THE DEVELOPMENT AND CROSS-VALIDATION OF RAW ACCELEROMETER SEDENTARY AND PHYSICAL ACTIVITY THRESHOLDS IN UNIVERSITY STUDENTS

Ruth Brady

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the Degree of Master of Philosophy

October 2019

Contents

List of Tables

List of Figures

Figure 13: Dominant wrist placement cross validation results from 95% equivalence testing between thresholds and criterion reference in Step 3.

List of Abbreviations

AUC:	Area Under the Curve
ENMO:	Euclidean Norm Minus One
LPA:	Light intensity physical activity
MAPE:	Mean absolute percent error
MPA:	Moderate intensity physical activity
MVPA:	Moderate to vigorous intensity physical activity
PA:	Physical Activity
ROC Curve:	Receiver Operating Characteristic Curve
SB:	Sedentary Behaviour
UK:	United Kingdom
VPA:	Vigorous intensity physical activity

Abstract

Background: Regular participation in physical activity (PA) has been associated with the primary prevention of 25 chronic medical conditions. University is considered as a major period of transition as it is associated with full independence from parents for the first time. The prevalence of inactivity in university students is reported to be as high as 60%, though few studies have focussed on university students or used device-based methods to assess physical activity. Previously, researchers in the discipline have used device specific, proprietary, dimensionless units called counts to process and report accelerometer data. Researchers can now process raw acceleration signals rather than rely on proprietary counts. This is advantageous as these transparent methods improve comparability across different accelerometers and studies. Currently there is a lack of raw accelerometer thresholds for use in university students. **Aim:** The primary aim of this study was to calculate and cross-validate accelerometer thresholds to classify PA and SB in university students. **Methods:** Thirty-five university students enrolled during the 2018/2019 academic year at Liverpool John Moores University with a mean age of 21.4 years (*n*=21 females) completed a circuit of 12 activities in laboratory conditions. Each participant wore 3 ActiGraph GT9X accelerometers (both wrists and hip). Thresholds were generated using Receiver Operating Characteristic (ROC) curve analysis in a calibration group (*n*= 21) using indirect calorimetry as the criterion reference. Sensitivity, specificity, percentage agreement, mean absolute percent error (MAPE), and Cohens' Kappa coefficients were investigated. These thresholds were then fine-turned using equivalency analysis. These resultant thresholds were then cross-validated in an independent sample (12 participants, *n*=7 females). Once again, sensitivity, specificity, MAPE, percentage agreement and Cohens' Kappa coefficients were investigated. **Results:** The final SB thresholds range from <8 (hip) to <40 mg (dominant) (sensitivity: 0.76 (non-dominant) to 0.84 (hip), specificity: 0.81 (hip) to 0.92 (non-dominant), MAPE: 9.5 (dominant) to 20.7 (hip), percentage agreement: 82.2 (hip) to 87.2 (dominant), Cohens' Kappa coefficients: 0.61 (hip) to 0.70 (dominant)). LPA ≥8 to ≥40 mg (sensitivity: 0.25 (hip) to 0.39 (dominant),

specificity: 0.82 (dominant) to 0.88 (hip), MAPE: 9.8 (dominant) to 22.7 (hip), percentage agreement: 71.6 (non-dominant) to 72.8 (hip), Cohens' Kappa coefficients: 0.15 (hip) to 0.21 (non-dominant and dominant)). MPA ≥50 to ≥110 mg (sensitivity: 0.45 (dominant) to 0.68 (hip), specificity: 0.76 (non-dominant) to 0.83 (hip), MAPE: 6.8 (hip) to 12.7 (non-dominant), percentage agreement: 69.6 (non-dominant) to 78.8 (hip), Cohens' Kappa coefficients: 0.24 (dominant) to 0.49 (hip)). VPA ≥225 to ≥315 mg (sensitivity: 0.64 (hip) to 0.72 (dominant), specificity: 0.92 (dominant) to 0.95 (hip), MAPE: 11.3 (hip) to 12.1 (non-dominant), percentage agreement: 88.5 (dominant) to 90.1 (hip), Cohens' Kappa coefficients: 0.60 non-dominant and dominant) to 0.62 (hip)). MVPA ≥50 to ≥110 mg (sensitivity: 0.74 (dominant) to 0.84 (hip), specificity: 0.74 (non-dominant) to 0.85 (hip), MAPE: 8.8 (hip) to 13.6 (non-dominant), percentage agreement: 75.0 (dominant) to 84.6 (hip), Cohens' Kappa coefficients: 0.50 (non-dominant) to 0.69 (hip)). In comparison to the criterion reference, for the non-dominant wrist, the SB threshold provided equivalent estimates at ±20%. For the hip placement, MVPA estimates were equivalent at ±15% and MPA was equivalent at ±20%. For the dominant wrist placement, SB and MPA were equivalent at ±15%. LPA and MVPA estimates were equivalent at ±20%. VPA estimates were not statistically equivalent in comparison to the criterion for any placement. **Conclusion:** The thresholds for SB, LPA, MPA and MVPA showed acceptable levels of agreement between the accelerometer and criterion reference in regards to specificity, MAPE and percentage agreement. VPA was the only threshold to show no equivalency between accelerometer and criterion reference (Metamax) on any placement site. The hip placement generally provided better agreement between the criterion reference and accelerometer and therefore could be considered the optimum placement to provide accurate estimates of SB and PA levels of university students. The thresholds generated in this study can be used to help researchers estimate the habitual activity levels of university students. Future research should aim to validate these thresholds in a free-living or simulated free-living situations to examine their performance in an ecologically valid setting.

Declaration

I declare that the work within this thesis is entirely my own.

Acknowledgements

I would like to thank the following people for their contribution over the last year. Your support and guidance have been crucial in carrying out this research.

Firstly, Dr Lynne Boddy, my director of studies. Your support and guidance during the project have resulted in the completion of this. Thank you especially for all the time you have put into this project and your feedback is always greatly received. Thank you for also moulding me into the researcher I am today.

My supervisory team, Dr Lee Graves and Dr Lawrence Foweather for the continued support. Your input and feedback is greatly appreciated. It has been a pleasure to work with you both. Thank you to all my supervisors for seeing my potential and giving me this fantastic opportunity.

I would like to express my thanks to Gemma and Dean. Thank you for all your training and laboratory support over the last year.

I would like to thank all the participants who took part in this study. Their cooperation was instrumental in the completion of this.

I would like to say a massive thank you to my family and friends who have emotionally supported me through this journey. I hope I have made you proud!

Finally, I would like to thank the amazing post-graduate community who I am lucky to now call friends (Physical Activity Exchange PGRs). Thank you for being so welcoming and guiding me throughout the last year. The bake off was the highlight of every week and I hope lunch club continues forever. Thank you for the distractions, telling me to take everything one-step at a time and pushing me to believe in myself when I needed it the most. I would like to give a little shout out to Matteo, thank you for everything you have shown and helped me with this year. Thank you for answering all my questions and importantly showing me how to use GGIR.

Chapter 1: Introduction

This chapter provides a brief introduction to the programme of research. A more comprehensive overview of the literature and topic is provided in Chapter 2: Literature Review.

Definitions

For the purposes of this thesis, physical activity (PA) is defined as “any bodily movement produced by contraction of skeletal muscle that substantially increases energy expenditure” (Howley, 2001, p.364). Sedentary behaviour (SB) is defined as “any waking behaviour characterised by an energy expenditure <1.5 MET while in a seated, lying or reclined position” (Tremblay et al., 2017, p.2).

SB, PA and health outcomes

Physical inactivity has been identified as the fourth leading risk factor for global mortality (Kohl et al., 2012; Newtonraj et al., 2017). It is associated with many chronic diseases, non-communicable diseases and premature mortality (Wilmot et al., 2012; Alves et al., 2016; Peterson et al., 2018). These non-communicable diseases include diabetes, obesity and several types of cancers (Alves et al., 2016). However, regular participation in PA has been associated with the primary prevention of 25 chronic medical conditions (Rhodes et al., 2017). Sedentary behaviour is associated with cardiovascular disease and cancer mortality, and these risks are independent of PA (Ekelund et al., 2019a). Physical inactivity reportedly costs $53.8 billion worldwide, of which the UK costs are approximately £900 million (Ding et al., 2016). Therefore, physical inactivity and sedentary behaviour are key public health concerns in contemporary society.

University and PA levels

University is considered a major period of transition within the life course as it is associated with moving from the parental home to full residential independence (Rouse and Biddle, 2010; Deliens et al., 2015). The number of individuals enrolled in higher education during the 2017/18 academic year in the UK was 2.3 million (HESA, 2019). PA levels decline in adults aged between 18-24 years old and this occurs at the same time as many young people attend higher educational institutions (Gibson et al., 2016). There is a lack of consensus in the literature with regards to how active/inactive university students are and how much time they spend sedentary (Clemente et al., 2016; Calestine et al., 2017). This could be down to the differing data collection methods (self-report or device-based) used within this population. Similar to within the general adult population, much of the previous research conducted with university students used self-report measures, which may result in individuals to over-estimating the amount of PA achieved (Gibson et al., 2016) due recall errors and social desirability bias (Sedgwick, 2014; Arias-Palencia et al., 2015; Ekelund et al., 2019b). The reported proportion of students meeting PA guidelines ranges from 0.5% (Clemente et al., 2016) to 80% (Rouse and Biddle, 2010) depending on the research study and methods employed, therefore there is a lack of consensus related to the activity levels of university students.

PA and SB measurement methods

There are several ways in which PA and SB can be measured or estimated. Each PA and SB measure varies in simplicity, precision and information gathered. PA has often been estimated using self-report methods such as questionnaires and activity logs (e.g International Physical Activity Questionnaire and Global Physical Activity Questionnaire (Chastin, Culhane and Dall, 2014; Shim, Oh and Kim, 2014). Device-based PA assessment tools include accelerometers, pedometers and heart rate monitors, which do not require participants to self-report their activity behaviours and therefore are considered

to be less open to social desirability bias (Sedgwick, 2014; Arias-Palencia et al., 2015; Ekelund et al., 2019b). Accelerometers are small, device-based methods to collect PA data that are used in both free-living and laboratory conditions (de Almeida Mendes et al., 2018). ActiGraph accelerometers are commercially available and are the most frequently used. ActiGraph accelerometers account for approximately 50% of studies published to date (Migueles et al., 2019a). ActiGraph acelerometers (various models) has shown high intrainstrument and interinstrument reliability (Montoye et al., 2016). Accelerometers can be attached to a variety of bodily locations, including the wrist, hip, thigh, ankle and chest (Aadland and Anderssen, 2012; Kamada et al., 2016). Some researchers have advocated the use of the wrist placement in recent years due to convenience for participants and comfort and improvements in wear compliance (Tudor-Locke et al., 2015). It has however, been reported that wrist worn accelerometers may over-estimate overall energy expenditure when compared to the hip or waist due to arm movements which do not involve whole body movement (Ellis et al., 2014). Despite the disadvantages of accelerometers being placed on the wrist, this placement has become more popular due to reported superior compliance (Ellis et al., 2014; Wolpern et al., 2019).

Raw data analysis

Previously, researchers used only proprietary “activity counts” to process the acceleration signals collected from accelerometers. Activity counts are device specific, proprietary, dimensionless units, which compress the acceleration signals into units over a user specified time period, known as an epoch (Innerd, Harrison and Coulson, 2018; Sanders et al., 2019). Due to device storage capacities, lack of user-friendly software and lack of computer processing abilities, raw acceleration signal analysis was not possible (Troiano et al., 2014). Advances in accelerometer technology, computer processing and the availability of open-source software, it is now possible to access and process the raw acceleration signals from three of the most commonly used accelerometers (GENEActiv, Axivity and ActiGraph GT3X+, GT9X). This means that once the raw acceleration data are

collected, raw data analysis and processing is subject to researcher-driven data decisions, rather than proprietary methods (Fairclough et al., 2016). With greater transparency of processing methods between studies, it has potential for comparability across different studies and devices. As more researchers use raw data processing and analysis, there is an increased need for thresholds (at the wrist and hip) using this processing and analysis technique (Fairclough et al., 2016). These thresholds can be developed using calibration studies, where participants complete daily activities, specific to their population.

This chapter has provided a brief introduction to the programme of research. The next chapter (2: Literature review) will provide a concise overview of the literature and topic. It also outlines the aims and objectives of the research programme.

Chapter 2: Literature Review

This chapter provides a concise overview of the literature and topic. It also outlines the aims and objectives of the research programme.

Physical Activity and Sedentary Behaviour

Physical activity (PA) is defined as “any bodily movement produced by contraction of skeletal muscle that substantially increases energy expenditure” (Howley, 2001, p.364). PA is divided into four intensity categories based on Metabolic Equivalent of Task (METs) thresholds: light intensity PA (1.5-2.9 METs, LPA), moderate intensity PA (3.0–5.9 METs, MPA), vigorous intensity PA (≥6 METs, VPA) and moderate to vigorous intensity PA (≥3 METs, MVPA) (de Almeida Mendes et al., 2018). Sedentary behaviour (SB) is defined as “any waking behaviour characterised by an energy expenditure <1.5 MET while in a seated, lying or reclined position” (Tremblay et al., 2017, p.2). Previously, SB has been confused and used interchangeably with the term ‘physical inactivity’ which describes individuals who do not meet government PA guidelines. To help with terminology, individuals who do not meet PA guidelines are referred to as “inactive” rather than sedentary. An individual can be both highly sedentary and considered active because they meet the PA guidelines. A new term has been developed to describe this: “active coach potatoes” (Lepp and Barkley, 2019, p.2).

PA Guidelines

In 2019, revised PA guidelines were released (CMO, 2019). It is now recommended adults should be physically active every day to promote good mental and physical health and any PA is better than none. Adults should do activities to develop or maintain strength in the major muscle groups; these could include heavy gardening, carrying heavy shopping, or resistance based exercises. Muscle strengthening activities should be done at least two days a week, but any strengthening activity is better than none. Each week,

adults should accumulate at least 150 minutes (2 1/2 hours) of MPA (such as brisk walking or cycling); or 75 minutes of VPA (such as running); or even shorter durations of very VPA (such as sprinting or stair climbing); or a combination of moderate, vigorous and very vigorous intensity activities. Adults should aim to minimise the amount of time spent being sedentary, and when physically possible should break up long periods of inactivity with a minimum of LPA (CMO, 2019). These guidelines are based on the available evidence that documents the associations between PA, SB and health outcomes.

PA, SB and Economic burden

Physical inactivity is positively associated with many chronic conditions but also have substantial economic burden (Ekelund et al., 2019a). Physical inactivity reportedly costs $53.8 billion worldwide, of which the UK costs are approximately £900 million (Ding et al., 2016). However, this figure represents the direct costs of physical inactivity; the indirect costs of physical inactivity could make this figure considerably higher. In the UK, £455 million (0.3%) of health care costs are related to the economic burden of physical inactivity (Ding et al., 2016). This has decreased since 2006/07 where the estimated cost of PA was £0.9 billion (Scarborough et al., 2011). It was reported that 11.6% of all-cause mortality deaths are related to SB and direct costs to the NHS were between £0.7-0.8 billion in 2016/17 (Heron et al., 2019). Getting individuals more active and less sedentary will decrease this economic burden on health services such as the NHS. This figure will only increase if individuals continue to live a highly sedentary lifestyle and do not become more active.

PA, SB and physical health outcomes

PA and SB are associated with health outcomes, which can affect an individual's overall health. It should be noted that PA is negatively associated with disease and premature

mortality, whereas SB is positively associated with these outcomes (Nicolson, Hayes and Darker, 2019). Physical inactivity has been identified as the fourth leading risk factor for global mortality (Kohl et al., 2012; Newtonraj et al., 2017). Physical inactivity is associated with many chronic diseases, non-communicable diseases and premature mortality (Wilmot et al., 2012; Alves et al., 2016; Peterson et al., 2018). These non-communicable diseases include diabetes, obesity and several types of cancers (Alves et al., 2016). Physical inactivity causes more than 5 million deaths per year and 9% of these are categorised as premature deaths (Ding et al., 2016). Physical inactivity is currently responsible for as many deaths as smoking and 6% of the total deaths worldwide (Alves et al., 2016). However, regular participation in PA has been associated with the primary prevention of 25 chronic medical conditions (Rhodes et al., 2017). It has been observed that those who are active (met the government guidelines) have a 31% lower risk of all-cause mortality compared to those who do not meet the guidelines (Rhodes et al., 2017). Individuals who walk for three or more hours per week, at a moderate pace, reduce the risk of a coronary event by 35% compare to those individuals who do not. However, individuals who become active later in life still receive health benefits of PA and have a lower risk of cardiovascular events (Alves et al., 2016).

SB and mortality have a positive dose-response relationship, meaning the more time spent sedentary (engaging in activities such as watching TV and sitting), increases the risk of mortality (Ekelund et al., 2019a). It should also be noted that SB and its relationship to all-cause mortality is independent of PA, as it has been shown that excessive amounts of TV viewing are related to all-cause mortality even when individuals have a high level of PA (Alves et al., 2016). It is reported that 30% of adults in the UK are sedentary for at least 6 hours a day and this is associated with an increased risk of all-cause mortality and cardiovascular mortality (Heron et al., 2019). In a recent study by Ekelund et al. (2019a), the risk of cardiovascular disease was 32% higher for those individuals who were sedentary for longer than 8 hours a day (Ekelund et al., 2019a). When investigating TV

viewing time as a sub-category of SB, cardiovascular disease risk was higher for those who watched TV for over 5 hours a day (Ekelund et al., 2019a). However, these risks were eliminated in individuals who were classified as the most active (>35.5MET-hour/week) (Ekelund et al., 2019a).

The dose-response relationship between LPA and health outcomes is unclear mainly because many studies and self-report tools do not report LPA (Ekelund et al., 2019b). However, with the shift towards device-based measures in PA research (e.g. accelerometers) this relationship can be investigated. Ekelund et al. (2019b) conducted a meta-analysis that investigated dose-response associations between accelerometry measured PA, sedentary time and all-cause mortality. The meta-analysis was categorised into exposure variables, with the first quarter as reference (the least active individuals). The analysis found that having a high sedentary lifestyle (7.5 hours is the reference value; most sedentary) combined with high MVPA time (4^{th} quarter (24 minutes a day); most active) can benefit health (0.52 min/day; when sedentary time is adjusted for). The largest observed mortality risk reduction occurred in those individuals who accumulated 375 minutes a day of LPA (when investigating LPA only; with no other variables controlled for). Individuals who spent ≥9.5 hours sedentary time exhibited a significantly higher risk of death. This study demonstrated the protective effect of PA and the detrimental effect of SB in an adult population when measured with a device.

University Students

University is a transitional stage for individuals (Deliens et al., 2015) and the amount of students enrolled in higher education during the 2017/18 academic year in the UK was 2.3 million (HESA, 2019). In England, 1.2 million individuals were enrolled for the start of their degree on a full-time mode of study with 56% of these being female and 79% of the overall university population in England being ≤20 years old (HESA, 2019). University is

considered a major period of transition as it is associated with moving from the parental home to full residential independence (Rouse and Biddle, 2010). In 2016, almost half a million individuals moved away from home to attend their first year of university, with 55.8% moving up to 55 miles away from the family home (Donnelly and Gamsu, 2018). PA levels decline in adults aged between 18-24 years old and this occurs at the same time as many young people attend higher educational institutions (Gibson et al., 2016). However, those who graduate from university are more active (meeting the guidelines) than those with a high school education (University: men: 54.6% and women: 53.3% versus High school: men: 37.2% and women: 37.1%). However, the reason PA is adhered better in university graduates is unclear (Towne et al., 2017). Higher education is a period characterised by change, uncertainty and adjustment, which can have a negative impact on health and well-being (Gibson et al., 2016). The term “unhealthy behaviours” can refer to the following; smoking, drug use, alcohol consumption, sedentary lifestyles and malnutrition (Lazzeri et al., 2014; Farhud, 2015). These unhealthy behaviours such as unhealthy diet and reduced PA levels/increased sedentary lifestyles could be due to the progression from controlled environments (“family home” or “school”) to more autonomous and self-motivated university environments (Gibson et al., 2016). University students are at a stage where they are more responsible and autonomous in the decisions they make compared to during childhood and adolescence (Maselli et al., 2018). As such, it is important to understand why behaviours change in this sub-group as it represents a large proportion of young adults (Rouse and Biddle, 2010)

SB and PA in university students

There is a lack of consensus in the literature with regards to how active/inactive university students are and how much time they spend sedentary. However, this could be due to the varied data collection methods used to assess PA and SB. Similar to within the general adult population, much of the previous research conducted with university students has

used self-report measures, which may result in individuals over-estimating or underestimating the amount of PA and SB engaged in (Gibson et al., 2016). Self-report measures are prone to recall errors and social desirability bias (Sedgwick, 2014; Arias-Palencia et al., 2015; Ekelund et al., 2019a) which influence the accuracy of the reported PA or SB levels. There are a few studies which have used device-based methods to examine PA and SB levels within a university student population. One of these was conducted by Arigo, Pasko and Mogle (2019) who reported that 40% of American students met the guidelines (the guidelines were a minimum of 5 x 30 mins/day of MPA or 3 x 20 mins/day of VPA or a combination of MPA and VPA) (ACSM, 2013). Males were more active than females, with 42% of males reaching the MVPA guidelines compared to 23% of females. On average individuals spent 26.4 minutes/day engaged in MVPA. Similarly to what is reported in child and other adult populations, females were reported to have a lower PA levels than their male counterparts (Fagaras, Radu and Vanvu, 2015; Arigo, Pasko and Mogle, 2019). The Arigo, Pasko and Mogle (2019) study is in contrast to the results of Clemente et al. (2016) who reported that only 5.4% of Portuguese students met the MVPA or VPA guidelines (measured using an accelerometer). However, as the guidelines at the time stipulated that PA should be completed in bouts of 10 minutes or more to benefit health, when this was taken into account an even smaller percentage of students (0.5%) met these guidelines. Rouse and Biddle (2010) reported a considerably higher figure (80%) meeting the PA guidelines. However, the data were collected using a self-report tool (Ecological Momentary Assessment Tool; a diary requesting what behaviour was occurring every 15 minutes). Another study that used a self-report measure (IPAQ) reported similar results as Rouse and Biddle (2010) with 76% of English students meeting the guidelines (Calestine et al., 2017). This study also reported sedentary time, which was estimated to be 3 hours a day, which is in contrast to a study conducted by Clemente et al. (2016), who used device-based methods and reported that university students spent 12 hours per weekday sedentary. Other self-reported evidence, for example the study conducted by Calestine et al. (2017), reported that students were sedentary for only 3 hours/day. Clemente et al. (2016) also investigated the difference

between weekend and weekday SB and PA and differences between genders in a population of university students (aged 18-23 years). Findings reported higher SB on the weekend in comparison to weekdays (770.87 mins/day vs 751.05 mins/day; investigating the overall population), with females also spending significantly more time sedentary (5%) than their male counterparts. MPA and VPA differed between genders, with males reporting higher PA levels in comparison to females (males spending 65 minutes compared to females with 51 minutes in MPA per weekday). Overall, 59% of men and 35% of females met the guidelines, this figure did not control for bouts of ≥10 minutes. When previous studies take into account PA being conducted in bouts of ≥10 minutes, the adherence and amount of individuals reaching the guidelines drops dramatically. In general, it is difficult to interpret PA and SB levels within the student population due to the varied data collection methods used to measure PA and SB. It is also difficult to compare studies due to the different methods used. This is one reason standardisation of methods are needed.

PA and SB assessment methods

Measurement of both SB and PA is important for describing the prevalence of disease and evaluating the effectiveness of interventions. It is important also to describe the dose-response relationship because it allows researchers to see how much SB and intensities of PA can benefit health and which diseases are associated with different movement behaviours (Husu, Vaha-Ypya and Vasankari, 2016). There are several ways in which PA and SB can be estimated. Each PA and SB measure varies in simplicity, precision and information gathered. PA has often been estimated using self-report methods such as questionnaires and activity logs. Self-report measures are cost effective and require low participant burden, however, this type of data collection has limitations (Shim, Oh and Kim, 2014). The limitations of self-report tools are these are subjective, leading to the possible underestimation of SB and overestimation of PA due to recall errors and social desirability bias (Sedgwick, 2014; Arias-Palencia et al., 2015; Ekelund et al., 2019b). Self-

report PA measures, such as questionnaires, are highly prone to inaccuracy due to reliance on the ability of participants being able to recall what they have done (Freedson, Melanson and Sirard, 1998). This is especially apparent when recall periods are long, with previous studies in adults showing that individuals under-estimated sitting time by up to 4 hours when questionnaires are compared to an accelerometer (Clark et al., 2015). One specific type of subjective method (International Physical Activity Questionnaire; IPAQ) has been reported to under-estimate SB by between 2 to 3.5 hours. The IPAQ only has two questions relating to SB, which means it is difficult to capture the variability and total time of SB that occurs in different contexts (Ku et al. 2018). Questionnaires usually contain a set of specific close ended questions, which can impact the precision of the PA reported (Freedson, Melanson and Sirard, 1998). Daily activities may be harder to recall due to incidental activities being generally less structured and less memorable than structure PA. However, even though accelerometers overcome these limitations, accelerometers do not provide information on the context (e.g. leisure, occupation and transport) the PA is being completed in (Freedson, Melanson and Sirard, 1998; Romanzini et al., 2019). This may result in researchers combining both self-report and device-based methods to eliminate some of the limitations and provide the most representative set of PA data than either method can provide alone.

Device-based PA assessment tools include accelerometers, pedometers and heart rate monitors. Using device-based devices can reduce human error (reporting the incorrect amount of PA or omitting completely) and issues associated with recall bias (Ainsworth et al., 2011). The research field has moved towards the use of accelerometers as they estimate time spent in PA and many components of PA behaviours (frequency, intensity and time) but there remains no gold standard accelerometer, which can capture type and context of PA (Pfister et al., 2017). The choice of monitor relies on researcher decisions such as cost, the specific component of PA of interest and target population (Ainsworth et al., 2011).

Accelerometers and compliance

Accelerometers are small, device-based methods to collect PA data that are used in both free-living and laboratory conditions (de Almeida Mendes et al., 2018). As technology has advanced, so have accelerometers. Older models of accelerometers could only measure acceleration in a single plane of movement (McGarty, Penpraze and Melville, 2016). More recently, several models of accelerometers are able to measure accelerations on all three axis of movement: vertical (Y), horizontal right-left (X) and horizontal front back (Z) (de Almeida Mendes et al., 2018). ActiGraph accelerometers are the most frequently used of commercially available accelerometers and account for approximately 50% of studies published (Migueles et al., 2019a). ActiGraph acelerometers (various models) has shown high intrainstrument and interinstrument reliability (Montoye et al., 2016).

Accelerometers can be attached to a variety of bodily locations, including the wrist, hip, thigh, ankle and chest (Aadland and Anderssen, 2012; Kamada et al., 2016). Historically, the hip placement was the most popular, however, there has now been a shift towards placement on the wrist. Wrist worn accelerometers are increasing in popularity to attempt to increase wear compliance for two main reasons: convenience for participants and comfort as they are small and unnoticeable compared to hip worn accelerometers (Tudor-Locke et al., 2015; White et al., 2019). Issues such as compliance have been noted as a reason for the change and wrist placements have the ability to be worn 24 hours a day which provides an opportunity to examine 24hr movement profiles and assess sleep (Jungquist et al., 2015). The wrist placement site has become more popular especially as large population surveillance studies such as National Health and Nutrition Examination Survey (NHANES) demonstrated improved compliance to monitoring protocols (Matthews et al., 2012). However, compliance can vary depending on the age of participants (Matthews et al., 2012). Compliance has been an issue in accelerometer data collection for a few different reasons (e.g. refusal to wear an accelerometer for the required length of time or removal of devices). Having greater compliance is important as this allows the device to better capture habitual behaviours and provides the most reliable estimates of

PA levels (Lee, Macfarlane and Lam, 2013). Despite the advantages gained by increasing participant compliance, it has been reported that wrist worn accelerometers may over-estimate overall energy expenditure when compared to the hip or waist due to additional arm movements that are not associated with ambulation or whole body movements (Ellis et al., 2014). Wrist worn accelerometers are also suggested to have limited ability to differentiate between postural states, which is important when considering SB assessment (Duncan et al., 2018). However, despite the disadvantages of wrist worn accelerometer placements; researchers are still using this placement site due to the superior compliance (Ellis et al., 2014; Wolpern et al., 2019), which will reduce the risk of misclassification and selection bias due to the exclusion of participants who do not wear the monitor for the required hours per day or days per week (Rowlands et al., 2018a).

Activity Counts and Raw data analysis

Previously, researchers used “activity counts” to analyse the data collected from accelerometers. Activity counts are device specific, proprietary, dimensionless units, which compress the acceleration signals into units over a user specified time period, known as an epoch (Innerd, Harrison and Coulson, 2018; Sanders et al., 2019). Activity counts provide an overview of movement, but a fundamental challenge is to determine how activity counts equate to more meaningful indicators, such as energy expenditure or time spent in MVPA (Welk, 2007). It is challenging to compare between and across different monitors when using activity counts due to the differences in how the acceleration data is collected, processed, filtered and scaled (Migueles et al., 2017). Previously, raw acceleration signal analysis was not possible due to device storage capacities, lack of user-friendly software and lack of computer processing abilities (Troiano et al., 2014). Now this is possible due to some accelerometers’ (e.g. GENEActiv and ActiGraph GT3X+ and GT9X) ability to capture raw acceleration signals and make these available for researchers to process. These accelerometers (e.g. GENEActiv and ActiGraph GT3X+ and GT9X) have been designed to be worn on both the wrist and hip

and can record raw accelerations up to 100Hz (Fairclough et al., 2016; Rowlands et al., 2016). Saving and processing accelerometer signals for raw data processing removes the proprietary nature of activity counts and enables more transparency and the use of replicable methods (Rowlands et al., 2016). Once the data are collected, raw data analysis is subject to researcher-driven data decisions, rather than proprietary methods (Fairclough et al., 2016). There is an increase in comparability between studies as data processing methods can be the same. Raw data analysis adds a new layer of complexity to the evaluation of device-based research but has clear advantages with respect to data standardisation and harmonisation (Welk et al 2017). Raw acceleration signals can be analysed using open source packages and a range of methods exist (packages such as R) (Innerd, Harrison and Coulson, 2018). GGIR (Migueles et al., 2019b) is one software package that facilitates the processing and analysis of data from three most used commonly accelerometer sensors (ActiGraph, GENEActiv, Axivity). Data harmonisation would facilitate a change in the ability to compare prevalence or levels of activity/inactivity across populations, quantify dose-response associations between activity and health and identify the factors that impact on these associations (Rowlands et al., 2018b). The availability of efficient raw signal data analytic approaches will ultimately encourage researchers towards new models of accelerometer data analysis (Troiano et al., 2014).

R software and GGIR

R is free to download and GGIR is an open source R software package (Rowlands et al., 2016; Migueles et al., 2019b). GGIR comes with core functionalities: load data, extract signal metrics and detection of non-wear time and detection of sleep periods. Part 1 in GGIR searches for the data and detects from the file which accelerometer has been used and then uses this to investigate the calibration error. In calibration studies, only part 1 is used (Migueles et al., 2019a). The signal processing includes automatic calibration, detection of sustained abnormally high values, detection of non-wear and calculation of the average magnitude of dynamic acceleration (Euclidean Norm Minus One, ENMO).

GGIR is suitable for a wide range of research applications and unlike other computer science software, can be operated without programming expertise. A positive of this is it allows researchers to adapt GGIR for their specific needs. GGIR also facilitates a reproducible analysis of raw data which is needed in the comparison of published studies (Migueles et al., 2019a). This transparency, in theory, enables comparisons between acceleration data, regardless of monitor brand (Hildebrand et al., 2014). Raw data analysis has clear advantages with respect to data standardisation and harmonisation (Welk et al 2017). Data harmonisation would facilitate a change in the ability to compare prevalence or levels of activity/inactivity across populations, quantify dose-response associations between activity and health and identify the factors that impact on these associations (Rowlands et al., 2018b). The availability of efficient raw signal data analytic approaches will also ultimately encourage researchers towards new models of accelerometer data analysis (Troiano et al., 2014).

A range of outcome variables to describe the activity profile, MVPA and sleep can also be calculated (Rowlands et al., 2016). Autocalibration is an important step during signal processing as the ENMO statistic is vulnerable to calibration error due to the inherent assumption that gravity is measured as 1 g (van Hees et al., 2014; Rowlands et al., 2016). However, autocalibration can only occur in studies where the length of monitoring time is >24 hours (van Hees et al., 2014). Periods shorter than this may present slight calibration errors (Montoye et al. 2018), though one study protocol that was >24 hours used derived coefficient from free-living situations matched to the laboratory conditions, in cases were the laboratory data did not hold enough non-movement periods to facilitate direct auto-calibration (Hildebrand et al., 2017).

Determination of transparent, replicated steps to improve raw data comparability across different accelerometers are important because they offer the best potential for pooling

raw data from multiple studies, comparison across studies and developing highly accurate prediction models across different populations and a variety of accelerometer brands (Montoye et al., 2016). Pooling data from studies is more viable now than ever before because of the transparency raw data analysis has provided (Rowlands et al., 2016). Therefore, raw data signal processing provides substantial advantages over traditional activity counts-based methods.

Calibration studies

It is becoming increasingly important to ascertain precise and accurate PA measurement methods (e.g. accelerometers) to assess the impact of physical inactivity, SB and associated health outcomes. Calibration studies can help to achieve this (Bassett, Rowlands and Trost, 2012; Arvidsson, Fridolfsson and Borjesson, 2019). Calibration studies allow data to be scaled or adjusted to produce more accurate and usable estimates (Saint-Maurice et al., 2014). Such studies and the generated thresholds or cutpoints help researchers estimate more precisely the activity/inactivity levels of the population of interest. This in turn helps researchers understand the prevalence in each population and the associated health outcomes (Saint-Maurice et al., 2014).

There are two types of calibration in reference to wearable activity monitors, namely unit and value. Unit calibration is performed to reduce inter-instrument variability and to ensure individual activity monitors are correctly measuring the direct signals (e.g. accelerations). Value calibration of wearable monitors refers to the process used to convert the direct signals into other established measurement units (Bassett, Rowlands and Trost, 2012). Value calibration is performed to ensure that a wearable monitor gives the intended values for outcome variables. This process involves collecting data on multiple individuals as they perform different activities and simultaneously collecting criterion data (Bassett, Rowlands and Trost, 2012). Accelerometer calibration is age- and population-specific because of maturation and between group differences such as cardiorespiratory fitness

and differences in movement behaviours (McGarty, Penpraze and Melville, 2016). Therefore, generalised thresholds applied to all populations introduce systematic measurement errors and reduce the validity of the results (McGarty, Penpraze and Melville, 2016). Population specific thresholds help to better quantify PA as the thresholds generated in this population are applied to individuals with similar characteristics and will have mirrored typical daily activities (Roscoe, James and Duncan, 2017).

Calibration studies commonly occur in laboratory conditions due to the difficulty of using a criterion measure such as energy expenditure in the field (Rowlands and Eston, 2007; Hills, Mokhtar and Byrne, 2014). In early laboratory calibration studies, the participants would walk at progressively increasing speeds on a treadmill and this would be used to calculate accelerometer thresholds (Troiano et al., 2014). Studies that only used treadmill-based activities were questioned due to the limited range of activities and the lack of lifestyle or population specific activities involved, thus not reflecting 'real-life' behaviours (Arvidsson et al., 2019). It has been reported in previous calibration studies that the protocols that use ambulatory movements only, produce considerably lower estimates of MVPA compared to calibration studies using a variety of daily activities and then how they performed in free-living situations (Welk et al., 2019). However, more recent calibration studies have addressed limitations of previous studies and it is now common practice that the activities should mirror daily tasks which the population of interest typically engages in (Matthew, 2005; Migueles et al., 2017). The activities should range from SB to VPA to develop thresholds which provide the optimum estimates of PA (Matthew, 2005; Bassett, Rowlands and Trost, 2012). It is also important to conduct cross-validation against an independent group to reduce bias and increase the validity of the results (McGarty, Penpraze and Melville, 2016). In laboratory calibration studies, indirect calorimetry is one of the most commonly used criterion measures in adults (de Almeida Mendes et al., 2018). This method measures respiratory gas exchange (oxygen uptake and carbon dioxide production) to allow for the calculation of energy expenditure (Bassett, Rowlands

and Trost, 2012). There are a small number of calibration studies defining raw data thresholds in specific populations (Innerd, Harrison and Coulson (2018) (overweight and obese adults); Hildebrand et al. (2014) (children and adults); Rowlands et al. (2016) (adults)). These studies have started to help the research field by publishing thresholds that can be used and compared to other studies. However, no previous study has involved specifically young adults or the university students to develop thresholds, which shows a gap in the literature. As previously stated university students represent a large proportion of young adults, so this is an important group and there is a considerable need for these thresholds.

Considering all the points discussed in this literature review surrounding the prevalence of inactivity in university students, the associations between SB and PA and health status outcomes, the development of raw data analysis and the lack of thresholds in a university student population. The aims and objectives of the study were as follows:

Aim

The primary aim of this study was to calculate and cross-validate accelerometer thresholds to classify PA and SB in university students.

Objectives

Objective 1. To design a calibration study to generate accelerometer and energy expenditure data across a range of activities, that simulated daily living in university students.

Objective 2. To calculate and cross-validate accelerometer thresholds to classify university students' SB and PA behaviours generated using a calibration circuit.

Chapter 3: Study Methods

Participants: After gaining ethical approval from the University Research Ethics Committee (19/SPS/005), participants were invited to take part in the study via recruitment emails to course/module leader and posters put up in university social zones. A convenience sample of 35 (*n*= 21 female) undergraduate or postgraduate students enrolled at Liverpool John Moores University (LJMU) during the 2018/2019 academic year, studying any course at any academic level provided written informed to consent to take part. Participants were excluded if they were not aged between 18-25 years, and/or not enrolled at LJMU during 2018/2019 academic year, and/or not mobile at the time of data collection, and/or were not able to complete vigorous activities, and/or were not able to provide informed consent. The participants received a £10 voucher for their participation.

Study design: Cross-sectional study design using a calibration circuit of standardised activities (table 1) conducted between February and May 2019.

Calibration Circuit Protocol: The participants were instructed to refrain from eating, consuming caffeine or alcohol, smoking and exercise in the 2 hours prior to data collection. The laboratory based calibration circuit took approximately 2.5 hours to complete. Table 1 displays the activities included in the calibration circuit. The activities selected were informed by previous accelerometer calibration studies, the 2011 Compendium of Physical Activities and the typical activities that the population of interest engages in (Ainsworth et al., 2011). Three SBs (quiet supine rest, seated watching TV and seated typing at a computer), three LPA (standing folding clothes, light walk and standing stacking items), three MPA (brisk walk, climbing stairs and cleaning) and three VPA (running, activity circuit and dribbling a basketball) activities were included, and were completed by all participants (table 1). Each activity lasted 5 minutes and was followed by a 5 minute rest period, with the exception of a 15 minute quiet supine rest which was completed at the start of each data collection session. Each participant completed the

remaining activities in a randomised order. The randomisation of activities occurred using a research grade randomisation tool (https://randomizer.org). The participants were given a demonstration of the next activity during the preceding rest period. The lead researcher recorded the time (hours, minutes and seconds) at the start and end of each activity or rest period to allow the synchronisation of the accelerometer, calorimetry and activity data. Apart from quiet supine rest, the first and last 30 seconds of each activity were discarded in case of transitional movements and the remaining 4 minutes was used for analysis (Hurter et al., 2018).

For all treadmill activities (set at a 0% grade/incline; h/p/cosmos; Nussdorf, Germany) (light walk, brisk walk and run), the researcher recorded the start of the activity, once the participant had reached the intended speed. To account for individual differences in lower limb length the Froude Number (Fr) was calculated to standardise treadmill speeds between individuals. Fr is a dimensionless variable that allows the comparisons of motion between individuals with different gait (Minetti, 2001). By anchoring treadmill speed to Fr numbers the relative intensity of the activity was standardised between participants (Arvidsson et al., 2019). An Fr of 0.25 represents an individual's optimal walking speed and 0.5 represents the transition from walk to run. After pilot testing (data not shown), the following Fr numbers were assigned to each treadmill activity: 'light walk' was anchored to Fr = 0.175, 'brisk walk' Fr = 0.3 and 'run' Fr = 0.65. To calculate the participants' individual treadmill speed for each activity the formula: treadmill speed = (limb length * (Gravity 9.81 * Fr)) was used.

For consistency, during the climbing stairs activity the researcher walked up and down the stairs with each participant to ensure they kept in time with a metronome, which was set at 80 bpm.

Table 1: Standardised activity protocol

Activity	Description
Sedentary	
15 minutes of quiet supine Rest	Supine lying position on a hospital bed, asked to avoid bodily movements and talking
Seated watching TV	Sitting comfortably in a chair, watching TV on a computer screen
Seated typing at a computer	Sitting comfortably in a computer, typing a paragraph of text
Light	
Standing stacking items	Standing, putting 6 cans of food varying in weight on a shelf and putting them on a desk and repeating
Standing folding clothes	Standing at a table, folding a pile of clothing (6 t-shirts)
Light walk	Walking on a treadmill at individually calculated speed
Moderate	
Brisk Walk	Walking on a treadmill at individually calculated speed
Climbing stairs	Climbing stairs, following a metronome to consistently have all participants walk at the same speed
Cleaning	Sweeping up confetti with a broom
Vigorous	
Run	Running on a treadmill at individually calculated speed
Activity Circuit	Following standardised video containing 5 minutes of continuous upright activities, each lasting approximately 30 seconds.
Dribbling a Basketball	Dribbling a basketball around cones.

Outcome measures:

Demographic questionnaire: Participants' date of birth (then converted into age on day of data collection), gender, ethnicity, permanent home postcode, university home postcode and parents' highest level of educational attainment were collected using a demographic questionnaire.

Anthropometrics: Participants removed their shoes before anthropometric measures were taken. Stature and seated stature were assessed to the nearest 1 cm using a Stadiometer (model 213; SECA, Hamburg, Germany) and body mass was assessed to the nearest 0.1 kg (SECA scale 704; SECA, Hamburg, Germany) using standard methods

(Lohman et al. 1988). Each measurement was taken twice, and an average of both measurements was calculated. Body Mass Index (BMI) was calculated via the formula: (Weight/(Stature X Stature)).

Indirect Calorimetry: The Metamax (3B-R2; Cortex, Leipzig, Germany) was used to assess energy expenditure during the calibration protocol. The Metamax is a small, lightweight, portable indirect calorimeter, which measures the flow of air via a turbine connected to a facemask (7450 series silicone V2™ Oro-Nasal Mask, Hans Rudolph, Kanas, USA). Similar devices have been recommended for use in calibration studies (Welk et al., 2019). Prior to use, the Metamax was calibrated according to the manufacturer's instructions. Participant characteristics (date of birth, gender, body mass and stature) and facemask size were inputted into the Metamax software prior to data collection. After all sessions were completed, breath-by-breath data were converted into time-stamped second-by-second data to allow synchronisation with accelerometer data. The Metabolic Equivalent of Tasks (MET) values were calculated for each of the activities. This was completed by removing the first and last 30 seconds, averaging the remaining 4 minutes and finally dividing this by the participant's measured resting metabolic rate. The resting metabolic rate was defined as the mean VO_2 value observed during minutes 9-14 of the quiet supine rest activity. This was used to classify the intensity of the activities for threshold generation. The MET thresholds for activity classification were as follows: SB (<1.5 METs), LPA (≥1.5-2.9 METs), MPA (≥3.0–5.9 METs), VPA (≥6 METs) and MVPA (≥3 METs) (Mendes et al. 2018).

Accelerometers: All participants wore three ActiGraph GT9x accelerometers (ActiGraph Corp, Pensacola, FL, USA) throughout data collection: one on the dominant wrist (observed by the researcher when completing the consent form and verbally confirmed by the participants), one on the non-dominant wrist and one on the right hip. All accelerometers were attached by the researcher prior to the start of the calibration activity protocol. The wrist accelerometers were attached via two Link wristbands and the hip was placed in a Link belt pouch attached to the participant via an elastic belt. The monitors

were initialised to collect data using a sampling frequency of 100Hz. In the current study, the epoch length was set at 1 second. Monitors were initialised and downloaded using ActiLife (6.13.3; ActiGraph Corp, Pensacola, FL, USA). Once data collection was completed, all data was downloaded as .gt3x (raw) files and then converted into .csv files to facilitate raw data analysis. The files were then processed using the GGIR package (version 1.9-1) (Migueles et al., 2019b) in R (version 1.2.1335; Boston, USA; https://www.rstudio.com/). GGIR converted the raw accelerometer signals into an omnidirectional measure, which corrected for gravity referred to as Euclidean Norm Minus One (ENMO) (Migueles et al., 2019b). For the ENMO metric, negative values were rounded up to zero and due to the short data collect period auto-calibration was not completed (van Hees et al., 2013).

Data analysis and threshold generation: ENMO thresholds were generated for SB through to VPA using data from 21 participants (calibration group, females *n*=12) and 12 participants (females, *n*=7) were randomly assigned to a cross-validation group. Random allocation was achieved by selecting participant numbers “out of a hat”. These groups were proportionate to the total sample based on gender, with 63% of the sample allocated to the calibration group.

Threshold generation and cross-validation followed a 3-step process for each placement location (non-dominant wrist, dominant wrist, hip). *Step 1*. Consistent with other studies, Receiver Operating Characteristic (ROC) curve analysis was completed using indirect calorimetry data as the criterion reference to generate a threshold, Area Under the Curve (AUC), sensitivity and specificity for all activity intensities (SB, LPA, MPA, VPA and MVPA) using the calibration group data. The novel approach reported in Crotti et al. (2019, *in review*) was also utilised in this study whereby two versions of the ROC curve analysis were completed. The pairs analysis is a new novel approach which has not be tested in many previous accelerometer studies. This helps better classify SB and PA behaviours due to the disproportionate amount of data, as this would influence the ROC curve analysis. The traditional approach is often considered arbitrary, unstandardized and

can reduce information by pooling findings together. Paired designs allow fair comparison between tests (Obuchowski and Bullen, 2018). The first approach used intensity pairs (e.g. when calculating a threshold for SB data were included for SB and LPA only, referred to as the pairs analysis) and the second approach used all the data from the calibration protocol across all intensities. Crotti et al. (2019, *in review*) reported the pairs analysis reduced bias associated with unequal distributions of PA behaviours and provided a better estimate of time spent in SB, MVPA and VPA in comparison to the 'all data' approach. The pairs used for analysis were as follows: SB – LPA, LPA – MPA, MPA – VPA.

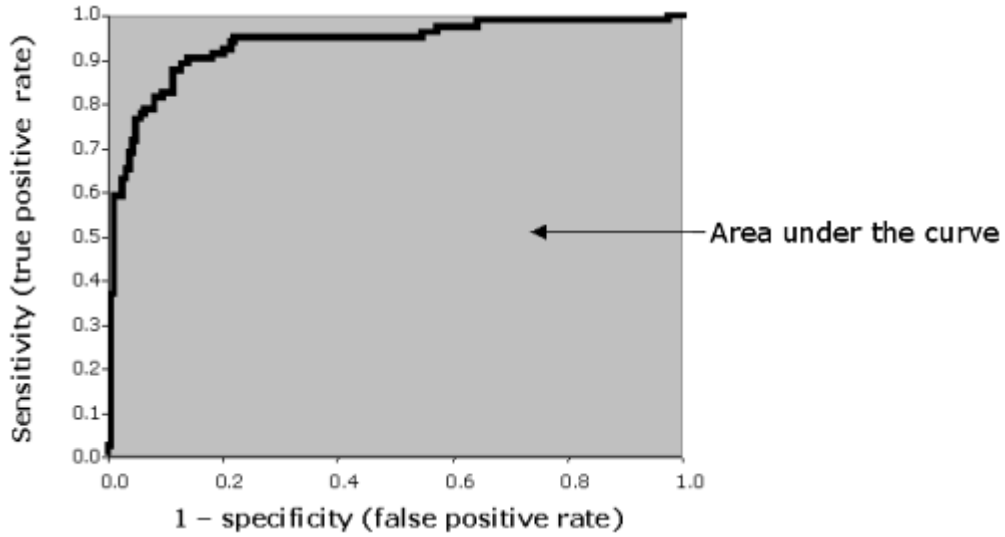


Figure 1: Illustrative ROC Curve figure

The Youden Index was used to identify the optimal threshold (Ruopp et al., 2008). Agreement between the estimates generated by the thresholds and the criterion measure was examined at the individual and group levels using sensitivity, specificity, percentage agreement, mean absolute percent error (MAPE), Cohens' Kappa coefficients and equivalency analysis. Cohens' Kappa coefficients were interpreted as follows: <0.0 no agreement, 0-0.2 slight agreement, 0.21-0.4 fair agreement, 0.41-0.6 moderate agreement, 0.61-0.8 substantial agreement and 0.81-1 almost perfect agreement (Landis and Koch, 1977). Equivalency analysis was used in this study to determine whether PA estimates from the monitor were equivalent to the estimates from the criterion measure on

average at the group level (Lee, Kim and Welk, 2014). Equivalence testing has been increasingly used in PA research (Kim and Welk, 2015; Boddy et al., 2018; Dixon et al., 2018). A 95% equivalency test was completed to establish whether 90% CI of the PA estimates fell within the zone of equivalence, defined as ±10%, ±15% and ±20% of the mean for the criterion reference. This study used paired-data CI because the data from the accelerometer and criterion measure were simultaneously measured (Dixon et al., 2018).

Step 2. To improve the accuracy of the thresholds, the thresholds identified in Step 1 were increased or decreased by 2- m*g* to improve the performance of thresholds based on the balance between group and individual level equivalency. Firstly, thresholds 5 m*g* lower/higher than the threshold identified in Step 1 were examined. From this, equivalency analysis was repeated and thresholds were then further refined. Sensitivity, specificity, percentage agreement, MAPE, Cohens' Kappa coefficients and group level equivalency were re-examined.

Step 3. Cross-validation analysis was completed using the thresholds identified in Step 2, that were deemed the most appropriate for use based on the balance of individual and group level comparability observed. Cross-validation was completed with the 12 participants not included in the main calibration analysis. Sensitivity, specificity, percentage agreement, MAPE, Cohens' Kappa coefficients and group level equivalency were examined. Descriptive statistics and ROC curve analyses were completed using SPSS (SPSS windows version 26.0; IBM, Armonk, NY) and all other analyses were completed using Microsoft Excel (windows version 16; Microsoft Corporation, Washington, USA).

Chapter 4: Study Results

Out of the thirty-five participants who took part in the calibration study, thirty-two participants had complete datasets and were included in all aspects of analysis. Three participants were removed for the following reasons: two participants were excluded where indirect calorimetry data was not captured (device or calibration error), and one due to missing accelerometer data files. The descriptive statistics for the thirty-two participants included in the final sample can be viewed in Table 2. The participants' calculated Fr speeds for the treadmill activities (light walk, brisk walk and run) can be viewed in Table 3. The intensity classifications of the activities based on MET values, VO_2 and ENMO for each placement location can be viewed in Table 4.

Table 2: Means (±SD) or percentage anthropometrics and demographic characteristics included in analysis of participants

Variables	**Mean (SD)**
Age (y)	21 (0.4)
Right handed	87.5%
Sex	60% female
Stature (m)	1.7 (0.1)
Body mass (kg)	66.2 (9.1)
BMI (kg/m^2)	22.8 (3.1)
Degree type	23% on a Bachelor of Science 44% on a postgraduate course
Year of course	65% on the first year of their respective degrees 22% on the second year of their respective degrees 16% on the third year of their respective degrees

Table 3: Means (±SD) and range of Froude speeds for each treadmill activity

Activity	**Mean treadmill speed (km/h)**
Light walk	4.3 (0.2) km/h (ranging from 4.0-4.5 km/h)
Brisk walk	5.6 (0.2) km/h (ranging from 5.1-6.0 km/h)
Run	8.3 (0.3) km/h (ranging from 7.6-8.8 km/h).

Table 4: Means (±SD) Intensity classifications of the standardised activity protocol based on MET values, VO_2 output and ENMO for each placement location

Activity	MET Value	VO_2 (mL O^2.kg^{-1}.min^{-1})	Non-Dominant ENMO (m*g*)	Hip ENMO (m*g*)	Dominant ENMO (m*g*)
Sedentary					
15 minutes of quiet supine rest	1.0 (0.0)	4.2 (0.8)	4.1 (5.4)	1.4 (2.7)	5.1 (7.0)
Seated watching TV	1.0 (0.0)	4.2 (1.0)	6.3 (8.0)	7.3 (9.4)	5.8 (6.5)
Seated typing at a computer	1.2 (0.0)	4.8 (0.8)	14.8 (8.9)	6.1 (9.2)	18.6 (9.0)
Light					
Standing stacking items	1.9 (0.1)	7.8 (1.9)	94.0 (37.6)	7.2 (6.4)	84.4 (30.6)
Standing folding Clothes	1.9 (0.1)	7.7 (1.7)	104.2 (21.8)	8.3 (6.5)	104.4 (20.7)
Cleaning	2.8 (0.9)	11.5 (3.3)	104.1 (38.9)	29.1 (35.8)	121.5 (52.8)
Moderate					
Light walk	3.3 (0.1)	13.2 (1.4)	116.2 (42.2)	124.2 (18.6)	108.9 (34.8)
Brisk Walk	4.3 (0.2)	17.6 (3.8)	191.2 (78.3)	197.1 (25.0)	174.0 (77.7)
Climbing stairs	5.1 (0.2)	20.7 (4.2)	136.7 (28.3)	107.8 (16.1)	123.9 (19.4)
Activity Circuit	5.8 (0.3)	23.2 (4.6)	271.0 (59.9)	154.8 (68.4)	269.3 (64.2)
Vigorous					
Run	8.0 (0.4)	32.4 (6.6)	659.0 (158.3)	552.1 (112.2)	612.6 (180.2)
Dribbling a Basketball	6.8 (0.4)	27.8 (9.9)	473.3 (241.3)	243.3 (133.1)	532.8 (169.8)

Threshold Generation:

Step 1. The ROC Curve generated thresholds developed using pairs of intensities (e.g SB-LPA) using the Youden Index presented values ranging from <3 to <40 m*g* for the

upper threshold for SB, ≥3 to ≥40 m*g* for the lower threshold of LPA, ≥40 to ≥126 m*g* for MPA and ≥250 to ≥328 m*g* for VPA intensity activity (table 5). The highest thresholds were observed for the non-dominant wrist placement, followed by the dominant wrist and lastly the hip. The diagnostic accuracy of the thresholds identifying SB or intensity of PA for all placements was better than what would be expected by chance (all AUC ≥0.59, $p<0.05$). All thresholds produced higher specificity compared to sensitivity values. Sensitivity, specificity, percentage agreement, MAPE and Cohens' Kappa coefficients were generally better for the hip placement (table 6: red indicates poor values, yellow represents acceptable values and green indicates good values. This is the same for all following tables) than the wrist placements.

Table 5: ROC curve analysis for the pairs analysis

	ENMO Value (m*g*)	**Area Under the Curve**	**95% CI**	**Sensitivity**	**Specificity**
Non-Dominant Wrist					
SB	<30	0.98	0.98-0.98	88.6	96.2
LPA	≥30	0.98	0.98-0.98	88.6	96.2
MPA	≥126	0.69	0.68-0.69	55.3	74.2
VPA	≥328	0.87	0.87-0.88	76.5	90.6
MVPA	≥126	0.67	0.66-0.68	44.0	87.6
Hip					
SB	<3	0.73	0.72-0.74	73.9	62.8
LPA	≥3	0.73	0.72-0.74	73.9	62.8
MPA	≥40	0.97	0.97-0.97	89.3	95.6
VPA	≥250	0.84	0.84-0.85	69.9	94.7
MVPA	≥40	0.59	0.58-0.60	36.3	95.8
Dominant Wrist					
SB	<40	0.96	0.96-0.97	83.0	96.2
LPA	≥40	0.96	0.96-0.97	83.0	96.2
MPA	≥112	0.67	0.66-0.67	58.7	67.3
VPA	≥310	0.89	0.89-0.90	82.3	89.3

MVPA	≥112	0.73	0.72-0.74	57.3	85.6

Table 6: Sensitivity, specificity, MAPE, percentage agreement, Cohens' Kappa coefficients and Interpretation of Kappa for the pairs analysis

	Sensitivity	Specificity	MAPE (%)	Agreement (%)	Cohens' Kappa coefficients	Interpretation of Kappa (Landis & Koch, 1977)
Non-dominant Wrist						
SB	0.76	0.91	11.0	85.9	0.67	Substantial Agreement
LPA	0.50	0.75	17.8	68.7	0.22	Fair Agreement
MPA	0.42	0.83	14.7	71.6	0.25	Fair Agreement
VPA	0.54	0.94	18.7	87.3	0.53	Moderate Agreement
MVPA	0.64	0.83	18.4	74.3	0.47	Moderate agreement
Hip						
SB	0.52	0.92	24.4	79.2	0.48	Moderate Agreement
LPA	0.48	0.73	18.6	67.2	0.19	Slight Agreement
MPA	0.68	0.79	12.3	75.6	0.43	Moderate Agreement
VPA	0.50	0.96	19.4	87.7	0.52	Moderate Agreement
MVPA	0.84	0.84	11.9	84.0	0.68	Substantial agreement
Dominant Wrist						
SB	0.72	0.90	11.0	83.7	0.62	Substantial agreement
LPA	0.38	0.80	11.3	70.1	0.17	Slight Agreement
MPA	0.44	0.81	13.3	71.2	0.25	Fair Agreement
VPA	0.56	0.93	18.2	86.3	0.53	Moderate Agreement
MVPA	0.69	0.80	13.4	75.1	0.49	Moderate agreement

The equivalency analysis found non-dominant wrist placement, SB was equivalent at ±20. However, no other threshold for this placement exhibited statistically significant equivalence in comparison to the criterion reference (Figure 2). For the hip placement, MVPA was equivalent at ±15. However, no other threshold estimates exhibited statistically significant equivalence in comparison to the criterion reference (Figure 3). For the dominant wrist placement, LPA and MVPA were equivalent at ±15. All other placements apart from VPA were equivalent at ±20 in comparison to the criterion reference (Figure 4).

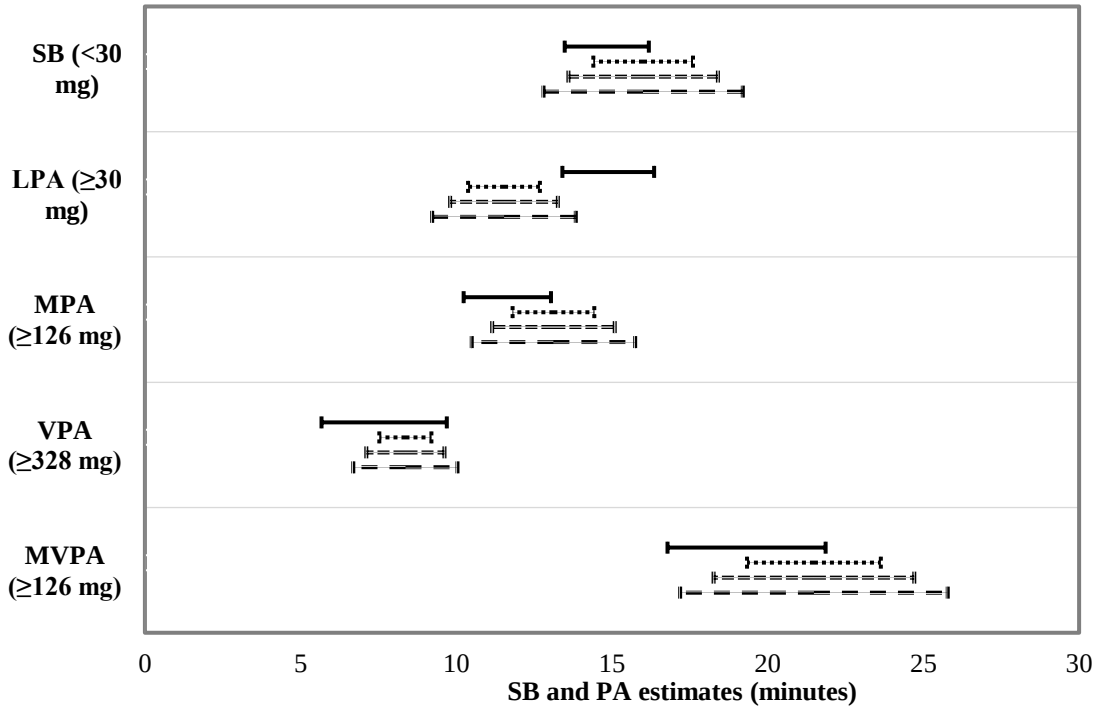


Figure 2: pairs analysis results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) on the non-dominant wrist placement in Step 1. The solid line represents the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.

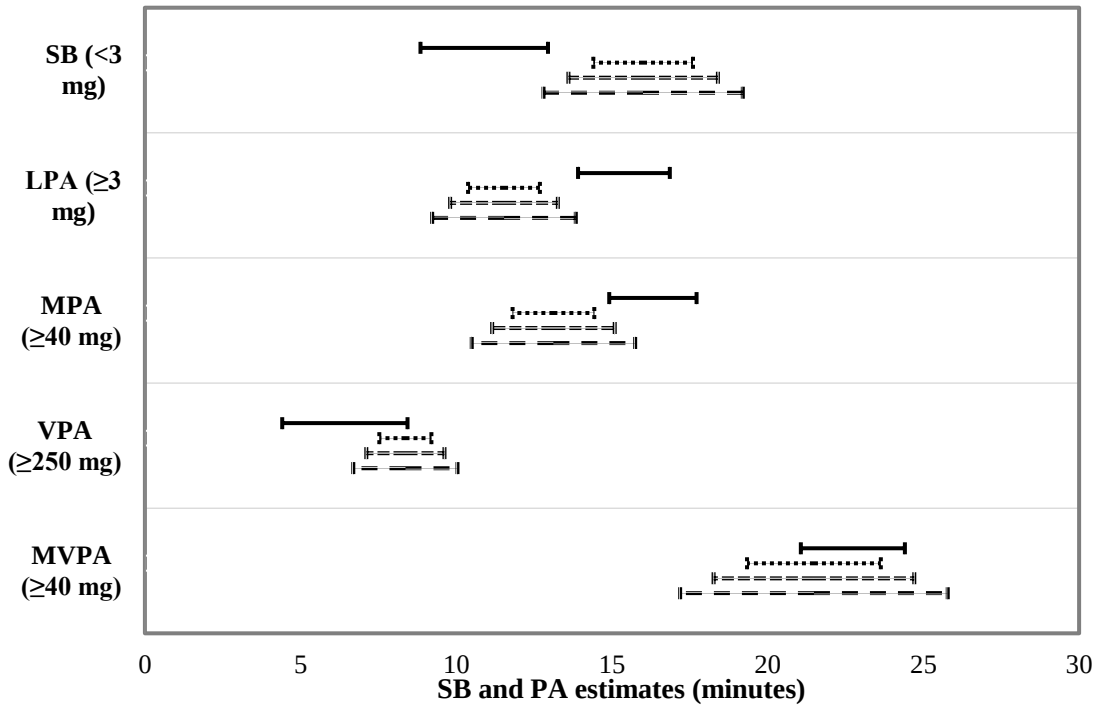


Figure 3: pairs analysis results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) on the hip placement in Step 1. The solid line represents the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.

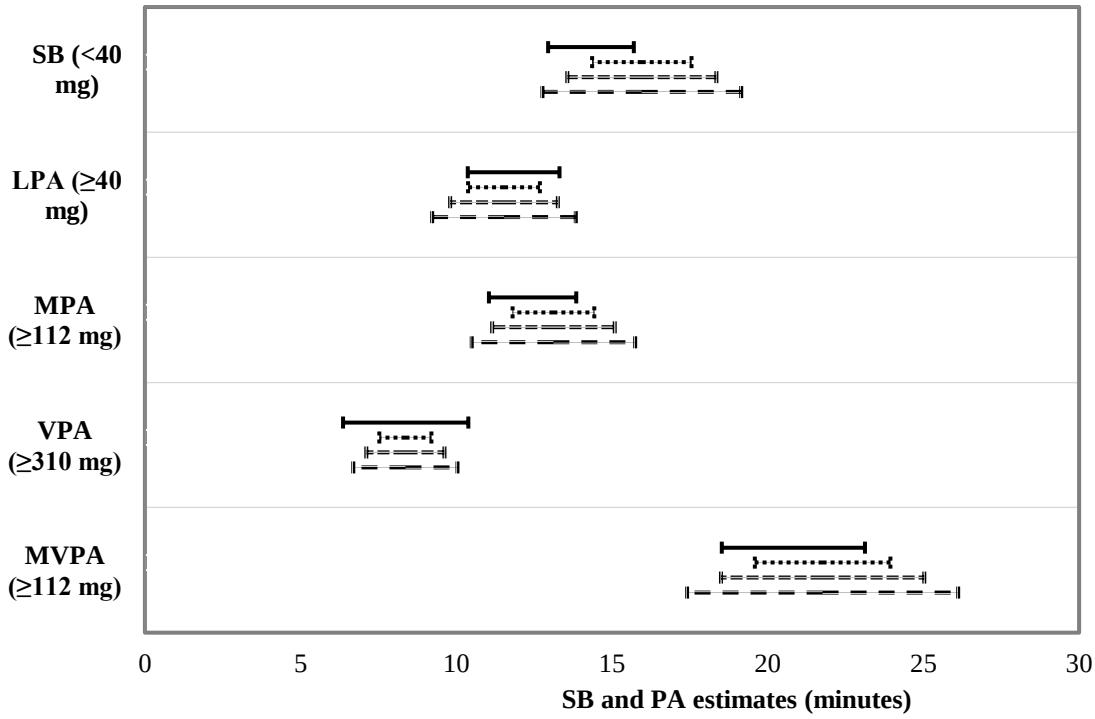


Figure 4: pairs analysis results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) on the dominant wrist placement in Step 1. The solid line represents the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.

The ROC Curve generated thresholds developed for SB and PA variables using all the data (i.e. not pairs, all intensities of PA) can be viewed in table 7. The diagnostic accuracy of the thresholds for identifying SB or intensity of PA for all placements was close to what would be expected by chance or better (AUC ≥0.51, $p<0.05$). This diagnostic accuracy was poorer than observed in the pairs analysis. This analysis produced similar sensitivity and specificity values to the pairs analysis and specificity was nearly always larger than sensitivity in contrast to what was observed for the pairs analysis. Sensitivity faired best at the hip 0.90 (SB) and MAPE faired best at the dominant wrist 9.2% (SB). Sensitivity, percentage agreement and Cohen's Kappa were greatest for the non-dominant wrist placement, 0.99 (LPA), 86.3% (VPA) and 0.67 (SB), respectively (table 8). Sensitivity values were lower than those in the pairs analysis, whereas specificity values were higher than those in the pairs analysis. Percentage agreement and MAPE were similar between the two analyses.

Table 7: ROC curve analysis for all intensities analysis

	ENMO Value (m*g*)	Area Under the Curve	95% CI	Sensitivity	Specificity
Non-Dominant Wrist					
SB	<30	0.67	0.67-0.68	94.1	42.5
LPA	≥30	0.99	0.99-0.99	92.8	97.2
MPA	≥183	0.51	0.50-0.51	33.1	91.6
VPA	≥270	0.93	0.93-0.94	79.7	94.1
MVPA	≥183	0.87	0.87-0.87	91.4	62.2
Hip					
SB	<28	0.90	0.90-0.91	67.3	97.0
LPA	≥28	0.78	0.78-0.79	92.9	71.3
MPA	≥50	0.73	0.73-0.74	58.6	97.6
VPA	≥214	0.94	0.94-0.94	73.5	95.3
MVPA	≥50	0.93	0.92-0.93	89.6	85.3
Dominant Wrist					
SB	<42	0.98	0.98-0.98	88.5	97.5
LPA	≥42	0.65	0.65-0.66	86.1	48.2
MPA	≥220	0.51	0.50-0.51	26.8	96.6
VPA	≥249	0.94	0.94-0.95	85.3	93.7
MVPA	≥220	0.88	0.88-0.88	88.1	68.0

Table 8: Sensitivity, specificity, MAPE, percentage agreement, Cohens' Kappa coefficients and Interpretation of Kappa for all intensities analysis

	Sensitivity	Specificity	MAPE (%)	Agreement (%)	Cohens' Kappa coefficients	Interpretation of Kappa (Landis & Koch, 1977)
Non-Dominant Wrist						
SB	0.76	0.91	11.0	85.9	0.67	Substantial agreement
LPA	0.68	0.68	35.6	64.8	0.25	Fair Agreement
MPA	0.17	0.94	39.1	73.4	0.15	Slight Agreement
VPA	0.59	0.92	18.7	86.3	0.52	Moderate Agreement
MVPA	0.49	0.93	29.7	73.1	0.43	Moderate Agreement
Hip						
SB	0.90	0.71	28.9	77.1	0.54	Moderate Agreement
LPA	0.48	0.73	17.9	67.2	0.19	Slight Agreement
MPA	0.07	0.95	50.6	71.1	0.02	Slight Agreement
VPA	0.53	0.93	21.4	86.1	0.49	Moderate Agreement
MVPA	0.81	0.85	12.0	83.2	0.66	Substantial agreement
Dominant Wrist						
SB	0.79	0.86	9.2	84.1	0.64	Substantial agreement
LPA	0.68	0.62	40.7	63.2	0.22	Slight Agreement
MPA	0.62	0.65	27.2	64.1	0.23	Fair Agreement
VPA	0.61	0.91	17.8	85.2	0.51	Moderate Agreement
MVPA	0.41	0.97	41.8	71.5	0.40	Fair Agreement

For the equivalency analysis for the non-dominant wrist placement for all intensities (i.e. not pairs), SB was equivalent at ±20%. However, no other threshold for this placement exhibited statistically significant equivalence in comparison to the criterion reference (figure 5). For the hip placement, MVPA was equivalent at ±20%. However, no other threshold for this placement exhibited estimates that were equivalent in comparison to the criterion reference (figure 6). For the dominant wrist, SB was equivalent at ±20 and no other threshold for this placement exhibited statistically significant equivalence in comparison to the criterion reference (figure 7). Overall, the pairs analysis provided a better estimate of time spent in SB and intensities of PA at the individual and group level analysis due to the slightly better AUC, greater Cohen's Kappa coefficients and percentage agreement and these thresholds were carried on into step 2 of the analysis.

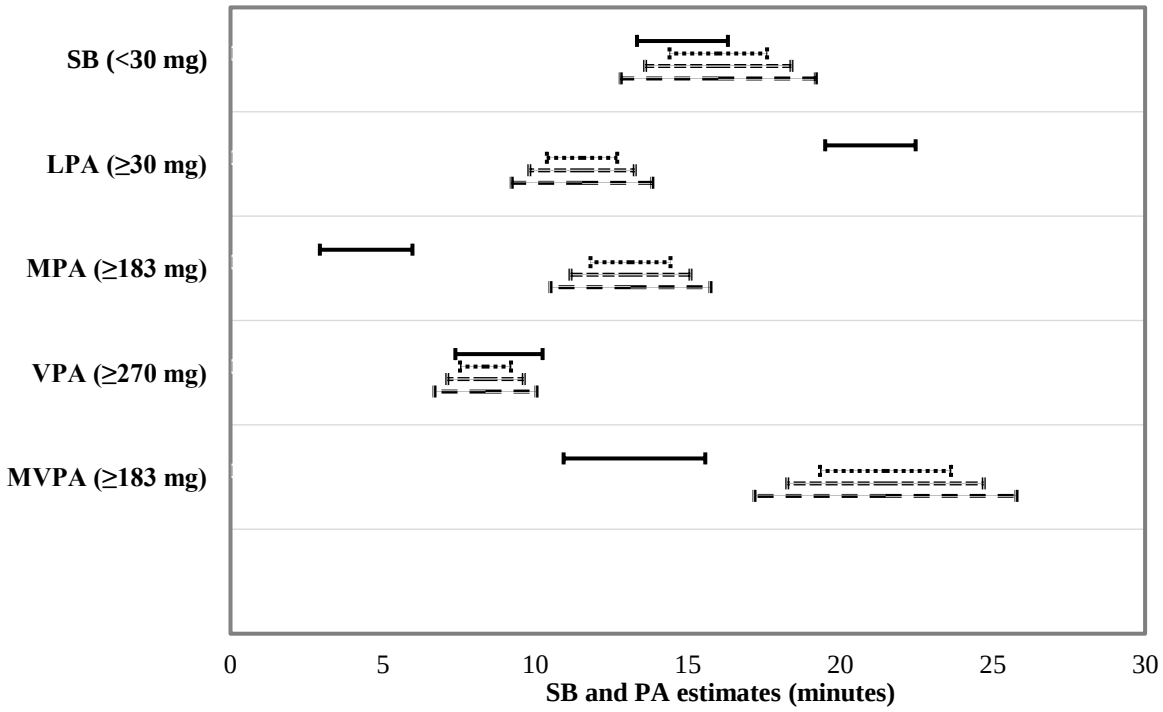


Figure 5: all intensities analysis results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) on the non-dominant wrist placement in Step 1. The solid line represent the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.

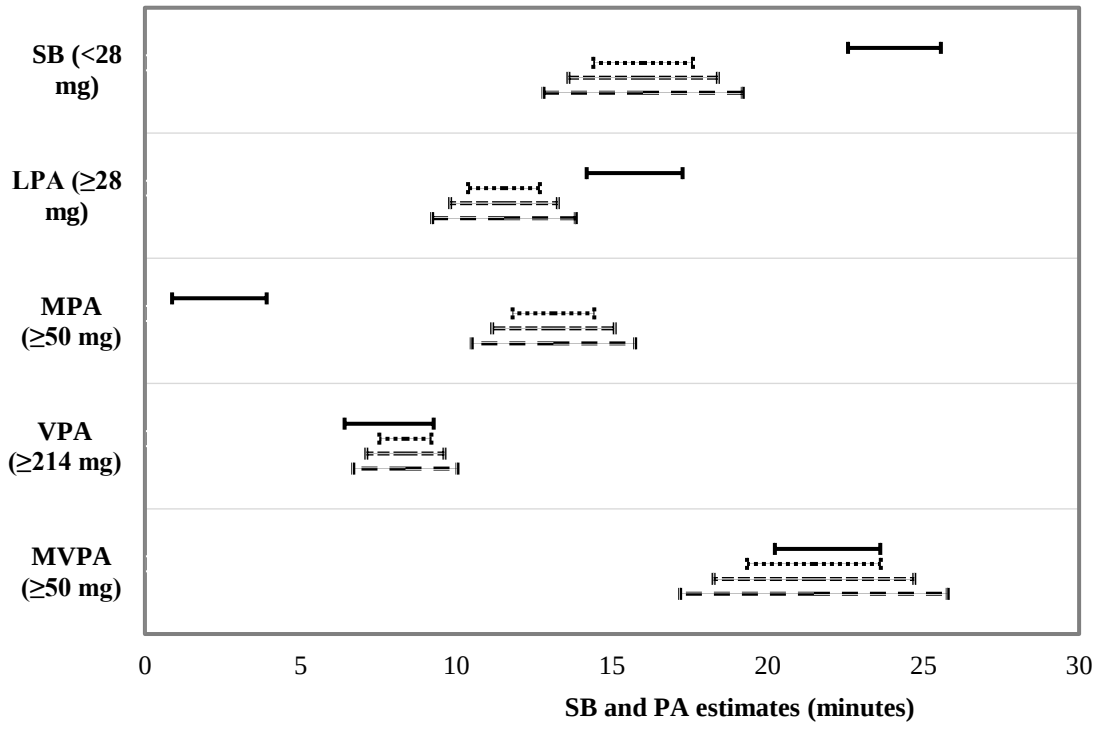


Figure 6: all intensities analysis results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) on the hip placement in Step 1. The solid line represent the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.

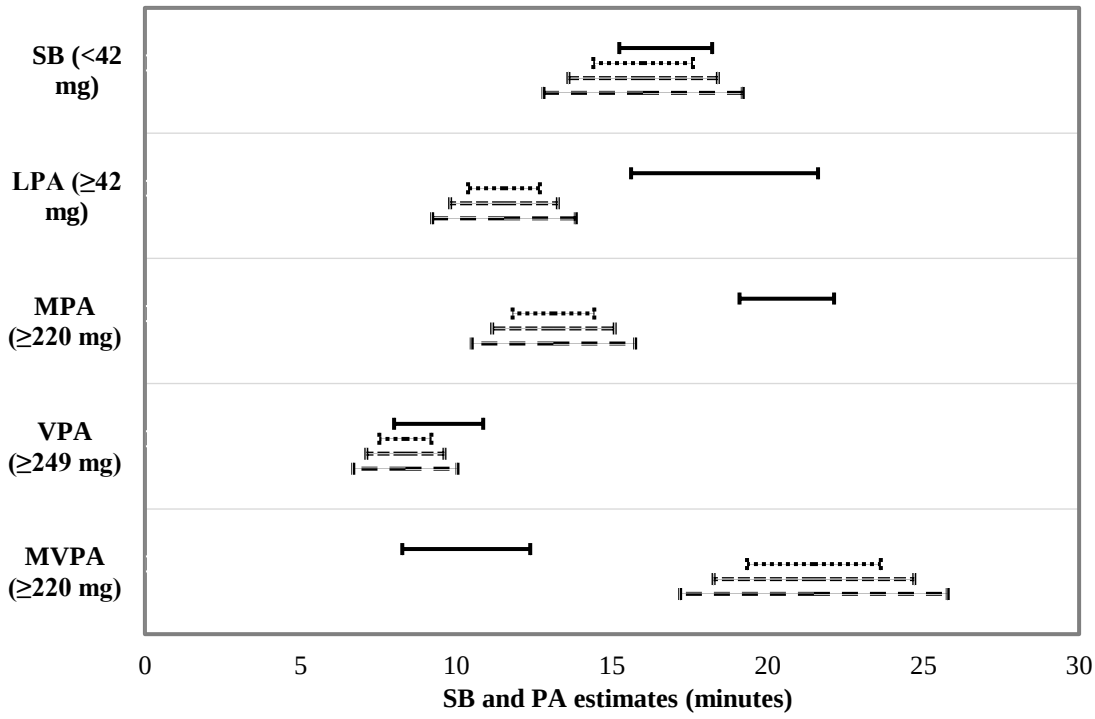


Figure 7: all intensities analysis results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) on the dominant wrist placement in Step 1. The solid line represent the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference

The thresholds that performed better in the equivalency analysis in step 1 were: LPA: ≥30, ≥3 and ≥40, MPA: ≥126, ≥40 and ≥112 and VPA: ≥328, ≥250 and ≥310 for non-dominant wrist, hip and dominant wrist placements respectively.

Step 2. The thresholds identified in Step 1 were fine-tuned by increasing or decreasing by a variety of different m*g* values to create the most appropriate thresholds based on the balance between group and individual level equivalency. Firstly, thresholds 5 m*g* lower/higher than the threshold in Step 1 were examined. Between two and four thresholds for SB, MPA and VPA were examined in this analysis step, depending on the equivalency results obtained for each m*g* increment. The LPA thresholds were the product of the optimum upper threshold for SB and lower threshold for MPA and MVPA was defined by the MPA threshold therefore LPA and MVPA were not examined separately within this analysis step (all figures in the appendix). Based on the balance of individual and group level comparability observed (tables 9, 10 and 11), the optimal thresholds identified in Step 2 were taken forward to the cross-validation analysis (Step 3). The thresholds taken forward were SB: <35, <8 and <40, LPA (by default): ≥35, ≥8 and ≥40, MPA: ≥110, ≥50 and ≥110 and VPA: ≥315, ≥225 and ≥315 and MVPA (by default): ≥110, ≥50 and ≥110 for the non-dominant wrist, hip and dominant wrist placements respectively.

Table 9: Step 2 Non-dominant thresholds developed to increase individual and group level agreement

Non-dominant wrist (mg)	Sensitivity	Specificity	MAPE (%)	Agreement (%)	Cohens' Kappa coefficients	Interpretation of Kappa (Landis & Koch, 1977)
SB						
<30	0.76	0.91	11.0	85.9	0.67	Substantial Agreement
<35	0.78	0.89	10.1	85.2	0.66	Substantial Agreement
<40	0.80	0.86	10.3	84.1	0.64	Substantial Agreement
<45	0.80	0.84	11.1	83.1	0.63	Substantial Agreement
LPA						
≥35	0.40	0.80	13.8	70.5	0.19	Slight Agreement
MPA						
≥110	0.47	0.80	13.8	71.2	0.27	Fair agreement
≥116	0.44	0.82	13.5	71.6	0.26	Fair agreement
≥121	0.42	0.83	13.7	72.0	0.26	Fair agreement
≥126	0.40	0.84	14.6	72.2	0.25	Fair agreement
VPA						
≥300	0.56	0.94	18.5	86.9	0.54	Moderate agreement
≥315	0.55	0.94	18.7	87.0	0.54	Moderate agreement
≥320	0.55	0.95	18.6	87.1	0.54	Moderate agreement
≥325	0.54	0.95	18.7	87.1	0.54	Moderate agreement
MVPA						
≥110	0.68	0.80	16.2	74.6	0.49	Moderate agreement

Table 10: Hip thresholds developed to increase individual and group level agreement

Hip (mg)	Sensitivity	Specificity	MAPE (%)	Agreement (%)	Cohens' Kappa coefficients	Interpretation of Kappa (Landis & Koch, 1977)
SB						
<3	0.50	0.92	25.5	78.8	0.46	Moderate agreement
<8	0.64	0.87	27.0	79.6	0.52	Moderate agreement
<13	0.69	0.82	26.1	78.1	0.50	Moderate agreement
<18	0.77	0.77	25.1	77.2	0.51	Moderate agreement
LPA						
≥8	0.40	0.77	23.1	68.7	0.16	Slight agreement
MPA						
≥45	0.64	0.80	8.3	75.1	0.41	Moderate agreement
≥50	0.55	0.82	8.2	75.1	0.37	Fair agreement
≥55	0.60	0.81	8.4	75.8	0.40	Fair agreement
VPA						
≥225	0.52	0.95	23.1	86.8	0.52	Moderate agreement
≥240	0.50	0.97	19.2	87.5	0.53	Moderate agreement
≥245	0.50	0.96	19.2	87.6	0.54	Moderate agreement
≥249	0.50	0.97	19.2	87.7	0.54	Moderate agreement
MVPA						
≥50	0.81	0.86	11.6	83.4	0.67	Substantial agreement

Table 11: Dominant thresholds developed to increase individual and group level agreement

Dominant wrist (m*g*)	Sensitivity	Specificity	MAPE (%)	Agreement (%)	Cohens' Kappa coefficients	Interpretation of Kappa (Landis & Koch, 1977)
SB						
<40	0.78	0.87	9.1	84.1	0.64	Substantial agreement
<45	0.81	0.86	10.7	84.0	0.64	Substantial agreement
LPA						
≥40	0.37	0.81	11.1	70.6	0.18	Slight agreement
MPA						
≥107	0.45	0.80	13.2	70.7	0.25	Fair agreement
≥110	0.45	0.80	13.4	70.7	0.25	Fair agreement
≥112	0.44	0.81	13.5	70.8	0.25	Fair agreement
VPA						
≥305	0.57	0.93	35.9	85.9	0.51	Moderate agreement
≥310	0.56	0.93	34.9	85.9	0.51	Moderate agreement
≥315	0.56	0.93	36.0	86.0	0.51	Moderate agreement
MVPA						
≥110	0.70	0.80	13.4	75.1	0.50	Moderate Agreement

Figures 8, 9 and 10 display the equivalency analysis for the final, refined thresholds. For the non-dominant wrist, the SB threshold provided equivalent estimates to the ±10% zone of equivalency in comparison to the criterion reference. MPA and MVPA were equivalent at ±15% and LPA was equivalent at ±20%. The VPA threshold did not provide statistically equivalent estimates in comparison to the criterion reference. For the hip placement, MVPA was equivalent at ±10%. No other threshold provided statistically equivalent

estimates in comparison to the criterion reference for the ±10%, ±15% or ±20% equivalency zones. For the dominant wrist placement, SB was equivalent at ±10%. LPA and MVPA were equivalent at ±15% and MPA was equivalent at ±20%. The VPA threshold did not provide statistically equivalent estimates in comparison to the criterion reference.

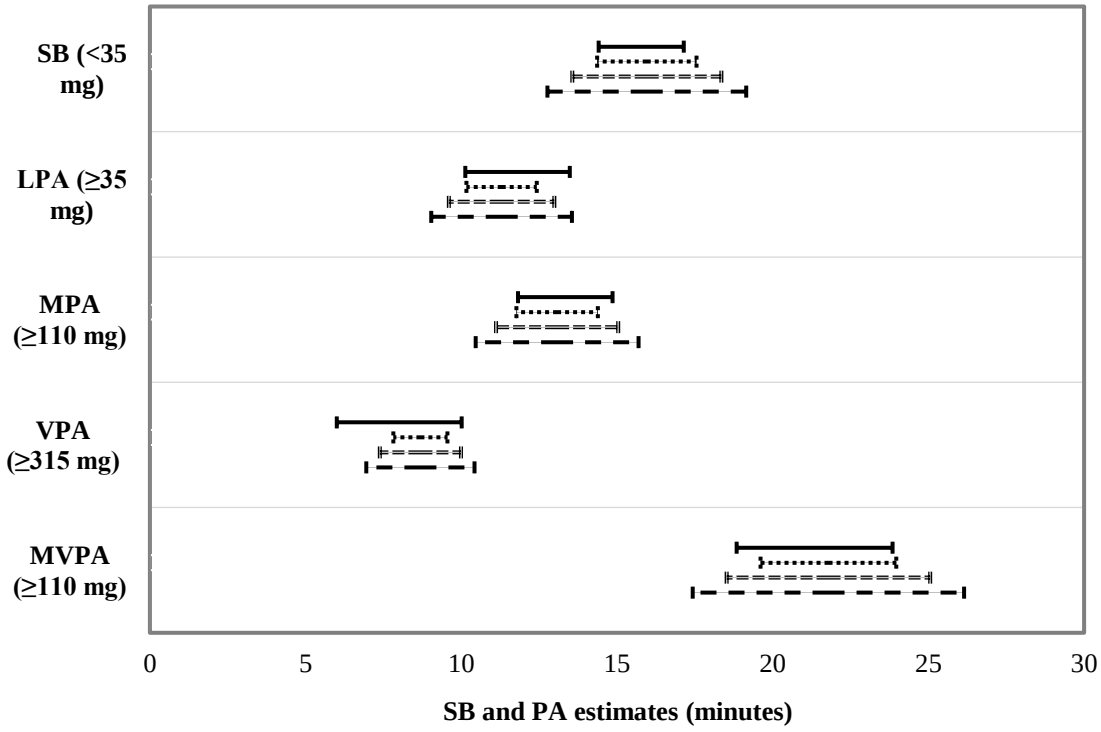


Figure 8: results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) on the non-dominant wrist placement in Step 2. The solid line represents the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.

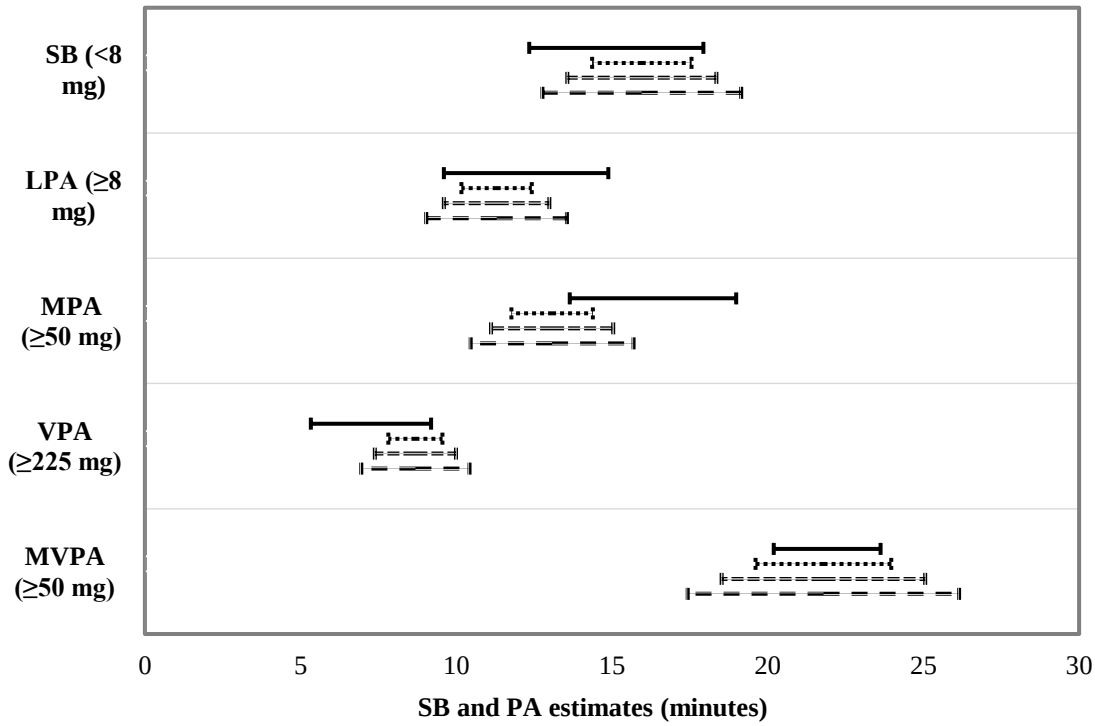


Figure 9: results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) on the hip placement in Step 2. The solid line represent the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.

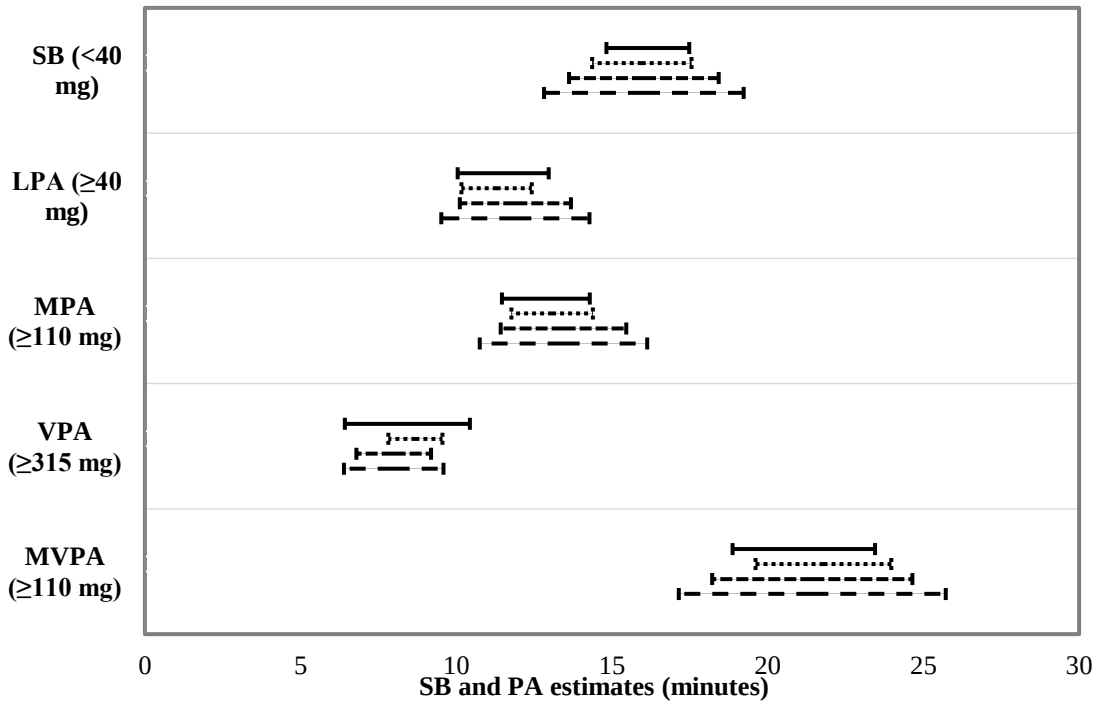


Figure 10: results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) on the dominant wrist placement in Step 2. The solid line represent the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.

Step 3. The optimum thresholds identified in Step 2 were then cross-validated in the independent cross-validation group (*n*=12, females, *n*=7). Table 12 displays the threshold value, sensitivity, specificity, MAPE and Cohens' Kappa coefficients, which were very similar to those to the previous stage. Percentage agreement and MAPE performed better in the cross-validation stage compared to the previous stage. In the cross validation, Cohens' Kappa coefficients were highest at the non-dominant wrist placement: 0.69 (SB). Sensitivity, specificity and percentage agreement: 0.76 (SB), 0.94 (MVPA) and 89.1% (VPA) respectively. MAPE was best at the hip placement: 6.8% (MPA). Cohens' Kappa coefficients of the thresholds in this stage were similar or better to those observed in Step 2. The lowest Cohen's Kappa value was for the LPA threshold (0.15, slight agreement) on the hip placement and the highest value for SB (0.70, substantial agreement) from the dominant wrist placement. On average according to the Cohen's kappa value and interpretation, the hip placement performed the best, followed by dominant wrist and non-dominant wrist respectively (table 12).

Table 12: Cut-points performance on cross-validation group

	Threshold (*mg*)	Sensitivity	Specificity	MAPE (%)	Agreement (%)	Cohens' Kappa coefficients	Interpretation of Kappa (Landis & Koch, 1977)
Non-Dominant Wrist							
SB	<35	0.76	0.92	10.9	87.0	0.69	Substantial agreement
LPA	≥35	0.38	0.83	13.6	71.9	0.21	Fair agreement
MPA	≥110	0.52	0.76	12.7	69.8	0.27	Fair agreement
VPA	≥315	0.67	0.94	12.1	89.1	0.60	Moderate agreement
MVPA	≥110	0.77	0.74	13.6	75.3	0.50	Moderate agreement
Hip							
SB	<8	0.84	0.81	20.7	82.2	0.61	Substantial agreement
LPA	≥8	0.25	0.88	22.7	72.8	0.15	Slight agreement
MPA	≥50	0.68	0.83	6.8	78.8	0.49	Moderate agreement
VPA	≥225	0.64	0.95	11.3	90.1	0.62	Substantial agreement
MVPA	≥50	0.84	0.85	8.8	84.6	0.69	Substantial agreement
Dominant Wrist							
SB	<40	0.79	0.91	9.5	87.2	0.70	Substantial agreement
LPA	≥40	0.39	0.82	9.8	71.6	0.21	Fair agreement
MPA	≥110	0.45	0.79	9.8	69.6	0.24	Fair agreement
VPA	≥315	0.72	0.92	11.9	88.5	0.60	Moderate agreement
MVPA	≥110	0.74	0.76	12.5	75.0	0.50	Moderate agreement

Figures 11, 12 and 13 display the cross-validation equivalency analysis for Step 3. For the non-dominant wrist, the SB threshold provided equivalent estimates at ±20% in comparison to the criterion reference, no other threshold provided statistically equivalent estimates in comparison to the criterion reference. For the hip placement, MVPA estimates were equivalent at ±15% and MPA was equivalent at ±20%. No other hip threshold provided statistically equivalent estimates in comparison to the criterion reference. For the dominant wrist placement, SB and MPA were equivalent at ±15%. LPA and MVPA estimates were equivalent at ±20%. VPA was the only threshold that did not

provide statistically equivalent estimates in comparison to the criterion reference for the ±10%, ±15% or ±20% equivalency zones.

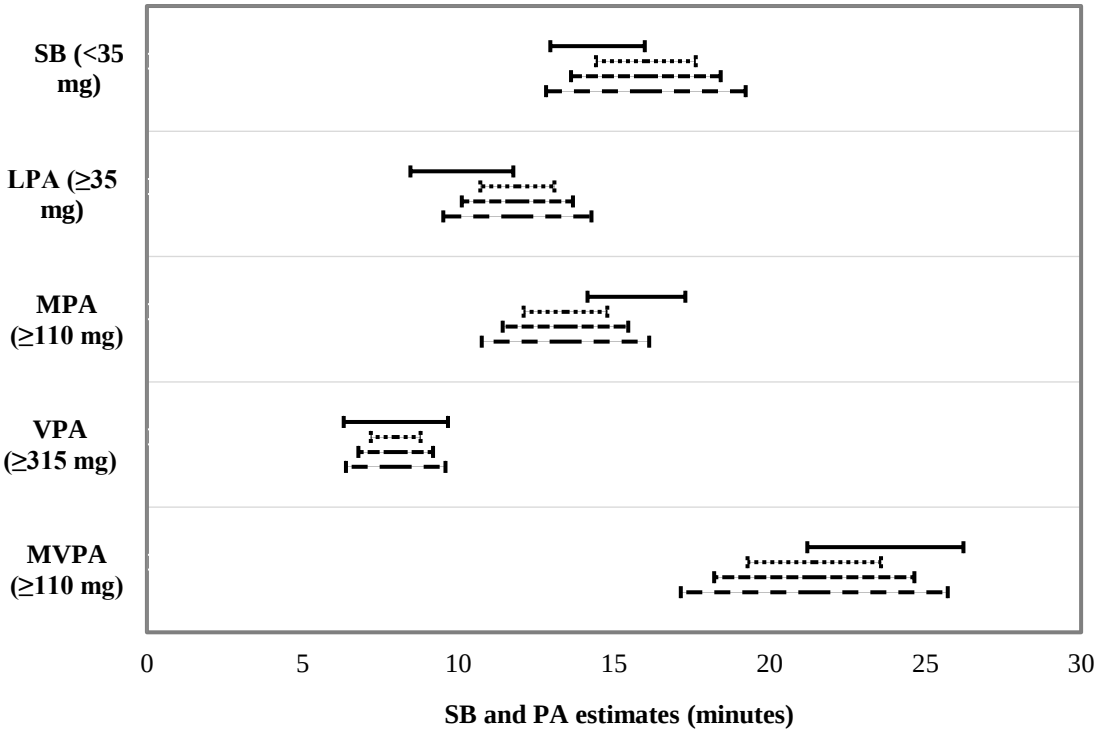


Figure 11: Non-dominant wrist placement cross validation results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) in Step 3. The solid line represents the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.

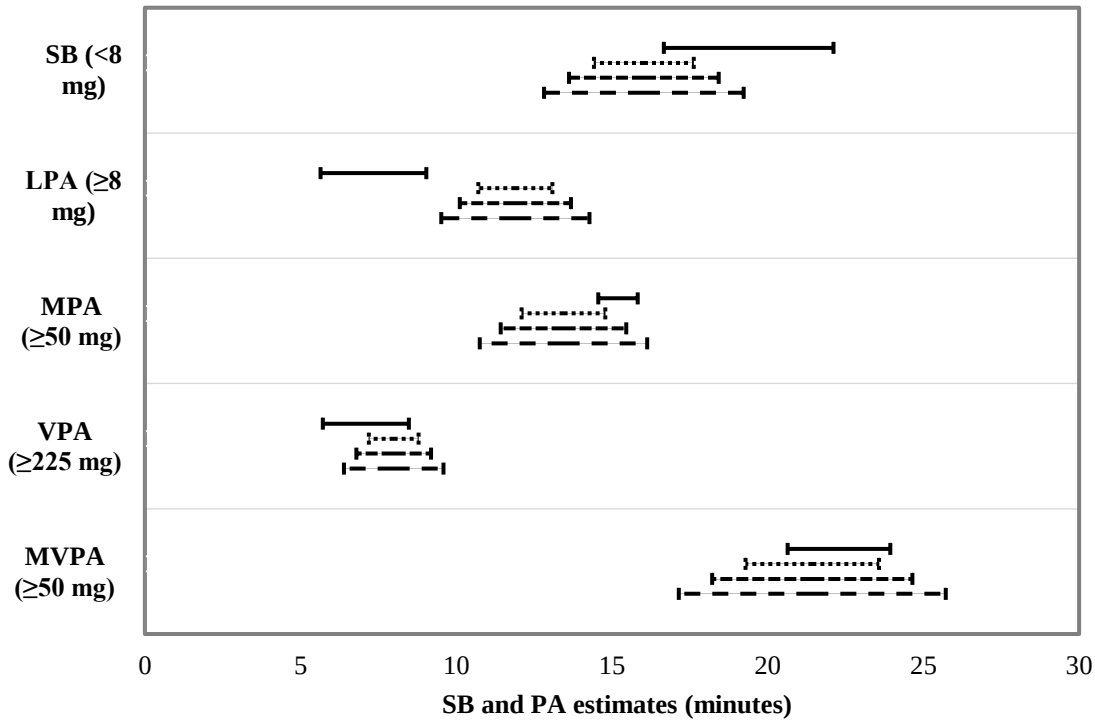


Figure 12: Hip placement cross validation results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) in Step 3. The solid line represents the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.

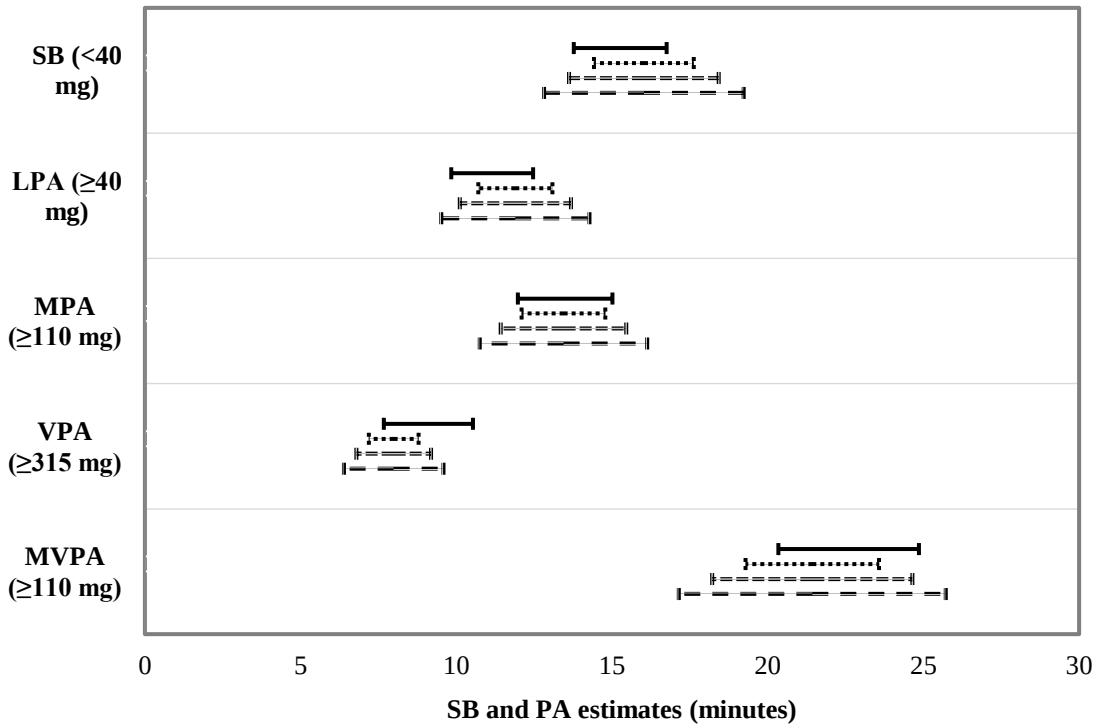


Figure 13: Dominant wrist placement cross validation results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) in Step 3. The solid line represents the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.

Chapter 5: Study Discussion

This study aimed to calculate and cross-validate accelerometer thresholds to classify SB and PA behaviours in university students. The thresholds for SB, LPA, MPA and MVPA showed acceptable levels of agreement between the accelerometer and criterion reference in regards to specificity, MAPE and percentage agreement. VPA was the only threshold to show no equivalency between accelerometer and criterion (Metamax) on any placement site. The hip placement thresholds were the most accurate in regard to sensitivity, specificity, percentage agreement, MAPE and Cohens' Kappa coefficients. Both the wrist placements (non-dominant and dominant) also demonstrated acceptable levels of agreement. In terms of percentage agreement, the dominant wrist placement demonstrated the lowest agreement between accelerometer and the criterion reference. On the other hand, the dominant wrist placement performed best in the equivalency analysis though, it should be noted that none of the placement sites had a threshold that produced estimates that were equivalent at ±10% in comparison to the criterion reference. This indicates that the accelerometer underestimated or overestimated the amount of time an individual spent in SB or any intensity of PA. Consistently, LPA was underestimated in the cross-validation group, whereas MVPA was consistently overestimated within the same group. VPA was the only threshold that did not demonstrate equivalence at any level (±10%, ±15% or ±20%) across any of the placement sites, despite this a high level of agreement between accelerometer and criterion reference (≥88.5%), moderate and substantial agreement based on Cohens' Kappa coefficients values (≥0.60) and good level of MAPE (≤12.1) were observed. The high agreement between accelerometer and criterion reference is important because quantities of time accumulated within VPA, MPA and MVPA are the most commonly cited outcome measures.

One interesting and unexpected finding in this study is both wrist placements (non-dominant and dominant) resulted in the same thresholds for MPA, VPA and MVPA. This

could be due to participants using both wrists for all these activities. It might be recommended that future studies have activities, which examine the differences between wrists, for example by using a writing task. The ENMO values in table 4 show large difference between the ENMO values at the wrist compared to that at the hip. This is referred to as “decoupling” and refers to the differences observed in values between the wrist and hip (Rowlands et al., 2014). The decoupling between the wrist and hip is especially apparent in the LPA activities, with the wrist being disproportionally higher than the hip during “housework” activities. For example, standing folding clothes, an activity which would be considered “housework” had a large decoupling between wrists and hip (Hip: 8.3 m*g*, non-dominant wrist: 104.2 m*g* and dominant wrist: 104.4 m*g*). The wrists (non-dominant and dominant) ENMO values are approximately 13 times more than that observed at the hip. The extent of decoupling of the wrist and hip accelerations and which the data from the wrist relate to the data from the hip is suggested to be population specific (Rowlands et al., 2014).

Analysis metrics and techniques

Sensitivity:

On average, SB across all placement sites performed the best when referring to sensitivity (non-dominant wrist: 0.76, hip: 0.84 and dominant wrist: 0.79). With the hip SB performing the best across all thresholds and placement sites (<8 m*g*) (0.84). Across all placement sites, on average LPA performed poorly sensitivity values (non-dominant wrist: 0.38, hip: 0.25 and dominant wrist: 0.38). LPA performed the worst at the hip in regards to sensitivity (≥8 m*g*), only correct identifying sensitivity 25% of the time. Sensitivity is the ability to correctly identifying whether a behaviour is occurring (Dziak et al., 2019). Even though LPA performed poorly across all placement sites (non-dominant: 0.38 and dominant: 0.39), with the highest placement for LPA (dominant) only capturing this intensity 38% of the time. One reason sensitivity for the LPA threshold at the hip was poor could be due to

the lack of lower body movement required during two of the activities included in the calibration circuit (standing stacking items and folding clothes; table 1). The accelerometer therefore could have interpreted these activities at the hip as sedentary due to the lack of registered acceleration signals. Montoye et al. (2016) study discovered sensitivity values at LPA were better than SB or MVPA. Although, Montoye et al. (2016) used direct observation as the criterion reference and had an activity where participants stood without completing any other movements. The comparisons between studies might therefore be limited, as the current study had no activities where participants stood still.

Specificity:

Across all placement sites VPA performed the best in regards to specificity (non-dominant wrist: 0.94, hip: 0.95 and dominant wrist: 0.92). VPA on the hip placement performed the best across all thresholds and placement sites (≥225 m*g*) (0.95). Across all placement sites, on average MVPA performed the worst in regards to specificity (non-dominant wrist: 0.74, hip: 0.85 and dominant wrist: 0.76). When looking into sensitivity and specificity, it is apparent the accelerometer estimates in this study were better at estimating the absence of a behaviour rather than detecting the presence of one, as the specificity values on average were generally higher than the sensitivity values. Sensitivity and specificity are inversely proportionate to each other. In practice, higher specificity would be preferable to allow the correct detection of health-enhancing behaviours and therefore allow the accurate reporting of PA behaviours however, it is unlikely that these behaviours would be overestimated using the generated thresholds, which would reduce the risk of misclassified as active when they are inactive.

PA guideline thresholds:

PA guidelines are concerned specifically with MPA, VPA and MVPA, so it is important to highlight these thresholds (MPA/MVPA: ≥50 m*g* and ≥110 m*g*, VPA: ≥225 m*g* and ≥315

m*g*). In the current study, MPA was identified poorly with 45%-68% of the time spent in MPA correctly estimated. When these thresholds are used in free-living situations, it could be assumed that a maximum of 68% of MPA could be estimated correctly. VPA did slightly better, with this being correctly identified between 64%-72%. It could be assumed that a maximum of 72% of VPA could be estimated correctly in free-living situations. Out of the highlighted thresholds, MVPA performed the best (75%-84%) and would be assumed in free-living situations up to 84% of MVPA can be estimated correctly. In this study, it should be mentioned that MVPA was consistently overestimated in the group level equivalency analysis. This could impact the results collected in free-living studies as higher reported MVPA levels than what the participants actually achieve. This would influence the accuracy of estimates related to the proportion of a population meeting guidelines. It was reported in Hildebrand et al. (2017) sensitivity values were always higher than the specificity values. The current study on average produced higher specificity values than sensitivity. In the pairs ROC curve approach, there was one threshold (hip SB: <3) which had a higher sensitivity value than specificity (sensitivity: 73.9, specificity: 62.8). In the all activity intensities ROC curve approach, only one SB threshold produced a higher sensitivity value than specificity value and this was at the non-dominant wrist placement (sensitivity: 94.1 and specificity: 42.5). Hildebrand et al. (2017), this study would better classify behaviours than the ones reported in the current study with higher sensitivity values.

MAPE:

Across all the placements, on average MPA demonstrated the lowest MAPE in comparison to SB and other intensities of PA, suggesting better individual-level agreement (non-dominant wrist: 12.7%, hip: 6.8% and dominant wrist: 9.8%). Across all thresholds and placement sites, MPA on the hip performed the best overall (≥50 m*g*) (6.8%). LPA performed the worst (non-dominant wrist: 13.6% and hip: 22.7%). Across all thresholds and placement sites, LPA on the hip (≥8 m*g*) performed the worse (22.7%).

MAPE provides an indication of overall measurement error and presents the error as a percentage (Nelson et al., 2016). It is also a more conservative estimate of error that takes into account both overestimation and underestimation because the absolute value of the error is used in the calculation (Lee, Kim and Welk, 2014). In Nelson et al. (2016) sedentary activities has the lowest MAPE values across all accelerometer brands (≤17%), however, in the current study this was not the same, MPA produced the best MAPE value (≤12.7%). However looking at one specific activity (stairs, included in MPA), MAPE was lower than (Nelson et al., 2016) (Nelson et al., 2016): 14% vs 6.8%, when looking at the hip placement). However, the differences could come from the choice of monitor (research grade accelerometer vs commercial monitor), in addition the participants in Nelson et al. (2016) self-paced their stairs activities whereas the participants in the current study were standardised and the analysis of MAPE for the current study was combined with the other MPA activities. The MAPE value of stairs alone is unknown in the current study. When comparing the current study to a previous study using an ActiGraph accelerometer (adults aged between 18-60 years) in controlled laboratory conditions (Bai et al., 2016), MAPE for SB was approximately 45% on the hip (reported on a figure and not mentioned in text). In the current study, MAPE for SB was considerably lower (20.7%). However, the difference may come from the activities selected as SB, as Bai et al. (2016) use only one sedentary posture (seated) which only fulfils one of the characteristics of SB, whereas the current study fulfilled two (sitting and lying).

Percentage agreement:

In reference to percentage agreement, across all placement sites, VPA performed the best (non-dominant wrist: 89.1%, hip: 90.1% and dominant wrist: 88.5%). Across all thresholds and placement sites, VPA on the hip performed the best (≥225 m*g*) (90.1%). On average, LPA performed the worst across all placement sites (non-dominant wrist: 71.9, hip: 72.8, dominant wrist: 71.6). Across all thresholds and placement sites, MPA on the dominant wrist overall performed the worse (≥110 m*g*) (69.8%). Percent agreement

has been criticized for its inability to account for chance agreement (McHugh, 2012). Percent agreement, due to the criticism, is not a widely used metric in adult studies, so comparison is limited and many of the previous studies are in habitual free-living situations using activity counts. There is a trend between the two studies when looking specifically at percent agreement, the agreement at LPA is lower than SB and other intensities of PA (Duncan et al., 2018). However, the highest percent agreement in Duncan et al. (2018) was always higher than the highest agreement in the current study. In both studies VPA had the highest agreement (99.3 vs the current study: 90.1).

Cohens' Kappa coefficients:

Across all the placement sites SB demonstrated the highest Cohens' Kappa coefficients (non-dominant wrist: 0.69, hip: 0.61 and dominant wrist: 0.70), with all being interpreted as demonstrating substantial agreement. SB on the dominant wrist had the best Cohens' Kappa coefficient value across all thresholds and placement sites (0.70; substantial agreement). On average LPA performed the worst across the placements (non-dominant and dominant wrist: 0.21 and hip: 0.15) representing only fair or slight agreement. LPA on the hip performed the worst across of thresholds and placement sites (0.15, slight agreement). The Cohen's Kappa coefficient examines intrarater reliability and this can aid comparisons between studies as this is standardised and the interpretation is consistent (McHugh, 2012). The Cohen's Kappa coefficients observed in this study are slightly lower than observed in previous calibration studies on the hip placement (0.75) (Vaha-Ypya et al., 2018). However, both of these Cohen's Kappa coefficients would be interpreted as demonstrating substantial agreement. Montoye et al. (2016) conducted a validation in laboratory conditions (activities were from SB, LPA and MVPA intensity categories) and participant wore an ActiGraph GT3X+ accelerometer on the hip and GENEActiv on the wrists, which could already account for some differences. Cohen's Kappa coefficients in (Montoye et al., 2016) were better than any observed in the cross validation group (≥0.75). Although, Montoye et al. (2016) used direct observation as the criterion reference,

20 Hz sampling frequency and used weighted Cohen's Kappa coefficients, which allow disagreement to be weighted differently

Previous traditional ROC Curve studies:

One study that used the traditional ROC curve approach in adults was Miller et al. (2015). The optimal thresholds in this study were MPA at 91 m*g* and 414 m*g* for VPA. In the pairs ROC curve analysis approach, the thresholds in the current study were considerably lower (MPA: 126 m*g* and VPA: 328 m*g*). The differences between studies could be because in Miller et al. (2015) only ambulatory (running) activities were completed, whereas the current study had a variety of daily activities. Having only ambulatory activities (such as running) can be viewed as a limitation because they do not truly reflect 'real-life' behaviours (Arvidsson et al. 2019). Previous studies reported calibration studies using only ambulatory activities produce considerably lower estimates of MVPA compared to studies, which mirror daily life. If the thresholds in Miller et al. (2015) and the current study were tested, Miller et al. (2015) could produce an underestimation of MVPA as higher thresholds will have been used.

New ROC Curve approach:

This study followed a different ROC curve analysis approach in comparison to many previous calibration studies, where intensity pairs were used to calculate the thresholds (e.g. when calculating a threshold for SB data were included for SB and LPA only) rather than accelerometer data from across the whole SB-VPA continuum. This method was used as it in theory removed disproportionate amounts of data, which influenced the ROC curve analysis (Obuchowski and Bullen, 2018). The thresholds developed using this approach provided a better estimate of time spent in SB and intensities of PA at the individual and group level analysis as demonstrated by the superior AUC, Cohen's Kappa coefficients and percentage agreement observed. In previous studies, it was reported the

pairs analysis provided a better estimate of time spent in SB, MVPA and VPA (Crotti et al. 2019, in review). The current study agrees with Crotti et al. (2019, in review) as better estimates of time spent in SB and intensities. This ROC curve approach has not been tested in many previous studies, therefore it would be suggested that future calibration studies adopt both approaches and investigate the difference between the two techniques.

Area Under the Curve:

When compared to Hildebrand et al. (2017) a higher AUC for SB in the ROC curve analysis was observed at the hip in the Hildebrand et al. (2017) study in comparison to the present study (0.92 vs 0.73; pairs) suggesting better diagnostic accuracy for SB at the hip in the Hildebrand study. However, when comparing the traditional ROC curve approach (the analysis which included SB and all intensities of PA), in the current study a better AUC for the hip at SB (0.90) a closer to Hildebrand et al. (2017) value was discovered. As Hildebrand et al, (2017) investigated SB thresholds only, it was reported that the AUC was better at the hip compared to the wrist (AUC hip: 0.92 compared to 0.87 at the non-dominant wrist). Looking into the first phase of the current study (using the ROC curve), it disagrees with Hildebrand et al. (2017) as generally the non-dominant wrist performed better in the pairs analysis (the current study: non-dominant wrist: 0.98 compare to 0.73 at the hip). When looking at the all intensities analysis, the hip AUC was better when compared to the non-dominant wrist, which is the same as Hildebrand et al. (2017) (hip: 0.90 compared to the non-dominant wrist: 0.67). When looking into the stepping activities (referred to as MPA in the current study), the hip placement produced higher AUC values on both the pairs and all activity intensities ROC curve approaches (0.93; pairs and 0.73; all intensities). The differences between the current study and Hildebrand et al. (2017) is the criterion reference. The current study used indirect calorimetry and Hildebrand et al. (2017) used another accelerometer (activPAL). The activPAL is a thigh-worn device which uses accelerometer-derived information about thigh position to determine the start and

end of each period spent sitting/lying, standing, and stepping, as well as stepping speed, step counts, and postural transitions (Edwardson et al. 2017). ActivPAL has been validated in other previous calibration studies investigation SB thresholds (using direct observation as the criterion reference) and has produced accurate results. However, the use this accelerometer should be considered as this type of accelerometer is primarily an inclinometer (which specifically measures angles of a slope and elevation) which may present issues when investigating PA as this is not based on postural classifications or when PA is in a SB postural classification (Lyden et al. 2017).

ENMO Thresholds:

Hildebrand et al. (2014) and Hildebrand et al. (2017) used the same data collection methods, however MPA and VPA thresholds were reported in 2014 and SB thresholds were reported in 2017. The participants in this study were a convenience sample of staff and students from the Norwegian School of Sports Sciences with the mean age 34.2 years. However, one difference between the two published studies was the criterion reference: (Hildebrand et al. (2014): indirect calorimetry) and Hildebrand et al. (2017): ActivPAL). The ENMO values in Hildebrand et al. (2014) and Hildebrand et al. (2017) are both on average higher than the ENMO values in the current study. Hildebrand et al. (2014): (hip: 258.7 m*g* and wrist: 428.8 m*g*) compared to the current study: (hip: 225 m*g* and wrist: 315 m*g*), however, the MPA wrist threshold were higher in this study (100.6 m*g* compared to 110 m*g*). Values reported in Hildebrand et al. (2017): hip 18.7 m*g* and wrist 8.4 m*g*, compared to the current study: hip 14.8 m*g* and wrist 6.1 m*g*. Both studies used the same data reduction method by downloading the accelerometer data onto its proprietary software and converting this into raw data files to process through R, which helps with the comparison between studies. However, Hildebrand et al. (2014) used linear regression to establish a relationship between the output and VO_2. The performance of these equations was assessed using 10-cross-validation mode (leave-one-out cross-validation) and the thresholds were calculated from the regression equations. Unlike the

current study, there was no further development or refinement of thresholds (finding the optimal thresholds based on individual and group level equivalency, step 2). In Buchan, Boddy and McLellan (2019), it was reported an ENMO value on the wrist of 233.9 mg. This value is lower than reported in Hildebrand et al. (2017) however is still considerably higher than reported in the current study (116.2 mg when compared on similar treadmill speeds alone). Hildebrand et al. (2017) used derived coefficient from free-living situations matched to the laboratory conditions, in cases were the laboratory data did not hold enough non-movement periods to facilitate direct auto-calibration. No auto calibration or integration of lab coefficients were used in this study, so could account for some of the differences observed. Hildebrand et al. (2017) only used the ROC curve approach, whereas the current study analysed the data two steps further to try and obtain the optimal thresholds. The difference in accelerometer outputs may also be due to the model of accelerometer being different.

The ENMO values recorded in this study for the treadmill activities were much lower than Hildebrand et al. (2017) even though the speeds in the current study were faster (current study: light walk: 124.2 mg, brisk walk: 197.1 mg and run: 552.1 mg), compared to Hildebrand et al. 2017 who only reported one ENMO value for all stepping activities (240.5 m*g*). This is very similar for the non-dominant wrist also; the ENMO values in the current study were substantially lower in comparison to Hildebrand et al. (2017) (354.8 m*g*), as the current study reported 116.2 m*g*, 191.2 m*g* and 659.0 m*g* across all treadmill activities (on the non-dominant wrist placement). However, Hildebrand et al. (2017) also measure free-living stepping and when compared to the current study, is the values were more similar (the current study (non-dominant wrist): 191.2 m*g* and Hildebrand et al. (2017) (non-dominant wrist: 110.8 m*g*). This could suggest participants in the current study walked/ran similarly, to how they would in habitual situations. This could be related to using Fr speeds, developed for each participant individually, instead of standardising the speed for all participants. When comparing the ENMO values to a different study during treadmill activities, the values in the present study are substantially lower.

Resting Metabolic Rate:

It is commonly accepted that 1 MET is 3.5 mL/kg/min in adults (Melzer et al., 2016). However, the mean resting metabolic rate in this study was 4.2 mL/kg/min. The widely accepted 1 MET is 3.5 mL/kg/min has been criticised because resting metabolic rate is individualised due to factors such as age, weight status classification and gender (McMurray et al., 2014). If studies use 3.5 mL/kg/min to represent 1 MET it is likely to misrepresent the expected energy costs of PA in populations. It is therefore recommended that researchers use equations that take into account individual characteristics to estimate resting metabolic rate, if measures such as indirect calorimetry cannot be taken (Melzer et al., 2016).

VO_2 differences:

When comparing VO_2 between calibration studies, the VO_2 in the current study was always higher than Hildebrand et al. (2014). On average, the difference in reported VO_2 between the two studies increased as the intensity of activities increased. The greatest difference between VO_2 values in the studies was during the step activity. On average the VO_2 for participants in the current study was 7.7 mL $O^2.kg^{-1}.min^{-1}$ greater than reported by Hildebrand et al. (2014) (current study: 20.7 mL $O^2.kg^{-1}.min^{-}1$ compared to 13.0 mL $O^2.kg^{-1}.min^{-1}$). However this discrepancy could be due to the indirect calorimeter used. In the current study, the Metamax was used, whereas Hildebrand et al. (2014) used the ergospirometry system. This could reduce the comparability between the two studies as these two machines may process and measure gases differently. The individual machines were calibrated to the manufacturers' recommendations. However, the concentration of gases varied between the two studies which could have impacted the results collected (Hildebrand et al. 2014; 95% oxygen and 5% carbon dioxide, the current study; 5.03%

oxygen and 15.08% carbon dioxide). Without comparing these machines independently it would not be known how they differ in practice.

Criterion References:

A study calibrating raw accelerometer thresholds using indirect calorimetry was Montoye et al. (2016). The mean age of the sample was similar to the one in the current study (22 compared to 21 years). The protocol reported in Montoye et al. (2016) is comparable to the one in the current study, with both ambulatory and non-ambulatory movements. Participants in both studies completed activities such as laundry, sweeping and climbing stairs. Montoye et al. (2016) combined MPA and VPA activities into MVPA, whereas the current study looked into MPA and VPA separately. The participants in Montoye et al. (2016) completed each activities for between 3-10 minutes in a simulated free-living environment, whereas the participants in current study completed each activity for a structured 5 minutes. When comparing the hip location in Montoye et al. (2016) to the final thresholds in the current study, specificity in the current study was always higher than sensitivity, which is the same for Montoye et al. (2016). The sensitivity values in Montoye et al. (2016) for SB (hip: 88.3, left hand: 97.5 and right hand: 93.1) are higher than observed in this study (hip: 0.84; non-dominant: 0.76 and dominant: 0.76). This is the same for LPA and MVPA was better in Montoye et al. (2016) compared to the current study. This is also apparent for specificity; Montoye et al. (2016) had better specificity values compared to the current study. However the discrepancies between Montoye et al. (2016) and the current study could be down to different indirect calorimetry (Oxycon Mobile) and accelerometer (GENEActiv) being used. It should be also noted Montoye et al. (2016) did not use individual MET values and used the standardised 3.5 mL/kg/min. As it was shown in this study, young adults have a higher resting metabolic rate than the standardised 3.5 mL/kg/min and variability was observed between participants. Another difference between the two studies was Montoye et al. (2016) divided the wrists in right or left instead of non-dominant and dominant. This may influence some of the difference

found in Montoye et al. (2016) as the participants who are left handed would be used to doing the activities with this wrist, and may take preference in their daily lives to use this wrist. Whereas, in the current study, this was controlled for and the wrist, dependent on the hand the participants' write with, was analysed as either the non-dominant or the dominant wrist. In the current study, there was slight differences between non-dominant and dominant wrists (approximately 20 m*g* across all activities).

Other studies have used indirect calorimetry as a criterion reference (Staudenmayer et al., 2015), where participants completed structured activities, however, no raw accelerometer threshold were developed. Staudenmayer et al. (2015) study used the Oxycon mobile indirect calorimetry system and the standardised 3.5 mL/kg/min to define 1 MET. There are large discrepancies between some activity and METs values in the current study and Staudenmayer et al. (2015). The only similarity in MET values between the two studies is the activity slow/light walk (3.3 METs in both studies). However, when looking at climbing stairs and "fast"/brisk walk, there is a large difference between values reported (stairs; Staudenmayer et al. (2015): 8.08 vs the current study: 5.1 METs) ("fast"/brisk walk; Staudenmayer et al. (2015): 5.2 vs the current study: 4.3 METs). It should be noted that in Staudenmayer et al. (2015) only minutes 3-5 (each activity was performed for 6 minutes) of each activity was the data analysed. As it was recommended in the previous paragraph, individual MET values should be used instead of the standardised 3.5 mL/kg/min as this value does not take into account individual characteristics such as stature and body mass.

Hip Placement

In the current study, it was determined that the hip placement provided the best estimates of SB and PA. However, as discussed in the literature review (chapter 2) researchers have turned away from hip worn accelerometers as they have been reported in large-

scale studies such as NHANES to have less wear time than data collection cycles with wrist worn accelerometers. Having studies with greater compliance will provide better estimates of habitual behaviours and provides the most reliable estimates of PA levels (Lee, Macfarlane and Lam, 2013; Hassani et al., 2014). Some hip worn accelerometers cannot be worn for 24 hours a day and have to be removed at night time (due to comfort). However, a wrist worn accelerometer can be worn for 24 hours a day and provides the opportunity to examine sleep, which may not be possible with hip worn accelerometers (Jungquist et al., 2015). It is possible when selecting wrist thresholds generated in the current study may reduce the accuracy of PA estimates, as agreement was poorer when compared to the hip. However, overall the MAPE values on the non-dominant and dominant wrist are better than those observed at the hip. Cohens' Kappa coefficients values were also similar between hip (≤0.69), non-dominant (≤0.69) and dominant wrists (≤0.70), therefore though the differences were observed, the magnitude of those differences may not be meaningful in practice.

Wider context:

A recent framework has been release providing a theoretical context for the current study (Keadle et al. 2019). After a rapid growth of new devices and analysis techniques, it is now suggested measurement field adopts a phase-based framework for developing and evaluating device-based methods for physical behaviours assessment. This framework will help facilitate the development and validation of processing methods to predict physical behaviour from devices, which require access to the raw output, rather than already processed summary estimates. The current study has followed the recommended phase I: laboratory development, where the environment is highly controlled. These conditions are optimal for identifying features of the signal that may be valuable for distinguishing between different activity intensities or types. However, these controlled environments are not reflective of real-world conditions and the results will not reflect how a device may perform in a real-world setting. This is due to the transitions between

activities not being included and the change between activities happening at irregular intervals. The intermediate phases (I and II) focus in the development of new methods under controlled laboratory or semi-structured conditions. The thresholds will be more reflective of real-world conditions as they progress along the framework, characterised by the increasing individual variability. Although, the earlier stages do not reflect real-world human behaviour, they are necessary and useful steps in the development process, particularly as a device or type of signal from the device is evaluated for the first time. As the environment and protocol becomes increasingly variable, we expect performance to decline, which may necessitate return to an earlier phase for further refinements and optimisation (Keadle et al. 2019).

The next two paragraphs will discuss the following topics: Machine learning approaches and threshold free techniques. These paragraphs will outline their purpose, how they overcome the issues surrounding thresholds and how they can be implemented.

Alternative approaches

There are alternative approaches to thresholds or cut-points to classify movement behaviours. Machine learning is one of these is one of the emerging techniques used in the discipline and has the potential to be a more accurate method to measure PA than tradition threshold methods (Ahmadi et al. 2019). Machine learning is an area of research concerned with the design and development of algorisms that allow computers to 'learn' from data. Machine learning has the ability to recognise complex patterns and make intelligent decisions based on data is the main focus of this research area (Hagenbuchner et al. 2015). Machine learning has the potential to significantly improve the accuracy of accelerometer-based assessments of PA. Errors from traditional energy expenditure prediction models can be between 25%-50% (Ahmadi et al., 2018). Machine learning approaches are suggested to improve the estimates of accelerometer based PA metrics

(Freedson et al., 2011). Machine learning approaches allow the identification of activity types, which is not possible with simple regression or ROC curve analysis methods. It is reported regardless of the data processing approach, wrist worn accelerometers may still be more vulnerable to misclassification errors when an activity produces significant arm movements (standing folding clothes or cleaning) (Trost, Zheng and Wong, 2014). It is also suggested machine learning approaches at the hip and wrist provided highly accurate recognition of sedentary behaviours (Kantoch, 2018). However, the current study provided highly accurate (≥87% percentage agreement) SB thresholds. It should be highlighted Trost, Zheng and Wong, (2014) was a study investigating children and adolescents. There have been other machine learning studies completed in a variety of populations such as older adults (Sasaki et al., 2016). Machine learning is still a relatively new concept in PA research, so has not been widely adopted. However, it should be noted machine learning requires extensive expertise so the thresholds approach is more valuable to researchers in the PA field without these extensive expertise. Machine learning techniques have traditionally been tested in controlled laboratory conditions, however when these models are tested under free-living conditions, they perform poorly. One reason for this could be due to PA being more incidental and specifically walking could be slower than tested in a laboratory condition. This could “confuse” the algorithms and therefore reducing the accuracy found in the laboratory condition.

To overcome some of the issues of accelerometer data, metrics such as average acceleration and intensity gradient can be adopted (Rowlands et al., 2018a). These metrics provide an indication of PA volume, and a PA intensity profile over a time period. However, it should be noted, these metrics have only been tested in school-aged children, adolescent girls and adults with type 2 diabetes (Rowlands et al., 2018a; Fairclough et al., 2019). A new accelerometer metric can be used to describe the minimum acceleration value above which a person’s most active X minutes are accumulated (Fairclough et al., 2019). This does not rely on thresholds produced in calibration studies and all participants achieve something, compared to the government guidelines, where people either achieve

or do not achieve these recommendations. PA intensity is usually expressed as time based on cut points, which have been developed in validation studies. However, these cut points are dependent on the sample and protocol, which can leave outcome comparisons problematic across population and studies (Rowlands et al., 2018a). This is a relatively new concept and it is unknown how this metric is associated with health and well-being of individuals compared to those which use the government guidelines. Studies that adopt this approach will be highly comparable due to the minimal researcher decision making and the standardisation of the data processing methods, and has been suggested to show key advantages over thresholds (Rowlands et al., 2018a; Fairclough et al., 2019). Even though, this study developed raw accelerometer thresholds, the data is presented in a way to allow the use of this new accelerometer metric in the university student population. Average ENMO and mean MET values (averaged over all data points within the activities period of time) were provided in table 4 in the current study. This will allow comparisons between future studies using non-threshold methods.

Strengths and limitations

The current study has several strengths and limitations which should be highlighted. This study used indirect calorimetry as the criterion reference which is a robust measure of energy expenditure. Individual MET values were calculated for all individuals across all activities. This would have accounted for differing body size and composition. However, it should be highlighted that data collection occurred across a variety of time points during the day, which could have affected the indirect calorimetry data as resting metabolic rate differs throughout the day. Previous literature reported afternoon resting metabolic rate being 100 kcal/day higher than in the morning (Haugen et al., 2003). It would be recommended that the time of data collection is standardised in future studies to account for this circadian variation. The study used multiple accelerometer placement sites, which has allowed the development of thresholds for the most commonly used placements (non-dominant and dominant wrist and hip), which may increase the utility of the thresholds for

practitioners. The study developed raw accelerometer thresholds which will allow transparency and comparisons between studies, whilst also reporting average ENMO and energy expenditure for each activity allowing comparisons to non-threshold derived metrics. The study also used ambulatory and non-ambulatory activities to simulate a range of daily activities indicative of university students' typical behaviours. This allowed for a thorough examination of accelerometer raw data thresholds across a wide variety of activities and intensities. Despite this, the calibration activities took place in a laboratory environment, which therefore influences the ecological validity of the study. Furthermore, the thresholds were not examined in a free-living situation, therefore their accuracy and validity for habitual physical activity monitoring requires further investigation in a free-living setting. In addition, Step 2 of the analysis involved fine-tuning the thresholds- this approach is relatively new and the utility yet to be fully established. It is unknown whether a more traditional 'hold one out' cross validation method would have resulted in superior performance of the resultant thresholds and warrants further investigation in future studies. This study employed the Fr calculation which accounted for the differing limb lengths of participants when completing the treadmill activities, which in theory should help standardise the effort required across participants. An additional limitation is the short length of the data collection period. This did not allow in the use of auto calibration of acceleration signals in GGIR. This may have produced a slight calibration error in the results due to the short protocol length as it is suggested the ENMO metric is sensitive to poor calibration. Future studies should integrate coefficients from similar samples/studies to reduce the potential for calibration errors. Finally, the cohort involved in the study was a convenience sample consisting of healthy individuals, therefore the generated thresholds may only be applicable to similar participant groups.

Conclusion

The thresholds for SB and across all intensities of PA showed acceptable levels of agreement between the accelerometer and criterion reference (Metamax) in all placement sites in regards to specificity, MAPE and percentage agreement. The most equivalent

thresholds in the cross validation group were SB, MPA and MVPA when compared to indirect calorimetry. The non-dominant wrist placement demonstrated the poorest group-level equivalency, with only one threshold being demonstrating group-level equivalence. LPA was the overall worst performing threshold across all the placement sites (lowest sensitivity, specificity and Cohens' Kappa coefficients value). The hip placement generally provided better agreement between criterion reference and accelerometer and therefore could be considered the optimum placement to provide estimates of SB and PA levels of university students. These developed thresholds should be examined in free-living studies to assess validity and accuracy when estimating the activity levels of university students. Lastly, the study protocol, methods and analysis can inform the development of rigorous calibration studies and analysis to determine thresholds in the future for a variety of populations of interest.

Chapter 6: Research Synthesis

The purpose of this short synthesis is to consider the findings in relation to the original aim and objectives, key outcomes, key take home messages and what has been learnt throughout process.

Achievement of aim and objectives:

The main aim of this research study was to calculate and cross-validate accelerometer thresholds to classify PA and SB in university students. This was completed and addressed through one study (Chapter 3, 4 and 5) investigating the following objectives in the same study and chapters:

Objective 1. To design a calibration study to generate accelerometer and energy expenditure data across a range of activities, that simulated daily living in university students.

Objective 2. To calculate and cross-validate accelerometer thresholds to classify university students' SB and PA behaviours generated using a calibration circuit.

Prior to this research being undertaken, there were no raw accelerometer thresholds for use in a university student population. This study has developed these thresholds which can now be taken forward into future research studies. The final thresholds from this study were as follows: SB ranged from <8 to <40 mg, LPA ≥8 to ≥40 m*g*, MPA ≥50 to ≥110, VPA ≥225 to ≥315 m*g* and MVPA ≥50 to ≥110 m*g*.

Table 13: Summary of wear site, threshold and equivalence

Threshold (mg)	Equivalence level
Non-Dominant Wrist	
SB (<35)	±20%
LPA (≥35)	>20%
MPA (≥110)	>20%
VPA (≥315)	>20%
MVPA (≥110)	>20%
Hip	
SB (<8)	>20%
LPA (≥8)	>20%
MPA (≥50)	±20%
VPA (≥225)	>20%
MVPA (≥50)	±15%
Dominant Wrist	
SB (<40)	±15%
LPA (≥40)	±20%
MPA (≥110)	±15%
VPA (≥315)	±20%

Key take home messages:

The thresholds for SB, LPA, MPA and MVPA showed acceptable levels of agreement between the accelerometer and criterion reference in regards to specificity, MAPE and percentage agreement. VPA was the only threshold to show no equivalency between accelerometer and criterion (Metamax) on any placement site at the group level. This study has developed these thresholds, which can now be taken forward into future research studies. This study used the traditional ROC curve analysis approach (where SB and all levels of PA were included in the analysis) and also used a new approach to compare how this affected the thresholds produced (e.g. where SB was paired with LPA and these were the only thresholds used in one analysis). Overall, it was determined the pairs analysis provided a better estimate of time spent in SB and intensities of PA at the individual and group level analysis. This was due to the slightly better AUC, greater Cohen's Kappa coefficients and percentage agreement values observed. These thresholds were carried on into step 2 of the analysis to develop the thresholds further to find the optimal thresholds based on individual and group level equivalency. In this study, it also used Fr numbers to created individual treadmill speeds for each of the treadmill activities. This accounted for the differences in limb length of all participants. This in

theory should have helped standardise the effort required across all participants. Using the indirect calorimetry data, individual resting metabolic rate and MET values for each activity were created and this informed the average intensity of each activity. In this study, it showed the standardised 3.5 mL O2.kg-1.min-1 is lower than the results discovered in this study.

What has been learnt from the process:

There have been many learning processes during the completion of this degree and study. This is the first time I had used specific analysis techniques (e.g. ROC curve and R Studio). This was a learning curve and took some time to get used to and a lot of trial and error. For future projects that require analysis in R, I will need to develop these skills to get the most out of the analysis software. However, for the need of this project, my R skills are what were required. At the start of the degree, I found the transition from a structured taught degree (e.g. BSc) to a research degree (e.g. MPhil) difficult. This degree relied on me to be more independent; however, completing this degree has shaped me to be a better researcher as I can better manage my own time. I have learned a multitude of different personal and professional skills which I will take forward into future research and daily life.

Chapter 7: Recommendations

This study has been the first to develop raw accelerometer thresholds in a university student population. From this study, it is possible for this research to be developed and the recommendations for future research are outlined below:

- Future research should examine the performance of these thresholds in a free-living situation or simulated free-living situation. The simulated free-living situation will allow researchers to observe behaviours (e.g. activities) being completed and use this to assess if the thresholds estimated a similar amount of time (e.g. minutes and seconds) in SB and each intensity of PA.
- The thresholds should be developed using the framework (phase II-IV) in Keadle et al. (2019) as the framework is intended to facilitate the development and validation of processing methods to predict PA behaviours in research devices. The framework is characterised by flexible and progressive processes, prespecified milestones for advancement, and allows return to earlier stages for refinement and optimisation where necessary.
- Another study in university students should use these thresholds to assess and estimate time spent in SB and PA intensities in habitual settings, as this would allow a better understanding of the activity levels of this population.
- Future studies may compare different degree courses and degree types to assess whether this has an impact on time spent in SB and intensities of PA.
- Raw data approaches and/or standardised approaches should be used in further research in this area using accelerometers to increase the comparability between studies.
- It should be recommended to use machine learning approaches to overcome some of the limitations of thresholds.

- Future studies should consider implementing multiple accelerometers to increase the classification accuracy and allow for comparison between models and manufacturers.

Chapter 8: References

Aadland, E. and Anderssen, S.A. (2012) Treadmill Calibration of the Actigraph GT1M in Young-to-Middle-Aged Obese-to-Severely Obese Subjects. *J Obes*, 2012, 318176.

ACSM (2013) *ACSM's Guidelines for Exercise Testing and Prescripition*. Baltimore: ACSM

Ahmadi, M.N., Brookes, D., Chowdhury, A., Pavey, T. and Trost, S.G. (2019) Free-living Evaluation of Laboratory-based Activity Classifiers in Preschoolers. *Medicine & Science in Sports & Exercise.* DOI: 10.1249/MSS.0000000000002221

Ahmadi, M., O'Neil, M., Fragala-Pinkham, M., Lennon, N. and Trost, S. (2018) Machine learning algorithms for activity recognition in ambulant children and adolescents with cerebral palsy. *J Neuroeng Rehabil*, 15 (1), 105.

Ainsworth, B.E., Haskell, W.L., Whitt, M.C., Irwin, M.L., Swartz, A.M., Strath, S.J., O'Brien, W.L., Bassett, D.R., Schmitz, K.H., Emplaincourt, P.O., Jacobs, D.R. and Leon, A.S. (2011) Compendium of Physical Activities: an update of activity codes and MET intensities. *Medicine & Science in Sports & Exercise*, 32 (Supplement), S498-S516.

Alves, A.J., Viana, J.L., Cavalcante, S.L., Oliveira, N.L., Duarte, J.A., Mota, J., Oliveira, J. and Ribeiro, F. (2016) Physical activity in primary and secondary prevention of cardiovascular disease: Overview updated. *World J Cardiol*, 8 (10), 575-583.

Arias-Palencia, N.M., Solera-Martinez, M., Gracia-Marco, L., Silva, P., Martinez-Vizcaino, V., Canete-Garcia-Prieto, J. and Sanchez-Lopez, M. (2015) Levels and Patterns of Objectively Assessed Physical Activity and Compliance with Different Public Health Guidelines in University Students. *PLoS One*, 10 (11), e0141977.

Arigo, D., Pasko, K. and Mogle, J.A. (2019) Daily relations between social perceptions and physical activity among college women. *Psychology of Sport and Exercise*.

Arvidsson, D., Fridolfsson, J. and Borjesson, M. (2019) Measurement of physical activity in clinical practice using accelerometers. *J Intern Med*, 286 (2), 137-153.

Arvidsson, D., Fridolfsson, J., Borjesson, M., Andersen, L.B., Ekblom, O., Dencker, M. and Brond, J.C. (2019) Re-examination of accelerometer data processing and calibration for the assessment of physical activity intensity. *Scand J Med Sci Sports*.

Bai, Y., Welk, G.J., Nam, Y.H., Lee, J.A., Lee, J.M., Kim, Y., Meier, N.F. and Dixon, P.M. (2016) Comparison of Consumer and Research Monitors under Semistructured Settings. *Med Sci Sports Exerc*, 48 (1), 151-158.

Bakrania, K., Yates, T., Rowlands, A.V., Esliger, D.W., Bunnewell, S., Sanders, J., Davies, M., Khunti, K. and Edwardson, C.L. (2016) Intensity Thresholds on Raw Acceleration Data: Euclidean Norm Minus One (ENMO) and Mean Amplitude Deviation (MAD) Approaches. *PLoS One*, 11 (10), e0164045.

Bassett, D.R., Jr., Rowlands, A. and Trost, S.G. (2012) Calibration and validation of wearable monitors. *Med Sci Sports Exerc*, 44 (1 Suppl 1), S32-38.

Boddy, L.M., Noonan, R.J., Kim, Y., Rowlands, A.V., Welk, G.J., Knowles, Z.R. and Fairclough, S.J. (2018) Comparability of children's sedentary time estimates derived from wrist worn GENEActiv and hip worn ActiGraph accelerometer thresholds. *J Sci Med Sport*, 21 (10), 1045-1049.

Calestine, J., Bopp, M., Bopp, C.M. and Papalia, Z. (2017) College Student Work Habits are Related to Physical Activity and

Fitness. *International Journal of Exercise Science*, 10 (7), 1009-1017.

Chastin, S.F., Culhane, B. and Dall, P.M. (2014) Comparison of self-reported measure of sitting time (IPAQ) with objective measurement (activPAL). *Physiol Meas*, 35 (11), 2319-2328.

Clark, B.K., Lynch, B.M., Winkler, E.A., Gardiner, P.A., Healy, G.N., Dunstan, D.W. and Owen, N. (2015) Validity of a multi-context sitting questionnaire across demographically diverse population groups: AusDiab3. *Int J Behav Nutr Phys Act*, 12, 148.

Clemente, F.M., Nikolaidis, P.T., Martins, F.M. and Mendes, R.S. (2016) Physical Activity Patterns in University Students: Do They Follow the Public Health Guidelines? *PLoS One*, 11 (3), e0152516.

CMO, U. (2019) UK Chief Medical Officers' Physical Activity Guidelines.

de Almeida Mendes, M., da Silva, I.C.M., Ramires, V.V., Reichert, F.F., Martins, R.C. and Tomasi, E. (2018) Calibration of raw accelerometer data to measure physical activity: A systematic review. *Gait Posture*, 61, 98-110.

Deliens, T., Deforche, B., De Bourdeaudhuij, I. and Clarys, P. (2015) Determinants of physical activity and sedentary behaviour in university students: a qualitative study using focus group discussions. *BMC Public Health*, 15, 201.

Ding, D., Lawson, K.D., Kolbe-Alexander, T.L., Finkelstein, E.A., Katzmarzyk, P.T., van Mechelen, W. and Pratt, M. (2016) The economic burden of physical inactivity: a global analysis of major non-communicable diseases. *The Lancet*, 388 (10051), 1311-1324.

Dixon, P.M., Saint-Maurice, P.F., Kim, Y., Hibbing, P., Bai, Y. and Welk, G.J. (2018) A Primer on the Use of Equivalence Testing for Evaluating Measurement Agreement. *Med Sci Sports Exerc*, 50 (4), 837-845.

Donnelly, M. and Gamsu, S. (2018) Social, ethnic and spatial inequalities in student mobility.

Duncan, S., Stewart, T., Bo Schneller, M., Godbole, S., Cain, K. and Kerr, J. (2018) Convergent validity of ActiGraph and Actical accelerometers for estimating physical activity in adults. *PLoS One*, 13 (6), e0198587.

Dziak, J.J., Coffman, D.L., Lanza, S.T., Li, R. and Jermiin, L.S. (2019) Sensitivity and Specificity of Information Criteria.

Edwardson, C.L., Winkler, E.A.H., Bodicoat, D.H., Yates, T., Davies, M.J., Dunstan, D.W and Healy, G.N. (2017) Considerations when using the activPAL monitor in field-based research with adult populations. *Journal of Sport and Health Science,* 6 (2), 162-178.

Ekelund, U., Brown, W.J., Steene-Johannessen, J., Fagerland, M.W., Owen, N., Powell, K.E., Bauman, A.E. and Lee, I.M. (2019a) Do the associations of sedentary behaviour with cardiovascular disease mortality and cancer mortality differ by physical activity level? A systematic review and harmonised meta-analysis of data from 850 060 participants. *Br J Sports Med*, 53 (14), 886-894.

Ekelund, U., Tarp, J., Steene-Johannessen, J., Hansen, B.H., Jefferis, B., Fagerland, M.W., Whincup, P., Diaz, K.M., Hooker, S.P., Chernofsky, A., Larson, M.G., Spartano, N., Vasan, R.S., Dohrn, I.M., Hagstromer, M., Edwardson, C., Yates, T., Shiroma, E., Anderssen, S.A. and Lee, I.M. (2019b) Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. *BMJ*, 366, l4570.

Ellis, K., Kerr, J., Godbole, S., Lanckriet, G., Wing, D. and Marshall, S. (2014) A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiol Meas*, 35 (11), 2191-2203.

Fagaras, S.-P., Radu, L.-E. and Vanvu, G. (2015) The Level of Physical Activity of University Students. *Procedia - Social and Behavioral Sciences*, 197, 1454-1457.

Fairclough, S.J., Noonan, R., Rowlands, A.V., Van Hees, V., Knowles, Z. and Boddy, L.M. (2016) Wear Compliance and Activity in Children Wearing Wrist- and Hip-Mounted Accelerometers. *Medicine and Science in Sports and Exercise*, 48 (2), 245-253.

Fairclough, S.J., Rowlands, A.V., Taylor, S. and Boddy, L.M. (2019) Cut-point-free accelerometer metrics to assess children's physical activity: An example using the school day. *Scand J Med Sci Sports*.

Farhud, D.D. (2015) Impact of Lifestyle on Health. *Iran Journal of Public Health*, 44 (11), 1442-1444.

Freedson, P.S., Lyden, K., Kozey-Keadle, S. and Staudenmayer, J. (2011) Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: validation on an independent sample. *J Appl Physiol (1985)*, 111 (6), 1804-1812.

Freedson, P.S., Melanson, E.L. and Sirard, J.R. (1998) Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc*, 30 (5), 777-781.

Gibson, A.-M., Shaw, J., Hewitt, A., Easton, C., Robertson, S. and Gibson, N. (2016) A longitudinal examination of students' health behaviours during their first year at university. *Journal of Further and Higher Education*, 42 (1), 36-45.

Hagenbuchner, M., Cliff, D.P., Trost, S.G., Van Tuc, N. and Peoples, G.E (2015) Prediction of activity type in preschool children using machine learning techniques. *Journal of Science and Medicine in Sport,* 18 (4), 426-431.

Hassani, M., Kivimaki, M., Elbaz, A., Shipley, M., Singh-Manoux, A. and Sabia, S. (2014) Non-consent to a wrist-worn accelerometer in older adults: the role of socio-demographic, behavioural and health factors. *PLoS One*, 9 (10), e110816.

Haugen, H.A., Melanson, E.L., Vu Tran, Z., Kearney, J.T. and Hill, J.O. (2003) Variability of measured resting metabolic rate. *American Society for Clinical Nutrition*, 78, 1141-1144.

Heron, L., O'Neill, C., McAneney, H., Kee, F. and Tully, M.A. (2019) Direct healthcare costs of sedentary behaviour in the UK. *J Epidemiol Community Health*, 73 (7), 625-629.

HESA (2019) Higher Education Student Statistics: UK, 2017/18.

Hildebrand, M., Hansen, B.H., van Hees, V.T. and Ekelund, U. (2017) Evaluation of raw acceleration sedentary thresholds in children and adults. *Scand J Med Sci Sports*, 27 (12), 1814-1823.

Hildebrand, M., VT, V.A.N.H., Hansen, B.H. and Ekelund, U. (2014) Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. *Med Sci Sports Exerc*, 46 (9), 1816-1824.

Hills, A.P., Mokhtar, N. and Byrne, N.M. (2014) Assessment of physical activity and energy expenditure: an overview of objective measures. *Front Nutr*, 1, 5.

Howley, E.T. (2001) Type of activity: resistance, aerobic and leisure versus occupational physical activity. *Medicine & Science in Sports & Exercise*, 33 (6), 364-369.

Hurter, L., Fairclough, S.J., Knowles, Z.R., Porcellato, L.A., Cooper-Ryan, A.M. and Boddy, L.M. (2018) Establishing Raw Acceleration Thresholds to Classify Sedentary and Stationary Behaviour in Children. *Children (Basel)*, 5 (12).

Husu, P., Vaha-Ypya, H. and Vasankari, T. (2016) Objectively measured sedentary behavior and physical activity of Finnish 7- to 14-year-old children- associations with perceived health status: a cross-sectional study. *BMC Public Health*, 16, 338.

Innerd, P., Harrison, R. and Coulson, M. (2018) Using open source accelerometer analysis to assess physical activity and sedentary behaviour in overweight and obese adults. *BMC Public Health*, 18 (1), 543.

Jungquist, C.R., Pender, J.J., Klingman, K.J. and Mund, J. (2015) Validation of Capturing Sleep Diary Data via a Wrist-Worn Device. *Sleep Disord*, 2015, 758937.

Kamada, M., Shiroma, E.J., Harris, T.B. and Lee, I.M. (2016) Comparison of physical activity assessed using hip- and wrist-worn accelerometers. *Gait Posture*, 44, 23-28.

Kantoch, E. (2018) Recognition of Sedentary Behavior by Machine Learning Analysis of Wearable Sensors during Activities of Daily Living for Telemedical Assessment of Cardiovascular Risk. *Sensors (Basel)*, 18 (10).

Keadle, S.K., Lyden, K.A., Strath, S.J., Staudenmayer, J.W. and Freedson, P.S. (2019) A Framework to Evaluate Devices That Assess Physical Behavior. *Exercise and Sport Sciences Reviews*, 47 (4), 206-214.

Kim, Y. and Welk, G.J. (2015) Criterion Validity of Competing Accelerometry-Based Activity Monitoring Devices. *Med Sci Sports Exerc*, 47 (11), 2456-2463.

Kohl, H.W., Craig, C.L., Lambert, E.V., Inoue, S., Alkandari, J.R., Leetongin, G. and Kahlmeier, S. (2012) The pandemic of physical inactivity: global action for public health. *The Lancet*, 380 (9838), 294-305.

Landis, R.J. and Koch, G.G. (1977) The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33 (1), 159-174.

Lazzeri, G., Azzolini, E., Pammolli, A., Simi, R., Meoni, V. and Giacchi, M.V. (2014) Factors associated with unhealthy behaviours and health outcomes: a cross-sectional study among tuscan adolescents (Italy). *International Journal for Equity in Health*, 13 (83), 1-10.

Lee, J.M., Kim, Y. and Welk, G.J. (2014) Validity of consumer-based physical activity monitors. *Med Sci Sports Exerc*, 46 (9), 1840-1848.

Lee, P.H., Macfarlane, D.J. and Lam, T.H. (2013) Factors associated with participant compliance in studies using accelerometers. *Gait Posture*, 38 (4), 912-917.

Lepp, A. and Barkley, J.E. (2019) Cell phone use predicts being an "active couch potato": results from a cross-sectional survey of sufficiently active college students. *Digit Health*, 5, 2055207619844870.

Lyden, K., Keadle, S. K., Staudenmayer, J. and Freedson, P. (2017) The activPAL™ Accurately Classifies Activity Intensity Categories in Healthy Adults. *Med Sci Sports Exerc,* 49 (5), 1022-1028.

Maselli, M., Ward, P.B., Gobbi, E. and Carraro, A. (2018) Promoting Physical Activity Among University Students: A Systematic Review of Controlled Trials. *Am J Health Promot*, 32 (7), 1602-1612.

Matthew, C.E. (2005) Calibration of accelerometer output for adults. *Med Sci Sports Exerc*, 37 (11 Suppl), S512-522.

Matthews, C.E., Hagstromer, M., Pober, D.M. and Bowles, H.R. (2012) Best practices for using physical activity monitors in population-based research. *Med Sci Sports Exerc*, 44 (1 Suppl 1), S68-76.

McGarty, A.M., Penpraze, V. and Melville, C.A. (2016) Calibration and Cross-Validation of the ActiGraph wGT3X+ Accelerometer for the Estimation of Physical Activity Intensity in Children with Intellectual Disabilities. *PLoS One*, 11 (10), e0164928.

McHugh, M.L. (2012) Interrater reliability: the kappa statistic. *Biochemia Medica*, 22 (3), 276-282.

McMurray, R.G., Soares, J., Caspersen, C.J. and McCurdy, T. (2014) Examining variations of resting metabolic rate of adults: a public health perspective. *Med Sci Sports Exerc*, 46 (7), 1352-1358.

Melzer, K., Heydenreich, J., Schutz, Y., Renaud, A., Kayser, B. and Mader, U. (2016) Metabolic Equivalent in Adolescents, Active Adults and Pregnant Women. *Nutrients*, 8 (7).

Migueles, J.H., Cadenas-Sanchez, C., Ekelund, U., Delisle Nystrom, C., Mora-Gonzalez, J., Lof, M., Labayen, I., Ruiz, J.R. and Ortega, F.B. (2017) Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. *Sports Med*, 47 (9), 1821-1845.

Migueles, J.H., Cadenas-Sanchez, C., Tudor-Locke, C., Lof, M., Esteban-Cornejo, I., Molina-Garcia, P., Mora-Gonzalez, J., Rodriguez-Ayllon, M., Garcia-Marmol, E., Ekelund, U. and Ortega, F.B. (2019a) Comparability of published cut-points for the assessment of physical activity: Implications for data harmonization. *Scand J Med Sci Sports*, 29 (4), 566-574.

Migueles, J.H., Rowlands, A.V., Huber, F., Sabia, S. and van Hees, V.T. (2019b) GGIR: A Research Community–Driven Open Source R Package for Generating Physical Activity and Sleep Outcomes From Multi-Day Raw Accelerometer Data. *Journal for the Measurement of Physical Behaviour*, 2 (3), 188-196.

Miller, P.J.O., Vähä-Ypyä, H., Vasankari, T., Husu, P., Mänttäri, A., Vuorimaa, T., Suni, J. and Sievänen, H. (2015) Validation of Cut-Points for Evaluating the Intensity of Physical Activity with Accelerometry-Based Mean Amplitude Deviation (MAD). *PLoS One*, 10 (8).

Minetti, A.E. (2001) Biomechanics: walking on other planets. *Nature*, 409 (6819), 467.

Montoye, A.H.K., Pivarnik, J.M., Mudd, L.M., Biswas, S. and Pfeiffer, K.A. (2016) Validation and Comparison of Accelerometers Worn on the Hip, Thigh, and Wrists for Measuring Physical Activity and Sedentary Behavior. *AIMS Public Health*, 3 (2), 298-312.

Nelson, M.B., Kaminsky, L.A., Dickin, D.C. and Montoye, A.H. (2016) Validity of Consumer-Based Physical Activity Monitors for Specific Activity Types. *Med Sci Sports Exerc*, 48 (8), 1619-1628.

Newtonraj, A., Murugan, N., Singh, Z., Chauhan, R.C., Velavan, A. and Mani, M. (2017) Factors Associated with Physical Inactivity among Adult Urban Population of Puducherry, India: A Population Based Cross-sectional Study. *J Clin Diagn Res*, 11 (5), LC15-LC17.

Nicolson, G., Hayes, C. and Darker, C. (2019) Examining total and domain-specific sedentary behaviour using the socio-ecological model - a cross-sectional study of Irish adults. *BMC Public Health*, 19 (1), 1155.

Obuchowski, N.A. and Bullen, J.A. (2018) Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Phys Med Biol*, 63 (7), 07TR01.

Peterson, N.E., Sirard, J.R., Kulbok, P.A., DeBoer, M.D. and Erickson, J.M. (2018) Sedentary behavior and physical activity of young adult university students. *Res Nurs Health*, 41 (1), 30-38.

Pfister, T., Matthews, C.E., Wang, Q., Kopciuk, K.A., Courneya, K. and Friedenreich, C. (2017) Comparison of two accelerometers for measuring physical activity and sedentary behaviour. *BMJ Open Sport Exerc Med*, 3 (1), e000227.

Rhodes, R.E., Janssen, I., Bredin, S.S.D., Warburton, D.E.R. and Bauman, A. (2017) Physical activity: Health impact, prevalence, correlates and interventions. *Psychol Health*, 32 (8), 942-975.

Romanzini, C.L.P., Romanzini, M., Barbosa, C.C.L., Batista, M.B., Shigaki, G.B. and Ronque, E.R.V. (2019) Characterization and Agreement Between Application of Mobile Ecological Momentary Assessment (mEMA) and Accelerometry in the Identification of Prevalence of Sedentary Behavior (SB) in Young Adults. *Front Psychol*, 10, 720.

Roscoe, C.M.P., James, R.S. and Duncan, M.J. (2017) Calibration of GENEActiv accelerometer wrist cut-points for the assessment of physical activity intensity of preschool aged children. *Eur J Pediatr*, 176 (8), 1093-1098.

Rouse, P.C. and Biddle, S.J.H. (2010) An ecological momentary assessment of the physical activity and sedentary behaviour patterns of university students. *Health Education Journal*, 69 (1), 116-125.

Rowlands, A. and Eston, R.G. (2007) The measurement and interpretation of children's physical activity. *Journal of Sports Science and Medicine*, 6, 270-276.

Rowlands, A.V., Edwardson, C.L., Davies, M.J., Khunti, K., Harrington, D.M. and Yates, T. (2018a) Beyond Cut Points: Accelerometer Metrics that Capture the Physical Activity Profile. *Med Sci Sports Exerc*, 50 (6), 1323-1332.

Rowlands, A.V., Mirkes, E.M., Yates, T., Clemes, S., Davies, M., Khunti, K. and Edwardson, C.L. (2018b) Accelerometer-assessed Physical Activity in Epidemiology: Are Monitors Equivalent? *Med Sci Sports Exerc*, 50 (2), 257-265.

Rowlands, A.V., Rennie, K., Kozarski, R., Stanley, R.M., Eston, R.G., Parfitt, G.C. and Olds, T.S. (2014) Children's Physical Activity Assessed with Wrist- and Hip-Worn Accelerometers. *Medicine & Science in Sports & Exercise*, 46 (12), 2308-2316.

Rowlands, A.V., Yates, T., Olds, T.S., Davies, M., Khunti, K. and Edwardson, C.L. (2016) Wrist-Worn Accelerometer-Brand Independent Posture Classification. *Med Sci Sports Exerc*, 48 (4), 748-754.

Ruopp, M.D., Perkins, N.J., Whitcomb, B.W. and Schisterman, E.F. (2008) Youden Index and Optimal Cut-Point Estimated from

Observations Affected by a Lower Limit of Detection. *Biometrical Journal*, 50 (3), 419-430.

Saint-Maurice, P.F., welk, G.J., Beyler, N.K., Bartee, R.T. and Heelan, K., A. (2014) Calibration of self-report tools for physical activity

research: the Physical Activity Questionnaire

(PAQ). *BMC Public Health*, 14 (1-9).

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

Sasaki, J.E., Hickey, A.M., Staudenmayer, J.W., John, D., Kent, J.A. and Freedson, P.S. (2016) Performance of Activity Classification Algorithms in Free-Living Older Adults. *Med Sci Sports Exerc*, 48 (5), 941-950.

Scarborough, P., Bhatnagar, P., Wickramasinghe, K.K., Allender, S., Foster, C. and Rayner, M. (2011) The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006-07 NHS costs. *J Public Health (Oxf)*, 33 (4), 527-535.

Sedgwick, P. (2014) Cross sectional studies: advantages and disadvantages. *BMJ*, 348 (mar26 2), g2276-g2276.

Shim, J.-S., Oh, K. and Kim, H.C. (2014) Dietary assessment methods in epidemiologic studies. *Epidemiology and Health*.

Staudenmayer, J., He, S., Hickey, A., Sasaki, J. and Freedson, P. (2015) Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. *J Appl Physiol (1985)*, 119 (4), 396-403.

Towne, S.D., Ory, M.G., Smith, M.L., Peres, S.C., Pickens, A.W., Mehta, R.K. and Benden, M. (2017) Accessing physical activity among young adults attending a university: the role of sex, race/ethnicity, technology use, and sleep. *BMC Public Health*, 17 (1).

Tremblay, M.S., Aubert, S., Barnes, J.D., Saunders, T.J., Carson, V., Latimer-Cheung, A.E., Chastin, S.F.M., Altenburg, T.M., Chinapaw, M.J.M. and Participants, S.T.C.P.

(2017) Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. *Int J Behav Nutr Phys Act*, 14 (1), 75.

Troiano, R.P., McClain, J.J., Brychta, R.J. and Chen, K.Y. (2014) Evolution of accelerometer methods for physical activity research. *Br J Sports Med*, 48 (13), 1019-1023.

Trost, S.G., Zheng, Y. and Wong, W.-K. (2014) Machine learning for activity recognition: hip versus wrist data. *Physiological Measurement*, 35 (11), 2183-2189.

Tudor-Locke, C., Barreira, T.V., Schuna, J.M., Jr., Mire, E.F., Chaput, J.P., Fogelholm, M., Hu, G., Kuriyan, R., Kurpad, A., Lambert, E.V., Maher, C., Maia, J., Matsudo, V., Olds, T., Onywera, V., Sarmiento, O.L., Standage, M., Tremblay, M.S., Zhao, P., Church, T.S., Katzmarzyk, P.T. and Group, I.R. (2015) Improving wear time compliance with a 24-hour waist-worn accelerometer protocol in the International Study of Childhood Obesity, Lifestyle and the Environment (ISCOLE). *Int J Behav Nutr Phys Act*, 12, 11.

Vaha-Ypya, H., Husu, P., Suni, J., Vasankari, T. and Sievanen, H. (2018) Reliable recognition of lying, sitting, and standing with a hip-worn accelerometer. *Scand J Med Sci Sports*, 28 (3), 1092-1102.

van Hees, V.T., Fang, Z., Langford, J., Assah, F., Mohammad, A., da Silva, I.C., Trenell, M.I., White, T., Wareham, N.J. and Brage, S. (2014) Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. *J Appl Physiol (1985)*, 117 (7), 738-744.

van Hees, V.T., Gorzelniak, L., Dean Leon, E.C., Eder, M., Pias, M., Taherian, S., Ekelund, U., Renstrom, F., Franks, P.W., Horsch, A. and Brage, S. (2013) Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. *PLoS One*, 8 (4), e61691.

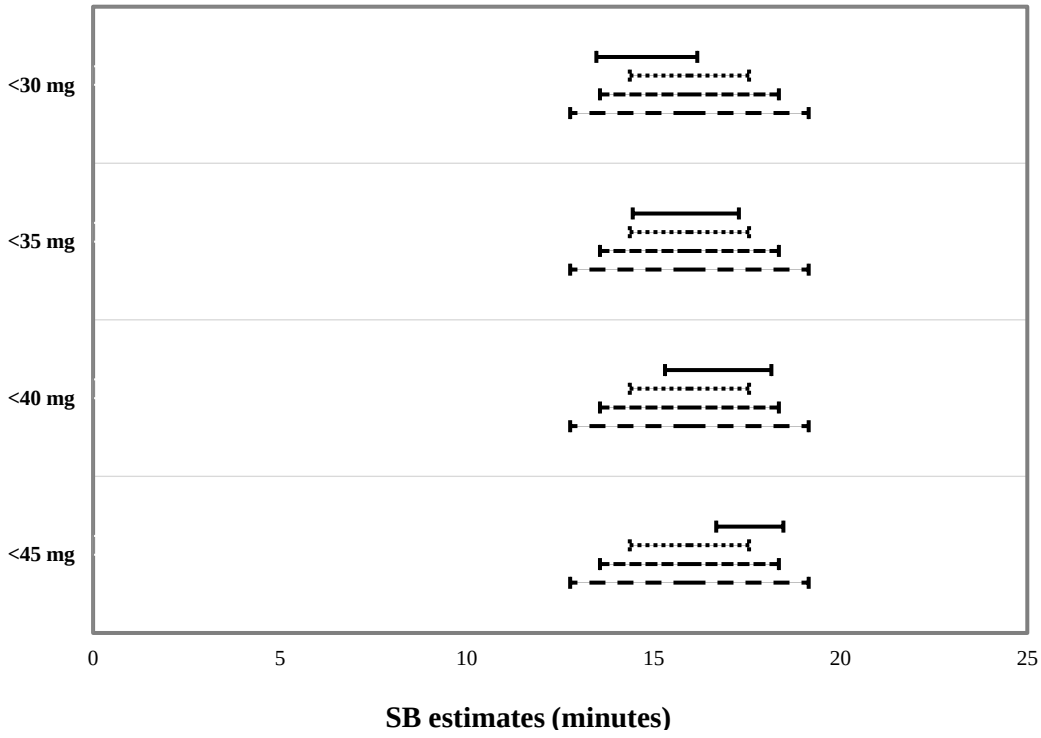
Welk, G.J., Bai, Y., Lee, J.-M., Godino, J.O.B., Saint-Maurice, P.F. and Carr, L. (2019) Standardizing Analytic Methods and Reporting in Activity Monitor Validation Studies. *Medicine & Science in Sports & Exercise*, 51 (8), 1767-1780.

White, T., Westgate, K., Hollidge, S., Venables, M., Olivier, P., Wareham, N. and Brage, S. (2019) Estimating energy expenditure from wrist and thigh accelerometry in free-living adults: a doubly labelled water study. *International Journal of Obesity*.

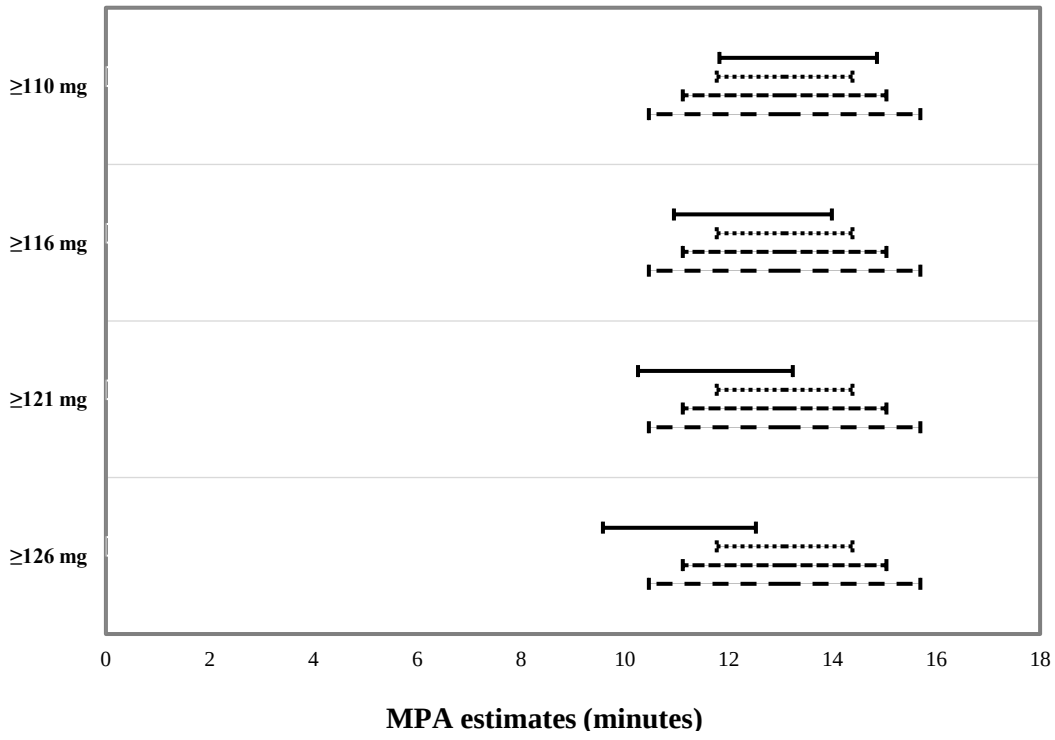
Wilmot, E.G., Edwardson, C.L., Achana, F.A., Davies, M.J., Gorely, T., Gray, L.J., Khunti, K., Yates, T. and Biddle, S.J. (2012) Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. *Diabetologia*, 55 (11), 2895-2905.

Wolpern, A.E., Sherwin, K.J., Moss, W.D., Nygaard, I.E., Egger, M.J., Brusseau, T.A. and Shaw, J.M. (2019) Compliance with wrist-worn accelerometers in primiparous early postpartum women. *Heliyon*, 5 (1), e01193.

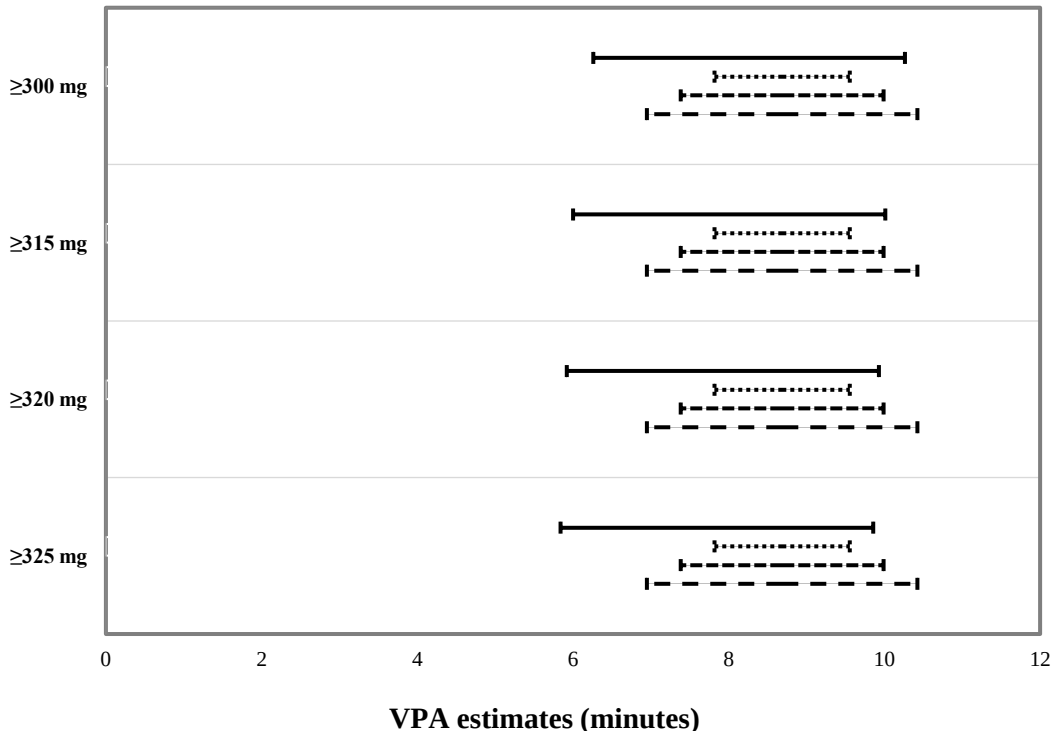
Appendices



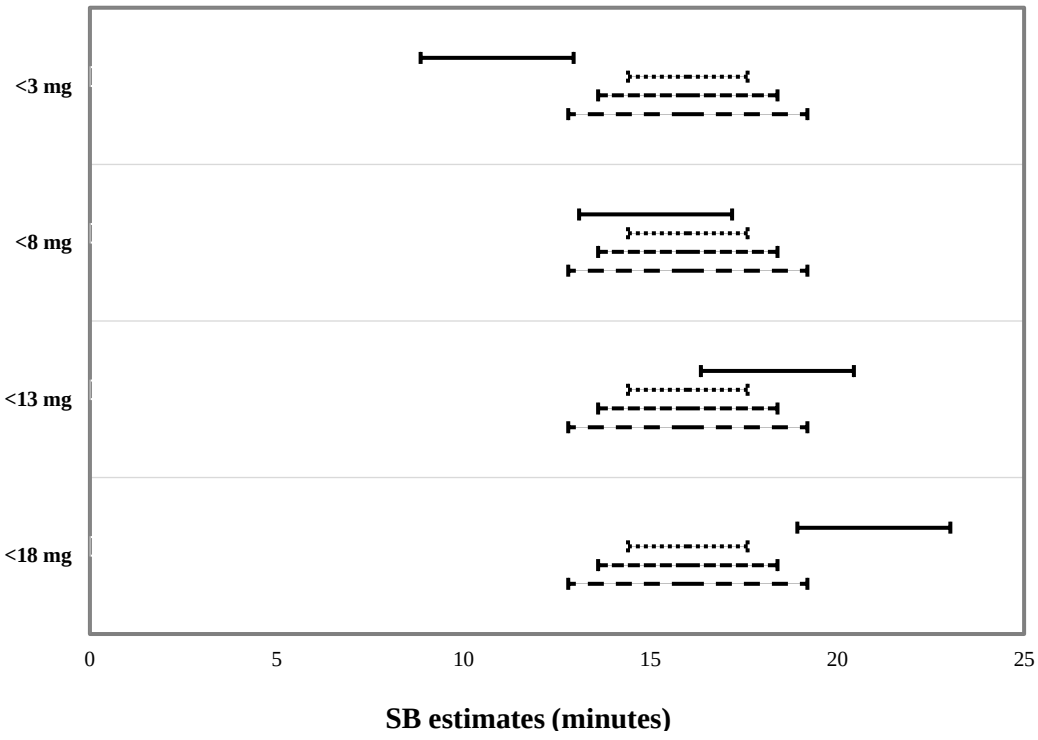
Appendix Figure 1: SB results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) on the non-dominant wrist placement in Step 2. The solid line represents the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference



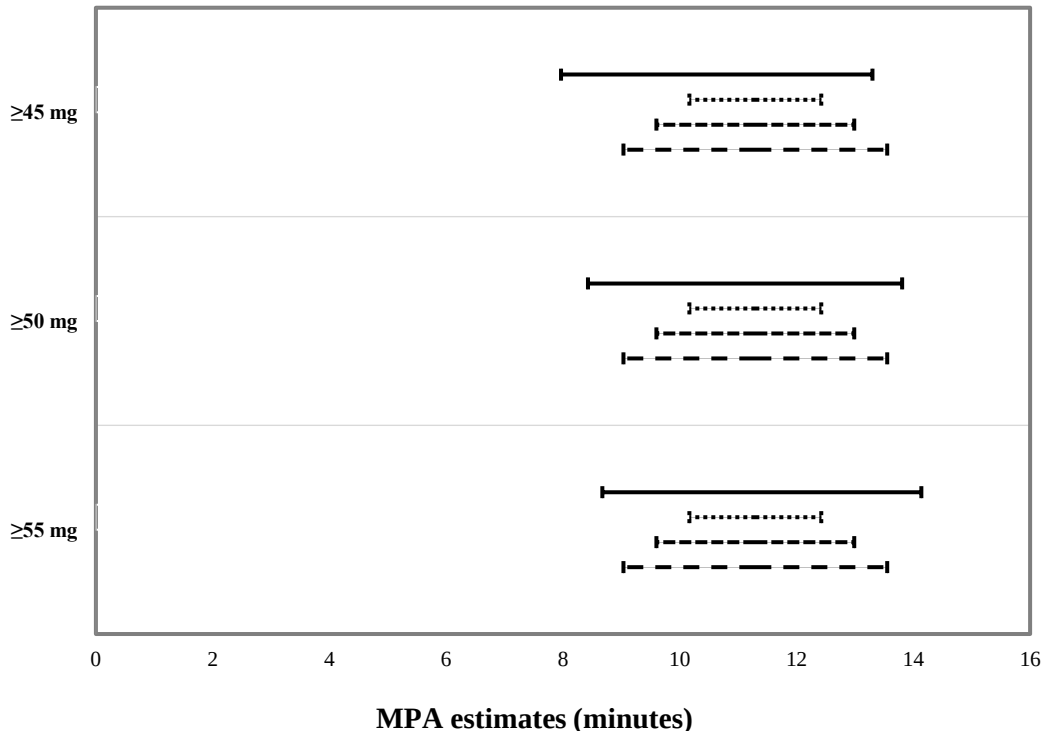
Appendix Figure 2: MPA results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) on the non-dominant wrist placement in Step 2. The solid line represents the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.



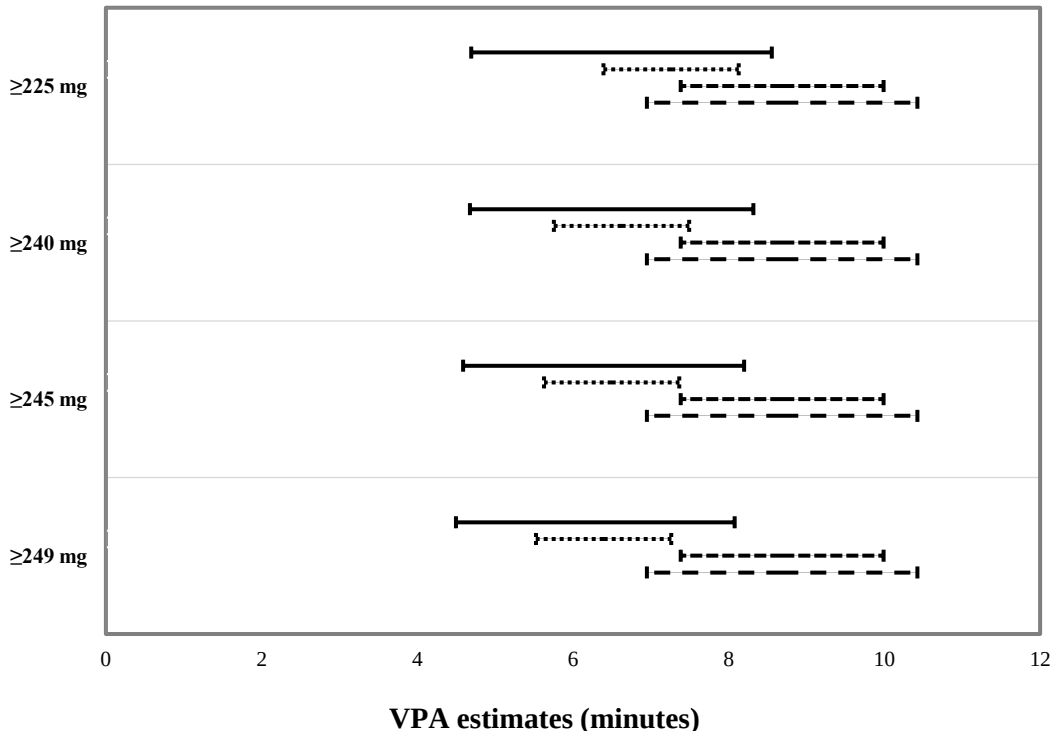
Appendix Figure 3: VPA results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) on the non-dominant wrist placement in Step 2. The solid line represents the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.



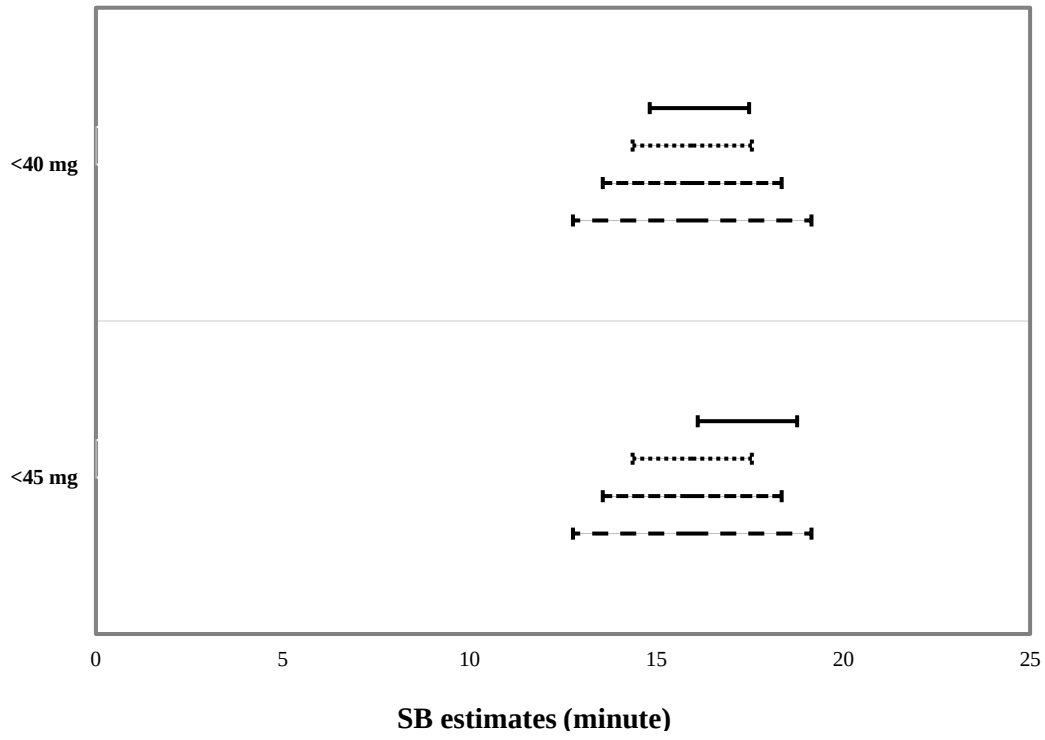
Appendix Figure 4: SB results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) on the hip placement in Step 2. The solid line represents the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.



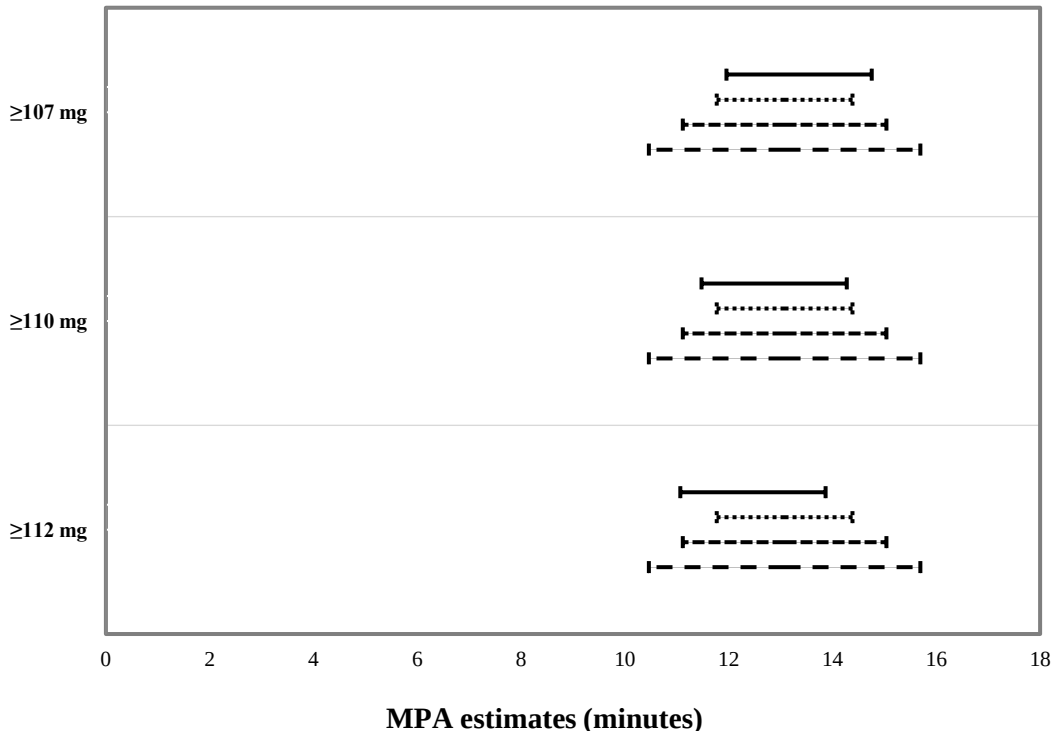
Appendix Figure 5: MPA results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) on the hip placement in Step 2. The solid line represents the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.



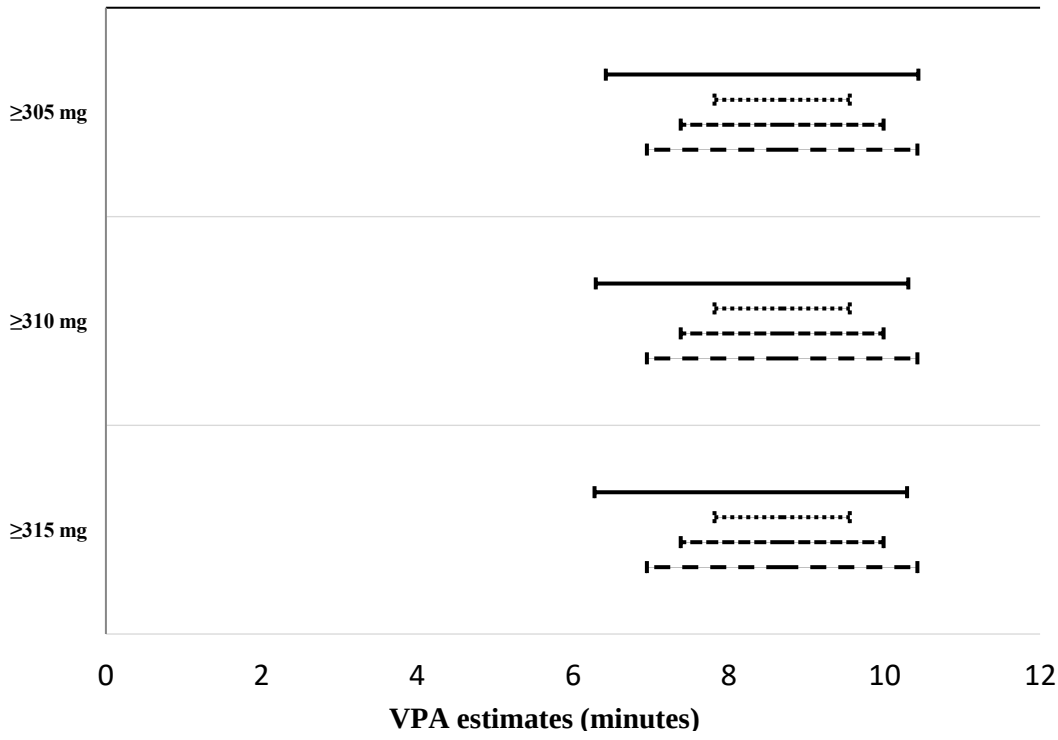
Appendix Figure 6: VPA results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) on the hip placement in Step 2. The solid line represents the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.



Appendix Figure 7: SB results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) on the dominant wrist placement in Step 2. The solid line represents the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.



Appendix Figure 8: MPA results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) on the dominant wrist placement in Step 2. The solid line represents the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.



Appendix Figure 9: VPA results from 95% equivalence testing between thresholds and criterion reference (i.e. Metamax; indirect calorimetry) on the dominant wrist placement in Step 2. The solid line represents the ActiGraph estimates, under which in descending order the dashed lines represent ±10%, ±15% and ±20% equivalency for the criterion reference.

Ethics Email:

Dear Ruth

With reference to your application for Ethical Approval

19/SPS/005 - Ruth Brady, PGR - LABORATORY CALIBRATION OF WRIST WORN ACCELEROMETERS IN UNIVERSITY STUDENTS (Lynne Boddy)

UREC decision: Approved

The University Research Ethics Committee (UREC) has considered the above and I am pleased to inform you that ethical approval has been granted.

Approval is given on the understanding that:

- any adverse reactions/events which take place during the course of the project are reported to the Committee immediately by emailing ethicspr@ljmu.ac.uk;
- any unforeseen ethical issues arising during the course of the project will be reported to the Committee immediately emailing ethicspr@ljmu.ac.uk;
- the LJMU logo is used for all documentation relating to participant recruitment and participation eg poster, information sheets, consent forms, questionnaires. The LJMU logo can be accessed at http://www2.ljmu.ac.uk/corporatecommunications/60486.htm

Where any substantive amendments are proposed to the protocol or study procedures further ethical approval must be sought (https://www2.ljmu.ac.uk/RGSO/93205.htm)

Applicants should note that where relevant appropriate gatekeeper / management permission must be obtained prior to the study commencing at the study site concerned.

Please note that ethical approval is given for a period of five years from the date granted and therefore the expiry date for this project will be **21st January 2024**. An application for extension of approval must be submitted if the project continues after this date.

LIVERPOOL JOHN MOORES UNIVERSITY

Mandy Williams, Research Support Officer

(Research Ethics and Governance)
Research and Innovation Services
Exchange Station, Tithebarn Street, L2 2QP
t: 01519046467 e: a.f.williams@ljmu.ac.uk