brage et al., 2004; Romero-Ugalde et al., 2017). However, this work has primarily used accelerometers located on the wrist, hip, chest, or thigh. While ankle-worn accelerometers have not been used to the same extent as wrist- and hip-worn sensors, one study (Pärkkä et al., 2007) found that ankle accelerometer data was more highly correlated with MET levels over a variety of prescribed tasks than were wrist- or hip-worn accelerometer data in a sample of adults aged 25-60 years. Therefore, it would be beneficial to develop a combined heart rate-accelerometer model that uses ankle-worn accelerometers due to the potential to improve intensity estimates compared to common current methods, as well as the additional gait and balance analyses that can be conducted with these data. As a result, the overarching objective of this thesis is to investigate the differences in measured activity intensity and volume by four wearables models.

1.2 Introduction to Accelerometry

Although it may seem more intuitive to measure speed of movement to describe physical activity, using acceleration (the rate of change of speed with respect to time) more closely reflects the energy requirements of movement since it is related to skeletal muscle force production (Hills et al., 2014). Various accelerometer outcome measures have been shown to be highly correlated with ambulatory energy expenditure (Hills et al., 2014) and many different physical activities (Esliger et al., 2011) making accelerometers a useful tool for measuring physical activity in free-living. However, accelerometers do not always accurately measure activity as detected acceleration does not always relate directly to intensity of movement, and accelerometers worn on different body segments may not accurately represent all types of activity based on which body segment(s) is/are active (Chen & Bassett, 2005). Accelerometers have been used to quantify physical activity since the early 1980s but did not become popular until the late 1990s/early 2000s as prices decreased and functionality increased (Troiano, McClain, Brychta, & Chen, 2014). Early accelerometers relied on piezoelectric materials to measure acceleration. When an internal weight accelerated, the piezoelectric material became strained and generated an electric current proportional to the acceleration. This type of accelerometer required a beam to attach the piezoelectric material to the weight which created a uniaxial accelerometer as the sensor was most sensitive to accelerations along the long axis of the beam (however, accelerations along any axis would have a small effect) (Chen & Bassett, 2005). More recently, integrated chip technology was developed which can measure acceleration in all three axes. These triaxial accelerometers have demonstrated higher correlations with energy expenditure than their uniaxial counterparts (Chen & Bassett, 2005; Hendelman, Miller, Baggett, Debold, & Freedson, 2002).

1.2.1 How Accelerometers Are Currently Used

Accelerometers can be used to describe physical activity using the FITT-V principle but despite the potential of accelerometers to capture high resolution, richly detailed characteristics of physical activity, summary measures are most often used due to the overwhelming amount of raw data obtained during multi-day collections. For wrist- and hip-worn accelerometers, the most common summary measure is the activity count which can be calculated two ways. The first is determined by counting the number of times the acceleration signal exceeds an arbitrary internal threshold. The second is determined by taking the area under the curve of the magnitude of the acceleration signal. Both these methods then involve summing those values over a specific time window (an epoch), most commonly 60 seconds (Chen & Bassett, 2005). These calculations yield dimensionless “count” units and the sum of gravitational units per epoch, respectively.

Without additional context, activity counts cannot be interpreted as they have no physiological meaning (Hills et al., 2014). To determine activity intensity from activity counts, thresholds (cut-points) are developed using regression or receiver operating characteristics techniques to determine count values that best represent the MET levels that correspond to light (1.5 METs), moderate (3.0 METs), and vigorous (6.0 METs) activity (Powell et al., 2017; Assah et al., 2011). Once epoched count values are compared to the cut-points, activity duration in each intensity category can be calculated using the number of epochs and the epoch duration (activity volume can be subsequently calculated). Data derived from epoched accelerometer data have temporal resolution equal to the epoch length. Although resolution is lost compared to the raw data, epoching data still leads to resolution far greater than the resolution of self-report measures and doubly-labelled water. It should be noted that if activities of varying intensities are performed during a single epoch that the calculated intensity would represent the average intensity during the epoch (Chen & Bassett, 2005). However, metabolic measures are commonly averaged over 30- or 60-second periods so this loss of resolution is somewhat unavoidable.

Cut-points are developed primarily in laboratory environments using two types of protocols. One type is treadmill based where participants walk and/or run at several speeds while their VO_2 is measured. For example, one of the most popular sets of cut-points which were developed for a hip-worn accelerometer was developed by Freedson, Melanson, and Sirard (1998) using this type of protocol. Participants walked at 4.8, 6.4, and 9.7 km h^{-1} on a treadmill. Using this development protocol may lead to low ecological validity. The other protocol type attempts to simulate free-living by having participants perform semi-structured tasks in the laboratory. These tasks often include housework (doing dishes or laundry, cleaning, etc.) or leisure activities (working on a computer, reading, etc.). This type of protocol may include treadmill or over-ground walking. (See Esliger et al., 2011 and Powell et al., 2017 for examples of this protocol type).

1.2.2 Gap in the Literature: The Use of Ankle Accelerometers

To date, cut-points have been developed for many age groups using accelerometers from many different manufacturers for the wrist and hip, however, none have been developed for use with ankle-worn accelerometers. Despite this gap in the literature, similar reasoning that has been applied to wrist and hip accelerometers could be applied to ankle accelerometers. Like the hip, ankle accelerometers capture lower and whole-body movement. Focusing specifically on gait since it is the most common physical activity (Hulteen et al., 2017), heart rate increases linearly with gait speed during both walking and running (Rotstein, Inbar, Berginsky, & Meckel, 2005). Additionally, VO_2 increases linearly with gait speed. Waters and colleagues (1988), the

ACSM (ACSM, 2014), and others describe this relationship using separate linear equations for walking and running.

Gait speed is increased by two main methods: by increasing cadence and/or increasing step length (Tudor-Locke et al., 2019). Until gait transitions from walking to running, cadence and stride length contribute equally to increasing speed. Once running, stride length is the greater contributor to further increases in speed (Terrier & Schutz, 2003). Prior to this point, there is a linear relationship between cadence and gait speed (Tudor-Locke et al., 2019). Although not a focus of their article, the relationship between cadence and gait speed during walking can clearly be seen using data reported in (Tudor-Locke, et al., 2019) (see Figure 1). Further, VO_2 is strongly correlated with cadence at walking speeds

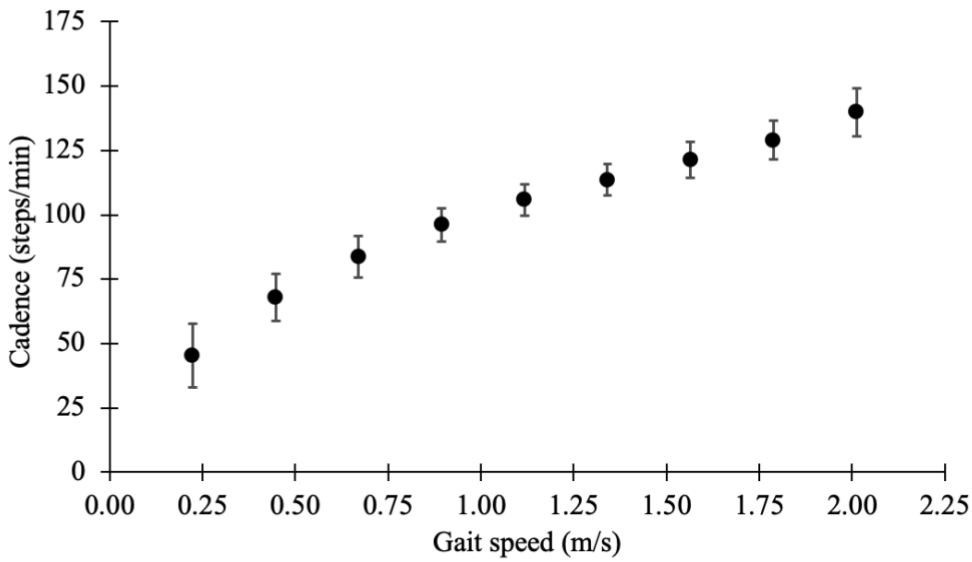

Figure 1: The relationship between gait speed and stepping cadence in adults aged 21 to 40 years (data from Tudor-Locke et al., 2019). Values are means ± 1 SD.

(Tudor-Locke et al., 2019) and regression equations have been developed to predict VO_2 using gait speed (Tudor-Locke et al., 2019; ACSM, 2014; Waters et al., 1988). Despite the potential influence of an individual's height which could affect their stride length for a given speed, Tudor-Locke and colleagues (2019) found that including leg length, as well as BMI, do not lead to improved performance of the gait speed-VO_2 regression. There is a clear opportunity to use ankle accelerometers to estimate activity intensity in ways similar to what has been done using wrist and hip accelerometers.

1.2.3 The Use of Stand-Alone Accelerometry

Despite its wide adoption as the gold standard for measuring free-living activity, there are several major shortcomings of using accelerometry on its own. Although using cut-points is convenient because it greatly reduces the data volume, the post-processing burden, and interpretation difficulty, it introduces several sources of error. Firstly, cut-points are specific to the device (Trost, McIver, & Pate, 2005; Bassett et al., 2015), wear location, population and types of tasks used in the protocol in which they were developed (Migueles et al., 2017) making it difficult to make valid comparisons between studies. Further, cut-points rely on the length of the epoch in which they were developed. Mathematically, cut-points can be linearly scaled to different epoch lengths, but concerns have been raised about the validity of this approach (Aguilar-Farias, Brown, & Peeters, 2014; Nilsson, Ekelund, Yngve, & Sjostrom, 2002). Epochs of different lengths are considered more or less appropriate for different age groups based on how long typical activity bouts last for that population (Migueles et al., 2017) so to accurately measure activity intensity, cut-points need to be developed for multiple epoch lengths and for different populations.

Secondly, the protocols used to develop cut-points often rely on the assumption that resting VO_2 (1 MET) is 3.5 mL kg^{-1} min^{-1}. Byrne and colleagues (2005) found that this definition overestimated resting VO_2 in a sample of 769 adults between the ages of 18 and 74 years by 35% and similarly, Hall and colleagues (2013) measured the average resting VO_2 of 20 older adults (aged 60-90 years) to be 2.66 mL kg^{-1} $minute^{-1}$. However, in the development of their cut-points for the GENEActiv accelerometer (ActivInsights, Kimbolton, UK), Powell and colleagues (2017) measured the average resting VO_2 of 56 adults with a mean age of 39.9 years to be 3.27 ± 0.62 mL kg^{-1} $minute^{-1}$, showing that the assumed 3.5 mL kg^{-1} min^{-1} does not always overestimate resting VO_2 by such a large margin. Further, individual fitness levels are not accounted for when using standard MET values and ranges. Depending on maximum aerobic capacity, the MET ranges and their intensity classification do not always correspond to the common definition of moderate activity as 40-60% heart rate reserve (HRR) and a given MET level for unfit individuals is relatively more intense due to their lower cardiorespiratory capacity (Ozemek, Cochran, Strath, Byun, & Kaminksy, 2013). Since standard MET ranges and resting VO_2 values may not be valid on a population level (Hills et al., 2014; McCracken et al., 2018), calibration based on an individual's resting and maximal VO_2 values is suggested to obtain more valid results at the individual level (McCracken et al., 2018) but this is an expensive and time-consuming task (Villars et al., 2012).

Thirdly, although strong correlations have been found between activity counts from the hip and gait speed (Trost et al., 1998; Rowlands, 2007), estimates of energy expenditure during running are likely underestimated (Hills et al., 2014) and standard error in predicting energy expenditure is greater during running than walking (Trost et al., 1998). Energy expenditure and activity counts during different activities also vary depending on which body segments undergo the most movement during the activity and where the accelerometer is worn. For example, wrist- or hip-worn accelerometers do not capture cycling activity accurately (Welk, 2002) and hip-worn accelerometers become less accurate during slow gait (Storti et al., 2008). Ankle-worn accelerometers remain accurate during both these activities (Storti et al., 2008; Foster et al., 2005). Multiple accelerometers can be used to gain a more comprehensive understanding of context but activities that include load carrying or changes in elevation are not reflected by activity counts (Hills et al., 2014) regardless of the number or location of accelerometers. The issues of decreased accuracy at high intensity activity or during increased workload without a change in total acceleration quantity warrant approaches that combine accelerometry with physiological measures to better capture activity intensity during these unique situations.

1.3 Heart Rate

A relatively simple and inexpensive physiological measure related to activity intensity is heart rate. While heart rate is not a direct measure of intensity or energy expenditure, similar to accelerometry, heart rate is related to VO_2 (Swain & Leutholtz, 1996; Hills et al., 2014). With increased energy requirements during activity, oxygen demand is met by increasing cardiac output through increasing heart rate (Plowman & Smith, 2007). Across most submaximal activity, VO_2 and heart rate both increase linearly until they plateau at maximum intensity (Laughlin, 1999; Opondo, Sarma, & Levine, 2015). Although the slope of this relationship varies between people, within an individual it has been shown to be consistent independent of the activity type (Hills et al., 2014). Using regression, heart rate can predict VO_2 accurately with a standard error of estimate of less than 6% of maximum VO_2 (Londeree & Ames, 1976).

The most basic way to report heart rate is beats per minute (bpm). When used to describe activity intensity, this absolute heart rate measure does not allow comparisons between individuals since changes in beats per minute for a given change in absolute intensity will vary between individuals due to factors including age and cardiovascular fitness (Hills et al., 2014). Heart rate can be expressed in relative terms which allow between-person comparisons to be made. Since the development of the Karvonen method in 1957 (Karvonen, Kentala, & Mustala, 1957), heart rate has commonly been reported as a percent of heart rate reserve (HRR). This variable measures heart rate as a percentage of the difference between resting heart rate and maximum heart rate (Swain, Leutholtz, King, & Branch, 1998): two variables which partially account for the individual's fitness level and age, respectively.

$$\text{Eq. (1): } \%\ \text{HRR} = \frac{\text{HR} - \text{resting HR}}{\text{maximum HR} - \text{resting HR}} \times 100$$

While it has widely been assumed that percent heart rate reserve and percent VO_2max are equivalent (Swain et al., 1998), in two studies by Swain and colleagues (1996 and 1998), the authors found that the regression of percent heart rate reserve on percent VO_2 reserve (VO_2 as a percent of the difference between resting VO_2 and VO_2max) had a slope closer to 1 and an intercept closer to 0 than the regression of percent heart rate reserve on percent VO_2max during both cycling and treadmill activity. In both studies, percent heart rate reserve explained more than 98% of the variance in percent VO_2 reserve ($r \geq 0.99$), suggesting that prediction using the percent heart rate reserve to percent VO_2 reserve relationship is valid and accurate across multiple activity types. Similar to how accelerometer cut-points classify activity intensity based on MET ranges, heart rate can be used to quantify activity intensity as a percentage of VO_2max or heart rate reserve by categorizing it into light (<45% VO_2max, < 40% HRR), moderate (45-75% VO_2max, 40-60% HRR) and vigorous (>75% VO_2max, >60% HRR) activity (ACSM, 2014; Hawley, Hargreaves, Joyner, & Zierath, 2014).

1.3.1 Equation-Based Estimations of Maximum Heart Rate

While maximum heart rate can be measured directly using maximal effort exercise protocols (Londeree & Moeschberger, 1984), it is possible to predict maximum heart rate using different equations that do not require exercise testing. This is particularly useful in studies where participants have physical limitations and would not want to or should not perform an exercise test due to physical of medical conditions. It is also useful in largescale studies where individually testing each participant would be too time consuming. Despite variability in maximum heart rate between individuals (Londeree & Ames, 1976; Nes, Janszky, Wisløff, Støylen, & Karlsen, 2013), the most common way to predict maximum heart rate includes only age as a factor: *HRmax = 220 – age* (Fox, Naughton, & Haskell, 1971). Although very simplistic, Londeree and Moeschberger (1984) found that age alone accounted for 71.4% of the variability in maximum heart rate. More recently, Tanaka and colleagues' meta-analysis (2001) (n=18 712) found that age accounted for 80% of the variability in maximum heart rate using the equation *HRmax = 208 – 0.7* x *age* and that maximum heart rate is not significantly affected by sex or cardiovascular fitness level. Further, both Tanaka and Franckowiak and colleagues (2011) found that predicting maximum heart rate using the Fox equation tended to be less accurate for individuals above the age of 40 years. Building upon the work of Tanaka and colleagues, Nes and colleagues (2013) found in a sample of 3320 apparently healthy participants that maximum heart rate could be predicted by the similar equation *HRmax = 211 – 0.64* x *age* with an r^2 value of 0.36 and standard error of estimate of 10.8 bpm. They also found that maximum heart rate was not significantly affected by gender, smoking status, or body mass index. Conversely, Whyte and colleagues (2008) developed four equations to predict maximum heart rate depending on sex and training status since they found that maximum heart rate declines at a different rate for males and females. Maximum heart rate can be predicted by *HRmax = 202 – 0.55* x *age* and *HRmax = 207 – 0.55* x *age* for trained and sedentary males, respectively, and by *HRmax = 216 – 1.09* x *age* and *HRmax = 221 – 1.09* x *age* for trained and sedentary females, respectively. Overall, their equations had an r^2 value of 0.330.

The prediction of maximum heart rate has also been studied in individuals who are overweight (body mass index [BMI] between 25 and 30 kg m^{-2}) and obese (BMI $\geq$ 30 kg m^{-2}). Miller and colleagues (1993) derived equations for healthy and obese populations using a modified Balke protocol and found that the Fox equation was "very similar" (p. 1080) to their equation $HRmax = 217 - 0.85 \times age$. They suggest that the Fox equation should be used clinically due to its accuracy and simplicity. Further, they found that maximum heart rate in obese individuals was better calculated using the equation $HR_{max} = 200 - 0.48 \times age$. The latter equation, as well as the Fox and Tanaka equations, were tested by Franckowiak and colleagues (2011) in a sample of individuals with a BMI greater than 25 kg m^{-2}. They found that the Fox equation was accurate across weight categories and in the 41-to-60-year-old subgroup but that it significantly overestimated maximum heart rate for those aged 20-40 years by approximately 6 bpm. The Tanaka equation was found to be accurate across all weight and age groups. The Miller equation significantly overestimated maximum heart rate for all weight groups by a mean of approximately 3 bpm and for all age groups. The authors concluded that using Tanaka's $HR_{max} = 208 - 0.7 \times age$ was the best predictor regardless of age, sex, or BMI. Figure 2 shows the differences in predicted maximum heart rate using these six equations across a range of ages. These equations tend to be more consistent for younger adults and predicted max heart rate values diverge with aging.

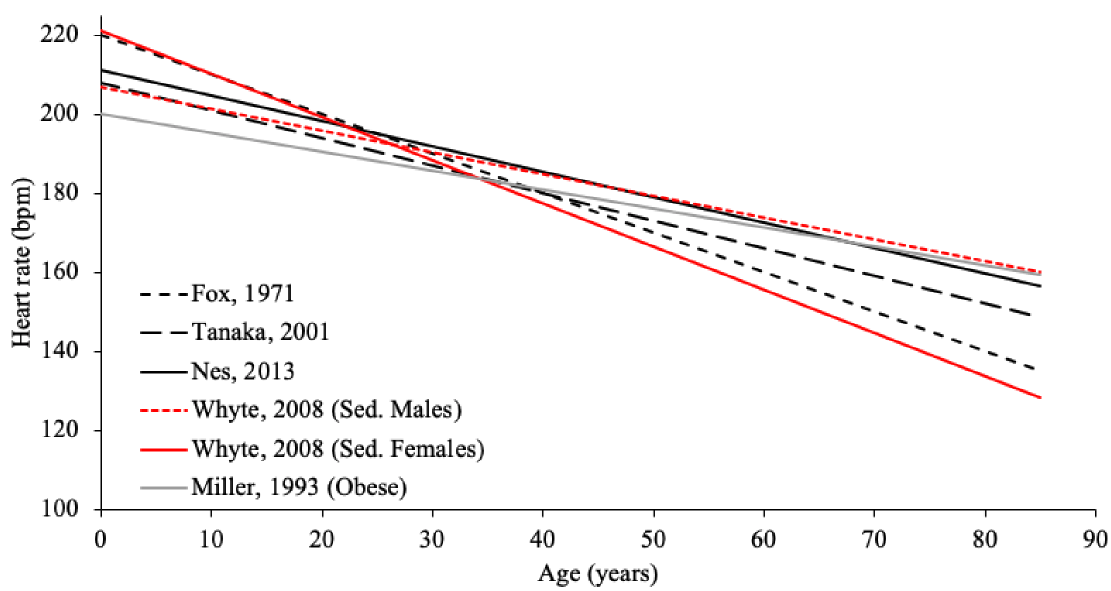

Figure 2: Maximum heart rate over the lifespan as predicted by six equations.

1.3.2 The Use of Stand-Alone Heart Rate

Although the heart rate to VO_2 relationship is quite strong, using heart rate on its own to estimate intensity and energy expenditure can lead to errors. Firstly, the slope of this relationship differs for upper- and lower-limb activities within an individual (Strath et al., 2001). Secondly, heart rate is less closely coupled to energy expenditure during rest and activity intensities that elicit lower heart rates (Hills et al., 2014) because it is affected by many non-activity factors. These factors include transient changes in environmental temperature, hydration, psychological stress as well as less transient changes such as cardiovascular fitness level (Brage et al., 2004). Because of this transient variability, it is difficult to determine whether changes in heart rate occur as a response to physical activity, from a change in output of the autonomic nervous system, or from some other variable that changes moment-to-moment.

One potential way to reduce error during rest or low intensity activity is the use of flex heart rate (Hills et al., 2014). Flex heart rate attempts to separate the non-linear low intensity portion from the linear higher intensity portion of the heart rate-VO_2 curve (Hills et al., 2014). Flex heart rate is commonly calculated as the average of the highest heart rate during rest and the lowest heart rate during a low intensity activity (Leonard, 2003; Ceesay et al., 1989) although there is no standardized definition or calibration protocol (Villars et al., 2012). The performance of heart rate on energy expenditure regression can be improved using flex heart rate. When heart rate is above flex heart rate, the typical heart rate on energy expenditure regression equation is used (Leonard, 2003). When heart rate is below flex heart rate, energy expenditure can either be substituted with the resting energy expenditure value (Freedson & Miller, 2000) or the average heart rate from lying supine, sitting, and standing postures can be used in the regression equation (Spurr, 1990). Figure 3 shows the heart rate-energy expenditure relationship with the flex heart rate marked.

After its development in the 1980s, several initial studies were conducted to assess the validity of using flex heart rate to estimate energy expenditure. Leonard's 2003 review provides details on three studies of male and female adults conducted between 1988 and 1993. Two of those studies compared flex heart rate to direct calorimetry and found that total energy expenditure calculated by the flex heart rate method was overestimated by 2.7% (Spurr et al., 1988) and underestimated by 1.2% (Ceesay et al., 1989), respectively. Similarly, the third study, which compared flex heart rate to doubly labelled water, found an overestimation of 2.0%. Leonard also cites several more recent studies on older adults with error ranging from a 9.7% underestimation to 5.9% overestimation. Validation studies have found that the flex heart rate method is accurate to ± 2 to 3% at a group level, that it is strongly correlated with energy expenditure measured by either calorimetry or doubly labelled water (r = 0.88), and that the regression of energy expenditure on heart rate has an intercept of approximately 120 kcal day^{-1} (Leonard, 2003). Other work has suggested that an individual's error can be up to ± 20% (Johansson, Rossander-Hulthén, Slinde, & Ekblom, 2006; Freedson & Miller, 2000) despite the accuracy of using flex heart rate at a group level. This disparity is potentially due to individuals whose heart rate is more often close to their flex heart rate where this method is least accurate (Freedson & Miller, 2000).

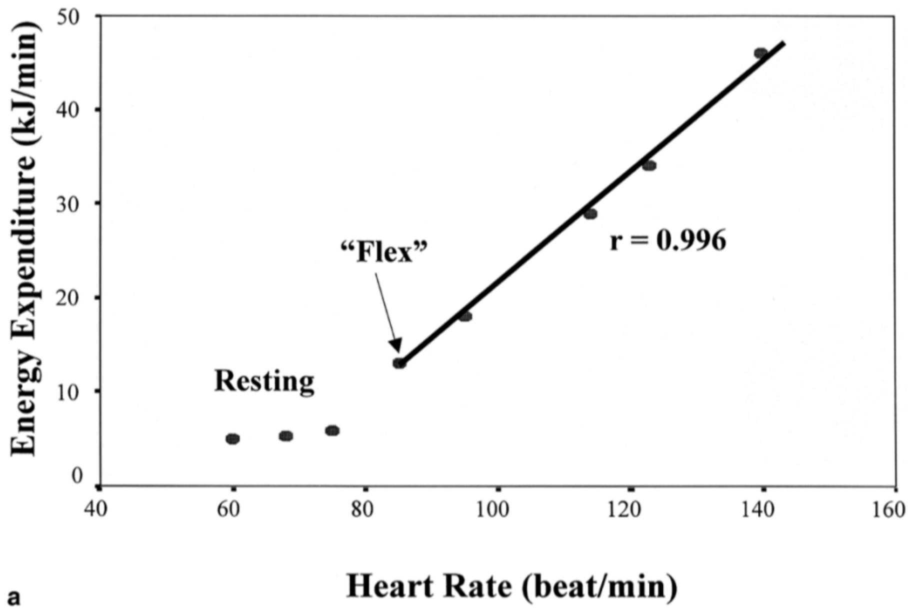

Figure 3: The relationship between absolute heart rate, energy expenditure, and flex heart rate (image from Leonard, 2003).

Although the use of flex heart rate reduces some of the error associated with measuring energy expenditure compared to using percent heart rate reserve or percent maximum heart rate, there still remain several challenges when using any heart rate-based measure. ECG signal quality is expected to remain consistent over short collection periods. However, for multi-day collections, signal quality can become an issue as the electrode-skin contact quality lessens to a point where the data may become unusable. For example, one study (Strath et al., 2002) measured heart rate using a chest strap for 6 hours and found that on average, participants had 329 minutes of usable data; this represents 8.6% of the collection being lost to poor signal quality. Another major limitation that remains is the lack of context as to why changes in heart rate are occurring. Flex heart rate reduces some of the physiological "noise" due to autonomic changes from stress or arousal that increase heart rate in the absence of movement, but the transition between rest and activity cannot be differentiated by a single heart rate (Freedson & Miller, 2000). Additionally, the heart rate response at the onset of physical activity is not instantaneous; heart rate takes approximately two minutes to reach steady state during submaximal, constant intensity activity (Plowman & Smith, 2007; Strath et al., 2000). Very rapid parasympathetic withdrawal occurs with the onset of movement and can increase heart rate by approximately 30 bpm in 4 seconds (Nobrega & Araújo, 1993). If energy demand is not met by decreasing parasympathetic output, heart rate will continue increasing through sympathetic activity although there is a delay of approximately 20 seconds (Hughson, Tschakovsky, & Houston, 2001). Once activity ends, heart rate does not instantly return to its resting level (Strath et al., 2000) and may decrease at a different rate than that at which it increased at the onset of physical activity. Taken without movement context, using heart rate alone may misrepresent the timing, duration, and intensity of a given bout of physical activity.

1.4 Combining Accelerometry and Heart Rate

By combining accelerometry with heart rate, there is the potential to further reduce errors generated by using either on its own since their sources of error are not positively correlated (Brage et al., 2004). Briefly, the rationale is that accelerometry can both provide movement context and predict energy expenditure with greater accuracy than heart rate at lower intensities (Meijer, Westerterp, Koper, & ten Hoor, 1989) while heart rate can provide greater accuracy at high intensities (Romero-Ugalde et al., 2017). Figure 4 shows the relationships of accelerometer counts and heart rate to oxygen consumption and demonstrates how these two techniques complement each other. (In the figure, "COHR" and "COACC" refer to the cut-offs of heart rate and accelerometer counts, respectively; the labels were cropped).

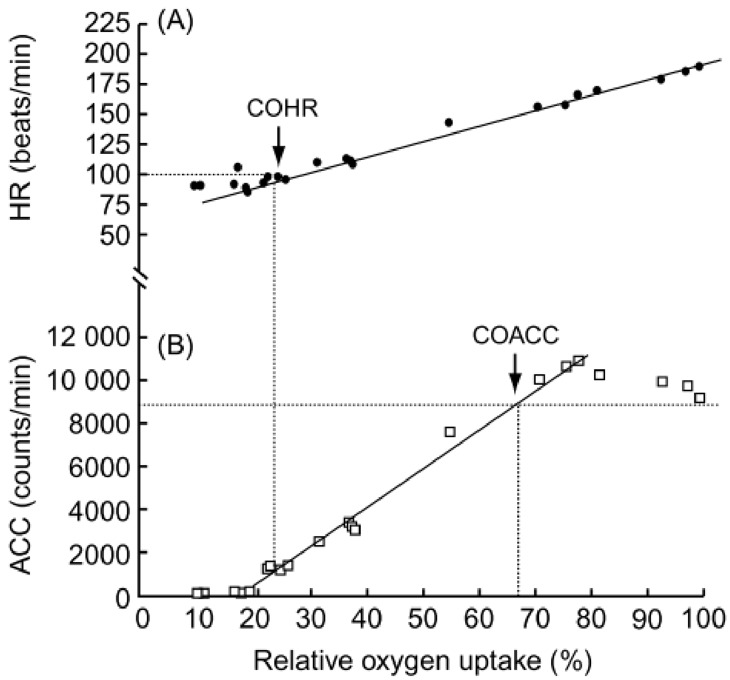

Figure 4: The relationship between oxygen consumption, hip accelerometer counts, and heart rate, showing the complementary accuracy ranges in the measurement of energy expenditure (image from Johansson et al., 2006).

1.4.1 Combined Heart Rate-Accelerometer Models

The simplest heart rate-accelerometer models use the accelerometer to ensure that changes in heart rate are caused by physical activity. If multiple accelerometers are worn, their counts can be compared to determine which body segment(s) was/were most active. Strath and colleagues (2001) created a model using this logic in which each participant performed leg and arm ergometer calibration protocols to create a regression equation of heart rate on VO_2 for each activity. Their model uses the ratio between the counts from the wrist- and thigh-worn accelerometers to determine whether activity was primarily upper- or lower-body. The appropriate regression equation is then used to estimate VO_2 from heart rate. Using this method improved the accuracy compared to heart rate alone (0.4 MET overestimation, $r^2 = 0.53$) or a hip-worn accelerometer alone (1.1 MET overestimation, $r^2 = 0.45$) and achieved an r^2 of 0.81 with a non-significant bias of 0.1 METs. In their follow-up study (Strath et al., 2002), an accelerometer threshold of 500 counts per minute was added as the cut-off between activity and rest and there was a more complicated selection process for which equation to use (see Figure 5). If one of the wrist- or thigh-worn accelerometers were above 500 counts per minute, the regression equation for that body segment was used. If both wrist and thigh counts were above 500 counts per minute, their ratio was used to determine which heart rate-VO_2 regression equation to use. In the case of whole-body physical activity, the lower-body regression was used because the heart rate-VO_2 relationship for lower-body physical is very close to that of whole body (Haskell, Lee, Evans, & Irby, 1993). In the validation of their model,

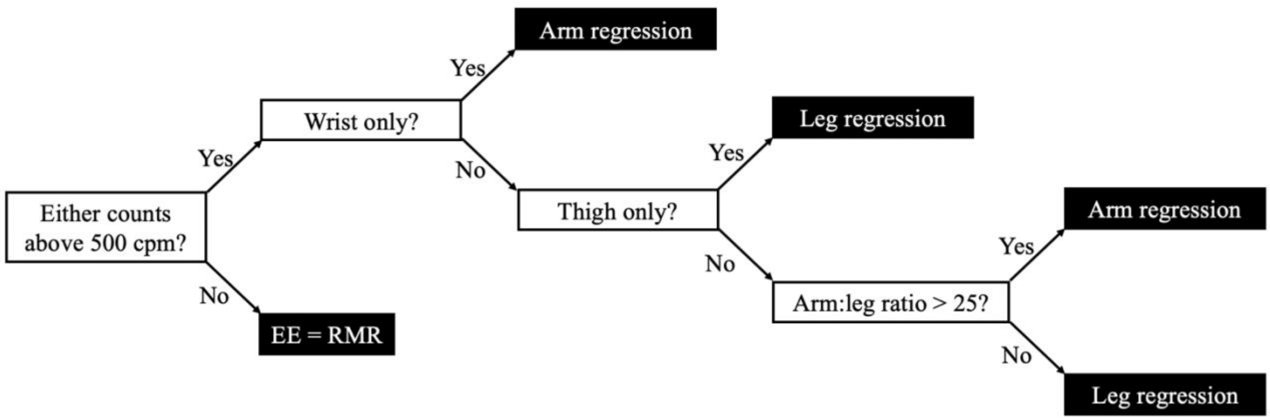

Figure 5: Combined heart rate-accelerometer model from Strath and colleagues (adapted from Strath et al., 2002).

participants had indirect calorimetry measured for 6 hours in free-living. On average, total activity volume calculated by the heart rate-accelerometer model measured 1 MET minute less than the criterion measure (748 ± 178 compared to 749 ± 138 MET minutes, respectively), and the combined model was more accurate than using flex heart rate on its own.

Similar to the two Strath studies, Brage and colleagues (2004) developed a model that uses a hip accelerometer threshold to verify that physical activity is occurring. Brage incorporated regression equations from individual calibration protocols into a branched model. In the algorithm that determines which equation to use, both accelerometer counts and heart rate are compared to respective thresholds and one of four equations is used depending on the which values are above/below threshold. This model also proved to be highly accurate. With individual calibration, error in energy expenditure prediction was an underestimation of 2.36% and with group-level calibration, error was an overestimation of 0.54% with r^2 values of 0.61 and 0.78, respectively. Interesting, the authors attribute the improved performance of the group-calibrated model to errors in individual calibration being larger than the between-person variance.

Johansson and colleagues (2006) took a similar approach in the development of their model (see Figure 6). Using an accelerometer placed on the low back, they found that two linear

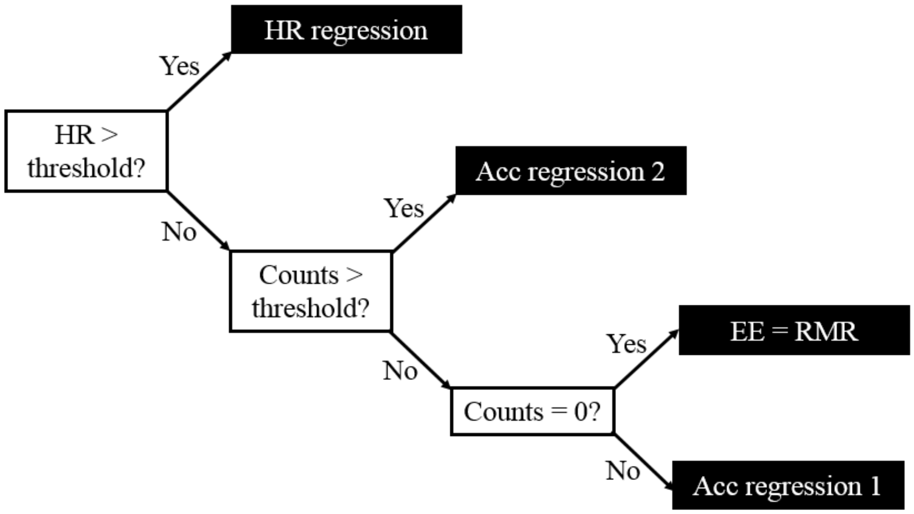

Figure 6: Combined heart rate-accelerometer model from Johansson and colleagues (adapted from Johansson et al., 2006).

accelerometer-energy expenditure regression equations provided better estimates of energy expenditure than did using a single equation. One of these equations corresponded to activity less intense than walking at 2 km h^{-1} while the other corresponded to the intensity range between walking at 2 km h^{-1} up to approximately 75% of VO_2max. They also measured flex heart rate and added an additional 10 bpm to the calculated value to ensure the heart rate-VO_2 relationship was linear above that threshold. For heart rates below flex heart rate, resting metabolic rate was used when accelerometer counts were zero (it was assumed that the participant was asleep), otherwise the appropriate accelerometer equation was used depending on the accelerometer count. For heart rates above flex heart rate, the heart rate-VO_2 regression equation was used. This model achieved a root-mean-square error (RMSE) of 2.99 MJ per day (714 kcal day^{-1}) which was the same as accelerometry alone (RMSE = 2.99MJ day^{-1}) and better than flex heart rate alone (RMSE = 3.99 MJ day^{-1} or 953 kcal day^{-1}).

Lastly, Romero-Ugalde and colleagues (2017) developed a three-equation branched model for a hip-worn accelerometer using twenty-five activities in the individual calibration procedures. Although this model was created using individual calibration, the regression coefficients that were used were from the group-level calibration. Cut-off values were normalized at the level of the individual by including their maximum counts per minute and maximum attained heart rate values as scaling factors. Heart rate was quantified in this model as the number of beats above resting heart rate ("net heart rate") and +40 bpm was used as their flex heart rate. This was done to represent physical activity intensity in the moderate and above range. Validation of this model found that the median R^2 value of 0.87 was higher than the concurrently tested nonlinear accelerometer-only model ($r^2 = 0.66$), linear accelerometer-only model ($r^2 = 0.56$), and heart rate-only model ($r^2 = 0.81$). Further, the branched model accurately classified physical activity into light or moderate-to-vigorous intensity 81.6% of the time; this was the best performance of the tested models (range = 72.6% to 80.5%). While the branched model and heart rate-only model both performed well, the authors stated that the main differences between the models was the improvement in accuracy of the branched model at low intensity physical activity (Romero-Ugalde et al., 2017).

These four combined heart rate-accelerometer models use the data differently but do so using the same underlying reasoning. They also have several similarities. First, they all implement regression equations for the heart rate-VO_2 and accelerometer counts-VO_2 relationships that were calibrated either at an individual or group level. Second, either a heart rate or accelerometer threshold was used to mark the transition point from rest to activity or similarly, between the nonlinear and linear portions of the heart rate-VO_2 relationship. Third, below this threshold, accelerometer data is used to predict energy expenditure while heart rate is used if it's above this threshold. Finally, the use of a combined heart rate-accelerometer model improves prediction accuracy compared to heart rate-only and accelerometer-only models.

1.5 Thesis Objectives and Hypotheses

This thesis aims to answer three questions which will be described as three separate objectives. To answer these questions, activity profiles will be generated using an existing wrist-worn accelerometer model, by developing an ankle accelerometer-based model, by a heart rate model, and by developing a combined heart rate and ankle accelerometer model. Activity profiles are defined as here as an epoch-by-epoch classification of activity intensity for each model. These data can then be summed to calculate activity volumes.

Objective 1: to determine if four wearables models measure the same amount of activity at each intensity. It was hypothesized that the models would measure a significantly different amount of total activity at each of the activity intensities.

Objective 2: to determine to what extent wearables models agree on their classification of activity intensity on an epoch-by-epoch basis. It was hypothesized that models would have a moderate agreement ($.40 \leq$ Cohen's kappa $\leq .60$) (Cohen, 1960) in intensity classification.

Objective 3: to determine whether an individual's overall activity level influences the between-model differences in activity volume (Objective 3A) and epoch-by-epoch agreement in intensity classification (Objective 3B). It was hypothesized that overall activity level would have a significant effect on the between-model differences.

Chapter 2: Methods

2.1 Project Overview and Protocol

2.1.0 Project Overview: ONDRI@Home

The data used in this thesis is part of the Ontario Neurodegenerative Disease Research Initiative (ONDRI) "@Home" project. The primary objectives of this project are to develop tools and data management systems to improve early diagnosis and to track disease progression for those at risk of or living with neurodegenerative disease. These data then need to be reported to patients and healthcare practitioners in meaningful ways. The data used in the present thesis were collected as part of the ONDRI@Home control cohort.

The pilot stage of the ONDRI@Home project began in the Fall of 2017. While not the focus of the work in this thesis, my role in the project included being the primary team member in charge of determining what wearable devices to use, working on protocol development including researching study collection lengths, and data processing methods. I also led the initial pilot project which collected data from retirement homes and laid the groundwork for the collection protocol used in data used in this thesis.

2.1.1 Participant Recruitment

Participants were recruited by word of mouth and through posters around the research facility. These participants included family, friends, University staff, participants in exercise programs held at the Centre for Community, Clinical and Applied Research Excellence, and undergraduate students. Data collections took place between December 2018 and March 2020. To be eligible to participate, prospective participants needed to provide informed written consent and have no diagnosis of neurodegenerative disease. Participants provided medical history which included diagnoses and medications. Effects of medications were checked to ensure they would not have an effect on heart rate. No participants were excluded due to medication use.

Any participants taking medications that would affect their heart rate response to physical activity (e.g. beta-blockers) were excluded from analyses. This study was approved by the University of Waterloo's Office of Research Ethics (ORE #31943).

2.1.2 Protocol and Equipment

On the first day of the study, participants filled out demographic, medical history, and medication forms. Weight was taken using a physician's scale and height was taken using a sliding scale fixed to the wall. Measures were rounded to the nearest tenth of a kilogram and nearest centimeter, respectively. A research assistant familiarized participants with the wearables including instructions on how and when to remove and re-attach the devices. Participants wore multiple devices for a period of 4-7 days (due to protocol changes as the project evolved), including overnight, and were instructed only to remove the wearables for bathing or water-based activities and to maintain their normal daily activities. The collection duration varied due to some protocol changes that occurred during the study period.

GENEActiv accelerometers (ActivInsights, Kimbolton, UK) were worn on both wrists and ankles. GENEActivs are small (43 x 40 x 13mm), lightweight (16 g), triaxial accelerometers that measure raw acceleration in the range of ± 8 G (where 1 G = 9.81 m s^{-2}). Accelerometer data can be sampled at frequencies from 10 to 100 Hz. Data was collected at a frequency of 75 Hz to maximize temporal resolution while maintaining adequate battery life for the collection period. All GENEActivs were initialized to begin recording at the same time and for a set duration according to when the participant was scheduled for their end-of-study meeting with the research assistant.

Figure 7: Ankle and wrist GENEActiv accelerometer attachment methods showing the use of custom-made sleeves (left) and original watchband (right).

The wrist worn GENEActivs were worn on the original rubber strap on the posterior aspect of the distal forearm as a wristwatch would be worn. Ankle-worn GENEActivs had the strap removed and were fitted into custom-made medical grade tensor wraps which were held on using hook-and-loop fasteners (see Figure 7). Participants were given the option of wearing the ankle GENEActivs medially or laterally depending on perceived comfort. These accelerometers were worn proximal to the malleoli. For this thesis, no raw data was used so device orientation was not important.

Participants also wore a Bittium Faros 180º (Bittium Corporation, Oulu, Finland) on their torso. The Bittium Faros contains an electrocardiogram (ECG) and measures raw triaxial acceleration (data not used in the present thesis). It is 48 x 29 x 12 mm and 16 g in size. ECG can be sampled at 125, 250, 500, or 1000 Hz. 250 Hz was selected to allow a battery life of approximately 4-5 days while maintaining adequate signal resolution. Once initialized, the Bittium Faros collected data until its battery died. Each participant was given two single-lead FastFix (Bittium Corporation) electrodes that attach to the Bittium Faros via micro USB. Participants were given the choice to wear the electrode vertically (on the left lateral border of the sternum), diagonally (at approximately the level of the left 5th and 6th ribs with the medial end of the electrode superior to the lateral end) or horizontally (at approximately the level of the 2nd and 3rd ribs) depending on their anatomy and perceived comfort. Figure 8 shows the attachment options for the Bittium Faros. The second electrode was given in case the participant

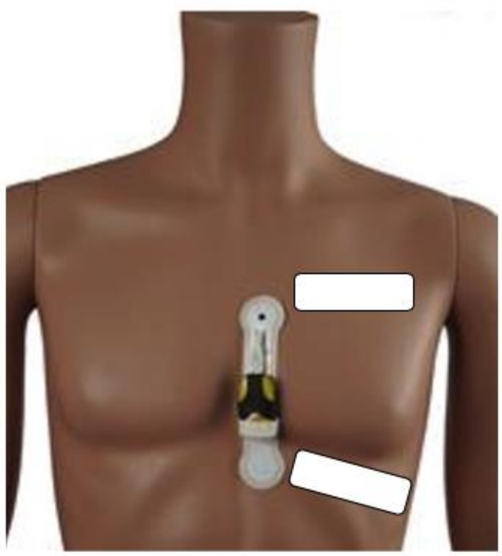

Figure 8: Bittium Faros attachment locations using the FastFix electrode (image adapted from the Bittium Faros user manual, Bittium Corporation).

experienced poor adhesion or needed to remove the first electrode. These electrodes can get wet; however, participants were instructed to remove the Bittium Faros from the FastFix electrode and to re-attach it once water activities were complete.

Partway through data collections, a treadmill protocol was implemented for the purpose of developing an ankle accelerometer-based activity model. At a time that was convenient during their collection period, participants performed a series of walks on a treadmill with zero incline and without using the handrails. Participants began by self-selecting their preferred/comfortable gait speed. They then walked for 2 minutes at both 60% and 80% of this speed, followed by 4 minutes at 100%, 120%, and 140% of preferred pace. Approximately 90 to 120 seconds of rest were given between each speed. A cool-down period at a self-selected speed was performed following the final walk. These speeds were selected to approximate a range of speeds likely to occur in free-living but without the need for participants to jog or run. This was done to alleviate potential safety concerns if this protocol were to be implemented in cohorts with

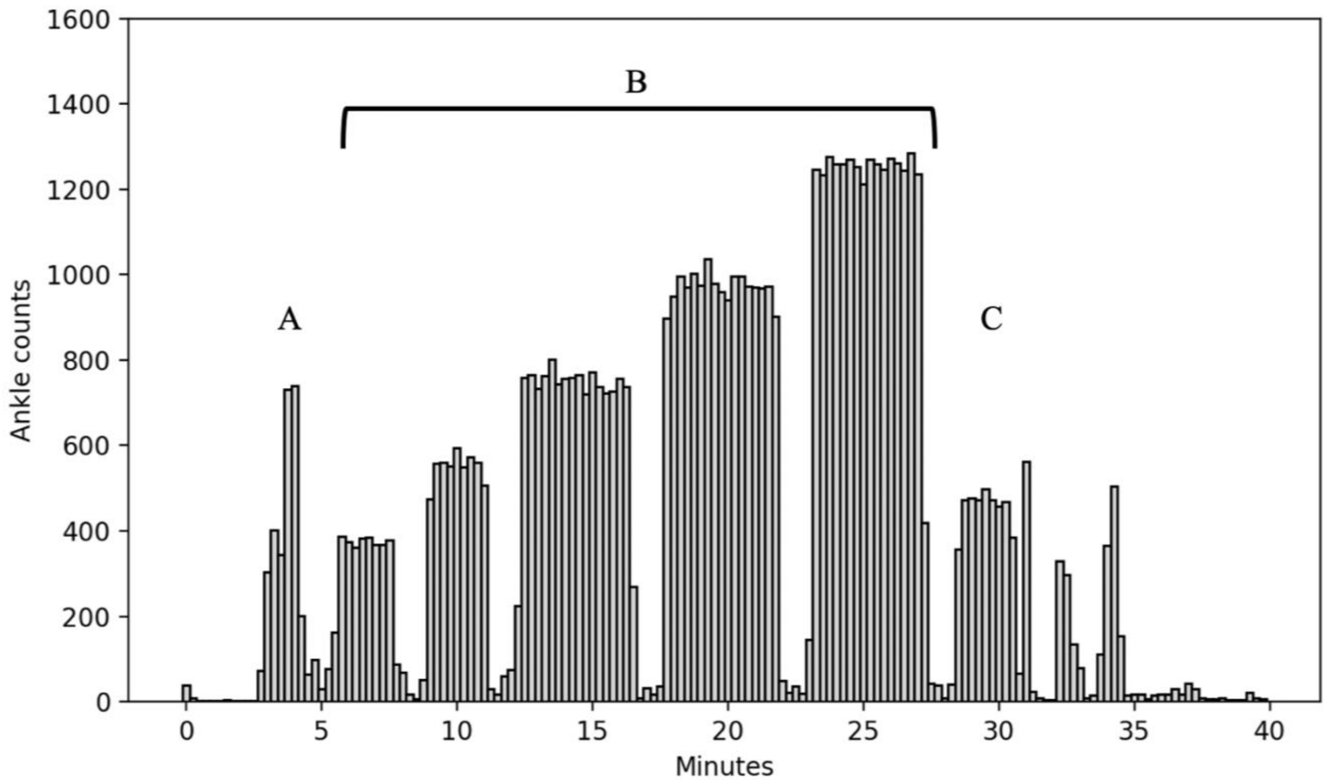

Figure 9: Annotated ankle accelerometer data during the treadmill protocol. Each bar represents one 15-second epoch. Protocol stages and corresponding walking bouts: determining preferred pace (A), walking at speeds ranging from 60% to 140% of preferred pace (B), and cool-down (C).

neurodegenerative disease and/or musculoskeletal limitations. Figure 9 shows annotated epoched ankle data for the treadmill protocol from one participant.

Participants were also given device removal and sleep logs to fill out during the course of data collections. Sleep and device removal logs were used in the present work to help determine when participants went to bed, took naps, and removed any or all devices.

2.1.3 A Priori Power Calculation

An a priori sample size/power analysis was conducted using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) to determine how many participants would be required to detect differences in activity profiles generated from the four models with a power of $\beta = .80$ and $\alpha = .05$. This analysis was conducted using the between-model effect sizes from Brage and colleagues (2004) and Johansson and colleagues (2006), which both compared energy expenditure measured by accelerometer-only, HR-only, and HRAcc models. These effect sizes ranged from $d = .65$ to 2.78; the effect size averaged across all model comparisons was $d = 1.79$. Using these data, sample size calculations ranged from 4 to 21 participants, with 5 of the 6 between-model differences requiring a sample size of 10 participants or fewer. Since differences between groups with different activity levels will be analyzed for Objective 3, power analysis was also conducted based on the expected differences in activity levels between these two groups. This analysis was conducted using data from previous work (Westerterp, 2001) which assessed the physical activity level ratio (ratio of total energy expenditure to resting energy expenditure) of 173 participants between the ages of 20 and 50 years. Using the difference in physical activity level ratio between the first and fourth quartiles of 1.34 standard deviations, it was determined that 7 participants per group would be needed to attain $\beta = .80$ with $\alpha = .05$. A sample size of 28 participants was estimated to be necessary to have sufficient power to address both research questions.

2.2 Data Pre-Processing

2.2.1 Accelerometry Pre-Processing

After the collection was finished, raw GENEActiv data were extracted using the GENEActiv software version 3.3 (ActivInsights) and saved to the NiMBal Lab's secure network drive. Upon extracting the raw data, the GENEActiv software calculates the amount of clock drift experienced by the on-board clock of each device relative to the computer's clock. Interpolated datapoints were added or subtracted periodically to ensure proper sample timing. This ensured that timestamps from different GENEActivs matched throughout the entire collection period. The clock drift quoted by the manufacturer is ± 1.7 seconds per day (11.9 seconds per week); a different part of the ONDRI@Home project found similar results.

Using custom-made Python software (see Acknowledgments), data files were converted to European Data Format (EDF) to create a standardized data format for the ONDRI project. Accelerometer data were not filtered to maintain consistency with the methods used in the existing Wrist model (see section 2.5.1).

2.2.2 ECG Pre-Processing

Bittium Faros data were downloaded and stored in the same location as the GENEActiv data. No file conversion was required as the data are already stored in EDF format.

Work was conducted to determine an appropriate QRS peak detection algorithm to calculate heart rate from raw ECG data. Several algorithms showed high accuracy with a clean ECG. However, ECG signal quality showed a high degree of variability, often to the point where the signal was unusable despite filtering. It is suspected that the decrease in signal quality was from degrading skin-electrode contact or from residual moisture following the FastFix electrode getting wet. Different techniques were implemented to improve the performance of peak detection algorithms including running the algorithms on windowed data (15-second sections) and adjusting the temporal and magnitude parameters of the peak detection algorithms; these attempts were not successful due to extremely noisy ECG sections (see Figure 10 for an example of various ECG signals from a single participant). Ultimately, an algorithm developed by

Orphanidou and colleagues (2015) to detect periods of usable ECG signals was implemented. This algorithm was developed using single-lead ECG data from several datasets using multiple ECG devices. Data used in the algorithm development included 24-hour wear as well as isolated 10-second segments of ECG data. In total, the authors validated their algorithm using 1500 ten-second segments of data. Algorithm performance was quantified by comparing the output from the algorithm with visual inspection from two researchers. For these inspections, the researchers designated ECG segments as usable if they could "confidently derive a reliable [heart rate] from it, by counting the number of salient features (such as R-peaks) … over fixed time intervals" (p. 834). Overall, it was found that 64% of the data were usable and the algorithm achieved a sensitivity of 94% and specificity of 97% in its ability to detect usable segments of data.

ECG data in the present thesis were filtered using a 1-30Hz, 2nd order Butterworth bandpass filter to reduce baseline wander and high-frequency noise generated from muscle activity or electrical interference (Pouryayevali, Wahabi, Hari, & Hatzinakos, 2014). The Orphanidou algorithm does not specify signal filtering. The filtered data were input into the Orphanidou algorithm in 15-second segments. The algorithm operates in two stages: peak detection and a series of condition checks. Peaks were detected using the Python package

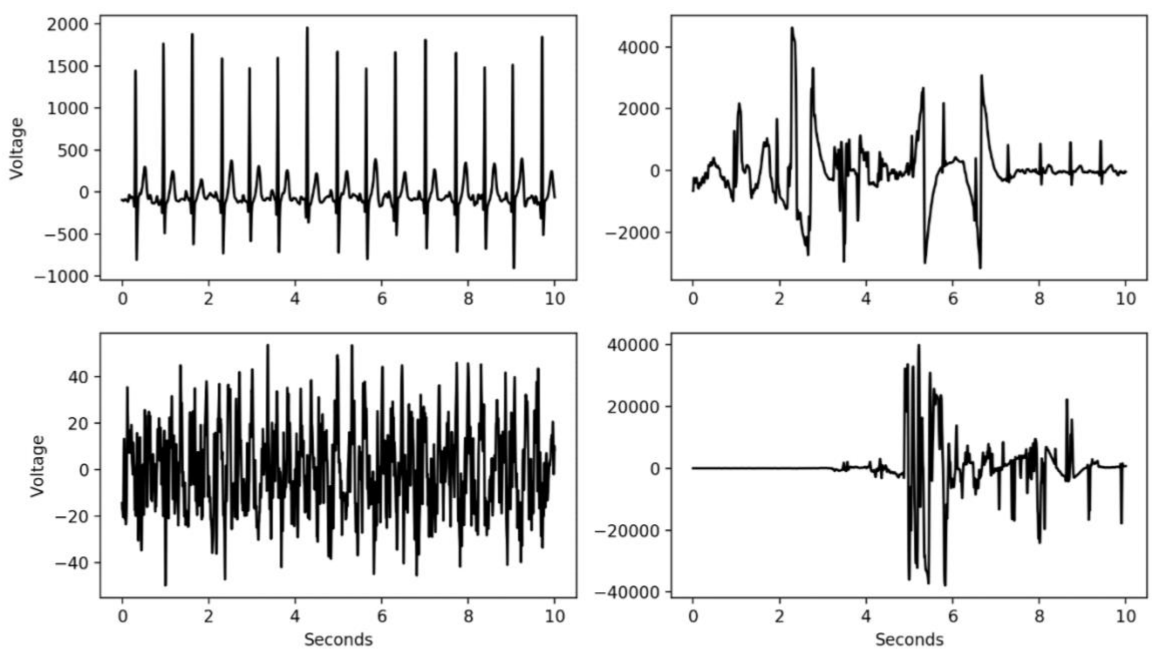

Figure 10: Four 10-second segments showing the variability in filtered ECG signal quality from one participant. Shown are a clean segment (top left), segment with a combination of noise and clean signal (top right), a period when the device was not worn (bottom left), and a highly noisy region (bottom right; note the voltage amplitude relative to other segments).

ecgdetectors that implements a wavelet transformation and the Pan-Tompkins peak detection algorithm (Pan & Tompkins, 1985).

The second stage of the algorithm involves passing a series of conditions. All conditions must be passed for the data window to be deemed usable. First, based on the number of detected peaks and their timing, the average heart rate must be between 40 and 180 bpm. Second, no R-R interval can be greater than 3 seconds. Third, the ratio of the longest-to-shortest R-R intervals must be less than 2.2. Fourth, a correlational analysis is performed. A "template" heartbeat is created by taking a window of data centered around each detected peak with a width of the duration of the median R-R interval. If a window extended passed the start or end of the 15-second data segment, its peak was not used in the template. These windows are then laid on top

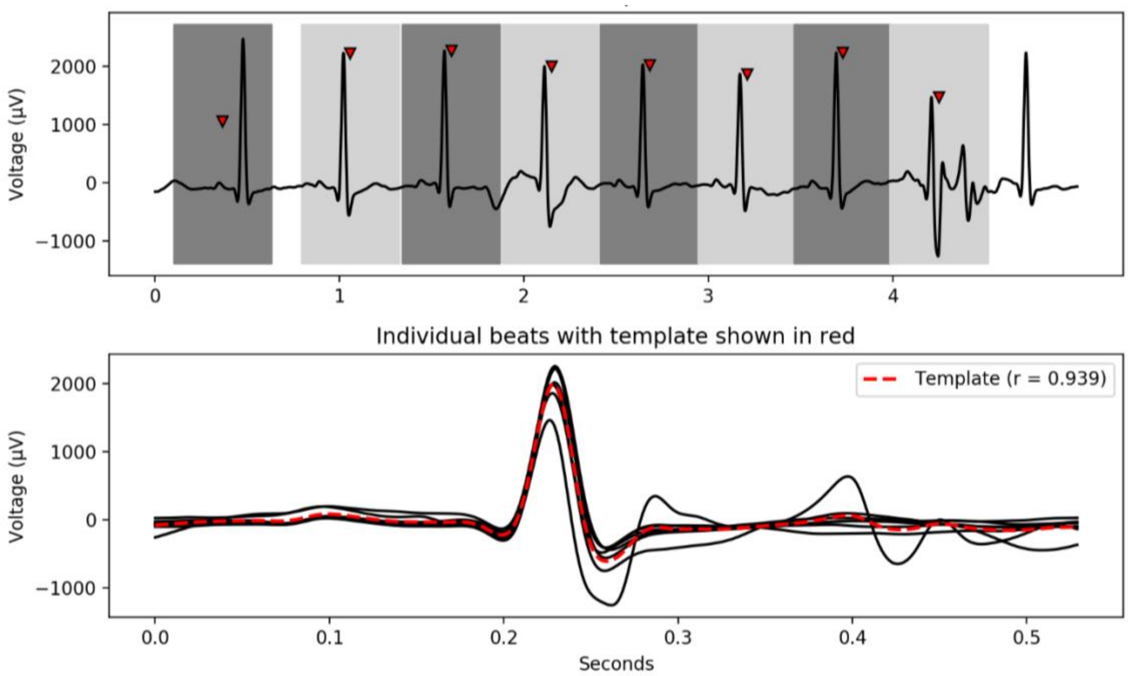

Figure 11: A 5-second segment of filtered ECG data showing detected beats (top) and the resulting QRS template used in the Orphanidou et al. quality check algorithm (bottom). Individual beats are drawn in black with the template in red.

of each other, so to speak, and the template is created by taking the average voltage at each time point. The top subplot in Figure 11 shows a sample 10-second filtered ECG segment with approximate peak locations marked and peak windows shaded in grey (there are 8 windows; the final beat was excluded because it extended beyond the data segment). The bottom subplot shows each peak window overlaid in black and the template in red. Then, a Pearson correlation is calculated between each peak window and the template. These values are then averaged. If the average Pearson correlation is greater than .66, this condition is met. Figure 12 shows the algorithm processing steps.

Further testing showed that this algorithm often failed to reject periods when the Bittium Faros was not worn. These regions are partially differentiated from periods when the device was

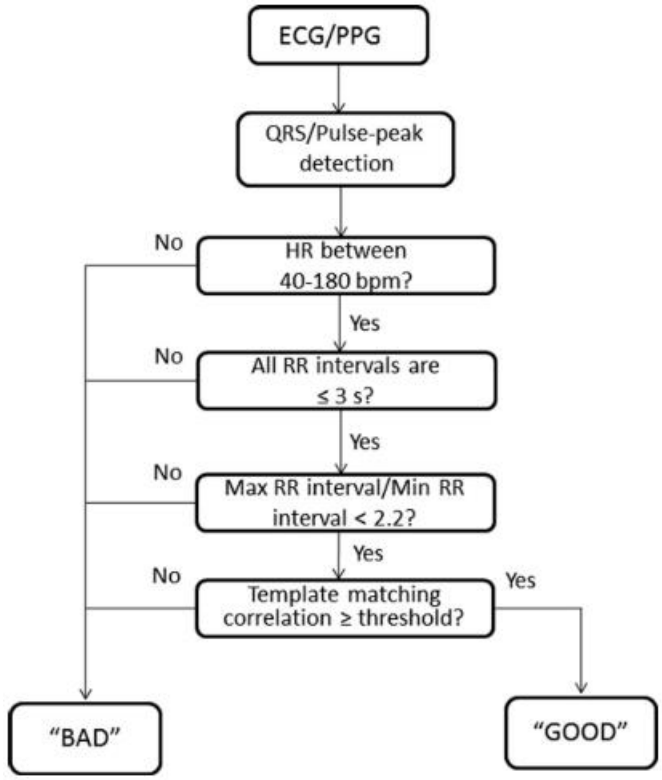

Figure 12: Flowchart showing the processing steps in the original signal quality algorithm (image from Orphanidou et al., 2015). The added voltage range rule is not shown.

worn by a relatively low signal amplitude. An additional condition (not included in Figure 12) was added to improve the identification of these periods; a voltage range of ≥ 250 μV was required for each data segment. If all five of these conditions were met, the 15-second segment was classified as usable.

The modified algorithm was validated using 1 000 randomly generated 15-second segments of ECG data where the author determined whether the signal was contaminated by noise to a degree which would affect the reliability of beat detection or if the signal was relatively noise-free. These judgements were then compared to the output from the algorithm to quantify algorithm performance.

2.2.3 Device Synchronization

Synchronization of the timestamps on the GENEActivs was described in section 2.2.1. Because the Bittium Faros was started manually (unlike the GENEActiv, it cannot be set to start recording at a specific time), files from each device needed to be cropped so that the epochs from all devices contained the same time periods. The last device to start collecting data was not cropped; the other files were cropped to excluded data before this point in time. Similarly, all devices were cropped to exclude timepoints once any device stopped collecting data. Data cropping at the start of collection was typically less than a few minutes' worth of data. Due to the difference in battery life between the GENEActivs and Bittium Faros, a few days' worth of data were cropped from the GENEActiv files from the end of their collection once the Bittium Faros stopped collecting.

While the GENEActivs timestamps were correct for clock drift, the Bittium Faros' on-board clock was taken as correct because the amount of clock drift is not calculated by the Bittium Faros.

2.2.4 Data Epoching

Methods used by Powell and colleagues (2017) served as the starting point for the development of data analytics in this thesis due to it being an established method. Consistencies in data processing and analytics were maintained where possible and where appropriate to improve the validity of making between-model comparisons. This model will be further explained in section 2.5.1.

Data from the synchronized accelerometer and ECG data were windowed into 15-second epochs. Accelerometer epoching was calculated using the equation found in the GENEActiv software manual. This is the standard way to epoch GENEActiv data and was the method used by Powell and colleagues. The sum of vector magnitudes (SVM; synonymous to activity counts) value for each epoch was calculated with Equation 2 where i is the epoch index number, n is the index of the first raw data point in the epoch, l is the length of the epoch in seconds, f is the sampling frequency in Hz, and g is gravitational acceleration.

$$\text{Eq. (2): SVM}\left(\text{g s epoch}^{-1}\right) = \sum_{n = i \times 1 \times f}^{(i+1) \times 1 \times f - 1} \left| \sqrt{x_n^2 + y_n^2 + z_n^2} - 1\text{g} \right|$$

ECG data were also epoched using the same 15-second interval. Peaks were given in datapoint indexes which were converted to time in seconds by dividing by the sampling rate. The average heart rate for each epoch was calculated using Equation 3 where b is the number of detected beats, f is the sampling rate in Hz, and n_{end} and n_{start} are the index numbers of the first and last data points, respectively.

$$\text{Eq. (3): Heart rate}\left(\text{beats min}^{-1}\right) = \frac{\text{b - 1}}{\text{n}_{\text{end}} - \text{n}_{\text{start}}} \times \frac{\text{f samples}}{\text{1 second}} \times \frac{\text{60 seconds}}{\text{1 minute}}$$

2.2.5 Invalid Epoch Detection

Consistent with the literature, periods of device non-wear and sleep were removed from the analyses. Ideally, both sleep and non-wear periods would be detected automatically, however, study timelines did not permit the completion of the final testing of the validated analytical tools. Non-wear periods were determined by visual inspection with the help of the device removal logs with the key criteria being absence of change accelerometer output for ≥ 5 minutes. Wrist and ankle accelerometer data were visualized simultaneously which facilitated non-wear period detection as the majority of periods involved the removal of both devices. The ECG data did not undergo visual non-wear detection as these periods were already accounted for using the signal quality algorithm. In conjunction with the sleep logs, visual inspection was used to determine when the participant went to bed each night. Sleep onset was defined as the marked decrease in movement for an extended duration around the time the participant said they went to bed. Because heart rate data were not always available and no other physiological measures were taken, the periods marked as sleep represent sedentary time in bed in addition to actual sleep.

Periods with invalid ECG data were also omitted from analyses. Therefore, only periods with valid ECG data, while all devices were worn, and when the participants were awake were included in the analyses. These periods will be referred to as “valid epochs”.

2.2.6 Cardiorespiratory Measures

2.2.6.1 Resting Heart Rate

An in-person resting heart rate was not taken. These measures have been shown to be influenced by the “white coat” effect and because continuous heart rate data was measured which can lead to a more accurate calculation of resting heart rate (Palatini, 2009). No standardized methods to determine resting heart rate from free-living data were found so published recommendations for calculating resting heart rate clinically (Palatini, 2009) were followed. These recommendations were integrated with an approach similar to that found in (Logan, Reilly, Grant, & Patton, 2000). To determine resting heart rate, average heart rate was calculated over a 60-second window (four consecutive 15-second epochs) using a rolling average calculation. If the one-minute window contained an invalid epoch (either from signal loss or from sleep), it was omitted. The 60-second windows were sorted in ascending order and the average of the first 30 windows was taken as the participant’s resting heart rate.

2.2.6.2 Maximum Heart Rate

Further reducing participant burden, maximum heart rate was estimated using an equation instead of being calculated directly using a graded exercise test. Maximum heart rate was calculated using the predictive equation from Tanaka and colleagues (2001).

$$\text{Eq. (4): } HRmax = 208 - 0.7 \times \text{age [years]}$$

2.2.6.3 Estimation of Resting VO_2

Resting VO_2 was estimated using data from Kwan and Kwok (2004) that factors in both sex and age (see Table 1). These values helped maintain consistency with the methods of Powell and colleagues by not assuming resting VO_2 is 3.5 mL kg^{-1} min^{-1}. Powell and colleagues measured the average resting VO_2 of their sample to be 3.27 mL kg^{-1} min^{-1}. Their value is closer to the values reported by Kwan & Kwok than the standard 3.5 mL kg^{-1} min^{-1}.

Table 1: Mean resting VO_2 by age and sex. Values in mL kg^{-1} min^{-1}. Data from Kwan & Kwok, 2004.

Age Group	Men	Women
16-64 years	3.03	3.32
65-89 years	2.84	2.82

2.3 Model-Generated Outcome Measures

For each model, two main outcome measures were generated. First was the epoch-by-epoch classification into sedentary, light, moderate, or vigorous activity. Secondly, total time spent in each intensity category was calculated by counting the number of epochs spent in that intensity and dividing by 4 to convert from the number of epochs to number of minutes (15 seconds per epoch = 4 epochs per minute). Activity totals were also reported as a percentage of the valid data (awake with all devices worn and valid ECG data) to allow between-subject comparisons to be made while accounting for the differences in quantity of valid data and data collection duration. Lastly, a fifth intensity category, moderate-to-vigorous activity, was created by summing moderate and vigorous activity. This was done to represent the intensity commonly used in activity guidelines.

2.4 Participant Inclusion Based on Usable Data Volume

Analyses were run to determine how much usable data was obtained from each participant. This was done using the data classified as valid as described in section 2.2.5.

To be included in analyses, a threshold of 30 hours of usable data was set. The purpose of selecting a threshold was to include a quantity of usable data similar to what is required from common protocols in the literature. This threshold was determined from a review of 57 studies (Skender et al., 2016) that used accelerometers and questionnaires to quantify physical activity. Of these 57 articles, a collection duration of seven days was the most common (65% of studies). 82% of studies defined a valid day as at least 10 hours of accelerometer wear time. Lastly, nearly half (47%) of the studies required at least 4 valid days. Requiring ≥ 10 hours per day for 4 days in a 7-day collection is a minimum of approximately 25% of the entire collection period. The current data collection length was limited by the Bittium Faros' battery life of approximately 5 days. The threshold for number of usable hours was set at 25% of 5 days which is 30 hours. Therefore, only participants who provided ≥ 30 hours of valid data were included in the analyses.

2.5 Model Descriptions

The following section describes each wearables model. Specific nomenclature is used from here on. To avoid confusion with body segments, all model names are capitalized (i.e. "wrist" is the body part and "Wrist" is the wearables model). Model comparisons are denoted using a "vs." between model names (e.g. "Wrist vs. Ankle" refers to a comparison between the Wrist and Ankle models). The samples used in the secondary analysis (see section 2.7) were named based on which models provided enough usable data; these samples are named the AnkleWrist and WristHR samples.

2.5.1 Wrist Accelerometer Model

The wrist accelerometer (Wrist) model implemented the cut-points of Powell and colleagues (2017). These cut-points consist of two sets: one for the dominant wrist and one for the non-dominant. Only data from the non-dominant wrist were included in this thesis. Powell's cut-points were created using GENEActiv accelerometers sampling at a frequency of 30 Hz. Since the present study's data was collected at 75 Hz and activity count calculations are affected by the total number of data points, the cut-points were multiplied by a factor of 2.5 ($75 \div 30 = 2.5$) to scale them for use with 75 Hz data. Unpublished pilot work by the author has shown that this method does not affect total measured activity minutes. Accelerometer data were not filtered since the data used in the cut-point development were not filtered.

To generate the activity profiles, each epoch's activity count value was compared to the Powell cut-points that mark the boundaries for light (1.5 METs), moderate (3.0 METs), and vigorous (6.0 METs) activity, and the epoch's intensity was classified based on the MET range within which it fell. These MET-range definitions were kept as consistent as possible in the subsequent models to allow for more valid between-model comparisons. Figure 13 shows the Wrist model processing flowchart.

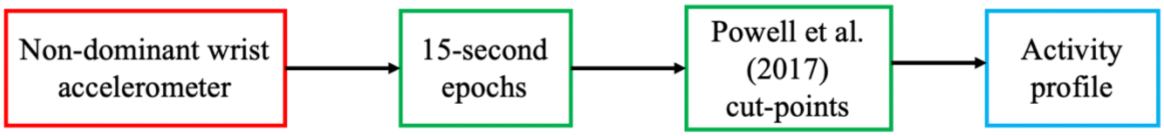

Figure 13: Data processing flowchart for the Wrist model by Powell et al (2017).

2.5.2 Ankle Accelerometer Model

For the ankle accelerometer (Ankle) model, each participant's treadmill protocol data and the ankle accelerometer data from the same side of the body as the non-dominant wrist were used. In this model, the average ankle activity count from each of the five treadmill walks were used in a linear regression equation to predict gait speed. Multivariate regression using demographic variables such as height were investigated but were not used due to the extremely high coefficients of determination attained with individually calibrated univariate regression equations.

Physiologically, ankle counts should be approximately zero when no movement is occurring. However, the regression was not forced through the origin which led to an improvement in the regression's performance. Instead, a threshold was set to differentiate a potentially meaningful bout of movement or gait from transient movement such as shifting position in a chair or small amplitude leg tapping. There is no standard definition of what constitutes a bout of gait and therefore the time duration of a bout is not defined either. Using a combination of definitions and data about gait bout durations (Orendurff, Schoen, Bernatz, Segal, & Klute, 2008; Awais, Chiari, Ihlen, Helbostad, & Palmerini, 2018), average preferred speed, cadence, and different gait bout definitions (Waters et al., 1988; Prajapati, Mansfield, Gage, Brooks, & McIlroy, 2011; Roos, Rudolph, & Reisman, 2012; Danks, Roos, McCoy, & Reisman, 2014), a temporal threshold of 5 seconds was selected to differentiate between transient and meaningful movement. The threshold was converted to an equivalent activity count value by dividing the average activity counts from the participant's walk at their preferred gait speed from the treadmill protocol by 3 (since 5 seconds is one-third of the 15-second epoch).

For each epoch, a gait speed was estimated using the activity count value with the participant's individual regression equation. For epochs that were below the threshold described above, a gait speed of zero was given. VO_2 was then calculated using the ACSM's walking and running equations (see Equations 5a and 5b). These equations are most accurate in the ranges of 50-100 m min^{-1} and greater than 135 m min^{-1}, respectively. The running equation was used for speeds above 100 m min^{-1}.

$$\text{Eq. (5a): } VO_2\ (\text{mL kg}^{-1}\ \text{min}^{-1}) = 3.5\ \text{mL kg}^{-1}\ \text{min}^{-1} + 0.1 \times \text{speed [m min}^{-1}] + 1.8 \times \text{speed [m min}^{-1}] \times \text{grade [decimal]}$$

$$\text{Eq. (5b): } VO_2\ (\text{mL kg}^{-1}\ \text{min}^{-1}) = 3.5\ \text{mL kg}^{-1}\ \text{min}^{-1} + 0.2 \times \text{speed [m min}^{-1}] + 0.9 \times \text{speed [m min}^{-1}] \times \text{grade [decimal]}$$

In these equations, an assumed resting VO_2 value of 3.5 mL kg^{-1} min^{-1} is used and they include a vertical component used when moving on an incline. The present thesis replaced the resting VO_2 value with that determined from Kwan and Kwok's data and the vertical component was removed by assuming level-ground walking (orientation calculations are not possible using epoched data). The modified equations used to predict gait speed is shown in Equations 5c and 5d.

$$\text{Eq. (5c): } VO_2\ (\text{mL kg}^{-1}\ \text{min}^{-1}) = \text{estimated resting } VO_2 + 0.1 \times \text{speed [m min}^{-1}]$$

$$\text{Eq. (5d): } VO_2\ (\text{mL kg}^{-1}\ \text{min}^{-1}) = \text{estimated resting } VO_2 + 0.2 \times \text{speed [m min}^{-1}]$$

Once VO_2 was estimated for each epoch, this value was divided by the participant's resting VO_2 value to obtain a MET level. Lastly, an intensity classification was assigned using the definitions of sedentary (< 1.50 METs), light (1.50 – 2.99 METs), moderate (3.00 – 5.99 METs), and vigorous (≥ 6.00 METs) based on the predicted MET value. Figure 14 shows the Ankle model processing flowchart.

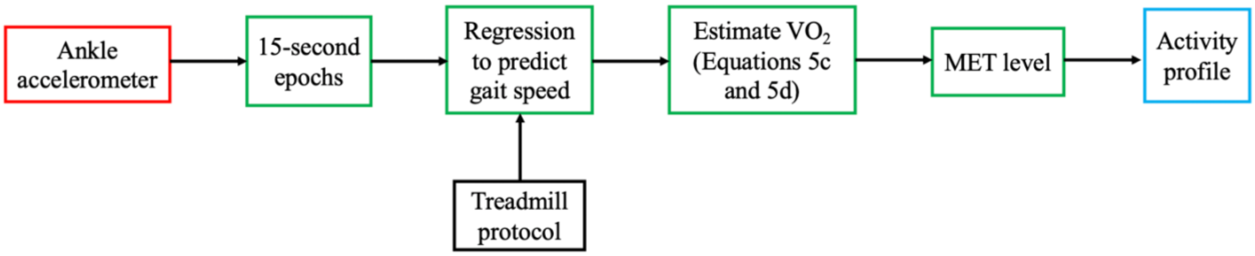

Figure 14: Data processing flowchart for the novel Ankle model.

2.5.3 Heart Rate Model

The heart rate (HR) model used percent heart rate reserve to quantify activity intensity. For each epoch, the average heart rate was calculated as a percent of heart rate reserve based on the individual's derived resting heart rate and predicted maximum heart rate using Equation 3. Since resting heart rate was calculated with a rolling average method, it was possible that an individual epoch's average heart rate was below resting heart rate, which would result in a negative percent heart rate reserve. In such instances, a value of zero percent heart rate reserved was assigned.

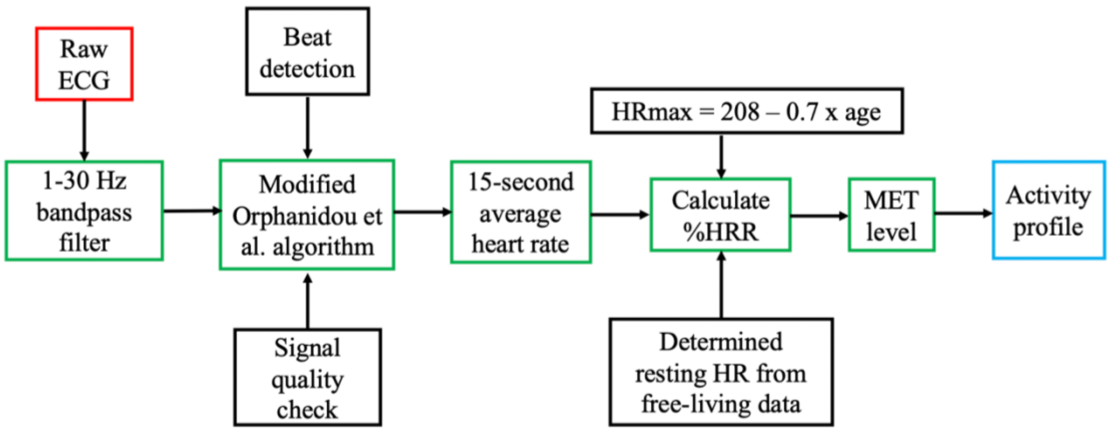

Figure 15: Data processing flowchart for the HR model.

Activity intensity was classified as sedentary (< 30% HRR), light (30-39.99% HRR), moderate (40-59.99% HRR), and vigorous (≥ 60% HRR). These intensity ranges correspond to <2.00, 2.00-2.99, 3.00-5.99, and ≥ 6.00 METs, respectively (ACSM, 2014). Figure 15 shows the HR model processing flowchart.

2.5.4 Combined Heart Rate-Ankle Accelerometer Model

The combined heart rate and ankle accelerometer (HRAcc) model used logic derived from the literature where the accelerometer data are used at low intensity and heart rate data are used at higher intensities (Brage et al., 2004; Johansson et al., 2006; Romero-Ugalde et al., 2017; Strath et al., 2002). The present model's threshold was implemented using a relative heart rate measure as opposed to an absolute threshold (i.e. flex heart rate in beats per minute). The objective when selecting a threshold was to determine a value where heart rate was highly correlated with lower/whole-body movement, thereby eliminating the use of heart rate data when heart rate may have been primarily under the influence of non-activity factors.

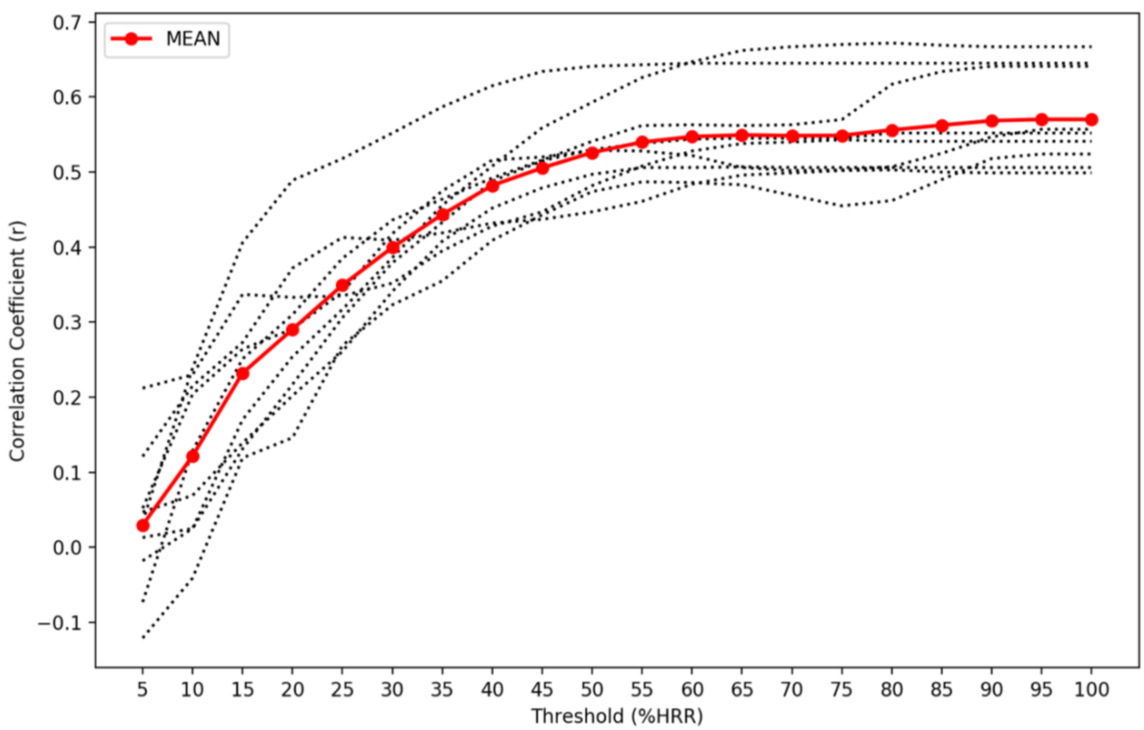

Figure 16: Correlation data between relative heart rate and ankle activity counts used in the generation of the HRAcc model threshold. Individual participant data are shown as dotted lines. Mean correlation values for each HRR increment are plotted in red.

To determine this threshold, each participant's data (n=10) were analyzed with a series of Pearson correlation coefficients. In increments of 5% heart rate reserve, data subsets were created that only included epochs where heart rate was above the threshold and when the participant was awake. This was done for both heart rate and ankle activity count data. The correlation coefficient between percent heart rate reserve and ankle activity counts was calculated using these data subsets. For example, with a threshold of 15% heart rate reserve, the correlation coefficient between percent heart rate reserve and ankle counts was calculated for all epochs where heart rate reserve was above 15%. This was done for each participant using thresholds between 5 and 100% heart rate reserve. For each threshold, the average correlation was calculated across subjects. The threshold of 30% heart rate reserve was selected using the graph shown in Figure 16. This threshold corresponded to an average value of $r = 0.40$. A threshold of 40% heart rate reserve was also considered due to its proximity to where the curve plateaus (an approximate correlation coefficient of $r = 0.55$). However, selecting a threshold of 40% heart rate reserve led to a considerable decrease in the amount of time for which the HR model would be used, which would reduce the usefulness of such a combined model. For the participants included in this analysis, individuals spent $11.1 \pm 5.3\%$ of valid epochs above 30% heart rate reserve while this number fell to $4.3 \pm 2.7\%$ for a threshold of 40% heart rate reserve. The threshold of 30% heart rate reserve led to results consistent with the combined HRAcc model proposed by Johansson and colleagues (2006) who reported that 11.9 ± 2.2 of waking hours were spent at an intensity that elicited the use of heart rate as opposed to accelerometry in

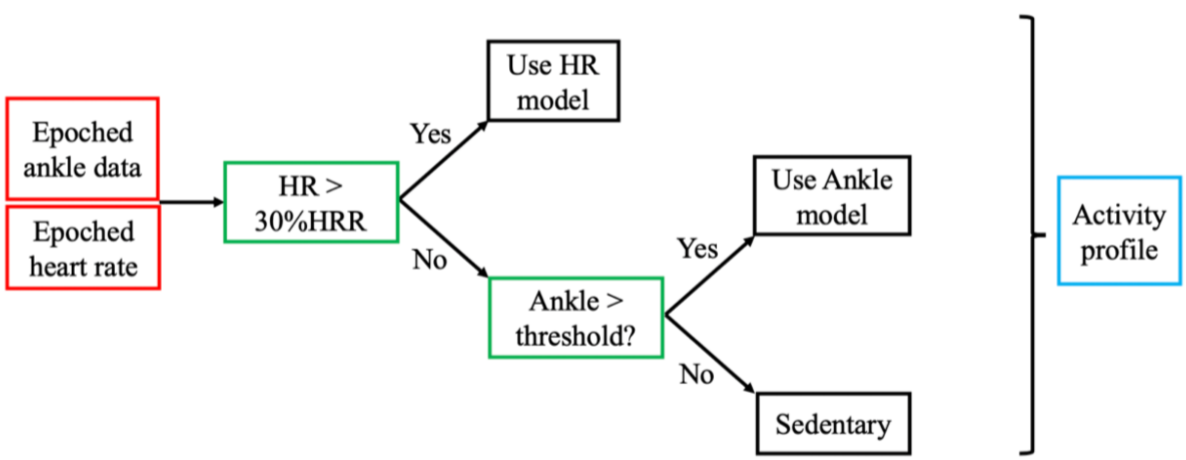

Figure 17: Data processing flowchart for the combined heart rate-accelerometer model.

his model. Although this is not the typical approach to determine a heart rate threshold, a flex heart rate based on intensity has been used previously (Romero-Ugalde et al., 2017).

For the HRAcc model, if an epoch's percent heart rate reserve was below 30%, the Ankle model was used. For those epochs above 30% heart rate reserve, the HR model was used. The activity profile was generated using the appropriate model. Figure 17 shows the HRAcc model processing flowchart.

2.6 Group Stratification for Objective 3

To address Objective 3, two groups needed to be created to separate participants into low- and high-activity groups. Based on the a priori power analysis, the plan was to stratify participants into two groups of seven by using the most and least active seven participants. However, only 10 participants had enough valid data; these 10 participants were split into two activity groups with 5 participants in each group.

Two stratification methods were investigated with the goal of creating groups based on whole-body activity in a way that did not bias results towards any of the four models. The two methods were grouping by average ankle counts and by average wrist counts. The averages were calculated without removing invalid ECG epochs. Sleep epochs were not included in the calculation. Independent sample t-tests were conducted to determine if these two stratification methods led to a statistically significant difference in average activity counts between groups and to determine if the groups differed in age, height, or weight. Importantly, both stratification methods created groups that had 4 of the 5 participants in common. Due to the similarities, the analysis was conducted using the groups created by average ankle counts as this measure likely better reflects whole-body activity level.

2.7 Additional Sample Generation and Secondary Analysis

A secondary analysis was conducted due to the Primary sample being smaller than the target calculated in the a priori power analysis as an attempt to replicate the findings using a larger sample size. This analysis was not conducted to address an a priori objective but was found to be necessary and is referred to as Objective 4.

Two additional samples (Secondary samples) were created by including all participants that provided valid data for both the Ankle and Wrist, and HR and Wrist data, respectively, without the requirement of having provided valid data for all four models. These samples are referred to as the AnkleWrist and WristHR samples, respectively, and each only contain data from the two models in their names. Participants added to the WristAnkle sample included participants who had too much invalid ECG data to be included in the Primary sample but had enough valid data from the Ankle and Wrist models. Participants added to the WristHR sample were collected in the first half of the study prior to the treadmill protocol implementation and therefore did not have data to generate the activity count-gait speed regression needed for the Ankle model. In this secondary analysis, 20 participants were included for the AnkleWrist sample and 18 for the WristHR sample. To include a larger volume of data than the Primary sample, data for the AnkleWrist sample were processed without removing epochs where the ECG data were not valid. The WristHR sample's data were not reprocessed. Both samples were stratified into activity groups using the same method as in Objective 3 resulting in two groups of 10 for the AnkleWrist sample and of 9 for the WristHR sample.

2.8 Statistical Analyses

2.8.1 Objectives 1 to 4

Statistical analyses were conducted in Python and R using the packages *pingouin* and *ezANOVA*, respectively. Statistical significance was set at $p < .05$. Violations for the assumption of sphericity were tested using the Mauchly test. The Greenhouse-Geisser epsilon correction was applied for violations of sphericity where applicable.

Data are reported as *mean ± standard deviation [95% confidence interval]* and error bars on graphs are the 95% confidence interval unless otherwise noted. Effect size statistics for F-tests were selected using the criteria suggested in the supplemental information from Läkens (2013). These will be explained in the relevant sections. For pairwise comparisons, Hedges' g was selected over Cohen's d to account for potential bias due to the small sample sizes (Läkens, 2013). Colour-coding of effect sizes in data tables represent the following effect size magnitudes: negligible (red), small (orange), medium (yellow), large (green) (Cohen, 1988).

No corrections were made for multiple comparisons for the post-hoc tests. This was done for several reasons. First, when comparing results from the Primary and Secondary samples (Objective 4), there are a different number of comparisons being made in each analysis so a one-to-one comparison could not be made with any corrected values. Second, due to the exploratory nature of this thesis and the ramifications of Type I and II errors, type I errors were viewed as being less detrimental than type II errors. And thirdly, the results of individual tests are more important than maintaining the family-wise significance level.

For Objectives 1 and 3A, the original plan was to analyze four activity intensities: sedentary, light, moderate, and vigorous. However, once the data were processed, moderate and vigorous activity were collapses into moderate-to-vigorous physical activity (MVPA) due to the very limited amount of vigorous activity. For Objectives 2 and 3B, model agreement was calculated using four intensities: sedentary, light, moderate, and vigorous. Objective 4 used the equivalent data from the Secondary samples.

For Objectives 2, 3B, and 4B (the Objectives that assess model agreement), Cohen's kappa was selected over a simple percent agreement since it accounts for agreement by chance (McHugh, 2012) which may have a large effect on data that is classified into only four categories. However, percent agreements are also reported as they are easier to interpret (McHugh, 2012).

2.8.1.1 Objective 1: Between-Model Comparison in Total Activity

To determine if the four models measured the same amount of total activity, a one-way repeated measures ANOVA was conducted for each of the three intensity categories. Separate ANOVAs were conducted because the measurements of intensity are not independent from one another as their sum is fixed by the collection duration; measurement of one intensity affects the others.

The factor Model had four levels: Wrist, Ankle, HR, and HRAcc. Effect sizes are reported using partial eta squared (η_p^2) because Model was manipulated (as opposed to observed) between all participants (Läkens, 2013). Post-hoc analysis was conducted to determine which model(s) measured different volumes of activity using pairwise dependent sample t-tests. Effect sizes are reported using Hedges' g.

2.8.1.2 Objective 2: Epoch-by-Epoch Agreement

To determine the level of agreement in epoch-by-epoch intensity classification between the four models, inter-model reliability was calculated for each of the six model pairs using Cohen's kappa. A one-way repeated measures ANOVA was conducted with the factor Model Comparison which had 6 levels: Wrist vs. Ankle, Wrist vs. HR, Wrist vs. HRAcc, Ankle vs. HR, Ankle vs. HRAcc, and HR vs. HRAcc. Effect size is reported using partial eta squared for the same reasons as Objective 1. Post-hoc analysis was conducted using pairwise dependent sample t-tests to determine if pairs of models had different levels of agreement.

2.8.1.3 Objective 3: The Effect of Activity Level on Model Performance

The same dataset from Objectives 1 and 2 was used for Objective 3 but participants were stratified into activity level groups as described in section 2.6.

2.8.1.4 Objective 3A: Time Spent in Each Activity Intensity

To determine if overall activity level had an effect on how the four different models measured total activity, a 4 x 2 mixed ANOVA was conducted for each of the three activity intensities. The between-subjects factor of Group had two levels: low and high activity. The within-subjects factor of Model had four levels: Wrist, Ankle, HR, and HRAcc. Separate ANOVAs were conducted for the same reason as in Objective 1. For this analysis, the focus was the Group x Model interaction so main effects are not reported. Effect sizes are reported using generalized eta squared (η_G^2) since not all factors are manipulated (Group was observed while Model was manipulated) (Läkens, 2013). Post-hoc analysis on the Group x Model interaction was performed using pairwise independent sample t-tests.

2.8.1.5 Objective 3B: Model Agreement in Activity Intensity Classification

To determine if overall activity level had an effect on the epoch-by-epoch agreement in activity intensity between models, a 6 x 2 mixed ANOVA was conducted. The between-subjects factor of Group had two levels: low and high activity. The within-subjects factor of Comparison had six levels (one for each between-model comparison): Wrist vs. Ankle, Wrist vs. HR, Wrist vs. HRAcc, Ankle vs. HR, Ankle vs. HRAcc, and HR vs. HRAcc. Effect sizes are reported using generalized eta squared since not all factors are manipulated (Group was observed, Model Comparison was manipulated) (Läkens, 2013). The main effect of Model Comparison was not of interest as it was addressed in Objective 2. Post-hoc analysis on the Group x Model interaction was performed using pairwise independent sample t-tests.

2.8.1.6 Objective 4: Secondary Analyses of Ankle vs. Wrist and Wrist vs. HR Models

A portion of the analyses from Objective 3 were repeated separately on the two Secondary samples generated by including more participants in an attempt to verify the results from the primary analysis in a larger sample. These Secondary samples included all participants in the primary analysis. Due to the change in inclusion criteria related to the amount of usable data, the secondary analysis was limited to the Ankle vs. Wrist and Wrist vs. HR comparisons.

2.8.1.7 Objective 4A: Time Spent in Each Activity Intensity

The same analysis from Objective 3A was used for the Secondary samples except the factor Model was reduced to two levels (either Ankle and Wrist or Wrist and HR), leading to a 2 x 2 mixed ANOVA instead of a 4 x 2 mixed ANOVA. Separate analyses were conducted on the WristAnkle and WristHR samples.

2.8.1.8 Objective 4B: Model Agreement in Activity Intensity Classification

Due to the reduction to two levels of Comparison (either Ankle vs. Wrist or Wrist vs. HR), an independent samples t-test was conducted for each sample to determine if there was a difference between the high- and low- activity groups in Cohen's kappa values.

2.8.2 Performance of the ECG Signal Quality Algorithm

Since an additional condition was added to the ECG signal quality algorithm in the present thesis to improve detection of non-wear periods, a sensitivity/specificity analysis was conducted. One thousand 15-second segments of ECG data were generated at random from random participants; this is the same volume of data that was used in the original algorithm validation as the authors used one thousand five hundred 10-second data segments. Data were plotted and I decided whether the segment of ECG data contained a reasonably clean signal where QRS peaks could be confidently located visually. The sensitivity/specificity analysis was conducted by comparing these judgements to the output of the modified Orphanidou algorithm using an Excel (version 16.39, Microsoft Corp., Seattle, Washington) spreadsheet. Algorithm performance was quantified using sensitivity, specificity, percent accuracy, and Cohen's kappa.

CHAPTER 3: RESULTS

3.1 ECG Signal Quality Algorithm Results

Compared to the performance described in the original paper (Orphanidou et al., 2015), the incorporation of the additional condition to detect periods of non-wear marginally improved performance. The test of one thousand 15-second sections of ECG data resulted in a sensitivity of 94.9% and specificity of 97.5%. Overall accuracy was 96.1% with almost perfect agreement (Cohen's kappa = .921) (Cohen, 1988). It should be noted that the data used in the present work were likely much more variable in signal noise than that used in the original study; ECG was collected over a much longer interval (5 days compared to ≤ 24 hours) and the original dataset likely did not have to deal with non-wear periods or residual moisture post-bathing. Individual participant data in the present work were found to be 53.1 ± 25.1 [45.4 to 60.7] % usable compared to 64% from the original paper. This decrease in usable data was expected due to the aforementioned signal quality issues.

Despite only the marginal statistical improvement in performance with the addition of the fifth condition compared to the original study, its addition was critical to the algorithm's performance in this more challenging dataset. Without it, non-wear periods were being frequently identified as being usable data. Table 2 shows the confusion matrix of results from the validation of the modified Orphanidou algorithm. The unaltered algorithm's performance was not formally assessed in this work.

Table 2: Confusion matrix showing the performance of the modified Orphanidou et al. (2015) ECG signal quality algorithm validation procedure.

		Researcher decision	
		Usable	**Unusable**
Algorithm output	**Usable**	52.4%	1.1%
	Unusable	2.8%	43.7%

3.2 Usable Data

For Objectives 1, 2, and 3, participants required at least 30 hours of valid data from all four models to be included. Of the 44 total participants, 5 were excluded due to missing device data (2 from a protocol change shortly after data collection started, 2 from device malfunction, and 1 participant selected not to wear an ankle GENEActiv). Further participants were excluded for having less than 30 valid hours of ECG data (n=11), not performing the treadmill protocol (n=11), having less than 30 hours of usable data once sleep and device non-wear were accounted for (n=6), and for a repeat collection (n=1). The remaining sample consisted of 10 participants. Table 3 contains the demographic characteristics for this sample. Considering each exclusion criterion independently, participants were excluded for missing data (n=5), having less than 30 hours of valid ECG data (n=11), not performing the treadmill protocol (n=23), and having less than 30 hours of valid data after accounting for sleep and device non-wear (n=21). As a percentage of the accelerometer collection duration, these 10 participants slept for 34.7 ± 5.0 % and removed one or both accelerometers for 3.8 ± 4.3% of the time. 60.8 ± 18.0 % of the total ECG data volume was usable. Once unusable ECG, sleep, and device non-wear data were combined, 38.7 ± 9.4 % (38.9 ± 8.6 hours) of the ECG collection period remained as usable.

Table 3: Demographic characteristics of the Primary sample.

	Value
Age (mean ± SD, years)	22.1 ± 4.4
Females (n, %)	6, 60%
Weight (mean ± SD, kg)	73.9 ± 14.5
Height (mean ± SD, cm)	172.9 ± 8.1
Right-handed (n, %)	8, 80%
Volume of usable data (mean ± SD, hours)	38.7 ± 8.6

3.3 Ankle Model Regression Equations

There was a strong linear relationship (r > .7) between activity counts and gait speed when all participants' data were pooled which can be seen in Figure 18.

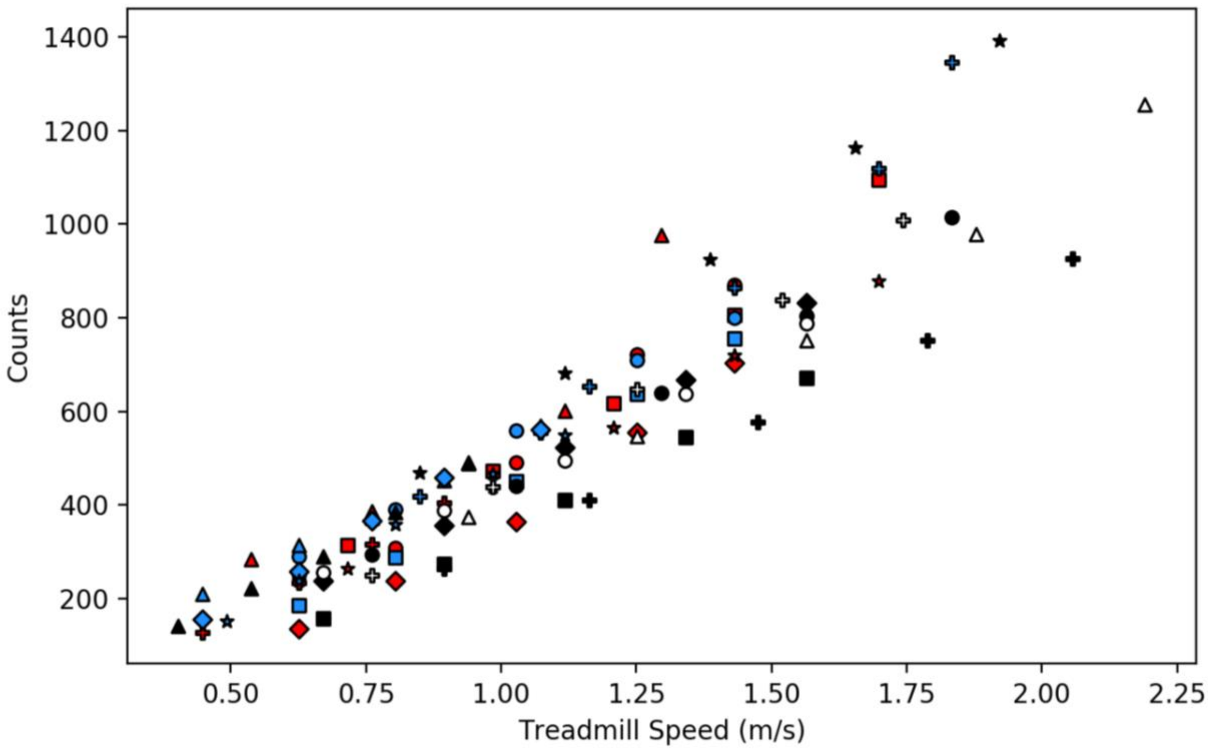

Figure 18: Pooled treadmill protocol ankle accelerometer data from all participants (n=21). Each participant is represented using a unique marker colour and shape combination.

Individual simple linear regression equations were created to further improve this relationship. The individual regression equations (n=21) generated models with such good fit that no additional predictors were needed. Individual coefficients of determination ranged from .879 to > .999 (.989 ± .026 [0.983 to .994]) and standard error of estimate values ranged from .007 to .134 (.036 ± .030 [.023 to .050]) m s^{-1}. Figure 19 shows individual regression lines over the range of counts observed counts during all treadmill protocols; this figure highlights the consistency in slopes between participants

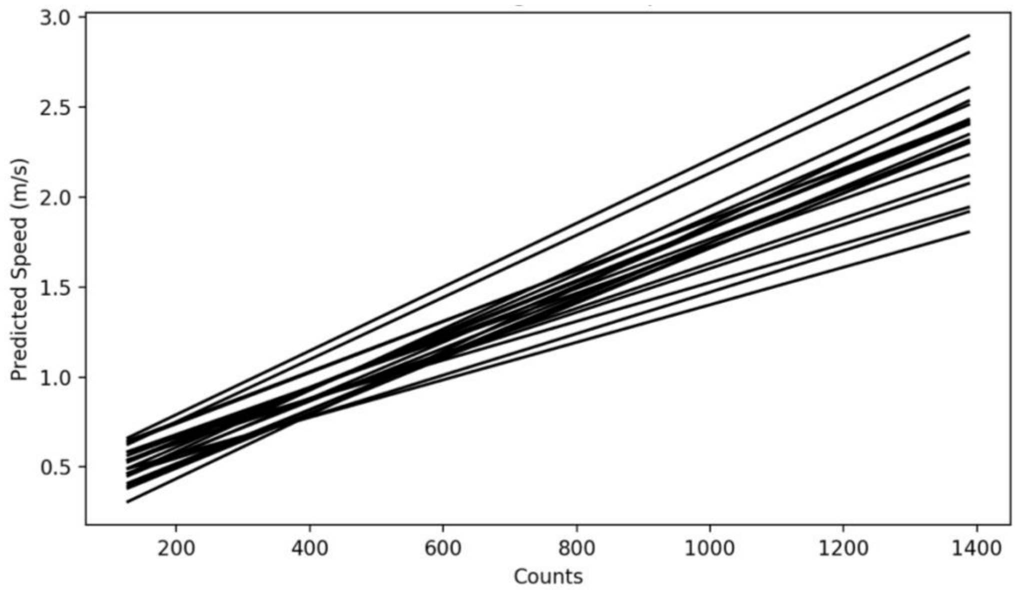

Figure 19: Individual regression lines from each participant's Ankle model regression equation.

Prediction accuracy remained consistent across gait speeds; there was a very weak relationship between measured and residual gait speeds (r = .077, p = .435). This consistency is seen in Figure 20 which shows a modified Bland-Altman graph (Bland & Altman, 1986).

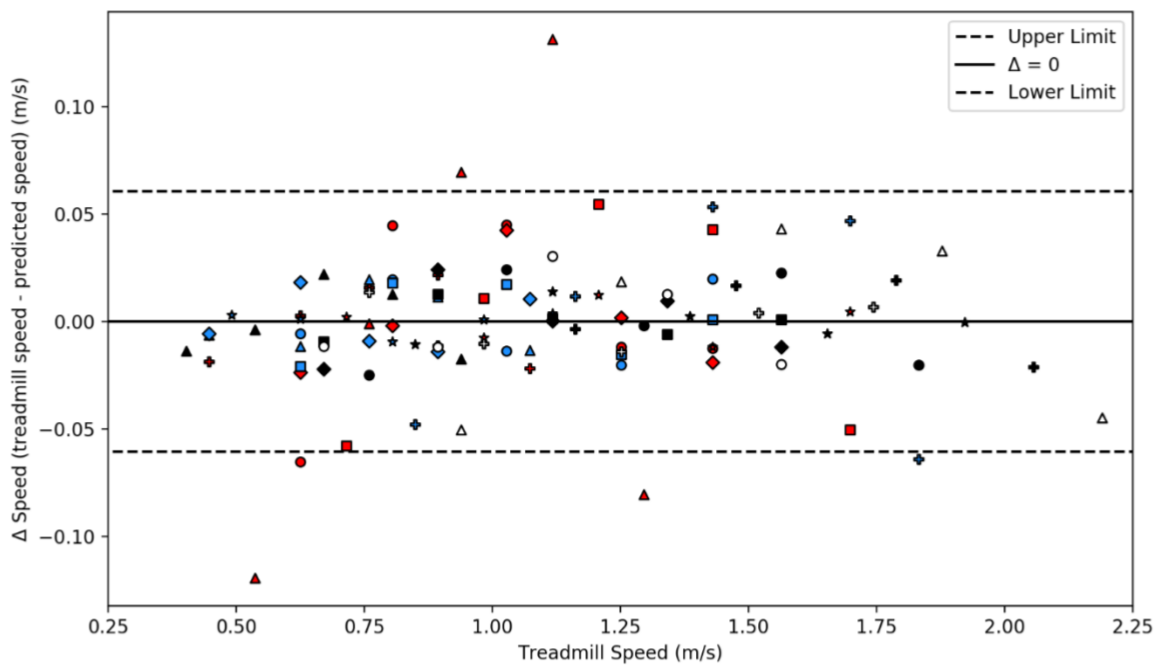

Figure 20: Modified Bland-Altman plot showing the differences between measured and predicted gait speed using each participant's individual regression equation. Limits of agreement are ± 1.96 standard deviations of the pooled residuals. Symbols are consistent with Figure 19.

3.4 Objective 1 Results: Average Activity Volume Across Models

Figure 21 shows the mean activity time measured by each model. Results from all activity intensities are summarized in Tables 4 and 5.

3.4.1 Sedentary Behaviour

There was a significant main effect of Model for sedentary time measured as a percent of valid epochs ($F_{(3, 27)} = 13.58$, $p[GG\varepsilon] = .002$, $GG\varepsilon = .428$, $\eta_p^2 = .601$). There were significant increases in measured sedentary time using the Ankle model (90.7 ± 4.3 [87.7 to 93.8] %) compared to Wrist model (87.5 ± 5.5 [83.6 to 91.5] %) ($t_{(9)} = 4.13$, $p = .003$, $g = .621$), compared to the HR model (84.4 ± 7.5 [79.1 to 89.8] %) ($t_{(9)} = 3.85$, $p = .004$, $g = .989$), and compared to the HRAcc model (81.1 ± 8.3 [75.2 to 87.1] %) ($t_{(9)} = 5.90$, $p < .001$, $g = 1.386$). There was a significant increase in measured sedentary time using the Wrist model compared to the HRAcc model ($t_{(9)} = 3.23$, $p = .010$, $g = 0.867$), and using the HR model compared to the HRAcc model ($t_{(9)} = 4.47$, $p = .002$, $g = .399$). There was no significant difference in measured sedentary time between the Wrist and HR models ($t_{(9)} = 1.44$, $p = .183$, $g = .451$).

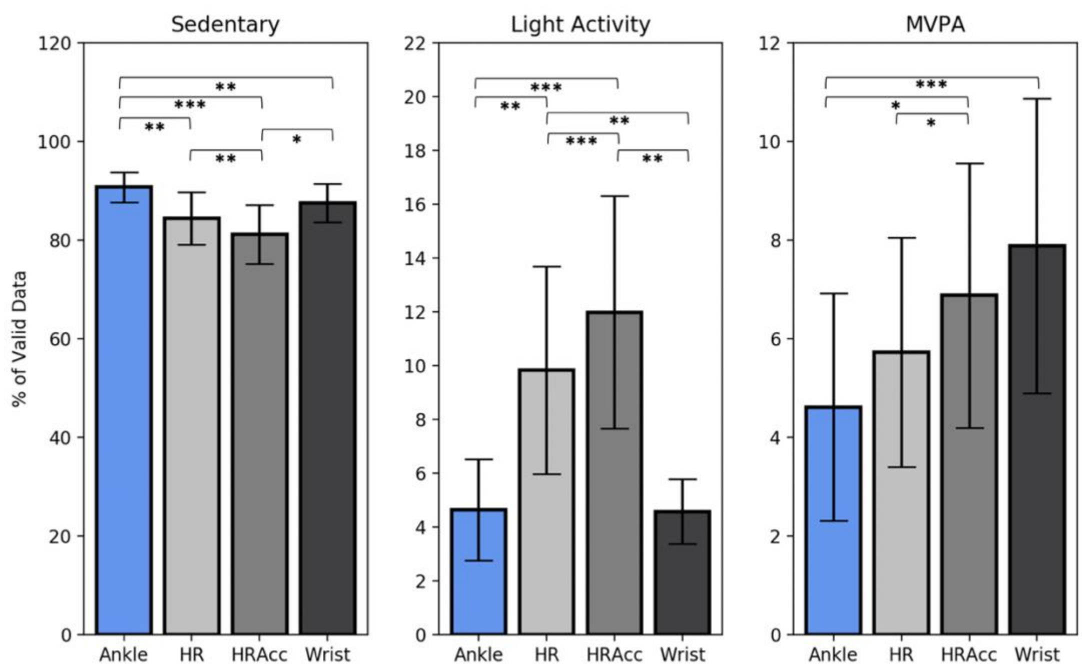

Figure 21: Results from Objective 1 showing Primary sample model means with 95% confidence intervals for each activity intensity (* p < .05, ** p < .01, *** p < .001).

3.4.2 Light Activity

There was a significant main effect of Model for light intensity activity measured as a percent of valid epochs ($F_{(3, 27)} = 19.49$, $p[GG\varepsilon] < .001$, $GG\varepsilon = .396$, $\eta_p^2 = .684$). There were significant increases in measured light activity using the HR model (9.8 ± 5.4 [6.0 to 13.7] %) compared to the Ankle model (4.6 ± 2.6 [2.8 to 6.5] %) ($t_{(9)} = 4.09$, $p = .003$, $g = 1.171$), using the HRAcc model (12.0 ± 6.0 [7.7 to 16.3] %) compared to the Ankle model ($t_{(9)} = 5.40$, $p < .001$, $g = 1.510$), using the HR model compared to the Wrist model (4.6 ± 1.7 [3.4 to 5.8] %) ($t_{(9)} = 3.57$, $p = .006$, $g = 1.257$), using the HRAcc model compared to the Wrist model ($t_{(9)} = 4.63$, $p < .001$, $g = 1.598$), and using the HRAcc model compared to the HR model ($t_{(9)} = 5.63$, $p < .001$, $g = .359$). There was no significant difference in light activity between the Ankle and Wrist models ($t_{(9)} = .11$, $p = .913$, $g = 0.026$).

3.4.3 Moderate-to-Vigorous Activity

There was a significant main effect of Model for moderate-to-vigorous intensity activity measured as a percent of valid epochs ($F_{(3, 27)} = 5.05$, $p[GG\varepsilon] = .029$, $GG\varepsilon = .512$, $\eta_p^2 = .359$). There were significant increases in MVPA using the Wrist model (7.9 ± 4.2 [4.9 to 10.9] %) compared to the Ankle model (4.6 ± 3.2 [2.3 to 6.9] %) ($t_{(9)} = 5.16$, $p < .001$, $g = .840$), using the HRAcc model (6.9 ± 3.8 [4.2 to 9.6] %) compared to the Ankle model ($t_{(9)} = 2.79$, $p = .021$, $g = .620$), and using the HRAcc model compared to the HR model (5.7 ± 3.2 [3.4 to 8.0] %) ($t_{(9)} = 2.45$, $p = .037$, $g = .315$). There were no significant differences in MVPA between the Ankle and HR models ($t_{(9)} = 1.12$, $p = .294$, $g = .330$), between the Wrist and HR models ($t_{(9)} = 1.75$, $p = .113$, $g = .553$), or between the Wrist and HRAcc models ($t_{(9)} = 1.02$, $p = .334$, $g = .243$).

Table 4: Summary of Objective 1's one-way repeated measures ANOVA.

	$P < .05$	ηp^2
Sedentary	*	.601
Light	*	.684
MVPA	*	.359

Table 5: Summary of results from Objective 1 post-hoc analysis (pairwise dependent samples t-tests).

	Sedentary		Light		MVPA	
Model Pair	**p < .05**	**Hedges' g**	**p < .05**	**Hedges' g**	**p < .05**	**Hedges' g**
Ankle vs. Wrist	*	.621		.026	*	- .840
Ankle vs. HR	*	.989	*	-1.171		- .330
Ankle vs. HRAcc	*	1.386	*	-1.510	*	- .620
Wrist vs. HR		.451	*	-1.257		.553
Wrist vs. HRAcc	*	.867	*	-1.598		.243
HR vs. HRAcc	*	.399	*	- .359	*	.315

3.5 Objective 2 Results: Epoch-by-Epoch Agreement Between Models

There was a significant main effect of Model Comparison on level of agreement ($F_{(5, 45)}$ = 83.65, p[GGε] < .001, GGε = .482, η_p^2 = .903). Group mean Cohen's kappa values are shown in Figure 22 and results are summarized in Table 6.

The Ankle vs. Wrist comparison (κ = .502 ± .116 [.415 to .589]; percent agreement = 90.5 ± 3.6 [87.9 to 93.1] %) had greater agreement than the Wrist vs. HR comparison (κ = .305 ± .082 [.246 to .364]; percent agreement = 82.6 ± 6.8 [77.8 to 87.5] %) ($t_{(9)}$ = 6.42, p < .001, g = 1.873), greater agreement than the Wrist vs. HRAcc comparison (κ = .380 ± .083 [.318 to .442]; percent agreement = 83.0 ± 6.6 [78.2 to 87.7] %) ($t_{(9)}$ = 4.93, p < .001, g = 1.161), greater agreement than the Ankle vs. HR comparison (κ = .296 ± .090 [.228 to .364]; percent agreement = 84.2 ± 6.5 [79.3 to 89.0] %) ($t_{(9)}$ = 6.19, p < .001, g = 1.899), and lesser agreement than the HR vs. HRAcc comparison (κ = .884 ± .072 [.830 to .939]; percent agreement = 96.7 ± 2.3 [95.1 to 98.4] %) ($t_{(9)}$ = 9.32, p < .001, g = 3.789). There was no difference between the Ankle vs. Wrist agreement and the Ankle vs. HRAcc agreement (κ = .505 ± .105 [.425 to .584]; percent agreement = 87.4 ± 5.8 [83.2 to 91.6] %) ($t_{(9)}$ = .05, p = .958, g = .022).

The Wrist vs. HR agreement was lesser than the Wrist vs. HRAcc agreement ($t_{(9)} = 3.40$, $p = .008$, $g = .867$), lesser than the Ankle vs. HRAcc agreement ($t_{(9)} = 4.85$, $p < .001$, $g = 2.019$), and lesser than the HR vs. HRAcc agreement ($t_{(9)} = 24.44$, $p < .001$, $g = 7.153$). There was no difference between the Wrist vs. HR agreement and the Ankle vs. HR agreement ($t_{(9)} = .80$, $p = .446$, $g = .105$).

The Wrist vs. HRAcc agreement was greater than the Ankle vs. HR agreement ($t_{(9)} = 3.49$, $p = .007$, $g = .931$), lesser than the Ankle vs. HRAcc agreement ($t_{(9)} = 4.67$, $p = .001$, $g = 1.261$), and lesser than the HR vs. HRAcc agreement ($t_{(9)} = 13.09$, $p < .001$, $g = 6.223$).

The Ankle vs. HR agreement was lesser than the Ankle vs. HRAcc agreement ($t_{(9)} = 5.54$, $p < .001$, $g = 2.038$), and lesser than the HR vs. HRAcc agreement ($t_{(9)} = 20.73$, $p < .001$, $g = 6.884$).

The Ankle vs. HRAcc agreement was lesser than the HR vs. HRAcc agreement ($t_{(9)} = 7.20$, $p < .001$, $g = 4.029$).

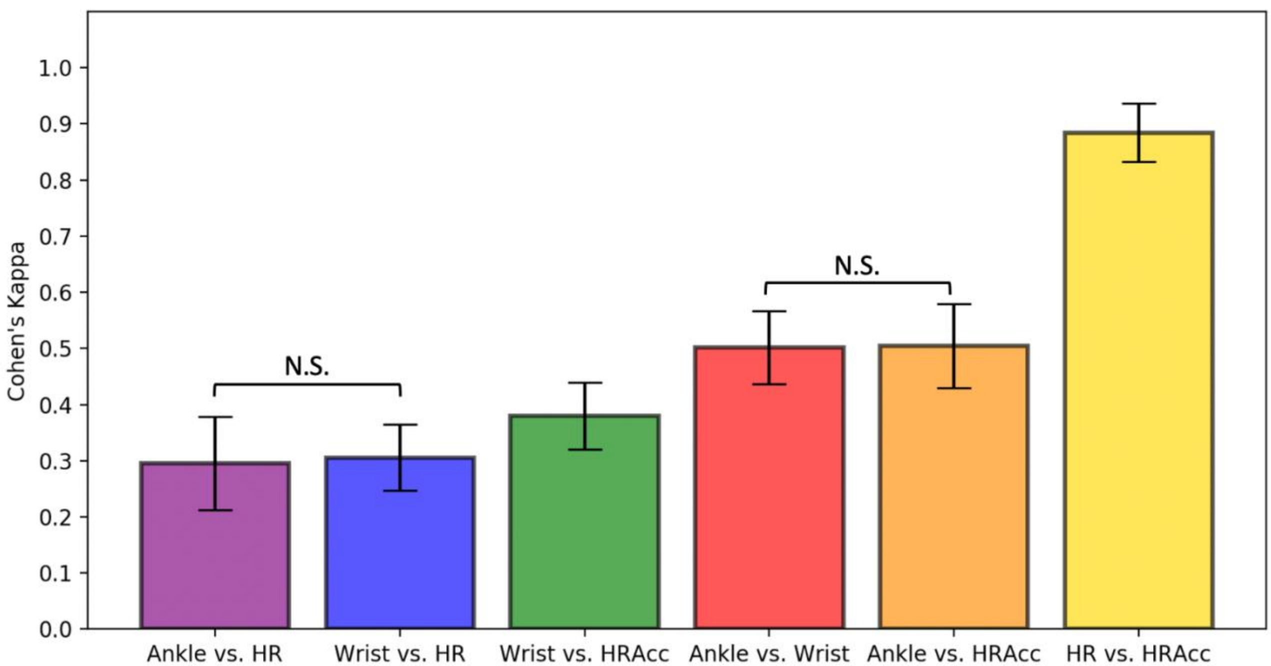

Figure 22: Primary sample mean Cohen's kappa values by Model Comparison with 95% confidence intervals. All comparisons were significantly different from each other ($p < .05$) unless marked with N.S. (not significant; $p \geq .05$).

Table 6: Summary of results from Objective 2 post-hoc analysis (pairwise dependent t-tests) to determine which model pairs demonstrated the same level of agreement.

Pair A	Pair B	p < .05	Hedges' g
Ankle vs. Wrist	**Wrist vs. HR**	*	1.873
Ankle vs. Wrist	**Wrist vs. HRAcc**	*	1.161
Ankle vs. Wrist	**Ankle vs. HR**	*	1.899
Ankle vs. Wrist	**Ankle vs. HRAcc**		- .022
Ankle vs. Wrist	**HR vs. HRAcc**	*	-3.789
Wrist vs. HR	**Wrist vs. HRAcc**	*	- .867
Wrist vs. HR	**Ankle vs. HR**		.105
Wrist vs. HR	**Ankle vs. HRAcc**	*	-2.019
Wrist vs. HR	**HR vs. HRAcc**	*	-7.153
Wrist vs. HRAcc	**Ankle vs. HR**	*	.931
Wrist vs. HRAcc	**Ankle vs. HRAcc**	*	-1.261
Wrist vs. HRAcc	**HR vs. HRAcc**	*	-6.223
Ankle vs. HR	**Ankle vs. HRAcc**	*	-2.038
Ankle vs. HR	**HR vs. HRAcc**	*	-6.884
Ankle vs. HRAcc	**HR vs. HRAcc**	*	-4.029

3.6 Objective 3 Results: The Effect of Activity Level

3.6.1 Activity Groups Comparison

There were significantly more ankle activity counts in the high- (85.6 ± 16.3 [65.4 to 105.8]) compared to low-activity (52.8 ± 12.9 [36.8 to 68.8]) groups ($t_{(8)}$ = 3.54, p = .008, g = 2.019). There were no significant differences between high- and low-activity groups in age (23.8 ± 5.7 [16.7 to 30.9] and 20.4 ± 0.5 [19.8 to 21.0] years, respectively) ($t_{(8)}$ = 1.26, p = .244), weight (79.9 ± 19.4 [55.8 to 104.0] and 67.9 ± 2.2 [65.2 to 70.6] kg, respectively) ($t_{(8)}$ = 1.37, p = . 207), and height (174.4 ± 5.0 [168.2 to 180.6] and 171.3 ± 10.7 [158.0 to 184.6] cm, respectively) ($t_{(8)}$ = .57, p = .584). Notably, stratifying participants by ankle counts also led to a significant difference in wrist counts between the high- (71.2 ± 6.1 [63.8 to 78.7]) and low-activity (49.4 ± 9.53 [37.6 to 61.2]) groups ($t_{(8)} = 4.32$, $p = .003$)

3.6.2 Objective 3A Results: Activity Level and Measured Activity Volume

Since the main effect of Model in this analysis is the same as the analysis conducted for Objective 1, it will not be reported again in this section. Results are summarized in Table 7.

The Model by Activity Group interaction was not significant for sedentary time ($F_{(3, 24)} = 1.63$, $p[GG\varepsilon] = .201$, $GG\varepsilon = .443$, $\eta_G^2 = .038$), light activity ($F_{(3, 24)} = .67$, $p[GG\varepsilon] = .459$, $GG\varepsilon = .397$, $\eta_G^2 = .023$), or MVPA ($F_{(3, 24)} = 1.32$, $p = .290$, $\eta_G^2 = .046$). Figure 23 shows the means for each Model by Activity Group group.

The main effect of Activity Group was not significant for sedentary ($F_{(1, 8)} = .40$, $p = .545$, $\eta_G^2 = .039$), light activity ($F_{(1, 8)} = .30$, $p =.600$, $\eta_G^2 =.027$), or MVPA ($F_{(1, 8)} = 4.98$, $p = .056$, $\eta_G^2 = .306$).

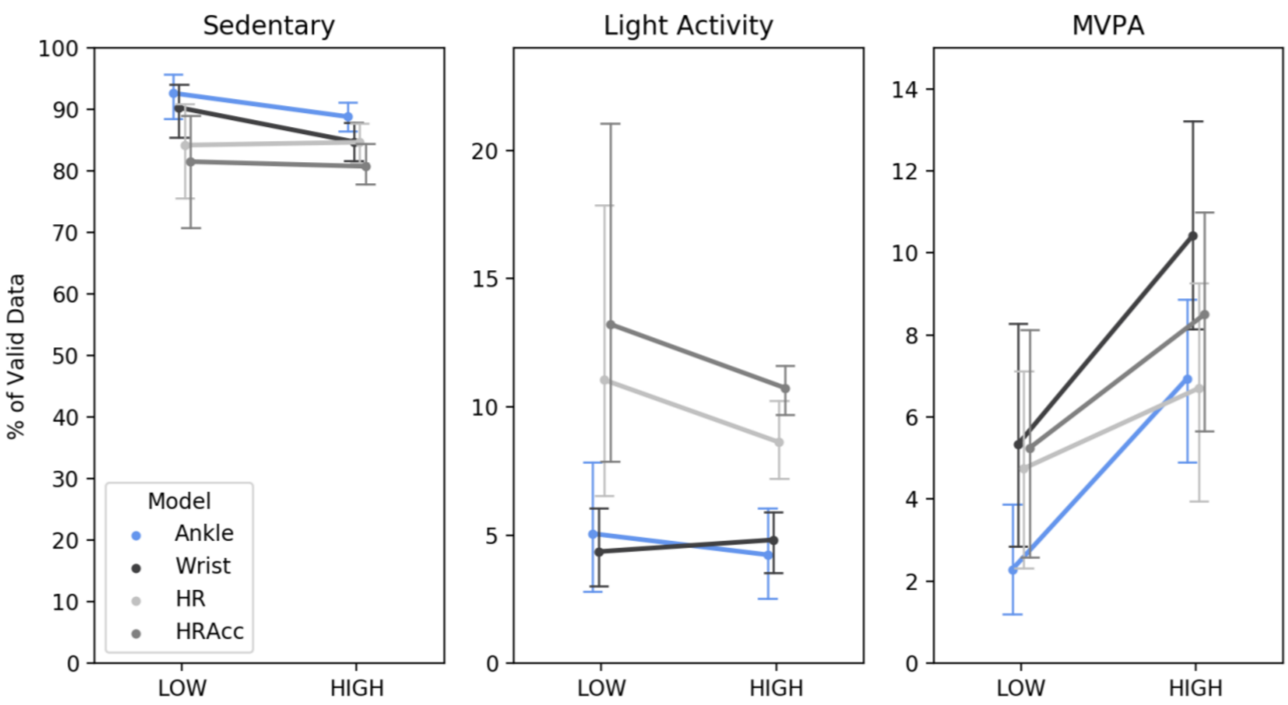

Figure 23: Primary sample activity volume means by Model and Activity Group. There were no significant Model by Activity Group interactions (all p > .05).

Table 7: Summary of Objective 3A's two-way mixed ANOVA.

Effect	Sedentary		Light		MVPA	
	P < .05	ηp²	P < .05	ηp²	P < .05	ηp²
Activity Group		.039		.027		.306
Model	*	.259	*	.394	*	.160
Interaction		.038		.023		.046

3.6.3 Objective 3B Results: Activity Level and Epoch-by-Epoch Agreement

The main effect of Model Comparison was not reported in this section since it is the same analysis that was conducted in Objective 2.

The Model Comparison by Activity Group interaction was not significant ($F_{(5, 40)} = 2.60$, p[GGε] = .105, GGε = .401, $\eta_G^2 = .175$) (see Figure 24).

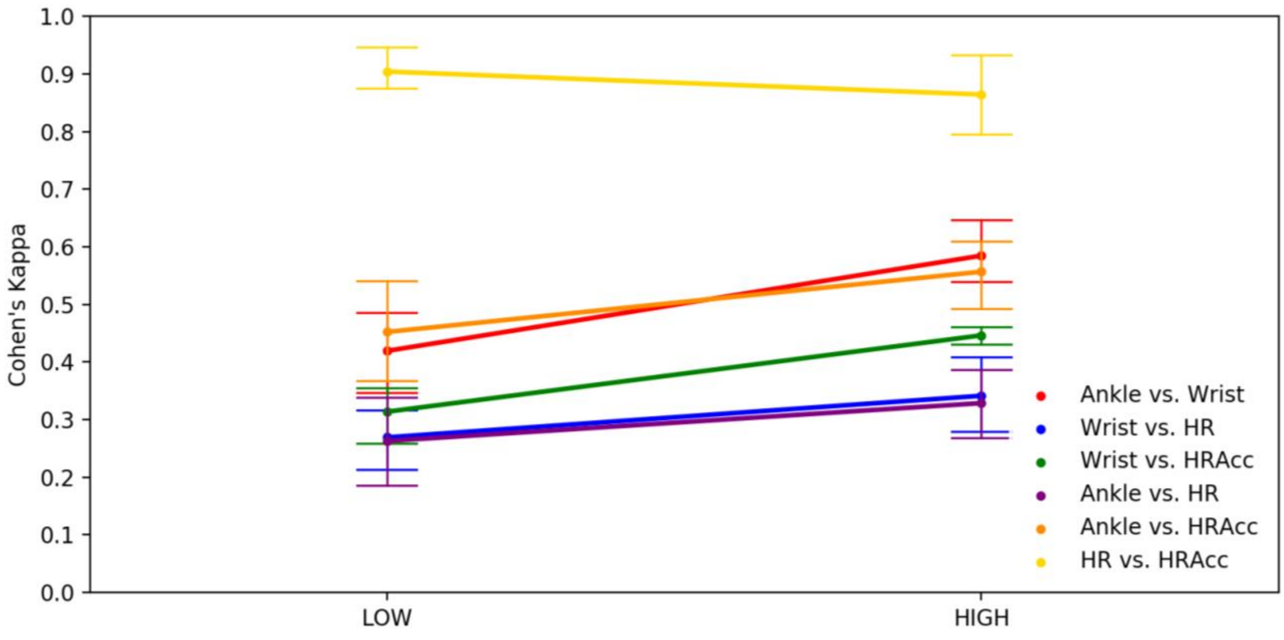

Figure 24: Primary sample mean Cohen's kappa values by Model Comparison and Activity Group with 95% confidence intervals (n = 5 per group).

However, there was a significant main effect of Activity Group on epoch-by-epoch agreement ($F_{(1, 8)} = 8.18$, p = .021, $\eta_G^2 = .261$) (Figure 25 shows the Activity Group means). The high-activity group (κ = 0.520 ± .197 [.449 to .592]; percent agreement = 87.3 ± 5.7 [85.2 to 89.4] %) had significantly higher Cohen's kappa values than the low-activity (κ = 0.437 ± .237

[.351 to .523]; percent agreement = 87.5 ± 8.8 [84.3 to 90.7] %) group ($t_{(8)}$ = 2.86, p = .021, g = 1.634).

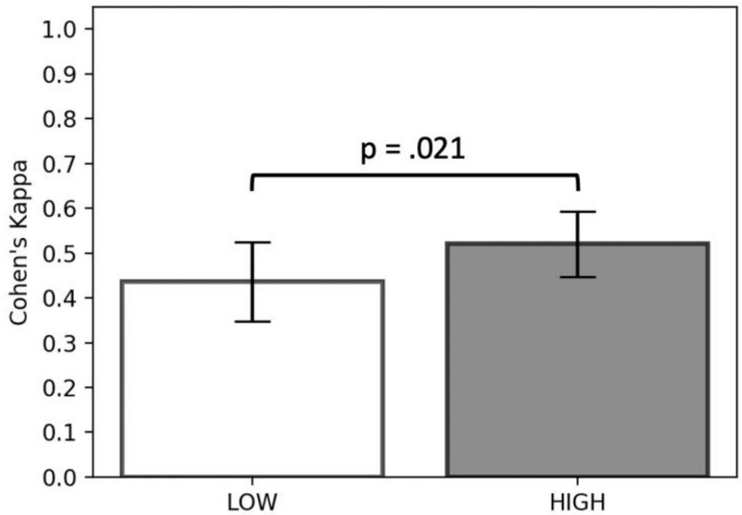

Figure 25: Primary sample mean Cohen's kappa values by Activity Group with 95% confidence intervals (n = 5 per group).

3.7 Objective 4 Results

Table 8 shows the demographic information of the AnkleWrist and WristHR samples used to confirm the primary findings.

Table 8: Demographic characteristics for the AnkleWrist and WristHR samples.

	AnkleWrist sample	WristHR sample
n	20	18
Age (mean ± SD, years)	21.8 ± 3.3	37.8 ± 21.1
Females (n, %)	13, 65%	12, 67%
Weight (mean ± SD, kg)	69.3 ± 13.6	73.0 ± 13.2
Height (mean ± SD, cm)	170.8 ± 7.9	171.3 ± 7.5
Right-handed (n, %)	16, 80%	16, 89%

Similar to Objective 3, the Model by Activity Group interaction was the focus so main effects are not reported. Activity volume data are shown in Figures 26 and 28 for the AnkleWrist and WristHR samples, respectively. Epoch-by-epoch agreement for both samples (Objective 4B) is shown in Figure 27. Test results are summarized in Table 9.

3.7.1 Objective 4 Results: AnkleWrist Sample

3.7.1.1 Activity Groups Comparison

As noted, 10 participants from the AnkleWrist sample were assigned to each of the high- and low- activity groups. The high-activity group had significantly more ankle counts (93.7 ± 16.6 [81.8 to 105.6]) than the low-activity (58.4 ± 11.1 [50.5 to 66.3]) group ($t_{(18)}$ = 5.60, p = .001, g = 2.399) as would be expected due to the stratification method. With respect to demographic characteristics, there were no significant differences between high- and low-activity groups in age (22.9 ± 4.2 [19.9 to 25.9] and 20.6 ± 1.5 [19.5 to 21.7] years, respectively) ($t_{(18)}$ = 1.62, p = .123), weight (73.4 ± 15.0 [62.7 to 84.1] and 65.3 ± 11.1 [57.2 to 73.4] kg, respectively) ($t_{(18)}$ = 1.36, p = .191), and height (169.9 ± 5.2 [166.2 to 173.6] and 171.7 ± 10.2 [164.4 to 179.0] cm, respectively) ($t_{(18)}$ = .52, p = .613). The high-activity group had 9 (90%) females and the low-activity group had 5 (50%).

3.7.1.2 Objective 4A: AnkleWrist Sample and Activity Volume

The Activity Group by Model interaction was not significant for sedentary time ($F_{(1, 18)}$ = .31, p = .583, η_G^2 = .006), light activity ($F_{(1, 18)}$ = .33, p = .573, η_G^2 = .005), or MVPA ($F_{(1, 18)}$ = 1.00, p = .329, η_G^2 = .027). Figure 26 shows the group means for each activity intensity.

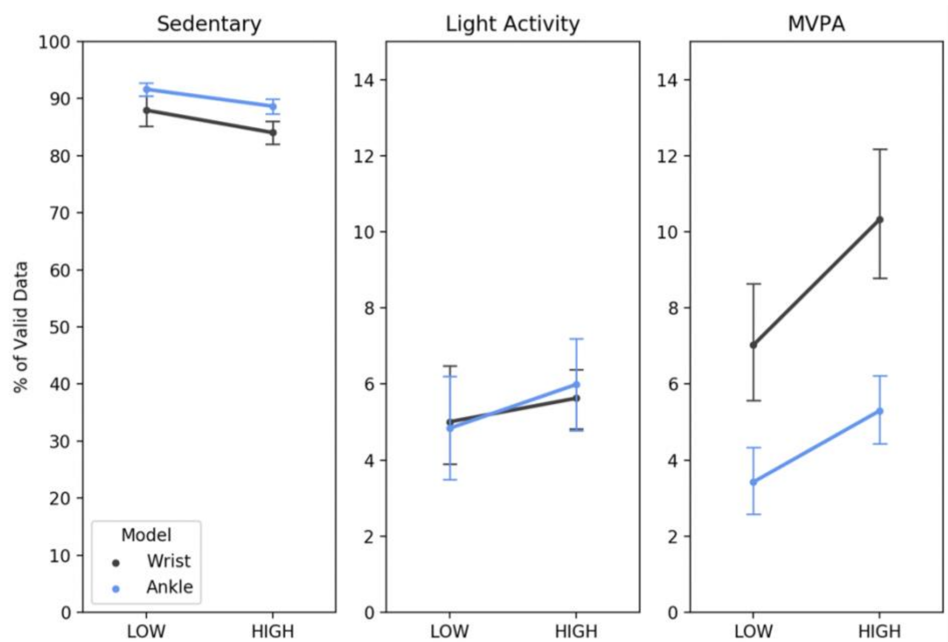

Figure 26: AnkleWrist sample activity volume by Model and Activity Group with 95% confidence intervals (n = 10 per group).

3.7.1.3 Objective 4B: AnkleWrist Sample and Epoch-by-Epoch Agreement

There was not a significant difference in the Ankle vs. Wrist agreement between high- (κ =.487 ± .059 [.445 to .529]; percent agreement = 88.4 ± 2.4 [86.7 to 90.1] %) and low- (κ = .433 ± .101 [.361 to .505]; percent agreement = 89.5 ± 3.8 [86.8 to 92.3] %) activity groups ($t_{(18)}$ =

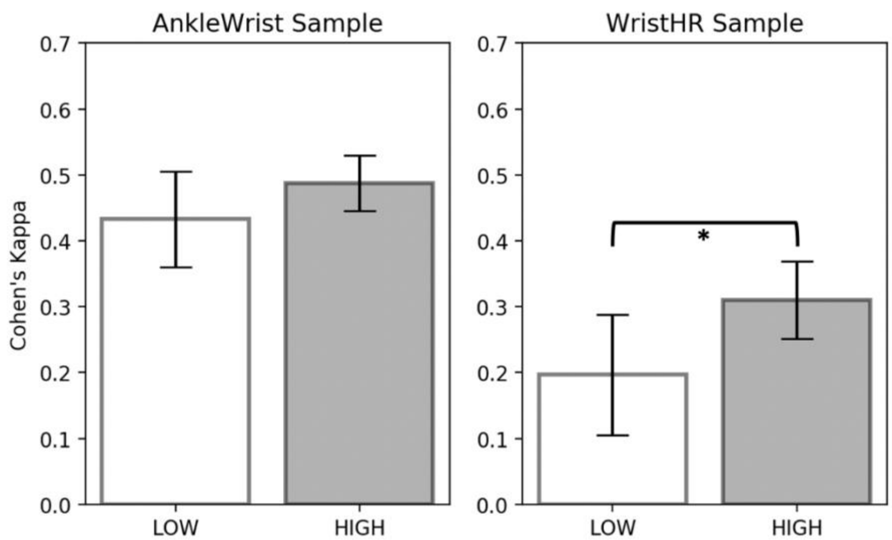

Figure 27: Secondary samples' mean Cohen's kappa values by Activity Group with 95% confidence intervals (* p < .05).

1.46, $p = .161$, $g = .626$). Figure 27 shows group means for both the AnkleWrist and WristHR samples.

3.7.2 Objective 4 Results: WristHR Sample

3.7.2.1 Activity Groups Comparison

Nine participants from the WristHR sample were assigned to each of the high- and low-activity groups. As expected, there was a significant increase in ankle activity counts for the high-activity group (81.8 ± 12.8 [71.1 to 92.5]) compared to the low-activity group (50.1 ± 11.5 [40.5 to 59.7]) ($t_{(16)} = 5.51$, $p < .001$, $g = 2.475$). With respect to demographic characteristics, there were no significant differences between high- and low-activity groups in age (35.7 ± 18.3 [20.4 to 51.0] and 40.0 ± 24.5 [19.5 to 60.5] years, respectively) ($t_{(16)} = .43$, $p = .676$), weight (72.1 ± 17.0 [57.9 to 86.3] and 73.9 ± 9.0 [66.4 to 81.4] kg, respectively) ($t_{(16)} = .28$, $p = .787$), and height (170.4 ± 6.8 [164.7 to 176.1] and 172.2 ± 8.5 [165.1 to 179.3] cm, respectively) ($t_{(16)} = .49$, $p = .632$). The high-activity group had 7 (77.8%) females and the low-activity group had 5 (55.6%).

3.7.2.2 Objective 4A: WristHR Sample

Similar to Objective 3, the Activity Group by Model interaction was the focus so main effects of Activity Group and Model are not reported.

The Activity Group by Model interaction was not significant for sedentary time ($F_{(1, 16)} = 1.94$, $p = .183$, $\eta_G^2 = .042$), light activity ($F_{(1, 16)} = 1.18$, $p = .294$, $\eta_G^2 = .024$), or MVPA ($F_{(1, 16)} = .56$, $p = .464$, $\eta_G^2 = .015$). Figure 28 the group means for each activity intensity.

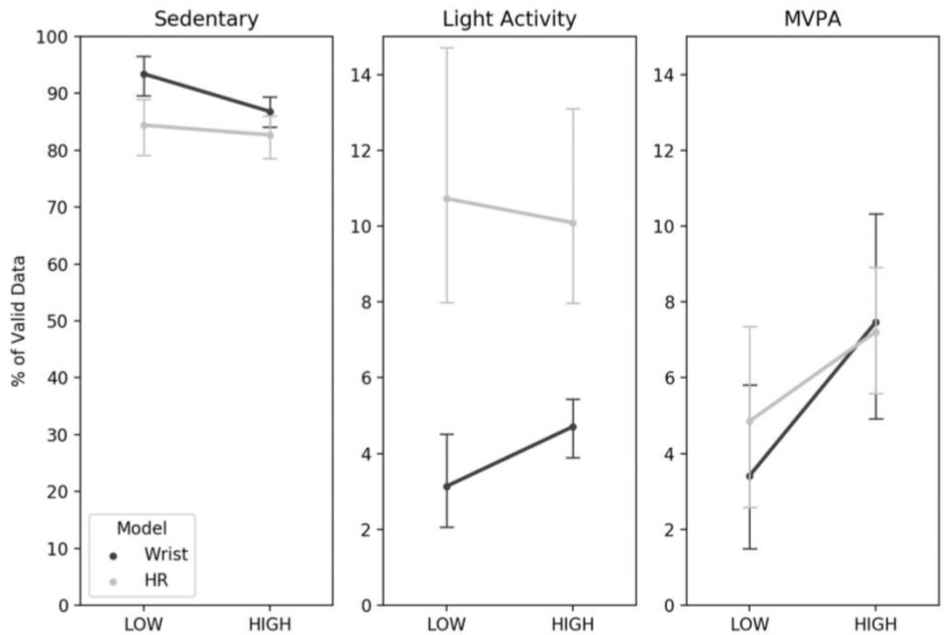

Figure 28: WristHR sample activity volume means by Model and Activity Group with 95% confidence intervals (n = 9 per group).

3.7.2.3 Objective 4B: WristHR Sample

There was a significant difference in the Wrist vs. HR agreement between high- (κ = .310 ± .076 [.252 to .369]; percent agreement = 81.3 ± 5.0 [77.5 to 85.0] %) and low- (κ = .197 ± .120 [.105 to .289]; percent agreement = 84.2 ± 5.5 [78.7 to 89.7] %) activity groups ($t_{(16)} = 2.40$, $p = .029$, $g = 1.078$). Refer back to Figure 27 for the group means for both the AnkleWrist and WristHR samples.

Table 9: Summary of activity volume results for the AnkleWrist and WristHR samples for Objective 4's mixed ANOVA.

		Sedentary		Light		MVPA	
Sample	**Effect**	**P < .05**	**η_G^2**	**P < .05**	**η_G^2**	**P < .05**	**η_G^2**
AnkleWrist	Activity Group	*	.232		.052	*	.267
	Model	*	.305		.001	*	.502
	Interaction		.006		.005		.027
WristHR	Activity Group		.112		.004	*	.180
	Model	*	.239	*	.456		.007
	Interaction		.042		.024		.015

3.7.3 Comparison Between the Primary and Secondary Samples

3.7.3.1 Activity Levels

The AnkleWrist sample resulted in the same conclusions as the Primary sample with repsect to activity volume for all three activity intensities. The WristHR sample agreed with the Primary sample in measured activity for light and MVPA but found a significant difference in sedentary time. Table 10 summarizes these results.

For the Activity Group by Model interactions for the Ankle vs. Wrist and Wrist vs. HR comparisons, none of the samples found a significant effect for any of the three activity intensities. Table 11 summarizes these results.

Table 10: Comparison of activity volume results between Primary, AnkleWrist, and WristHR samples.

		Primary		AnkleWrist		WristHR		Same Result
Intensity	**Contrast**	**p < .05**	**Hedges' g**	**p < .05**	**Hedges' g**	**p < .05**	**Hedges' g**	
Sedentary	Ankle vs. Wrist	*	.621	*	1.107			Yes
	Wrist vs. HR		.451			*	.984	**No**
Light	Ankle vs. Wrist		.026		.047			Yes
	Wrist vs. HR	*	-1.257			*	-1.716	Yes
MVPA	Ankle vs. Wrist	*	- .840	*	-1.626			Yes
	Wrist vs. HR		.553				.148	Yes

Table 11: Comparison of the Activity Group by Model interaction on activity volume between the Primary, AnkleWrist, and WristHR samples.

		Primary		AnkleWrist		WristHR		Same Result
Intensity	**Effect**	**p < .05**	**η_G^2**	**p < .05**	**η_G^2**	**p < .05**	**η_G^2**	
Sedentary	Group x Model		.038		.006		.042	Yes
Light			.023		.005		.024	Yes
MVPA			.046		.027		.015	Yes

3.7.3.2 Epoch-by-Epoch Model Agreement

Cohen's kappa values were similar between the Primary and Secondary samples for both the Ankle vs. Wrist (κ = .502 and .460, respectively) and Wrist vs. HR (κ = .305 and .253, respectively) comparisons.

The Primary and AnkleWrist samples did not find the same result for the effect of Activity Group on Ankle vs. Wrist agreement; a significant difference was found in the Primary sample but not in the AnkleWrist sample. There was also disagreement between the Primary and WristHR samples on whether Activity Group affected the Wrist vs. HR agreement; no significant difference was found in the Primary sample, but the WristHR sample found that the high-activity group had a significantly greater Wrist vs. HR agreement than the low-activity group. Table 12 summarizes these results.

Table 12: Comparison of epoch-by-epoch agreement between the high- and low- activity groups in the Primary, AnkleWrist, and WristHR samples.

		Primary		**AnkleWrist**		**WristHR**		
Effect	**Contrast**	**P < .05**	**Hedges' g**	**P < .05**	**Hedges' g**	**P < .05**	**Hedges' g**	**Same Result**
Group	High-Low: Ankle vs. Wrist	*	1.841		.626			No
	High-Low: Wrist vs. HR		.842			*	1.078	No

CHAPTER 4: DISCUSSION

4.1 Objective 1: Activity Volume

The purpose of Objective 1 was to determine if the four wearables models measured the same volume of sedentary, light activity, and MVPA. The volumes of activity measured in the present study were comparable to those found in the literature using a variety of methods including hip-worn accelerometers (Ayabe, Kumahara, Morimura, & Tanaka, 2014) and two different wrist-worn accelerometers (Rowlands, Yates, Davies, Khunti, & Edwardson, 2016) in adults below the age of 72 years. While this thesis is not a validation study, these similarities are important as they suggest that the models used in the present work performed comparably to other existing models and that the participants' activity levels are representative of a broader population.

The data supported the hypothesis that the models would measure different amounts of activity at each intensity. The Ankle model measured the greatest proportion of sedentary time, followed in descending order by the Wrist, HR, and HRAcc models. The HRAcc model measured the greatest proportion of light activity, followed in descending order by the HR, Ankle, and Wrist models. The Wrist model measured the greatest proportion of MVPA, followed in descending order by the HRAcc, HR, and Ankle models. Notably, only the Wrist and HR models measured the same volume of activity in at least two of the three intensities (sedentary and MVPA) but these models demonstrated the lowest epoch-by-epoch agreement. Conversely, both the HR vs. HRAcc and Ankle vs. HRAcc model pairs measured a significantly different volume of activity at all three intensities but demonstrated the highest and second highest degree of epoch-by-epoch agreement, respectively.

Between-model differences in activity volumes were large. Based on the interpretation of effect sizes by Cohen (1988), at least one large effect size was found in each activity intensity. Effect sizes (Hedges' g) ranged from .399 (HR vs. HRAcc) to 1.386 (Ankle vs. HRAcc) for sedentary time, .026 (Ankle vs. Wrist) to 1.598 (Wrist vs. HRAcc) for light activity, and .243 (Wrist vs. HRAcc) to .840 (Ankle vs. Wrist) for MVPA.

In addition to determining statistical significance, it is imperative to determine if the differences in measured activity are clinically significant. To do so, the normalized volume data can be extrapolated to number of hours per week. By assuming 15 hours of valid data per day (8 hours for sleep and one hour of device non-wear), the differences between the model that measured the most and the least time spent in each intensity category are 10.1 hours per week sedentary, 7.8 hours per week light activity, and 3.4 hours per week MVPA. Figure 29 shows the extrapolated activity volumes for all models.

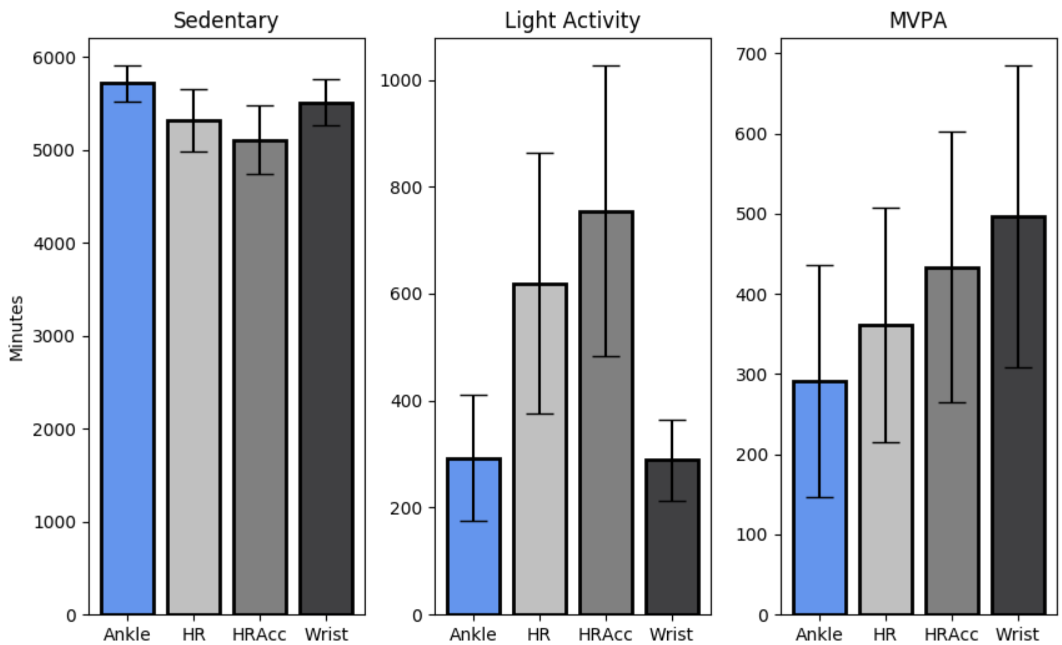

Figure 29: Activity volume extrapolated to the equivalent of 15 hours per day over a 7-day period.

Although guidelines do not report recommended amounts of light activity, one study found a 16% decrease in all-cause mortality for each hour of light activity per day (Loprinzi, 2017). Activity guidelines recommend 150 minutes per week (2.5 hours) of MVPA to prevent weight gain and 225 minutes per week (3.75 hours) to improve the chances of losing a clinically significant amount of weight (≥ 5% body weight) (Swift, Johannsen, Lavie, Earnest, & Church, 2014). The difference of 3.4 hours per week represents the equivalent of over 9 days' worth of recommended MVPA for weight maintenance and nearly the recommended weekly MVPA for clinically significant weight loss. Due to the health outcomes associated with these volumes of activity, these results are believed to be clinically significant.

The following sections discuss factors that may have affected the performance of each model.

4.1.1 Wrist Model

The Wrist model, developed by Powell and colleagues (2017), measured the second most sedentary time, the least light activity, and the most MVPA. The combination of least light activity and most MVPA suggests that the threshold for MVPA is the lowest of the four models. The cut-points for this model were originally developed in a sample aged 39.9 ± 11.5 years. The Primary sample had an age of 22.1 ± 4.4 years: much younger than the cut-point development sample. Because accelerometer counts are an absolute measure and aerobic capacity declines with age, it is likely that the relative intensity for a given activity count in a younger individual would be lower (Miller, Strath, Swartz, & Cashin, 2010) due to the failure of absolute measures to account for changes in aerobic capacity. Similarly, if the true resting VO_2 values in the Primary sample were lower than the measured resting VO_2 in the cut-point development sample, this could have led to a further overestimation in MET levels using the Wrist model. Using the resting VO_2 values found by Kwan & Kwok (2004), on average, this would be true for males but not for females in the Primary sample (females would have a slight underestimation in intensity). For males with a resting VO_2 equal to that reported by Kwan & Kwok, this difference would lead to a .22 MET overestimation when the Wrist model measured 3 METs and a .44 MET overestimation at 6 METs. For females, these differences would be negligible (underestimations of .06 and .13 METs, respectively).

Inspection of the Bland-Altman plots in the Powell et al. paper clearly shows that the variability of the difference scores between measured and predicted METs increases as their mean increases. For the MET ranges associated with sedentary and light activity, difference scores are quite low. However, for moderate and vigorous activity, these difference scores are more variable. Powell and colleagues discuss these data by highlighting the specificity values for the 3-MET cut-point. Average specificity was .902 for 1.5 METs and .965 for 6.0 METs, while the specificity for 3.0 METs was much lower with a value of .781. This lower specificity for 3.0 METs suggests that relatively more epochs that were truly light activity were classified as moderate activity than epochs that were truly sedentary being classified as light activity or truly moderate intensity epochs being classified as vigorous. Similar results, including increased variability in difference scores and lower specificity values for moderate activity, have been found in cut-points developed for adults using walking, running, and a variety of household activities (Esliger et al., 2011). The incorrect categorization of light activity as moderate activity is a potential explanation for why the Wrist model measured the most MVPA and least light activity.

The differences in activity volume between the Wrist model and the other models may be associated with performing different types of common activities. Depending on activity type, the same activity counts could be measured but the energy expenditure of those activities could be very different. For example, Powell and colleagues report an activity count of 57 during dish handling (i.e. an isolated upper limb activity) with a measured intensity of 1.52 METs. During walking between 2.5 and 4.5 km h^{-1} (i.e. a whole-body activity), the count value was 72.5 and the measured intensity was 3.14 METs. Although activity counts only differed by 27 percent between these two activities, energy expenditure doubled. Even with the inclusion of walking and tasks that mimic activities of daily living in the cut-point development protocol, without the implementation of pattern recognition to determine activity type, the energy expenditure measured by the model will be based on averaged values across different activities. This is an inherent limitation of using cut-points and single-device models.

4.1.2 Heart Rate Model

The HR model measured the third most sedentary time, second most light activity, and third most MVPA. The relatively low amount of sedentary time and relatively high amount of light activity can be partially explained by the limitations of using heart rate as a measure of intensity on its own. Heart rate can be increased by autonomic nervous system activity that is not caused by physical activity (Freedson & Miller, 2000; Villars et al., 2012; Brage et al., 2004). During periods of inactivity, these changes may increase heart rate above the 30% heart rate reserve threshold for light activity. Moss and Wynar (1970) found that psychological stress prior to delivering a presentation can even increase heart rate into the moderate intensity range. In the Primary sample, the average heart rate at 30% heart rate reserve was 95 bpm. Although this model did not use a flex heart rate *per se*, 95 bpm falls within the ranges of flex heart rates of 83 to 101 bpm and 68 to 118 bpm determined in two studies (Strath et al., 2002; Johansson et al., 2006), and close to the median flex heart rate of the latter of 90 bpm. Flex heart rate acts to safeguard against measuring activity from increased heart rate caused by non-activity factors. Since there is some overlap between the threshold for light activity and the flex heart rates that have been used in the literature, it is reasonable to suspect that some epochs under the influence of non-activity factors could have been measured as light activity using the HR model. It is less likely that measured MVPA was affected by non-activity factors. The average heart rate corresponding to 40% heart rate reserve in the Primary sample was 109 bpm. This value exceeds the range of flex heart rates found in (Strath et al., 2002) and approaches the upper limit found by (Johansson et al., 2006).

The HR model measured more light activity and more total activity (non-sedentary time) than the Wrist and Ankle models. This can be explained by what type of activities are measured by each model. The Wrist model measures both isolated upper limb and whole-body movements, and the Ankle model measures both isolated lower limb and whole-body movements. The HR model is able to detect isolated upper limb, isolated lower limb, and whole-body activities. It is also able to measure increased intensity in the absence of proportional increases in movement, such as during weight-bearing activities or while walking uphill. Due to its ability to measure all activity types, it is not surprising that the HR model measured the greatest proportion of light activity (as well as total activity) of the three single-device models.

Since the HR model quantified heart rate as a percent of heart rate reserve, its performance relied partly on the measurement of resting heart rate. The average resting heart rate measured from the free-living data was 53.0 bpm. This value is lower than what has been reported in largescale studies for males (71 bpm) and females (76 bpm) in the 20-to-39-year-old age group (Ostchega, Porter, Hughes, Dillon, & Nwankwo, 2011). Lower resting heart rate values increase the percent heart rate reserve value for a given absolute heart rate. Due to the relatively low measured resting heart rate, the current methods could have led to an increased intensity for the same absolute heart rate compared to more common clinical methods for measuring resting heart rate. However, due to the demographic characteristics of the sample, lower resting heart rate values can be expected. This sample was comprised mainly of undergraduate students studying Kinesiology (n=8; 80%). These students are likely to be more active and to have greater cardiovascular fitness than the sample in the Ostchega et al. study which was representative of the general population. This potential increased activity level is supported by the finding that this sample performed 361 and 497 minutes per week (extrapolated) according to the HR and Wrist models, respectively. Since fitter individuals have lower resting heart rates (Jensen, Suadicani, Hein, & Gyntelberg, 2013), this could explain the lower resting heart rate values found in the Primary sample. This apparent difference decreases when using resting heart rate data from more comparable samples. Melanson (2000) measured resting heart rate in a convenience sample of males recruited around a university campus. He found that individuals who were moderately active (an average activity energy expenditure of 1400 kcal per week) and aged 29.4 ± 3.1 years, had a resting heart rate of 53.7 ± 9.5 bpm. The study's low-activity group had a resting heart rate of 63.4 ± 8.3 bpm – both lower than the resting heart rates found by Ostchega and colleagues.

In addition to fitness, part of the differences in measured resting heart rates can be explained by posture during measurement. The Melanson study measured resting heart rate while supine and the Ostchega et al. study measured while seated. In the literature, these methods are used with the same frequency, but the expected difference between the two is a lower heart rate while supine of 1-2 bpm (Palatini, 2009). A postural requirement was not included in the derivation of resting heart rate in the current study, but it is suspected that the epochs used to derive resting heart rate were while participants were in supine or prone postures. Considering posture and the potential effect of fitness, the derived resting heart rate values are reasonable.

For different types of activities that elicit the same VO_2, heart rate tends to be higher during arm activity than during leg or whole-body activity (Strath et al., 2002). Therefore, it would be expected that for a given VO_2, the HR model would measure a higher intensity for upper limb compared to lower limb or whole-body activity. It is difficult to conclude whether this notion was supported by the results given the data were collected continuously without knowledge of what activity type was being performed and because of the nominal intensity scale. As mentioned, the HR model may have measured more activity due to its ability to measure all types of activity which the Wrist and Ankle models could not do. However, the expected trend given the heart rate-VO_2 relationships for upper and lower limb activity were supported by the data for MVPA; the HR model measured less MVPA than the Wrist model but more MVPA than the Ankle model.

4.1.2.1 ECG Signal Quality and Movement

Due to 53.1 ± 25.1% of individual participant's ECG data being unusable, an unplanned investigation was conducted to determine factors that may have affected signal quality. For the 36 participants included in this analysis, the average wrist and ankle activity counts during waking hours were calculated for epochs with valid and invalid ECG. 25 participants (65.8%) had higher average wrist counts and 21 participants (57.9%) had higher average ankle counts during invalid ECG epochs compared to valid ECG epochs. This is a conservative assessment as it did not account for device removal periods which would decrease the activity counts measured during invalid ECG epochs and lessened the difference. This finding suggests that movement may have affected the ECG signal quality due to motion artefact or by decreasing the quality of the electrode-skin connection. Unfortunately, the frequency content of typical human movement and ECG content overlap so some of this noise cannot be removed without affecting the ECG signal as well (Winter, 2009; Luo & Johnston, 2010).

If higher activity led to an increased chance of the ECG data becoming unusable, bias would be introduced into the calculation of intensity for all models as epochs with high intensity would be excluded from analyses. Overall, this would decrease the amount of measured activity for all models. Upper limb movement may have had a larger impact than lower limb movement on the ECG signal quality. The percent increase in activity counts during invalid epochs compared to valid epochs was 2.3 times greater for the wrist than the ankle. If movement truly did affect ECG signal quality, activity type would have played a major role in the bias that was introduced into these models. For example, running, where the ankles move much more than the upper limbs, may have had a lower chance of rendering the ECG signal unusable and therefore, running may have been measured more reliably by all models than a more upper limb-dominant activity such as activities of daily living.

4.1.3 Ankle Model

The Ankle model measured the most sedentary time, third most light activity, and the least MVPA. Of the Ankle, Wrist, and HR models, the Ankle model is the only model that requires whole body movement or cycling measure activity; upper limb activity can be performed without whole-body activity and heart rate can be influenced by non-activity factors or upper limb activity. Small lower limb movements would measure some activity, but the threshold that corresponded to walking at one's preferred speed for 5 seconds would have helped reduce these probably unmeaningful movements. The limited type of activities detected by the Ankle compared to Wrist and HR models may explain the low amount of measured activity. The Ankle model measured the least light activity, though this value was not significantly different than the Wrist model. Since the Ankle model predominantly measures walking, it is probable that the majority of its measured MVPA was from ambulation. In a sample of 20 to 59-year-olds, average cadence at preferred walking speed was found to be 112.5 ± 13.6 steps per minute and that walking at one's preferred pace required 12.1 mL O_2 kg^{-1} min^{-1} (Waters et al., 1988). Depending on what resting VO_2 value is used, this would equate to a value of approximately 3.5 to 4.0 METs. More recently, Tudor-Locke and colleagues (2019) measured cadence and intensity in adults aged 21 to 40 years and determined that a cadence of 102 steps per minute was equal to 3 METs. Both these studies suggest that walking at one's preferred pace would be considered moderate physical activity. Walking bouts that did not span a full 15-second epoch and were preceded or followed by sedentary time would have contributed to light activity since the intensity measured using activity counts reflects the average intensity of that epoch (Chen & Bassett, 2005). Since walking typically takes up such a small percentage of waking hours, it was not surprising that the Ankle model measured the least MVPA. Similarly, the Ankle model's inability to account for increased intensity at a given speed when moving up hill or during weight-bearing activities may have erroneously measured some slower walking with increased energy demand as light activity.

The relationship between heart rate and VO_2 differs between upper-body, lower-body, and whole-body activity (Strath et al., 2002). However, lower-body activity is more similar to whole-body activity than is upper-body activity (Haskell et al., 1993). Lower-body and whole-body energy demands are similar enough that lower-body activity is used to predict whole-body activity even when both upper-body and lower-body data are available (Strath et al., 2002). This is supported with the present results for MVPA; the difference between measured MVPA by the HR and Ankle models (5.7% and 4.6%, respectively, Hedges' g = .330) is half as large as the difference between MVPA measured by the HR and Wrist models (5.7% and 7.9%, respectively, Hedges' g = .553).

4.1.4 Combined Heart Rate-Accelerometer Model

The HRAcc model measured the least sedentary time, the most light activity, and the second most MVPA. Theoretically, the benefit of using a combined heart rate-accelerometer model is to ensure that any measured activity during periods of low heart rate occur in the context of movement. However, due to the elevated heart rate that remains once movement stops and the ever-present potential influence of non-activity factors on heart rate, this was not the case with the present HRAcc model.

The extent of this benefit depends on the selected flex heart rate threshold. By increasing the threshold, the likelihood of heart rates under the influence of non-activity factors getting classified as activity decreases, and the threshold would be crossed more quickly once movement stops which would mark the end of measured activity. However, increasing the flex heart rate threshold also leads to the HR portion of the model being used less frequently; this reduces the usefulness of including heart rate in the model. By selecting a flex heart rate of 30% heart rate reserve, which corresponded to an absolute heart rate near the middle of flex heart rates found in the literature, the effect of non-activity factors on heart rate would have been greater than if a higher flex heart rate had been selected. This threshold led to the HR model being used for an average of 11.0% of the valid epochs. This usage decreases to 4.3% with a 40% heart rate reserve threshold. Additionally, a 30% heart rate reserve threshold would lead to a larger volume of activity measured by the HR portion of the HRAcc model once movement had stopped, as it would take longer for heart rate to decrease below this threshold. Ultimately, selecting a flex heart rate is a trade-off between managing the influence of non-activity factors and how often the HR model is used. In contrast to the theoretical benefit of HRAcc models, due to the aforementioned limitations, the primary benefit of the HRAcc model in this thesis was its ability to detect activity using the Ankle model that were not associated with significant increases in heart rate.

The HRAcc model was unique in the requirements that needed to be met to be classified into the different activity intensity categories. To be considered sedentary, heart rate needed to be below 30% heart rate reserve and the Ankle model needed to measure sedentary time. This may be the most accurate measure of sedentary time due to its dual condition requirement. Light activity could be measured under two circumstances. First, if heart rate was between 30% and

40% heart rate reserve, regardless of the intensity measured by the Ankle model. Second, if heart rate was below 30% heart rate reserve but the Ankle model measured light activity; this likely occurred at the start of an activity bout before the autonomic nervous system evoked the required increase in heart rate. MVPA was also measured under two circumstances: with a heart rate above 40% heart rate reserve, or with a heart rate below 30% heart rate reserve with the Ankle model measuring MVPA. The use of the Ankle model is especially important at the onset of physical activity or when workload increases. With an increase in activity intensity, heart rate increases fairly quickly to meet the increased oxygen demand. The initial parasympathetic withdrawal and sympathetic activation has been shown to increase heart rate by 33 bpm within 4 seconds during maximal effort, unloaded cycling (Nobrega & Araújo, 1993). However, the sympathetic response needed to increase heart rate beyond its intrinsic rate lags behind the onset of activity by 20 seconds (Hughson et al., 2001). During this time, movement that should be measured as activity is occurring without an equivalent increase in heart rate which may lead to less activity being measured when using heart rate on its own. Depending on the timing and intensity of the start of movement relative to the start of an epoch, the use of the Ankle portion of the HRAcc model could measure MVPA while the HR model measures sedentary or light activity. Over an extended data collection period, these small differences can accumulate due to the many short walking bouts (Orendurff et al., 2008) that occur in free-living. The use of the Ankle portion of the combined model also allows detection of short walking bouts that may not have elicited a sufficient increase in heart rate to qualify as activity according to the HR model.

However, the combined heart rate-accelerometer method may overestimate activity duration. Once activity ends, heart rate does not instantly return to its resting rate. Depending on the intensity of the activity and its impact on body temperature and blood lactate levels, the return to resting heart rate takes approximately 100 seconds and remains elevated during this time (Zakynthinaki, 2015). Assuming no further activity is initiated during this recovery period, the HR portion of the combined model can overestimate activity duration and intensity until heart rate decreases below each intensity threshold. This is the likely explanation for why the HRAcc model measured less sedentary time and more light and MVPA than the independent Ankle and HR models.

4.2 Objective 2: Model Agreement

In contrast to the differences in total activity volume analyzed in Objective 1, Objective 2 aimed to determine the extent to which models agree on their epoch-by-epoch measurement of activity intensity. It was hypothesized that models would demonstrate moderate agreement (Cohen's κ between .40 and .60). The hypothesis was supported by the Ankle vs. Wrist and Ankle vs. HRAcc comparisons. Three of the six model comparisons demonstrated lower-than-hypothesized agreement (the Wrist vs. HR, Ankle vs. HR, and Wrist vs. HRAcc comparisons). The HR vs. HRAcc comparison demonstrated higher-than-hypothesized agreement.

4.2.1 HR vs. HRAcc and Ankle vs. HRAcc Agreement

The highest epoch-by-epoch agreement was found between the HR and HRAcc models with a Cohen's kappa of .884, and the second highest agreement was found between the Ankle and HRAcc models with a Cohen's kappa value of .505. These two comparisons were unique because the single-device model in each pair (the HR or Ankle model, respectively) was used in the HRAcc model.

The HR vs. HRAcc agreement was much greater than the Ankle vs. HRAcc agreement. The greater agreement between the HR and HRAcc models was surprising since, on average, the HRAcc model used the HR model approximately 11% of the time and the Ankle model 89% of the time. This discrepancy in agreement may be attributed to different situations which led to disagreement between the single-model and HRAcc models. In the HR vs. HRAcc comparison, the models would agree on intensity if the HR model measured light, moderate, or vigorous activity. The only source of disagreement would be epochs when heart rate was below 30% heart rate reserve and the Ankle model measured non-sedentary activity; this situation is likely to occur at the start of a whole-body or isolated lower limb activity bout. The impact of these frequent, short activity bouts accumulates over the course of the day to have a marked effect on the relationship between heart rate and ankle movement. This relationship can clearly be seen in Figure 16; the correlation is much weaker when heart rate is in the sedentary range (< 30% heart rate reserve). Although the HR and HRAcc models would agree on intensity for all epochs with heart rate above 30% heart rate reserve (i.e. the threshold for light intensity activity), active time represented a much smaller fraction of the total collection period than sedentary time (sedentary time ranged from 81.1% to 90.5% of the collection period, depending on model). However, for the Ankle vs. HRAcc comparison, there are many more opportunities for disagreement including any time the Ankle accelerometer measures a change in intensity. This led to the Ankle vs. HRAcc agreement being lower than the HR vs. HRAcc agreement.

4.2.2 Ankle vs. Wrist Agreement

Of the models that did not have common data, the Ankle vs. Wrist comparison was the only one to demonstrate moderate agreement ($\kappa = .502$). The largest advantage this model comparison had over the other comparisons was the use of accelerometers which change their output as soon as movement occurs, in contrast to physiological measures like heart rate whose responses to changes in workload are delayed in time. During whole-body activity bouts, upper- and lower-limb movements start at the same time and likely with the same intensities. For example, at the start of a walking bout, both the ankle and wrist accelerometers would immediately measure the intensity related to walking at that speed. Conversely, when comparing accelerometers to the HR model at the start of a bout, the measured changes in intensity would likely not be immediately equivalent, leading to disagreement in intensity classification. Both these examples hold true at the end of an activity bout as well.

Due to the types of activities that the Ankle and Wrist models can measure, the observed agreement relative to the other model comparisons was a bit surprising. Both models use accelerometers to measure movement as an estimate of intensity, but because the Wrist model can measure isolated upper limb activity, it was expected that many epochs would be categorized as light or MVPA by the Wrist model and as sedentary by the Ankle model. This was not supported by the data; both models measured the same volume of light activity and the moderate level of agreement suggests many of these epochs were categorized as the same intensity. These results suggest that whole-body movement, which would be captured by both accelerometers, occurred often enough to lead to the similar performance between these two models.

4.2.3 Wrist vs. HR Agreement

The Wrist vs. HR comparison had the fifth strongest agreement ($\kappa = .305$), which was not significantly different than the Ankle vs. HR comparison which showed the lowest agreement. Given the accuracy of the Wrist model as described by Powell and colleagues, this agreement was lower than expected. As discussed before, there is the potential to have “out of sync” accelerometer and heart rate data, especially at the onset and end of an activity bout. Although the Wrist and HR models did not measure significantly different amounts of sedentary time and MVPA, if the timing of numerous activity bouts were offset between models, this could explain the low level of agreement despite having measured similar activity volumes.

The Wrist model can measure isolated upper limb activities. For activity that requires a given VO_2, upper body activity results in a higher heart rate than lower limb or whole-body activity (Strath et al., 2002). This may have led to the HR model measuring a higher intensity than the Wrist model. The HR model can measure any type of activity. During epochs where the upper limb movement does not reflect overall intensity (stationary cycling or during weight-bearing activities, for example), this would have led to a disagreement in measured intensity between the two models. Conversely, non-activity factors can increase heart rate into the light or moderate intensity ranges (Moss & Wynar, 1970) in the absence of movement. This could have also led to the HR model measuring a higher intensity than the Wrist model. However, it is unlikely that non-activity factors would have had a significant effect on agreement as those periods of increased heart rate would likely represent a very small fraction of the total measurement period. The greater concern would be the frequent periods of isolated upper limb activity that exceed the acceleration threshold for activity but that do not elicit a large increase in energy expenditure, leading to misclassification as activity by the Wrist model.

4.2.4 Wrist vs. HRAcc Agreement

The Wrist vs. HRAcc comparison had the fourth strongest agreement ($\kappa = .380$). Logically, this level of agreement fell between the observed agreements between the Wrist vs. HR and Ankle vs. Wrist comparisons. The increase in agreement for the Wrist vs. HRAcc model over the Wrist vs. HR agreement could be caused by the use of the Ankle model. As discussed before, with the exclusion of isolated limb movements, wrist and ankle movement is typically synchronized in the onset of activity which increases the chance of agreement when comparing either accelerometer model to the HR model. The temporal synchronization of the Ankle and Wrist model was likely the factor that drove the increased agreement of the Wrist vs. HRAcc comparison relative to the Wrist vs. HR comparison.

4.2.5 Ankle vs. HR Agreement

Surprisingly, the Ankle vs. HR comparison had the lowest agreement (although it was not significantly lower than the Wrist vs. HR agreement) with a Cohen's kappa of .296. This agreement was expected to be better due to the ability of both models to measure walking – the most common activity (Hulteen et al., 2017). Once again, the offset in measured intensity at the start and end of activity or during isolated upper-limb activity that led to increases in heart rate may have led to a decrease in Ankle vs. HR agreement. This relatively low level of agreement highlights the importance of a combined HRAcc model that has the ability to better capture all types of activity while maintaining the ability to measure activity as its onset using accelerometry.

4.3 Objectives 3 and 4: The Effect of Activity Level on Model Performance

The purpose of Objective 3 was to determine if activity level has an effect on measured activity volume and epoch-by-epoch intensity agreement from four wearables models. In contrast to the hypothesis, there was no statistical evidence that activity level influenced the performance of any model in either the primary or second analyses (the latter with larger samples of participants having provided data for only two of the four models); the Activity Group by Model interaction was not statistically significant for any of the three activity intensities, nor for the epoch-by-epoch intensity agreement. However, it should be noted again that due to the Primary sample being comprised mainly of Kinesiology undergraduate students, this sample was likely more fit and had less variability in fitness than if the sample had been randomly sampled from the general population.

4.3.1 Objective 3: Activity Volume

It was hypothesized that there would be a significant interaction effect on activity volume for the same reasons that advocate for the use of combined HRAcc models; at low intensities, heart rate is susceptible to the influence of non-activity factors, and at high intensities, accelerometer counts may plateau. By stratifying the samples into high- and low-activity groups, the aim was to compare groups that spend different amounts of time in those low and high intensity ranges that could affect the performance of different models.

While there was a significant difference between activity groups in average ankle activity counts, this difference did not translate to the expected main effect of Activity Group on activity volumes. While this main effect was not required to observe a significant interaction effect, it does raise the concern of whether there was a true difference in activity levels between these groups despite the significant difference in average activity counts.

These results may have been confounded by cardiorespiratory capacity as it is likely related to activity level and model performance but was not assessed in the present study. After the age of approximately 30 years, cardiorespiratory capacity declines by 10% per decade (Plowman & Smith, 2007). By comparing data from the Primary and WristHR samples, the latter of which was older (37.9 ± 21.1 compared to 22.1 ± 4.5 years) and had a greater age range (18 to 76 compared to 18 to 34 years), the differences in model performance are consistent with the differences that would occur with decreases in cardiorespiratory capacity (refer back to Tables 3 and 8 for sample demographic information). The difference in activity volume measured by the HR model in both samples was smaller in magnitude than the difference measured by the Wrist models; this was true for all three intensities. Since the HR model uses a relative measure of intensity and the Wrist model uses an absolute measure, differences in cardiorespiratory capacity between samples would be partially accounted for with the HR model but not with the Wrist model, leading to the observed larger differences with the absolute measure. Further work is needed to determine if either or both of these variables affect model performance.

4.3.2 Objective 3: Model Agreement

Although there was no Activity Group by Model Comparison interaction on epoch-by-epoch agreement, there was a main effect of Activity Group. The high-activity group had a significantly greater overall level of agreement than the low-intensity group by a margin of κ = .083. Along with the non-significant interaction effect of Activity Group and Model on activity volume, this difference in agreement suggests that the selection of one model over another may have less of an effect on activity outcome measures for those with higher activity levels than those who are less active.

The largest observed difference between activity groups was the Ankle vs. Wrist agreement with a difference of κ = .165. Since groups were stratified using ankle accelerometer counts, it is possible that group stratification could have generated groups with different levels of ankle counts but the same amount of wrist counts which would have biased results. However, ankle and wrist counts were both significantly higher for the high-activity group so the observed difference in Ankle vs. Wrist agreement could be due to activity type differences. For example, those who are more active overall would be more likely to spend more time walking or performing whole-body exercise. This would lead to increased Ankle vs. Wrist agreement during periods of synchronized movement. Conversely, the greater agreement in the high-activity group may be a result of bias due to the removal of epochs with large amounts of movement which led to unusable ECG data. These periods were not removed in the secondary analysis of the Ankle vs. Wrist models, and the high- and low-activity groups' levels of agreement were much closer than in the Primary sample which did remove periods of unusable ECG signal. Notably, between samples, the low-activity groups had very similar values ($\Delta\kappa$ = .014), but the high-activity groups' kappa values differed by κ = .097. The secondary analysis' data may have provided a more robust test of agreement due to the inclusion of periods with more movement. The improved agreement due to more whole-body movement cannot be confirmed at this time, but a similar between-group trend for the Wrist vs. HRAcc comparison ($\Delta\kappa$ = .133), which also includes the Wrist and Ankle models, supports the notion that potentially increased levels of whole-body movement for more active individuals led to better model agreement between the Wrist and Ankle in the high-activity group.

The between-group differences for Wrist vs. HR and Ankle vs. HR agreement were much smaller in magnitude than the Ankle vs. Wrist difference; kappa values were higher in the high-activity group by .072 and .066, respectively. In addition to the potential differences in measured intensity between accelerometer models and the HR model at the onset and end of activity and the differences in the type of data captured by each model (i.e. acceleration or heart rate), fitness can also play a role in these differences. Compared to less fit individuals, higher cardiovascular fitness leads to a faster increase in heart rate at the start of activity, reaching steady state sooner, and to a faster decrease in heart rate once activity stops (Zakynthinaki, 2015). For high-fitness individuals, the apparent lag in heart rate at the onset and end of activity compared to the immediate change in accelerometer counts may not be as pronounced and could lead to better model agreement between the Ankle or Wrist and HR models. Fitness was not assessed in the present study, but assuming fitter individuals move more and generate more ankle activity counts, increased fitness could have contributed to the small improvement in the Ankle vs. HR and Wrist vs. HR agreements observed in the high-activity group compared to the low-activity group. The relative increase in the agreement between the Wrist and HR models in high-activity group in the secondary sample, which had a much larger age range and therefore likely a larger range of cardiovascular fitness than the primary sample, provides further evidence of the effect of fitness on improving agreement between HR and accelerometer models.

4.3.3 Objectives 3 and 4: General Discussion

The classification of intensity into three (for activity volume) and four (for model agreement) intensity categories may have prevented the finding of a significant interaction, at least for higher-intensity activity. Brage and colleagues (2004) are commonly cited for their finding that activity counts for a hip-worn accelerometer increase linearly during walking but eventually plateau when running at speeds above 9 km h^{-1}. With the collapse of moderate and vigorous activity into MVPA, this plateau would occur well beyond the threshold for moderate activity and would not be detectable in the three-category intensity data used to measure activity volume. The four intensity categories used in determining model agreement had an improved chance of being affected by the potential accelerometer plateau. However, using the ACSM's equation for running VO_2, a speed of 9 km h^{-1} would correspond to approximately 10 METs – well above the 6-MET threshold for vigorous activity.

Further, it is currently not known whether wrist or ankle accelerometer counts plateau or if this phenomenon only occurs in hip accelerometers. More recently, several studies have found that the plateau described by Brage and colleagues was due to two factors which would not occur using the methods in this thesis. First, the data were filtered in a way that reduced the signal amplitude at very high stepping cadences (Rowlands, Stone, & Eston, 2007). Second, the accelerometer that was used was uniaxial and only measured acceleration in the vertical axis; acceleration in the horizontal planes were not captured (Rowlands et al., 2007). Both (Rowlands et al., 2007) and (Fudge et al., 2007) found that activity counts derived from vector magnitude data from triaxial hip accelerometers did not experience this plateau. Cadence continues to increase as running speed increases (Rowlands et al., 2007), so ankle accelerometer counts would likely not plateau until running at near-maximum speed. The relationship between wrist accelerometer counts and running speed has not been well established, but Esliger and colleagues found a decrease in counts when increasing speed from 10 km h^{-1} to 12 km h^{-1} on the left wrist. However, an increase was observed on the right wrist, so it is not clear whether wrist-worn accelerometer counts plateau during running.

All considered, the impact of any potential plateau on activity intensity and volume measurements would likely be quite small for most individuals as the activities associated with that level of accelerometer counts represent a small portion of one's time and may even be unlikely to be attained by many. Due to the lack of evidence that activity level and the unlikeliness that near-maximal intensity activity on its own would lead to statistically significant differences in model performance, it is more plausible that model performance would be affected by fitness level. Further work is needed to determine the effect of fitness level on model performance.

4.4 Performance of the Ankle Model

The purpose of developing an ankle accelerometer model to quantify activity intensity was to increase the utility of ankle accelerometers in the hopes of reducing the number of sensors participants need to wear in order to provide a large volume of high-quality data. This model uses a novel technique by using ankle activity counts to predict gait speed and then predicting VO_2 using the ACSM's equations. A strong relationship between counts and speed was expected due to the linear relationship between cadence and gait speed during walking (speeds below 2.1 m s^{-1}) (Hansen, Kristensen, Nielsen, Voigt, & Madeleine, 2017; Latt, Menz, Fung, & Lord, 2008). By having each participant undergo the treadmill protocol and creating individual regression equations, individual variability in cadence for a given gait speed was accounted for which likely improved the performance of the equations over a single group-level equation.

Individual linear regression equations showed extremely high coefficients of determination without any additional predictor variables; r^2 values were greater than .975 for all but one participant ($r^2 = .878$). Prediction accuracy was also very high as the average standard error of estimate was less than 0.04 m s^{-1}. This is an acceptable amount of error as it is less than the clinically meaningful difference in gait speed associated with reduced self-reported mobility (.05 to .10 m s^{-1}) (Perera, Mody, Woodman, & Studenski, 2006) and reduced disability following a stroke (.16 m s^{-1}) (Tilson et al., 2010).

However, there are several potential sources of error. First, in the development of the Ankle model, a treadmill protocol was used to develop a regression equation that would ultimately predict gait speed and VO_2 during over-ground walking in free-living. Walking on a treadmill leads to a slower preferred speed compared to during over-ground walking (Sloot, van der Krogt, Harlaar, 2014). If preferred gait speed on a treadmill were slower than over-ground walking, the threshold for meaningful activity in the Ankle model could be too low; this would increase the chance of classifying unmeaningful movement as meaningful activity.

Second, previous work has examined spatiotemporal differences between treadmill and over-ground walking with conflicting results. For a given gait speed, Song and Hidler (2008) found no differences between stride length and cadence while Stolze and colleagues (1997) found a significant decrease in stride length of 4% and a significant increase in cadence of 6% when walking on a treadmill compared to over-ground walking. Song and Hidler (2008) also found that the impact at heel strike was smaller when on a treadmill. With larger impact forces during over-ground walking, activity counts could increase for a given speed compared to treadmill walking. If cadence at a given speed increases during treadmill walking and/or heel strike impact is lessened on a treadmill, over-ground walking speed could be overestimated when relying on activity counts.

A third potential source of error is the acceleration that occurs during stance phase. During over-ground walking, little acceleration is experienced by an ankle-worn accelerometer during stance phase; the foot is on the ground and the ankle dorsiflexes as the contralateral limb takes its step. However, during treadmill walking, the motion of the treadmill belt causes the entire lower limb to translate posteriorly during stance phase in addition to generating dorsiflexion. It is unclear if the additional acceleration from the treadmill would be apparent in epoched data. Further analysis of the acceleration profiles for treadmill and over-ground walking is required to determine if there are practically significant differences and if potential differences remain when data are expressed as activity counts.

The validity of using the ACSM equations to predict energy expenditure should also be considered. Hall and colleagues (2004) investigated the differences in measured and equation-predicted energy expenditure during treadmill and over-ground walking. They found no significant differences in total energy expenditure between over-ground and treadmill walking and running. Over a 1600 m bout at 1.41 m s^{-1}, they found that the ACSM walking equation underestimated total energy expenditure by 3.8%. For an individual of 70.5 kg (their average participant's mass) and assuming 1L O_2 per 5 kcal (Hills et al., 2014), this underestimation equates to less than 0.3 METs. The magnitude of error was similar for running at 2.82 m s^{-1} but in the opposite direction: an overestimation of 4.3% which equates to 0.5 METs. These results support the validity of using the ACSM equations to predict activity intensity in free-living. However, the ACSM describes the walking and running equations as being most accurate for gait speeds between 0.83 and 1.66 m s^{-1} and greater than 2.25 m s^{-1}, respectively. Because of the gap between these ranges, speeds between 1.66 and 2.25 m s^{-1} may not predict VO_2 as accurately as speeds that fall within these two ranges. The present thesis used the running equation for all speeds above 1.66 m s^{-1}. Because the slope of the running equation is double that of the walking equation, speeds between 1.66 and 2.25 m s^{-1} while walking likely led to overestimations in VO_2, especially given that the typical walk-to-run transition occurs between 1.88 (Talor, Heglund, & Maloiy, 1982) and 2.24 m s^{-1} (Hansen et al., 2017). Using the walking equation, the range of 1.66 to 2.25 m s^{-1} predicts approximately 4.0 to 5.1 METs, depending on the value used for resting VO_2. Due to the collapse into MVPA in the present work and 1.66 m s^{-1} falling into the MVPA range ($\geq$ 3 METs), the discrepancy between the walking and running equations in VO_2

calculation would not be apparent in the activity volume data but it may have had an effect on model agreement since moderate and vigorous activity were not combined in that analysis.

Two sources of error related to changes made to the ACSM equations may have influenced the activity intensity measured by the Ankle model. First, resting VO_2 was estimated using data from (Kwan & Kwok, 2004). Compared to the typical resting VO_2 value of 3.5 mL kg^{-1} min^{-1}, the lower values reported by Kwan & Kwok would increase MET level estimates for a given VO_2. Given the underestimation in predicted energy expenditure during walking when using the ACSM equation (Hall et al., 2004), the decrease in the value of resting VO_2 may have compensated for this underestimation and led to more accurate MET values. However, it is possible that the increase in MET levels may have overcompensated and led to an overestimation in MET levels. Secondly, the components of the ACSM equations that account for vertical movement were not included because accelerometer orientation data, which would be used to calculate incline angle, cannot be calculated from epoched accelerometer data. Even at a shallow incline of 5%, predicted VO_2 is approximately 50% higher when the vertical component is included for both walking and running. Although it would be possible to calculate incline angles using raw accelerometer data, this would greatly increase the data processing burden and would require advanced analytics to accurately calculate the incline's angle.

4.5 Measuring Resting Heart Rate

Activity measures that are relative to cardiorespiratory capacity provide a more accurate representation of intensity, especially for older individuals or those with lower cardiorespiratory capacities (ACSM, 2014). To allow the use of a relative measure in the HR model while limiting the burden placed on participants, maximum heart rate was predicted and resting heart rate was derived from free-living data instead of measuring it in the laboratory. No standardized methods were found for deriving resting heart rate using free-living, continuous data. Guidelines for taking heart rate clinically (Palatini, 2009) were adapted for use in continuous data using methods similar to those found in (Logan et al., 2000). For the present work, resting heart rate was derived from 30 minutes' worth of data when the participants were awake; this is much longer than the clinically recommended 30 seconds measured on two occasions (Palatini, 2009).

Since heart rates were averaged over one-minute intervals, this method also reduces the magnitude of potential errors in calculated heart rate due to inaccurate ECG peak detection.

Because there are no standardized guidelines for deriving resting heart rate from continuous data, it should be noted that changes in the interval over which heart rates are averaged or how many of these intervals are included in the calculation can affect the derived resting heart rate. Logan and colleagues (2000) found a difference of 8 bpm in resting heart rate depending on calculation method in a sample of pre-school children. All derived heart rates were lower than the heart rate measured during 5 minutes of rest shortly after waking up. Although these differences are large on an absolute scale (beats per minute), they are scaled down when calculating percent heart rate reserve by a factor of *100 - % HRR*. For example, if resting heart rate is calculated using two methods and a difference of 10 bpm is found, for an individual in their early 20s (i.e. maximum heart rate around 190 bpm), the heart rate needed to reach 30% heart rate reserve differs by approximately 7 bpm between these two methods (that is, (100% – 30%) x 10 bpm). Practically, underestimating resting heart rate leads to a lower absolute heart rate needed to attain a given percent heart rate reserve which can lead to overestimations of activity intensity. This highlights the importance of how resting heart rate is calculated and the need to standardize its derivation using continuous data.

4.6 Strengths

This thesis developed the novel method of using activity counts from the ankle to predict gait speed. Given the very promising results, it is worth expanding on this proof of concept with a validation study. If ankle accelerometers can be used for both activity intensity and additional analyses such as gait or balance assessments, the need for accelerometers worn on other body segments may not be necessary which could reduce participant burden. Stemming from the development of the Ankle model, this study created the first HRAcc model that uses an ankle accelerometer, allowing comparisons between models that have not been described previously.

This thesis addresses the issue of determining the usability of ECG data during multi-day, continuous wear. The algorithm that was adapted in the present work provided the foundation for the ECG signal quality check, but it did not perform well in identifying unusable data when the

Bittium Faros was removed. The addition of the fifth condition to the algorithm improved its performance in a free-living ECG dataset.

This study also provides a novel comparison between wearables models with its description of differences in total activity volume in addition to epoch-by-epoch activity intensity in free-living. Agreement measures are fairly common in the literature but only for total activity volume or energy expenditure, or during short, laboratory tests. While these comparisons are important, understanding the epoch-by-epoch changes in intensity between models increases in value as data analytics advance and incorporate additional measures such as social interactions or speech which can also occur sporadically and for short durations.

4.7 Limitations

Notable limitations include the relative homogeneity in the Primary sample which limits the external validity of the results of Objectives 1 through 3 to young, healthy, adults. However, with the inclusion of the secondary samples, this limitation was partially overcome as the WristHR sample had a much larger age range. For Activity Group comparisons, age, sex, and fitness were not accounted for and could have confounded results.

The reliable detection of QRS complexes in very noisy ECG data was not possible using the current peak detection methods and led to the exclusion of 48.3% of the total collected ECG data. Alternative techniques to measure heart rate, such as photoplethysmography (PPG), can be used, however, these are also susceptible to noise. Orphanidou and colleagues also tested PPG data and found that less PPG data were usable than ECG data (55.2% and 66.0%, respectively). Their PPG algorithm was not as accurate in classifying data segments as usable or unusable as the ECG algorithm. This does not make PPG a viable alternative. The influence of noise in the ECG signal due to movement artefact or from electrode adherence issues that occur during continuous wear will need to be further addressed by improved analytic techniques to better extract the ECG signal or reduced proactively by improving electrode adhesion materials.

Due to project timeline restrictions, exclusion of periods of sleep and non-wear were determined through visual inspection by the author with the help of participant logs instead of through the use of existing validated algorithms. Visual inspection was performed while looking for similar characteristics to those that algorithms use such as extended periods of no, if any, change in accelerometer counts. A common non-wear detection algorithm (Choi, Liu, Matthews, & Buchowski, 2011) relies on activity count data, as did the present study. Due to the relative simplicity of finding non-wear periods visually, visual detection should yield similar results to the algorithm. Compared to visual sleep detection, visual non-wear detection likely led to less error or bias due to the difficulty in determining when a participant fell asleep using activity count data in the absence of other measures such as body temperature, ambient lighting, or heart rate. This resulted in classifying the sedentary period in bed before falling asleep as “asleep”. This method would therefore reduce the participant’s sedentary time by the amount of time spent in bed while awake. The period of sleep was also continuous until the participant appeared to awake the next morning; any activity bouts that occurred in the night were not considered.

Lastly, the study was underpowered for several of the analyses. The original target sample size was calculated with the goal of stratifying participants into activity groups using the most and least active 25% of the sample to ensure that these groups were truly different in their activity levels. This method led to a target sample size of n=28 with activity groups of n=7. Due to this sample size not being attained in the Primary sample (n=10), groups were created using the most and least active 50% of participants, leading to two activity groups of n=5. Post-hoc power analysis revealed that the power associated with the Activity Group x Model effect for sedentary and light activity did not reach the desired level of 80% (63.0% and 28.6%, respectively). However, 80% power was attained for the Activity Group x Model interaction for vigorous activity and epoch-by-epoch agreement.

4.8 Future Directions

This thesis proposed a novel model for measuring free-living activity using ankle accelerometer counts to predict gait speed. Further work is needed to validate this model during over-ground walking, during running, and in populations with mobility impairments such as asymmetric gait and for those who use a gait aid. Further, the use of individually calibrated regression equations should be compared to a group-level regression equation. If this model proves accurate on a group level, participant burden could be further reduced by eliminating the need for each participant to perform the treadmill protocol. Similarly, the investigation into multivariate regression may make this endeavour more successful.

Future research should work towards the standardization of methods to improve the ability to accurately quantify activity and volume, and to compare between studies. For example, working towards the standardization of epoch length, which has been shown to affect measured activity volume (Ayabe et al., 2014) and may affect epoch-by-epoch agreement, or the measurement of resting heart rate from continuous, free-living data.

This thesis provides a valuable description and discussion about the epoch-by-epoch agreement in intensity which could be used in the development of new multi-device models. Such models could measure activity intensity using a more complex algorithm to choose which model to use for a given epoch; this strategy could overcome the limitations of the HRAcc model in the present work. One such limitation is that when an activity bout ends, the Ankle model measures sedentary time, but the current HRAcc model would not measure sedentary time until the heart rate drops below 30% heart rate reserve. A more advanced model could use the accelerometer to ensure that a certain level of movement is maintained during a specifically defined activity bout (while measuring intensity using the HR model if above its threshold), but would revert back to using the accelerometer model at the end of activity while heart rate remained above its model use threshold before returning to its resting rate. This latter type of model could also incorporate a second accelerometer to differentiate upper-body from lower- or whole-body activity, similar to what was done in the model from Strath and colleagues (2002). Alternatively, data epoching could be performed relative to activity bouts instead of using fixed duration intervals. These options may better reflect the intensity at the onset and offset of a bout and the bout’s duration.

As wearables become more advanced and feature multiple devices and sensors, new opportunities in measuring activity arise. However, these opportunities are not without their challenges. Measuring activity data in ways that are both meaningful to clinicians and to participants/patients is crucial in creating a framework that helps inform clinical decisions in addition to actionable goals for participants and patients. Future work should expand its focus to the reporting of FITT-V principles in addition to variables measured on a continuous scale, such as daily energy expenditure. These outcome measures can be more easily interpreted by patients or participants and can be used to guide clinical decision making related to physical activity interventions or lifestyle changes. As the ONDRI projects continue, maximizing the effectiveness of data acquisition and delivering meaningful data to both clinicians and participants/patients will be a continued focus.

Chapter 5: Summary and Conclusion

The present study confirms the hypothesis and supports the literature in that different amounts of physical activity are measured by different wearables models. In the Primary sample, which was comprised mainly of young, apparently healthy, undergraduate students, 5 of the 6 model comparisons measured statistically different amounts of sedentary and light activity while 3 of the 6 comparisons measured different amounts of MPVA. Only the Wrist and HR models measured the same total amount of more than one activity intensity. These differences are large enough to be clinically significant. Cohen's kappa values used to quantify epoch-by-epoch model agreement in intensity classification ranged from fair to almost perfect agreement. The HR vs. HRAcc models demonstrated the highest agreement while the Ankle vs. HR and Wrist vs. HR models demonstrated the lowest agreement. Activity level, as measured by average ankle counts, did not appear to have an effect on the differences in measured activity volume or epoch-by-epoch agreement between models, however, those who were more active had higher overall epoch-by-epoch agreement. These results were confirmed in the secondary analyses.

The present work cannot recommend using any of the four activity models inter-changeably as no two models measured the same amount of sedentary, light, and moderate-to-vigorous physical activity. Epoch-by-epoch agreement in activity intensity classification was also lower than hypothesized for several pairs of models, so not all model pairs tell the same "story" on an epoch-by-epoch basis. With its use of individually calibrated regression equations using a standardized treadmill protocol, this study provides evidence that ankle activity counts can be used to measure activity intensity using methods similar to those commonly used with wrist- and hip-worn accelerometers. Considering the limitations of using single-device models and the current HRAcc model, the results from this study advocate for the advancement of more complex combined heart rate-accelerometer models which have the potential to more accurately measure activity duration and intensity across all types of activities.

References

Aguilar-Farías, N., Brown, W. J., & Peeters, G. M. E. E. G. (2014). ActiGraph GT3X+ cut-points for identifying sedentary behaviour in older adults in free-living environments. *Journal of Science and Medicine in Sport*, *17*(3), 293–299.

Ainsworth, B. E. (2009). How do I measure physical activity in my patients? Questionnaires and objective methods. *Br J Sports Med*, *43*(1), 6-9.

American College of Sports Medicine. (2014). *ACSM Guidelines for Exercise Testing and Prescription*. *American College of Sports Medicine* (Vol. 27).

Assah, F. K., Ekelund, U., Brage, S., Wright, A., Mbanya, J. C., & Wareham, N. J. (2011). Accuracy and validity of a combined heart rate and motion sensor for the measurement of free-living physical activity energy expenditure in adults in Cameroon. *International Journal of Epidemiology*, *40*(1), 112–120.

Awais, M., Chiari, L., Ihlen, E. A. F., Helbostad, J., & Palmerini, L. (2018). Physical Activity Classification for Elderly People in Free Living Conditions. *IEEE Journal of Biomedical and Health Informatics*, *23*(1). 197-207.

Ayabe, M., Kumahara, H., Morimura, K., & Tanaka, H. (2014). Epoch length and the physical activity bout analysis: An accelerometry research issue. *BMC Research Notes*, *7*(1).

Bassett, D. R., Troiano, R. P., Mcclain, J. J., & Wolff, D. L. (2015). Accelerometer-based physical activity: Total volume per day and standardized measures. *Medicine and Science in Sports and Exercise*, *47*(4), 833–838.

Berger, B. G., & Tobar, D. A. (2011). Exercise and quality of life. *New Sport and Exercise Psychology Companion*, 483–505.

Bize, R., Johnson, J. A., & Plotnikoff, R. C. (2007). Physical activity level and health-related quality of life in the general adult population: A systematic review. *Preventive Medicine*, *45*(6), 401–415.

Bland, J., & Altman, G. (1986). Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *The Lancet*, *327*(8476), 307-310.

Borg, G. A. V. (1982). Psychophysical bases of perceived exertion. *Medicine & Science in Sports & Exercise*, *14*(5), 377-381.

Brage, S., Brage, N., Franks, P. W., Ekelund, U., Wong, M.-Y., Andersen, L. B., et al. (2004). Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. *Journal of Applied Physiology*, *96*(1), 343–351.

Brooks, A. G., Withers, R. T., Gore, C. J., Vogler, A. J., Plummer, J., & Cormack, J. (2004). Measurement and prediction of METs during household activities in 35- to 45-year-old females. *European Journal of Applied Physiology, 91*(5–6), 638–648.

Byrne, N.M., Hills, A. P., Hunter, G. R., Weinsier, R. L., & Schutz, Y. (2005). Metabolic equivalents: one size does not fit all. *J Appl Physiol, 99*(3), 1112-1119.

Canadian Society for Exercise Physiology (2011). Canadian Physical Activity Guidelines: For Adults - 18-64 Years. Retrieved from https://www.csep.ca/CMFiles/Guidelines/CSEP_PAGuidelines_adults_en.pdf

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical Activitiy, Exercise, and Physical Fitness: Definitions and Distinctions for Health-Related Research. *Public Health Reports*, *100*(2), 126–131.

Ceesay, S. M., Prentice, A. M., Day, K. C., Murgatroyd, P. R., Golderb, G. R., et al. (1989). The use of heart rate monitoring in the estimation of energy expenditure: a validation study using indirect whole-body calorimetry. *Br J Nutr*, *61*(2), 175-186.

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

Choi, L., Liu, Z., Matthews, C. E., & Buchowski, M. S. (2011). Validation of accelerometer wear and nonwear time classification algorithm. *Medicine and Science in Sports and Exercise*, *43*(2), 357–364.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement. 20*(1).

Cohen, J. (1988). *Statistical Power Analysis for Behavioral Sciences.* 2nd edition. Lawrence Erlbuam Associates.

Copeland, J., & Esliger, D. (2009). Accelerometer assessment of physical activity in active, healthy older adults. *Journal of Aging & Physical Activity*, *17*(1), 17–30.

Danks, K. A., Roos, M. A., McCoy, D., & Reisman, D. S. (2014). A step activity monitoring program improves real world walking activity post stroke. *Disability and Rehabilitation, 36*(26), 2233–2236.

Dillon, C. B., Fitzgerald, A. P., Kearney, P. M., Perry, I. J., Rennie, K. L., Kozarski, R., &

Phillips, C. M. (2016). Number of Days Required to Estimate Habitual Activity Using Wrist-Worn GENEActiv Accelerometer: A Cross-Sectional Study. *PloS One*, *11*(5), e0109913.

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

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

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Foster, R. C., Lanningham-Foster, L. M., Manohar, C., McCrady, S. K., Nysse, L. J., Kaufman, K. R., … Levine, J. A. (2005). Precision and accuracy of an ankle-worn accelerometer-based pedometer in step counting and energy expenditure. *Preventive Medicine*, *41*(3–4), 778–783.

Fox, S. M. III, Naughton, J. P., & Haskell, W. L. (1971). Physical activity and the prevention of coronary heart disease. *Ann Clin Res*. *3*(6). 404-432.

Franckowiak, S. C., Dobrosielski, D. A., Reilley, S. M., Walston, J. D., & Andersen, R. E. (2011). Maximal Heart Rate Prediction in Adults that are Overweight or Obese. *J Strength Cond Res*, *25*(5), 1407–1412.

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

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

Fudge, B. W., Wilson, J., Easton, C., Irwin, L., Clark, J., Haddow, O., … Pitsiladis, Y. P. (2007). Estimation of oxygen uptake during fast running using accelerometry and heart rate. *Medicine and Science in Sports and Exercise, 39*(1), 192–198.

Hall, C., Figueroa, A., Fernhall, B., & Kanaley, J. A. (2004). Energy expenditure of walking and running: Comparison with prediction equations. *Medicine and Science in Sports and Exercise*, *36*(12), 2128–2134.

Hall, K. S., Howe, C. A., Rana, S. R., Martin, C. L., & Morey, M. C. (2013). METs and Accelerometry of Walking in Older Adults: Standard versus Measured Energy Cost. *Med Sci Sports Exerc*, *34*(3), 574–582.

Hansen, E. A., Kristensen, L. A. R., Nielsen, A. M., Voigt, M., & Madeleine, P. (2017). The role of stride frequency for walk-to-run transition in humans. *Scientific Reports*, *7*(1), 1–8.

Haskell, W. L., Yee, M. C., Evans, A., & Irby, P. J. (1993). Simultaneous measurement of heart rate and body motion to quantitate physical activity. *Medicine and Science in Sports and Exercise*, 25(1), 109-115.

Hawley, J. A., Hargreaves, M., Joyner, M. J., & Zierath, J. R. (2014). Integrative biology of exercise. *Cell*, *159*(4), 738–749.

Hendelmen, D., Miller, K., Baggett, C., Debold, E., & Freedson, P. (2000). Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Med Sci Sports Exerc, 32*(9), 442-449.

Hills, A. P., Mokhtar, N., & Byrne, N. M. (2014). Assessment of Physical Activity and Energy Expenditure: An Overview of Objective Measures. *Frontiers in Nutrition*, *1*(June), 1–16.

Hughson, R. L., Tschakovsky, M. E., & Houston, M. E. (2001). Regulation of oxygen consumption at the onset of exercise. *Exercise and Sport Sciences Reviews*, *29*(3), 129–133.

Hulteen, R. M., Smith, J. J., Morgan, P. J., Barnett, L. M., Hallal, P.C., et al. (2017). Global participation in sport and leisure-time physical activities: a systematic review and meta-analysis. *Prev Med, 95*, 14-25.

Janssen, I., & LeBlanc, A. G. (2010). Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. *International Journal of Behavioral Nutrition and Physical Activity*, *7*(40).

Jensen, M. T., Suadicani, P., Hein, H. O., & Gyntelberg, F. (2013). Elevated resting heart rate, physical fitness and all-cause mortality: A 16-year follow-up in the Copenhagen Male Study. *Heart*, *99*(12), 882–887.

Johansson, H. P., Rossander-Hulthén, L., Slinde, F., & Ekblom, B. (2006). Accelerometry combined with heart rate telemetry in the assessment of total energy expenditure. *British Journal of Nutrition*, *95*(3), 631–639.

Karvonen, M. J., Kentala, E., & Mustala, O. (1957). The effects of training on heart rate; a longitudinal study. *Ann Med Exp Biol Fenn*, *35*(3), 307-315.

Klassen, T. D., Simpson, L. A., Lim, S. B., Louie, D. R., Parappilly, B., Sakakibara, B. M., … Eng, J. J. (2016). “Stepping Up” Activity Poststroke: Ankle-Positioned Accelerometer Can Accurately Record Steps During Slow Walking. *Physical Therapy*, *96*(3), 355–360.

Kwan, M., Woo, J., & Kwok, T. (2004). The standard oxygen consumption value equivalent to one metabolic equivalent (3.5 ml/min/kg) is not appropriate for elderly people. *International Journal of Food Sciences and Nutrition*, *55*(3), 179–182.

Läkens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*(NOV), 1–12.

Latt, M. D., Menz, H. B., Fung, V. S., & Lord, S. R. (2008). Walking speed, cadence and step length are selected to optimize the stability of head and pelvis accelerations. *Experimental Brain Research*, *184*(2), 201–209.

Laughlin, M. H. (1999). Cardiovascular response to exercise. *Advances in Physiology Education Can*, *277*(22), 244–259.

Lee, J., Cho, S., Lee, J., Lee, K., & Yang, H. (2007). Wearable Accelerometer System for Measuring the Temporal Parameters of Gait. *Conference of the IEEE EMBS*, 483–486.

Leonard, W. R. (2003). Measuring human energy expenditure: What have we learned from the flex-heart rate method? *American Journal of Human Biology*, *15*(4), 479–489.

Logan, N., Reilly, J. J., Grant, S., & Paton, J. Y. (2000). Resting heart rate definition and its effect on apparent levels of physical activity in young children. *Medicine and Science in Sports and Exercise*, *32*(1), 162–166.

Londeree, B. R., & Ames, S.A. (1976). Trend analysis of the %VO2max-HR regression. *Medicine Sciences Sports*. *8*(2). 123-125.

Londeree, B. R., & Moeschberger, M. L. (1984). Influence of age and other factors on maximal heart rate. *Journal of Cardiac Rehabilitation*, *4*(2), 44-49.

Loprinzi, P. D. (2017). Light-Intensity Physical Activity and All-Cause Mortality. *American Journal of Health Promotion*, *31*(4), 340–342.

Luo, S., & Johnston, P. (2010). A review of electrocardiogram filtering. *Journal of Electrocardiology*, *43*(6), 486–496.

McCracken, L. A., Ma, J. K., Voss, C., Chan, F. H., Martin Ginis, K. A., & West, C. R. (2018).

Wrist Accelerometry for Physical Activity Measurement in Individuals With Spinal Cord Injury—A Need for Individually Calibrated Cut-Points. *Archives of Physical Medicine and Rehabilitation*, *99*(4), 684–689.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochem Med*, *22*(3), 276–282.

Meijer, G. A., Westerterp, K. R., Koper, H., & ten Hoor, F. (1989). Assessment of energy expenditure by recording heart rate and body acceleration. *Med Sci Sports Exerc*, *221*, 343-347.

Melanson, E. L. (2000). Resting heart rate variability in men varying in habitual physical activity. *Medicine & Science in Sports & Exercise*, *32*(11), 1894–1901.

Migueles, J. H., Cadenas-Sanchez, C., Ekelund, U., Delisle Nyström, C., Mora-Gonzalez, J., Löf, M., … Ortega, F. B. (2017). Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. *Sports Medicine*, *47*(9), 1821–1845.

Miller, N. E., Strath, S. J., Swartz, A.M., & Cashin, S. E. (2010). Estimating absolute and relative physical activity intensity across age via accelerometer in adults. *J Aging Phys Act*, 18(2), 158-170.

Miller, W. C., Wallace, J. P., & Eggert, K. E. (1993). Predicting max HR and the HR-VO2 relationship for exercise prescription in obesity. *Medicine and Science in Sports and Exercise*, 25(9), 1077-1081.

Moss, A. & Wynar, B. (1970). Tachycardia in House Officers Presenting Cases as Grand Rounds. *Annals of Internal Medicine*, *72*(2), 255-256.

Nelson, M. B., Kaminsky, L. A., Dickin, D. C., & Montoye, A. H. K. (2016). Validity of Consumer-Based Physical Activity Monitors for Specific Activity Types. *Medicine and Science in Sports and Exercise*, *48*(8), 1619–1628.

Nes, B. M., Janszky, I., Wisløff, U., Støylen, A., & Karlsen, T. (2013). Age-predicted maximal heart rate in healthy subjects : The HUNT Fitness Study. *Scand J Med Sci Sports*, 23(6), 697–704.

Nilsson, A., Ekelund, U., Yngve, A., & Sjostrom, M. (2002). Assessing physical activity among children with accelerometers using different time sampling intervals and placements. *Pediatric Exercise Science*, *14*(1), 87-96.

Nobrega, A. C. L., & Araujo, C. G. S. (1993). Heart rate transient at the onset of active and passive dynamic exercise. *Medicine and Science in Sports and Exercise*, *25*(1), 37–41.

Opondo, M. A., Sarma, S., & Levine, B. D. (2015). The Cardiovascular Physiology of Sports and Exercise. *Clin Sports Med*, *34*, 391–404.

Orendurff, M. S., Schoen, J. A., Bernatz, G. C., Segal, A. D., & Klute, G. K. (2008). How humans walk: bout duration, steps per bout, and rest duration. *Journal of Rehabilitation Research and Development*, *45*(7), 1077–1089.

Orphanidou, C., Bonnici, T., Charlton, P., Clifton, D., Vallance, D., & Tarassenko, L. (2015). Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring. *IEEE Journal of Biomedical and Health Informatics*, *19*(3), 832–838.

Ostchega, Y., Porter, K. S., Hughes, J., Dillon, C. F., & Nwankwo, T. (2011). Resting pulse rate reference data for children, adolescents, and adults: United States, 1999-2008. *National Health Statistics Reports*, 41, 1-16.

Ozemek, C., Cochran, H. L., Strath, S. J., Byun, W., & Kaminsky, L. A. (2013). Estimating relative intensity using individualized accelerometer cutpoints: The importance of fitness level. *BMC Medical Research Methodology*, *13*(53).

Palatini, P. (2009). Recommendations on how to measure resting heart rate. *Medicographia, 31*, 414-419.

Pan, J., & Tompkins, W. J. (1985). A simple real-time QRS detection algorithm. *IEE Transactions on Biomedical Engineering*, *32*(3), 230-235.

Pärkkä, J., Ermes, M., Antila, K., van Gils, M., Manttari, A., & Nieminen, H. (2007). Estimating Intensity of Physical Activity: A Comparison of Wearable Accelerometer and Gyro Sensors and 3 Sensor Locations. *Conf Proc IEEE Eng Med Biol Soc*, 1511–1514.

Pedišić, Ž., & Bauman, A. (2015). Accelerometer-based measures in physical activity surveillance: Current practices and issues. *British Journal of Sports Medicine*, *49*(4), 219–223.

Perera, S., Mody, S. H., Woodman, R. C., & Studenski, S. A. (2006). Meaningful change and responsiveness in common physical performance measures in older adults. *Journal of the American Geriatrics Society*, *54*(5), 743–749.

Plowman, S., & Smith, D. (2007). Cardiovascular responses to exercise. In *Exercise Physiology for Health, Fitness, and Performance, Second Edition* (pp. 351–382). Wolters Kluwer.

Pouryayevali, S., Wahabi, S., Hari, S., & Hatzinakos, D. (2014). On establishing evaluation standards for ECG biometrics. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 3774–3778.

Powell, C., Carson, B. P., Dowd, K. P., & Donnelly, A. E. (2017). Simultaneous validation of five activity monitors for use in adult populations. *Scandinavian Journal of Medicine and Science in Sports*, *27*(12), 1881–1892.

Prajapati, S. K., Gage, W. H., Brooks, D., Black, S. E., & McIlroy, W. E. (2011). A novel approach to ambulatory monitoring: Investigation into the quantity and control of everyday walking in patients with subacute stroke. *Neurorehabilitation and Neural Repair*, *25*(1), 6–14.

Rejeski, W. J., Marsh, A. P., Brubaker, P. H., Buman, M., Fielding, R. A., Hire, D., … Miller, M. E. (2016). Analysis and interpretation of accelerometry data in older adults: The LIFE study. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, *71*(4), 521–528.

Romero-Ugalde, H. M., Garnotel, M., Doron, M., Jallon, P., Charpentier, G., Franc, S., … Bonnet, S. (2017). An original piecewise model for computing energy expenditure from accelerometer and heart rate signals. *Physiological Measurement*, *38*(8), 1599–1615.

Roos, M. A., Rudolph, K. S., & Reisman, D. S. (2012). The Structure of Walking Activity in People After Stroke Compared With Older Adults Without Disability: A Cross-Sectional Study, *92*(September), 1141–1147.

Rotstein, A., Inbar, O., Berginsky, T., & Meckel, Y. (2005). Preferred transition speed between walking and running: Effects of training status. *Medicine and Science in Sports and Exercise*, *37*(11), 1864–1870.

Rowlands, A. V. (2007). Accelerometer assessment of physical activity in children: an update. *Pediatr Exerc Sci*, *19*(3), 252-266.

Rowlands, A. V., Stone, M. R., & Eston, R. G. (2007). Influence of speed and step frequency during walking and running on motion sensor output. *Medicine and Science in Sports and Exercise*, *39*(4), 716–727.

Rowlands, A. V., Yates, T., Davies, M., Khunti, K., & Edwardson, C. L. (2016). Raw Accelerometer Data Analysis with GGIR R-package: Does Accelerometer Brand Matter? *Medicine and Science in Sports and Exercise*, *48*(10), 1935–1941.

Ryan, D. J., Wullems, J. A., Stebbings, G. K., Morse, C. I., Stewart, C. E., & Onambele-Pearson, G. L. (2018). Reliability and validity of the international physical activity questionnaire compared to calibrated accelerometer cut-off points in the quantification of sedentary behaviour and physical activity in older adults. *PLoS ONE*, *13*(3), 1–17.

Sanders, G. J., Boddy, L. M., Sparks, S. A., Curry, W. B., Roe, B., Kaehne, A., & Fairclough, S. J. (2019). Evaluation of wrist and hip sedentary behaviour and moderate-to-vigorous physical activity raw acceleration cutpoints in older adults. *Journal of Sports Sciences*, *37*(11), 1270–1279.

Sasaki, J. E., Hickey, A., Staudenmayer, J., John, D., Kent, J. A., & Freedson, P. S. (2017). Performance of Activity Classification Algorithms in Free-Living Older Adults, *48*(5), 941–950.

Shephard, R. J., & Aoyagi, Y. (2012). Measurement of human energy expenditure, with particular reference to field studies: an historical perspective. *Eur J Appl Physiol, 112,* 2785-2815.

Simpson, L. A., Eng, J. J., Klassen, T. D., Lim, S. B., Louie, D. R., Parappilly, B., … Zbogar, D. (2015). Capturing step counts at slow walking speeds in older adults: Comparison of ankle and waist placement of measuring device. *Journal of Rehabilitation Medicine*, *47*(9), 830–835.

Skender, S., Ose, J., Chang-Claude, J., Paskow, M., Brühmann, B., Siegel, E. M., … Ulrich, C. M. (2016). Accelerometry and physical activity questionnaires - A systematic review. *BMC Public Health*, *16*(1), 1–10.

Sloot, L. H., van der Krogt, M. M., & Harlaar, J. (2014). Self-paced versus fixed speed treadmill walking. *Gait and Posture*, *39*(1), 478–484.

Song, J. L., & Hidler, J. (2008). Biomechanics of overground vs. treadmill walking in healthy individuals. *Journal of Applied Physiology*, *104*(3), 747–755.

Spurr, G. B., Prentice, A. M., Murgatroyd, G. R., Reina, J. C., & Christman, N. T. (1988). Energy expenditure from minute-by-minute heart-rate recording: comparison with indirect calorimetry. *American Journal of Clinical Nutrition*, *48*(3), 552-559.

Spurr, G. B. (1990). Physical activity and energy expenditure in undernutrition. *Progr Food Nutr Sci*, *14*(2-3), 139-192.

Storti, K. L., Pettee, K. K., Brach, J. S., Talkowski, J. B., Richardson, C. R., & Kriska, A. M.

(2008). Gait speed and step-count monitor accuracy in community-dwelling older adults. *Medicine and Science in Sports and Exercise*, *40*(1), 59–64.

Strath, Scott J., Swartz, A. M., Bassett, D. R. J., O'Brien, W. L., King, G. A., & Ainsworth, B. E. (2000). Evaluation of heart rate as a method for assessing moderate intensity physical activity. *Medicine & Science in Sports & Exercise, 32*(Supplement), S465–S470.

Strath, S. J., Bassett, J., Swartz, A. M., & Thompson, D. L. (2001). Simultaneous heart rate-motion sensor technique to estimate energy expenditure. *Medicine and Science in Sports and Exercise, 33*(12), 2118–2123.

Strath, Scott J., Bassett, D. R., Thompson, D. L., & Swartz, A. M. (2002). Validity of the simultaneous heart rate-motion sensor technique for measuring energy expenditure. *Medicine and Science in Sports and Exercise, 34*(5), 888–894.

Stolze, H., Kuhtz-Buschbeck, J. P., Mondwurf, C., Boczek-Funcke, A., Jöhnk, K., Deuschl, G., & Illert, M. (1997). Gait analysis during treadmill and overground locomotion in children and adults. *Electroencephalography and Clinical Neurophysiology - Electromyography and Motor Control*, *105*(6), 490–497.

Swain, D. P., & Leutholtz, B. C. (1996). Heart rate reserve is equivalent to % VO2 reserve, not to % VO2max. *Medicine & Science in Sports & Exercise*, *29*(3), 410–414.

Swain, D. P., Leutholtz, B. C., King, M. E., & Branch, D. J. (1998). Relationship between % heart rate reserve and % VO2 reserve in treadmill exercise. *Medicine*, *30*(2), 318–321.

Swift, D. L., Johannsen, N. M., Lavie, C. J., Earnest, C. P., & Church, T. S. (2014). The role of exercise and physical activity in weight loss and maintenance. *Progress in Cardiovascular Diseases*, *56*(4), 441–447.

Talor, C. R., Heglund, N. C., & Maloiy, G. M. (1982). Energetics and mechanics in terrestrial locomotion. 1. Metabolic energy consumption as a function of speed and body size in birds and mammals. *J Exper Bill, 97(*1), 1-23.

Tanaka, H., Monahan, K. D., & Seals, D. R. (2001). Age-predicted maximal heart rate revisited. *Journal of the American College of Cardiology, 37*(1), 153–156.

Terrier, P., & Schutz, P. (2003). Variability of gait patterns during unconstrained walking assessed by satellite positioning (GPS). *Eur J Appl Physiol, 90*(5-6), 554-561

Tilson, J. K., Sullivan, K. J., Cen, S. Y., Rose, D. K., Koradia, C. H., Azen, S. P., & Duncan, P. W. (2010). During the First 60 Days Poststroke: Minimal Clinically Important Difference.

Physical Therapy, *90*(2), 196-208.

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

Trost, S. G., Mciver, K. L., & Pate, R. R. (2005). Conducting accelerometer-based activity assessments in field-based research. *Medicine and Science in Sports and Exercise*, *37*(11 SUPPL.), 531–543.

Trost, S. G., Ward, D. S., Moorehead, S. M., Watson, P. D., Riner, W., & Burke, J. R. (1998). Validity of the computer science and applications (CSA) activity monitor in children. *Medicine & Science in Sports & Exercise*, *30*(4), 629–633.

Tudor-Locke, C., Aguiar, E. J., Han, H., Ducharme, S. W., Schuna, J. M., Barreira, T. V., … Staudenmayer, J. (2019). Walking cadence (steps/min) and intensity in 21-40 year olds: CADENCE-adults. *International Journal of Behavioral Nutrition and Physical Activity*, *16*(1), 1–11.

Tudor-Locke, C., & Rowe, D. A. (2012). Using cadence to study free-living ambulatory behaviour. *Sports Medicine*, *42*(5), 381–398.

Turcot, K., Allet, L., Hoffmeyer, P., & Armand, S. (2009). Investigation of standing balance in diabetic patients with and without peripheral neuropathy using accelerometers. *Clin Biomech*, *24*(9), 716-721.

Van Blarigan, E. L., Kenfield, S. A., Tantum, L., Cadmus-Bertram, L. A., Carroll, P. R., & Chan, J. M. (2017). The Fitbit One Physical Activity Tracker in Men With Prostate Cancer: Validation Study. *JMIR Cancer*, *3*(1), e5.

Villars, C., Bergouignan, A., Dugas, J., Antoun, E., Schoeller, D. A., Roth, H., … Simon, C. (2012). Validity of combining heart rate and uniaxial acceleration to measure free-living physical activity energy expenditure in young men. *Journal of Applied Physiology*, *113*(11), 1763–1771.

Warren, J. M., Ekelund, U., Besson, H., Mezzani, A., Geladas, N., & Vanhees, L. (2010). Assessment of physical activity - A review of methodologies with reference to epidemiological research: A report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. *European Journal of Cardiovascular Prevention and Rehabilitation*, *17*(2), 127–139.

Waters, R. L., Lunsford, B. R., Perry, J., & Byrd, R. (1988). Energy-speed relationship of walking: Standard tables. *Journal of Orthopaedic Research*, *6*(2), 215–222.

Welk, G. J. (2002). Physical Activity Assessment for Health-Related Research. Human Kinetics Publishers, Inc.

Westerterp, K. R. (2001). Limits to sustainable human metabolic rate. *Journal of Experimental Biology*, *204*(18), 3183–3187.

Westerterp, K. R. (2009). Assessment of physical activity: a critical appraisal. *Eur J Appl Physiol*, *105*(6), 823-828.

Whyte, G. P., George, K., Shave, R., Middleton, N., & Nevill, A. M. (2008). Training induced changes in maximum heart rate. *International Journal of Sports Medicine*, *29*(2), 129–133.

Winter, D. (2009) Biomechanics and Motor Control of Human Movement (4th ed.). John Wiley & Sons, Inc.

World Health Organization (n.d.). Background: The global burden of chronic disease. Retrieved from https://www.who.int/nutrition/topics/2_background/en/.

Zakynthinaki, M. S. (2015). Modelling heart rate kinetics. *PLoS ONE*, *10*(4), 1–26.