MEASURING PHYSICAL ACTIVITY IN OBESE POPULATIONS USING ACCELEROMETRY

Simon Peter Gerrard-Longworth

School of Health Sciences,

University of Salford,

Salford, UK.

2014

Thesis Submitted for the Degree of

Doctor of Philosophy

November 2014

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank, my supervisor Dr Stephen Preece for his continual support and seemingly limitless patience over the past few years. I would also like to thank my second supervisor Professor Lindsey Dugdill for her support and input, particularly in the area of public health; she will be sadly missed. I am grateful to Dr Yannis Goulermas for his advice on activity classification, and for providing a highly useful library of MATLAB functions pertaining to activity classification and feature selection. Thanks go to Steve Horton and Laura Smith of the Human Performance Laboratory for their technical expertise and practical help. I would also like to thank my personal tutor Dr Anita Williams for her support and encouragement in the final stages of producing this thesis.

My family have been very supportive throughout my time as a PhD candidate. I thank my wife, Linda, for her love and understanding throughout this undertaking. Thanks also go to my two young daughters, Anna Lucia and Freya Star, who have often allowed me to work undisturbed, have always provided a hug and a smile, and have been otherwise delightful. I thank my son, Oliver, for his repeated encouragement. My mother deserves thanks for the great support she has given to me and my family throughout this time, and thanks also go to my father who has given some practical assistance along the way.

External Funding

The initial two years of my PhD programme was funded by SSHOES: European Community's Seventh Framework Programme (FP7/2007-2013) under SSHOES project, Grant Agreement no. NMP2-SE-2009-229261

I performed a certain amount of research to meet some of the objectives required by this project. The research was not directly related to this thesis and therefore is not documented here.

Abstract

The thesis is concerned with objectively measuring human physical activity through accelerometry, and compares the effectiveness of algorithms between obese and non-obese groups. The thesis comprises three studies:

Classification of Aerobic and Gym-based Exercises from Accelerometer Output. This study investigated whether accurate classification could be achieved from hip- or ankle-mounted accelerometers for a programme of aerobic exercises and free-living activities. It also examined whether accuracy was affected by obesity, and whether a single classifier could be applied across BMI groups. The study achieved high classification accuracies (85% for hip and 94% for ankle) for both obese and normal BMI groups using the same approach across groups.

Walking Speed Estimation Using Accelerometry. This study aimed to develop a speed estimation model that was applicable across BMI groups, and which utilised a hip-mounted accelerometer. To achieve this, multiple accelerometer signal features were evaluated for use in a linear speed estimation model, and performance was compared between obese and normal BMI groups. The speed estimation algorithm achieved overall RMSE of $0.08ms^{-1}$ for a mixed BMI group, which is comparable with previous research using homogeneous groups.

Prediction of Energy Expenditure from Accelerometer Output. This study aimed to identify physiological and anthropometric parameters for use in an improved energy expenditure estimation model. Model performance was tested on a mixed BMI group. The energy expenditure prediction model incorporating subject attributes showed around 20% improvement over the standard model.

This research found that current approaches to activity classification using accelerometry are equally applicable to obese groups and normal BMI groups. Walking speed prediction was shown to be possible from a hip-mounted accelerometer for both obese and normal BMI groups. Energy expenditure estimation is improved by including subject-specific parameters in the prediction model. Accelerometry is, therefore, a suitable tool for measuring different aspects of physical activity for obese and mixed BMI groups.

1 Introduction

Obesity prevalence has reached the level of a global epidemic, with over half a billion adults worldwide being obese in 2008, the figure having nearly doubled since 1980 (1). In 2012, the annual Health Survey for England reported that around a quarter of adults were obese (2). There are 2.8 million deaths each year due to diseases that can be attributed to being obese or overweight (1) including diabetes, heart disease, many cancers, and stroke. Obesity is consequently an increasing drain on public health resources (3), and estimated costs to the NHS due to obesity and overweight were as great as £4.2 billion in 2007 (4). Indirect costs of obesity in the United Kingdom, such as reduced productivity experienced by businesses due to absenteeism and increased reliance on state benefits for obesity related conditions, were estimated at £15.8 billion and up to £6 billion respectively (4). Similarly in the United States, where in 2008 obesity prevalence was at 33.8% and more than two thirds of the populace was overweight (5), estimates of obesity-related medical costs for 2008 were as high as $147 billion per year (6).

Causes of obesity on a wide scale are complex. In their report Foresight Tackling Obesities: Future Choices Project (4), Butland et al. describe the following factors which help to create an “obesogenic environment”: the biological imperative to accumulate and store energy from food; modern high rates of food production; economic forces determining which foods are manufactured, leading to use of cheaper, unhealthy ingredients; a greater reliance on mechanised transportation and other technologies which reduce the need for physical effort; and social, psychological and environmental influences on food and physical activity choices. The report also identifies four key determinants of obesity: “the level of primary appetite control, the force of dietary habits, the level of physical activity and the level of psychological ambivalence” (4). Other studies have specified that a decline in physical activity levels (7) and large food portions are contributing to the apparent obesity epidemic (8).

At the level of the individual, weight gain results from a net intake of energy (through diet) greater than that which is expended (through metabolic processes and physical activity). Obesity is a consequence of this state persisting over a prolonged period where, gradually, excess food energy is turned into body fat (9). To reduce obesity in individuals there are, therefore, two main approaches to consider: reducing calorie intake through diet, and increasing energy expenditure through physical activity. Research shows that either of these approaches alone can cause weight loss in individuals (10), though a combination of the two is the most effective (11-12). Diet, however, is beyond the scope of this thesis which focuses on the objective measurement of physical activity.

It is widely accepted that the benefits of physical activity (PA) are many, and that PA has a positive effect on numerous health outcomes. It has also been shown that PA protects against illnesses associated with obesity (13). PA is recommended for use in weight management by public health bodies worldwide including the ACSM (14) and American Heart Association (14). Exercise has been shown to aid weight loss (15-16) and reduce body fat (17). For adults aged 18-65 years, at least thirty minutes per day of moderate-intensity activity for at least five days per week is recommended by the National Institute of Clinical Excellence (NICE) (18), and to prevent obesity forty-five to sixty minutes of moderate-intensity activity is recommended (18). Similarly, the U.S. Department of Health and Human Services recommend at least 150 minutes per week of moderate-intensity, or 75 minutes per week of vigorous-intensity aerobic exercise (19).

Many studies have investigated the affects of physical exercise interventions on a number of health outcomes for the overweight and obese. Outcomes have included changes in weight (20-32) and BMI(33-34), changes in cholesterol levels (29, 35), and changes in blood pressure(36-37) . These interventions have involved several different activities including overground walking at both brisk (21, 35, 38) and slow speeds (39-40), treadmill walking (20, 25), jogging (30, 41), cycling (exercise bicycle) (20, 24, 39, 42), rowing (rowing machine) (43), stair stepping (20, 24), stretching (37), dance (44), resistance/weight training (23, 26-27, 45-46), step aerobics (26, 47) and calisthenics (48). These interventions have prescribed between 15 and 60 minutes of exercise per day for between three and five days per week. Some have specified intensities of between 40% and 70% of maximal aerobic capacity, or between 40% and 85% of maximum heart rate.

The physical activities undertaken by participants in interventions such as those discussed above, are measured in a number of ways. Where exercises take place in the laboratory or are supervised by a researcher (20, 34, 49), physical activity may be quantified using a variety of measuring equipment, and there is no question as whether the programme of exercise has been adhered to. Before the advent of body-worn activity monitors, exercise undertaken in the field was measured subjectively by the participants themselves. However, self-reporting methods such as activity diaries, questionnaires and interviews have been shown to be unreliable (50-52) due to their reliance on participants' ability to recall their PA accurately. Also, self-reporting overweight and obese subjects have been found to overestimate their activity levels (53-54). Another limitation of these self-reporting tools is a lack of precision when quantifying activity intensity. For example, an individual is not able to estimate the number of calories they have used or their average walking speed. It is important to be able to accurately measure physical activity in order to assess health outcomes according to the activity patterns of individuals, and subjective tools are not adequate for this purpose.

An objective alternative to self-reporting, able to provide data on the type, intensity, frequency and duration of activities performed by individuals, would allow a researcher to more fully evaluate the effectiveness of interventions prescribing PA to obese individuals. Body-worn activity monitors have demonstrated the ability to provide objective measurements of several aspects of PA, and therefore have the potential to become such a tool. Accelerometry is a key technology employed within activity monitors. Accelerometer-based monitors can record the accelerations of a person's body segments over a period of time, and by applying various analytical techniques to these data, quantities of PA undertaken can be estimated. There are generally two main approaches to quantifying PA from accelerometer output: the first is to estimate energy expenditure (EE) through correlating accelerometer output with a proxy measure of EE such as oxygen consumption (discussed in section 2.5); the second is to identify which types of activity have been undertaken by the participant by applying machine learning techniques to accelerometer output (discussed in section 2.3). There have been several studies investigating these approaches to measuring PA, but the subject groups involved have generally been of lower or mixed BMIs. Furthermore, there has been little research on whether the existing methods of quantifying physical activity through accelerometry need to be modified for obese populations.

The ultimate goal of activity monitoring research is to produce an objective tool capable of reliably measuring aspects of individuals' physical activity under free-living conditions, obviating the need for subjective input. Current research into accelerometry, however, is still some way off delivering such a tool. This is partially because there are a large variety of activities that humans perform in free-living conditions, and algorithms developed in the laboratory are not currently able to account for all of these. Also, physical and physiological differences between individuals make current algorithms inadequate in a varied population.

The thesis aims to address some of the potential challenges presented when measuring different aspects of physical activity using accelerometry. To this end, a scenario is envisaged where accelerometry may be applied to measuring PA in individuals undertaking a weight loss programme involving several prescribed exercises. Such an application would need to distinguish both the exercises and free-living activities from accelerometer data. Additionally, as walking is the most common physical activity and the most frequently prescribed for weight loss, the thesis also considers the measurement of walking speed and energy expenditure. Additionally, the thesis considers the needs of epidemiological studies which may involve subject groups that represent a cross-section of the wider population, and are consequently involve individuals of different BMIs. It should be noted that a high BMI, although generally indicative of high adiposity levels, may also result from high muscle mass, which further illustrates the differences between individuals and how these differences need to be accounted for by the algorithms which provide measures of physical activity from accelerometers.

The thesis investigates how accelerometry may be applied to objectively measure PA, particularly with respect to obese populations. There are three areas of measurement under consideration, each of which is investigated by a separate study as follows: classification of activity type, estimation of walking speed, and prediction of energy expenditure. These areas relate to the type and intensity of activities, and measurements of the frequency and duration of activities are an easily obtained secondary outcome provided by this analysis. These studies when considered *in toto* present accelerometry as a single tool able to provide measures in each of the core areas of PA. Furthermore, such a tool would provide a rich amount of data on physical activity patterns under free-living conditions. This data would aid epidemiological research aimed at better understanding the types of behaviours which affect obesity, and may also help identify risk factors associated with obesity.

The studies described in the thesis have two particular practical considerations which have informed their design. First, the types of activities used to test the classification algorithm have been chosen to reflect those that an obese person may perform under free-living or as part of a weight loss programme. Walking has additional focus as it is the most commonly performed physical activity, and is recommended for weight loss. Second, accelerometer placement has been chosen primarily at the hip, as this is an unobtrusive position unlikely to interfere with natural movement and daily living.

The rest of the thesis is structured as follows.

Chapter 2 reviews previous literature relevant to the thesis, and first discusses currently available methods of quantifying physical activity (section 2.1) including accelerometry. This is followed by a discussion of how accelerometer output may be affected by obesity (section 2.2). Sections 2.3, 2.4 and 2.5 discuss techniques and previous research related to the three studies presented in the thesis in chapter 3 (activity classification), chapter 4 (walking speed estimation) and chapter 5 (energy expenditure prediction) respectively.

Activity Classification:.

An overview of current machine learning techniques used to classify physical activity from accelerometer output, and related work in this area, is presented in 2.3. Chapter 3 describes a study where classification techniques are applied to a mixed BMI group and classification accuracy is compared between obese and non-obese subgroups.

Walking Speed Estimation:

In section 2.4 current methods of estimating walking speed from accelerometer output are discussed and compared with alternative approaches. Chapter 4 describes a study where a walking speed estimation model is identified which can be applied across BMI groups, and results are compared between obese and normal BMI subgroups.

Energy Expenditure Prediction:

Section 2.5 reviews the current models used to predict energy expenditure from accelerometer output, and discusses a number of physiological parameters which may be added to the model to improve prediction accuracy. Chapter 5 describes a study which identifies an energy expenditure prediction model that incorporates these additional parameters, and compares performance with traditional models for a mixed BMI group.

These studies represent a step towards an integrated and comprehensive PA measurement system using accelerometry, which may be used under free-living conditions for applications such as weight loss interventions and epidemiological studies. This is further discussed in Conclusions (chapter 6).

2 Background and Literature Review

2.1 Methods of quantifying PA

There are four principal characteristics of physical activity: intensity; type; duration; frequency (55). Methods of measuring and quantifying each of these characteristics are necessary in order to obtain a comprehensive picture of individuals' activity patterns under free-living conditions. There are two available approaches to obtaining measures of physical activity. The first is to rely on self-reporting methods such as questionnaires (51), and the second is to apply objective measurement tools such as pedometers (56) or indirect calorimetry (57).

The intensity of physical activity is commonly quantified using a measure of energy expenditure such as the number of calories or METs[1] that have been expended (58). There are a number of methods that may be used to obtain an individual's energy expenditure for a period of activity, these are discussed in subsection 2.1.2.

The concept of what constitutes an activity *type* varies depending on the particular motivation behind measuring the activities of individuals. For example, the activity of "sitting reading a book" is a subcategory of "sitting" which in turn may also be regarded as a subcategory of "sedentary behaviour". Within the context of measuring exercise behaviour the degree of detail required may allow "sitting" to sufficiently describe the activity, whereas a neuroscientist may wish to distinguish between "sitting reading a book" and "sitting watching television". Duration also plays a part in defining an activity type. For example, an individual may take a few steps between products when shopping in a supermarket, but this may or may not be considered to be "walking". Similarly, walking at different speeds may be considered as a single activity or may be broken down into categories such as "brisk walking" and "slow walking".

2.1.1 Subjective methods of measuring PA

With the absence of a gold standard for objective activity measurement under free-living conditions, subjective self-reporting methods have been necessary to obtain information on activity patterns. Activity diaries have been used (59), and questionnaires were devised to extract data on physical activity (60-61). The Compendium of Physical Activities (58) is a resource which can be used in conjunction with physical activity questionnaires to quantify

[1] One MET (metabolic equivalent) represents the energy used by an individual when at rest.

the recorded activities in terms of METs. Contained within the compendium are a wide variety of activities which have been described, categorised and coded, and each has been given a corresponding energy cost in METs.

Due to their reliance on subjects accurately recalling their physical activities, questionnaires have been found to be unreliable in giving accurate measures of PA (50-52). Vigorous activities have been sometimes found to be more accurately measured by questionnaire than lower intensity activities (62), which are typically underestimated due to the questions being unable to capture certain types of activity (63). In contrast, a study by Boon et al. (64) compared data from two PA questionnaires and concluded that both significantly overestimated activity levels. Similarly, activity diaries can also underestimate energy expenditure (65).

In addition to the questionable reliability of activity diaries and questionnaires, they are also limited to broad descriptive categories of intensity. Walking speeds, for example, are reduced to categories such as “slow”, “normal”, “brisk”, and “fast”, and energy expenditure is limited to “light”, “moderate” and “vigorous”. The ability to obtain more precise estimates of intensity would better inform studies investigating the effects of physical activity on health outcomes.

2.1.2 Objective Measures of PA

Objective approaches to measuring PA are traditionally limited to quantifying energy expenditure (EE). EE is commonly measured through indirect calorimetry, which is an approach that measures the amount of oxygen consumed and/or carbon dioxide produced by respiration and subsequently uses this data to estimate EE through standard models (66). The gold standard for EE estimation is the doubly labelled water (DLW) technique which is used to obtain the average EE for a given period of time. The efficacy of using DLW for EE estimation in human subjects was first demonstrated by Schoeller and Van Santen in 1982 (57). For this technique, subjects ingest a quantity of an unnatural isotope of water such as deuterium oxide ($D_2{}^{18}O$). The human body’s respiratory process uses the oxygen isotope from the ingested water when producing carbon dioxide, and the deuterium is primarily lost through urination. By periodically sampling urine, estimates of CO_2 production can be calculated by considering the amount of the original water isotope remaining in samples. From this, oxygen consumption, and therefore energy expenditure may be obtained. However, although the DLW technique is accurate, it is expensive to implement, due to the cost of the isotopes involved, and cannot breakdown energy expenditure for different activities.

Breath-by-breath gas analysis techniques can also be used to quantify EE under laboratory conditions. With this approach experimental subjects wear a gas mask linked to an analysis system which measures oxygen consumption and CO_2 production. Analysis of changes in gas concentrations enable the measurement of energy cost of individual activities provided a steady state can be reached, which takes around one to four minutes (67). Although effective, this approach is impractical under free living conditions as it requires the subject to be connected to cumbersome equipment, and the mask may be uncomfortable and limits daily activities such as eating and drinking. Room calorimetry (68) can be a better alternative to breath-by-breath analysis as the subject is not constrained by measuring equipment. Instead subjects spend time within a sealed room where the incoming air is managed and measured. Air samples are periodically taken and analysed to obtain EE estimates. This is also an effective approach to quantifying PA but is clearly unsuitable for use in studies incorporating free-living activities as subjects are constrained to the test environment.

Body-worn activity monitors may provide a practical solution to objectively quantifying PA in a free-living environment. Over recent years, they have become small and unobtrusive enough to be worn by an individual for long periods of time. The widespread availability of these devices, their size, and their relatively low cost makes them well suited to measuring PA for clinical interventions and epidemiological research. Activity monitors record continuous data collected through on-board sensors such as accelerometers, gyroscopes, and magnetometers. Accelerometers record body acceleration data in up to three dimensions, gyroscopes measure changes in orientation and the angular velocity of body segments, and magnetometers measure absolute orientation in relation to the Earth’s magnetic field. However, the latter two sensors have limited applications in activity monitoring and feature less frequently in the literature than accelerometers. Additionally, gyroscopes consume more power than accelerometers, which means that operational times are lower when running from battery. This thesis focuses on the use of accelerometers to measure physical activity. Compared with a device such as a pedometer, which merely keeps a running count of how many times a threshold has been exceeded, accelerometers return rich data. Tacitly contained within these data is information relating to the different characteristics of PA, and through analysis of these data, estimates may be made for measures of PA.

An accelerometer is an electronic device which measures accelerations in relation to a single axis. A triaxial accelerometer combines three orthogonally placed accelerometers to enable measurement of acceleration in three dimensions. Accelerometers may be attached to a particular body site of an individual. The accelerations measured are those undergone by the body segment to which the accelerometer is affixed. Most accelerometers used in activity monitoring utilise a piezoelectric sensor to measure acceleration (69). Deformation in a piezoelectric element generates a voltage in proportion to the magnitude of the force (70). This voltage is sampled at discrete intervals and stored onboard as a digital signal.

Accelerometry has been applied to measuring many aspects of human activity. A common application is to estimate energy expenditure from the accelerometer signal (71-78). Another is to identify postures (79-81) and the types of physical activity that are being performed (82-89). Accelerometry has also been applied to identifying gait parameters (90-94) such as cadence, speed and step length. Other applications include falls detection (95-98) and assessing balance (99-100). This thesis is concerned with activity classification, walking speed estimation, and energy expenditure prediction, and methods of measuring these though accelerometry are explored in the following chapters.

2.1.2.1 The Actigraph GT3X+

The Actigraph GT3X+ activity monitor (Actigraph LLC, Pensacola, FL, USA) (Figure 1) was chosen for use in the studies described in this thesis. Actigraph accelerometers have featured in many studies involving quantifying physical activity (101-104). At the time of writing, the GT3X+ is the current model in a range of Actigraph activity monitors that have been used in research for the past several years, and has been validated by several studies (105-108). A number of alternative activity monitors were considered for use, such as PAL Technologies ActivPAL, the Philips Actical, and the Tri-Trac RT3. However, the overall specification of the GT3X+ was superior to these alternatives – in terms of sample rate, onboard memory size, acceleration range, and battery life – and this, coupled with the popularity of the Actigraph devices in previous literature made it suitable for the studies described below.

Figure 1: The Actigraph GT3X+

The GT3X+ is able to sample accelerations of ± 6G at between 30Hz and 100Hz. It has sufficient battery life to collect data for a maximum of around one month on a single charge, and has enough internal memory for a maximum of around 42 days. When sampling at 50Hz estimated battery life and memory capacity converge at around 24 days. The GT3X+ specification describes accelerometer data being available in “raw” form. However, consultation with Actigraph revealed that there is some proprietary pre-processing of the acceleration signal that occurs on-board the device, for which they were unwilling to give details. This is likely to be an anti-aliasing filter, which aims to remove signal frequencies that are beyond that of the sampling frequency. For the purposes of the studies below, the signals were regarded as raw.

2.1.3 Terminology for the Accelerometer Coordinate System

The studies in this thesis each utilise one or more Actigraph GT3X+ activity monitors affixed to the waist and/or the hip of the study participant. When the devices are fitted the participant is standing, and the three axes on the devices are aligned as closely as possible with the vertical, anteroposterior and mediolateral directions in relation to the individual. Throughout the thesis, this original alignment is used as a naming convention when referring to an accelerometer axis or signal. For example, the accelerometer axis that was initially aligned to the vertical in relation to an individual will still be referred to as such if the individual is lying down, but now the vertical accelerometer axis would be horizontal in relation to the absolute vertical as determined by gravity. In chapters 4 and 5 the vertical, anteroposterior and mediolateral accelerometer axes are also labelled X, Y and Z respectively. Where the term ‘vertical’ is used without reference to the accelerometer signal or accelerometer axis, then this refers to the absolute vertical.

2.1.4 Activity Monitor Placement

It was intended that a single activity monitor should be placed at an appropriate body site in order to measure the different aspects of physical activity under consideration in the three studies presented in the thesis. That is, activity type, energy expenditure and walking speed should be obtained from accelerometer data collected at a single site, which allows the potential of producing measurements in each area using the same dataset. There were a number of candidate body sites such as the ankle, wrist, hip/waist, and chest. The chosen site was a compromise between the level of burden to the wearer, and its utility in being applied to measuring these aspects. The hip was considered a lesser burden than the chest and ankle, and there was an abundance of literature in these areas for hip and waist which exceeded that found for chest and ankle. The wrist would be preferable in terms of burden, but based on the literature this placement did not appear best suited for measuring all the aspects of physical activity under consideration in the thesis, this is discussed further below.

In recent years large scale epidemiological research has shifted focus to collecting data through wrist-mounted accelerometers, as in the case of UK Biobank and NHANES, and for the general public there has been increased interest in commercial activity monitoring systems using wrist-mounted devices. The primary reason that the wrist has been chosen for such studies has been to increase compliance, as the wrist is already a familiar placement site for watches and bracelets, and places little burden on the wearer. However, the characteristics of data collected at the wrist are significantly different from those collected at the hip, as the arm is able to move independently to the torso and, therefore, the wrist-mounted accelerometer may be subject to two kinds of body movement simultaneously. This may mean that certain aspects of physical activity may not easily be inferred from a wrist-mounted accelerometer signal. For example, when considering estimation of walking speed from accelerometer output, there are established speed-prediction algorithms in the literature based on the movement of the centre of mass (CoM). Such algorithms may utilise hip-mounted accelerometer data due to the proximity of the hip to the CoM. However, it is not likely that CoM movement may be estimated from the wrist – particularly if the wearer were to use a mobile phone or to eat while walking. This is not to say an alternative approach may not be applicable for wrist data. Additionally, wrist accelerometers may not be suitable to accurately estimate energy expenditure, as shown in a study by van Hees et al. who found that acceleration explained only 24% of the variation in physical activity EE at best (109). For the studies in this thesis, the hip was chosen as the primary accelerometer site as a compromise between the burden to the wearer and its use in previous research relevant to the three studies described in this thesis.

2.2 Effects of Obesity on Accelerometer Output

The ability to classify physical activities using accelerometry relies on the premise that accelerometer signals exhibit common characteristics for like activities, and that different activities generate their own distinctive signals. Similarly, in order to estimate energy expenditure from accelerometer output, the signal magnitude must increase with the intensity of an activity. Given that these premises are true for an individual, there is still the question as to whether a single approach to PA measurement can be applied to a wider population due to differences between individuals. Specifically, it is not known if signals generated by obese and normal BMI groups are sufficiently similar to allow such an approach.

There are two main reasons to postulate that the same approaches to PA measurement using accelerometry may not apply across BMI groups. First, the movements of obese individuals differ from their normal BMI counterparts, and this may result in significant differences between accelerometer signal characteristics obtained from obese and normal groups. Second, an accelerometer may experience unwanted movement due to excess adipose tissue at the body site where it is placed, which may introduce noise to the signal. If the accelerometer signal characteristics diverge significantly between BMI groups, then an alternative approach to PA measurement may be required depending on BMI.

Previous studies have shown that PA measurement such as activity classification (83, 110-113) and energy expenditure estimation (71, 104, 114-115) is possible using accelerometry. However, very few of these studies take into account the effects by obesity on approaches to PA measurement. Most studies are based on either BMI groups in the accepted normal range (85, 116-117), or mixed BMI groups that are considered as a whole with relatively small numbers of obese participants (84, 101, 118), and there is little or no comparison made between BMI subgroups.

2.2.1 Obese Movement

Previous research into obese movement is principally concerned with spatiotemporal parameters and biomechanical aspects of gait. Several studies note that the preferred walking speed of obese individuals is slower than non-obese individuals (119-121). Many studies found that obese individuals exhibited shorter stride lengths (119-121), and cadence was also observed to be lower (121-122). Step widths were found to be greater in obese individuals (119) resulting in greater mediolateral sway of the centre of mass (123). The stance phase and double support phase was longer for obese individuals (120-121). Also, a recent study found that obese individuals walked with a straighter leg (124). Many biomechanical differences in gait have been observed between obese and normal groups. Hip abduction/adduction angles differ between BMI groups (119-120). Obese participants have a higher toe-out angle (121, 125) and greater eversion at the ankle (120, 125).

There is little research comparing BMI groups for movements other than gait. One study considered differences between obese and normal BMI groups when performing whole body reaching tasks, and found there was a greater centre of mass displacement in the obese group (126). Studies investigating sit to stand movements have found that there is greater hip flexion in the non-obese group (127-128) and that foot placement differed between groups for the task (128). Differences between BMI groups for simple tasks such as these imply that there may also be differences for other types of movement. A systematic review by Runhaar et al. (129) focussed on differences in lower extremity joint biomechanics between obese and non-obese – within the context of the effects of obesity on osteoarthritis – and showed that obese individuals exhibited altered biomechanics for everyday tasks.

In some cases it is apparent that the different movement styles will have an effect on the accelerometer signal. Increased mediolateral sway of the centre of mass will certainly cause greater accelerations to be recorded by a hip-mounted accelerometer. For other differences the effect on accelerometer output is not as apparent. For example, the several altered gait parameters observed in obese individuals may result in an accelerometer signal which is different to those generated by non-obese individuals, yet these signals may retain key characteristics, common to both groups, that distinguish them as walking.

2.2.2 Excess Adipose Tissue

There is no previous research on how excess adipose tissue affects accelerometer output. Though, it is evident that some common accelerometer sites, such as the waist/hip, may have a thick layer of adipose tissue which, when performing activities, could move independently to the movement of the body segment to which the accelerometer is attached, and which would in turn cause unwanted movement of the accelerometer. The accelerometer would subsequently be subject to two sources of movement – the body segment movement and the movement of the adipose tissue – and the accelerometer signal would reflect this.

2.2.3 Summary

It is clear from previous research that there are distinct differences between obese and non-obese movements. When measuring PA using accelerometers in a mixed BMI group, these differences in movement may result in inconsistent accelerometer signal characteristics between the two BMI groups. The signal may be further influenced by unwanted accelerometer movement due to excess adipose tissue at accelerometer sites. When algorithms designed to estimate measures of PA from accelerometer signals are applied to mixed BMI groups, these differences in signal characteristics have the potential to affect the accuracy of estimated PA.

2.3 Classification of Activities from Accelerometer Output

In order to gain a greater understanding of how physical activities affect individuals, it is necessary to be able to quantify several aspects of those activities. Many studies choose to evaluate physical activity in terms of energy expenditure, as this is a good measure against which outcomes such as weight loss may be assessed. However, energy expenditure measurements alone give a limited insight into the wider aspects of PA. The ability to identify the type of activities a person undertakes would help assess the role that each activity plays in affecting outcome measures in clinical studies. Additionally, the mode of activity is an important element in understanding individuals' activity patterns, and a practical tool able to provide this information under free-living conditions would be of use in epidemiological research.

As discussed previously, exercise interventions have included several varied activities including overground walking (21, 35, 38-40), treadmill walking (20, 25), jogging (30, 41), stationary cycling (20, 24, 39, 42), rowing (43), stair stepping (20, 24), stretching (37), resistance training (23, 26-27, 45-46), step aerobics (26, 47) and calisthenics (48). A prescribed exercise program may contain a single (42, 130-132) or multiple activities (20, 24, 43). These activities may be performed at light (39), moderate (40, 133) and vigorous (40, 134) intensities. An activity classification algorithm for use in exercise interventions would, therefore, need to be able to recognise many different activity modes.

It is envisaged that an activity monitor may be used to generate a periodic profile of the activities performed by obese individuals participating in an intervention such as a weight-loss programme. The activity profile would show how well an individual had adhered to the prescribed programme of activities, and this then may be compared with the amount of weight lost. Each activity could be assessed for its effectiveness in weight loss, and how well it is adopted by participants. In addition to prescribed exercise, an activity profile could also return detailed lifestyle activity which may not otherwise be accounted for. With information such as this, a large scale weight loss intervention could be better assessed on its overall effectiveness. Detailed activity profiles obtained using an activity monitor may also be of use in epidemiological research aiming to identify which behaviours are risk factors associated with obesity.

As will be discussed in section 2.5, EE estimation equations based on accelerometer output are less effective for multiple activity types than for single activities in isolation. Therefore, separate EE estimation algorithms are recommended for each activity type. A classification stage is thus required in order to decide which of the EE estimation equations is appropriate to the accelerometer output.

2.3.1 Activity Classification

Signals generated by body-worn accelerometers vary according to the type of activity being undertaken by the wearer. The movements of a body, and the accelerations it undergoes, are clearly similar for a single activity type, but vary across different activities; for example, jumping always involves a significant vertical acceleration, whereas using a rowing machine entails very little vertical movement. Consequently, body-worn accelerometer signals for the same activity type share signal characteristics that may not be common to other activity types. By exploiting the distinguishing characteristics of accelerometer signals, machine-learning techniques are able to identify which activities have taken place in the accelerometer record, and can assign to each an activity label accordingly. This process is known as *classification*. Figure 2 illustrates how different physical activities generate distinctive accelerometer signals; the cross-trainer can be seen to have significantly higher changes in vertical acceleration than cycling and rowing, and, similarly, rowing shows greater anteroposterior accelerations than both cross-trainer and cycling. However, the differences in the accelerometer signals between different activities are not always so pronounced, and the effectiveness of the machine-learning algorithms to classify activity from accelerometer signals relies on which signal characteristics can best distinguish activities.

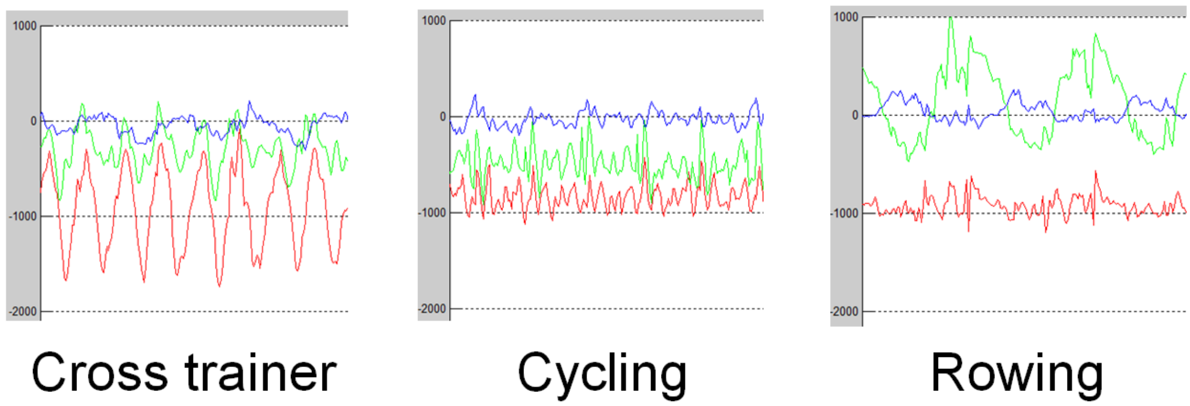

Figure 2: Example accelerometer signals for selected activities. Signals shown in units of g x 1000 against time in 0.02s increments. Accelerometer axes: Red = vertical axis, Green = anteroposterior axis, Blue = mediolateral axis. (This data was collected as part of the study described in section 3.)

2.3.2 Overview of Classification Process

There are a number of stages involved in the process of classifying physical activity from body-worn accelerometer data. The main stages, shown in Figure 3, are as follows: 1) data collected using accelerometers is used as input to the classifier; 2) input data is segmented into manageable and meaningful parts; 3) signal characteristics, known as *features*, are extracted from each data segment; 4) a classification algorithm is applied to the feature data; 5) estimates of activity type are produced for each data segment. Segmentation, feature extraction and classification are described in more detail in sections 2.3.3, 2.3.4 and 2.3.5.

input → segmentation → feature extraction → classification → output

Figure 3: The classification process

A classification scheme may employ either a supervised (135-136) or an unsupervised (137-138) machine-learning approach to identifying activities from an accelerometer signal. Supervised machine-learning algorithms first require training by a dataset of example accelerometer signals for different activities in order to learn how to associate accelerometer signal characteristics with activity types (111, 139-140). Once trained, these algorithms may then produce estimates (also known as *predictions*) for which activities are represented within a new set of activity data. Unsupervised machine-learning attempts to distinguish between different activity types and group together like activity types without the use of training data (141-143). The main disadvantage of supervised machine-learning is that physical activities which are not trained for will necessarily be misclassified as one of the possible activities contained within the training set, whereas unsupervised machine-learning is not limited by a predetermined activity set. However, the majority of previous activity monitoring research has used the supervised-learning approach to test whether a specific set of activities may be identified. Unsupervised learning is beyond the scope of the thesis, and all future references to and descriptions of classification techniques will be in the context of supervised machine-learning.

Classification may be applied on an inter- or intra-subject basis. For intra-subject classification (144), a single subject provides training data for the classifier, which is subsequently applied to test data for the same subject. Inter-subject classification (84) uses a group of several subjects to train the classifier, which is then applied to a subject that does not belong to the training group. Intra-subject classification usually leads to higher prediction accuracy, as there are no variations in accelerometer output that are due to subject differences. However, in a real world application such as a clinical intervention, or a study where many subjects may be taking part, it is not practical to train for every subject individually, and, therefore, inter-subject classification is necessary.

2.3.3 Segmentation

A dataset collected from a body-worn accelerometer over a period of time will typically contain a number of different activities. The aim of an activity classification scheme is to provide estimates for the types of activity that have been performed over this time period. In order to achieve this, the accelerometer data must first be divided into manageable segments known as *windows*. Classification then takes place on a window-by-window basis. Windows generally contain a fixed number of acceleration samples, and thus represent a fixed duration. This duration ranges between 0.2s (145) and 12.4s (85). As far as possible, each window should contain representative accelerometer data for a single activity type, though windows containing transitions between activities are inevitable. If the chosen window length is too long, however, then it may cover two or more activities. This will cause misclassifications in situations where smaller windows may have been sufficient to correctly identify each activity. Conversely, a window that is too short may not have adequate data within it to distinguish the activity.

2.3.4 Feature Extraction

The feature extraction process reduces each data window to a number of characteristics; these are usually statistical in nature and are chosen with the aim to best distinguish differences between activities. Features may be time-domain, such as mean and standard deviation (110-

111, 144), or frequency-domain such as spectral energy (111, 139, 146), and fast Fourier components (FFT) (147). Additionally, a heuristic approach may be employed, where insight into the problem domain informs feature choice – for example, the angle of the sensor with respect to the vertical (in relation to gravity) may be used to distinguish body orientation and thus infer types of posture (148). Also, wavelet analysis (149-151) has been applied to feature extraction (152-154); this method combines both frequency and time information by subdividing the signal using a sequence of high and low pass filters .

It can be seen that for the three activities in Figure 2 the vertical signal variation, shown in red, is much greater for the cross-trainer than the other two activities, and the green anteroposterior signal varies more for rowing than the other two activities. A two-dimensional feature set produced by calculating the variance of the vertical and anteroposterior would sufficiently distinguish the three activities. By applying a set of rules based on thresholds between high and low variances, estimations of activity type could be returned. However, when considering a greater number of activities, the differences between signals are not so apparent, and a more complex feature set would be required to allow for this.

If a feature set is not able to adequately characterise the accelerometer signals for different activities, then the classifier accuracy will be limited regardless of which algorithm is chosen. Too few features may contain insufficient information for the classifier to perform satisfactorily, whereas too many features can confound the classifier and reduce classification accuracy (155-156). Some studies may initially select a large number of features then apply feature reduction techniques such as Principal Component Analysis (PCA) (157) which transforms a number of potentially correlated features into a smaller number of uncorrelated features. Also, a feature selection algorithm (158) may be applied to reduce the number of features by identifying and eliminating redundant features.

2.3.4.1 Feature Space

Feature values extracted from a window of data can be considered as representing a point within a multi-dimensional *feature space*. Taking the mean and standard deviation of a window, for example, would place it within a two dimensional feature space. Well chosen features would allow windows of like activities to form groups within the feature space, and would result in more distinct groupings between different activities. In the example in Figure 4, it can be seen that the three activities have formed clear groups in the feature space, which suggests that this feature set may be effective in distinguishing those activities. A greater number of features can help mitigate overlap between groups, and consequently lead to higher classification accuracy. However, if too many features are selected, then this can lead to a phenomenon known as the *curse of dimensionality*, where “the demand for a large number of samples grows exponentially with the dimensionality of the feature space” (135). This means that the feature space becomes more sparsely populated with training data as dimensionality increases with every feature added, thus necessitating a greater amount of training data. If there is insufficient training data, then this will detrimentally affect the ability of the classifier to produce accurate predictions.

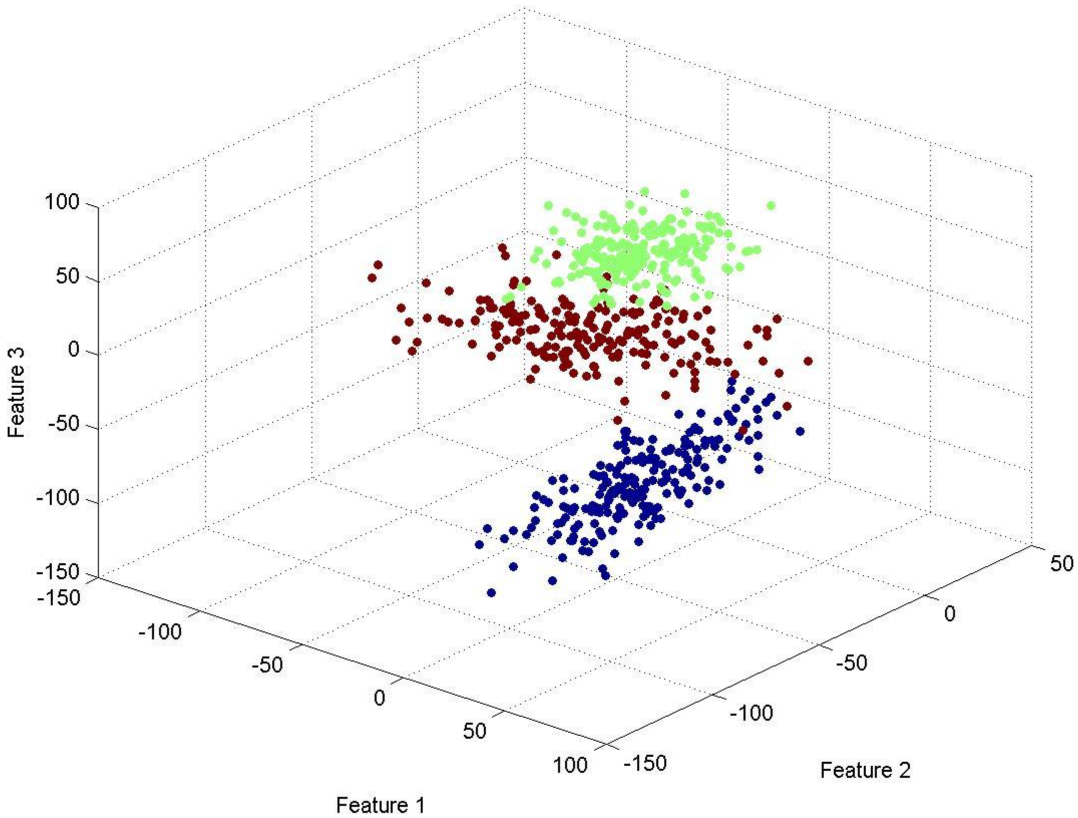

Figure 4: Feature space example. Windows of three different activities (red, green, and blue) projected into a three dimensional feature space. (Mock data to illustrate feature space.)

2.3.5 Classification Algorithms

The classification algorithm is the engine that processes the training and test data, and produces predictions of which activities are present in the test dataset, as shown in Figure 5.

The classifier must first be trained to recognise activities before the test data can be processed. The training data that is supplied to the classification algorithm is a collection of features extracted from each data window, and labelled according to which activity the window represents. The test data contains accelerometer signals which represent an unknown set of activities for which predictions are to be made. The test data is divided into windows, and features are extracted before being passed to the classifier. The classifier generates an activity type prediction for each window in the test dataset based on the training data it has received.

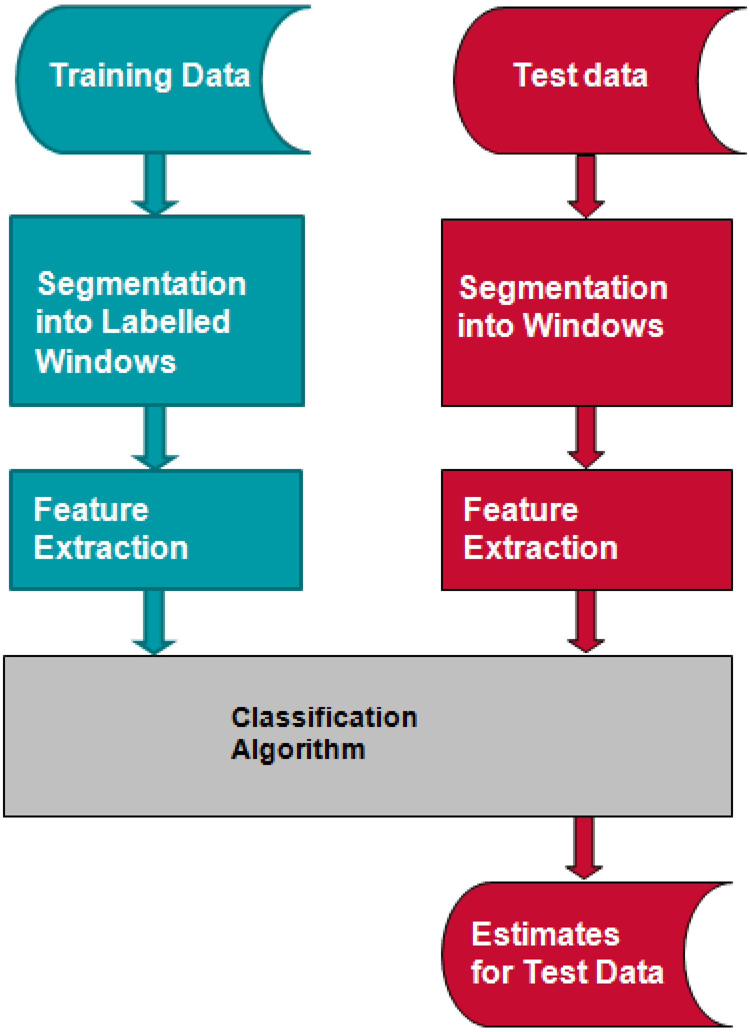

Figure 5: Supervised machine learning

There are several machine-learning algorithms that have been used effectively in the field of activity monitoring. A brief description of some relevant algorithms follows. The reader is directed to Preece et al. (159) for a more comprehensive summary.

Decision trees (84, 111) are structures built to model a particular decision making process. The tree consists of a number of nodes following a parent-child hierarchy. Parent nodes represent decisions, each of which have child nodes which may either be further decisions or terminate at a conclusion to the decision process. The structure is traversed from the top, following the sequence of decisions, until a terminal node is reached. In the context of activity classification each decision is based on features from a window of monitor output and the terminal node is used as the prediction for the class. Figure 6 shows a decision tree based on the example accelerometer output in Figure 2: first the variance of the vertical signal is compared with a threshold which has been established through the analysis of training data, if the threshold is exceeded, then the decision tree returns "cross-trainer" as its prediction, otherwise a comparison is made between the anteroposterior variance and a similarly obtained threshold, and "rowing" or "cycling" is returned accordingly. Decision trees may be defined manually but there are also algorithms, such as the C4.5 algorithm, able to automatically construct optimal decision trees from a set of data

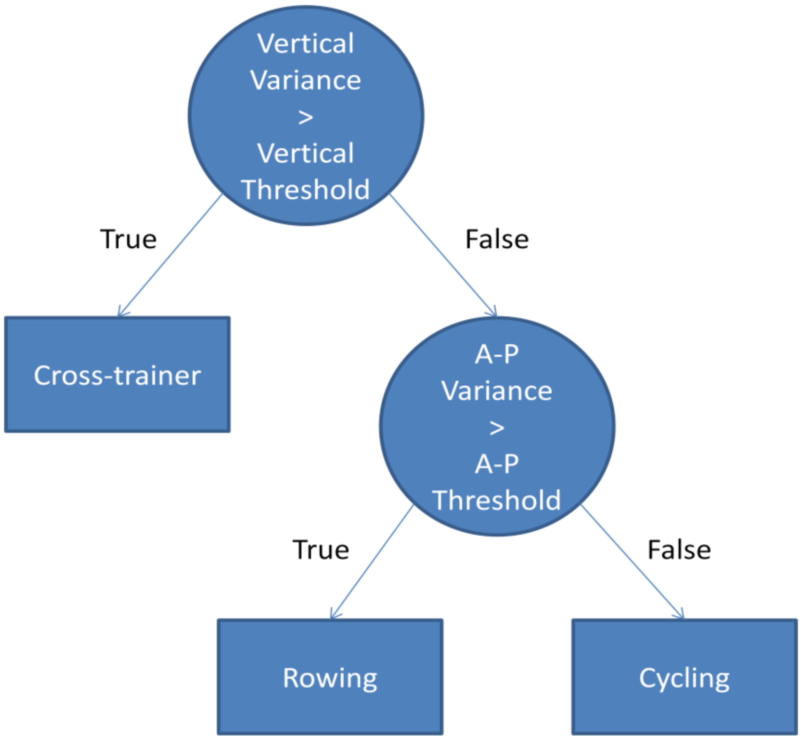

Figure 6: A simple decision tree. The circular nodes represent decisions, and the square terminal nodes represent an activity classification which depends on the truth of the criteria in the decision nodes. In this example, activity type is decided first on whether the variance of the vertical accelerometer signal exceeds a predetermined threshold, and, if not, whether the anteroposterior (A-P) signal variance exceeds the threshold.

Another common machine-learning algorithm is k-Nearest Neighbours (kNN) (135). For this approach, features are first extracted from labelled training data and these are used to populate an n-dimensional feature space, such as the example in Figure 4. A window from an unknown dataset is mapped to the feature space and its distance from each labelled point in the training set is determined. A classification is arrived at by taking the first k nearest points, or neighbours, in the training set and choosing the most frequently occurring class (as labelled). The value of k varies; a small value for k can mean that the classifier is more susceptible to noise in the data, but a larger value will increase computational time. The kNN algorithm was first applied to activity recognition by Foerster et al. (160) who were able to classify nine activities, and a number of subsequent studies have successfully built on this approach (161-162).

Quadratic Discriminant Analysis and Linear Discriminant Analysis (QDA and LDA) have been applied to activity classification (140). Probabilities for each class of activity are defined by multivariate Gaussian distributions (163), and a discriminant function (which is a simplification of the distribution), when applied to a window of unknown activity, provides a likelihood value for each possible activity (164). The activity with the maximum likelihood is chosen as the activity prediction for the window. LDA is a special case of QDA where the covariances of each activity distribution are assumed to be equal, resulting in linear decision boundaries between activity classes (165).

Other common algorithms include the following: threshold-based classification (154, 166), which is a simple classification scheme that compares feature values with predefined thresholds to determine which activity is chosen; Naïve Bayes classification and Gaussian Mixture Model (GMM) (80, 111, 167), which are probabilistic schemes based on Bayes theorem; artificial neural networks (aNN) (84, 117), which return predictions using a mathematical model designed to process information in a similar way to the human brain; Support Vector Machine (SVM) (144, 168), which aims to differentiate between two activities by finding a hyperplane that separates the two with the greatest margin; fuzzy logic (169-170), which is a type of logic which assigns a measure of truth ranging between 0 and 1, allowing input data to have partial membership of fuzzy sets, and returns predictions from rules based on set membership; and Hidden Markov Models (HMM) (140, 146, 171), which return predictions based on the likelihood of transitions between states.

2.3.6 Assessing Classifier Accuracy

Before a classifier is ready to be applied to activity data obtained in the field, as part of a clinical intervention or a study involving measurement of physical activity, its effectiveness at identifying physical activity must first be evaluated. To achieve this, the classifier is trained and then applied to a set of test data where the activities being performed are already known, and by comparing the classification results with the known activity labels, measures of prediction accuracy may be calculated.

The accuracy of a classification algorithm is assessed using a confusion matrix, the format of which is shown in Table 1. This method compares the labelled data with the classifications generated by the algorithm and presents the results in a grid with columns representing predictions and rows representing the actual activities or "Ground Truth". The intersection of row and column contains the total value for the prediction/ground truth pair. In Table 1, P represents a prediction, the first part of the subscript represents the ground truth activity, and the second part of subscript represents the predicted activity – for example, P_{AC} represents an incorrect prediction of Activity C for the actual Activity A. The diagonal ($P_{AA,}$ $P_{BB,}$ P_{CC}) represents totals where the predictions have successfully matched the ground truth, other totals in the grid represent misclassifications.

Ground Truth\Prediction	Activity A	Activity B	Activity C
Activity A	P_{AA}	P_{AB}	P_{AC}
Activity B	P_{BA}	P_{BB}	P_{BC}
Activity C	P_{CA}	P_{CB}	P_{CC}

Table 1: The format of a confusion matrix.

Overall classification accuracy is calculated by dividing the total number of correct classifications by the total number of predictions made. Further to this, measures of classifier accuracy known as precision, specificity, sensitivity (or recall) and F-score may be calculated. These measures are calculated from the total number of true positive (tp), false positive (fp), true negative (tn) and false negative (fn) results. How results are placed within these four categories can be illustrated using the example confusion matrix in Table 1. True positive values, where the prediction matches the ground truth, are P_{AA}, P_{BB}, and P_{CC}. False positives are predictions that have incorrectly been predicted as being a particular activity; for example, P_{BA} and P_{CA} are false positives for Activity A. False negative values are where a specific activity has incorrect predictions; for Activity A these are P_{AB} and P_{AC} where Activity A has been incorrectly predicted as Activity B and Activity C respectively. True negative values are where the predictions do not contain true positives, false positives or false negatives; for activity A these are P_{BB}, P_{BC}, P_{CB}, P_{CC}. Precision, specificity and sensitivity, and F-score are calculated in the following way:

Precision = tp/(tp + fp) *Specificity = tn/(tn + fp)* *Sensitivity = tp/(tp + fn)*

F-score = 2 x (Precision x Sensitivity) / (Precision + Sensitivity)

A classification algorithm is validated for inter-subject capability using a cross-validation technique. One such technique is n-fold cross-validation where data are divided into equal segments, known as folds, and all but one fold is used as training data to test the remaining fold. This is repeated until each fold is tested. Similarly, for leave-one-out cross-validation (Figure 7), training data are obtained from all but one dataset and the algorithm is applied to the remaining dataset. This is repeated systematically so that each dataset is classified in this way. The results may be compiled for all subjects then assessed for accuracy using a confusion matrix and measures, such as sensitivity and precision, as appropriate.

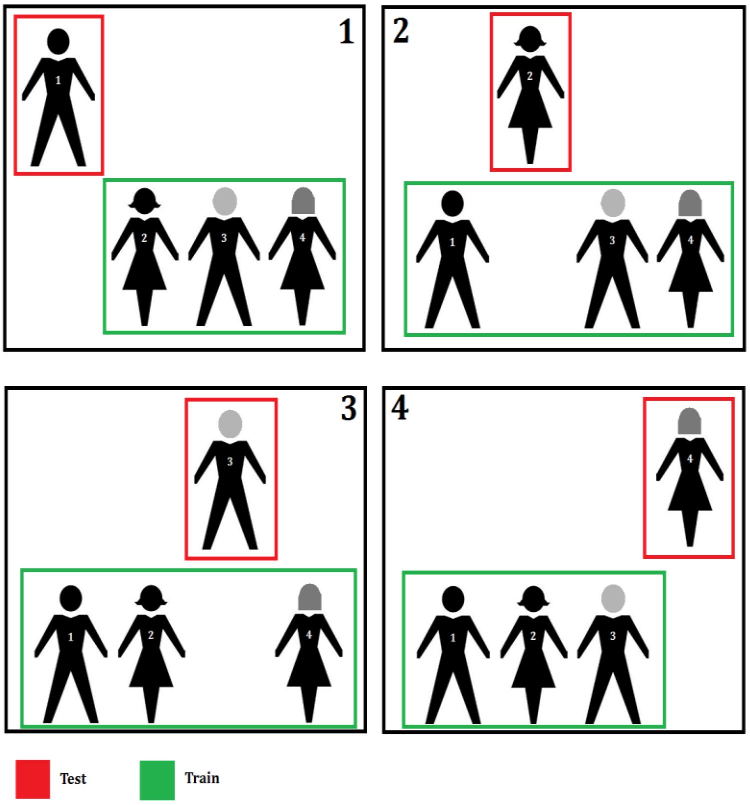

Figure 7: Leave-one-out cross-validation. In the example there are four participants to be tested. For each of the four tests the classifier is trained by data from all but one of the participants and the classifier is applied to the excluded participant's dataset. The results of the four tests are collected to obtain an overall assessment of the classifier's performance.

2.3.7 Related Work

Previous activity classification research has tested different approaches to classification, with the general aim of improving classification accuracy. Studies have often compared the performance of different classification algorithms, and occasionally the effects of different feature sets. However, the findings of this research depend on not only the classification scheme employed, but also on the number and placement of accelerometers, the number of participants, the physical characteristics of the participants(such as BMI), and the types of activity studied.

The sections below present a review of previous activity classification research with respect to the following: the different activity types considered (2.3.7.1), the number and positioning of accelerometers (2.3.7.2), the number of participants in the obese and non-obese BMI ranges (2.3.7.3), and the types of features that have been employed (2.3.7.4). In addition, these areas are considered in relation to their implications for identifying the activities of obese individuals undertaking a weight-loss intervention under free-living conditions.

2.3.7.1 Activity Types

When developing an activity classification scheme there should be sufficient activities to test its capability; the fewer the activities in the chosen activity set, and the more varied the movement characteristics are between activities, the more the classification accuracy will increase, but the classifier will be of limited use – distinguishing between two dissimilar activities such as running and sitting, for example, would certainly give rise to high accuracies without gaining much insight into the general effectiveness of the classification scheme. The activity set is generally aimed at a particular application or domain, such as that of *free-living*. Many studies have incorporated the core free-living activities of standing, sitting, walking (172), and lying (80, 84). Some studies have additionally included stair walking (111, 117, 144, 171) . More vigorous activities such as bicycling (111, 117), jogging/running (84, 111, 117, 144, 171) and Nordic walking (84, 117) are also common. Household chores such as vacuuming (111, 144) and folding laundry (111) have sometimes been considered. Occasionally gymnasium activities such as strength training (111) and rowing (84, 117) are included in the activity set. Certain other activities studied are a specific case of a more general activity, as in the case of *sitting talking* (160) and *watching TV* (111) which may both be considered examples of sitting.

Previous research has generally been arbitrary in its choice of activities, having no aim beyond testing the algorithms with a set of common activities found in everyday life. Specifically, no study has selected an activity set expressly chosen to represent a weight-loss programme suitable for overweight and obese participants. A weight-loss programme may be performed unsupervised at home and in the gym, and accelerometry has the potential to measure adherence to such a program, and to objectively quantify time spent in each activity in order to assess the effectiveness of the weight-loss program. It is, therefore, important to ascertain whether such an activity set can be effectively classified from accelerometer output.

Activity types may be divided into two categories: static activities, such as sitting, standing and lying; and dynamic activities, such as walking, running and cycling. The accelerometer signals produced by dynamic activities can be complex, and a feature set able to capture their distinguishing characteristics is required in order to generate accurate predictions. Conversely, the static activities of sitting, standing and lying produce simple accelerometer signals which are dominated by the constant acceleration due to gravity. These signals show little variance for the same type of static activities, but clear distinctions between different static activity types may be observed; for example, for upper body mounted accelerometers the magnitude of the vertical acceleration signal is significantly greater in sitting than in lying. The goal of classifying static activities such as these is, therefore, relatively easily achieved: static activity is indicated if the signal magnitude area is lower than a predefined threshold (80), and once identified can be further classified using a simple decision tree based on the angle of the accelerometer (80), or through a secondary set of pre-established thresholds between standing, sitting and lying (82).

In order to measure health outcomes based on physical activity patterns, in addition to measuring the level and type of exercise being performed by individuals, it is important to determine the amount of time spent in sedentary activity. A comprehensive activity classification system intended for practical use under free-living conditions would, therefore, need to be able to identify both static and dynamic activities. To achieve this would require the implementation of a two-phase approach to activity classification. The first phase would be to calculate the magnitude of the accelerometer signal. This would be compared to a predetermined threshold representing the boundary between static and dynamic activity. In the second phase, a dynamic classifier algorithm or static classifier algorithm would be employed accordingly (see also Appendix C for further details).

2.3.7.2 Number and Positioning of Accelerometers

The number and positioning of sensors on the body significantly affects how well the activity classification algorithm performs, and also the types of activities that can be identified. Body site sensor placements that have been considered in previous research include the hip (101, 111, 117, 140, 173), waist (80, 110, 142, 146), thigh (111, 160, 174), chest (84, 175), ankle (101, 111), wrist (84, 111, 117, 146, 160, 173), arm (111), head (160), and back (82, 85), and combinations thereof. Additionally, some studies have affixed accelerometers to body-worn items such as a backpack (147), shirt and trouser pockets (167), a bag (167) and a necklace (167, 176). The number of sensors under consideration in previous research ranges between one (80, 140, 144) and thirty (177), though most studies have fewer than six. Having a greater number of sensors placed at different body sites yields a richer dataset, which generally improves classifier accuracy (177-178), but there is a point where improvement is small (111) compared to the monetary cost of each additional sensor and the additional computational resources required. Activities involving distinctive upper body movements are more easily detected with wrist- or arm-mounted sensors, and, similarly, lower body-mounted sensors better identify locomotive activities. For this reason, classifier accuracy differs depending on the combination of accelerometer site and activity set (101). A corollary to this is that there is no single best accelerometer site for optimum classifier accuracy across all possible activities. However, a major goal of activity monitoring is to be able to record activity patterns under free-living conditions in a manner that is as unobtrusive as possible to the person being monitored. For this reason a single sensor placed at an appropriate body site would be desirable.

2.3.7.3 Activity Classification and Obesity

Many activity classification studies have demonstrated high levels of accuracy when testing participant groups with BMIs predominantly in the normal range (18.5-25 kg/m^2). For example, Parkka et al. (84) tested sixteen participants with BMIs of 24.1 ± 3.0 kg/m^2, Ermes et al. (117) tested twelve subjects with BMIs of 23.8 ± 1.9 kg/m^2, and Bonomi et al. (85) tested twenty participants with BMIs of 23.6 ± 3.2 kg/m^2. Each of these studies reported an overall accuracy of between 86% and 93%. However, as these studies have included few obese participants, it is not clear whether the same approaches to activity classification would yield similar results when applied to an obese group.

Another limitation of activity classification research is the small number of participants used to develop and test the algorithms. Studies have been conducted with as few as eight (174), six (80, 140), two (144) and one participant (89). Having very few participants in these studies makes it less likely that these approaches will generalise to the wider population (179), and they either include insufficient numbers to allow a comparison between normal and obese groups, or they have no obese participants. There are, however, some activity classification studies which have taken participant weight and height or BMI into consideration; these are discussed below.

A recent study by Oudre et al. (180) included twenty-four participants, eight of which were overweight, and four were obese. The study reported high accuracies for detection of four activities: walking, biking, running and “other” which included non-periodic and static activities. The overweight group scored the highest accuracies overall, followed by the normal BMI group. The obese group scored highest for running but lowest for biking, making their overall score the lowest. The authors concluded that the variances between BMI groups were reasonable, and that their classification method was still justifiable for mixed BMI groups. However, it is arguable that there was an insufficient number of obese participants to allow any meaningful conclusions to be drawn. Additionally, the study reported that a visual representation of the frequencies of the accelerometer signal (spectrogram) for dynamic activities were visibly different between the obese group and the normal weight group. This may imply that a more pronounced difference between classification results for obese and non-obese groups may become apparent with greater numbers of obese participants, but this was not one of the conclusions of the study.

One study by Zhang et al. (181) included seventy-six participants with BMIs averaging 24.7 ± 4.4kg/m^2 (ranging from 18.4kg/m^2 to 41.0kg/m^2). Sixty-nine of the participants performed five dynamic activities: walking, running, climbing stairs, descending stairs, and jumping. Classification accuracy was compared between the over 25 BMI group and the under 25 BMI group. The authors reported that although there was a significant effect of BMI on classification ($p = 0.045$) it was only correlated with the “running” detection rate ($r = -0.25$, $p = 0.031$) and none of the other activities; as detection rates for running were still high (>99%) they concluded that accuracy was not greatly affected by BMI. However, because the overweight group included the obese participants, no distinction could be made between the separate effects of the overweight and obese groups on classifier accuracy, or how the normal group results differed to the obese group alone.

A study by De Vries et al. (101) considered the effects of accelerometer position on the classification accuracy of an artificial neural network, and also tested the effect of including demographic information such as height and weight. Forty-nine individuals participated with BMI values averaging 23.8 ± 3.4 kg/m^2. Two sets of activities were used: one consisted of five activities comprising general activity types such as walking and using the stairs, the other totalled nine activities which included more specific activity types such as *brisk* or *regular* walking and *ascending* and *descending* stairs. Using three ANN models based on which accelerometers were to be used – 1) hip alone, 2) ankle alone, and 3) hip and ankle combined – they returned accuracies of 80.4%, 77.7% and 83.0% for models 1), 2) and 3) respectively against the five activities, and 60.3%, 64.2%, 69.1% for models 1), 2) and 3) respectively against the nine activity set. For the nine activities the models were unable to satisfactorily differentiate between activities of the same general type but of different speeds, and also standing still was frequently mistaken for sitting, and stair ascent was confused with stair descent. To investigate whether classification accuracy was improved by including demographic variables, the researchers included body height and weight, gender, and age as additional inputs to the ANN. Surprisingly, they found that classifier accuracy actually decreased slightly for all three models. However, that demographic variables do not improve classifier accuracy cannot be regarded as conclusive in light of some fundamental issues. Within the five component activity set, two of the activities were static (standing and sitting), and therefore easier to detect (as discussed in 2.3.7.1), leaving only three dynamic activities (walking, stair walking, and cycling), and for this varied combination a high overall accuracy might be expected. Under these circumstances, the prediction accuracies for the five activities of 77%-83% may possibly be regarded as lower than expected. This result, and the relatively low reported accuracy for the nine component activity set (60%-69%), may be due to the accelerometer data being sampled at only 1Hz; as human activity may involve movements of up to 20Hz. This low sample rate may not be sufficient to capture the subtlety of movement involved within the activities, the extracted features are consequently less characteristic of the activity types in question, and thus the ANN's prediction accuracy is compromised.

There is reason to hypothesise that classification systems developed from principally non-obese groups may not perform as expected when applied to the obese. Obese gait and movement characteristics differ from those of their lower BMI counterparts, as discussed in 2.2.1. Additionally, an excess of adipose tissue at sensor placement sites may allow extraneous movement of the accelerometer, which in turn may add noise to the accelerometer signal that could affect the classification process. Previous research has not sufficiently investigated the effect of obesity on activity classification accuracy, and, therefore, more research is needed to compare obese and normal BMI groups.

2.3.7.4 Features Considered by Previous Research

As was discussed briefly in 2.3.4, activity classification studies employ statistical features to characterise each window of activity data. Also, several studies have sought to exploit the periodic nature of some activities, such as walking and cycling, by transforming the accelerometer data to the frequency domain. A selection of features that have been used in activity monitoring studies is described below.

Two of the most popular choices of feature are the mean of the signal (111, 117, 140, 146, 160), and the standard deviation or variance (84, 89, 140, 142, 146) . Bao and Intille's 2004 study (111) became the most comprehensive in terms of the number of participants and activities involved; the features they employed were mean, energy, frequency-domain entropy, and also a measure of correlation between accelerometer axes. Energy was calculated as the sum of the squared magnitudes of the discrete FFT components of the signal, excluding the DC component (which is effectively the mean of the signal), and was normalised by dividing by the window length. Similarly, frequency-domain entropy was calculated as the normalised entropy of the FFT components of the signal, again excluding the DC component. The features used by Bao and Intille (111) have subsequently been used, either alone or as a subset of a larger feature set, in studies such as Ravi et al. (144), Mannini et al. (88), Bonomi et al. (85), Huynh and Schiele (147) and Lester et al. (146). A number of studies have used skewness and kurtosis (84, 110, 117), and eccentricity (110) of the signal. Percentile values have been used as features (84, 101, 117) – these features are typically comprised of two or more of the following: 5^{th}, 10^{th}, 25^{th}, 75^{th}, 90^{th} and 95^{th} percentiles. The signal magnitude area (SMA), which is the sum of the high-pass filtered rectified signal, and is more commonly used in estimating energy expenditure as described in section 2.5.1, has also been considered as a feature in activity classification (80, 142). The Fast Fourier transform (FFT) has been used to extract features based on Fourier coefficients (85, 87, 146-147) – these may be used as individual features, placed into frequency bands, or used to compute measures such as entropy and energy. Similarly, discrete cosine transformation (DCT) (112) and Cepstral analysis (146) have also been used to generate frequency domain features for application to activity recognition – a cepstrum of a signal is defined as the inverse Fourier transform of the log of the Fourier transform of the original signal.

Although there are several studies that compare the capabilities of different classification algorithms, there are few which compare different feature sets. Maurer et al. (167) compared eight feature sets for six activities and six sensor sites for six participants. The results showed a range of accuracies for each permutation of feature set and sensor site, but the researchers had deliberately optimised the feature sets for use with a watch-based sensor worn at the wrist. In addition, there were only four dynamic activities (running, walking, ascending and descending stairs) all of which were locomotive in nature, and the classifier returned poor accuracies for ascending and descending stairs for all feature sets. There was, therefore, no conclusion to be drawn as to which feature set performed best. Huynh and Schiele (147) compared thirty-two features and concluded that no single feature performed best for the six activities under consideration; although FFT coefficients consistently ranked highly, the particular coefficients responsible for the highest precision differed across activities. Preece et al. (182) also found that FFT components yielded the highest prediction accuracies for eight different dynamic activities. These results were obtained using ten feature sets, including that proposed by Bao and Intille (111), and combinations of three sensor placement sites (ankle, waist and thigh) with the highest accuracy resulting from the ankle sensor data.

In section 2.2 it was hypothesised that the movement characteristics of obese individuals, and excess adipose tissue at sensor placement sites, may give rise to accelerometer signals that differ from those observed in normal weight subjects. As a consequence, this may result in a disparity in feature values between the different BMI groups for like activities, which would lead to potential misclassifications. As discussed above, there have been few studies that have considered the effect of BMI on classification, and there are none which have compared the performance of feature sets between obese and non-obese groups.

2.3.8 Research Questions

Previous research has not adequately addressed the effect of obesity on activity classification, nor has it considered an activity set developed specifically for weight-loss purposes. Furthermore, there are no studies which have compared classifier accuracy between obese and normal BMI participants, or compared the performance of different feature sets for these groups. A single, unobtrusive accelerometer site is desirable for free-living applications such as monitoring an obese person's adherence to a prescribed set of activities designed for weight loss purposes. However, it is not known how a classification scheme will perform with a single accelerometer against a weight-loss oriented activity set. Given the limitations of previous studies, the following research questions are proposed:

1) Can a set of aerobic exercises and free-living activities be identified from data collected by a single accelerometer mounted at the hip or at the ankle?
2) Does activity classification accuracy differ between obese and normal BMI groups?
3) Do the same accelerometer features apply to obese and normal BMI groups, or do they require different accelerometer features to characterise their physical activities?

2.4 Walking Speed Estimation from Accelerometer Output

Walking is a key activity in maintaining health (183) and can help reduce excess weight (184-185). NICE guidelines recommend at least five days of moderate intensity exercise, such as brisk walking, for at least thirty minutes per day (18), and walking is a preferred activity among sedentary individuals wishing to increase their activity levels (186). Walking is an easily performed aerobic exercise which can be incorporated into daily life. It does not require special equipment or professional oversight. It can be performed at low speeds to suit the ability of the individual, and can be increased in intensity and duration over time as fitness improves. Walking has been described as a the "nearest activity to perfect exercise", having multiple benefits and minimal adverse affects (187), and is recommended by public health bodies for maintaining and losing weight (183). Walking is, therefore, an area of special interest in activity monitoring research, particularly in the context of obesity management.

Walking speed is a measure of intensity which may be considered as an alternative or complement to measures of energy expenditure. Ainsworth et al. (188) assert that speed is a key factor in categorising an activity such as walking into light, moderate or vigorous intensity. A clinician implementing an intervention which recommends brisk walking would benefit from the ability to identify when bouts of walking are occurring, their frequency and duration, and what speeds are being achieved. From this information it would be possible to measure adherence and evaluate the effectiveness of the intervention. Also, these measures may be useful to a participant in such in an intervention, as it would provide motivational feedback and a cumulative record of how well they were meeting exercise targets.

Various techniques exist for objectively estimating walking speed. Global positioning systems (GPS), for example, have demonstrated this ability. Schutz and Chambaz found a high correlation between GPS-measured and actual walking speed (189), though they concluded that the level of error they experienced made it unsuitable for intervention studies.

La Faucher et al. (190) applied processing techniques to GPS signals and achieved a very high correlation between estimated and actual speed (R^2=0.987). However, although GPS may be accurate – and also a common technology available through smart phones – it is limited to outdoor areas within the line of sight of four satellites, which is problematic in heavily developed areas such as cities. It also is not possible to assess walking speeds of individuals exercising on a treadmill by means of GPS.

Gyroscopes and inertial measurement units (IMUs), which combine gyroscopes and accelerometers, have previously been applied to estimating walking speed (191-194). Gyroscopes are first discussed, and accelerometry is considered separately below. The gyroscope approach to walking speed estimation first obtains angular rotation of the leg about the hip from leg-mounted gyroscopes. Using this value, in conjunction with the length of the leg, the stride length is calculated through trigonometry. This can then be divided by the duration of the gait cycle to obtain walking speed. Studies differ in their approaches to gyroscope placement, how the angle of rotation is obtained, the model used to estimate stride length, and how the length of the gait cycle is calculated.

Miyazaki (195) took the maximum and minimum angles recorded by a thigh-mounted gyroscope to obtain the angle through which the leg has moved. Using these angles and the length of the subject's leg, the stride length was approximated through trigonometry. The gait cycle duration was the time elapsed between the points at which the leg was oriented at these maximum and minimum angles. From the stride length and duration, an estimate for walking speed was obtained. Aminian et al. (196) used three gyroscopes – one attached to each shank and one on the right thigh – to measure the angular rate of rotation parallel to the mediolateral axis. Heel strike and toe off were obtained through identifying negative peaks in the angular velocity signal, and the duration of the gait cycle was calculated as the time between consecutive toe-off events for the right leg. The angular velocity of shank and thigh was integrated to obtain estimates of the angular rotation of shank and thigh, and a double pendulum model was applied to these angles, along with thigh length and shank length, to obtain stride length. Velocity was then calculated as the derived stride length divided by the estimated gait cycle duration. The RMSE for speed estimation returned by the study was 0.06m/s, which represents good accuracy.

Although gyroscopes have been shown to accurately measure walking speed in the laboratory, they use a relatively high level of power for their operation, which limits the amount of time they can be used before their power source requires recharging. This may be a problem for studies in the field. Gyroscopes can also be subject to drift over time, which can lead to inaccuracies when computing angular velocities (197) and, consequently, affect speed estimates. For these reasons, gyroscopes may not be ideal tools for measuring walking speed under free-living conditions and over prolonged periods.

Body-worn accelerometers have been shown to be capable of estimating walking speeds in various studies (91, 198-201). Accelerometers consume less power than gyroscopes, and therefore have longer operational times in the field, and their use is not limited by geographical location as in the case of GPS. An additional advantage of using accelerometers is that, when analysing data collected under free-living, an activity classification phase (as described in section 2.3) may precede walking speed estimation to help ensure that the data has indeed been generated by walking. This phase would not require additional data or sensors, and would mean that periods of activity other than walking were not estimated for speed. Previous research in estimating walking speed using accelerometry is discussed below.

2.4.1 Related Work

Several studies have investigated the potential for accelerometry to be used to estimate walking speed. These studies fall into two main categories: those based on a biomechanical model, and those modelling the relationship between accelerometer signals and walking speed by using abstracted models such as machine-learning algorithms. The biomechanical approach first attempts to estimate one or more gait parameters from the accelerometer output, such as vertical displacement, then utilises an established biomechanical walking model to calculate speed from these parameters (90). Abstracted approaches attempt to model a relationship between the accelerometer signals and walking speed. This approach usually involves an initial phase where example accelerometer data is collected at different walking speeds. From this, a multi-linear speed prediction model may be derived though linear regression (200). Alternatively, a machine learning algorithm, such as an artificial neural network, is trained to associate characteristics of the accelerometer signal with the measured speeds (198). A selection of walking speed estimation studies are discussed in more detail below.

Human walking can be considered as adhering to an inverted pendulum model (202-203). For each leg in stance phase the centre of mass moves through an arc in relation to the foot (pivot) as shown in Figure 8.If suitable parameters can be obtained from an accelerometer signal, then spatiotemporal gait parameters may be calculated using the model. Walking speed estimation methods employing a biomechanical model are generally split into two phases. First, the gait cycle duration is calculated from the accelerometer signal by identifying the times of gait events through techniques such as heel strike detection (90, 204). Second, stride length is obtained from the model by applying integration and geometry, as will be discussed below. From these values walking speed may be subsequently obtained. In a study by Zijlstra and Hof (90), an accelerometer was affixed to the lower back, and the vertical accelerometer signal was double integrated to obtain vertical displacement of the centre of mass (see Figure 8). From the vertical displacement and the subject's leg length the step length was calculated using the inverted pendulum model according to the formula: $step\ length = 2\sqrt{2lh - h^2}$ (where h is the vertical displacement and l is the leg length). This result was divided by the gait cycle duration to obtain walking speed.

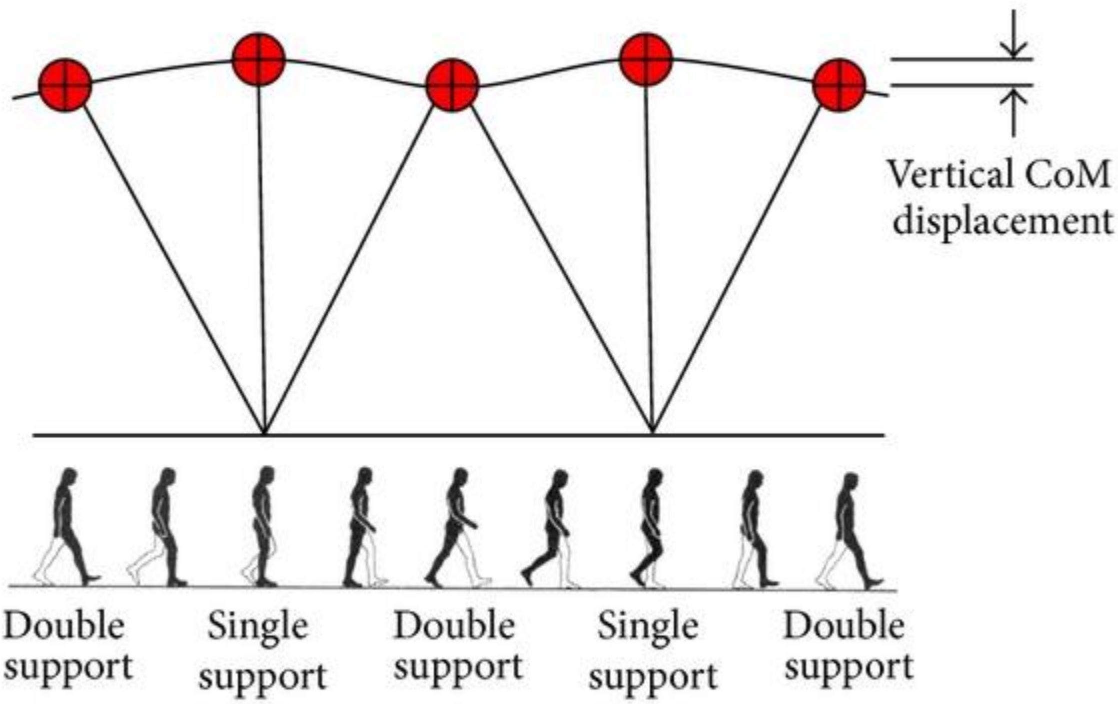

Figure 8: Inverted pendulum model of gait, showing the trajectory of the centre of mass (CoM) through stages of the gait cycle (205)

A recent study by Bishop and Li (206) also exploited the pendulum-like movement which characterises walking . Subjects wore two biaxial accelerometers laterally mounted on their right shank. Throughout each stride cycle the angle of the vertical accelerometer axis changes in relation to the global vertical, relative to the direction of gravity. Bishop and Li (206) used this angle to transform the accelerations experienced by the accelerometers into the global coordinate system. Double integrating these accelerations gave global vertical and horizontal displacements. From these, stride length was calculated. Stride cycles were identified as occurring between consecutive time points where the shank was parallel with the direction of gravity, and this allowed the gait cycle duration to be determined.

There are a number of studies which have applied correlation and line fitting techniques to accelerometer signals to obtain walking speed prediction models. These studies have generally first extracted one or more features (as described in 2.3.4 and 2.3.7.4) from the accelerometer signals, and the relationship between the set of features and walking speed has then been analysed using methods such as regression (85, 200). Schutz et al. (200) were the first to use features of an accelerometer signal to estimate speed through regression, in this case a single feature (root mean square) was used. Bonomi et al. (85) extracted the average, standard deviation, peak-to-peak distance, and cross-correlations between axes as features of the signal and employed multiple linear regression to estimate walking speed. Panagiota et al. (199) developed a linear model based on ten features. In this study, features obtained from the accelerometer signal included step cycle duration and the mean magnitude of the signal, and additional features were based on subject attributes such as height, weight and BMI. In contrast, the model used by Barnett and Cerin (207) was obtained by simply correlating accelerometer counts with walking speed[2]. An alternative approach by Yeoh et al. (208) took the sum of the net acceleration values of two thigh-mounted accelerometers for a number of walking speeds and applied the method of least squares to fit a third order polynomial model for speed estimation.

An artificial neural network (aNN), which is a form of non-linear regression, was first applied to walking speed prediction from accelerometer output by Aminian et al. (198). This approach uses accelerometer signal features as input parameters to the aNN. Parameters are usually obtained after segmenting the accelerometer signal, commonly into gait cycles (198), and are then calculated segment by segment. Parameters taken from the signal are usually statistical, such as the root mean square (209) or median (198). Subject demographics such as height (204, 209) and weight (204) may also be included within the parameters. The parameters and their corresponding speed values are used to train the neural network. Predictions for speed may then be generated by the aNN from a given set of input parameters.

[2] Accelerometer counts are the sum of the filtered and rectified accelerometer signal over a fixed period, and are frequently used in energy expenditure estimation equations. Counts are described in more detail in 2.5.1.

Previous research has considered treadmill walking and overground walking when developing and testing speed prediction algorithms. Many studies have considered only overground walking (85, 94, 199, 207, 209), while several studies have used treadmill data alone to both train and test an algorithm (204, 206, 210). Various other studies have trained algorithms with treadmill data, and applied this to overground test data (118, 198). The advantage of using a treadmill is that a constant speed may be imposed and maintained, and therefore reliable walking data is easy collected for each speed dictated by the study protocol. Overground walking is more naturalistic and allows the subject to set the walking pace.

2.4.2 Limitations of current approaches

In 2.2.1 it was discussed that obese gait differs from that of lower BMI individuals. However, most previous research into estimating walking speed from accelerometry has not considered the effect of obesity on the outcomes. It is possible, for example, that the vertical displacement of the centre of mass in obese individuals is lessened due to wider strides and greater mediolateral sway when walking (119-121). It is not clear, therefore, whether a speed estimation algorithm which uses this displacement, such as that employed by Zijlstra and Hof (90), would generalise across BMI groups. Again, the study by Bishop and Li (206) focussed on developing a novel method to predict walking speed from accelerometry based on a biomechanical model, but did not consider the effects of altered gait patterns observed in obese individuals. It is, therefore, unclear whether speed estimates would be equally accurate for obese individuals as for non-obese individuals when employing these techniques. Approaches using artificial neural networks may be able to compensate for differences in gait, as the accelerometer signals contain tacit information pertaining to gait characteristics, which may be recognised by an aNN. The aNN would, however, need sufficient training with data from a heterogeneous group of walkers, and this has not previously been investigated.

Studies or clinical interventions which require measurements of walking speed may involve many participants. For this reason, the measurement procedure needs to be practical and cost effective. A limitation of certain previous walking speed studies involving accelerometry is that prediction algorithms have been developed on an intra-subject basis (198, 200). This approach requires an initial calibration phase where the participant provides sample accelerometer data by walking at a number of speeds in the laboratory. The speed estimation algorithm can subsequently make predictions for that particular individual based on this calibration data. This approach is useful in tailoring the prediction algorithm to the subject, resulting in higher accuracy rates, and as a result it may neutralise the problem posed by obese gait. However, performing this initial phase adds cost in terms of calibration time and data processing time, and also requires the participant to attend a laboratory appointment. It is therefore not practical in larger studies. Conversely, inter-subject speed prediction models aim to be applied to the wider population without an individual calibration phase. In order to achieve this, the algorithms need to be adequately pre-trained with a suitable heterogeneous subject group, and the prediction models must account for inter-subject differences in gait.

One study was identified which involved many obese participants. Schutz et al. (200) conducted a number of experiments in order to develop and test their speed prediction model, which correlated the RMS of a belt-worn uniaxial accelerometer signal with speed. In one of these experiments, a significant number of obese participants (n=50 females, BMI: 31.4 ± 5.1 kg/m^2) took part. However, the aim of this experiment was to demonstrate that speed prediction models require individual calibration due to inter-subject differences. This was successfully demonstrated for the model in question as, although correlations between RMS and speed were high for each individual in isolation, there was a large amount of variability observed across the subject group. Consequently, they chose to employ individual calibrated algorithms in their model. A subsequent experiment validated their speed prediction model against six non-obese male subjects (BMI: 23.6 ± 2.5 kg/m^2). The final experiment applied the model in a free living environment to a group containing a number of obese members (n=28 females, BMI: 30.0 ± 3.8 kg/m^2). This experiment aimed to demonstrate how the model could be applied to both normal and obese women. However, there was nothing in place to measure actual walking speed in the period of testing, which means that it is uncertain whether the results returned were accurate. Although this study involved several participants with high BMIs, it did not validate the speed prediction model for obese individuals. Also, individual calibration was required by the model, which, as discussed, is impractical in large scale studies or interventions.

Many approaches to walking speed estimation require accelerometers to be placed at specific body sites, and these may not be practical in a free-living environment. The lower back close to the centre of mass is a common placement (85, 91, 94, 198). However, this may prove uncomfortable for individuals when sitting. The method employed by Bishop and Li (206) produced accurate speed predictions, but the two accelerometers need to be carefully placed so that both align with gravity when the shank is vertical, and the algorithm accuracy was also affected by the distance between the two devices. In a study or intervention employing this approach the perceived burden to a participant of having to carefully affix two accelerometers to the shank may result in reduced compliance. Another study investigated the use of a chest-mounted accelerometer (204), but again this is not an optimum site for long term studies outside the laboratory. The hip is an ideal accelerometer placement site for walking speed estimation in the field due to its proximity to the centre of mass and relative unobtrusiveness, but there are few studies which have investigated whether accelerometer data collected at this site can be used to accurately predict walking speed..

A study by Panagiota et al. (199) used hip-mounted accelerometers and applied a multi-linear model to predict walking speed. The study also acknowledged that height and weight can affect gait. For this reason the prediction model incorporated height, weight and BMI in addition to accelerometer features. However, there was not sufficient diversity in the subject group to test for the effects of BMI on the model, and a comparison of results between BMI groups was not made. A previous study by Vathsangam et al. (118) also used a hip-mounted accelerometer and tested three linear regression approaches to walking speed estimation. Subjects were selected for this study with varying BMIs; values ranged from 22kg/m^2 to 34.5kg/m^2 with a mean of 26.4 ± 5.3 kg/m^2. However, there were only eight participants in total, which means that there were insufficient obese participants to test the effects of BMI on the algorithms. Additionally, the analysis was performed on an intra-subject basis, and, therefore, may not generalise for applications where this training phase is not practical.

2.4.3 Research question

There are several approaches that may be applied to estimating walking speed in individuals. GPS can be accurate but is limited to outdoor areas within the line of sight of four satellites. Gyroscopes consume more power than accelerometers, and therefore have lower operational times in the field. Accelerometry presents a low cost, practical solution to these problems. However, there are two main limitations to current approaches of estimating walking speed using accelerometry that are relevant to this thesis. First, no study has considered how obese gait characteristics may affect the accuracy of speed estimation algorithms. Second, accelerometer placement in previous research may be impractical or potentially intrusive in free-living conditions.

The research question is asked: can a hip-mounted accelerometer be used to accurately estimate walking speed for an obese group?

2.5 Energy Expenditure Estimation (from Accelerometer)

Energy expenditure (EE) is a quantifiable aspect of PA against which health outcomes may be measured. EE may characterise PA more suitably than other measures – such as step counts or walking speeds – for a purpose such as a weight loss intervention. Also, for interventions combining diet and exercise, a comparison between calorific intake and energy expenditure would provide useful information on the energy deficit or surplus associated with changes in weight. It is important, therefore, to have tools able to provide objective and accurate measurements of EE. This is particularly the case for overweight and obese groups who are most likely to take part in weight-loss interventions. Accelerometry may be a tool suitable for this purpose.

Accelerometers are used to measure accelerations of body segments. As the intensity of an activity increases, augmented body movements impose greater accelerations upon the accelerometer, resulting in an increase in amplitudes and a change in frequencies within the accelerometer signal. The accelerometer signal characteristics, therefore, relate to activity intensity and, consequently, energy expenditure (115), which is a quantification of intensity. Studies which have used accelerometers to measure energy expenditure in a free-living environment have often employed only simple algorithms to place activities into categories of intensity (211) such as light, moderate, and vigorous. However, it is possible to obtain more precise estimates of EE in terms of units of EE such as calories, and these may be more appropriate for applications.

2.5.1 Current Models for Predicting Energy Expenditure from Accelerometer Output

For the purposes of predicting energy expenditure, accelerometer output is typically quantified in units known as "counts", which are calculated over an arbitrarily chosen period known as an "epoch". Various ways of computing a count have been applied, such as the zero-crossing method where counts represent the number of times the accelerometer signal passes through zero within the epoch (212), or the vector magnitude method (213) where counts are calculated as the square root of the sum of the squared accelerations for each accelerometer axis. However, the most common method is to find the area under the rectified curve across the epoch, and then to convert this value to counts per minute by dividing by the sample rate multiplied by sixty. To calculate the area under the rectified curve the signal is first high-pass filtered to remove the offset from zero caused by gravity. The filtered signal is then rectified – that is, all negative values are converted to positives – and finally the sum of the rectified filtered signal is taken over the epoch. An example of this process is illustrated in Figure 9. Epoch length can vary, but is typically around one minute (212).

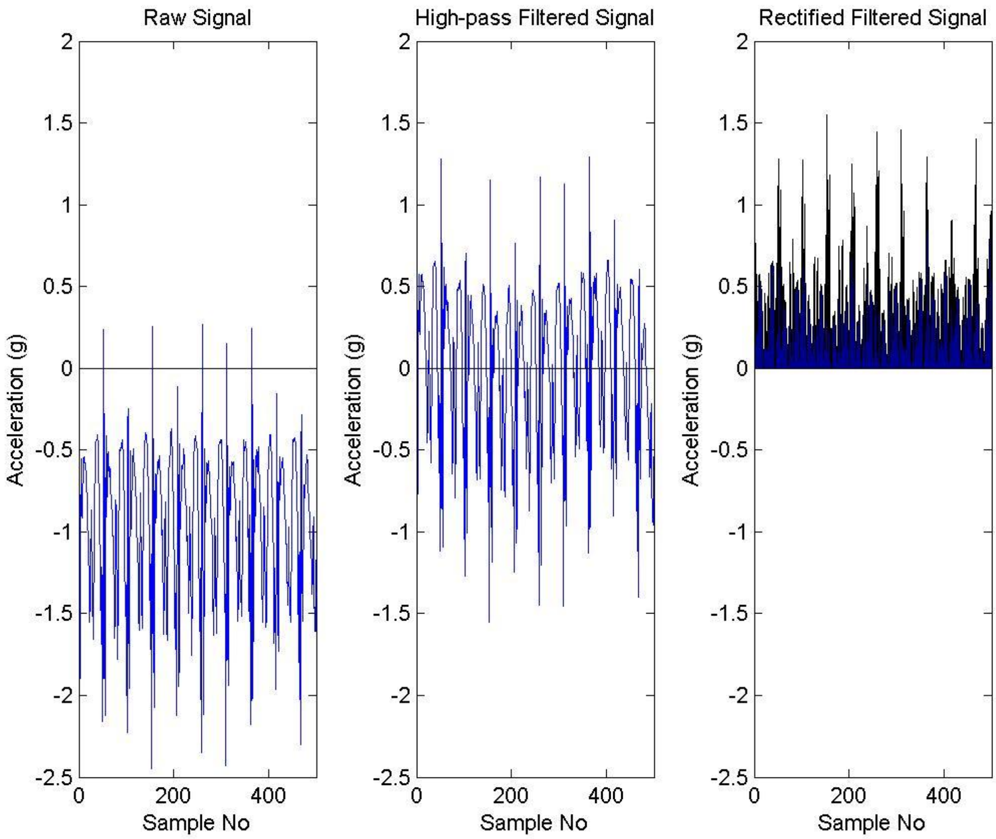

Figure 9: Obtaining accelerometer counts from the raw signal. First, the raw signal is subject to a high pass filter to remove the offset from zero caused by gravity. The filtered signal is rectified to allow the summation of values across the epoch to obtain the area under the curve. In this example a 10s epoch is shown (500 samples at 50Hz). (The figure shows data collected as part of the study described in section 5.)

The potential to estimate energy expenditure from accelerometer output was first investigated by Montoye et al. (115) using a simple linear prediction model. Many subsequent studies have examined the relationship between accelerometer counts and a criterion measure of EE (71, 76, 116, 214-218). During these experiments participants are asked to perform activities at various intensities while both energy expenditure data and accelerometer data are collected. Previous studies have used a range of different placement sites for the accelerometer, such as the waist/hip (71), ankle (219), wrist (73), or lower back (69). The most common activities used to derive EE prediction equations are walking (76, 216), or both walking and running (71, 214-215), though lifestyle activities have also been considered (73, 217).

Indirect calorimetry is the most commonly used method for quantifying energy expenditure (212). With this technique, oxygen consumption and CO_2 production are obtained either by measuring breath-by-breath gas exchange (214), or by room calorimetry (114). Due to the specialised equipment involved in room calorimetry, or the need to wear a face mask or mouthpiece throughout testing during breath-by-breath analysis, these techniques are limited to laboratory conditions. Longer term studies in the field may employ the doubly labelled water (DLW) technique (described in 2.1.2) (220), which is considered the gold standard for energy expenditure measurement under free-living conditions (221). The DLW technique provides the cumulative energy expenditure over the testing period. It is, however, unable to return fine-grained information regarding the energy cost of specific activities.

Oxygen consumption (VO_2) is a proxy measure of energy expenditure, and can be used to examine the correlation between EE and accelerometer counts (71). Most studies, however, elect to convert VO_2 to kilocalories (using $kcal = 5 \cdot VO_2\,(\mathrm{L/min})$ (222) or Weir's equation(223) $kg.cal = 3.9 \cdot VO_2\,(\mathrm{L/min}) + 1.1 \cdot CO_2\,(\mathrm{L/min})$) or METs ($METs = \frac{VO_2\,(mL/min)}{3.5 \cdot body\ weight\ (kg)}$) (222). The MET (metabolic equivalent) value returned by the calculation is representative of the amount of energy used relative to the resting metabolic rate (RMR). The RMR is assumed to be $3.5\mathrm{mLkg^{-1}min^{-1}}$ (224), though an actual measure of an individual's RMR may be substituted for this value. METs are used to quantify EE levels in terms of sedentary (1.0-1.5 METs), light (1.6-2.9 METs), moderate (3.0-5.9 METs) and vigorous levels (>=6 METs) (58). NICE guidelines suggest forty-five to sixty minutes per day of moderate-intensity activity for at least five days per week to prevent obesity (18). With accurate predictions of MET values from accelerometer output under free living, individual activity levels may be compared with guidelines such as these. This may be particularly of use in epidemiological studies which aim to assess the effects of PA on health outcomes (225). Alternatively, kilocalories represent absolute EE. Estimates of kilocalories may be useful as an alternative to METs in cases where the specific energy cost of activities is of interest to the researcher; for example, when a comparison between calorie intake and energy expenditure is made (226-227).

With both accelerometer data and a criterion measure of EE for physical activities performed by subjects at a number of intensities, it is possible to model the relationship between accelerometer counts and EE. Several studies have explored this relationship and have reported encouraging correlation coefficients in excess of r=0.85 (71, 116, 217-218).

Although more recent research has investigated non-linear approaches to predict EE from accelerometer output, these high linear correlations have led to linear prediction equations being the most widely developed. In order to formulate these equations, the measured energy expenditure and accelerometer data for all subjects are collated and linear regression is applied. These basic linear models can be developed to predict METs or kilocalories (with the addition of subject weight as an independent variable).

The linear model for predicting METs from accelerometer counts can be expressed mathematically as follows:

$$METs = a + bK \qquad [1]$$

Where a and b are constants, and K represents accelerometer counts per minute. The constants a and b are obtained through a linear regression between accelerometer counts and a measure of EE expressed in METs.

Similarly, the linear model for prediction kilocalories from accelerometer counts is as follows:

$$kcal/min = a + bK + cW \qquad [2]$$

Where a,b and c are constants, K represents accelerometer counts per minute and W represents body weight. The constants a,b and c are obtained through multiple linear regression. Equations of this form are widely used by commercial systems to provide estimates for energy expenditure under free-living conditions. For example, at the time of writing, the proprietary Actilife software for use with the Actigraph accelerometer has the option to use twelve MET and five kilocalorie prediction equations, all of which are based on linear regressions from previously published research.

2.5.2 Factors which may Explain More of the Variance in the Linear EE Prediction Model

The basic linear prediction models described by equations [1] and [2] above express simple relationships between accelerometer counts, body weight and energy expenditure. Inherent in these models is the assumption that the variance in energy expenditure between individuals is described by accelerometer counts and weight alone. This assumption would appear at first to be supported by the high correlation coefficients discussed above. However, these correlations were obtained from small homogeneous subject groups with BMI levels predominantly in the normal range (71, 116). Physiological differences between obese individuals and those of normal BMI may lead to differences in EE which cannot be explained by differences in body weight and accelerometer counts alone. These differences and the impact they may have on the prediction models is discussed in the following paragraphs.

There are three main components of energy expenditure in individuals which combine to give total energy expenditure, these are as follows (228): resting energy expenditure (REE), or basal metabolic rate (BMR); the thermic effect of food (TEF), which is the energy cost of digesting and processing food; and the energy expenditure due to physical activity. BMR is the minimum amount of energy the body requires to sustain vital functions when at rest under comfortable and warm conditions and more than twelve hours after eating (229). In practice accurate values for BMR may not be easily obtainable, therefore resting metabolic rate (RMR) or sleeping metabolic rate (SMR) may be measured for use as an approximation for BMR. It has been shown by previous research that the magnitudes of body-worn accelerometer signals increase as exercise becomes more vigorous. However, when considering total energy expenditure, accelerometry alone does not bear any relationship with BMR or TEF, as these remain constant for an individual regardless of how the accelerometer signal changes with physical activity; accelerometer output alone is only able to estimate energy expenditure due to physical activity. A first step in improving EE prediction from accelerometry would be to account for BMR in the model.

BMR can account for up to 70% of total energy expenditure (230), and because BMR varies between individuals the total energy expenditure measured for a specific activity will also differ. This will, therefore, affect the relationship between total energy expenditure and accelerometer counts. Because accelerometer counts alone are not able to predict BMR, this supports the need for additional physiological measurements to be incorporated into the EE prediction model to account for individual BMR levels. For example, fat-free mass is the primary determinant of BMR followed by fat mass (FM) (231). The MET and kilocalorie prediction equations [1] and [2], as used in previous prediction models, both presume that BMR is a constant between individuals. For METs this constant is incorporated into the unit when converting from VO2. The kilocalorie model does not distinguish how much of the total EE being estimated is due to BMR and how much is due to physical activity. Some previous research studies (232-233) have chosen to remove subjects' BMR values (or the similar but more easily measured RMR values) from the energy expenditure measurements in order to obtain the actual energy cost of the activity before formulating prediction equations.

This approach is acceptable for research studies, but it may not be practical to obtain individual RMR measurements in large scale implementations of EE estimation algorithms.

Other physiological factors which have the potential to influence energy expenditure during physical activity include age (234-237), gender (235, 237), and ethnicity (235, 237). Children have a higher metabolism than adults due to high cellular activity and growth (229). Fat free mass is known to reduce with age (234), and older adults have been found to have lower BMR levels (236). Men oxidise food more quickly than women which contributes to a higher BMR (229). A recent study comparing accelerometer-based energy estimation equations between Black and Caucasian concluded that ethnic-specific formulae are required due to lower accuracies being returned from the Black group (238).

Rising et al. (239) aimed to investigate the relationship between obesity and total energy expenditure in a small population of Pima Indian men. The study found the major determinant of total daily energy expenditure is fat free mass. Subsequently, Weyer et al. (237) considered a large group of Caucasians (n=416) and Pima Indians (n=500). One of the aims of the study was to provide equations for predicting 24 hour total EE, and for this purpose several demographic, anthropometric and physiological measurements were considered. The study found that age, gender, ethnicity, fat free mass, fat mass, and the ratio between waist and thigh circumferences were significant determinants of total energy expenditure.

For obese individuals the increased energy cost of physical activity resulting from carrying extra body weight is accounted for in the basic linear model. However, there are aspects of obesity which may otherwise increase energy expenditure. Obese gait characteristics, such as shorter strides and wider steps (see 2.2.1), may negatively affect walking economy. The distribution of fat over the body may contribute to greater energy being spent by obese individuals over those with lower BMIs. For example, there is research to suggest that the energy cost of walking in the obese may be increased due to greater weight of the legs (240).

A number of demographic, anthropometric and physiological factors were highlighted above which have the potential to influence EE for physical activities and could therefore give rise to between-subject differences in EE. Such differences are likely to be more pronounced between groups of obese individuals and those with normal BMI. If such differences do, in fact, impact significantly on EE in obese groups, then lower correlations between EE and accelerometer counts might be expected for cohorts with a large range of BMI. A preliminary study by Preece et al. (241) appears to support this argument. The study compared the correlations between accelerometer counts and oxygen consumption for an obese group with a normal BMI group and found a lower correlation in the obese group (r=0.76 compared with r=0.89 for the normal BMI group), though subject groups were relatively small, having ten participants in each. Brookes et al. (76) tested a group of seventy-two middle-aged participants with mean BMI of 26.0 ± 4kgm^{-2} and reported R^2=0.61 for overground walking activities, which is lower than the R^2=0.82 (71) and above reported by those using small homogeneous groups discussed above.

The participant group (n=70) in the study by Swartz et al. (73) was relatively varied in terms of age (41 ± 15 years) and BMI (26.0 ± 5.4kgm^{-2}), and this study returned an R^2 value of only 0.343. However, EE prediction models derived from experimental protocols involving multiple activities have returned significantly lower correlation coefficients (217) than those based on walking and running, and when linear prediction models are applied to multiple activities they tend to overestimate EE for certain activities while underestimating EE for others (242). The low R^2 value returned by Swartz et al. can be largely attributed to the number and variety of activities in the experimental protocol (the study considered twenty-eight varied activities including cooking, tennis, softball, and gardening activities). It is not clear, therefore, how the variance in the linear model was affected by more varied subject attributes. Crouter et al. (77) tested the validity of fourteen linear EE prediction equations using a subject group of forty-eight individuals with varied BMI (25.8 ± 5.2 kgm^{-2}). The study concluded that all equations underestimated vigorous physical activity, though once again multiple activities were considered and the effect of BMI on the results is therefore unclear.

When considering a heterogeneous population of individuals, the physical and physiological factors described above may explain more of the variance observed in the model relating accelerometer counts to EE. Prediction accuracy may consequently be improved by incorporating additional subject attributes in the EE prediction model. This would have positive implications for studies involving obese or mixed BMI groups, such as weight loss interventions. In order to identify which additional factors improve the EE prediction model, a diverse subject group is required, and a single activity must be considered in isolation, as multiple activities are a confounding factor in assessing the model's performance.

2.5.3 Attempts to Improve on the Basic Linear Model for Predicting Energy Expenditure

Nonlinear approaches to improving EE prediction have been investigated. Crouter et al. (243) used two separate nonlinear regression equations: one for lifestyle activities (including computer work, lawn mowing, washing dishes, basketball, and so on) and a separate nonlinear regression for locomotive activities (comprising walking and running). The algorithm used the coefficient of variation of the accelerometer signal to decide which equation to use. Considering eighteen activities in the experimental protocol, the study found improved accuracy over three linear models (these being the Freedson (71), Swartz (73) and Hendelman (217) linear equations). However, this improvement in accuracy may be due to the prediction equations being well suited to the chosen configuration of activity types and durations in the testing protocol. If the times spent in each activity mode were changed, it is not clear that there would still be an improvement over the linear models. Also, it is not clear whether the apparent improvements were due to the two-phase approach, or the nonlinear model innovation.

Su et al. (233) applied support vector regression techniques to accelerometer output and EE collected from treadmill walking. Using this approach they achieved an improvement in EE estimation over linear models. However, the subject group was small (n=11) and all subjects were below the threshold for obesity, so this model has not been sufficiently tested against a diverse subject group. Chen and Sun (114) applied linear and nonlinear models to accelerometer data collected over two days for a mixed protocol of sedentary and lifestyle activities (walking, stepping, static cycling, and everyday routine activities). The study also incorporated four parameters based on subject attributes, and used the novel approach of combining the horizontal accelerometer axes. They found improved energy expenditure estimates over the proprietary algorithms used by their accelerometer software. However, the extent to which each of their three innovations (nonlinear equation, subject attributes in the prediction model, and combined horizontal axes) was individually responsible for this improvement is not clear.

Rothney et al. (103) were the first to apply an artificial neural network to energy expenditure estimation. The study considered ten parameters from which the neural network made predictions, five of which were based on subject attributes, the other five were accelerometer signal features used as an alternative to accelerometer counts. The study reported improved accuracies over two linear prediction models. Staudenmayer et al. (104) also developed an artificial neural network to estimate MET energy expenditure for a programme of mixed intensity activities and achieved an RMSE of 1.22 METs. The study employed statistical accelerometer features as inputs to the ANN instead of accelerometer counts, and the model was subsequently validated against a larger (n=277) diverse subject group and mixed activity set by Freedson et al. (244), this time achieving an RMSE of 1.9 METs. Although these two studies show the promise of using an ANN to estimate EE, the RMSE values returned suggest that the level of accuracy is not currently adequate for field applications involving diverse participants.

There have been certain studies that have incorporated demographic, anthropometric and physiological attributes of individuals in addition to accelerometer counts and body mass. Heil (214) incorporated age, gender, height and weight in regression models, while Rothney et al. (103) included gender, age, height, body mass, and ethnicity as parameters to their neural network EE prediction model. The study by Chen and Sun (114) considered age, gender, height, body mass, body fat percentage, fat mass, fat-free mass, and residual lung volume; although, as a result of stepwise linear regression analysis, only the first four of these variables were included in their model. The additional parameters examined in these studies have been limited in number and variety, and many of the factors discussed in 2.5.2 have been neglected. In addition, the effect of these parameters is not clear as they are used in conjunction with other innovations, such as an ANN or a nonlinear equation.

Each of the studies described in the paragraphs above has tested prediction algorithms against a programme of multiple activities. However, the relationship between accelerometer counts and EE differs across modes of activity, which, for linear models, suggests that a separate regression equation should be applied to each individual activity mode (245). This also implies that a classification phase should first be employed to decide which prediction equation is chosen. The studies involving ANNs did not attempt to first identify activity type before applying an appropriate activity-specific EE prediction algorithm. However, because the accelerometer features may contain characteristic information pertaining to activity type, they may be implicitly accounting for activity type and consequently mitigate the problem of many activities, and this may to some extent explain any improvement in EE estimation reported. However, this effect is limited by the ability of the chosen features to both distinguish different activities and estimate energy expenditure.

Other attempts to improve EE prediction models have involved combining accelerometers with additional sensors. In particular, heart rate combined with accelerometer output has been shown to improve EE estimation over accelerometry alone (246-247). However, reliable heart rate monitoring requires electrodes to be fitted to the chest, which increases the burden to individuals being monitored. Heart rate is a dynamic physiological attribute requiring constant monitoring to be of use in EE estimation. Consequently, using combined accelerometer and heart rate data to estimate EE increases the complexity of the data analysis (248). In contrast, the physiological and anthropometric attributes discussed above remain static, or change gradually over prolonged periods, and therefore do not greatly increase the complexity of the prediction model. Additionally, there is no requirement for extra sensors in the field.

2.5.4 Research Question

There have been a number of attempts to improve on the basic linear model for estimating energy expenditure from accelerometer output, as described above. These studies have applied nonlinear equations and neural networks to this purpose. Where these studies have addressed the need for additional parameters based on demographic, physical and physiological attributes of individuals, they have not included a sufficient number of appropriate subject attributes in the prediction model to fully test their effect on prediction accuracy. Also, the effect of these additional parameters is often not distinguishable from the other innovations that each study has implemented in parallel, or is confounded by the issue of multiple activities. Further research is needed, therefore, to investigate whether the addition of anthropometric and physiological parameters to the prediction equations can improve the capacity of the basic linear model to estimate EE.

The research question is asked: can EE estimation accuracy be improved by the addition of anthropometric and physiological attributes to the prediction model?

In order to clearly assess the effects of subject attributes on the relationship between EE and accelerometer counts, a single activity needs to be considered in isolation. Walking is of primary interest as it is the most common physical activity, and is practical for obese participants to undertake in both experimental and free living settings. If physiological and anthropometric attributes are identified which improve EE prediction for walking, then it is likely that EE prediction models for other activities would benefit from the addition of such attributes, though the relevant attributes may differ between activities.

3 Classification of Aerobic and Gym-Based Exercises from Accelerometer Output

As discussed in detail in section 2.3.7, classification accuracy varies according to which activities are being tested (section 2.3.7.1), the types of accelerometer feature being used (section 2.3.7.4), and the number and placement of sensors (section 2.3.7.2). The study described in this chapter aimed to establish whether good classification accuracy could be obtained from hip- and ankle-mounted accelerometer data, for both obese and non-obese participants performing a set of activities suitable for an obesity management programme. The study also investigated whether a different approach to feature selection is needed for obese populations when compared to non-obese populations. The research questions were posed as follows:

Research question 1: can a set of aerobic exercises and free-living activities be identified from data collected by a single accelerometer mounted at the hip or at the ankle?

Different activity sets return varying classification accuracies, and this is the case even if the feature set and classification algorithm remain unchanged (160, 249). It is not clear, therefore, whether it is possible to accurately classify a particular set of activities. The study described in this chapter aimed to evaluate classification accuracy for an activity set comprising a variety of lifestyle activities and aerobic/gym exercises. The activity set is intended to be suitable for inclusion in a weight management programme aimed at obese participants. With this in mind, activities were chosen that should not be too difficult for an obese person to perform, that may be performed at low intensities and built up over time, and include common exercises that are used in the gym to lose weight. The activities are described in more detail in section 3.2.2.2.

Although the wearing of multiple sensors and cumbersome equipment may be acceptable in the laboratory, in a field setting activity monitoring equipment should be as unobtrusive as possible. For those taking part in a weight management programme, a single accelerometer placed at an unobtrusive site would minimise the burden experienced by the wearer, and thus aid compliance to the measurement regime. The study, therefore, examined classification accuracy for single-site mounted accelerometers. As discussed in 2.3.7.2, previous research does not provide consensus as to which single accelerometer placement site will provide the best overall accuracy for activity classification. However, hip and lower limb sites have proved effective for identifying activity sets involving whole-body dynamic activities (89, 110, 180). Currently, activity monitors are around the size of a matchbox and are worn using belts and straps, or in some cases affixed directly to skin through adhesives (250). The hip is an unobtrusive body site for use under free-living conditions, as the accelerometer may be attached to a belt and worn as an item of clothing. Lower limb mounted accelerometers may be more obtrusive than those worn at the hip, but still may be acceptable in free-living conditions. Continuing advances in technology mean increased miniaturisation of sensors, and thus reduced obtrusiveness of wearing these devices. However, the subject may perceive affixing multiple sensors as a greater burden than a single sensor, and this may affect compliance. Additionally, data from more sensors contributes to the computational cost of analysing the data, which supports the rationale for a single body site. For this study, two accelerometer sites were chosen for comparison: one accelerometer was affixed to a belt worn around the waist over the right hip, and the other accelerometer was worn at the ankle.

The present study addresses the research question by comparing the classification accuracies achieved for the two body sites from an activity set containing free-living activities, and aerobic and gym-based exercises.

Research question 2: Does activity classification accuracy differ between obese and normal BMI groups?

As discussed in 2.2, accelerometer signals produced by obese individuals performing physical activity may differ from signals generated by their non-obese counterparts. There are two main factors to consider: a surfeit of adipose tissue at accelerometer sites may influence the measured accelerometer signals; and body movements, such as gait, differ between obese individuals and those with lower BMIs. The waist, for example, exhibits higher adipose tissue levels for obese groups, which may affect accelerometer movement and introduce noise to the signal. Similarly, signals taken from ankle -mounted accelerometers may exhibit different characteristics which reflect the differences in how obese and non-obese persons move (119-123), as discussed in 2.2.1. To answer the research question, the study compared classification accuracy across BMI groups for the two accelerometer sites.

Research question 3: do the same accelerometer features apply to obese and normal BMI groups, or do they require different accelerometer features to characterise their physical activities?

For the purposes of activity classification, a feature set is chosen with the aim of exploiting the characteristic differences in the accelerometer output between activities. The prediction accuracy of the classification algorithm greatly depends on how well the feature set captures those characteristics. However, there may also be characteristic differences between obese and normal groups within single activities (as discussed in 2.2.1), and an alternative feature set may be required to effectively distinguish activities depending on BMI group. It is not clear, therefore, whether a particular set of features will apply equally to both obese and non-obese groups. To answer the research question, the study compared the effectiveness of a number of feature sets when applied to obese and normal BMI groups in order to determine whether different sets of features are better suited to one BMI group over the other.

3.1 Research Design

Hip- and ankle-mounted accelerometer data was collected for fifty subjects performing a range of aerobic and gym-based exercises and free-living activities. To answer research question one, classification techniques were applied to the processed accelerometer output to obtain estimates for activity type, and overall classification accuracy was tested for each accelerometer site. To address research question two, the subject group was split into subgroups according to BMI so that classification accuracy could be compared between obese (BMI>30), and normal (18.5<BMI<25) subgroups. Several features were selected, and machine learning techniques were applied. The classification accuracy was ranked for each feature set, and the rankings were compared between obese and normal BMI groups in order to answer research question three.

3.2 Methods

3.2.1 Recruitment and Subject Statistics

The study was given approval by the University of Salford ethics committee. Participants were recruited via a number of avenues: staff and students of the university were approached through email; an approved recruitment message was posted on several online forums and social media sites; local weight loss groups were contacted and were given details of the study. Interested parties were asked to complete the physical activity readiness questionnaire (PAR-Q. See appendix A.) to ascertain their eligibility to take part. It was necessary that physical exercise did not pose any risk to participants' health, and they should not suffer from any pathological conditions which may affect natural movement. Those eligible were asked to attend the university on one occasion for testing to take place.

Overall, the success rate in gaining interest for the study was low. For example, the online forums that were used to post the recruitment message had a total subscription population in the order of five figures, yet responses were to be counted in dozens, and only a fraction of these converted to a completed participant in the study due to either ineligibility or withdrawal. Another factor was that although some areas of pursuit garnered several potential volunteers, these tended to cluster into a number of mainly female groups with similar ages and BMIs in the 25 to 30 kg/m^2 range. This meant that some would need to be rejected to prevent the subject group becoming homogeneous.

Obese persons in particular were difficult to recruit. This was possibly due in part to a tactful approach in the recruitment messages, which merely asked for recruits interested in losing weight rather than those who were obese, but also, obese people, having a general predisposition for sedentary activity, may be less likely to volunteer for a study involving several exercises. Weight loss groups were targeted as they were likely to contain overweight and obese subjects who were motivated to perform physical activity. Attempts to contact several weight loss organisations and their franchises were made, including Weight Watchers, Lighter Life, Slimmer's World, and Rosemary Conley. Of these, a single Rosemary Conley franchise allowed access to their membership.

Other avenues for recruitment were pursued without success. Salford Weight Management Services advised that recruitment through them would require NHS approval, which would not be easy to obtain. Also, ABL Health, a private organisation subcontracted by the NHS to provide weight management services across the Northwest, denied a request for access to their clients.

Though the recruitment process was fraught with difficulty, through perseverance the target number of fifty participants was successfully achieved, with obese, overweight and normal BMI groups being well represented.

Fifty subjects (21 male, 29 female) completed the study. Average (mean ± SD) age was 34.6 ± 11.2 yrs, height was 168.6 ± 8.7 cm, body mass was 81.3 ± 16.7 kg, and BMI was 28.7 ± 6.2 kg/m^2. The subjects were considered as being in three groups: normal (BMI<25), overweight (BMI>=25 and BMI<30) and obese (BMI>30). The distribution of subjects across the groups were as follows: 17 normal subjects (9 male, 8 female), 14 overweight subjects (4 male, 10 female), and 19 obese subjects (8 male, 11 female). The difference in numbers between these groups should not have any great impact on accuracy statistics.

It should be noted that subjects who took part in this study were also participants in the studies described in chapters 4 and 5. Data collection for all three studies was carried out simultaneously or as part of the same laboratory visit. Further details regarding this are included in the relevant chapters.

3.2.2 Data Collection

3.2.2.1 Activity Monitors

Figure 10: Accelerometer placement at the ankle and at the waist above the hip.

The Actigraph GT3X+ Accelerometers (as described in 2.1.2.1) were charged and initialised using Actilife5, the Actigraph companion software provided by the manufacturer. The accelerometers initialised by ActiLife obtain their date and time settings. Two GT3X+ activity monitors were fitted to the participant (Figure 10): one at the waist above the right hip using an elasticated belt; the other at the lateral side of the lower leg, immediately above the ankle, using an elasticated strap, and further secured with a length of bandage. The Nyquist Sampling Theorem (251) states that is necessary that the sample rate be twice the maximum frequency occurring within the signal being sampled in order to fully capture the signal information and avoid issues such as aliasing. As the frequency of both trunk and ankle movement in humans lies below 25Hz, each GT3X+ was, therefore, initialised to sample at the rate of 50Hz. This frequency of 50Hz is considered sufficient to capture human movement (252).

3.2.2.2 Activity Set

A selection of common aerobic and gym activities were chosen according to their suitability for obese persons to perform. Aerobic activities have proved effective in weight loss (30, 35, 40) and may be performed at a range of intensities; therefore, the activities chosen for the study were predominantly of this type. To evaluate the effectiveness of a weight loss programme being undertaken in free-living conditions, a comprehensive picture of all participant activity is required over the period of the assessment. For this reason, a number of free-living activities were included in the activity set, so that both exercise and daily activity were represented when evaluating the classifier. The activities are described below.

Walking is the most common dynamic physical activity and consequently features in numerous previous activity classification studies (80, 84, 89, 113, 117, 140, 144, 146). An individual undertaking a weight-loss programme may perform walking as both a means of transportation and also as a gym exercise using a treadmill. Both overground walking and treadmill walking data were collected from participants in the present study. It was intended that the classification algorithm should identify walking regardless of whether it was performed on a treadmill or overground and irrespective of speed.

Treadmill walking was performed using the Ergo ELG55 treadmill (WOODWAY GmbH of Weil am Rheine, Germany). Each participant was required to walk at four different speeds, ranging from slow to fast (approximately between 1.0 ms^{-1} and 1.7 ms^{-1}), for five minutes each. These speeds were determined according to the participant's abilities, and based upon a timed walk to ascertain normal walking speed. Participants were not permitted to use the hand rail. Participants were also asked to walk a designated route through the university campus at a self-selected speed. The duration of the walk was approximately three minutes, depending on their walking speed. The walking surface was paved, and in some places uneven or sloped.

It should be noted that the same treadmill walking data were used for both the present study and the studies described in chapters 4 and 5. Also, for the purposes of the study described in chapter 5, a mask was worn to measure respiration while the treadmill walking data was collected.

Jogging, as considered by previous activity classification research (87, 147) and used in exercise interventions (30, 41), was performed at a self-selected speed for one minute on the treadmill. Subjects were allowed to alter the treadmill speed themselves until comfortable. Again, it is important that the classifier should identify jogging regardless of speed. In previous activity monitoring research jogging and running are often performed on a track (87). However, the treadmill was chosen for the following reasons: it was convenient that jogging should take place on the treadmill after the treadmill walking; a suitable track was not readily available; obese participants may have felt embarrassment while jogging in an open area instead of the private confines of the laboratory; also, individuals taking part in an obesity management programme involving gym exercises would be likely to use a treadmill for jogging.

Three common gym exercises that have been used previously in activity classification research were chosen: cross-trainer (otherwise known as elliptical trainer) (253-255), rowing machine (84, 117), and static cycle (84, 117, 256). These were selected as they are easily accessible at any gym, and can be performed at a range of intensities, making them ideal activities for obese people to perform. These exercise machines were set at low to moderate resistances according the participant’s preference, and each exercise was performed at a self-selected intensity for one minute each.

Step aerobics (stepping up and down on a step), as used in previous exercise interventions (26, 47), was performed for one minute at a regular pace set by the participant. The step can be adjusted to be one of two heights; the lower of the two heights was chosen in order to reduce the effort required, so that the heavier subjects were able to complete the activity. The activity was first demonstrated to the participant, but they were free to choose which foot to begin the activity with.

Stair climbing is a common free-living activity which has been considered by several previous activity classification studies (110, 144, 146, 160). Participants were asked to climb and descend stairs between four and six times (laps) depending on their fatigue level – after the four laps the participant was asked whether they felt able to perform two more laps – but all participants completed a minimum of four. They were asked to ascend a single flight of stairs, to turn around on reaching the top, to descend the same flight of stairs, then turn around again to repeat. The start time and end time of the stair walking activity as a whole was noted, but it was considered more practical to determine the start and end times for each individual ascent and descent by later examining the accelerometer record, rather than attempt to note them manually as they were performed. Participants were, therefore, asked to pause briefly between each transition to make it easier to isolate each instance of stair walking.

Two aerobic exercises were performed for one minute each: sidestepping, and side stretching. Sidestepping involves initially standing with legs apart, then moving one foot to meet the other before returning to its initial position. This is repeated for alternate feet so that the body performs a regular side to side movement. Side stretching begins with the participant standing with legs apart. An arm is raised and the body bends sideways towards the opposite side of the raised arm. The un-raised arm stretches towards the ground and the toe on the opposite leg becomes pointed. The movement is repeated with alternate arms. Each aerobic exercise was demonstrated to the participants.

Except for treadmill walking the intensity of all activities was set by the participant, though they were instructed to maintain a consistent level of intensity – treadmill walking intensity was determined by the four speeds, which varied between participants according to their fitness levels. Activity intensity was not predetermined because the aim was to develop a classifier able to identify activities irrespective of their intensity. This is important as individuals undertaking an exercise-based weight loss programme would not all exercise at the same intensity, and may increase exercise intensity as their fitness improved.

Static activities, specifically sitting, standing and lying, were not included as part of the activity set to be tested by the classifier. These activities do not contribute to weight loss, and an alternative approach to classification is more suitable to detect these activities, as discussed in 2.3.7.1.

3.2.2.3 Annotations

Throughout the programme of activities, start and end times were recorded for each activity using a wristwatch. Before testing, each Actigraph GT3X+ was initialised so that it was synchronised with the computer, and therefore the Actilife5 software. Any offset between the wristwatch time and the computer time was noted when initialising the accelerometer so that hand-annotated labels could be later synchronised with accelerometer output.

3.2.3 Data Analysis

3.2.3.1 Data Pre-processing

Accelerometer data was downloaded from each unit using Actilife5 software provided by Actigraph. The software converts the proprietary data format to comma separated values (CSV) so that it may be imported into MATLAB through standard file reading techniques. Ankle and hip output for each subject was stored as three vectors of acceleration values (in units of g) for the vertical, anteroposterior, and mediolateral accelerometer axes. Alongside these accelerometer data, the sample interval in seconds (being the inverse of the sample rate in Hertz) and the date and time of the first sample were stored; from these, each three-dimensional sample could be located in time and synchronised with other time-stamped data.

3.2.3.2 Labelling

From the paper-based record taken at the time of testing, a list of activity labels was manually compiled for each participant using an electronic spreadsheet application; this records the activity descriptions and timestamps for the start and end of each activity, and adjusts these timestamps according to the previously noted offset in time between the wristwatch and the accelerometer time, so that both synchronise. The spreadsheet information was imported to MATLAB and stored in a data structure which records the description of each activity and its start and end times.

Representative data samples of each activity were extracted on a subject-by-subject basis. To achieve this, the accelerometer output for a subject was plotted in MATLAB, and the activity labels were visually indicated on the graph according to the label file. A selection of each labelled activity was manually identified using on-screen clicks, and based on these selections, accelerometer output and corresponding label values were extracted.

3.2.3.3 Segmentation into Windows

A study by Bonomi et al. (85) compared the results of using six window sizes (0.4s, 0.8s, 1.6s, 3.2s, 6.4s, 12.8s) and found that classification accuracies tended to increase with window size, and the best accuracies were obtained from 6.4s and 12.8s windows. However, the lowest and highest accuracies were not greatly different, ranging from 90.4% for 0.4s windows to 93.1% for 6.4s windows. Alternatively, Huynh and Schiele (147) compared five window sizes (0.25s, 0.5s, 1s, 2s and 4s), and although marginally better classification results were obtained from one and two second windows, they found that precisions varied by individual activity depending on window length, and for the six activities under consideration they concluded that there was not “a single window length that would perform best across all activities”. For the present study, a window length of two seconds was chosen, as has been used in previous research (182, 257-258). This window length was considered to be enough time to capture a representative sample of an activity, and lies in the mid-range compared with most previous studies.

The dataset for each accelerometer axis was segmented into contiguous windows of two seconds in length. For each subject, the number of windows of activity data was limited to a maximum of thirty windows per activity; these windows were selected at random using MATLAB functions. The threshold of thirty windows was chosen as most activities had around one minute of data per subject, with the exception of walking and stair walking. The number of windows was limited in this way to avoid biasing the classifier towards an activity that has many windows of data. For example, there are many windows of walking data, whereas stair climbing has much fewer windows as this is an activity where participants fatigued quickly. A second consequence of limiting the number of windows is that the bias toward *subjects* with greater numbers of windows of activity is also mitigated.

Accelerometer data for each subject were stored in a MATLAB data structure suitable for processing by the classification algorithm. For each of the three accelerometer axes the data was stored as an n x 100 matrix, where n was the total number of windows of activity, and 100 was the number of acceleration samples per window –that is, 100 samples in a two second window of data sampled at 50Hz. For each window of 100 samples, a corresponding activity class label was stored. The labels served as ground truth against which the activity class estimates were compared to obtain measures of classification accuracy.

3.2.3.4 Feature Generation

A number of feature sets were chosen for the present study in order to compare classification accuracies returned for each. Many different combinations of feature choices have been used in previous studies, a selection of which is described in 2.3.7.4. For the present study, it was necessary that the chosen feature sets be sufficiently varied in order to investigate how different features affect classifier accuracy according to which BMI group is under consideration. Ten feature sets were considered for the study; these are shown in Table 2.

Feature Set		Details	Total No of Features
F1	Mean and SD	Mean and standard deviation	6
F2	Bao and Intille	Mean, energy, frequency-domain entropy, and correlation of acceleration data	12
F3	Baek et al.	Mean, standard deviation, kurtosis, skewness, and eccentricity	15
F4	Huynh and Schiele	Mean, variance, energy, entropy, pairwise addition of 20 FFT coefficients	34
F5	Fast Fourier Transform (FFT)	The first 25 FFT coefficients	75
F6	Discrete Cosine Transform (DCT)	The first 25 DCT coefficients	75
F7	Custom Feature Set 1	Interquartile range, signal magnitude area. 10th, 25th, 75th, and 90th percentiles, The first 5 DCT coefficients, mean, standard deviation	37
F8	Custom Feature Set 2	5th, 10th, 25th, 75th, 90th, and 95th percentiles, the first 5 FFT coefficients, the number of signal peaks, zero-crossing rate	39
F9	Custom Feature Set 3	Mean and standard deviation, the first five cepstrum coefficients, root mean square.	24
F10	Custom Feature Set 4	Eccentricity, kurtosis, skewness. 10th, 25th, 75th, and 90th percentiles. The first 5 DCT coefficients, mean, standard deviation, zero-crossing rate, signal magnitude area.	46

Table 2: Feature sets used to test the classifier on the activity set. Note that the features were applied to three accelerometer axes.

The features chosen in the custom features sets were based on those used in previous research. Mean (111, 117, 140, 146, 160) and standard deviation (84, 89, 140, 142, 146) (feature set F1) were chosen due to being two of the most frequently used features in previous research. The Bao and Intille feature set (111) used for feature set F2 was a clear choice due to its popularity in previous research (142, 144). Huynh and Schiele (147) considered some of the features used by Bao and Intille (111) (mean, energy and entropy) with the addition of variance and several values obtained from the FFT of the accelerometer signal. The majority of the features found in Huynh and Schiele (147) were used for feature set F4. Feature set F3 was taken from Baek et al.: in addition to mean and standard deviation, skewness, kurtosis and eccentricity were selected.

Fast Fourier transform (FFT) coefficients were selected to capture frequency characteristics of the accelerometer signals. Although FFT coefficients are commonly used to compute features such as entropy and energy (as discussed in 2.3.7.4) in the case of F5 each component is used as a separate feature. A study by Preece et al. (182) used the first five FFT coefficients per accelerometer signal as features and found this feature set returned the highest accuracy when compared with six other feature sets. This agreed with the previous study by Huynh and Schiele (147) who found that FFT coefficients consistently ranked among the best discriminating features, although they concluded that the highest performing FFT coefficient was different depending on activity. Preece et al. also investigated the effect on accuracy of increasing the number of FFT components, and concluded that optimal classification accuracy was obtained from six components, with little increase in accuracy to be gained by adding additional components. However, contrary to this, preliminary results from a pilot study achieved significantly higher accuracies when the number of components was increased from five to twenty-five; beyond twenty-five coefficients, increases in accuracy were not justified by the additional computational cost incurred. For this reason, feature set F5 comprised twenty-five FFT components per accelerometer axis.

Discrete cosine transformation (DCT) coefficients were also considered as an alternative frequency-domain feature to the FFT. An activity classification study by He and Jin (112) used the DCT to generate features and found that for sixteen, twenty-four, thirty-two, and forty-eight coefficients per accelerometer axis activity recognition accuracy increased as the number of coefficients increased. At sixty-four coefficients there was a slight decrease in accuracy compared to forty-eight, and they concluded that more coefficients would not necessarily mean better accuracies as higher frequency DCT components are closely related to signal noise. For feature set F6, twenty-five DCT coefficients were chosen to keep computational cost within acceptable levels and to match the number of components in the FFT feature set, thus allowing a more valid comparison between the two feature sets.

When devising the custom feature sets it was decided that they should incorporate both frequency-domain and time-domain features. The application of frequency-based features to the activity protocol is appropriate as many of the activities are cyclic in nature such as walking, rowing, cycling, and so on. Also, it was clear from a visual representation of the accelerometer output that statistics such as the mean and standard deviation would be useful in distinguishing activities. Further features used in the custom feature sets were taken from previous studies, and the four custom feature sets used combinations of these.

Custom feature set F7 incorporated several common statistical measures (mean, standard deviation, interquartile range, and percentile values). Also, for frequency-domain features DCT coefficients were chosen. These coefficients were limited to five in number so that they did not overwhelm the feature set, and also so that the feature set did not contain too many elements – each additional frequency coefficient would add three features as there are three accelerometer axes. The signal magnitude area (SMA), which is the sum of the absolute values of the accelerometer signal, has been used as a feature in activity classification (80, 142). The SMA was selected as the final feature in feature set F7, making thirty-seven features in total across the three accelerometer axes.

Custom feature set F8 also contained five frequency coefficients, though these were obtained from an FFT of the accelerometer signals. Again, percentile values were included. The remaining features were the number of peaks in the signal, and the zero crossing rate (172), giving a total of thirty-nine features across three accelerometer axes.

Custom feature set F9 used the first five cepstrum coefficients for frequency domain features. A cepstrum is the inverse Fourier transform of the log of the Fourier transform of the original signal. Cepstral analysis has been previously applied to activity classification by Lester et al. (146). Mean and standard deviation were again chosen. Also, the root mean square (167) of each accelerometer signal was selected.

Feature set F10 used the feature set selected by Baek et al. (110) as a basis (mean, standard deviation, eccentricity, kurtosis, and skewness) then added five DCT components to capture frequency characteristic, zero-crossing rate, signal magnitude area, and percentile values to extend the statistical information, with the aim of improving classification over the original feature set.

3.2.3.5 Classification

Linear Discriminant Analysis (LDA) was selected as the classification algorithm, as it had given high accuracies in previous pilot testing. As discussed 2.3.5, LDA defines a probability distribution for each possible class of activity. A window of activity data is classified by applying a discriminant function which returns the likelihoods of the window belonging to each activity distribution. The activity that is chosen by the classifier is the one returning the highest likelihood value from the discriminant function. The prior probabilities of each activity can be set to allow weighting of one activity over another, and thus make the classifier more likely to choose certain activities. For example, under free-living conditions, walking is more likely to occur than rowing, and this can be accounted for in the prior probabilities. For the present study, all activities were presumed to have equal prior probabilities as the number of windows of each activity was made as equal as possible, and the activity protocol was not designed to weight the times spent on each activity to reflect everyday life.

Classification was performed independently for the ankle and hip sites, and for each site the ten feature sets were evaluated separately. The results were obtained using leave-one-out cross-validation (see 2.3.6); each subject was tested individually using a training dataset comprising data from the remaining normal BMI, overweight, and obese subjects. From the results, a confusion matrix was constructed for the entire subject group. The sensitivity for each activity (as discussed in 2.3.6) was calculated in order to give an indication of which activities were better recognised by the classifier/feature set combination.

A table of results was constructed for each accelerometer site to show classification accuracies by BMI group, and by feature set. From these tables, overall classification accuracy was compared between hip and ankle sites in order to answer research question 1. The separate classification accuracies for each BMI group, as provided by these tables, were compared in order to answer research question 2. To ascertain whether the same feature sets apply equally to obese and normal BMI groups (according to research question 3), accuracy rankings for the ten feature sets were compared, and Kendall's Tau statistic was applied to obtain a measure of correlation between feature rankings for each BMI group. A good correlation between rankings would suggest that the same features may be applied across BMI groups.

3.3 Results

Table 3 below shows a matrix of results grouped by accelerometer site. The results are divided by the BMI group against which each feature set was tested. Table 4 shows the rankings of the feature set in terms of highest accuracy.

Accelerometer Site:	Test dataset:	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Hip	Entire Group (n=50)	63.0%	64.2%	73.2%	76.7%	78.3%	83.6%	81.8%	81.8%	66.6%	85.0%
	Normal (n=17)	64.4%	66.8%	72.9%	77.2%	75.7%	82.3%	81.3%	80.7%	66.9%	84.5%
	Obese (n=19)	61.0%	61.1%	70.3%	73.7%	76.7%	80.9%	78.6%	79.6%	64.9%	81.9%
	Overweight (n=14)	64.2%	65.1%	77.4%	80.2%	83.4%	88.8%	86.8%	86.3%	68.4%	90.0%
Ankle	Entire Group (n=50)	82.1%	84.5%	88.1%	93.0%	92.0%	92.3%	93.4%	94.1%	86.3%	93.9%
	Normal (n=17)	81.8%	83.3%	86.4%	90.9%	90.3%	90.5%	91.0%	91.4%	86.3%	91.6%
	Obese (n=19)	83.1%	84.2%	88.0%	93.8%	92.3%	93.3%	94.6%	95.7%	86.1%	95.0%
	Overweight (n=14)	81.2%	86.3%	90.3%	94.3%	93.7%	93.2%	94.5%	95.1%	86.8%	95.2%

Table 3: Accuracy % values for each feature set by placement, by test dataset group

Accelerometer Site:	Test dataset:	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Hip	Entire Group (n=50)	10	9	7	6	5	2	4	3	8	1
	Normal (n=17)	10	9	7	5	6	2	3	4	8	1
	Obese (n=19)	10	9	7	6	5	2	4	3	8	1
	Overweight (n=14)	10	9	7	6	5	2	3	4	8	1
Ankle	Entire Group (n=50)	10	9	7	4	6	5	3	1	8	2
	Normal (n=17)	10	9	7	4	6	5	3	2	8	1
	Obese (n=19)	10	9	7	4	6	5	3	1	8	2
	Overweight (n=14)	10	9	7	4	5	6	3	2	8	1

Table 4: accuracy ranked by feature set for combinations of accelerometer placement and test dataset

Table 5 and Table 6 below show confusion matrices for the two best performing features sets at the hip and ankle (F10 and F8 respectively).

F10 Hip	Walk	DownStairs	UpStairs	Cycling	Rowing	Cross-trainer	Jog	Stepping	Sidestepping	Sidestretching	Sensitivity
Walk	1309	71	72	19	0	0	3	4	7	15	87.3%
DownStairs	93	694	68	0	0	2	10	6	0	12	78.4%
UpStairs	56	42	741	1	0	14	10	43	9	15	79.6%
Cycling	9	0	22	1419	0	4	0	0	8	35	94.8%
Rowing	0	0	1	39	1419	0	0	4	0	35	94.7%
Cross-trainer	1	6	70	39	0	1249	10	0	79	41	83.5%
Jog	59	19	0	0	0	0	1386	0	0	30	92.8%
Stepping	18	34	105	9	0	0	0	1245	41	39	83.5%
Sidestepping	1	17	26	46	0	5	0	26	1190	189	79.3%
Sidestretching	0	0	2	213	0	0	0	36	173	1073	71.7%

Table 5: Confusion Matrix for Hip, for F10 feature set, for all subjects

F8 Ankle	Walk	DownStairs	UpStairs	Cycling	Rowing	Cross-trainer	Jog	Stepping	Sidestepping	Sidestretching	Sensitivity
Walk	1435	3	3	1	0	0	11	40	0	7	95.7%
DownStairs	0	852	12	0	0	0	0	18	0	0	96.6%
UpStairs	0	19	798	6	0	0	0	154	1	0	81.6%
Cycling	0	0	31	1355	0	104	0	0	0	2	90.8%
Rowing	0	0	0	0	1496	0	0	0	0	0	100.0%
Cross-trainer	0	0	0	5	30	1463	0	0	0	0	97.7%
Jog	4	0	0	0	0	0	1491	0	0	0	99.7%
Stepping	0	17	118	0	0	0	0	1356	0	9	90.4%
Sidestepping	0	0	0	0	0	0	0	4	1327	169	88.5%
Sidestretching	0	1	0	0	0	16	0	0	31	1435	96.8%

Table 6: Confusion Matrix for Ankle, for F8 feature set, for all subjects

Hip	All	Obese	Normal	Overweight
All	1.0000	1.0000	0.9111	0.9556
Obese		1.0000	0.9111	0.9556
Normal			1.0000	0.9556
Overweight				1.0000

Table 7: Kendall's Tau: hip accelerometer, trained on all subjects, correlation of rankings is shown between test groups

Ankle	All	Obese	Normal	Overweight
All	1.0000	1.0000	0.9556	0.9111
Obese		1.0000	0.9556	0.9111
Normal			1.0000	0.9556
Overweight				1.0000

Table 8: Kendall's Tau: ankle accelerometer, trained on all subjects, correlation of rankings is shown between test groups

Table 7 and Table 8 above show the Kendall's Tau statistic which measures the amount of correlation for the ranks of the ten feature sets between BMI groups. Each group can be seen to be well correlated in terms of feature ranks, and this is the case for both Ankle and Hip accelerometer sites.

3.4 Discussion

The study aimed to address the following research questions:

- Research question 1 (RQ1): can a set of aerobic exercises and free-living activities be identified from data collected by a single accelerometer mounted at the hip or at the ankle?
- Research question 2 (RQ2): Does activity classification accuracy differ between obese and normal BMI groups?
- Research question 3 (RQ3): do the same accelerometer features apply to obese and normal BMI groups, or do they require different accelerometer features to characterise their physical activities?

In order to answer the research questions, fifty subjects performed ten dynamic activities while accelerometer data was collected at the ankle and hip. Using the LDA classifier, ten sets of features were tested for classification accuracy across the following subject groups: all subjects (n=50); obese subjects (n=19); normal BMI subjects (n=17); overweight subjects (n=14). The results are shown in Table 3. Prediction accuracies of over 80% and over 90% were achieved for the hip and ankle accelerometer sites respectively, which suggest that it is possible to accurately identify an activity set such as this using a single accelerometer at either the ankle or the hip (RQ1). Overall accuracy did not differ greatly between BMI groups (RQ2), and the feature sets ranked similarly for each group when trained on mixed BMI groups, suggesting that a single feature set is sufficient across BMIs (RQ3).

3.4.1 Research Question 1

Can a set of aerobic exercises and free-living activities be identified from data collected by a single accelerometer mounted at the hip or at the ankle?

The entire subject group comprising the three BMI groups (normal, overweight, and obese) was used as training data, as the test subjects were also of mixed BMI. Ankle-mounted accelerometer data returned the best results returning an overall prediction accuracy of 94%. The recognition accuracy obtained from hip-mounted accelerometer data was lower, but still achieved 85% accuracy. As can be seen from Table 9, these results compare favourably with previous research. Direct comparison of accuracies between studies is difficult, however, due to the combined effects of several factors on the outcomes of an activity classification study. In order to consider the results of the present study in the context of the literature, it is necessary to discuss these factors.

3.4.1.1 Factors Affecting Accuracy Comparisons Between Studies

Classification accuracy varies depending on the number and types of activity being tested. In the present study, the activity set itself was intended to be novel, having been chosen with obesity management in mind, and therefore cannot be compared directly with previous research. Recognition accuracies for single activities that are common between studies, such as walking, still may not be compared, as individual accuracies are also dependent on the activity set as a whole. For example, if an activity set comprised only the two diverse activities of walking and sitting, then a high detection accuracy for walking would be expected, whereas given the three similar activities of walking, stair ascent, and stair descent, walking may not be distinguished as accurately. A between study comparison of the detection accuracies for walking (when considered in isolation from a mixed activity set) would, in fact, reveal little about the relative abilities of either classification scheme to detect walking.

Another particular issue when considering classifier accuracy is that the inclusion of static activities in the activity set can improve overall accuracy figures. As discussed in 2.3.7.1, static activities are more easily distinguished as they can be identified using simple thresholds, and do not require sophisticated classification approaches or complex accelerometer features to characterise them. Consequently, higher overall classification accuracies are likely to be reported for activity sets which include static activities. There are no static activities in the present study, so the activity set may be considered more challenging to classify with high accuracy.

The number and location of body-worn accelerometers also affects classification accuracy. The combined output from multiple accelerometers at different sites improves classification accuracy results over a single accelerometer site (177-178). Also, classification accuracy due to placement is closely connected to the activity set being tested. For example, a distinction between “sitting reading” and “sitting watching television” (111) would not be possible with a single hip-mounted accelerometer; a differentiating factor for these activities is the movement of the arm, thus requiring an arm-mounted sensor. The potential difficulty in classifying activities to this level of detail is, therefore, mitigated by the increased number and appropriate positioning of the sensors used. The present study considers classification accuracy for a single accelerometer site, and is consequently at a disadvantage in this area.

Prediction accuracy is affected by the number of subjects used to train and test the classification algorithm. Generally, classification accuracy improves when a greater number of subjects – and, therefore, more data – is used in the training dataset. In contrast, intra-subject classification, where the same subjects are used to both train and test a classification algorithm, results in higher accuracies because there are no inter-subject differences that the classifier must account for; consequently, an algorithm generated through this approach may not generalise to an inter-subject classification scheme. The high overall classification accuracy of 97.5% reported by Baek et al. (110) is greatly due to using a single subject to train and test the algorithm. Similarly, Ravi et al. (144) obtain accuracies between 90% and 99% for variations of intra-subject classification on two subjects, but this reduces to 73% when the algorithm is trained by one subject and tested on another. As the present study employed leave-one-out cross-validation of the classifier, it ensured that data for a subject was never tested using training data from the same subject. Also, a subject group of fifty was appropriate to test the algorithm as an inter-subject classifier.

The type of classification algorithm employed has a significant influence on recognition accuracy. However, as discussed, classification accuracy is greatly affected by the particular activity set under consideration, the selected feature set, the chosen configuration of sensors, among other factors. This means that it is not valid to compare classifiers between studies where these factors vary. To illustrate this, where Bao and Intille (111) achieved 84% accuracy from a decision tree and 52% from a naive Bayes classifier, Ravi et al. (144) reported 64% from naive Bayes and only 57% from a decision tree. Similarly, for Parkka et al. (84) a decision tree (86% accuracy) outperformed a neural network (82% accuracy), but the converse was the case for Ermes et al. (117), with a neural network achieving 87% compared to 60% returned by a decision tree. The present study was not concerned with comparing different classifiers as this would not have any bearing on the research questions; there is little reason to suppose that different classifiers would perform differently depending on the BMI of the participant group. The classification algorithm, therefore, needed only to perform well enough to answer the research questions. The chosen LDA algorithm returned high classification accuracies from the activity data, and was, therefore, suitably adequate.

The number and type of features selected greatly affects classification accuracy. As discussed in 2.3.4.1, classification accuracy can improve as the number of features increases provided that there is sufficient training data to avoid the “curse of dimensionality” issue. However, although an insufficient number of features will be detrimental to classifier accuracy, a small number of well chosen features may result in better performance than a large selection of poorly selected features. Ultimately, the optimum number of features for use in classifiers cannot be satisfactorily ascertained, as this varies depending on the classification algorithm and type of features utilised.

Additionally, the effect of the selected features on accuracy is connected with the number of axes per accelerometer. Studies have used one axis (uniaxial) (160, 259-260) and two axis (biaxial) accelerometers (261-262), but advances in the technology of activity monitors have meant that more recent research mainly utilises three axis (triaxial) accelerometers. Clearly, the effect of the number of accelerometer axes on accuracy depends on the number of accelerometers used, and also the number of features extracted from each. Bao and Intille (111) employed five biaxial accelerometers at key body sites and selected seven features from each, giving a rich dataset of thirty-five features. Ermes et al. (117), however, used two triaxial accelerometers and selected only six features in total from the six axes available. Between six and seventy-five features were extracted from a triaxial accelerometer in the present study.

The sampling rate can also have an effect on classification accuracy. The relatively low overall accuracy of 69% achieved by De Vries et al. (101), for example, is most likely attributed to the 1Hz sample rate which will necessarily give rise to an accelerometer signal lacking the detail required to effectively distinguish activities. On the other hand, signals sampled above 50Hz are unlikely to yield any additional information about the activities they represent. In fact, Maurer et al. (167) found that recognition accuracy improved as sample rate increased until it stabilised between 15-20Hz with little gain in accuracy being observed for higher rates. Related to sample rate, the window size can also influence classification accuracy, as discussed in 2.3.3. Comparison of studies is further impeded by the different approaches to the calculation of measures of accuracy that are employed; some studies elect to use overall accuracy (82, 89), whereas others have considered other accuracy measures such as sensitivity (180).

In some cases the reported recognition accuracy figures have a strong element of bias. As discussed in 2.3.6, overall accuracy is calculated by taking the total number of correctly predicted windows and dividing this by the sum of windows for all activities. This is a good measure of accuracy providing the number of windows for each activity is similar; otherwise the accuracy of the classifier is biased towards those activities that are represented the most. In the study by Bao and Intille (111), the number of windows for each activity varies between 180 and 1,441. There are 1,047 windows of “reading”, which has a sensitivity of 92%, whereas there are only 309 windows of “stretching", which returns only 41% sensitivity. Applying the accuracy calculation to these two activities in isolation would return a combined accuracy of 89%, which when compared to their average sensitivity measure of 67% illustrates how results can be biased towards the best represented activity. Similarly, scrutiny of the results obtained by Parkka et al. (84) revealed that the quoted figure of 86% accuracy was heavily biased by the large number of windows representing sit/stand, which made up around half of the total windows and individually returned over 95% sensitivity, whereas average sensitivity for the activities was closer to 80%. The study by Ermes et al. (117) which followed on from Parkka et al. also retained this heavy bias towards high performing static activities.

3.4.1.2 Comparison of Classification Accuracy with Previous Research

Having considered the factors affecting classification accuracy and the consequent problems in comparing studies, it may be useful to categorise these factors as either external or internal to the classifier; the external factors define the classification problem, and the internal factors define the approach to the problem solution. External factors include the number and placement of sensors, the activities under consideration, and the number of subjects. Internal factors include the classification algorithm, sample rate, window size, the number and type of features. The classifier is composed of the internal factors and is applied to the data influenced by the external factors. For each study, the classifier is being tested against the external factors. Therefore, if the external factors are similar between studies then results may be compared; the internal factors need not be considered individually as it is the classifier as a whole that is being evaluated against the external factors.

The present study had the disadvantage of a single accelerometer site, as opposed to five accelerometers at different body sites in the study by Bao and Intille (111) and four in the study by Foerster et al. (160). Both of these studies contained activity sets with several static activities, which would be expected to boost prediction accuracy. However, because these were categorised into subtypes such as “sitting operating a computer” (160) and “watching TV” (111) this effect may have been reduced to some extent. Foerster et al. achieved an overall accuracy of 67% for nine activities, though only four were dynamic, but this improved to 81% when selecting only instances of activity which exceeded forty seconds. The 85% and 94% accuracies returned by the present study, for hip and ankle respectively, may be considered superior to those obtained by Foerster et al., particularly when noting that, in addition to the four accelerometers, Foerster et al. used two more sensors to measure vertical movement of the head and also record speech. For a twenty component activity set, comprising detailed static and dynamic activities at various intensities, Bao and Intille (111) returned an overall accuracy of 84%. The present study included a similar number of dynamic activities to those of Bao and Intille, but given that their results were achieved using five sensors, then the results of the present study can be considered to compare favourably.

The activity set chosen in the study by Parkka et al. (84) contains several activities that may be comparable with the present study. There were seven activities, five of which were dynamic activities as follows: walking, Nordic walking, rowing on a rowing machine, cycling on an exercise bike, and running. The present study also included walking, rowing on a rowing machine and cycling on an exercise bike in the activity set. Additionally, jogging in the present study and running in Parkka et al. (84) may be comparable, as jogging can be considered as running at a lower intensity – the intensity threshold between these two activities is arbitrary, and the level of running intensity is not stated in Parkka et al. (84). The study by Ermes et al. (117) follows on from Parkka et al. (84) and considered nine activities, seven of which were identical to the previous study, with the addition of cycling on a bicycle (as opposed to an exercise bike) and playing football.

As the present study contains the majority of the dynamic activities in Parkka et al. (84), has additional dynamic activities, and does not include any static activities that contribute to higher prediction accuracy, it is arguable that this activity set presents a greater challenge to classification, and the two additional activities included by Ermes et al. (117) do not add sufficient complexity to change this. Additionally, Parkka et al. (84) used the combined output from accelerometers at the chest and wrist, and Ermes et al. used combined hip and wrist output, which puts both these studies at an advantage compared to the single accelerometer site in the present study. Furthermore, Ermes et al. (117) utilised GPS to obtain speed. Having a measure of speed is a great advantage over having accelerometer data alone – clearly this feature would overcome the difficulty of discerning between static cycling and travelling on a bicycle, and would also aid in discriminating between walking and running. For a real-world application, use of GPS in conjunction with accelerometry would not be practical, as it would mean the increased burden of additional equipment, and would only be applicable to certain activities – in the context of a gym-based program of activity, no advantage would be gained.

The activity prediction accuracies reported in Parkka et al. (84) and Ermes et al. (117) were enhanced due to the bias towards the sit/stand activity, as discussed. In the present study, a maximum of thirty windows was set per activity for each subject. Consequently, for eight out of the ten activities there were close to 1,500 windows, and only stair climbing and descent, where the data collection was limited by subject fatigue, contained less (between 880 and 980 windows each). Stair walking, therefore, exercised less influence on the overall accuracy than the other activities. Nonetheless, overall accuracy and average measures of sensitivity and precision were within one percentage point of each other, which suggests that the bias introduced due to fewer windows of stair walking is not significant.

There are more dynamic activities and fewer sensors in the present study than in Parkka et al. (84) and Ermes et al. (117), and there is no bias towards high performing static activities in the accuracy calculation. Considering these factors, the 85% hip and 94% ankle accuracies achieved by the present study compare well with the 86% accuracy reported by Parkka et al. and the 89% accuracy reported by Ermes et al.

Study	No. Of Subjects	No. Of Sensors	No. Of Activities	Placement	Best Accuracy
Allen et al. (80)	6	1	4 + 4 transitions	Waist	91%
Baek et al. (110)	1	1	8	Waist	98%
Bao and Intille (111)	20	5	20	Wrist + arm + ankle + hip + thigh	84%
Bonomi et al. (85)	20	1	7	Lower back	93%
De Vries et al. (101)	49	2	9	Hip + Ankle	69%
Edgar et al. (89)	1	2	8 household* / 8 athletic*	Wrist + foot	90% / 93%
Ermes et al. (117)	12	2	9	Hip + wrist	89%
Foerster et al. (160)	24	4	9	Sternum + wrist + thigh + lower leg	67%
Gyllensten and Bonomi (139)	52 / 20	1	5	Lower back	95%
Huynh and Schiele (147)	2	1	6	Shoulder strap on backpack	N/A
Lee et al. (82)	12	1	9	Back	95%
Maurer et al. (167)	6	6*	6	wrist/belt/neckla ce/ trouser pocket/shirt pocket/bag	73% - 87%
Oudre et al. (180)	24	1	4	Shin	96%**
Parkka et al. (84)	16	2	7	Chest + wrist	86%
Pober et al. (140)	6	1	4	Hip	81%
Ravi et al. (144)	2	1	8	Pelvis	73%
The present study	50	2*	10	Hip / Ankle	85% / 94%

Table 9: Accuracies achieved by selected previous research. * considered separately **average sensitivity

Several previous studies exceeded 90% recognition accuracy (Table 9). However, the high accuracies reported in these studies were invariably due to factors such as the limited activity set being tested rather than the ability of the classification scheme. The activity set used by Allen et al. (80) contained only one dynamic activity (walking) along with three static activities (sitting, standing and lying). Similarly, Gyllensten and Bonomi (139) included only two dynamic activities (walking and cycling) and two static activities (combined sit/stand and lying). Although in both of these studies the transitions between activities were also considered, these activity sets present a significantly lesser challenge to classification than the present study. Bonomi et al. (85) chose to analyse seven activities, three of which were static in nature, and three were included in the present study, giving the present study the disadvantage of having an additional six dynamic activities to recognise. Similarly, the activity set considered by Oudre et al. (180) was, again, a simpler classification problem, as it contained only three “periodic” activities (walking, running and cycling), and some “non-periodic” and static activities that were classed together as “other”.

Of the nine activities considered by Lee et al. (82), five were variations of sitting, standing and lying, and the remaining four were ambulatory (walking, running, stair ascent and stair descent). Again, this activity set is not as varied as the present study with several static activities, thus making it a simpler classification problem which may be expected to yield higher accuracies. Additionally, it is likely that the approach by Lee et al. (82) would not be able to distinguish several different dynamic activities due to the simple hierarchical algorithm employed. The algorithm relies on finding distinctive thresholds between activities using the AC and DC components of the accelerometer signals, and it is probable that thresholds such as these will not be possible between many dynamic activities.

It is not certain that those studies reporting over 90% accuracy would still return such high accuracies if the number of dynamic activities in their activity sets was greater. Additionally, in the case of both Baek et al. (110) and Edgar et al. (89), an intra-subject classification scheme was employed which, as discussed, greatly improves recognition accuracy but does not generalise to a wider population.

3.4.1.3 The Effect of Sensor Placement on Classification Accuracy

The best overall accuracy achieved by the present study for the hip-mounted accelerometer was 85% and for ankle-mounted accelerometer was 94%. The results show that overall the ankle-mounted accelerometer performs better than the hip-mounted accelerometer for this particular activity set, and this is true for all feature sets (Table 3). This agrees with a study by Preece et al. (263) which considered eight dynamic activities and found consistently higher accuracies from an ankle-mounted accelerometer compared to thigh and hip sites. The results of the present study, however, are contrary to findings by Bao and Intille, who found that hip alone performed better than ankle alone. De Vries et al. (101) also obtained marginally better overall accuracy from the hip-mounted accelerometer (80% accuracy) over an ankle-mounted sensor (78% accuracy) for five activities. However, when the activity set was increased to nine activities their accuracies decreased to 60% for the hip and 64% for the ankle. This may suggest that the ankle performs better with a greater number of activities, as in the present study, but it is more likely that the classification scheme devised by De Vries et al. did not perform well enough to allow a valid comparison of placement sites – several activities, for both accelerometer sites, reported results of less than 10% accuracy.

In the present study, ambulatory activities (walking, stair ascent and descent) are much better distinguished by the ankle accelerometer. Recognition of rowing returned higher accuracy for the ankle, which is unexpected as the foot is secured whereas for the hip sensor there is a clear, repetitive anteroposterior oscillation of the torso that generates distinctive signals which are not observed in any other activity. Only cycling returned a higher sensitivity value for the hip-mounted accelerometer, which is again surprising as the ankle clearly displays a distinctly cyclic movement whereas the hip undergoes only a small amount of sway for the less vigorous participants. The explanation for these unexpected findings may be that the low intensity of the signal helps better distinguish the activities; if most activities other than cycling have relatively high intensity signals at the hip, then cycling is indicated because of the low intensity of the signal, and this may also be the case with the ankle signal for rowing.

For the ankle-mounted accelerometer, there was confusion between stepping and stair ascent, and vice versa (Table 6). This might be expected due to the similarity of the two activities; both involve the climbing of a step, and for both the subject is facing the step being climbed or descended (as opposed to stair descent where the subject faces away from the step being descended). Under free living it is possible that the confusion between these two activities could be reduced by setting the prior probabilities to the LDA algorithm to favour stair ascent over stepping, as stair usage is likely to occur more frequently than the stepping exercise.

Results for the hip-mounted accelerometer show that the classifier confuses ambulatory activities (walking, stair ascent and stair descent) for this body placement. Stepping is also confused with stair climbing, as in the case of the ankle-mounted accelerometer, and again this is likely to be due to the similarities between the two activities. Side-stretching is confused with cycling, which is likely to be because of a similar hip sway movement for both – although, notably, cycling was not often mistaken for side-stretching, which suggests that a better distinction between these two activities may be possible by applying the appropriate features. Also, the side-stretching activity was interpreted quite differently between subjects, with some performing relatively vigorous movements involving shifting feet positions and others merely moving their arms above their heads. It may be that some individual interpretations of side-stretching generated accelerometer output which shared more characteristics with cycling than the overall group data for side-stretching. Additionally, some participants performed the cycling activity particularly slowly, showing barely perceptible movement in the torso, which could result in the accelerometer signals generated by the hip more closely matching those generated by the less vigorous interpretations of side-stretching.

3.4.1.4 Limitations in Respect to Research Question 1

A problem experienced by some activity classification studies has been that high recognition accuracies obtained in the laboratory have not been retained when the classifier has been applied to data obtained under free-living conditions. Foerster et al. (160) found that an accuracy of 95.8% for laboratory collected data dropped to 66.7% for data collected outside the laboratory. More recently, Gyllensten and Bonomi found that recognition accuracy dropped from 95.9% in the laboratory to 75.9% under free-living, although this figure was obtained using activity diaries to label accelerometer output, and self-reporting has been shown to be unreliable (50-52) which puts into question these results. Bao and Intille attempted to collect semi-naturalistic data by creating an "obstacle course" containing "goals" which were not directly related to the collection of activity data to encourage participants to act more naturalistically. This approach may or may not be effective, but ultimately is still a laboratory-based data collection protocol. It is not clear how the classification scheme defined in the present study would perform under free-living conditions. However, the gym-based activities of treadmill walking, rowing, using the cross-trainer, and static cycling are arguably less sensitive to laboratory influences because, to a certain extent, the machinery involved enforces adherence to particular motions. Nevertheless, the classification scheme would need to be tested under free-living conditions to ascertain to what extent classification performance was affected.

Another important limitation the current study has, in terms of its direct application to free-living, is that static activities have not been accounted for by the classification scheme. Static activities were omitted as preliminary pilot testing had shown near perfect accuracies for sitting, standing and lying using a hip-mounted accelerometer, and therefore inclusion of these would boost overall accuracy without adding any insight into the detection of the dynamic aerobic/gym activities that are of interest; the focus of the present study was to classify an activity set of less easily distinguished dynamic activities. However, it is possible that some of the aerobic/gym-based activities that were performed at very low intensities might not generate a signal with sufficient signal magnitude area to distinguish it from a static activity. For example, certain participants' slow cycling might be mistaken for sitting, and for some participants side-stretching might be seen as standing. Static activities would need to be included in the analysis to verify whether this was the case. Additionally, it is likely to be more difficult to find a threshold between standing and sitting when considering the ankle-mounted accelerometer. Some studies have experienced problems such as this, and one solution has been to consider sitting and standing as a single activity (84, 117, 139).

A fundamental issue with supervised machine-learning algorithms is that activities that have not been trained for are necessarily misclassified. For example, if the classifier for the present study was applied to data obtained from an individual performing skiing, then this would be incorrectly classified as one of the ten activities in the original training dataset. It is not practicable to train a classifier for all possible activities that might occur under free-living conditions. However, the focus of the present study has been aerobic/gym-based exercises and common free-living activities that might be undertaken by an obese person wishing to lose weight. Clearly, obese individuals are more limited in terms of which physical activities they may perform, so the pool of possible activities is somewhat diminished. The activity set chosen in the present study is still some way off being comprehensive – there are many gym-based activities that could be part of an obesity management programme that do not feature here – but the activity set chosen demonstrates that it is possible to recognise a variety of dynamic activities.

Another limitation of the present study is that activities that are predominantly defined by arm movements may not easily be identified. The present study focussed on locomotive and aerobic activities that could be part of an exercise programme, and the results showed that these types of activity can be recognised both from hip- and ankle-mounted accelerometer output. However, certain exercises eligible for a weight-loss programme primarily use arm movements. Weight-training, for example, can involve strenuous arm movement while the torso and legs remain relatively still. If exercises and activities such as these were to be undertaken, then it is not clear whether the hip- and ankle-mounted sensors would return adequate information to distinguish them.

Several studies have looked at transitions between activities, whereas the present study has not used any data representing transitions and has instead isolated instances of each activity so that each window of activity data contains only a single activity. This means that if the present classification scheme were applied to real-world data, then there would be misclassifications where transitions occurred between activities. As the classifier uses two second windows it means that such misclassifications will be minimised – that is, if a window were ten seconds long and contained a transition which lasted only one second, then this may misclassify the whole ten seconds of accelerometer data as opposed to two seconds' worth of data in the present scheme. However, to fully account for transitions, the present classification scheme would need some modification.

A self-imposed limitation of the study is that a single accelerometer site should be considered to minimise burden on the wearer. If this restriction was not in place, then it is likely that classifier accuracy would improve by using the combined output from the two accelerometer sites, as has been the case in previous research (178).

Although some walking was measured outside the laboratory on a path which was not uniformly level, it is not certain how the classification algorithm would classify walking at steeper gradients and on different surfaces. As walking energy expenditure increases with gradient (when ascending a slope) it may be useful to identify not only that walking was taking place, but also that the walking surface was sloped. There has been previous research which has estimated walking gradients from accelerometer output with some success (264). Future research may consider integrating the classification process with a second separate algorithm to estimate the angle of incline which would be applied when instances of walking were identified.

3.4.1.4.1 Application of the Classifier to Free-Living Data

The present study has demonstrated that several dynamic physical activities and exercises may be accurately distinguished by a classification scheme applied to hip- and ankle-mounted accelerometer data collected in the laboratory. However, because no static activities are considered by the classification scheme, it is currently inadequate for real-world applications. Also, as mentioned above, it is common that high performing classification algorithms in the laboratory may not perform as well under free-living conditions, and this is compounded by the inability of a supervised-learning classifier, such as that used in the present study, to accurately classify activities that lie beyond the training set.

To demonstrate these limitations, the classification scheme was applied to data made available from a previous pilot study which contained free-living hip-mounted accelerometer data and corresponding activity diaries maintained by the participants (see appendix B). The activity diaries were not sufficiently reliable to allow the The activity diaries did not report gym-based exercises such as rowing and cross-trainer, therefore these activities would not be expected in the results. Also, it would be expected that the most common dynamic activity would be that of walking.

To gain a more accurate picture of how the classifier would perform under free-living conditions, an initial step was added to the classification scheme which ascertained whether a window of activity was static or dynamic, based on a signal magnitude threshold (see Appendix C), before applying the dynamic classifier. This threshold was set as a value based on the distribution of SMA values aimed at ensuring dynamic activity was identified, rather than utilising the method described in appendix C. This was because the main study had not been designed to obtain an accurate value for the threshold.

It was clear from the results that there were anomalies in activity classification. The first issue was that the classifier often reported a disproportionate amount of stair walking in comparison to walking. This suggests that walking was being misclassified as stair walking. The ratio between walking upstairs and walking downstairs was also not as expected – the time spent in each of these two modes might be expected to be the same, or to weigh in favour of stair descent (as the more strenuous activity of stair climbing might be purposely avoided through use of elevator or escalator). However, according to the classifier results, the time spent in stair ascent in many cases far exceeded that of stair descent, which further suggests that level walking and stair ascent were not adequately distinguished by the classifier.

Cross-trainer and jogging were not reported by the classifier for any periods longer than four seconds in one day. This is in line with the participant diaries and is an encouraging result, particularly as using a cross-trainer is a cyclic activity which potentially shares a similar periodicity and intensity with walking. Rowing was reported by the classifier for all subjects, and for one dataset this activity totalled more than three minutes. It is unlikely that any of the participants performed rowing and, though three minutes is of small duration within a twenty-four hour period, this level of error may prove misleading when assessing an individual's daily activity patterns within the context of an exercised-based weight loss programme. Cycling was erroneously reported for all subjects as being up to 5.1% of the total dynamic activity time. For the participant who recorded thirty minutes of cycling in their activity diary, the classification algorithm underestimated this by twenty minutes. This may be due to cycling being misclassified as another dynamic activity (possibly side-stepping given the high incidence of this activity for this subject), but may also be due to the chosen threshold between static activity and dynamic activity being set two high to regard more leisurely cycling as dynamic activity. Additionally, the cycling dataset used to train the classifier was obtained from a static cycling machine as opposed to a moving bicycle, which may also have affected classification.

The stepping activity is higher than expected and is likely to be greatly due to its similarity to stair walking, as observed in the main study, but also appears to be resulting from a misclassification of walking. As discussed in section 2.1, many dynamic real-world activities may defy classification – for example, shuffling between work surfaces when cooking, or moving between products on a shelf in a supermarket. It may be that some of these share characteristics with the aerobic activities of side-stepping and side-stretching, which may explain the incidence of these in the classification results where they were unreported in the activity diaries.

The results of this initial test of the classifier against free-living data indicate that the study protocol in this chapter is not adequate to derive a robust classifier for use in real-world applications. This may be due to differences between how activities are performed under lab conditions versus free-living, and may also be as a result of issues such as terrain, but is also due to the limited number of activities that are represented in the classifier training set and the exclusion of static activities. In order to develop and test a reliable classifier suitable for free-living applications, one approach might be to train a classifier using a large quantity of accurately labelled accelerometer data that are representative of a multitude of real-world activities performed by a mixed BMI group. This approach would provide a suitably varied dataset for use as training data to the classifier, and for use in validating the algorithm. However, such a large-scale undertaking presents problems for the researcher as, in the absence of an objective means of recording human physical activity, it would likely require the participants to record a diary of their own activities, which is inherently unreliable, as discussed previously.

Other approaches to improving classification accuracy under free-living may be considered. To avoid the disproportionate incidence of less frequently performed activities (such as rowing) in the classifier output, it may be necessary to tailor algorithms to be weighted towards everyday activities (such as walking) rather than presuming that each activity has the same likelihood to occur. For example, in the case of the LDA algorithm used in the present study, it is possible to change prior probabilities to favour various activities over others. To account for activities that do not appear in the training set of activities, unsupervised-learning techniques may be required, such as clustering where activities are grouped together according to features extracted from their corresponding accelerometer signals, but the type of activity each represents is not necessarily known.

There are also other practical considerations which may impact classification accuracy when using accelerometry under free-living conditions. Accelerometers must be calibrated to account for local gravity (265), as gravity is not uniform in magnitude at each location on the Earth, otherwise classification accuracy may be impaired due to the quality of the accelerometer signal. Classifier algorithms may also need to be modified to account for this. Another consideration is that non-wear time should be distinguished from sedentary activity, as during this time the amount of physical activity being performed is unknown and should, therefore, be reported as such. This consideration should be accounted for when developing a classifier for static activity. Some activities may introduce noise to the accelerometer signal, such as riding a bus, which may be interpreted by a classification algorithm as dynamic physical activity. The classifier needs to account for activities such as these so that errors are not made.

3.4.1.5 Summary RQ1

Can a set of aerobic exercises and free-living activities be identified from data collected by a single accelerometer mounted at the hip or at the ankle?

The present study has demonstrated that an activity set comprising aerobic exercises and free-living activities can be classified using data from a single hip- or ankle-mounted accelerometer. Furthermore, both accelerometer sites have achieved a degree of accuracy that is comparable with or exceeds previous research. However, in its current form the classification scheme employed in the present study is not adequate for application to free-living.

3.4.2 Research Question 2

Does activity classification accuracy differ between obese and normal BMI groups?

The accuracies returned by the classifier for each of the ten feature sets were divided into four subject groups as follows: the entire subject group, obese subjects, overweight subjects, and normal BMI subjects. These are shown in Table 3. A comparison between classifier accuracies according to BMI group is made below.

3.4.2.1 Comparison of Classification Accuracies between BMI Groups

Ankle and hip classification accuracies, for each BMI group and feature set, are summarised in Table 3. First, the normal BMI group is considered in comparison to the obese group, as these groups have the greatest difference in BMI. For the hip-mounted accelerometer the best performing feature set (F10) returned 84.5% and 81.9% accuracies for the normal BMI and obese groups respectively, which is a difference of 2.6% accuracy in favour of the normal group. F8 was the overall best performing feature set for the ankle placement returning accuracies of 95.7% and 91.4% for the obese and normal BMI group respectively. This was a difference of 4.3% in favour of the obese group, and this was also the largest difference shown between normal BMI and obese groups for any of the feature sets. All but one of the feature sets for the hip-mounted accelerometer returned marginally higher accuracies for the normal BMI group. Conversely, all but one of the ankle accuracies were slightly higher for the obese group. However, the differences in accuracy are not great enough to conclude that hip-mounted accelerometers perform better for normal BMI groups and ankle-mounted accelerometers perform better for obese groups. The average absolute difference in accuracy between the two BMI groups across feature sets was 2.6% and 2.3% for hip and ankle sites respectively, and for the best performing features the difference in accuracy was less than 5% for both placement sites. More than 80% prediction accuracy was achieved for the hip-mounted accelerometer and over and 90% accuracy was achieved for the ankle-mounted accelerometer for both obese and normal BMI groups. These results suggest that a single classification scheme may be effectively applied across BMI groups for both ankle and hip placement sites.

There has been very little research which has considered how activity classification may be affected by BMI. Zhang et al. (181) compared classification accuracy between two BMI groups (<25 kg/m^2 and >=25 kg/m^2) and concluded that although there was a statistically significant effect of BMI on detection, BMI did not greatly affect prediction accuracy for five activities, as this averaged at over 99% for both BMI groups. The high accuracies returned for the ankle data for both BMI groups in the present study apparently corroborate the conclusion made by Zhang et al., but the comparison is not entirely valid due to the different number of accelerometers and activities under consideration – had there been a greater range of activities in Zhang et al., then a more pronounced affect of BMI may or may not have been observed. Oudre et al. (180) investigated classification accuracy for three BMI groups – normal (n=12), overweight (n=8) and obese (n=4) – and, having returned accuracies between

85% and 100% for three activities for each BMI group, concluded that their classification approach was justified across BMI groups. The results of the present study agree with the conclusion made by Oudre et al. (180), but again the comparison may be weak because of the few activities in their study, and also the low number of obese participants (n=4 compared to n=19 in this study).

In the present study it is notable that for the hip-mounted accelerometer the overweight group scores higher prediction accuracy over the entire group, the normal BMI group, and the obese group for all but two of the thirty accuracy comparisons for the ten feature sets, and in some cases this difference is greater than 5% in favour of the overweight group result. Similarly, the overweight group returned higher accuracies for all but six out of these thirty possible comparisons for the ankle-mounted sensor. These results match the findings in the study by Oudre et al. (180) where the overweight group also returned the highest accuracies. However, as there were only eight overweight participants in Oudre et al. and four obese subjects, it is not clear that the same findings would be repeated using a larger participant group. In the present study the small improvements in accuracy for the overweight group over the other groups may be natural fluctuations due to the difference in the BMI group sizes (n=17, n=19 and n=14 for normal BMI, obese and overweight groups respectively) rather than because of intrinsic differences between groups.

3.4.2.2 Limitations in Respect to Research Question 2

A comparison between activity recognition accuracies for the obese and normal BMI groups has been made for the activity set defined in 3.2.2.2, and the results suggest that accuracy is not greatly affected by BMI. These results were obtained when the classifier was trained using a mixed BMI group. It is not clear, however, whether improved results may be achieved by using a BMI-specific classification scheme. This would involve training a classifier using data from a single BMI group for application to individuals of the same BMI range; that is, an obese individual would be tested by the obese-specific classifier, and a separate classifier would apply to a normal BMI individual.

3.4.2.3 Summary RQ 2

Does activity classification accuracy differ between obese and normal BMI groups?

As far as the author is aware, the present study is the only one which compares classifier accuracy between obese and normal BMI groups that has sufficient numbers of obese subjects and activities. Overall, the results of the study show that accelerometer data collected from both normal BMI and obese subjects can be identified with a similar degree of accuracy when trained using a mixed BMI group. This suggests that a classification scheme such as the one employed in the present study may be applied to both obese and non-obese subject groups participating in aerobic/gym-based and free-living activities.

3.4.3 Research Question 3

Do the same accelerometer features apply to obese and normal BMI groups, or do they require different accelerometer features to characterise their physical activities?

To answer the research question, the rankings of the feature sets were compared between obese and normal BMI groups. If the feature sets are equally effective for obese groups as normal BMI groups, then the rankings of feature set accuracies should be similar for each BMI group. To determine how similar the ranks were between groups, the Kendall's Tau rank correlation coefficient was calculated.

3.4.3.1 Comparison of Feature Sets

From Table 4 it can be seen that the ranks for feature set accuracy remain largely the same within each test dataset, for both hip- and ankle-mounted accelerometer, where the classifier has been trained using the entire subject dataset. Kendall's Tau coefficients in Table 7 and Table 8 show high correlations to corroborate this for each subset of test data (the entire group dataset, and both normal and obese groups in isolation). The results obtained through this approach suggest that there is little difference between the types of feature which perform best against particular BMI groups, when the classifier is trained using mixed BMI data.

The feature set F1 comprising the mean and standard showed the lowest accuracies for both accelerometer placement sites. This was expected as it was the simplest feature set. Yet, the accuracy figures returned by this feature set were much higher than expected with 63% and 82.1% being returned from the hip and ankle respectively. The 82.1% accuracy at the ankle seemed abnormally high to the extent that it suggested an error may be occurring in the classification scheme. To check the algorithm, the classifier was tested with various combinations of data and input parameters. For example, the classifier was given spuriously labelled data which meant that the expected results should show around 10% accuracy for each of the ten activities based on probability. The results of the tests were as expected which suggested the classifier was functioning correctly. The high accuracies may be due to the effectiveness of the LDA classification algorithm when applied to this particular activity set, or several other parameters of the study, such as the window size, which in combination result in particularly high accuracies.

The Bao and Intille (111) feature set (F2) did not perform much better than the mean and standard deviation (F1), but this may be because in the present study there are only three accelerometer axes from which to derive features, giving a total of only twelve features for the F2 feature set, whereas in the study by Bao and Intille itself (111) there were ten accelerometer axes and a total feature set of thirty-five elements.

The DCT feature set (F6) consistently performed better than the FFT feature set (F5) at the hip but returned similarly high accuracies at the ankle. The Baek et al. feature set (F3) performed averagely but the F10 custom feature set, which incorporated the F3 features in conjunction with five frequency components and additional statistical features, was the best performing feature set for the hip-mounted accelerometer and the second best for the ankle-mounted accelerometer. The best performing feature set for the ankle placement was the F8 custom feature set. The custom feature set F7 was consistently ranked 3rd or 4th across all combination of accelerometer placements, training datasets and test groups. The Huynh and Schiele (147) feature set (F4) performed averagely at the hip but performed better at the ankle.

The feature sets returning the lowest accuracies tended to have fewer features. F1, F2 and F3 are three of the four poorest performing feature sets for both placement sites and these have six, twelve and fifteen features respectively. The fourth poorest performing feature set (F9) had twenty-four features, but these were not well chosen for maximum classification – the cepstrum frequency components are not as effective as the FFT and DCT, and there were only a few additional statistical features. The DCT is an effective feature set, comprising seventy-five features. The best performing feature sets F10 and F8 incorporated both frequency and statistical features and numbered forty-six and thirty-nine components respectively. The DCT feature set (F6) had seventy-five features and rivalled the best accuracies without using any time-domain features.

The success of the feature sets which included several frequency-based features may be due to the cyclic nature of the activities in activity set. Walking, jogging, stair climbing and descent all have cyclic patterns, and for the gym-based activities these patterns may be even more pronounced. The aerobic exercises, again, display regular repetitive patterns. Frequency domain features are, therefore, ideally suited to characterising each activity. Where there is similarity in the cyclic patterns between activities – walking and stair climbing, for example – the additional time-domain features may help make the distinction, or in the case of the FFT and DCT feature sets, the number of frequency components may represent each activity more finely.

3.4.3.2 Limitations in Respect to Research Question 3

The custom feature sets selected in the present study were not chosen using any feature selection or feature reduction techniques aimed at maximising classification accuracy or optimising computational cost. This is because the feature sets were not all designed to return the best accuracies, as this was not the sole aim of the research. Instead they were intended to test how differently each performed between BMI groups. Feature set F9 was designed to be different to the other feature sets and was, in fact, expected to return lower accuracies. Feature sets F7, F8 and F10 shared several elements and performed similarly – each returning high accuracies – but the relatively small differences between them helped test if rankings were different between BMI groups (the rankings of a poor feature set and a good feature set may be the same regardless of which BMI group they are applied to, whereas two subtly different feature sets may switch ranking depending on the attributes of the group that produced the accelerometer data – although it is also possible that a feature set that performs poorly for one group may perform well in another).

In the present study, features were tested for their ability to distinguish physical activities regardless of the physical attributes of the individuals performing them. Those feature sets which returned the best classification accuracies captured common accelerometer signal characteristics across individuals which did not vary greatly according to BMI. However, it is possible that the inclusion of certain BMI-related measurements may improve classification accuracy. For example, if higher BMI values correspond with greater mediolateral sway in individuals (as discussed in 2.2.1) then it is possible that activities involving mediolateral movement may be more easily distinguished if BMI values were to be included in the feature set, or used to normalise the mediolateral accelerometer signal. Future work may wish to investigate how a classification scheme may be improved by the inclusion of anthropometric measurements within the feature set when applied to a mixed BMI subject group.

3.4.3.3 Summary RQ3

The research question sought to ascertain whether the same features sets may effectively applied to obese and normal BMI groups, or whether different types of feature gave different results depending on the BMI of the subject group. Ten feature sets were tested and these were found to rank similarly for different BMI groups. It was thus concluded that the same features may be applied equally to both BMI groups.

3.4.4 Chapter Summary

Activity monitoring research has shown that human physical activity modes may be recognised from accelerometer signals (84, 144, 146, 162, 266). This has been demonstrated for various activities, but no previous study has tested an activity set designed to incorporate exercises specifically aimed at combating obesity. Findings from the present study imply that it is possible to accurately determine activities such as these. Furthermore, the study limited the number of sensors to a single sensor. This was to test how well a single sensor might perform in the field where it is desirable to minimise the burden imposed upon the wearer. Classification accuracies returned for both hip- and ankle-mounted accelerometers (85% and 94% overall accuracy respectively) were comparable with previous research, and in many cases exceeded the accuracies returned by studies which had used multiple sensors and less challenging activity sets. Also, previous research has not investigated how effectively activity classification can be applied to obese individuals compared to those who are not obese. The present study found that classification accuracy was comparable between obese and normal BMI groups with both returning high accuracies. A number of feature sets were compared to ascertain whether the same features are equally applicable to obese and normal BMI groups. It was found that these feature sets were similarly applicable across BMI groups.

The implication of these results is that a classification scheme may be effectively applied to subjects across a range of BMIs without needing to make alterations to the feature set according to the BMI of the target population. The classification scheme has been shown to be effective using a single activity monitor, as opposed to several sensors, and thus may be applied in the field with a lesser burden to the individual and a lower financial cost of implementation. The ankle was found to be the best placement site, though the hip also returned high accuracy and may be preferable for practical implementations in order to minimise burden and improve compliance. The present study has shown that, in principle, accelerometry may be used to identify bouts of the type of activities that an individual might perform while taking part in an obesity management programme. The motivation for this research has been to work towards an activity monitoring system that would allow a clinician to evaluate the effectiveness of an activity programme in terms of weight loss, and also to determine participants' adherence to such a programme. A system incorporating accelerometry to identify physical activity may also be used to provide motivational feedback to those taking part in an intervention aimed at weight loss; participants may be given regular access to their activity profiles, and this might provide motivation for further exercise. Additionally, this technology may be applied to epidemiological research aimed at determining which particular patterns of activity might influence obesity. However, the questions that have been answered by this research must be regarded only as a step towards such a system, and further research is required to improve classification accuracy under free-living conditions before it is suitable for use in the field.

4 Walking Speed Estimation Using Accelerometry

The ability to objectively measure physical activity would help evaluate weight loss programmes and provide motivational feedback to those trying to lose weight. It would also be of use in epidemiological research involving individuals' activity patterns. For implementation in large scale studies, a walking speed measurement tool should function effectively irrespective of the BMI of the participants and without requiring individual calibration. When considering walking in the context of weight management, the speed estimation algorithm should work for both overground and treadmill walking, the latter being an exercise which may be part of a weight loss programme. Additionally, when applied to a free-living environment, a body-worn sensor needs to be unobtrusive and should not interfere with the natural behaviour of the wearer. A single accelerometer worn on a belt above the hip would meet this criterion.

The research question was asked: can a hip-mounted accelerometer be used to accurately estimate walking speed for an obese group?

In order to answer the research question, the following objectives were set:

- Objective 1: to produce a speed estimation model that is accurate across BMI groups and walking modes.
- Objective 2: to investigate how walking speed estimation accuracy differs between obese and normal BMI groups.
- Objective 3: to explore how walking speed estimation accuracy is affected by walking mode.

4.1 Research Design

To meet the objectives of the study, and to answer the research question, hip-mounted accelerometer data was collected for twenty-two subjects performing treadmill walking at five speeds, and overground walking at four speeds. Several accelerometer features were chosen to characterise elements of walking intensity and gait characteristics. An algorithm was applied to these features in order to identify which of these may best be used as parameters in linear models to accurately estimate walking speed across BMI groups. A number of models were identified, and each was evaluated separately using leave-one-out cross-validation. The results were compared between the obese and normal BMI groups for both overground and treadmill walking.

4.2 Methods

4.2.1 Recruitment and Subject Statistics

The recruitment procedure was the same as that described in section 3.2.1. Twenty-two subjects completed the study (Table 10). There were eleven subjects with BMIs in the normal range (under 25kg/m^2) and a further eleven with BMIs in the obese range (greater than 30kg/m^2). The figure below (Figure 11) shows the relationship between height and leg length for the participants of the study.

	All Subjects (n=22)	Male (n=12)	Female (n=10)
Age (years)	31.5 (7.60)	32.5 (8.3)	30.4 (6.9)
Height (cm)	169.4 (8.8)	173.7 (9.2)	164.3 (4.6)
Weight (kg)	83.2 (21.3)	85.6 (18.9)	80.3 (24.5)
BMI (kg/m^2)	29.0 (7.2)	28.5 (6.2)	29.6 (8.6)

Table 10: Subject statistics. Figures are mean (standard deviation).

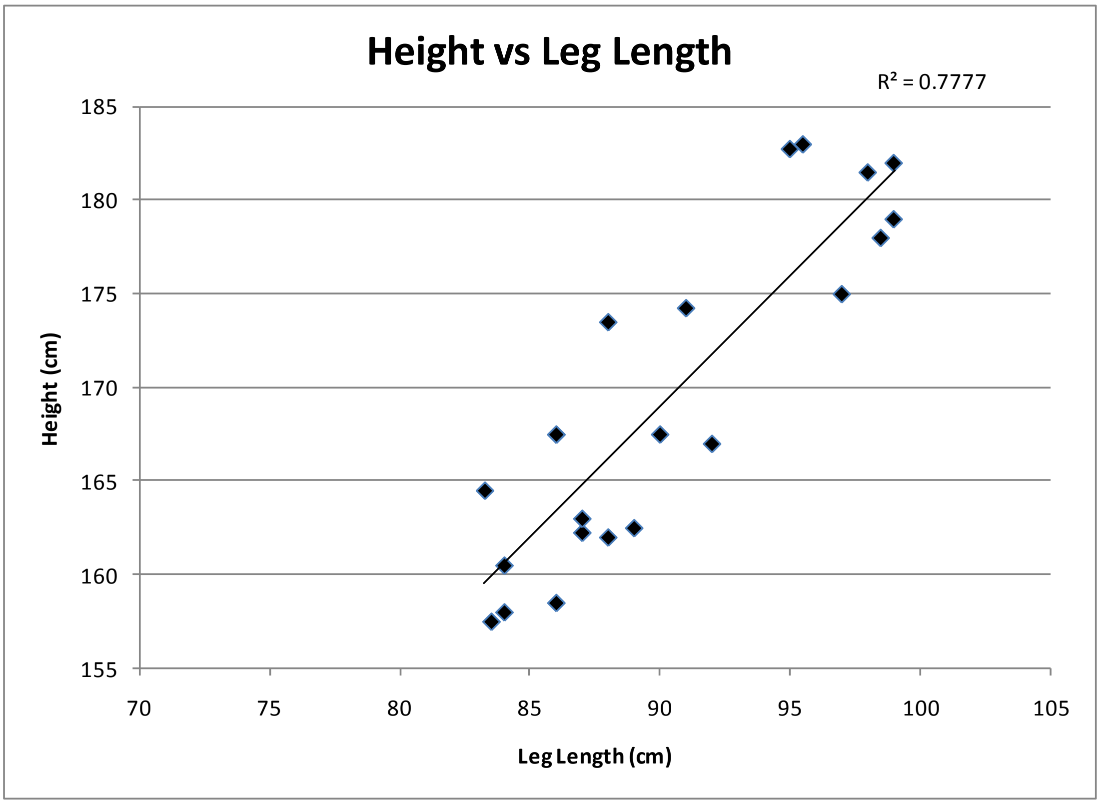

Figure 11: Height versus leg length for the participants of the study. The correlation was high (R^2 = 0.7777).

The majority of the subjects who took part in this study also participated in the studies described in chapters 3 and 5. These subjects include all eleven in the obese BMI range and five in the normal BMI range. Data collection for treadmill walking was taken simultaneously for all three studies. Overground walking data was measured part way through the programme of activities described in chapter 3. An additional six subjects in the normal BMI range performed treadmill and overground walking specifically for this study.

4.2.2 Data Collection

Data was collected at the Human Performance Laboratory, at the University of Salford.

4.2.2.1 Participant Anthropometric Measurements

Participants were weighed and their height was measured using a stadiometer. From these, BMI was calculated. Biomechanical models of walking rely on leg measurements. For this reason, the length of the participant's right leg was measured from the top of the greater trochanter to the floor. Shoes were worn for these measurements as the additional height they confer contributes to the pendulum length when considering walking as an inverted pendulum (202-203).

4.2.2.2 Activity Monitor

A single Actigraph GT3X+ monitor was affixed to the participant's right hip above the iliac crest (Figure 12). The device was set to sample at 50Hz in order to capture movements of up to 25Hz according to the Nyquist Sampling Theorem as discussed in subsection 3.2.2.1. This is sufficient for movement at the hip which is unlikely to produce acceleration signals approaching this frequency.

Figure 12: Hip-mounted Actigraph GT3X+ accelerometer

An additional Actigraph GT3X+ was affixed to the participant's right ankle for use in the study described in chapter 3, but this data was not used in the analysis for the present study. As discussed in section 2.1.4, the hip site presents a balance between burden to the participant, when considering future real-life applications, and the range of aspects of PA that previous research has used waist/hip-mounted accelerometers to measure, such as energy expenditure and activity type; one of the overall aims of the thesis has been to consider measuring multiple aspects of PA from a single accelerometer site. The hip also has close proximity to the centre of mass (CoM), which means it may be used to derive walking speed from CoM displacement, as in Zijlstra and Hof (90). The ankle site does not have the same advantages as the hip in these two respects, and although it may be possible to apply alternative approaches to estimating walking speed using the ankle site, the aims of the thesis are more easily met by the hip-mounted accelerometer.

4.2.2.3 Walking Protocol

The walking phase of the protocol was divided into two parts: treadmill walking and overground walking.

Overground walking took place on level ground on a twenty metre length track in the Human Performance Laboratory at the University of Salford. Optical timing gates were used to measure the overall time taken to walk twenty metres, and walking speed was later calculated from these. Participants were asked to walk at four speeds with the following descriptions: “slow”, “normal”, “brisk”, and “fast”. Differing interpretations of these speeds by the participants, influenced by the physical differences between them and their naturally preferred speeds, allowed a variation of speeds to be measured. A prescribed set of speeds was not desirable as this would not test the ability of the algorithm to estimate a range of speeds with precision; the estimation algorithm would merely need to be able to distinguish between four specific speeds, which may lead to artificially high estimation accuracies, and would not be as applicable to a free-living environment. Participants were asked to pause after each walking speed test was completed. This would show in the accelerometer record as a period of low activity, and, when the accelerometer signal was plotted, would make each walking test visually distinguishable from the participant returning to the start position.

Treadmill walking was performed using the Ergo ELG55 treadmill (WOODWAY GmbH of Weil am Rheine, Germany). Five speeds were selected and were performed for at least one minute to ensure sufficient data was collected. The treadmill speeds varied according to the participant and were based on their preferences and abilities. Participants were consulted as the testing took place, and speeds were adjusted accordingly. As far as possible, speed was increased in even increments in order to obtain a representative sample of speeds across the range. Participants were not permitted to use the treadmill hand rail while walking.

As mentioned previously, the same treadmill walking data were used for both the present study and the studies described in chapters 3 and 5. For the purposes of the study in chapter 5, the first four speeds were performed for five minutes each (see chapter 5) and a breath analysis mask was worn. Walking at the fifth, and fastest, speed was performed for one minute without the mask.

Five speeds were collected for treadmill walking, but only four were collected for overground walking. This was because a range of specific speeds can be easily imposed when using the treadmill, whereas individuals are not able to set their own speed levels so precisely when walking freely. It is difficult for the individual to grade their walking speed beyond four descriptive guidelines.

4.2.2.4 Annotations of Walking Times

Start times were recorded for each of the four over ground walking trials so that they could later be located in the accelerometer dataset. As each participant activated the timing gate placed at the zero metre mark, a wristwatch was used to record the start time of the trial. The elapsed time in seconds, as returned by the handset connected to the timing gates, was written down as the participant passed the timing gates at the 20m mark. For the five treadmill walking trials, the start and end times were handwritten, and the corresponding treadmill speed was recorded. When the Actigraph GT3X+ was initially synchronised with the computer, the offset between the wristwatch time and the computer time was noted so that hand written times could be later synchronised with accelerometer output.

4.2.3 Preliminary Processing

4.2.3.1 Data Labelling and Extraction

The proprietary software provided by the accelerometer manufacturer was used to download the acceleration data for each set of walking trials. The three-dimensional accelerometer data was converted into MATLAB format as described in subsection 3.2.3.1.

The hand written annotations for overground walking were transferred to a spreadsheet for each participant before being converted to MATLAB format. Timestamps were adjusted according to the previously noted offset between wristwatch and computer in order to obtain the absolute time. Accelerometer data was plotted and start and end times for each walking trial were indicated against the plot to allow visual inspection of how well they align. Adjustments to the start and end time stamps for each walking trial were made where necessary. From these timestamps the walking speeds were calculated for each trial as the total distance (20m) divided by time. Additionally, the times recorded in the spreadsheet were used to extract the accelerometer data corresponding to each walking trial. Similarly, the start and end times for treadmill walking were input into a spreadsheet and read into MATLAB. From these the corresponding accelerometer data were extracted. Treadmill walking speeds that were noted at the time of testing were transferred to a separate spreadsheet and subsequently read into MATLAB.

4.2.3.2 Computation of the Input Parameters for the Speed Estimation Models

The main objective of the study was to produce a walking speed estimation model which is appropriate for both obese and non-obese groups and applicable to both overground and treadmill walking. The speed estimation model depends on suitable input parameters in order to achieve this objective. A set of candidate parameters were selected according to two main categories. The first category of parameters comprised accelerometer features that correspond with the biomechanical walking model or gait cycle. The second category included statistical features of the accelerometer signal that may correlate with walking speed. The parameters are listed in Table 11 and are described in detail below.

Although the majority of the candidate parameters are accelerometer features, some are derived from a combination of accelerometer output and leg length, and others return gait cycle parameters from the signals based on a biomechanical model. For this reason the term "parameter" is favoured over the term "feature" in the context of this study.

Parameters MeanX, MeanY, MeanZ. These represent the mean of the vertical, anteroposterior and mediolateral signals respectively. Mean was chosen as it is a fundamental statistic and was also employed in the speed estimation model by Bonomi et al. (85). However, it was expected that these parameters would not correlate well with speed, as at constant walking speed there is no net acceleration in any direction.

Parameters STDx,STDy, STDz. These represent the standard deviation of the three accelerometer signals. Again, the standard deviation is a fundamental statistic and describes how much the signal varies from the mean. This was also employed in the speed estimation model by Bonomi et al. (85). It was expected that as speed increases there would be higher standard deviation values of the acceleration signals due to the more vigorous movements involved.

Type	Parameter	Description
Statistical	MeanX	Mean of the vertical accelerometer signal.
	MeanY	Mean of the anteroposterior accelerometer signal.
	MeanZ	Mean of the mediolateral accelerometer signal.
	CPSx	Sum of the absolute magnitude of the high-pass filtered vertical signal, divided by seconds.
	CPSy	Sum of the absolute magnitude of the high-pass filtered anteroposterior signal, divided by seconds.
	CPSz	Sum of the absolute magnitude of the high-pass filtered mediolateral signal, divided by seconds.
	RangeX	Range of the vertical acceleration values.
	RangeY	Range of the anteroposterior acceleration values.
	RangeZ	Range of the mediolateral acceleration values.
	RMSx	Root mean square of the vertical accelerometer signal.
	RMSy	Root mean square of the anteroposterior accelerometer signal.
	RMSz	Root mean square of the mediolateral accelerometer signal.
	STDx	Standard deviation of the vertical accelerometer signal.
	STDy	Standard deviation of the anteroposterior accelerometer signal.
	STDz	Standard deviation of the mediolateral accelerometer signal.
Biomechanical	MST	Mean step time.
	MSTn	Mean step time, normalised by leg length.
	MVD	Mean vertical displacement.
	MVDn	Mean vertical displacement, normalised by leg length.
	ZH0	Parameter based on the pendulum model used by Zijlstra and Hof.
	ZH1	A simplication of the Zijlstra and Hof formula.
	ZH2	MVDn divided by mean step time.

Table 11: Candidate parameters for the walking speed estimation model.

Parameters RangeX, RangeY, Range Z. These represent the range of acceleration for the three accelerometer signals. Faster speeds might be expected to produce a higher range between maximum and minimum acceleration values, particularly in the case of vertical acceleration due to the vertical movement of the centre of mass according to the pendulum model. Acceleration data from one of the subjects contained a large spike which was not consistent with the rest of the signal for this particular speed. A measure such as range is sensitive to this kind of noise; a single spike significantly changes the range value, whereas such a spike would not significantly affect a measure such as the median of the signal. When using measures that are sensitive in this way, it may be better to implement an initial smoothing function. For this reason the present study employed a median filter to the signal before calculating the range.

Parameters RMSx, RMSy, RMSz. These are the root mean square (RMS) values of the signal and are calculated by according to the formula below:

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {x_i}^2}$$

In this equation, x values represent acceleration samples, and n is the number of samples. The RMS of the signal was chosen as a candidate feature as it had been previously shown to correlate with speed by Shultz et al. (200).

Parameter MST. This parameter is the mean step time based on the average number of samples between estimated heel strikes. The value returned is the number of seconds between steps. A heel strike detection algorithm was employed which is described in subsection 4.2.3.3. **Parameter MSTn** was the mean step time normalised by leg length.

Parameter CPSx. This parameter bears similarity with accelerometer "counts" (described in 2.5.1) used by Barnett et al. in their speed estimation model (207) except instead of returning total counts over a predefined fixed epoch, the results are in counts per second, and it is, therefore, independent of epoch. The parameter is calculated by first high-pass filtering the vertical acceleration signal, then finding the sum of the absolute magnitude of the filtered signal. This is then divided by the number of seconds the signal represents so that the result may be compared for signals of different sample lengths. **Parameters CPSy, and CPSz** are calculated in the same manner as CPSx for the anteroposterior and mediolateral accelerometer signals respectively.

The remaining parameters rely on vertical displacement, which within the context of body movement implies a measure of the displacement of the centre of mass (CoM). Previous studies such as that by Zijlstra and Hof (90) have mounted an accelerometer on the lower back, close to the CoM; in which case the vertical accelerations relate to the CoM, and vertical displacement is obtained by double integrating the vertical accelerometer signal. An important element of the research question posed by the present study is whether speed estimation can accurately be made when the accelerometer is worn at the hip. A hip-mounted accelerometer is not able to measure accelerations of the CoM directly, as it is subject to the rotation of the hip in the frontal plane (267). This movement means that the hip height is greater during the swing phase of the gait cycle than in the stance phase. However, use of hip accelerations to measure vertical displacement may be justified if the calculated displacement is averaged over a number of strides in order to cancel the effect of these differences in hip height over the gait cycle. This is the procedure adopted by the present study.

Parameter MVD. This is the mean vertical displacement of the hip over the period of walking. This parameter first requires the identification of gait cycles contained within the accelerometer signal. In order to identify each gait cycle, the heel strike detection algorithm was first employed. The vertical accelerometer signal was twice numerically integrated according to the trapezium rule in order to calculate vertical displacement. To account for integration drift, the study applied a fourth order high pass Butterworth filter to the data. Zijlstra and Hof (90) used a cut off frequency of 0.1Hz, though the present study found that a higher correlation with speed could be achieved from higher cut off frequencies, as discussed in 4.2.3.4 below. **Parameter MVDn** is the mean vertical displacement normalised by leg length.

Parameters ZH0, ZH1, and ZH2. These were influenced by the inverted pendulum model used by Zijlstra and Hof (90). In their model the vertical displacement is first obtained, as described in the previous paragraph. The formula to obtain step length from the vertical displacement is $2\sqrt{2lh - h^2}$ where h is the vertical displacement, and l is the leg length of the walker. The speed is then calculated from the step length and the step duration which can be obtained from the time difference between heel strikes. **Parameter ZH0** has the formula: $ZH0 = 2\sqrt{2l\bar{h} - \bar{h}^2}/\bar{t}$ where l is leg length, $\bar{h}$ is the mean vertical displacement and $\bar{t}$ is the mean step time over the walking period. Both parameter ZH1 and parameter ZH2 were chosen to describe a simpler relationships between vertical displacement, leg length and step time than that represented by ZH0. **Parameter ZH1** is based on the following formula: $ZH1 = l\bar{h}/\bar{t}$. **Parameter ZH2** has the following formula: $ZH2 = \bar{h}/l\bar{t}$ which is effectively the MVDn divided by the mean step time.

4.2.3.3 Heel Strike Detection

A number of the biomechanical parameters derived for the speed estimation algorithm had to be calculated according to gait cycle. It was necessary, therefore, to develop an algorithm which could segment longer periods into individual gait cycles. Figure 13 shows the vertical accelerometer signal for a period of walking and illustrates the regular occurrence of heel strikes, seen as troughs due to the alignment of the vertical accelerometer axis (gravity is in the negative direction). It was not possible to apply a simple threshold algorithm to find these troughs, as signals varied too much between individuals and speeds. Instead, the heel strike detection algorithm first determines the frequency of the signal troughs, then traverses the signal according to this frequency to find each trough.

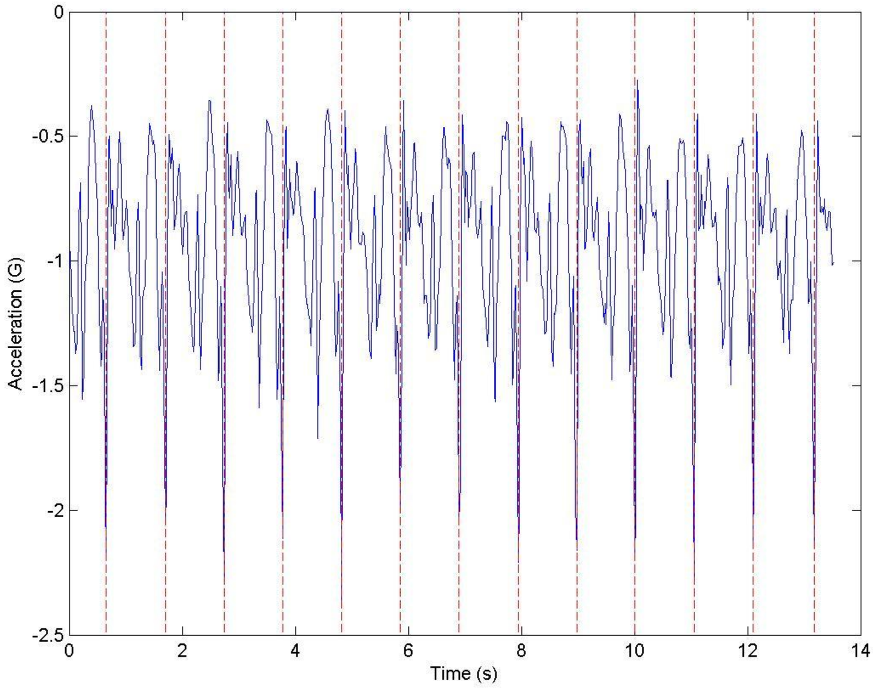

Figure 13: Example vertical acceleration data for walking at the preferred normal speed, taken from one of the participants of the study. The vertical dotted lines represent estimated right heel strikes. It can be seen that these align with troughs in the signal, as expected.

Figure 14 shows the frequency spectrum of a period of walking. As can be seen, there are a number of frequency peaks at amplitudes much higher than the mean amplitude. The first of these, which is often found to have the greatest amplitude, has a frequency that corresponds with the frequency of heel strikes for both feet. To find the value of this frequency, the algorithm first applies a fast Fourier transform to the accelerometer signal to obtain the frequency spectrum. It then isolates those frequencies that have amplitudes that are four or more standard deviations greater than the mean amplitude. The first of these frequencies, corresponding to the heel strike frequency for both feet, is selected. This frequency value is divided by two to obtain the frequency of heel strikes for a single foot – in this case it is the right foot as the accelerometer is fixed to the right hip – and this corresponds to the frequency of a complete gait cycle. It can often be seen from a graph of the FFT signal that the gait cycle frequency aligns with a peak in the frequency spectrum below approximately 1.25Hz (as illustrated by Figure 14). However, this frequency is not reliably obtainable directly from the FFT signal as it does not consistently have the highest amplitude within the range below 1.25Hz, whereas the frequency of heel strikes for both feet (at approximately double the frequency for a single foot) has a consistently high amplitude and is, therefore, more reliably identified.

From the computed gait cycle frequency value and the sample rate (in this case 50Hz), an approximation of the number of samples between heel strikes is obtained. The minimum value of the vertical accelerometer signal is considered to be a clear heel strike. From this starting point the signal is traversed in increments according to the number of samples between heel strikes previously calculated, and at each increment the minimum acceleration value within a predefined close range is designated as a heel strike (Figure 13).

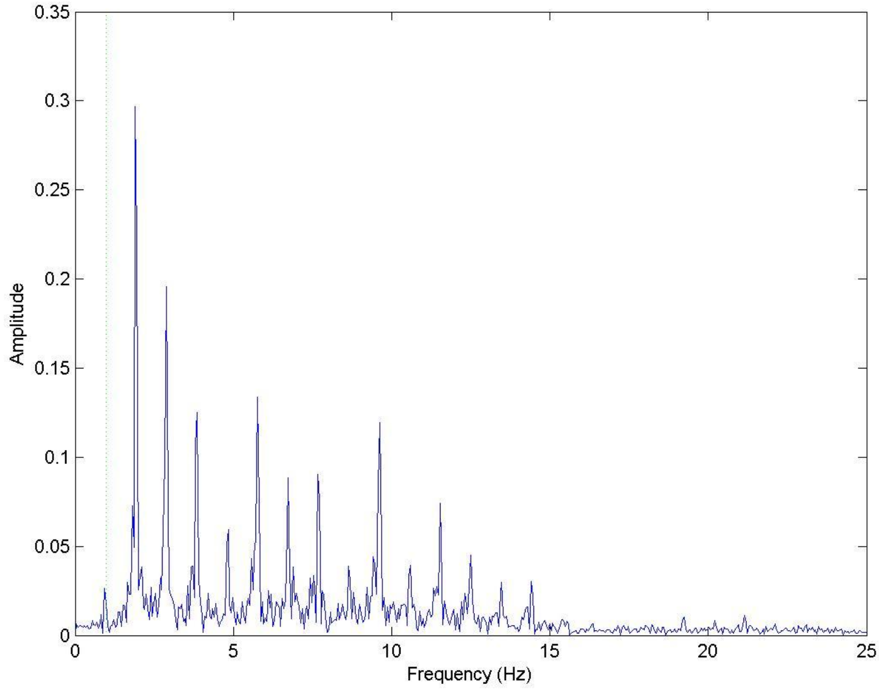

Figure 14: Example frequency spectrum for walking at preferred normal speed, taken from one of the participants of the study. The dotted vertical line represents the estimated frequency of right heel strikes, and in this case it can be seen to align with a minor peak in the frequency spectrum.

4.2.3.4 Optimisation of the Input Parameters

As discussed, the speed estimation models require a number of input parameters from which to derive their predictions for speed, and these parameters need to be selected in order that the speed estimation models are effective across BMI groups and walking modes. Before selecting the final parameters for the speed estimation models, some of these parameters were analysed to identify whether they could be optimised. To assist the optimisation procedure, individual correlations between each of the parameters and walking speed were calculated (Table 13).

The first area considered for optimisation was the cut off frequency for the high-pass signal filter. Of the twenty-two features under consideration (Table 11), eight require a high-pass filter before further processing. Using the entire dataset (all BMI groups, all walking modes), correlation was measured between each filtered parameter and walking speed. This measurement was repeated using high-pass filter cut off values between 0.1Hz and 3.0Hz in steps of 0.1Hz. All eight features returned higher correlation coefficients at around 1.4Hz (see Figure 15). Frequencies of 1.35Hz and 1.45Hz did not improve the results over 1.4Hz. For this reason 1.4Hz was selected as the high-pass cut off frequency for the eight features.

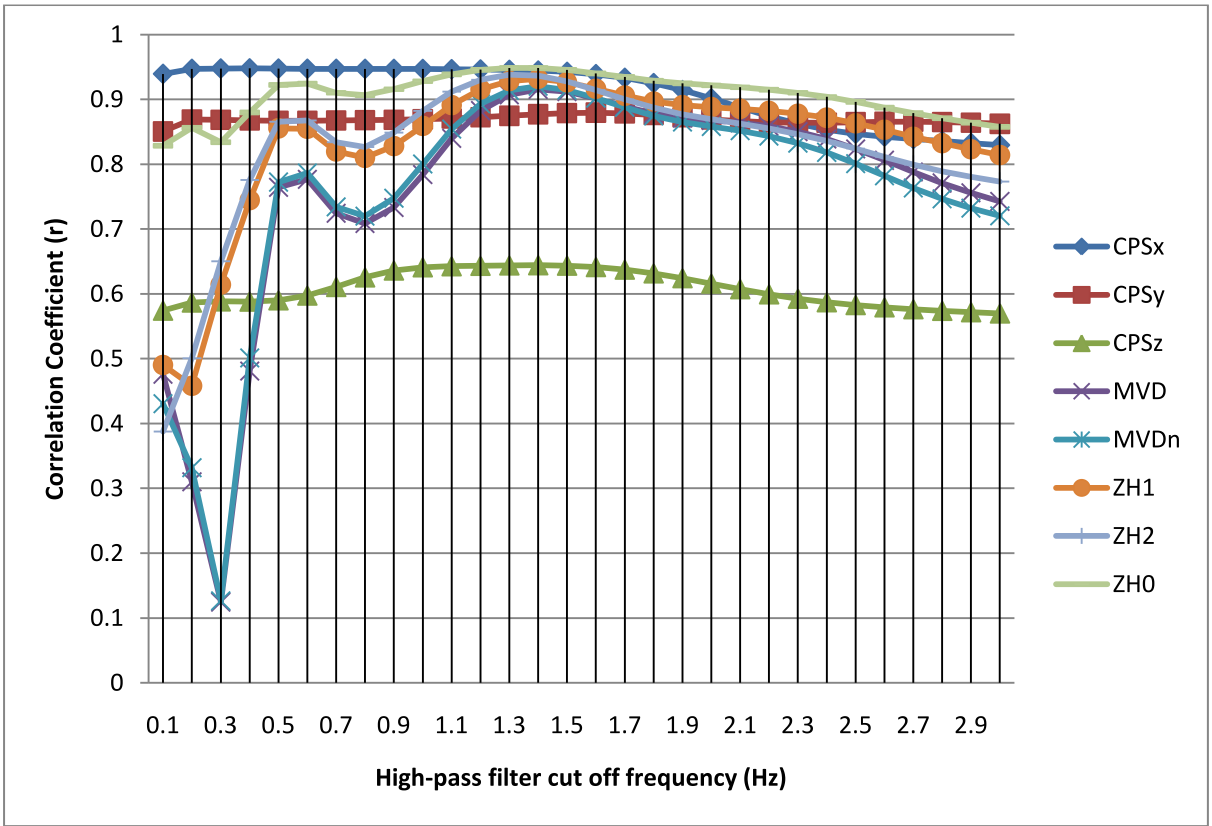

Figure 15: Correlation coefficients at different high-pass filter frequencies.

The second area under consideration was whether certain parameters would benefit from being normalised by leg length. Leg length was a factor which could particularly relate to two of the accelerometer features: mean step time and mean vertical displacement. Those with shorter leg lengths would be expected to exhibit a lower mean step time (parameter MST) to achieve the same speeds as those with longer legs. MST returned a high correlation with walking speed but it could be seen from a scatter plot (Figure 16) that, although individual correlations with speed were linear with similar gradient to the line of best fit for the group as a whole, there were outlying sets of walking speed trials for some individuals. The mean step time normalised by leg length (MSTn) barely improved the correlation coefficient but reduced the number of apparent outliers (see Figure 17). For this reason, the normalised MST value was selected for the analysis. Due to the geometry of the inverted pendulum, the mean vertical displacement (MVD) potentially will return greater values for those with longer leg lengths than those with shorter leg lengths, irrespective of walking speed. Therefore, it is arguable that MVD might be better normalised by leg length for the purposes of the speed estimation model. MVD returned a high correlation with walking speed, with r=0.916 (using the 1.4Hz limit for the high pass filter) for combined walking modes and combined BMI groups. The MVD normalised by leg length (MVDn) returned a similar correlation coefficient with r=0.92. Although there was little difference between the two correlation scores, the normalised version was chosen for analysis due to the initial motivations for the comparison.

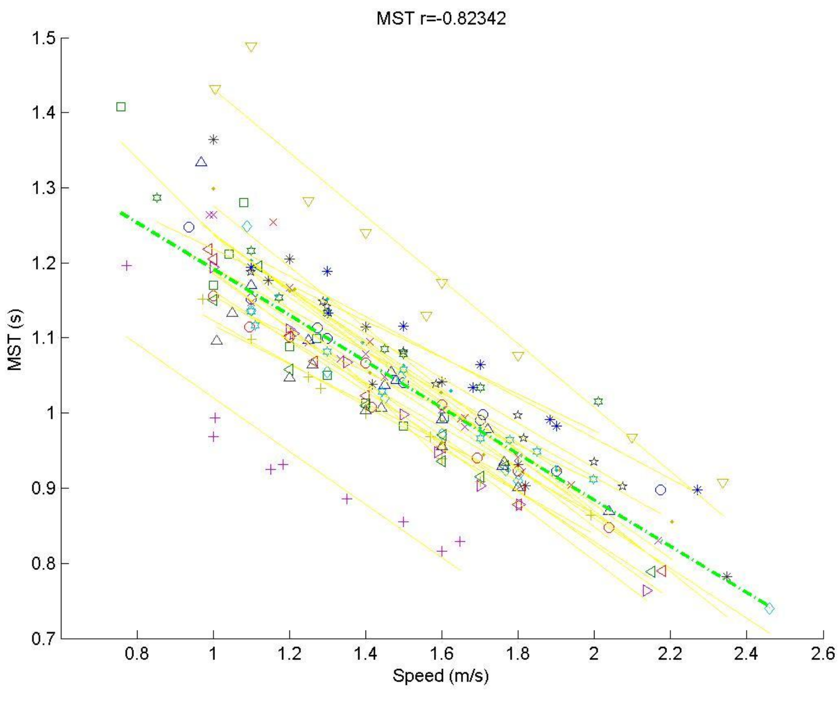

Figure 16: Mean Step Time versus walking speed. Each subject is represented by a different symbol. Each symbol appears nine times for the nine walking trials. Individual regression lines are shown in yellow. Overall line of best fit is green. Although the correlation coefficient is high, walking trials for individuals, such as the subject indicated by '+', were distanced from the line of best fit and added to the variance in the model.

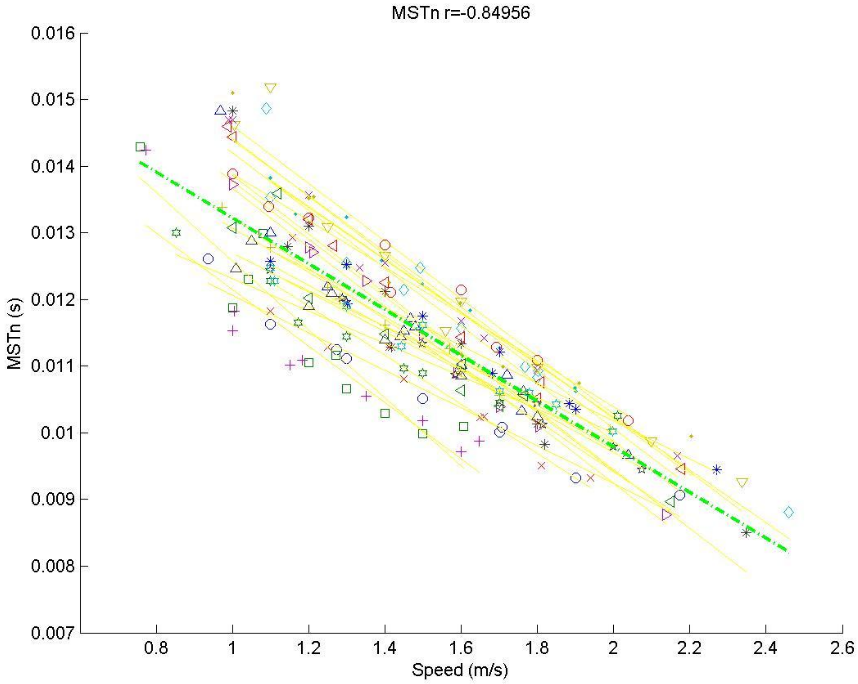

Figure 17: Mean Step Time normalised by leg length versus walking speed. Each subject is represented by a different symbol. Each symbol appears nine times for the nine walking trials. Individual regression lines are shown in yellow. Overall line of best fit is green.

4.2.4 Deriving the Linear Speed Estimation Models

Once the candidate parameters had been defined and optimised, the next phase was to generate linear speed estimation models using combinations of these parameters. A brute-force search [3] was implemented, where all possible combinations of parameters were systematically evaluated to identify the best performing speed estimation model. This technique has been used similarly in previous activity monitoring research to identify an optimum set of accelerometer features (268-269). In the present study, the approach was applied to a dataset containing both treadmill and overground data, for both obese and normal BMI groups, in order to meet the objective of identifying a speed estimation model which applies across BMI groups and walking modes (objective 1).

Twenty features were selected for the brute-force search algorithm. This number represents all the candidate parameters listed in Table 11 except the rejected non-normalised versions of the mean step time (parameter MST) and the mean vertical displacement (parameter MVD),

[3] The term "brute-force search" is used in the field of computer science to describe an exhaustive problem solving approach which tests all possible candidate solutions in order to identify which provide an answer the problem.

the reason for which was discussed above. There were 1,048,575 different combinations of the twenty candidate parameters. Each of these combinations represented the parameters for a linear speed estimation model. Leave-one-out cross-validation was applied for each successive set of parameters as follows. Initially, a single subject was selected for testing. Using the remaining subjects, multiple regression was applied to the subject data corresponding to the current set of parameters, in order to obtain the coefficients of each parameter for the linear speed estimation model. These coefficients were used in the linear model to estimate walking speeds for the individual who was initially selected. Estimated speeds were stored for later analysis, and the process was repeated for each subject in isolation until a set of results was generated for the entire group.

The estimated and actual speeds were collated for all subjects, and accuracy metrics were applied accordingly. The root mean square error (RMSE) was selected as the primary measure of accuracy. Additionally, the mean absolute percentage error (MAPE) was calculated. This measure is usually used to evaluate the accuracy of forecasting models. In this particular instance, MAPE would be able to give an indication of how far speed estimations were from actual speeds, and to help assess how this error would affect categorisation of walking into slow, normal, brisk, and fast modes. Maximum percentage error was also calculated. As each result was generated it was written to a comma separated values (CSV) text file. The results were transferred to a Microsoft SQL Server database system so that structured query language (SQL) queries could be executed to obtain specific sets of results from the more than one million records. In this way it allowed analysis of the models on a larger scale.

The best performing linear models were identified according to three criteria: the model with the lowest RMSE, the model with the lowest MAPE, and the model with the lowest maximum percentage error. Subsequently, these models were applied to the treadmill data in isolation and the overground walking in isolation in order to understand how walking mode affects speed estimation (objective 3). The results for the three walking modes were subdivided by BMI group to allow comparisons to be made (objective 2).

4.2.5 Further Analysis of the Speed Estimation Models

To further investigate the differences in accuracy between BMI groups, and how walking mode affects speed estimation (objectives 2 and 3) the database of one million results was used to identify underlying trends in accuracies according to walking mode and BMI group.

Data were extracted from the database of model results using a number of SQL queries. The first one thousand best performing models (according to RMSE) were selected in each case. One thousand was deemed a sufficient number of results that would allow relationships to be observed between BMI groups. The models that were returned by the queries depended on the criterion of which walking mode is required, and the results were ordered according to the RMSE of the required BMI group. These datasets were visually represented in graph form so that data trends may be inspected.

4.2.6 A Consideration of Repeated Measures

There were nine speed measurements per person (four overground walking speeds and five treadmill walking speeds). This means that the results of applying the speed prediction algorithm were calculated from repeated measures per subject. The algorithm needs to be robust in spite of the differences between the individuals being tested. However, the correlations between speed and accelerometer features may be higher for within-subject measurements than between subjects. Testing all nine measured speeds could, therefore, lessen the effects of inter-subject differences on the results, and the algorithm may consequently return lower RMSE values than it would without repeated measures. This would artificially present a more accurate speed prediction algorithm than is merited.

A secondary test of the algorithm was performed in order to ascertain the effect of the repeated measures on the results. This test was carried out in the same way as the main analysis, except when generating the test data for each subject a single speed was selected at random (using a standard random number generator within MATLAB). This was done separately for the treadmill and overground walking data combined, the treadmill data alone, and the overground walking data alone. Because the test speeds were selected at random, the results necessarily fluctuate according to which speeds have been chosen. The test was, therefore, performed ten times so that an average could be compared with the main results. Only one of the speed prediction models was tested in this way, as this is sufficient to highlight the differences in accuracy between the two approaches. The test was performed ten times for the treadmill and overground data combined, the treadmill alone, and the overground data alone, giving thirty rows of results. The number of iterations was limited by the time required to run the analysis, but this should be adequate to gain insight into the effect of repeated measures on the results.

4.2.7 Walking Speed Statistics

Twenty-two participants each performed walking at various speeds on the treadmill and overground. Table 12 shows the speed statistics by BMI groups and walking modes.

Group	Walking Mode	Mean Speed (m/s)	Std Dev (m/s)	Min Speed (m/s)	Max Speed (m/s)
All Participants	TM+OG	1.47	0.35	0.76	2.46
	Treadmill	1.41	0.27	1.00	2.00
	Overground	1.53	0.42	0.76	2.46
Obese Group	TM+OG	1.42	0.34	0.76	2.35
	Treadmill	1.38	0.25	1.00	1.90
	Overground	1.46	0.42	0.76	2.35
Normal BMI Group	TM+OG	1.52	0.35	0.94	2.46
	Treadmill	1.44	0.28	1.00	2.00
	Overground	1.61	0.40	0.94	2.46

Table 12: Walking speed statistics.

4.3 Results

4.3.1 Coefficients of the Correlations between Individual Parameters and Walking Speed

Correlation coefficients for the relationships between each individual parameter and walking speed are shown in Table 13.

Parameter	All Participants			Obese Group			Normal BMI Group		
	TM+OG	TM	OG	TM+OG	TM	OG	TM+OG	TM	OG
ZH0	0.949	0.931	0.962	0.964	0.950	0.974	0.939	0.923	0.958
CPSx	0.944	0.915	0.957	0.953	0.942	0.956	0.939	0.885	0.966
ZH2	0.937	0.924	0.944	0.958	0.944	0.964	0.915	0.904	0.931
STDx	0.933	0.895	0.950	0.950	0.946	0.950	0.922	0.851	0.957
ZH1	0.932	0.921	0.939	0.943	0.929	0.950	0.928	0.922	0.942
MVDn	0.920	0.924	0.926	0.943	0.940	0.948	0.894	0.905	0.909
MVD	0.916	0.922	0.921	0.934	0.932	0.938	0.898	0.913	0.910
STDy	0.882	0.845	0.902	0.903	0.890	0.913	0.856	0.785	0.887
RangeY	0.879	0.811	0.915	0.883	0.827	0.923	0.872	0.787	0.906
CPSy	0.877	0.834	0.899	0.890	0.865	0.904	0.860	0.792	0.892
RMSy	0.859	0.816	0.884	0.863	0.844	0.869	0.850	0.780	0.895
MSTn	0.850	0.775	0.904	0.883	0.838	0.918	0.852	0.755	0.914
RMSx	0.844	0.803	0.855	0.848	0.812	0.861	0.860	0.812	0.878
MST	0.823	0.733	0.884	0.856	0.781	0.910	0.803	0.700	0.865
RangeX	0.803	0.699	0.859	0.884	0.891	0.875	0.747	0.582	0.852
CPSz	0.644	0.531	0.699	0.836	0.923	0.802	0.668	0.533	0.743
STDz	0.571	0.456	0.618	0.741	0.874	0.699	0.638	0.504	0.706
RangeZ	0.443	0.361	0.494	0.752	0.792	0.716	0.503	0.384	0.583
RMSz	0.406	0.262	0.481	0.424	0.293	0.457	0.489	0.351	0.577
MeanY	0.252	0.204	0.329	0.232	0.273	0.183	0.291	0.174	0.505
MeanX	0.200	0.064	0.306	0.113	0.048	0.145	0.285	0.093	0.433
MeanZ	0.041	0.009	0.077	0.201	0.152	0.240	0.114	0.003	0.188

Table 13: Pearson's r values for correlations between features and walking speeds, for nine combinations of BMI group and walking mode.

4.3.2 Linear Models Identified by the Algorithm

The best performing linear models generated by the brute-force search algorithm are shown in Table 14. There were two models with equal lowest RMSE. Both were selected as they contained notably different features. The model with the lowest MAPE was also selected.

Model No	Feature Set	Description of Model
1	RangeY / MeanZ / RMSx / ZH0	Equal minimum overall RMSE for combined BMI groups and combined walking modes
2	MSTn / RangeY / CPSx / CPSz / ZH0	Equal minimum overall RMSE for combined BMI groups and combined walking modes
3	RangeY / STDz / RMSy / CPSz / ZH2 / ZH0	Lowest overall mean absolute percentage error for combined BMI groups and combined walking modes

Table 14: Linear models selected for further analysis.

4.3.3 Results of Applying the Linear Model

Table 15 presents the accuracies returned by the linear models that were selected through the brute-force search process. Each model was applied to the walking data for both treadmill and overground combined (TM+OG), the treadmill data in isolation, and the overground data in isolation. The accuracy metrics were calculated for each dataset according to three combinations of BMI group: combined obese and normal BMI (ALL), obese only, and normal BMI only. The model numbers correspond with the models listed in Table 14.

Walking Mode	Model	RMSE (ALL)	RMSE (Obese)	RMSE (Normal)	Mean PE	MAPE	MAX PE
TM+OG	1	0.0849	0.0897	0.0797	0.37%	4.80%	28.03%
Treadmill	1	0.0761	0.0804	0.0716	0.36%	4.49%	14.51%
Overground	1	0.0914	0.0992	0.0826	0.68%	5.08%	29.66%
TM+OG	2	0.0849	0.0894	0.0803	0.35%	4.82%	21.27%
Treadmill	2	0.0781	0.0826	0.0736	0.28%	4.61%	17.47%
Overground	2	0.0893	0.0925	0.0858	0.44%	4.75%	23.59%
TM+OG	3	0.0855	0.0864	0.0846	0.36%	4.75%	18.43%
Treadmill	3	0.0780	0.0796	0.0765	0.33%	4.56%	15.92%
Overground	3	0.0886	0.0888	0.0884	0.46%	4.73%	20.42%

Table 15: Accuracies returned by the linear models. RMSE values in ms^{-1}. PE = percentage error. MAPE = mean absolute percentage error.

The primary measure of accuracy adopted by the present study is RMSE. The following table provides addition accuracy metrics for comparison with other studies. These metrics are: standard error of the estimate (SEE) and coefficient of variation (CV). The results are reported for the three walking modes, and for each BMI group: all participants (ALL), the obese group (Obese), and the normal BMI group (Normal).

Walking Mode	Model No	SEE (ALL)	SEE (Obese)	SEE (Normal)	CV (all)	CV (obese)	CV (normal)
TM+OG	1	0.086	0.091	0.081	5.78%	6.33%	5.25%
Treadmill	1	0.077	0.081	0.073	5.38%	5.81%	4.96%
Overground	1	0.093	0.100	0.084	5.96%	6.80%	5.14%
TM+OG	2	0.086	0.091	0.082	5.79%	6.30%	5.29%
Treadmill	2	0.079	0.084	0.075	5.53%	5.97%	5.10%
Overground	2	0.091	0.094	0.087	5.83%	6.34%	5.34%
TM+OG	3	0.087	0.088	0.086	5.82%	6.09%	5.57%
Treadmill	3	0.079	0.081	0.078	5.52%	5.75%	5.30%
Overground	3	0.090	0.090	0.090	5.78%	6.08%	5.50%

Table 16: SEE and CV results for the linear models.

The following table shows the mean and standard deviation of percentage error for use in comparison with other studies which use these accuracy metrics.

Walking Mode	Model	Mean PE	Mean PE (Obese)	Mean PE (Normal)	Std Dev PE	Std Dev PE (Obese)	Std Dev PE (Normal)
TM+OG	1	0.37%	1.03%	-0.29%	6.17%	6.83%	5.40%
Treadmill	1	0.36%	0.97%	-0.24%	5.70%	6.16%	5.20%
Overground	1	0.68%	2.01%	-0.68%	6.80%	7.96%	5.11%
TM+OG	2	0.35%	0.90%	-0.20%	6.12%	6.71%	5.45%
Treadmill	2	0.28%	0.45%	0.12%	5.81%	6.39%	5.23%
Overground	2	0.44%	1.53%	-0.68%	6.53%	7.19%	5.64%
TM+OG	3	0.36%	1.31%	-0.58%	6.07%	6.35%	5.65%
Treadmill	3	0.33%	0.65%	0.01%	5.82%	6.26%	5.40%
Overground	3	0.46%	1.99%	-1.11%	6.35%	6.67%	5.66%

Table 17: mean and standard deviation of percentage error for the linear models.

The following table presents the mean and standard deviation of the error in ms^{-1} used for comparison with other studies.

Walking Mode	Model	Mean Err (m/s)	Mean Err (m/s) Obese	Mean Err (m/s) Normal	Std Dev Err (m/s)	Std Dev Err (m/s) Obese	Std Dev Err (m/s) Normal
TM+OG	1	-0.0002	0.0062	-0.0066	0.0851	0.0900	0.0798
Treadmill	1	0.0006	0.0084	-0.0071	0.0765	0.0807	0.0719
Overground	1	0.0022	0.0139	-0.0096	0.0919	0.0993	0.0830
TM+OG	2	-0.0002	0.0033	-0.0037	0.0852	0.0897	0.0806
Treadmill	2	-0.0002	0.0005	-0.0010	0.0785	0.0833	0.0742
Overground	2	-0.0003	0.0072	-0.0080	0.0898	0.0932	0.0864
TM+OG	3	0.0001	0.0099	-0.0097	0.0857	0.0863	0.0844
Treadmill	3	0.0005	0.0038	-0.0027	0.0784	0.0802	0.0772
Overground	3	0.0003	0.0157	-0.0154	0.0891	0.0883	0.0880

Table 18: mean and standard deviation of error in ms^{-1} for the linear models

The following tables (Table 19 and Table 20) present accuracy results returned by the linear models when the speeds are limited to two ranges: "normal" range from $1.0ms^{-1}$ to $1.3ms^{-1}$, and “brisk” range from $1.4ms^{-1}$ and $1.7ms^{-1}$. These results were calculated to allow a more direct comparison with other studies which use limited speed ranges.

Walking Mode	Model	RMSE (ALL)	RMSE (Obese)	RMSE (Normal)	Mean PE	MAPE	MAX PE
TM+OG	1	0.0986	0.1179	0.0709	-0.96%	6.25%	24.32%
Treadmill	1	0.0933	0.1105	0.0697	-0.85%	5.96%	23.25%
Overground	1	0.1067	0.1308	0.0681	-0.96%	6.85%	23.95%
TM+OG	2	0.0688	0.0759	0.0599	0.23%	5.05%	14.54%
Treadmill	2	0.0667	0.0745	0.0571	0.14%	4.82%	15.43%
Overground	2	0.0645	0.0750	0.0492	0.10%	4.65%	13.25%
TM+OG	3	0.0662	0.0635	0.0691	0.29%	4.72%	16.98%
Treadmill	3	0.0682	0.0656	0.0710	0.62%	4.78%	15.79%
Overground	3	0.0774	0.0883	0.0620	0.18%	5.48%	16.67%

Table 19: Accuracies returned by the linear models when speeds are limited to the range $1.0ms^{-1}$ to $1.3ms^{-1}$.

Walking Mode	Model	RMSE (ALL)	RMSE (Obese)	RMSE (Normal)	Mean PE	MAPE	MAX PE
TM+OG	1	0.0640	0.0650	0.0628	0.03%	3.28%	13.07%
Treadmill	1	0.0646	0.0677	0.0613	0.24%	3.28%	12.25%
Overground	1	0.0703	0.0861	0.0474	-0.35%	3.33%	13.76%
TM+OG	2	0.0629	0.0614	0.0645	0.09%	3.27%	12.98%
Treadmill	2	0.0642	0.0606	0.0679	0.16%	3.11%	12.48%
Overground	2	0.0535	0.0599	0.0455	0.08%	2.72%	7.45%
TM+OG	3	0.0637	0.0577	0.0695	0.12%	3.27%	13.78%
Treadmill	3	0.0669	0.0606	0.0729	0.16%	3.24%	12.75%
Overground	3	0.0693	0.0640	0.0748	0.20%	3.84%	8.42%

Table 20: Accuracies returned by the linear models when speeds are limited to the range $1.4ms^{-1}$ to $1.7ms^{-1}$.

The following table (Table 21) shows the results for model 1 when applied to one speed per person to eliminate the possible influences on the results of using repeated measures. Ten tests were repeated, for each person a single speed was selected at random from up to nine walking trials. The ten results are ordered by RMS for all subjects for each walking mode. The mean and standard deviation of the speeds chosen by the random selection process are also shown.

Walking Mode	RMS (ALL)	RMS (Obese)	RMS (Normal)	Mean PE	MAPE	MAX PE	Mean Test Speed (m/s)	Std Dev. Test Speed (m/s)
TM+OG	0.0774	0.0915	0.0602	0.88%	4.60%	13.50%	1.51	0.36
	0.0850	0.0957	0.0728	0.17%	4.72%	13.98%	1.56	0.36
	0.0897	0.0981	0.0804	0.20%	5.10%	11.17%	1.53	0.41
	0.0914	0.1093	0.0690	0.42%	5.79%	18.75%	1.47	0.33
	0.0945	0.1150	0.0679	0.20%	5.23%	16.39%	1.46	0.33
	0.0954	0.1009	0.0896	1.06%	5.47%	24.59%	1.54	0.41
	0.0981	0.1038	0.0921	0.05%	6.12%	14.44%	1.45	0.32
	0.0988	0.1207	0.0706	-0.29%	6.72%	21.44%	1.35	0.23
	0.1019	0.1232	0.0748	-0.32%	5.58%	11.85%	1.50	0.45
	0.1134	0.1365	0.0841	-0.43%	6.44%	24.61%	1.46	0.35
Mean:	0.0946	0.1095	0.0761	0.19%	5.58%	17.07%		
Treadmill	0.0632	0.0669	0.0593	0.15%	3.86%	8.97%	1.39	0.27
	0.0641	0.0613	0.0667	0.11%	3.86%	10.93%	1.41	0.29
	0.0742	0.0726	0.0757	0.52%	4.19%	11.93%	1.48	0.29
	0.0811	0.0885	0.0728	0.31%	3.98%	9.73%	1.55	0.20
	0.0913	0.0957	0.0867	0.96%	5.06%	14.13%	1.43	0.23
	0.0939	0.1054	0.0807	0.53%	5.92%	23.32%	1.42	0.30
	0.0948	0.0979	0.0916	0.45%	6.27%	17.00%	1.34	0.25
	0.0986	0.1070	0.0895	1.36%	6.40%	20.60%	1.36	0.28
	0.1070	0.1325	0.0729	0.96%	5.77%	27.23%	1.50	0.25
	0.1275	0.1631	0.0770	-0.17%	7.13%	29.03%	1.38	0.30
Mean:	0.0896	0.0991	0.0773	0.52%	5.25%	17.29%		

Continues …

Walking Mode	RMS (ALL)	RMS (Obese)	RMS (Normal)	Mean PE	MAPE	MAX PE	Mean Test Speed (m/s)	Std Dev. Test Speed (m/s)
… continued								
Overground	0.0810	0.0948	0.0643	0.14%	4.86%	11.17%	1.50	0.38
	0.0835	0.0905	0.0760	0.32%	5.14%	16.31%	1.53	0.37
	0.0870	0.0808	0.0927	0.33%	4.74%	11.40%	1.57	0.41
	0.0891	0.1015	0.0745	0.87%	5.96%	25.84%	1.31	0.31
	0.0914	0.1064	0.0734	0.18%	5.44%	23.80%	1.54	0.38
	0.0933	0.1057	0.0790	0.78%	5.20%	35.11%	1.48	0.41
	0.1026	0.1172	0.0856	0.53%	5.69%	29.84%	1.58	0.50
	0.1097	0.1190	0.0996	0.80%	6.05%	21.99%	1.54	0.48
	0.1139	0.1080	0.1195	0.39%	5.89%	13.16%	1.61	0.42
	0.1148	0.1285	0.0992	0.64%	6.11%	17.29%	1.64	0.49
Mean:	0.0966	0.1052	0.0864	0.50%	5.51%	20.59%		

Table 21: Results for model 1 when applied to single speeds for each subject.

4.3.4 Model Performance according to BMI Group and Walking Mode

The purpose of generating the results presented in this subsection was to investigate the differences in accuracy between BMI groups, and how walking mode affects this difference (objectives 2 and 3). There are seven graphs presented. The first three (Figure 18, Figure 19, and Figure 20) represent the differences in RMSE accuracy between BMI groups in relation to the RMSE (for combined treadmill and overground walking, overground only, and treadmill only data respectively).

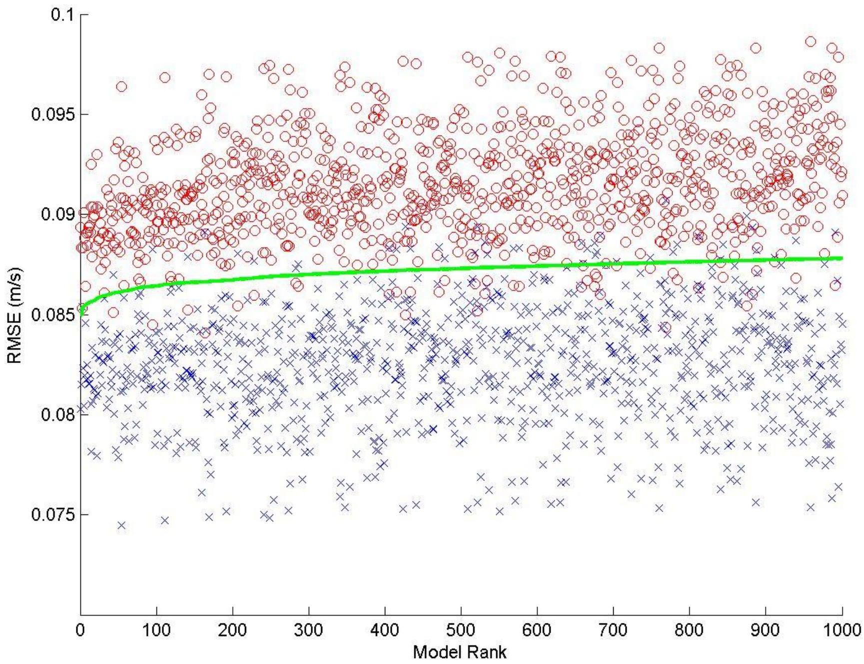

Figure 18: Accuracies for combined walking modes. RMSE values for the obese (o) and normal BMI (x) groups for a subset of 1,000 linear models ranked by overall RMSE (green line). Values calculated using combined obese and normal BMI data.

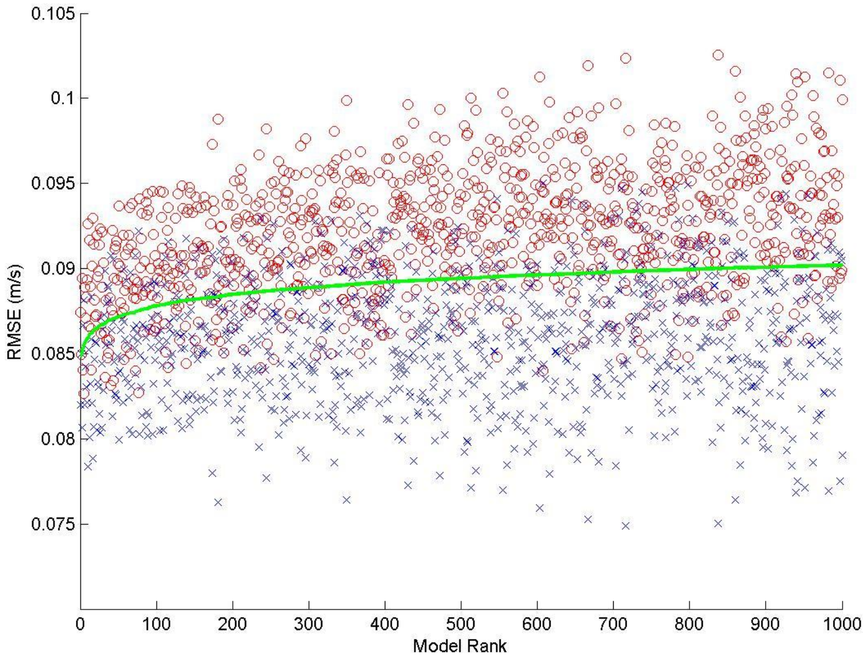

Figure 19: Accuracies for overground walking. RMSE values for the obese (o) and normal BMI (x) groups for a subset of 1,000 linear models ranked by overall RMSE (green line). Values calculated using combined obese and normal BMI data.

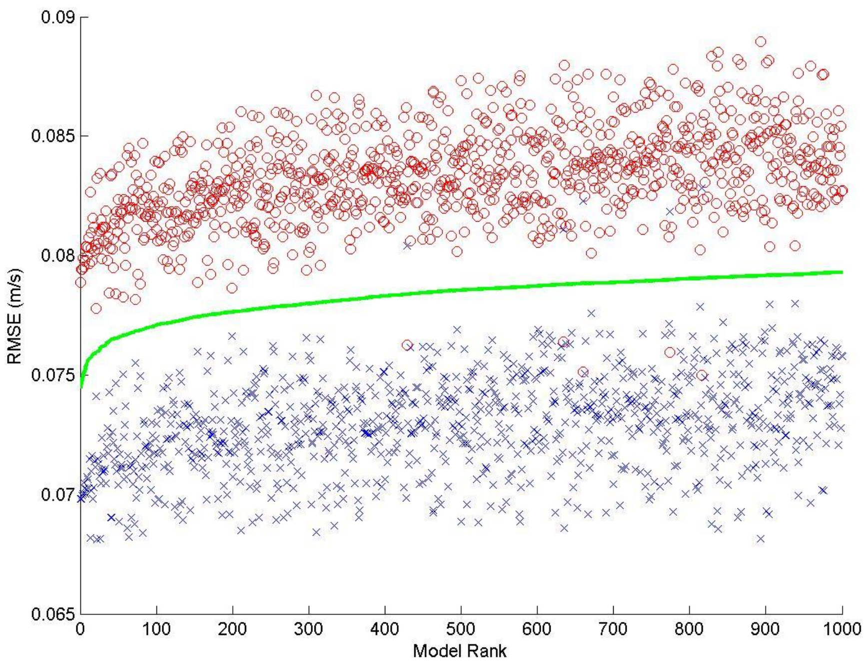

Figure 20: Accuracies for treadmill walking. RMSE values for the obese (o) and normal BMI (x) groups for a subset of 1,000 linear models ranked by overall RMSE (green line). Values calculated using combined obese and normal BMI data.

The following four graphs illustrate how one BMI group performs compared to another when the top 1,000 best performing models are selected according to the first BMI group. For the first of the two graphs representing this relationship (Figure 21 and Figure 22) overground walking data is used, and for the second two graphs (Figure 23 and Figure 24) treadmill walking data is used. These graphs highlight whether the same models are effective across BMI groups for different walking modes.

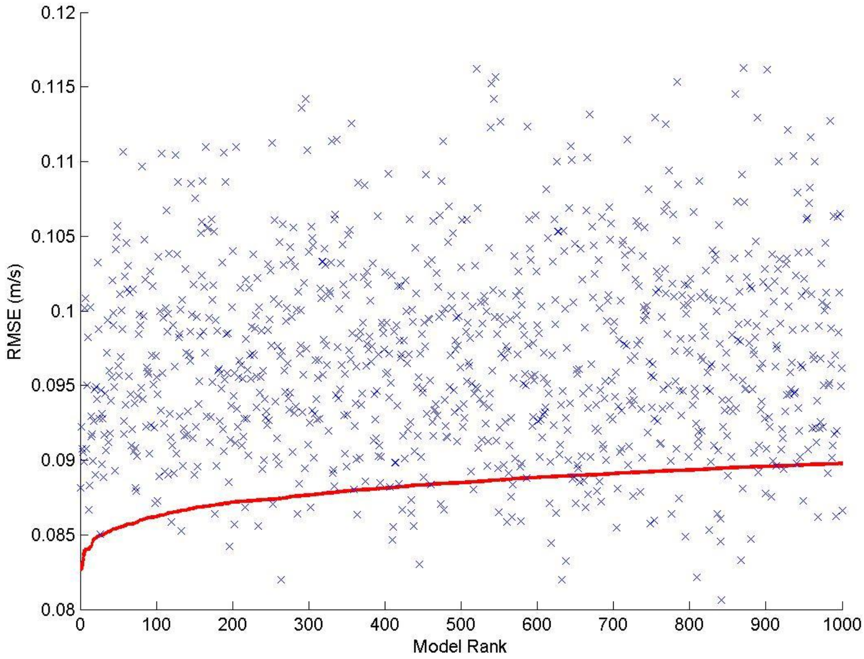

Figure 21: Accuracies for overground walking. RMSE values for the obese (red line) and normal BMI (x) groups for a subset of 1,000 linear models ranked by obese RMSE. Values calculated using combined obese and normal BMI data.

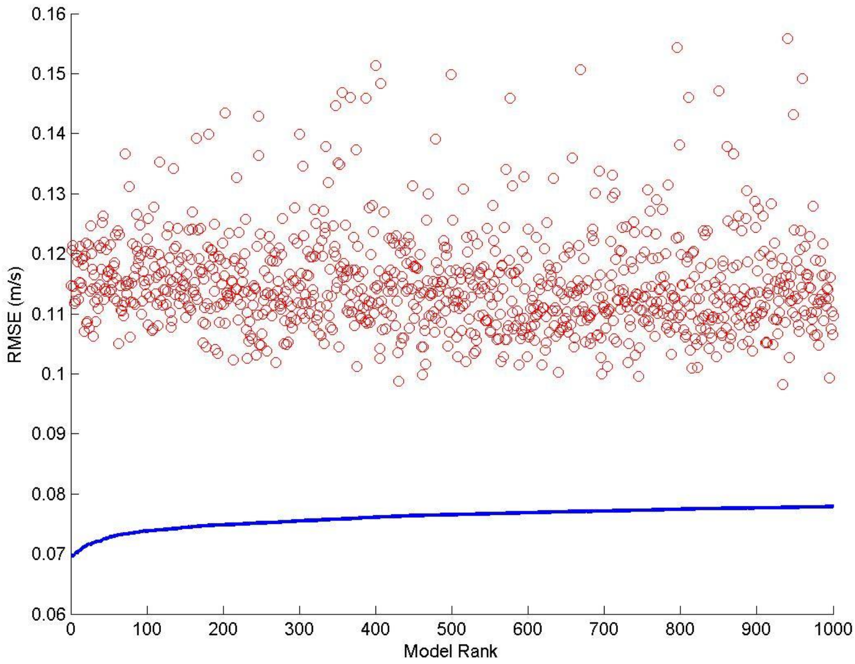

Figure 22: Accuracies for overground walking. RMSE values for the obese (o) and normal BMI (blue line) groups for a subset of 1,000 linear models ranked by normal BMI RMSE. Values calculated using combined obese and normal BMI data.

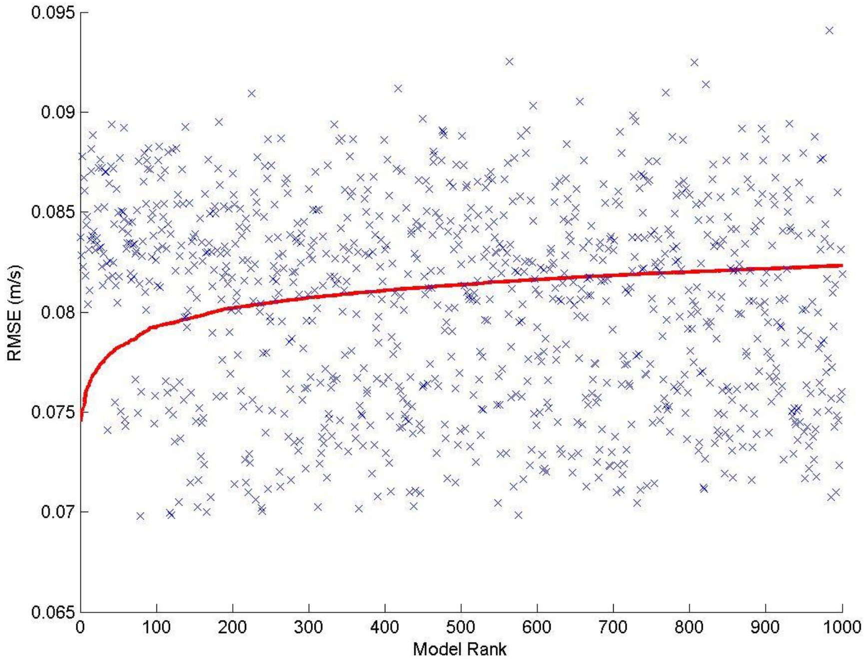

Figure 23: Accuracies for treadmill walking. RMSE values for the obese (red line) and normal BMI (x) groups for a subset of 1,000 linear models ranked by obese RMSE. Values calculated using combined obese and normal BMI data.

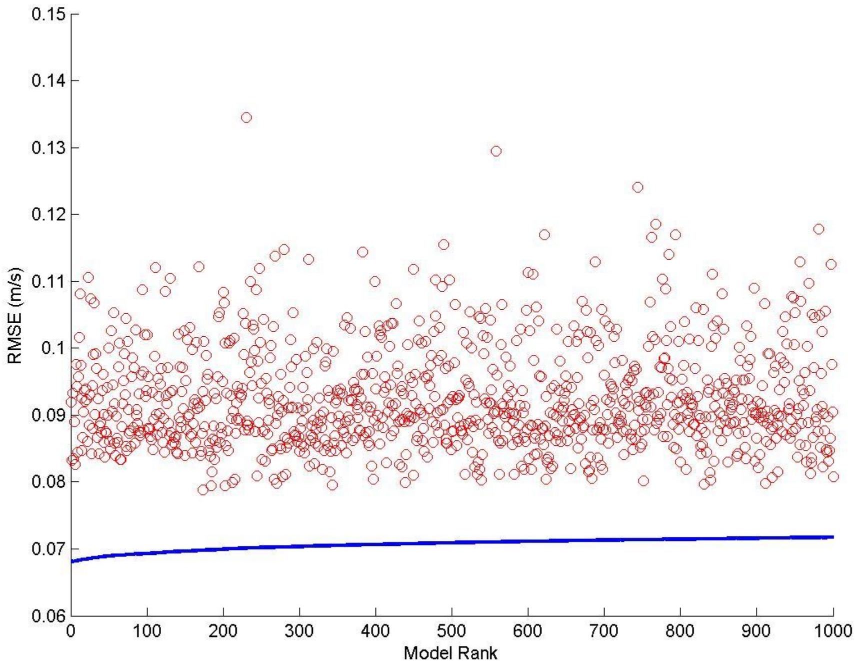

Figure 24: Accuracies for treadmill walking. RMSE values for the obese (o) and normal BMI (blue line) groups for a subset of 1,000 linear models ranked by normal BMI RMSE. Values calculated using combined obese and normal BMI data.

4.4 Discussion

The following research question was asked: can a hip-mounted accelerometer be used to accurately estimate walking speeds for obese groups?

The following objectives were set:

- **Objective 1:** to produce a speed estimation model that is accurate across BMI groups and walking modes.
- **Objective 2:** to investigate how walking speed estimation accuracy differs between obese and normal BMI groups.
- **Objective 3:** to explore how walking speed estimation accuracy is affected by walking mode.

To answer the question and meet the objectives of the study, twenty-two subjects (n=11 obese, n=11 normal BMI) wore hip-mounted accelerometers and performed overground and treadmill walking at a number of different speeds. Several features were derived from the accelerometer signals for use as candidate parameters to a linear speed estimation model. All possible combinations of these parameters were tested against the entire accelerometer dataset for their speed estimation capabilities in order to identify the best performing models (objective 1). This analysis identified three models which were then tested against treadmill data in isolation and overground walking in isolation in order to investigate how model accuracy was affected by walking mode (objective 3). Results were split according to the three combinations of BMI group to investigate how accuracy differs according to BMI (objective 2) Further analysis was performed to explore whether speed estimation was less accurate for subjects with high BMI levels when compared to normal BMI subjects (objective 2).

The main findings of the study are as follows. For the best performing linear models identified by the analysis, RMSE values of less than 0.085ms^{-1} (Table 15) were achieved across BMI groups for both modes of walking, also MAPE values of less than 5% were returned. These results compare with previous work, as will be discussed in 4.4.3 and 4.4.4 below, and suggest that it is possible to obtain accurate speed estimates from hip-mounted accelerometer data, for both obese and normal BMI groups, using a single speed estimation model. Despite this, a comparison of speed estimation models shows that accuracy changes according to BMI group, and the best performing models for the normal BMI group do not correspond with those for the obese group. This is particularly evident when considering how the best models for the normal BMI group perform when applied to the obese group, and this difference in accuracies is more pronounced for overground walking.

4.4.1 Accuracy of Brisk Walking

In weight-loss programmes involving walking it is important that a brisk walking pace is achieved by participants, as this corresponds with moderate exercise associated with weight loss (270). The average walking speed is around 1.4ms^{-1} (271-272), but the speed which constitutes brisk walking may differ among individuals according to their weight and leg length. However, the threshold for brisk walking is likely to fall in the range 1.4ms^{-1} to 1.7ms^{-1}. The results of applying the speed estimation models to this speed band (Table 20) show improved accuracies compared with the best performing models over the full range of speeds, with RMSE values of around 0.06ms^{-1} and MAPE of around 3%. These results show that this crucial range of normal and brisk speeds can be assessed with good accuracy. Additionally, under free-living conditions walking speeds are likely to lie predominantly within a limited range. Walking speeds of 1.8ms^{-1} and over are possible, but high speeds are not sustainable for long periods, especially for obese walkers. Consequently, this may increase confidence in the accuracy of the speed estimation model.

4.4.2 Analysis of the Parameters Used in the Best Performing Models

The best performing linear models, based on the metrics RMSE and MAPE, each included the ZH0 parameter taken from the Zijlstra and Hof (90) approach to measuring walking speed. This is, perhaps, expected as this parameter was highly correlated with walking speed (Table 13). In fact, the only features to consistently return correlation coefficients with speed above 0.9 for all the various combinations of BMI group and walking modes were the ones that were a function of mean vertical displacement, leg length, and mean step time (ZH0, ZH1 and ZH2). These results suggest that the ZH0 feature may be a robust predictor of walking speed regardless of walking mode or subject BMI.

RangeY (the range of anteroposterior accelerations) was also included in each of the linear models. RangeY returned r=0.879 when correlated with speed using combined BMI groups and combined walking mode data. For parameter RangeY there was a distinct difference between correlation values for treadmill and overground data, and this was the case for all BMI group combinations. The most disparate of the correlation coefficients for the RangeY parameter was between treadmill (r=0.787) and overground walking (r=0.906) in the normal BMI group, and in the combined BMI group and the obese group each showed a difference in r values of 0.1 between the two walking modes.

Model 1 (Table 14) included RMSx (the root mean square of the vertical accelerations) and MeanZ (the mean of the mediolateral accelerations) in addition to the ZH0 and RangeY parameters. RMSx when correlated with walking speed consistently returned r values over r=0.8 across the BMI groups and walking modes, and was ranked midway among the other parameters. MeanZ, however, was the lowest ranking parameter for combined BMI groups and walking modes with r=0.041, and returned a correlation coefficient as low as r=0.003 for treadmill walking in the normal BMI group. It is surprising, then, that the inclusion of this parameter improves the RMSE of the speed estimation model. It may be that this parameter contains information enough to improve speed estimates by a small margin over the same model with MeanZ removed.

Model 2 (Table 14) incorporated CPSx and CPSz (counts per second for the vertical and mediolateral accelerometer axes) in addition to MVDn (mean vertical displacement normalised by leg length) and the two parameters common to all models (ZH0 and RangeY). The CPSx parameter bears similarity to the RMSx parameter in model 1 – both are essentially a measure of magnitude of the vertical accelerometer axis – though CPSx has a higher correlation coefficient with speed, with r values in excess of 0.9 for all combinations of BMI group and walking mode except normal BMI treadmill walking (r=0.885). CPSz shows relatively low correlations with speed for most combinations of walking mode and BMI group, with r values as low as r=0.53. However, these values rise to over 0.8 for the obese group in isolation, and for obese treadmill walking a value of r=0.92 is returned. These CPSz correlation figures may be partially responsible for the performance of this model across BMI groups. MVDn, although correlated with walking speed, may be considered an unexpected parameter in this model, as mean vertical displacement and leg length are already accounted for in the ZH0 parameter.

Model 3 (Table 14) comprised STDz (the standard deviation of the mediolateral accelerometer axis), RMSy (the root mean square of the anteroposterior axis), ZH2 (mean vertical displacement normalised by leg length, divided by mean step time), and CPSz, in addition to RangeY and ZH0. STDz is another low ranking parameter in terms of correlation with speed, though in a similar manner to CPSz , correlation improves for the obese group, and treadmill walking in particular. It would seem unexpected that the STDz and CPSz would both appear in the same model, as both in some way represent variation of the mediolateral, and both have a similar profile of correlations with speed across the BMI groups and walking modes. Similarly, the inclusion both ZH2 with the ZH0 parameter would not necessarily be expected to improve the model compared with ZH0 alone, as both are functions of leg length, mean vertical displacement and step time.

The brute-force search approach which generated the linear models did not test for correlations between parameters. However, this approach thoroughly tested every possible combination of parameter. Furthermore, leave-one-out cross-validation was implemented when generating the evaluation metrics for each model, which means the results were not subject to overfitting. This method gives a good indication of how the algorithm would perform when applied to new subjects that were not in the original dataset. Therefore, the accuracies obtained may be indicative of those which might be expected in subsequent real-world setting.

4.4.3 Comparison with Previous Linear Approaches

Previous studies have used a number of metrics to evaluate the accuracy of their models. Among other metrics, RMSE (118, 210, 273), correlation (207), SEE (85, 200, 207), maximum percentage error (198), and average percentage accuracy (209) have been used.

The present study has calculated several accuracy metrics to allow direct comparison with previous work. However, there are a number of elements to consider which affect how results may be compared between studies. In the present study, three groupings of the participants were considered – the obese group, the normal BMI group, and the combined obese and normal BMI group – whereas previous research has not generally broken down results for specific BMI groups. Previous studies have used treadmill and overground walking in a variety of combinations to train and test their algorithms, as discussed in subsection 2.4.1. The present study calculated separate results based on three walking modes comprising treadmill, overground, and overground and treadmill combined. There are, therefore, nine sets of results for each speed estimation model. Also, the present study aimed to test whether speed estimation models may be applied to a population of mixed BMI individuals without the need to tailor the algorithms to the individual, whereas certain studies have chosen individual calibration.

Schutz et al. (200) investigated the correlation between the RMS of accelerometer output and walking speed. Both linear and quadratic models were applied to a group of fifty healthy women, a large proportion of which were obese (mean BMI was 31.4 ± 5.1kgm^{-2}). Although they observed high correlations between RMS and walking speed for individuals, there was a great amount of inter-subject variance apparent in the group data – for overground walking mean SEE values for the group model were 0.352ms^{-1} and 0.325ms^{-1} for the linear and quadratic models respectively, compared with 0.056ms^{-1} and 0.042ms^{-1} for the equivalent individually calibrated models. The study by Barnett et al. (207) also compared accuracy between a group model and individual calibration. In this case the relationship between accelerometer counts and walking speed was investigated, and their testing was limited to overground walking. The SEE values returned by Barnett et al. were 0.161ms^{-1} and 0.053ms^{-1} for group calibration and individual calibration models respectively, though a SEE score of 0.044ms^{-1} was also reported when the accelerometer itself was also calibrated to the individual. The present study returned SEE values of around 0.086ms^{-1} in several of the models tested against the combined walking mode and BMI data, and around 0.09ms^{-1} in the isolated overground walking for the combined BMI group (see Table 16). The lowest SEE returned by overground walking for the present study was 0.084ms^{-1} for the isolated normal BMI group, and the lowest SEE for the obese group for overground walking was 0.09ms^{-1}. The SEE values returned by the present study better the group calibration results of Barnett el al. and Schutz et al., and though they fall short of the individually calibrated results they still may be considered to compare well. Although individual calibration leads to higher speed estimation accuracies, it is not desirable when assessing large populations, as each subject is required to perform a calibration procedure under laboratory conditions. It is arguable that, when considering large scale studies or interventions, the expense of the calibration procedure is not worth the additional accuracy it may produce when compared with the present study.

A study by Bonomi et al. (85) investigated a multi-linear speed estimation model based on accelerometer features. The model was developed using fifteen participants and validated against another five. The study returned a SEE value of $0.056ms^{-1}$ for outdoor walking data. This result represents good accuracy comparable to the individually calibrated models discussed above. However, five participants in the validation group might be regarded as an insufficient number to reliably test the performance of the algorithm compared with the leave-one-out cross-validation approach used in the present study.

The study by Panagiota et al (199) used hip-mounted accelerometer to estimate walking speed, and also incorporated BMI in the speed estimation algorithm. This study merits detailed discussion, given the similarities with the present study. In the study the walking took place overground, and the participants were split into two groups to perform walking outdoors (n=20) and indoors (n=17). Two walking speed categories were considered: normal walking and brisk walking. Outdoors the participants were free to set the walking pace according to these categories. Indoor speeds were imposed at $1.33\ ms^{-1}$ and $1.55ms^{-1}$ for the two categories; participants were assisted in maintaining these speeds by a regular audible signal. The study used a multi-linear speed estimation model based on ten features, five of which were accelerometer features, and the other five incorporated anthropometric measurements including height, weight and BMI. In this sense their approach could be considered as accounting for obesity. There were, however, an insufficient number of obese participants to allow any conclusions to be drawn about the effects of obesity on the algorithm: there were n=37 with mean BMI of $24.96 \pm 3.24\ kg/m^2$, as opposed to the eleven out of twenty-two participants in the present study who exceeded a BMI of $30\ kg/m^2$.

Panagiota et al. applied leave-one-out cross-validation to their linear model, and accuracy was evaluated using the mean and standard deviation of the percentage error, and also the mean and standard deviation of the error in ms^{-1}. It was possible to split the accuracy evaluation between normal and brisk speeds due to the method of data collection, and also the standard deviation of speed was low for both mean walking speeds making overlap of speed categories less likely (mean outdoor speeds were 1.38 ± 0.08ms^{-1} and 1.78 ± 0.08ms^{-1}, and indoor speeds were 1.34 ± 0.03 ms^{-1} and 1.55 ± 0.02ms^{-1} for normal and brisk walking respectively). Most other studies, however, elect to report a single set of accuracy results for the full range of walking speeds. Panagiota et al. achieved an error of -0.01 ± 0.07ms^{-1} and a mean percentage error of -0.81 ± 4.90% for the normal walking pace. For the brisk walking pace an error of 0.02 ± 0.08ms^{-1} and a percentage error of 1.01 ± 4.94% was returned. The best results for overground walking in the present study was from model number 3 which returned a mean error of 0.000 ± 0.089ms^{-1} (Table 18) and a mean percentage error of 0.46 ± 6.35% for the mixed BMI group (Table 17). The obese group in isolation returned similar standard deviations, but the mean showed slight positive bias with an error of 0.016 ± 0.88ms^{-1} and a percentage error of 1.99 ± 6.67%. The normal BMI group has a small negative bias with an error of -0.015 ± 0.088ms^{-1} and a percentage error of -1.11 ± 5.66%.

Although the present study apparently returns marginally lower accuracies for the overall dataset than Panagiota et al., this may be greatly explained by the different approach to categorising speeds for evaluation. The present study has a greater variety of speeds represented, ranging from 0.75ms^{-1} to 2.46ms$^{-1,}$ (see Table 12) and accuracy is assessed across this full range of speeds. If we consider the results of the present study that were limited by speed bands, model 2 returned RMSE values around 0 ± 0.056ms^{-1} and mean percentage error of 0.08 ± 3.43% for overground walking in the 1.4ms^{-1} to 1.7ms^{-1} range (Table 20). The 1.0ms^{-1} to 1.3ms^{-1} range returned mean error values of around 0ms^{-1} with standard deviation of 0.066ms^{-1}, and mean percentage error of 0.1 ± 5.75% for model 2 (Table 19). Given these results, it is arguable that the estimation model in the current study may be superior to that in Panagiota et al. Also, because Panagiota limited testing to two relatively narrow bands of speeds, it is not certain that their model would perform as well across a broader range of speeds.

4.4.4 Comparison with Previous Non-Linear Speed Estimation Models

In 1995 Aminian et al. (264) first applied an artificial neural network to walking speed estimation from accelerometer output, and this was followed shortly afterwards by a similar study by the same researchers (198). Both studies used accelerometer features from treadmill walking at a range of speeds to train the ANN. The ANN was then used to estimate walking speeds from overground walking data. The first study reported that the maximum of the coefficient of variation (CV) of speed estimation was 6%, and the second study found the maximum of speed-predicted error was 16%. The present study reports several models with comparable CV values around 6% for all BMI groups across all modes of walking (Table 16). The lowest maximum error reported by the present study for combined overground and treadmill walking and combined BMI groups was 18.43%, though a maximum error value of 7.45% was recorded for model 2 for overground walking when speeds were limited between 1.4ms^{-1} and 1.7ms^{-1}. However, the maximum error is not a good measure of performance as it can be affected by a single outlier and does not give any indication as to the general ability of the estimation model.

A novel approach by Mannini et al. (210) applied a state vector machine (SVM) to speed estimation. The study employed a thigh-mounted accelerometer to collect treadmill data only. Their speed estimation model achieved an RMS error of around 0.08ms^{-1} which is comparable with the present study. However, the study by Mannini et al. incorporated both walking and running speeds between 0.33ms^{-1} and 2.67ms^{-1}. The walking speed results are not reported separately, but it is possible that a higher (or lower) accuracy may have been achieved for walking speed estimation alone.

4.4.5 Effect of BMI Group and Walking Modes on Accuracy

The present study has achieved an equivalent level of speed estimation accuracy with that of previous research for the combined BMI groups, and for the obese and normal groups in isolation. Within the study there are, however, a notable differences between the accuracies obtained for each BMI group and for each mode of walking.

The speed estimation models in the present study were derived from data containing both walking modes and both BMI groups and were chosen to give the best overall performance for the entire dataset of mixed BMI and walking mode data. A model derived this way necessitates a compromise in accuracy for individual walking modes and BMI groups in favour of overall accuracy.

When considering the best performing models based on combined BMI group and combined walking modes the RMSE for the obese group is generally higher than the normal group. Figure 18 shows a plot of the RMSE values for each BMI group for a selection of one thousand of the linear models generated by the analysis. The models are ranked in order of RMSE for the combined BMI group data. It can be seen that the majority of the obese RMSE values (o) are greater than the normal BMI values (x). There lower differences in RMSE between the two BMI groups for overground walking, though the RMSE remains generally higher for the obese group (Figure 19). For treadmill walking in isolation there is a distinct difference in RMSE values between groups (Figure 20).

The brute-force search technique that was implemented in this study returned accuracy metrics for over one million linear speed estimation models. These one million results were also subdivided by BMI group and walking mode. This means that accuracy performance between BMI groups could be compared on a larger scale. A sample of 1,000 of the best performing linear models for overground walking in isolation, ranked in order of performance for the obese group, returns varying results for the normal BMI group (as shown in Figure 21); though the normal BMI group has generally higher RMSE values than the obese group for this subset of data. When a sample of 1,000 of the best performing linear models for overground walking are ranked according to the normal BMI group performance, the RMSE values of the obese group are distinctly higher (Figure 22). When these comparisons are made for treadmill walking as shown in Figure 23, the normal BMI group RMSE values are distributed relatively evenly above and below the line representing the best ranking obese group models. In contrast, the highest ranking models for the normal BMI group return distinctly lower RMSE values when applied to the obese group (Figure 24).

These results imply that there is an inherent difference between the accelerometer output collected from obese and normal individuals while performing walking activities. There are two explanations for this: either there is an intrinsic difference between the walking styles of obese and normal BMI individuals; or the accelerometer movement is affected by a factor, other than gait, which differs between the BMI groups, such as the increased adipose tissue at the accelerometer site for the obese group. The difference is particularly evident in treadmill walking for the obese group (Figure 24). The results imply that differences in gait between BMI groups are responsible for this difference in accuracy; the accelerometer placement is a constant between the two walking modes as it was not removed between walking trials, yet results differ according to walking mode, which suggests that the differences are not caused by an alternative effect on accelerometer movement. This also apparently implies there may be a difference in walking styles between treadmill and overground walking for one or both of the BMI groups. However, this conclusion cannot be made for certain as the observed differences may have been influenced by the different speeds that were performed between the obese and the normal BMI groups. This is discussed further in section 4.4.7.

When speeds were divided into two bands (slower speeds between 1.0ms^{-1} and 1.3ms^{-1}, and faster speeds between 1.4ms^{-1} and 1.7ms^{-1}) and analysed separately, the walking speed estimation accuracies improved (Table 19 and Table 20). MAPE values were well under 4% for nearly all models. The difference in estimation accuracy between BMI groups was also affected. For the slower walking band, model 3 returned good accuracies across BMI groups, though the obese group showed better results for treadmill walking than the normal BMI group (RMSE was 0.66ms^{-1} and 0.71ms^{-1} respectively), and the normal BMI group showed better accuracy for overground walking than the obese group (RMSE was 0.62ms^{-1} and 0.88ms^{-1} respectively). For the faster walking band, RMSE values were generally similar for all combinations of model, BMI group and walking mode, except for overground walking which showed a greater disparity between BMI group accuracy for models 1 and 2. For overground walking in the faster speed band, model 1 returned RMSE values of 0.086ms^{-1} and 0.047ms^{-1} for the obese group and normal group respectively. In contrast, model 3 achieved RMSE values of 0.064 ms^{-1} and 0.075 ms^{-1} for the obese group and normal group respectively for overground walking. The differing accuracies between BMI groups according to estimation model (yet for the same walking mode) suggest that there is a difference between BMI groups that is accounted for by one model and not be the other. Because the data was divided into speed bands, these differences between BMI groups are less likely to be influenced by the speeds performed by each BMI group.

4.4.6 Effect of Repeated Measures on Accuracy

As discussed in section 4.2.6 the use of repeated measures to test the speed prediction algorithm accuracy had the potential to artificially increase accuracy results. To test the extent that this was occurring, model 1 was applied to a reduced dataset containing only a single speed for each participant of the study. The results are shown in table 21. From the table it can be seen that there are a range of results, which is a consequence of a different dataset being generated for each test. The test dataset has been reduced nine-fold to only twenty-two speeds, as opposed to the original number of around 200 speeds, which makes the results more susceptible to fluctuations depending on which speeds are selected at random.

For the mixed BMI group the RMSE values ranged from around 0.06m/s to 1.3m/s, and the average RMSE returned for the ten trials was close to 0.09m/s for all three walking modes.

These figures compare well with the original results for model 1 which at best report around 0.07m/s, though at worst report only around 0.09m/s for the mixed group (table 21). The single speed test was consistent with the original test in that the obese group generally returned lower speed estimation accuracies than the normal BMI group, though the disparity was greater in the single speed test. The single speed results for the normal BMI group were generally equal or better than the original test results. Conversely, the obese group reported markedly poorer accuracy in the single speed test than in the original test.

The single speed test apparently reports inferior accuracies to the original test. This supports the idea that the repeated measures have artificially improved accuracy results, in this particular instance. However, because there were fewer measurements in the reduced speed dataset, this made the results sensitive to the random speed selection process, and therefore it may not be possible to draw this conclusion, and this may also explain some of the other differences in results. Overall, in spite of these differences, the accuracies returned by the single speed test still compare favourably with previous research and suggest that this speed prediction approach has potential for use in research.

4.4.7 Limitations

In the present study, walking speeds were not dictated by the study protocol, and this has meant that the speeds are different for the two BMI groups. A consistent walking speed is not easy to enforce over level ground, and attempts to do so may affect natural walking. Subjects interpretations of fast walking were on average over 7% slower in the obese group than the normal BMI group (the obese group mean speed was 1.97m/s compared with 2.11m/s for the normal group), and for treadmill walking obese participants were 9% slower at the fastest speed (the obese group mean speed was 1.71m/s compared with 1.86m/s for the normal group) with three of the eleven obese participants failing to exceed 1.6m/s. Although this may appear to be an inconsistency in the methodology, it is arguable that the recorded speeds reflect those that would occur naturally; that is, under free living conditions we might expect slower speeds for obese walkers than for their lower BMI counterparts. However, the difference in the range of speeds between BMI groups may be responsible for some of the apparent differences in walking speed accuracies.

The present study was conducted using two distinct groups of participants according to their BMI: an obese participant group with BMI of $30kgm^{-2}$ or more, and those considered to be in the normal BMI group between $18kgm^{-2}$ and $25kgm^{-2}$. The disparity in BMIs between the two groups was intended to highlight the effect of BMI on the speed estimation algorithms. However, this means that the class of individuals categorised as overweight (in the range 25-30kgm^{-2}) were not tested, and it is, therefore, not certain how the speed estimation algorithms would perform for this group. On the other hand, it is likely that because the chosen groups bookend the overweight group, the best performing estimation models will apply equally well.

Some previous studies have considered incline in addition to speed, though estimation of incline has been generally considered as a separate problem; in the case of both Aminian et al. (198) and Herren et al. (273) incline was estimated independently to speed using a separate neural network. The present study was concerned with walking on level ground only. It is, therefore, unclear how the speed estimation algorithm would be affected by incline. A shallow incline may not significantly affect speed estimation accuracy of a model which has been trained by level walking alone. However, an individual's gait will be greater affected by a steep incline, and consequently the speed estimation accuracy of the model developed in the present study is likely to be reduced in this case. Similarly, the surfaces used for developing the algorithm in the present study were relatively smooth and even. Uneven ground, such as a poorly maintained pavement, may detrimentally affect the accuracy of the speed estimation model, as may non-smooth surfaces such as grass. Further testing of the algorithm under such conditions is required, and this may necessitate additional training data. Though incline and surface type were not relevant to the particular research question posed in this chapter, it may be that obese individuals and individuals of normal BMI may be affected differently by these factors. In which case, further investigation may be indicated for future work.

The brute-force search approach to testing over one million different linear models was effective at identifying the best performing models. The process was, however, time consuming and demanding of computer resources. For this reason the approach was only applied to the combined walking mode data. With unlimited time and resources the approach may have been used against many subsets of the data; each combination of walking mode and obese group may have been tested in isolation, and for each of these combinations the data may also have been limited by speed band. These results may have more clearly shown which models are most effective across BMI groups and walking modes, or whether separate models are required depending on these criteria. The results returned by the present study were sufficient to answer the research question, and also significantly richer in information than previous research. Also, the brute-force search approach is limited by the number of features under consideration, as each feature increases the possible combinations by a factor of two, which may soon become unfeasible to process regardless of the computer resources available – though, in the present study, twenty features were deemed sufficient to answer the research question.

A limitation of any walking speed estimation algorithm is that it may only be applied to accelerometer data which represents walking; otherwise speed estimates will be erroneously returned for non-walking activities. Under free-living conditions individuals are able to perform any number of different activities. This necessitates a preliminary activity classification procedure, such as that described in chapter 3, which must identify instances of walking before speed prediction may be made. However, the same data collected by an accelerometer for use in walking speed estimation may also be used to identify walking. Which supports the use of accelerometry to measure multiple categories of PA under free-living.

4.4.7.1 Application of the Walking Speed Algorithm to Free-Living Data

As in the case of activity classification algorithms, the performance of walking speed estimation models may degrade significantly when applied to data collected under free-living conditions. There are many reasons for this such as the type of surface being walked upon, and the inclination of the ground, as discussed above. As the walking speed prediction models in chapter 4 were not derived from or tested for different surfaces and inclines, it is likely that the level of speed prediction accuracy achieved in the study would not be matched under free-living conditions.

To obtain an indicative measure of how the models may perform under naturalistic conditions the speed prediction algorithm was applied retrospectively to the outdoor walking data collected as part of the protocol described in chapter 3 – this is described in more detail, and results are presented, in Appendix D. Results were generated by applying the speed estimation algorithm to the outdoor walking data using three sets of training data separately as follows: both treadmill and overground (laboratory) walking data; treadmill data alone; overground (laboratory) walking data alone. There were sixteen participants who completed both the walking speed and classification study protocols (eleven obese and five normal BMI). The results were returned in the form of Bland-Altman plots.

From the results it appears that there are no data trends and the differences are consistent as the average of the speeds increases – this is the case for all three training scenarios – which implies the results do not change according to speed. For all three sets of results there is a slight bias (-0.05ms^{-1} for the treadmill trained model, -0.1ms^{-1} for the combined training model, and -0.12ms^{-1} for the laboratory walking trained model), which implies that the prediction model may be systematically overestimating walking speed to a certain extent, though this is less apparent in the treadmill trained model. Nearly all points on all three Bland-Altman plots lie well within 1.96 standard deviations of the mean, and most are within or close to one standard deviation away from the mean. Excluding a single outlier (clearly visible in each of the plots) all the differences between walking speed measurements lie between -0.31ms^{-1} and 0.13ms^{-1}. These last two points suggest that there is good agreement between the estimates and the measured speeds. This level of agreement between predicted walking speeds and measured walking speeds may or may not be suitable for applications in free-living, depending on the requirements of the researcher. However, it does imply that this walking speed estimation model has the potential to be applied to free-living.

The results and corresponding implications should not be regarded as conclusive, however, for a number of reasons. First, the walking data was obtained from a single route on campus, which is inadequate to test the effectiveness of the speed prediction model over varying terrain. Additionally, the subjects were accompanied by the researcher while walking, and being observed may have affected individuals' natural walking patterns. Only sixteen subjects were available for the analysis, which is not a sufficient number to adequately test the algorithm. Additionally, because the original protocol from which the data was obtained was not originally intended for use in testing the walking speed model, there may be insufficient precision in the timings of the walks (as discussed in Appendix D) which may have introduced errors in the measured speeds. Although the walking speed prediction model shows promise, it is likely that more extensive testing under free-living conditions will highlight the failings of the algorithm. Further research is thus required, and it is probable that the model will need modification to allow for varying walking surfaces and inclines.

4.5 Chapter Summary

The research question was asked: can a hip-mounted accelerometer be used to accurately estimate walking speed for an obese group?

In order to answer the research question, the following objectives were set:

- Objective 1: to produce a speed estimation model that is accurate across BMI groups and walking modes.
- Objective 2: to investigate how walking speed estimation accuracy differs between obese and normal BMI groups.
- Objective 3: to explore how walking speed estimation accuracy is affected by walking mode.

The present study investigated whether it was possible to develop a walking speed estimation model from hip-mounted accelerometer data which may be applied across BMI groups and deliver a level of accuracy that is comparable with previous research. Overground and treadmill walking data, for a number of walking speeds, was obtained from hip-mounted accelerometers worn by a mixed BMI subject group. Twenty-two accelerometer-based parameters were chosen as candidates for use in a linear walking speed prediction model. An exhaustive "brute-force" search technique was employed to evaluate over one million linear prediction models for speed estimation accuracy, based on the candidate parameters, using the entire dataset of mixed BMI and combined walking modes. The best three models were selected according to RMSE and MAPE and these were applied again to the dataset to obtain accuracies for both treadmill data and overground walking data in isolation. Results were broken down by prediction model, walking mode, and BMI group, so that multiple comparisons could be made.

The highest RMSE accuracy value achieved for the entire subject dataset was $0.085ms^{-1}$ and an overall MAPE value of below 5% was achieved (research question 1 and objective 1). Additionally, higher accuracies could be attained by limiting the range of speeds. Although speed prediction accuracies across BMI groups were comparable with previous research, differences in results between obese and normal BMI groups suggest that there is an intrinsic difference between BMI groups that may not be captured by a single speed prediction model (objective 2). These differences appear to be more pronounced depending on walking mode (objective 3). The results of the study suggest that a single estimation model is able to estimate walking speed from accelerometry with good accuracy for a mixed BMI group, but differences in results between BMI groups imply that a separate model for each group may improve accuracy.

5 Prediction of Energy Expenditure from Accelerometer Output

It is well established that physical activity (PA) confers many health benefits (274). Of particular interest, within the context of this thesis, is the influence of PA on maintaining weight and reducing obesity. Energy expenditure (EE) is a key measure of physical activity which may be used to evaluate outcomes in exercise interventions aimed at weight loss, and also to gauge compliance to such interventions. However, for large populations under free-living conditions, objective and accurate measurement of EE is inherently difficult.

There is a large body of experimental evidence demonstrating strong correlations between accelerometer output and energy expenditure (71, 116, 217), as discussed in detail in 2.5.1. Linear regression is widely used to model this relationship for the purposes of predicting energy expenditure. However, although previous research has often accounted for the weight of individuals when deriving EE estimation equations, it has little considered the effect of obesity on the effectiveness of these equations. Physiological and anthropometric attributes of individuals affect their rate of energy expenditure (237), as discussed in 2.5.2, and also affect movement such as walking style (119-123), as discussed in 2.2.1. This means that the amount of energy expended when performing like activities may vary between individuals according to these attributes, which would have a detrimental effect on EE estimation accuracy as a consequence. This may be particularly the case when the individuals differ greatly in BMI, as many of their individual attributes will be significantly different. EE prediction equations in common use have been derived using groups of individuals predominantly in the normal BMI range (71, 116), and these have not been validated against obese groups. Furthermore, there is no previous research which adequately investigates the effect of using the attributes of individuals as parameters to accelerometry-based energy expenditure prediction models, as discussed in 2.5.3.

The research question was asked: can energy expenditure prediction using accelerometry be improved by the addition of physiological and anthropometric measurements?

In order to answer the research question, the present study focused on estimating the energy expenditure of treadmill walking using accelerometer output and additional anthropometric and physiological attributes from a group of mixed BMI subjects. As discussed in 2.5.4, a single activity (walking) was chosen in order that the effects of these attributes on EE estimation was clear and not confounded by the differing relationships between accelerometer counts according to activity type.

5.1 Research Design

Hip-mounted accelerometer data and breath-by-breath respiration data were collected from fifty subjects who performed treadmill walking at four different speeds. In addition, several anthropometric and physiological measurements were obtained for each subject. Two EE prediction equations were formulated: one for kilocalories and one for METs. First, oxygen consumption data was converted to kilocalories using standard equations, and stepwise regression was applied in order to identify the best predictor variables of kilocalories from the accelerometer data and subject measurements. Stepwise regression is a systematic method of selecting the best set of variables from a number of candidates for inclusion in a model (275). The stepwise regression algorithm chooses an initial model, then variables are systematically added and removed based on their statistical significance (276). The EE prediction model identified by the stepwise regression was tested against the subject dataset using leave-one-out cross-validation. The process was repeated to generate and test a MET prediction model, after converting the oxygen consumption data to METs using standard equations. Simple prediction models comprising only accelerometer counts and weight as predictor variables – as used in previous research – were tested against the subject dataset using leave-one-out cross-validation. Results were compared between the simple and enhanced models in order to ascertain whether the additional parameters improved EE estimation.

5.2 Methods

5.2.1 Recruitment and Subject Statistics

The recruitment procedure was the same as that described in 3.2.1. It was necessary to test a subject group which was sufficiently varied in physiological and anthropometric attributes, particularly in the case of those attributes which vary with BMI. Fifty subjects took part in the study, twenty of which exceeded the threshold for obesity, having a BMI of 30kgm^{-2} or greater. Age, weight, height and BMI of the subject group are shown in Table 22.

		All Subjects (n=50)		Males (n=21)		Females (n=29)	
Age	(years)	34.6	(11.2)	32.8	(10.5)	35.9	(11.6)
Height	(cm)	168.6	(8.7)	174.6	(9.7)	164.3	(4.5)
Weight	(kg)	81.3	(16.7)	84.4	(16.4)	79.0	(16.7)
BMI	(kgm^{-2})	28.7	(6.2)	27.8	(5.6)	29.3	(6.5)

Table 22: Subject attributes. Figures shown are mean (standard deviation).

The fifty subjects taking part in this study were the same as those who participated in the study described in chapter 3, and data collection took part simultaneously.

5.2.2 Data Collection

In order to answer the research question several anthropometric and physiological measurements were chosen as candidate predictor variables for use alongside accelerometer counts in the EE prediction model. The choice of measurements was influenced by previous research which identified determinants of energy expenditure (as discussed in 2.5.2), or factors which may affect gait such as fat distribution (as discussed in 2.2.1). Ethnicity was not included in the analysis as it proved too challenging to recruit healthy obese participants in representative numbers across ethnic groups, and most respondents were predominantly white Caucasians. Lean mass and body fat, as identified by Weyer et al. (237), were obtained through bio-impedance. Though, where Weyer et al. found waist-to-thigh circumference ratio was a determinant of energy expenditure, the present study considers waist circumference and thigh fat thickness independently, as each may have individual effects on both gait and the energy expenditure used in walking. Measures of fat distribution including hip circumference and fat thicknesses were chosen due to their potential influence on gait, and consequently on the energy spent in walking. The full set of measurements is summarised in Table 23, and the measurements are described in more detail below.

Parameter	Obtained By
Age	Questionnaire
Gender	Questionnaire
Blood pressure: Systolic	Sphygmomanometer
Blood pressure: Diastolic	Sphygmomanometer
Height	Stadiometer
Weight	Electronic scales
Waist Circumference	Tape measure
Hip Circumference	Tape measure
Body Fat %	BodyStat 1500 body composition analyser
Basal metabolic rate (BMR)	BodyStat 1500 body composition analyser
Lean Mass	BodyStat 1500 body composition analyser
Forced Expiratory Volume after 1s (FEV1)	Spirometer
Forced Vital Capacity (FVC)	Spirometer
Peak Flow	Peak flow meter
Resting Heart Rate (bpm)	Heart rate monitor
Axilla fat thickness	Ultrasound
Thigh fat thickness	Ultrasound
Triceps fat thickness	Ultrasound
Abdomen fat thickness	Ultrasound
BMI	Calculated from weight and height
Resting VO_2	Metamax gas analysis equipment

Table 23: Physiological and anthropometric measurements

Blood pressure was taken using an electronic sphygmomanometer. Subjects sat quietly for five minutes before a measurement was made. This was repeated over two more five minute periods to help ensure confidence in the readings. Any subject showing an apparent systolic figure greater than 160mmHg or a diastolic figure greater than 100mmHg was not allowed to continue with the testing protocol, as these levels are indicative of high blood pressure and taking part in the exercises may have posed a risk to the subject's health. Blood pressure was chosen as a candidate parameter for the EE estimation equations due to its association with physical fitness. Also, a relatively recent study found that obese individuals with high blood pressure showed a 9% increase in resting metabolic rate (RMR) over obese individuals with normal blood pressure levels (277). A study found RMR to be a significant predictor of systolic blood pressure (278), and another found significant correlations between RMR and both systolic and diastolic blood pressure (279) – though these two studies suggested that RMR is the causal element of the relationship.

The age and gender of the subject was recorded. Age and gender are recognised as determinants of energy expenditure (234-237). Weight was measured in kilograms to one decimal place using electronic scales. Weight has already been established as a key variable used in EE estimation through accelerometry. Height was measured in centimetres by stadiometer. Height is used in the long established Harris Benedict equation for predicting RMR (280). A recent study with the aim of developing a multivariate EE prediction model found that height related to total energy expenditure (281). Also, a study used a multiple regression involving mass, height and age to predict RMR (282). Height also relates to gait characteristics such as stride length, which affect energy expenditure when walking (283); as stride lengths become shorter a higher cadence is required to maintain a particular walking speed.

A number of measurements relating to fat distribution were made. Waist circumference was measured using a tape measure at the navel. Hip circumference was measured using a tape measure. Fat thicknesses were measured at different body sites. These body sites were taken from Jackson and Pollock (284) who formulated an estimation equation for total body fat density in men using seven body site measurements as follows: chest, axilla (below armpit), triceps, subscapula (below shoulder blade), abdomen, supra-iliac (above hip), and thigh. In Jackson and Pollock's study, fat thicknesses were taken using skin-fold fat callipers. In the present study ultrasound measurements were used. These two methods of measurement will yield different results; however, we would expect them to correlate closely. The present study did not employ the total density calculation and instead considered the fat thicknesses as separate inputs to the energy expenditure estimation model. For the analysis, triceps, thigh, axilla, and abdomen were selected to represent body fat distribution at the arm, leg, upper torso and lower torso. This choice was based on the within session test-retest data which showed them to have the most reliable readings. It was felt that this subset would constitute an adequate representation of upper and lower body sites. Body fat distribution measurements were made as they bear a relationship to overall body fat (284), which is a determinant of energy expenditure (237). Additionally, increased load over the hips due to excess fat distribution may cause increased mediolateral sway (285) which may require greater EE when walking.

The body fat thicknesses described above were measured using the BodyMetrix (IntelaMetrix, Inc) hand-held ultrasound scanner, connected to a laptop via USB. Ultrasound measurements were taken by the author for all subjects except three female subjects who preferred to be measured by a female researcher. Ultrasound gel was applied to the scanner lens to help ensure a good contact between the scanner and the skin. The scanner was held gently against the skin at the perpendicular, and a button was pressed to activate the ultrasound scan. The scan records the ultrasound image over a period of around three seconds. A real-time representation of the ultrasound image is available on screen at the time of scanning via the proprietary software. Where tissue boundaries were indistinguishable due to noise in the image, the scan was repeated. Ultrasound images were saved to the laptop hard drive for later measurement using the software. Divisions between types of tissue are visible in the ultrasound image as white bands. In the example image (Figure 25) the division between fat and muscle can clearly be seen. The uppermost reading of the white band was taken as the fat thickness – in the case of this example, the fat thickness is 10mm.

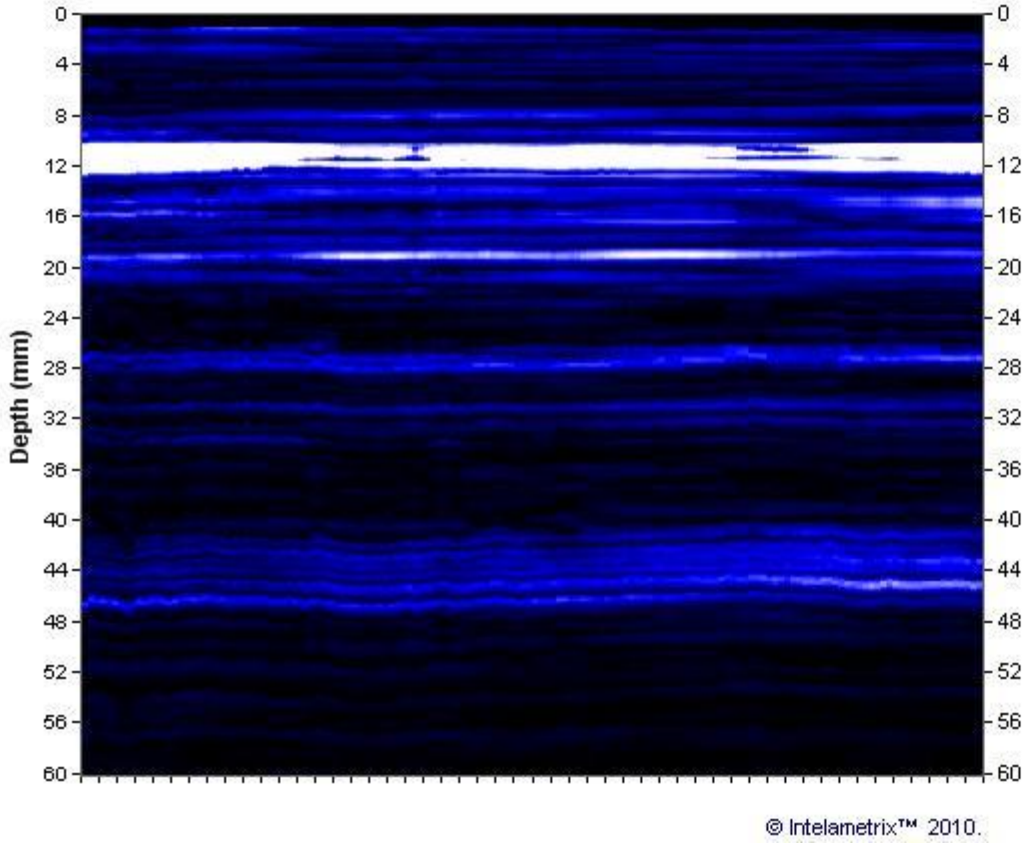

Figure 25: Example ultrasound scan of triceps from one of the study particpants. The thick white band represents the division between fat and muscle.

Several measurements related to body composition were returned by the Bodystat 1500 body composition analyser (BodyStat Ltd, Douglas, Isle of Man, British Isles). This instrument uses bio-impedance to calculate body fat percentage, dry lean weight, lean mass (fat free mass), water percentage, and basal metabolic rate (BMR). Participants were required to lie down, and electrodes were fitted at the right hand and right foot while the measurements were taken.

Chen and Sun had considered residual lung volume as an additional parameter to improve energy expenditure prediction from accelerometry (114) which motivated the inclusion of measures of lung capacity in the present study. Also, pathological lung function has been shown to increase RMR (286) which suggests that lung function in healthy subjects may also have some bearing on EE. In the present study, it was not possible to implement any of the techniques required to measure residual lung volume. Instead two other measures related to lung volume were made. Forced expiratory volume after one second (FEV1) and forced vital capacity (FVC) for each subject were obtained from the single use of a spirometer, for which the subject was required to exhale into the spirometer for a period of around six seconds.

Peak flow was also recorded as an alternative measure of lung function. Although peak flow does not measure lung volume, this is more easily measured using a hand held peak flow meter, and would therefore be preferable should it improve the EE estimation model to a similar extent to FEV1 or FVC. Subjects wore a heart rate monitor throughout testing so that measurements could be made when required, and also so that heart rates could be monitored for safety as the exercising grew more vigorous. Subjects sat quietly for five minutes before their resting heart rate was recorded. Resting heart rate may be used as an indicator of fitness (287) which was the motivation for its inclusion in the candidate predictor variables, as aerobic fitness has been shown to influence BMR (288).

The physiological and anthropometric measurements that were chosen due to their relationship with energy expenditure in individuals, as described above, in some cases may be causal (as in the case of lean mass which has significant influence on BMR (289)) or may merely correlate (as in the case of age, for instance, as lean mass declines with age (290)). Also, some physiological and anthropometric attributes – individually or in combination – may have an indirect relationship with energy expenditure, but may be easily measured and contribute to the estimation model (as mentioned, resting heart rate itself may not affect energy expenditure, but low heart rates suggest better physical fitness, which in turn may suggest lower fat mass). These relationships may not be apparent until the analysis has been carried out and the model derived. The majority of the chosen attributes have been shown to bear some relationship with resting energy expenditure or BMR (height, weight, fat mass, and lean mass (291)). However, there are fewer resources in the literature that describe the effect of individual attributes on energy expenditure due to physical activity itself (beyond evident relationships such as individuals with higher mass requiring more energy to move their own weight). Some measurements have been chosen due to their use in other studies that have attempted to improve an EE estimation model using accelerometry. The remaining measurements have been concerned with fat distribution which may affect energy expenditure when performing activities – for example, higher thigh fat thickness indicates heavier legs, which require more effort to move when walking, and also may affect gait economy.

The Actigraph GT3X+ was synchronised with the computer, and the offset between the wristwatch time and the computer time was noted, so that hand written times could be later synchronised with accelerometer output. A single Actigraph GT3X+ monitor was affixed to the participant's right hip above the iliac crest (Figure 26). The device was set to sample at

50Hz in order to capture movements of up to 25Hz according to the Nyquist Sampling Theorem as discussed in 3.2.2.1.

Figure 26: hip-mounted Actigraph GT3X+ accelerometer.

The Metamax 3B gas analyser (CORTEX Biophysik GmbH of Leipzig, Germany) was employed to measure breath-by-breath data. Software (MetaSoft version 3.9.7) is provided by the manufacturer to manage the collection, storage and analysis of the data. The analyser was calibrated according to the manufacturuer's instruction manual for each session of testing. The analyser was switched on for at least 30 minutes until operating temperature was reached. The Metasoft software calibration procedure requires barometric pressure readings; these were obtained from a digital barometer (Technoline Ltd, type WS-9032IT). The analyser is calibrated first using ambient air (20.93% O2 and 0.03% CO2) and then with gas from a cylinder (16.48% O2 and 4.98% CO2). The volume transducer is calibrated with a 3-Litre Hans Rudolph Series 5530 syringe; five good strokes are required for this, as determined by the software.

A gas collection mask was fitted to the participant and tested for leakage. Where necessary, petroleum jelly was used to stop air leakage coming from gaps in the mask where the fit was not exact. The gas analyser is controlled using the Metasoft software provided by the manufacturer. A calibration procedure was first initiated where the software records ambient levels of gas in the atmosphere. The mask was then connected to the gas analyser, and the Metasoft software was started and began recording gas exchange data for the participant. The participant sat quietly for five minutes while baseline oxygen consumption was measured (resting VO_2). Resting VO_2 is related to BMR which was identified as a determinant of energy expenditure.

Treadmill walking was performed at four different speeds for five minutes each. Five minutes was required to allow the participant time for their oxygen consumption to stabilise. The chosen speeds were determined by the fitness of the participant and their preferred normal walking speed, as established in a preliminary timed walk over twenty metres on level ground. Participants were not permitted to use the treadmill hand rail. Participants were not allowed to speak while gas analysis was taking place, except for safety reasons, but were able to respond to questions using head movements (nods and shakes) or hand movements (thumbs up or down). After each test speed the participant was asked whether a rest was required. Most participants declined this offer and continued directly with the next test speed. Participants were periodically informed of the elapsed time for each walking test. They were also asked whether they felt able to complete the walking as the speeds increased. Subsequent speeds were lowered for those apparently suffering fatigue at slower speeds. Treadmill speeds were hand annotated and time-stamped according to a digital wristwatch.

5.2.3 Walking Speed Statistics

A histogram of the speeds performed by the subjects is shown in Figure 27. Speeds generally started in the "slow" range at around 1.1ms^{-1} and increased with each five minute test to around 1.6ms^{-1}; though, speeds were as low as 0.8ms^{-1} and 0.9ms^{-1} and as high as 1.7ms^{-1} and 1.8ms^{-1} in some cases.

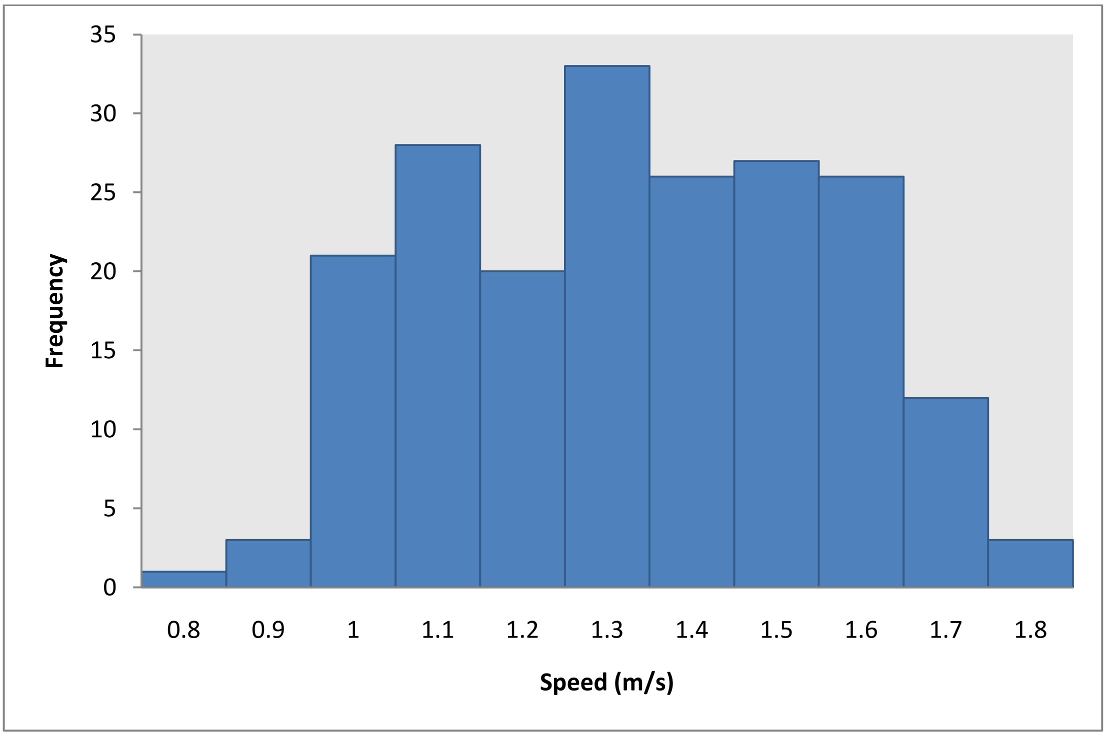

Figure 27: Histogram of collected speeds.

5.2.4 Data Analysis

5.2.4.1 Data Processing

Accelerometer data was downloaded, processed, and imported into MATLAB format, in the manner described in 3.2.3.1. Gas analysis data was exported to CSV file format using the Metasoft software. The software can provide the raw oxygen consumption (VO_2) and carbon dioxide production (VCO_2) data for each breath in litres per minute, but these breaths are not regularly spaced, and are of differing durations, making synchronisation with accelerometer data more difficult. The software gives the option to choose an epoch over which the gas data is averaged. A proprietary unpublished algorithm is used by the software to average the breath-by-breath data into epochs. To more easily synchronise accelerometer output with oxygen data, one second epochs were chosen. Except for the first few seconds of data collection, values averaged in one second epochs were found to align with one minute epoch and breath-by-breath data. The gas data CSV file was imported into MATLAB for processing using standard file reading techniques. The offset between the time on the computer running the Metasoft software and the wristwatch had been previously noted, so timestamps were adjusted accordingly. Oxygen data was converted to millilitres per minute as part of the import process.

Accelerometer output was aligned with gas data according to the timestamps in each dataset. The oxygen data, the accelerometer data, and the labels for each five minute bout of treadmill walking were plotted and visually inspected to check alignment (Figure 28). Timestamps were adjusted where necessary. One minute of oxygen data and the corresponding accelerometer output was extracted for each of the four speeds. Although the accelerometer signal characteristics remain similar within each five minute treadmill walking trial, the amount of oxygen being consumed by the participant does not represent the energy being spent until steady state is reached. For this reason one minute of data was chosen close to the end of each trial. For each speed the minute preceding the final fifteen seconds was used – that is, from around the 3.75 minute mark to the 4.75 minute mark – and the remaining data was discarded. Corresponding accelerometer counts were calculated for this same one minute time period using the sum of the rectified filtered signal, as described in 2.5.1.

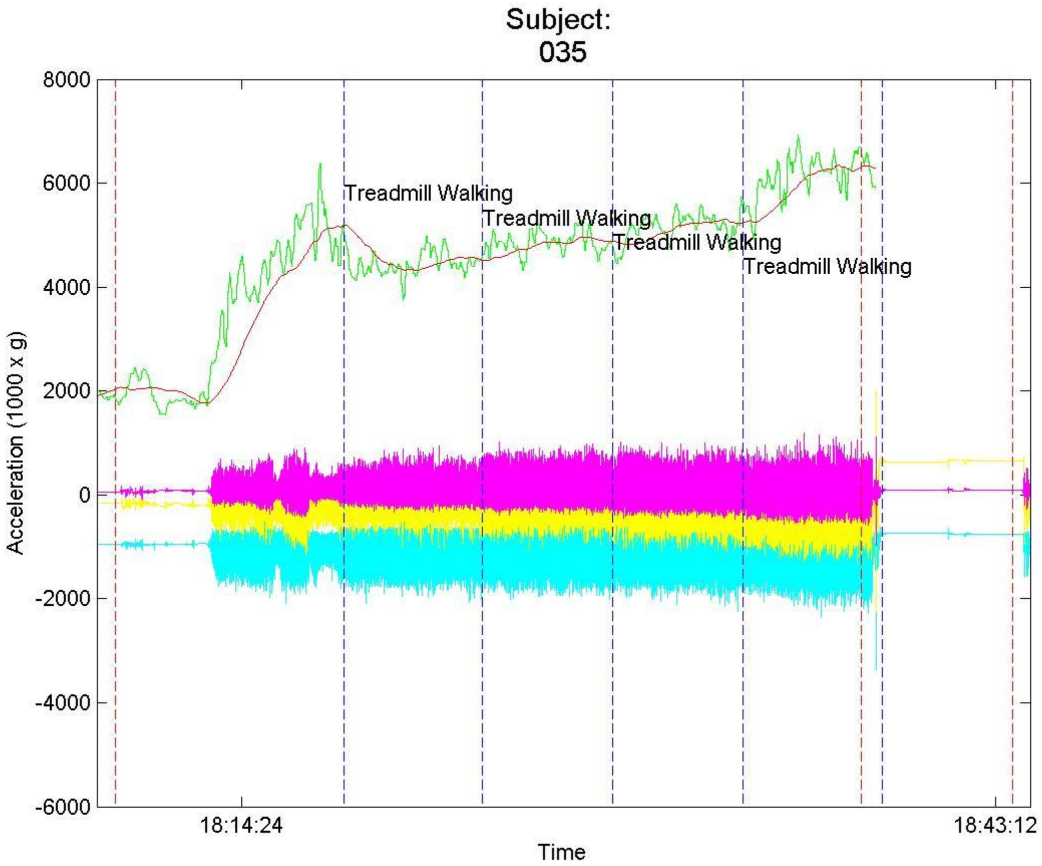

Figure 28: Visualisation of the accelerometer output and the corresponding oxygen consumption data taken from one of the study participants. Oxygen consumption data is shown in green. Moving average of oxygen consumption is in red. Magenta, yellow and cyan represent the accelerations for the three accelerometer axes. It can be seen that the oxygen data aligns with the four increasing speeds of treadmill walking (demarcated by the dotted vertical lines).

5.2.4.2 Energy Expenditure Units

Energy expenditure prediction equations are usually formulated to predict either METs or kilocalories, as discussed in 2.5.1. One MET, or metabolic equivalent, is equivalent to the resting metabolic rate (RMR) which is generally assumed to be 3.5mLkg^{-1}min^{-1}(224). METs approximated this way are known as standard METs. Related research predominantly uses this approach rather than using an individual's measured RMR. The majority of previous research has aimed to model the relationship between METs and accelerometer output (73, 104, 244), though kilocalories have also been considered (71). The method stated in the Compendium of Physical Activities to convert oxygen consumption values (in litres per minute) to kilocalories is to multiply by five (292). However the Weir equation (223) may provide a more accurate energy expenditure value ($kg.cal = 3.9 \cdot VO_2\,(\mathrm{L/min}) + 1.1 \cdot CO_2\,(\mathrm{L/min})$) and was thus used to convert respiratory readings to kilocalories. To convert from VO_2 to METs the absolute oxygen consumption recorded by the gas analysis equipment is first divided by the subject weight to derive relative oxygen consumption then this result is divided by 3.5 to obtain standard METs (292). EE prediction models based on MET values should not contain a weight parameter as this has been already factored into the dependent variable, whereas in a kilocalorie prediction model parameters may or may not include weight. This is a fundamental difference between the two types of model. The present study analysed models based on both types of unit in order to identify the best overall prediction model.

5.2.4.3 Basic Linear Energy Expenditure Prediction Models

In order to gauge the extent by which EE estimation accuracy was improved by using an enhanced EE prediction model, a number of basic linear models were required to provide comparative EE prediction accuracy levels. There were two main categories of basic model: MET and kilocalorie. For each of these categories, a model was chosen based on both a single vertical accelerometer approach, as in studies such as Freedson et al. (71), and a three dimensional accelerometer approach as in studies such as Bouten et al. (116), resulting in four linear models in total. These models are listed in Table 24 below.

To help further explain the rationale for the choice of these models, the general formulae for the MET and kilocalorie linear prediction models are reproduced below (from section 2.5.1).

> The linear model for predicting METs from accelerometer counts can be expressed mathematically as follows:
>
> $METs = a + bK$ [1]
>
> Where a and b are constants, and K represents accelerometer counts per minute. The constants a and b are obtained through a linear regression between accelerometer counts and a measure of EE expressed in METs.
>
> Similarly, the linear model for prediction kilocalories from accelerometer counts is as follows:
>
> $kcal/min = a + bK +cW$ [2]
>
> Where a, b and c are constants, K represents accelerometer counts per minute and W represents body weight. The constants a,b and c are obtained through multiple linear regression.

The first linear model (model M1 in Table 24) predicts MET values and corresponds with equation [1]. In equation [1] the K term is equivalent to the CountsX (vertical) parameter, and the constants a and b are calculated by multiple regression when processing the estimation results.

Model M2 is a variation on equation [1]; though, instead of a single value representing accelerometer counts, three axes are used separately. The equation thus becomes:

$$METs = a + b_xK_x + b_yK_y + b_zK_z \qquad [3]$$

where K_x, K_y, and K_z represent counts for the three accelerometer axes (these are CountsX, CountsY, and CountsZ), and the constants a, b_x, b_y, and b_z are obtained through multiple regression.

Model KC1 is the basic linear kilocalorie prediction model corresponding to equation [2] using the vertical accelerometer axis and weight alone. Model KC2 is a variation on this using three accelerometer axes. The format for this is:

$$kcal/min = a + b_xK_x + b_yK_y + b_zK_z + cW \qquad [4]$$

where K_x, K_y, and K_z represent counts for the three accelerometer axes (CountsX, CountsY, and CountsZ), W is weight, and the constants a, b_x, b_y, b_z and c are obtained through multiple regression.

Vector magnitude (VM) has been used as an alternative to traditional accelerometer counts. However, an additional baseline model using VM was considered unnecessary, as although a VM model may or may not improve prediction accuracy, it would not contribute to answering the research question. Additionally, a study has reported that no significant improvement was gained by using VM over vertical counts for this type of model (293).

Model	Parameters	Description	Applies to
M1	Counts X (vertical)	Vertical counts only	METs
M2	Counts X (vertical) / Counts Y (anteroposterior) / Counts Z (mediolateral)	Counts for all axes	METs
KC1	Counts X (vertical) / Weight	Vertical counts plus weight	kcal
KC2	Counts X (vertical) / Counts Y (anteroposterior) / Counts Z (mediolateral) / Weight	Counts for all axes plus weight	kcal

Table 24: basic linear models for use in comparison with the multiple parameter models

5.2.4.4 *Deriving and Testing the Enhanced Energy Expenditure Prediction Equations*

There were fifty subjects who each performed four treadmill walking trials, resulting in two hundred data records in total. For each data record, accelerometer counts were calculated for each accelerometer axis, and measured VO_2 was converted to METs and kilocalories. For the kilocalorie model, accelerometer counts (for the vertical, anteroposterior and mediolateral axes) and the twenty-one subject-specific measurements (Table 23) were considered as the candidate independent variables, and kilocalories represented the dependent variable in a multiple linear regression model. This resulted in a matrix of two hundred rows and twenty-four columns representing the dataset of independent variables, and a vector of two hundred corresponding kilocalorie values. To reduce the number of variables in the model, stepwise regression was applied, as in the study by Chen and Sun (114). The process was repeated using METs as the dependent variable, though there was one fewer independent variable used, as weight had already been used to convert VO_2 to MET values and was, therefore, already intrinsically part of the model. The MET and kilocalorie models derived through this process became models M_{SW} and KC_{SW} respectively (Table 25). Adjusted R^2 values were calculated for each model at this point for the purposes of comparison with previous research.

The two models (M_{SW} and KC_{SW}) were independently tested using leave-one-out cross-validation, and evaluated using RMSE. The way this was applied was as follows. Each of the fifty subjects was considered independently. Using only the independent variables identified by the stepwise regression, multiple linear regression was applied to the data from the remaining forty-nine subjects in order to formulate an EE prediction equation. A linear equation of the following form was returned:

$$EE = c_0 + c_1V_1 + c_2V_2 + \ldots + c_iV_i \qquad [5]$$

Where c_i represents the coefficient of the independent variable (V_i) returned by multiple linear regression, and *i* represents the number of independent variables in the model. EE is measured in METs or kilocalories depending on which is being considered. This equation was applied to the data from the single subject not used to formulate the equation in order to produce an EE estimate for each of their four treadmill trials. The difference between the estimated EE and the known actual EE value (that is, the error) was stored for each of the subject's four treadmill trials. The process was repeated for each subject, until a vector of two hundred error values had been generated. These values were each squared, the mean of the squares was calculated, then the square root was applied, according to the standard formula for RMSE. Results are shown for MET and kilocalorie models in Table 27 and Table 28 respectively. A similar process was applied to the four basic linear models (Table 24) in order to provide results for comparison with the stepwise regression model results.

To allow a greater insight into the effect of adding each parameter to the EE prediction models generated by stepwise regression, a procedure was executed to calculate the RMSE returned by the model at each step of adding a new parameter. This was done following the order that the parameters were selected by the stepwise regression. To clarify, a model that was generated by stepwise regression in the following order is considered: Parameter 1, Parameter 2, Parameter 3, …, Parameter N. The evaluation procedure first considers Parameter 1 alone and applies leave-one-out cross validation to obtain the RMSE for the single parameter model. The next step is to calculate the RMSE for a model containing Parameter 1 and Parameter 2. This is followed by evaluating the RMSE for Parameter 1, Parameter 2 and Parameter 3. This process continues until all parameters are exhausted. There are therefore N results for RMSE corresponding to the N steps taken, where the RMSE of each step *S* represents the accuracy of the cumulative addition of Parameter 1 to Parameter *S*. This procedure was repeated for MET and kilocalorie models. From these results it may become apparent at which point the accuracy gained by an additional parameter is outweighed by the effort of obtaining the parameter through subject measurement.

5.2.4.5 Energy Expenditure due to Physical Activity

The analysis described above focuses on an energy expenditure estimation model which returns an estimate for the total energy expenditure (TEE) of an individual undertaking physical activity (in this case, walking). This type of model may be useful, for example, when comparing total calorie intake with total energy expenditure. However, there are some matters regarding the model that should be considered. As discussed in 2.5.2, there are three main components of energy expenditure in humans: resting energy expenditure (REE), which is synonymous with the basal metabolic rate (BMR); the thermal effect of food (TEF); and physical activity energy expenditure (PAEE). TEE is the sum of REE, TEF and PAEE. However, a model comprised of accelerometer features alone will only be able to estimate the PAEE component of energy expenditure. A number of combined anthropometric and physiological attributes used in the above model may provide a proxy measure of BMR; the estimates generated by the linear regression equation may be offset according to these attributes in such a manner which approximates for BMR. However, from these equations, it is not clear which parameters maybe approximating BMR and which are helping explain further variance in the model for PAEE estimation. For this reason a secondary analysis was applied where REE was removed from the EE measurements before the model was derived. This model, therefore, provides estimation for the PAEE component alone. The analysis was performed in the same way as described in 5.2.4.4.

5.2.4.6 A Consideration of Repeated Measures

Energy expenditure and accelerometer output data were collected at four different speeds per subject. As in the case of the walking speed prediction analysis (discussed in 4.2.6) repeated measures in the test data may artificially boost EE estimation accuracy. For this reason, a secondary test was carried out, where a single speed per subject was selected at random for use in testing the EE estimation algorithm. This second test was performed ten times to allow some insight into how estimation accuracy changes when not using repeated measures. Each test was performed independently for each model.

5.3 Results

Results were generated for the estimation of TEE in terms of both METs and kilocalories. MET and kilocalorie estimation results were generated separately for PAEE. These results are presented below.

5.3.1 Multiple Parameter Models

5.3.1.1 TEE Models

The linear models derived for MET prediction of TEE (Model M_{SW}) and kilocalorie prediction of TEE (Model KC_{SW}) through stepwise regression are shown in Table 25. The EE prediction models using these parameters follow the format of equation [5] in 5.2.4.4 after multiple regression has been applied to the parameters during the leave-one-out cross-validation process.

Model	Parameters	Description	Applies to
M_{SW}	Counts X (vertical) / Triceps fat thickness / Waist Circumference / Counts Z (mediolateral) / Blood Pressure: diastolic / Resting VO2 / BMI	Stepwise regression model for METs	METs
KC_{SW}	Counts Z (mediolateral) / Counts X (vertical) / BMI / Triceps fat thickness / Blood Pressure: systolic / Waist Circumference / Weight / Resting Heart Rate / Resting VO2	Stepwise regression model for kcal	kcal

Table 25: Multiple parameter models for TEE estimation derived through stepwise regression.

5.3.1.2 PAEE Models

The linear models derived for MET prediction of PAEE (Model MPA_{SW}) and kilocalorie prediction of PAEE (Model $KCPA_{SW}$) through stepwise regression are shown in Table 26.

Model	Parameters	Description	Applies to
MPA_{SW}	Counts X (vertical) / Triceps fat thickness / Blood Pressure: diastolic / Counts Y (anteroposterior) / Age	Stepwise regression model for METs	METs
$KCPA_{SW}$	Counts Z (mediolateral) / Lean Mass / Counts X (vertical) / Body Fat % / Waist Circumference / Blood Pressure: diastolic / Triceps fat thickness / BMI / Peak Flow	Stepwise regression model for kcal for PAEE	kcal

Table 26: Multiple parameter models for PAEE estimation derived through stepwise regression.

5.3.2 Results of Applying the Energy Expenditure Estimation Models

5.3.2.1 TEE Estimation Results

Table 27 below compares the accuracy of the three TEE MET prediction models (models M1, M2 and M_{SW}). Results are shown in order of accuracy according to RMSE. Adjusted R^2 values are shown to allow comparison with the findings of previous research. It can be seen that the enhanced model (M_{SW}) shows an improvement over the basic linear models (M1 and M2).

Model	Adjusted R^2	RMSE (METs)
M_{SW}	0.748	0.427
M1	0.550	0.530
M2	0.557	0.537

Table 27: Energy expenditure estimation results for MET models.

Table 28 presents the accuracies of the three TEE kilocalorie prediction models (models KC1, KC2 and KC_{SW}). Again, RMSE and adjusted R^2 are included. Also, the RMSE returned in kilocalories has been converted to METs using the mean weight of the subject group to allow a comparison between MET and kilocalorie model accuracies. Again, it can be seen that the enhanced model returns higher accuracies over the basic linear models.

Model	Adjusted R^2	RMSE (kcal/min)	RMSE (converted to METs)
KC_{SW}	0.800	0.695	0.489
KC1	0.711	0.762	0.536
KC2	0.710	0.786	0.553

Table 28: Energy expenditure estimation results for TEE kilocalorie models. MET values have been estimated for comparison with MET models using the mean weight of the subject group.

Table 29 shows how RMSE improves with the cumulative addition of the parameters identified by the stepwise regression for the MET model. The process by which these results were obtained is described in 5.2.4.3. Each step shows the RMSE values returned for the combined parameters from all steps up to that point. The table also shows the percentage improvement in RMSE accuracy with the addition of each parameter, and the cumulative percentage improvement over the initial parameter alone. The corresponding results are shown for the kilocalorie model in Table 30. The steps follow the order that the parameters were added to the model by the stepwise regression. It can be seen that there is no great improvement in accuracy beyond step 3 in the MET estimation model (Table 29) or beyond step 4 in the kilocalorie estimation model (Table 30).

Step	Parameter	RMSE (METs)	Incremental % Improvement	Cumulative % Improvement
1	Counts X (vertical)	0.530	N/A	N/A
2	+ Triceps fat thickness	0.467	11.9%	11.9%
3	+ Waist Circumference	0.445	4.7%	16.0%
4	+ Counts Z (mediolateral)	0.440	1.1%	17.0%
5	+ Blood Pressure: diastolic	0.437	0.7%	17.5%
6	+ Resting VO2	0.426	2.5%	19.6%

Table 29: The incremental and cumulative improvement in RMSE with the addition of each parameter from the TEE MET model (Model M_{SW})

Step	Parameter	RMSE (kcal/ min)	Incremental % Improvement	Cumulative % Improvement
1	Counts Z (mediolateral)	1.526	N/A	N/A
2	+ Counts X (vertical)	0.989	6.1%	6.1%
3	+ BMI	0.942	4.7%	10.5%
4	+ Triceps fat thickness	0.838	11.0%	20.4%
5	+ Blood Pressure: systolic	0.812	3.1%	22.8%
6	+ Waist Circumference	0.833	-2.5%	20.9%
7	+ Weight	0.706	15.2%	32.9%
8	+ Resting Heart Rate	0.706	0.0%	32.9%
9	+ Resting VO2	0.695	1.5%	33.9%

Table 30: The incremental and cumulative improvement in RMSE with the addition of each parameter from the TEE kilocalorie model (Model KC_{SW})

5.3.2.2 PAEE Estimation Results

Table 31 below compares the accuracy of the three PAEE MET prediction models (models MPA1, MPA2 and MPA_{SW}). Results are shown in order of accuracy according to RMSE. Adjusted R^2 values are shown to allow comparison with the findings of previous research. As in the TEE model, it can be seen that the enhanced model (MPA_{SW}) shows a marginal improvement over the basic linear models (MPA1 and MPA2).

Model	Adjusted R^2	RMSE (METs)
MPA_{SW}	0.721	0.418
MPA1	0.633	0.459
MPA2	0.631	0.470

Table 31: Energy expenditure estimation results for PAEE MET models.

Table 32 presents the accuracies of the three PAEE kilocalorie prediction models (models KCPA1, KCPA2 and $KCPA_{SW}$). Again, RMSE and adjusted R^2 are included. As in the TEE results the RMSE returned in kilocalories has been converted to METs using the mean weight of the subject group to allow a comparison between MET and kilocalorie model accuracies. It can be seen that the enhanced model returns marginally higher accuracies over the basic linear models.

Model	Adjusted R^2	RMSE (kcal/min)	RMSE (converted to METs)
$KCPA_{SW}$	0.783	0.687	0.483
KCPA1	0. 717	0.713	0.501
KCPA2	0.718	0.731	0.514

Table 32: Energy expenditure estimation results for PAEE kilocalorie models. MET values have been estimated for comparison with MET models using the mean weight of the subject group.

Table 33 shows how RMSE improves with the cumulative addition of the parameters identified by the stepwise regression for the PAEE MET model, as was performed for the TEE model. The corresponding results are shown for the kilocalorie model in Table 34. The steps follow the order that the parameters were added to the model by the stepwise regression. It can be seen that there is no improvement in accuracy with the addition of age in the PAEE MET estimation model (Table 33) or beyond step 4 in the PAEE kilocalorie estimation model (Table 34).

Step	Parameter	RMSE (METs)	Incremental % Improvement	Cumulative % Improvement
1	Counts X (vertical)	0.459	N/A	N/A
2	+ Triceps fat thickness	0.440	4.1%	4.1%
3	+ Blood Pressure: diastolic	0.430	2.3%	6.3%
4	+ Counts Y (anteroposterior)	0.416	3.3%	9.4%
5	+ Age	0.418	-0.5%	8.9%

Table 33: The incremental and cumulative improvement in RMSE with the addition of each parameter from the PAEE MET model (Model MPA_{SW})

Step	Parameter	RMSE (kcal/min)	Incremental % Improvement	Cumulative % Improvement
1	Counts Z (mediolateral)	0.981	0.0%	0.0%
2	+ Lean Mass	0.898	8.5%	8.5%
3	+ Counts X (vertical)	0.765	14.8%	22.0%
4	+ Body Fat %	0.708	7.5%	27.8%
5	+ Waist Circumference	0.71	-0.3%	27.6%
6	+ Blood Pressure: diastolic	0.698	1.7%	28.8%
7	+ Triceps fat thickness	0.693	0.7%	29.4%
8	+ BMI	0.685	1.2%	30.2%
9	+ Peak Flow	0.687	-0.3%	30.0%

Table 34: The incremental and cumulative improvement in RMSE with the addition of each parameter from the PAEE kilocalorie model (Model $KCPA_{SW}$)

5.3.2.3 Non-repeated Measures Results

A single speed was tested for each subject by applying a random speed selection process to the data. Table 35 shows the results of testing the EE estimation algorithm against this data. The average adjusted R^2 and RMSE values are also shown.

Model:	M_{SW}		KC_{SW}		MPA_{SW}		$KCPA_{SW}$	
Rank by RMSE:	**Adj R^2**	**RMSE (METs)**	**Adj R^2**	**RMSE (kcal/min)**	**Adj R^2**	**RMSE (METs)**	**Adj R^2**	**RMSE (kcal/min)**
1	0.777	0.382	0.874	0.462	0.709	0.364	0.800	0.606
2	0.698	0.393	0.848	0.577	0.674	0.385	0.751	0.681
3	0.693	0.441	0.814	0.640	0.672	0.392	0.770	0.739
4	0.705	0.449	0.843	0.693	0.528	0.395	0.749	0.747
5	0.729	0.463	0.784	0.767	0.764	0.397	0.740	0.758
6	0.729	0.467	0.794	0.780	0.708	0.403	0.644	0.762
7	0.758	0.471	0.839	0.803	0.711	0.403	0.829	0.762
8	0.659	0.517	0.832	0.832	0.761	0.426	0.801	0.794
9	0.694	0.526	0.751	0.833	0.693	0.446	0.778	0.817
10	0.608	0.545	0.743	0.847	0.674	0.472	0.774	0.825
Mean:	0.705	0.465	0.812	0.724	0.689	0.408	0.764	0.749
Original:	0.721	0.427	0.800	0.695	0.721	0.418	0.783	0.687

Table 35: Results of applying the EE estimation to single speeds for each subject. The test was performed ten times for each of the four stepwise models. Results are shown in order of RMSE by model. The original results obtained through the repeated measures test are shown for comparison.

5.4 Discussion

The study aimed to answer the research question: can energy expenditure estimation using accelerometry be improved by the addition of anthropometric and physiological parameters?

In order to answer the research question fifty subjects each wore hip-mounted accelerometers and walked at four different speeds on a treadmill, while simultaneous oxygen consumption data were recorded via a face mask and gas analysis equipment. A number of anthropometric and physiological measurements were taken for each subject. Stepwise regression was applied to the accelerometer data and subject measurements in order to identify which variables most significantly contribute to EE estimation. These variables were used as input parameters to EE linear prediction models which were tested against the subject data using leave-one-out cross-validation. Both total energy expenditure (TEE) and physical activity energy expenditure (PAEE) was considered for both MET and kilocalorie estimation, resulting in four EE estimation models. For the TEE MET model (M_{sw}), energy expenditure estimation was improved by nearly twenty percent over models which considered only accelerometer counts and weight, and the TEE kilocalorie model (KC_{sw}) showed an improvement of around nine percent. These results suggest that EE prediction models can be improved by the addition of subject-specific anthropometric and physiological attributes, though these improvements were more modest for the PAEE MET (MPA_{sw}) and PAEE kilocalorie ($KCPA_{sw}$) models (around nine percent and four percent respectively).

Two different approaches to predicting EE per unit time were formulated. The first aimed to estimate METs (which is a measure of EE normalised by body mass) and the second aimed to estimate kilocalories which is an absolute measure of EE. To compare the two sets of results it is necessary to convert both to the same unit. To convert kilocalories to METs, the value must be multiplied by 200 and divided by 3.5 times body weight in kilograms (292). The kilocalorie results, however, were obtained from the entire subject group comprising many different body weights. In order to give an indicative figure, the mean subject body weight was used for the conversion. The MET prediction models return a marginally better RMSE (0.427 for TEE, 0.418 for PAEE) than the kilocalorie models when converted to METs (0.489 for TEE, 0.483 for PAEE). However, the conversion from kilocalories to METs was not exact, and the difference in model performance is not great. The researcher may make their own decision as to whether the kilocalorie or MET model is more appropriate to the research and expect similar levels of accuracy.

5.4.1 Parameters Selected by the Stepwise Regression

The vertical accelerometer counts parameter was selected by the stepwise regressions for both MET and kilocalorie models, for both TEE and PAEE. For the TEE kilocalorie model, BMI was chosen by the stepwise regression as the most significant parameter other than accelerometer counts, although weight was expected to fill this position. The selection of BMI over weight appears to be a consequence of applying the Weir equation (223) to calculate EE from both VO_2 and CO_2; preliminary testing of the MATLAB code had shown weight to be the second most significant parameter in the model when kilocalories were calculated using VO_2 alone. In terms of the MET models, weight was already factored in to the units of measurement of the independent variable, so weight related measurements are not expected to feature highly in the model. However, for the PA kilocalorie model, lean mass was selected instead of weight. It may be that lean mass in combination with parameters such as BMI and body fat percentage (as chosen by the model) is a better variable for the PAEE prediction model than weight in this particular case.

The triceps fat thickness is common to both the MET and kilocalorie models that were identified by stepwise regression, and was considered the most significant parameter other than accelerometer counts and weight in the TEE and PAEE MET models. It may be that the triceps fat thickness is a good representative quantity of upper body fat distribution. Additionally, it is possible that the triceps fat measurement is more reliable than the other ultrasound measurements. Triceps fat thickness did also factor in the TEE and PAEE kilocalorie models, though it was not considered as significant as in the MET models. Waist circumference appears in all models except the PAEE MET model. Waist circumference in conjunction with triceps fat thickness may give a good indication of how subjects' weight is distributed about their bodies. This in turn may influence energy expenditure directly or indirectly due to the effect of weight distribution on walking economy.

Resting VO_2 was common to the TEE MET and TEE kilocalorie models derived by stepwise regression. This parameter is a representation of the amount of energy that an individual consumes when at rest. The difference between resting VO_2 and the amount of oxygen consumed performing an activity such as treadmill walking represents the energy cost of the physical activity, and some studies have removed base level energy consumption in order to predict the physical activity energy expenditure directly (77). By including resting VO_2 in the prediction model, the model may be effectively accounting for the difference between resting EE and that which is due to physical activity, and a consequence of this may be an improvement in EE prediction. Unsurprisingly, the resting VO_2 parameter does not feature in the PAEE models as it has already been factored into the dependent variable (resting VO_2 was first removed from the measured VO_2 for the PAEE MET models, and a combination of resting VO_2 and resting CO_2 was removed from the measured PAEE in the PAEE kilocalorie model).

Diastolic blood pressure was selected by the stepwise regression for all models except the TEE kilocalorie model where systolic blood pressure was chosen. These parameters were added by the stepwise regression in the latter steps of the algorithm, which suggests that blood pressure measurements may have lower explanatory power in the model but high statistical significance. This reflects the findings of Snodgrass et al. (279) who reported statistically significant correlations between both systolic and diastolic blood pressure with the basal metabolic rate (BMR) of a population of Siberians.

It was unexpected that the mediolateral counts parameter was identified by the stepwise regression before the vertical counts parameter in the both TEE and PA kilocalorie models. However, this may be due to the models requiring a measure of both accelerometer output and weight before a good correlation between parameters and kilocalories is observed – to support this, it can be seen that once weight is added to the TEE model the RMSE reduces by 22% (Table 30). It was also unexpected that the thigh fat thickness did not appear in any of the models, as there is research to suggest that the energy cost of walking in the obese may be increased due to greater weight of the leg (240). It is possible, in this case, that differences in the EE model which are due to fat distribution may have been explained by other parameters, and that thigh fat thickness may not have any further explanatory power.

Gender and age do not feature in either the TEE MET or TEE kilocalorie model, which is also unexpected as previous research has identified these as factors affecting energy expenditure between individuals (234-237), and the subject group was sufficiently diverse in these areas to expect them to have a bearing on the estimation models. It may be that the other parameters have better explanatory power, and gender and age do not significantly improve the model once the other parameters have been selected. Age factored in the PAEE MET model, which again is unexpected because while there is a correlation between age and BMR (290), resting energy expenditure has already been accounted for in this model.

Lean mass was another parameter unexpected excluded from the TEE model, as it has been identified as a major determinant of energy expenditure (237). It may be that for the TEE models lean mass did not add greater predictive power than the weight parameter in combination with other parameters. However, lean mass was substituted for weight in the PAEE kilocalorie model. Several other candidate parameters, such as heart rate and height were also absent from the stepwise regression models, which may simply suggest that these parameters are not useful in improving EE prediction accuracy, or it may be that some are correlated with one of the selected model parameters, and therefore do not provide additional explanatory power to the regression.

5.4.2 Practical Implications of the Model Parameters

For practical application of the EE estimation algorithms, the number of parameters should be optimised such that a balance is reached between estimation accuracy and ease of execution of the measurement procedure. Ideally, for the practitioner or researcher implementing an energy estimation tool such as this, there should be as few measurements as possible, and the total time cost of measurements should be low. This is particularly important for large cohort studies where a time consuming measurement procedure may not be feasibly implemented. The number of measurements is, therefore, a concern, and the difficulty in obtaining certain measurements is also a factor. Future research may consider building a prediction model according to the ease by which measurements may be taken. The models derived from kilocalorie data are comprised of a greater number of parameters than the MET models, without providing greater EE estimation accuracy. For these reasons, the MET models would appear preferable.

Stepwise regression derives a multilinear model from a set of initial terms by systematically adding and removing terms according how well the model is improved by each addition/removal. This method returns a model which is a good fit for the data from which it was derived, but it does not consider the degree to which the model is improved by the addition of each parameter. This means that parameters are selected which, although statistically significant, do not greatly improve EE estimation accuracy. In these cases, the cost of collecting these anthropometric/physiological measurements is not justified by the gain in accuracy they confer. An example of this can be seen in Table 29 which shows that although the RMSE generally decreases with the addition of each parameter, the gain in accuracy is low for steps 4, 5 and 6, and RMSE actually increases for the final step. Overall, the enhanced TEE MET model (Model M_{SW}, see Table 25) improves EE estimation by 19.4% over vertical accelerometer counts alone, but some of these measurements, such as resting VO_2, are not easily made. A 16% improvement in accuracy may be obtained by the addition of the triceps fat thickness and waist circumference measurements (Table 29). Similarly, the addition of BMI and triceps fat thickness improves prediction accuracy of the TEE kilocalorie model (Model KC_{SW}) by around 15% over a model containing only vertical and mediolateral counts (Table 30). For practical purposes, this presents a significant improvement in energy expenditure estimation by the addition of two relatively easily made measurements. The improvement is more modest in the PAEE models (Table 31 and Table 32). This may suggest that the greatest gains are occurring for parameters which account for BMR in the model.

5.4.3 Comparisons with Previous Research

Comparison of results from the present study with previous research may be made using R^2 or RMSE. Only those studies that report results for locomotion may be compared directly.

Where r values are reported such as in Freedson et al. (71) (r=0.88 for TEE in METs), Hendelman et al. (217) (r=0.89 for TEE in METs), and Bouten et al. (116) (r=0.96 for PAEE in $W.kg^{-1}$), they are squared for comparison and become R^2=0.774, R^2=0.792 and R^2=0.922 respectively. These compare with R^2=0.803, R^2=0.748, R^2=0.721 and R^2=0.783 returned for the TEE kilocalorie, TEE MET, PAEE MET, and PAEE kilocalorie models respectively in the present study. As discussed previously, the results from these two studies were obtained from small homogeneous groups, which are likely to improve results. When considering vertical counts alone, as in Freedson et al. (71), the present study returns only R^2= 0.55. This is most likely due to the more diverse subject group in the present study, and this may more accurately reflect the relationship that might be obtained between accelerometer counts and EE from the population at large. With the addition of subject attributes to the model, the R^2 values between the present study and Freedson et al. (71) are comparable. Those R^2 values obtained by Bouten et al. (116) remain superior, but again they are likely to be inflated due to the small number (n=11) of subjects, all of low BMI (mean 20.5±1.9 kgm^{-2}).

An informative study was carried out by Lyden et al. (102), whose aim was to comprehensively evaluate a number of common EE prediction equations against a large, diverse population. The study used three accelerometers and eleven different prediction equations. Many different activities were considered, and each was reported separately for accuracy. The lowest returned RMSE was 0.5 METs for activities of doing dishes and dusting, and the lowest RMSE for locomotion was 0.6 METs, returned by the Freedson equation, for walking at $1.34ms^{-1}$ – though, this equation returned an RMSE of 0.9 METs for walking at $1.54ms^{-1}$, and banding the walking task into two separate speeds is likely to improve accuracy compared to considering a range of speeds at once. The present study reported an RMSE of 0.55 METs for treadmill walking using Model M1 which was derived in the same way as the Freedson equation through linear regression, and this measure improved to 0.43 METs with the addition of the subject attributes. When comparing results with Lyden et al. for locomotive activities, the present study apparently performs better than all eleven EE prediction equations tested. It must be noted that many of these equations were derived to predict EE across a range of different activities, which means that they were not optimised for locomotion, though this was not the case with the Freedson equation.

The study by Rothney et al. (103), which used a small number of subject attributes as input to an EE prediction model based on an ANN, achieved an RMSE value of 0.48 kcal/m for predicting TEE. The present study achieved an RMSE of 0.69kcal/m using the TEE kilocalorie model. However, a direct comparison is not possible as the study by Rothney et al. considered several different physical activities of varying intensities rather than walking in isolation. It might be expected that the present study might be at an advantage in this respect, as it has already been argued that an EE estimation equation applied to a mixed activity set returns lower accuracy than against an isolated activity. However, in the study by Rothney et al. it is possible that the low intensity activities (such as playing cards, typing and handwriting) returned very high prediction accuracies, because they will show less variation from baseline EE consumption levels and less variation in accelerometer output. High accuracies for low intensity activities over long periods would balance out low accuracies for short duration dynamic activities which have more scope for error. On the other hand, it is also possible that the accelerometer features used as parameters to the ANN were effective in improving EE prediction accuracy across activities.

5.4.4 Effect of Repeated Measures on Accuracy

As discussed in 5.2.4.6, because the EE estimation algorithm was tested using repeated measures per subject this may artificially increase apparent accuracy. For this reason, the algorithm was tested using accelerometer data collected from a single walking speed chosen at random for each subject. The test was carried out ten times to obtain indicative results of how accuracy is affected, and the results are shown in 5.3.2.3. The tests were performed independently for each of the four EE estimation models (M_{sw}, KCs_{w}, MPA_{sw} and $KCPA_{sw}$).

There were a range of results returned over the ten iterations for each model (see Table 35), some of which showed higher accuracy than the repeated measures results, though the majority were poorer. The differing results are largely due to the random selection process used to determine each dataset to be tested. On average the single speed tests produced marginally poorer accuracy results across all models except the PA MET model (MPA_{sw}) which returned a slightly higher RMSE on average. The results may suggest that using repeated measures gives higher accuracies than using single speeds per subject. However, this is not conclusive as the results are sensitive to the random selection process – which is also likely to explain the higher accuracy being returned for the single speed test of the MPA_{sw} model over the corresponding repeated measures results. However, presuming that the repeated measures do indeed artificially increase EE estimation accuracy, the extent of this indicated by the ten single speed tests implies that it should not have a significant effect on answering the research question.

5.5 Limitations

The present study has purposefully taken walking in isolation to consider if anthropometric and physiological parameters improve the EE estimation model. As concluded by several studies, each mode of activity requires an independent EE estimation model, and these may also be improved by the addition of such parameters. However, it is likely that the parameters which improve walking EE prediction will not be identical to those required in other modes of activity. For example, the lack of vertical movement when using a rowing machine suggests that a greater body weight would play less a part in increasing EE. This means that each activity mode would require separate analysis to identify which additional parameters are applicable to the EE prediction model. However, once EE estimation models are established for a number of activities, an initial classification phase (as described in Chapter 3) may be implemented in order to select the appropriate model. Accelerometry holds the advantage that the classification algorithm may be applied to the same dataset from which the energy expenditure data is to be estimated.

For the MET estimation model, standard METs were used as a measure of EE rather than individualised METs based on subject RMR. This means that the EE estimation model presumes that the MET is an absolute value rather than one which has a relationship to an individual's RMR. There is an established method to estimate RMR using the Harris-Benedict equation (280) which can then be used to calculate 'corrected' METs incorporating the estimated RMR. However, this equation uses height, weight and age, and these attributes were also used as parameters to the stepwise regression when generating the EE estimation model for METs. To use this approximation for RMR when generating the model could, therefore, potentially mean that height, weight and age feature in both the independent and the dependent variables. This was not desirable therefore the standard MET was used in the model, which is in line with previous research.

Treadmill walking was considered in order to answer the research question. However, walking under free living conditions where surfaces may be uneven or at an incline may affect which physiological and anthropometric parameters are most effective in improving EE estimation. Walking on an incline requires a greater amount of EE than walking on the flat (74), as does walking on muddy ground over walking on a paved path. These different environments may amplify the effect of particular attributes of individuals on EE, and may bring forth other attributes that have a bearing on EE under these conditions. This remains to be tested. The research in the present study, however, has established that EE prediction may be improved by such attributes.

Although ethnicity has been identified as a factor which affects EE in individuals (235, 237), it was not possible to recruit a sufficient number of individuals from diverse ethnic groups in order to test the effect of ethnicity on the EE prediction models. It is not clear how the addition of ethnicity, coupled with potential physiological or anthropometric differences between ethnic groups, would affect the prediction model.

The present study did not consider how different features of the accelerometer output might also improve the EE prediction model for walking. This was decided in order to keep the research question clear and focussed on how attributes of the individual might improve prediction. Differing characteristics between obese and non-obese gait may be captured through accelerometer features, which may help inform the prediction model, as was the case with walking speed prediction described in Chapter 4. Indeed, for walking, it may be that the same set of accelerometer features identified in Chapter 4 may also be used to improve the energy expenditure prediction model. However, accelerometer features would only account for those differences in movement between individuals, and would not be able to identify physiological factors affecting energy expenditure of individuals. The energy expenditure prediction model has been shown to improve with the addition of accelerometer features (103-104), but it is not likely that accelerometer features alone are sufficient to preclude the need for anthropometric and physiological parameters.

5.6 Chapter Summary

Energy expenditure in individuals is influenced by a number of demographic, anthropometric and physiological factors (234-237), and also may be affected by individual biomechanics for activities such as walking (240). Many of these effects are due to factors associated with obesity. However, current EE estimation models using accelerometry have not accounted for the potential effects of obesity on prediction accuracy. Where subject-specific attributes have been included as parameters in prediction models (103, 114, 214), their effect on EE prediction accuracy has not been conclusive due to one or more of the following issues: the model has not been adequately tested on diverse BMI groups; the model has been tested against multiple activities, making it unclear whether it is the choice of activities or the additional parameters that have affected prediction accuracy; the study has incorporated additional innovations in parallel, which has obscured the effect of the additional parameters;

or there have been insufficient parameters included in the prediction model to test the principle.

The present study sought to investigate whether the addition of subject-specific attributes to the EE prediction model may improve EE prediction using accelerometry. Twenty-one candidate demographic, anthropometric and physiological parameters were chosen, and stepwise regression was used to identify which parameters provide the best explanatory power to a linear EE prediction model. These enhanced models were compared with basic linear models incorporating only accelerometer counts and subject weight (as developed in previous research (71, 116)). The enhanced models returned around 20% improvement for MET prediction and around 11% improvement for kilocalorie estimation over the basic linear models. However, many of these measurements may be too time consuming for perform on large subjects groups, such as resting VO_2 levels. A 16% improvement in the MET model accuracy, and an 8.6% improvement in the kilocalorie model, could be achieved by the inclusion of only two additional parameters: triceps fat thickness, and waist circumference.

The present study has demonstrated that by including triceps fat thickness and waist circumference in a multilinear EE prediction model for the walking mode, improved estimates for EE estimation can be achieved over the standard linear equations currently widely in use. The time required to measure triceps by ultrasound and waist circumference by tape measure is in the order of minutes, making this a promising innovation for use in research under free-living. This has the potential to impact epidemiological research and health care applications where accurate measures of energy expenditure are required in the field.

6 Conclusions

The current global obesity epidemic is a mounting concern due to its detriment to public health and consequent economic cost. It has recently been reported that today more than 2.1 billion people are overweight, and based on current rates almost half the world population will be overweight by 2030 (294). Activities promoting physical activity are among strategies aimed at reducing weight (20-32). However, the report highlights that there is no systematic method of measuring the potential impact of interventions targeted at reducing obesity. To measure the impact of weight-loss interventions which prescribe increased activity levels requires tools capable of objectively measuring physical activity. Previous research has investigated how accelerometry may be used for this purpose, and has demonstrated how predictions of activity type and intensity can be generated by applying algorithms to accelerometer output. However, this research has generally been carried out using subject groups predominantly in the normal BMI range, and there is a lack of research which investigates the effects of obesity on the algorithms aimed at quantifying PA.

A comprehensive activity monitoring system (as described in 6.2 below) should be equally effective in measuring physical activity for both obese and non-obese individuals, and one of the primary aims of the research has been to investigate this principle. The classification study described in chapter 3 found that a high accuracy (85% for a hip-mounted accelerometer, and 94% for an ankle-mounted accelerometer) was achievable for a BMI subject group composed of normal BMI, overweight, and obese individuals. Similarly, the walking speed estimation study (chapter 4) returned accuracies comparable to previous research for both the normal BMI group and the obese group (though it was concluded that accuracy may be improved by using different models depending on BMI). The energy expenditure study (chapter 5) found that the addition of subject characteristics to the estimation model improves energy expenditure estimation accuracy; and in many cases these characteristics may be indicative of obesity level, such as triceps fat thickness which featured in all estimation models. These results suggest that, for each of the three categories of physical activity considered in the studies, a single approach to measuring physical activity is applicable across BMI groups, though it may be necessary to modify algorithms to incorporate subject-specific attributes.

A prescribed exercise programme may incorporate numerous activities that need to be distinguished by an activity monitoring system being used to evaluate the effectiveness of such a programme. To address this, the study in chapter 3 tested seven aerobic/gym-based exercises, and three everyday activities (walking, ascending stairs, and descending stairs). The high classification accuracies achieved by the study suggest that, in principle, it is possible to use accelerometry to create an activity profile for several dynamic activities, suitable for use in evaluating a programme of physical activity. However, it should be reiterated that the classification scheme tested in the study is currently inadequate to be deployed in the field (as discussed in section 3.4).

Measures of physical activity intensity have been considered by the studies in chapter 4 and chapter 5. Ultimately, it would be desirable to measure the intensity of multiple activities, but the studies represented here have been limited to measuring aspects of walking. It was found that, under experimental conditions, walking speed may be predicted with a degree of accuracy sufficient to distinguish bouts of brisk walking from slow walking, which is important in terms of measuring whether recommended walking guidelines for weight loss have been met. Chapter 5 found high correlation coefficients between walking energy expenditure and accelerometer output coupled with anthropometric and physiological measurements, and at best RMSE results were less than 0.5 standard METs (or the equivalent in kilocalories) which is comparable to previous research. These results indicate that it is possible to accurately estimate the intensity of walking from accelerometer output, and it is reasonable to surmise that this may be extended to types of activity other than walking.

The three studies presented in the thesis have each returned promising results using accelerometer data collected at the hip. This suggests that it is possible to develop a single integrated system able to measure multiple aspects of physical activity simultaneously. Furthermore, the techniques involved have been shown to be effective across BMI groups. The studies in combination represent a step towards such a system which could be implemented in many areas of research. Focus has also been given to pragmatic concerns, such as limiting the number of accelerometers to one, and considering an unobtrusive placement site for the accelerometer (the hip), to make these approaches affordable and practical for large scale real-life studies. However, before such a system may be fully realised further research is required. This is discussed in more detail below.

6.1 Future Work

The most important recommendation for further research is that each of the techniques developed in these studies requires adequate testing under free-living conditions. Many studies have reported success in the laboratory but have found that in the field their respective algorithms have not performed equally well (160, 218). It may be that some of the innovations implemented in these studies mitigate this problem. For example, the walking speed algorithms incorporate several accelerometer features which allow speeds to be estimated for both overground walking and treadmill walking. This may mean that they capture one or more essential characteristics of walking speed from the accelerometer signal which makes them less sensitive to changes in terrain. However, until these techniques are fully tested under free-living conditions, it is not clear whether the approaches to quantifying physical activity presented in this thesis will perform adequately for epidemiological research and health applications.

An important aspect of the research presented in this thesis is that the techniques investigated may be used in parallel to derive information on several different characteristics of activity from a single dataset of accelerometer output. The techniques may also be used in conjunction with each other. The ability to classify physical activities from accelerometry is in itself a useful tool to provide information on behaviours which may inform epidemiological research into obesity, but classification is also a necessary element in applying other techniques. For example, walking speed algorithms may only be applied to accelerometer signals that represent bouts of walking, which necessitates a preliminary phase of classification to identify those bouts. Furthermore, previous research into energy expenditure estimation from accelerometry has concluded that a single EE prediction equation does not apply across activities (242), which suggests that separate prediction equations are required depending on the activity being performed. Again, a preliminary classification phase is required in order to select the appropriate equation, and this may be applied to the same dataset from which the EE estimates are to be obtained.

The energy expenditure estimation study in chapter 5 demonstrated in principle that energy expenditure prediction could be improved by the addition of subject-specific anthropometric and physiological attributes. The study considered the activity of walking to test this principle. However, different activity modes do not necessarily exhibit the same relationship between accelerometer output and energy expenditure. Also, it is not certain whether different subject attributes may be required according to activity mode. Further research is recommended into deriving energy estimation prediction models for other modes of activity. It may not be practical to test all possible types of activity. However, certain subject attributes may emerge as significant which are common to many activities, and some activities may share a similar relationship between accelerometer counts and EE.

6.2 A Future Activity Monitoring System

It is envisaged that in the future it may be possible to create a comprehensive activity monitoring system which will be able to record multiple aspects of human physical activity using accelerometry. This system would take body-worn accelerometer data from an individual and produce an activity profile detailing the type of activities that have been performed (dynamic and sedentary) and the frequency and duration of each, the total energy expended (with a breakdown by activity), and the intensity of the activities undertaken (walking speed, for example). Figure 29 shows a basic overview of such a system.

The heart of the activity monitoring system is the activity classification module. This would determine the type of activities represented in the accelerometer output, and their frequency and duration, and would also interact with the other modules to inform the energy estimation and speed/intensity prediction algorithms. The study in chapter 3 corresponds with this module and has aimed to address some of the issues which may be encountered by a classifier, such as a mixed BMI group and multiple dynamic activities. However, the classification module would require training data for many more dynamic activities. It would also require a method of determining when activities were unknown to the training set, and have an alternative approach to classification to deal with these activities. It would need to be able to filter out noise in the signals generated by activities such as riding a bus, and would also require a separate classification algorithm for use in identifying different types of static activity (classification of static activities is described in more detail in Appendix C).

The energy expenditure estimation module would first utilise the classification module to ascertain the type of activity being performed. An appropriate energy expenditure estimation equation would be selected from a library of EE estimation equations according to the activity being performed. Also, the relevant anthropometric/physiological/demographic measurements would be retrieved for input into the equation. Currently, there is no such library of EE estimation equations. The study in chapter 5 derived equations for estimating the EE for walking based on accelerometer output and additional subject attributes. However, future studies would need to derive EE estimation equations for many activities. There would also need to be a contingency rule for when the detected activity is either unknown or does not have a corresponding EE estimation equation.

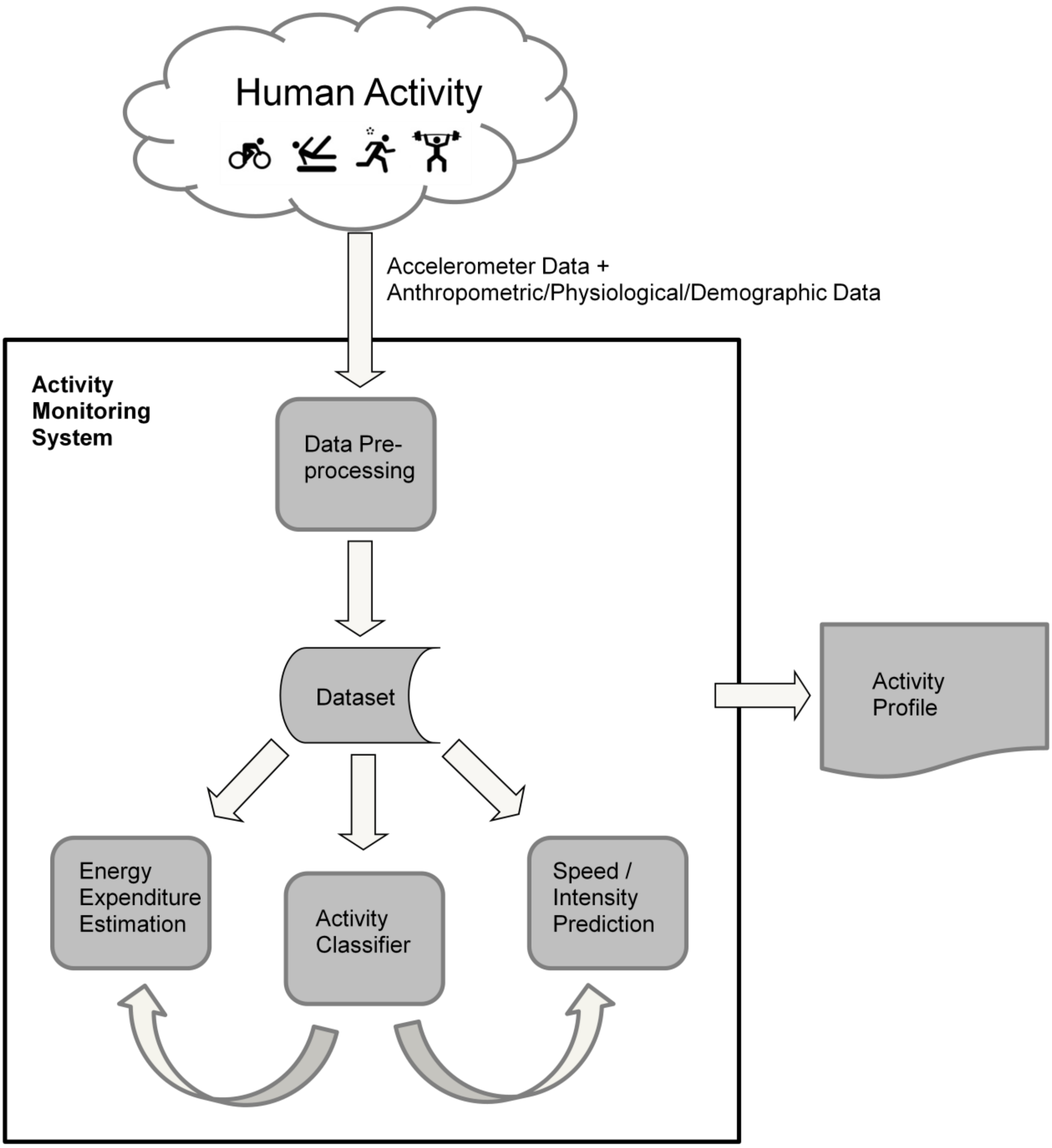

Figure 29: conceptual framework of envisaged activity monitoring system.

It is sometimes important to measure the intensity of an activity in terms other than energy expenditure. Walking speed, for example, may give greater insight into an individual's activity patterns over merely measuring the energy expenditure of walking. In chapter 5 it was investigated how walking speed for individuals of varying BMI may be estimated. However, in a comprehensive activity monitoring system measures of intensity would also be required for other activity modes. Again, a module designed to measure the intensity of an activity would first call on the classification module to determine the corresponding algorithms to be applied to the data from which to obtain intensity values. In addition to the speed of an activity, the intensity of activities such as rowing, for instance, might be measured in terms of cycles per minute. It would not be feasible to apply this module to all activities, however, and some issues may be insurmountable. It may not be possible, for example, to estimate cycling speed – even if the rate at which the pedals were being rotated by the cyclist was obtained, it would not be possible to calculate the speed of forward motion without additional information not available through the accelerometer, such as which gear the cyclist was using.

6.3 Summary

The main objective of the thesis was to investigate how the measurement of physical activity using accelerometry is affected by obesity. To achieve this goal, three studies have been undertaken using mixed BMI subjects in order to establish whether accelerometry-based PA measuring techniques are equally applicable to obese groups and non-obese groups. Three types of measurement technique were considered: activity type classification, walking speed prediction, and energy expenditure estimation. Each study reported results which show that it is possible to achieve similar results for obese groups and non-obese groups, though some modification of techniques is required. Furthermore, these studies achieved results that are comparable with, or an improvement upon, previous research. Taken in combination, these studies represent a step towards an integrated system, capable the simultaneous measurement of multiple aspects of PA. This type of system has a great deal of potential for use in a wide range of research areas and practical applications.

7 Appendices

7.1 Appendix A: Physical Activity Readiness Questionnaire (PAR-Q)

1.	Are you currently taking any medication that might affect your ability to participate in the test as outlined?	YES	NO
2.	Do you suffer, or have you ever suffered from, cardiovascular disorders? e.g. Chest pain, heart trouble, cholesterol etc.	YES	NO
3.	Do you suffer, or have you ever suffered from, high/low blood pressure?	YES	NO
4.	Has your doctor said that you have a condition and that you should only do physical activity recommended by a doctor?	YES	NO
5.	Have you had a cold or feverish illness in the last 2 weeks?	YES	NO
6.	Do you ever lose balance because of dizziness, or do you ever lose consciousness?	YES	NO

7.	Do you suffer, or have you ever suffered from, respiratory disorders? e.g. Asthma, bronchitis etc.	YES	NO
8.	Are you currently receiving advice from a medical advisor i.e. GP or Physiotherapist **not** to participate in physical activity because of back pain or any musculoskeletal (muscle, joint or bone) problems?	YES	NO
9.	Do you suffer, or have you ever suffered from diabetes?	YES	NO
10.	Do you suffer, or have you ever suffered from epilepsy/seizures?	YES	NO
11.	Do you know of any reason, not mentioned above, why you should not exercise? e.g. Head injury (within 12 months), pregnant or new mother, hangover, eye injury or anything else.	YES	NO

7.2 Appendix B: Application of the Classifier to Free-living Data

As part of a project that was independent to this thesis – SSHOES: European Community's Seventh Framework Programme (FP7/2007-2013) under SSHOES project, Grant Agreement no. NMP2-SE-2009-229261 – a pilot study was performed where a number of participants collected hip-mounted accelerometer data over a period of several days using the Actigraph GT3X+ activity monitor.

Each participant wore an activity monitor on the right hip above the iliac crest. The protocol for fitting the monitor to the hip was demonstrated to participants by a researcher. The participants each took away a monitor for a seven day period and each morning fitted the unit themselves. The monitors were worn during waking hours, but were removed for sleep, baths and showers. Accelerometer data was recorded continuously for the seven day period. Participants were asked to keep diaries of their daily activities. Participants recorded notable periods of activity such as cycling, walking, and travelling by car, but were not required to provide a comprehensive account of all activities within the period. This meant that there were numerous unlabelled activities in the dataset.

The classification algorithm described in Chapter 3 was tested against the above data. The algorithm was trained using the accelerometer data collected for the ten physical activities described in the classification study (Chapter 3). The algorithm was applied to a sample of the data collected in the seven day protocol outlined above. Datasets representing a twenty-four hour period of time were selected from the accelerometer record for four subjects. The selected datasets were segmented into two second windows. An additional step was added to the algorithm which first decided whether the window of activity was an example of static activity (sitting, standing or lying) or dynamic activity. This was determined according to whether the accelerometer signal magnitude exceeded a threshold chosen as the boundary between static and dynamic activity. For windows representing dynamic activity, the best performing feature set identified in chapter 3 was used to classify activity type.

7.2.1 Results

The number of minutes spent in each activity was estimated by the classifier, and the percentage of time in each activity (depending on the number of windows designated as dynamic activity) was also calculated (Table 36).

Subject:	1		2		3		4	
Activities Reported:	Walking / UpStairs / Downstairs / Driving		Sitting / Walking / Treadmill / Driving		Sitting / Walking / Car passenger		Driving / Cycling (0.5 hours)	
Feature Set: F10	Min.	Percent	Min.	Percent	Min.	Percent	Min.	Percent
Walk	27.57	65.9%	40.00	42.5%	76.07	51.1%	16.00	27.3%
DownStairs	1.50	3.6%	5.43	5.8%	12.23	8.2%	2.17	3.7%
UpStairs	3.77	9.0%	21.60	22.9%	29.67	19.9%	6.83	11.6%
Cycling	2.13	5.1%	3.00	3.2%	2.03	1.4%	10.37	17.7%
Rowing	1.13	2.7%	2.70	2.9%	3.33	2.2%	2.43	4.1%
Crosstrainer	0.00	0.0%	0.07	0.1%	0.00	0.0%	0.00	0.0%
Jog	0.07	0.2%	0.03	0.0%	0.00	0.0%	0.00	0.0%
Stepping	2.03	4.9%	10.57	11.2%	8.83	5.9%	5.47	9.3%
Sidestepping	2.40	5.7%	7.97	8.5%	12.10	8.1%	11.43	19.5%
Sidestretching	1.23	2.9%	2.83	3.0%	4.53	3.0%	3.97	6.8%
Total Minutes	41.83		94.20		148.80		58.67	

Table 36: Minutes spent in each dynamic activity, as estimated by the classification algorithm using feature set F10 described in chapter 3.

7.3 Appendix C: Classifying Static Activities

The focus of chapter 3 was to identify dynamic gym-based and free-living activities from body-worn accelerometer output obtained from a mixed BMI group. The research questions were concerned with how the classification of dynamic activities might be affected by subject BMI. Static activities (standing, sitting and lying) were not considered, as classifying these require a separate approach and may be performed independently to the classification of dynamic activity. The classification scheme outlined in the chapter, therefore, is not adequate to measure individuals under free-living conditions, as a large proportion of time is spent by individuals in sedentary activity modes. It may, therefore, be useful to describe how static activities may be incorporated into an activity classification system.

The first step is to establish a threshold between static activities and dynamic activities according to accelerometer output levels. The signal magnitude area (SMA) of the combined accelerometer signals is indicative of the intensity of the activity being performed by the accelerometer wearer (145). This is calculated as the sum of the high-pass filtered, rectified accelerometer signal (as described in section 2.5.1). The SMA threshold between static and dynamic activity may be established from labelled accelerometer data: by plotting a histogram of SMA values for windows of different static and dynamic activities, for multiple subjects, the SMA threshold becomes apparent (Figure 30). By considering the accelerometer axes separately, additional thresholds may be established between individual static activities. For example, Figure 31 shows histograms for windows of sitting, standing and lying activities for a triaxial accelerometer. Divisions between these activities can be clearly seen.

Once thresholds have been derived, the static classifier may be constructed and applied to unseen data. The classification scheme consists of an initial phase where windows of accelerometer data are assessed as being static or dynamic (Figure 32). Those windows deemed as dynamic are passed to a dynamic classification algorithm such as that described in chapter 3. Those that are denoted as static may be classified using a simple decision tree based on the pre-established thresholds between sitting, standing and lying (Figure 33).

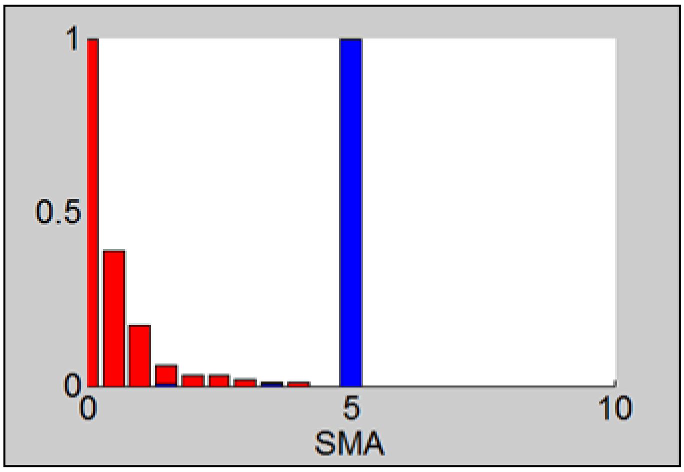

Figure 30: Example histogram of SMA for windows of static (red) and dynamic (blue) activities. The SMA threshold is apparently a little less than 5 SMA units.

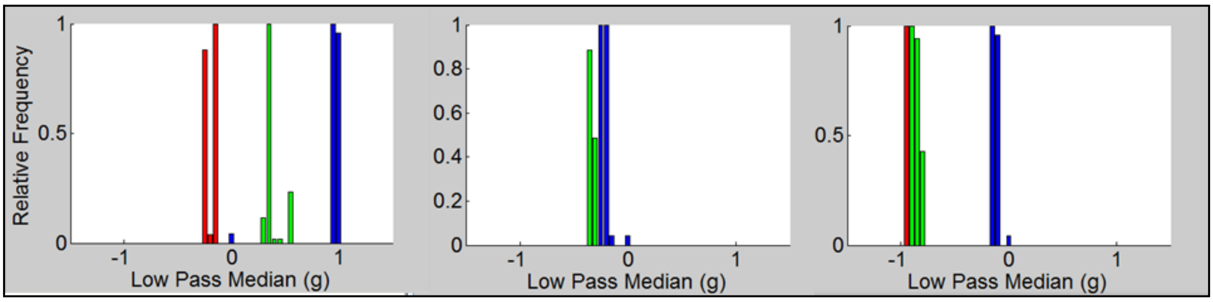

Figure 31: Example histograms of median g values for three low-pass filtered accelerometer signals (vertical signal left, mediolateral signal centre, anteroposterior signal right). Thresholds may be established for sit (red), stand (green) and lie (blue) activities. Typically, more than one axis is required to differentiate between static acitivities.

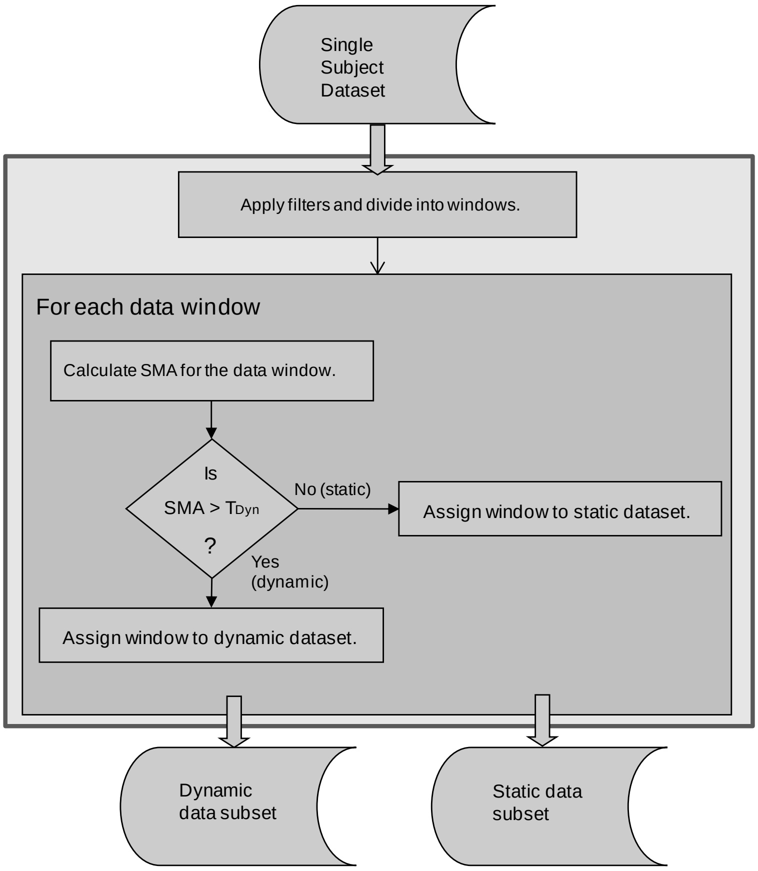

Figure 32: Initial phase of classification where windows of accelerometer data are designated as static or dynamic according to the pre-established SMA value (T_{Dyn})

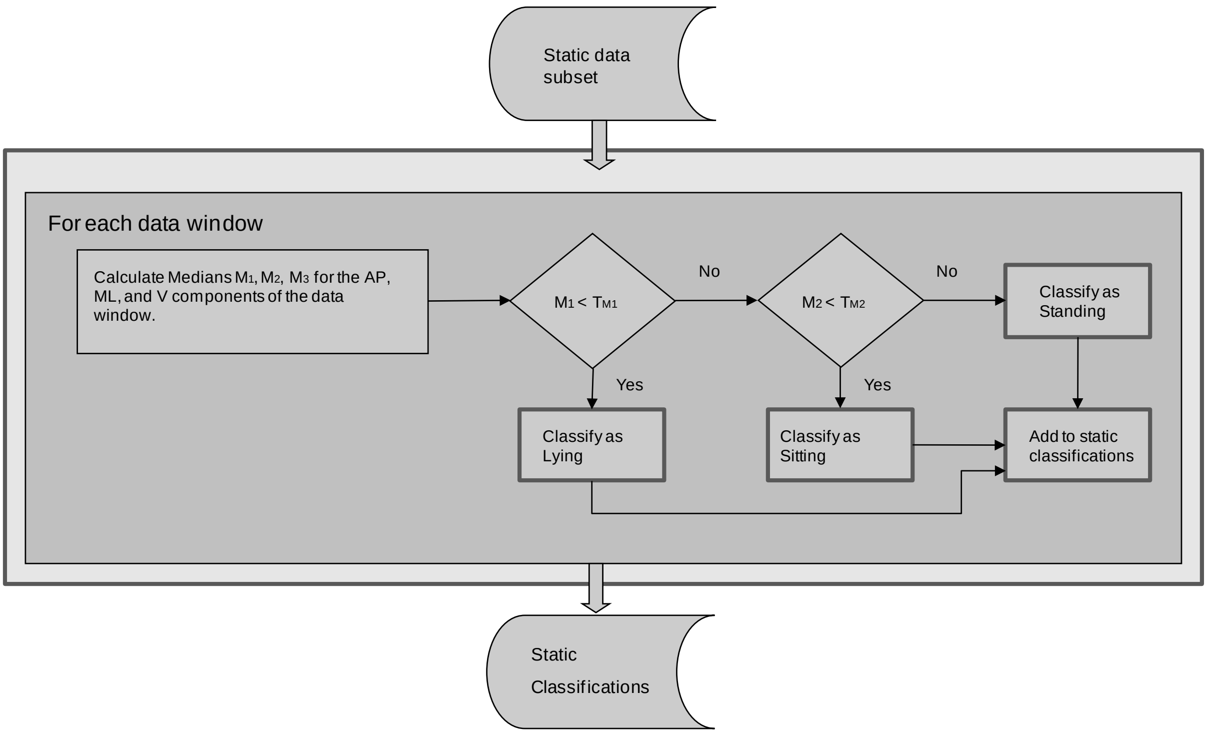

Figure 33: Example static activity classification algorithm. A simple decision tree classifies windows of activity according to predetermined thresholds (T_{M1} and T_{M2}) which are based on the median of the accelerometer signals for the three axes.

7.4 Appendix D: Application of the Speed Prediction Algorithm to Naturalistic Walking Data

In order to give an indicative performance level of the walking speed prediction algorithm (described in chapter 4) under free-living conditions, the algorithm was retrospectively applied to the outdoor walking accelerometer data which was collected as part of the classification study (described in chapter 3). Participants had performed approximately three minutes of walking on campus on a paved path which was uneven and undulating in places. The route had been the same for each participant and was of approximately 293m in length. This distance was measured retrospectively using a surveyor's wheel. Timings for the walking trial had been made using a wristwatch. High precision for timings was not required for the classification study – it needed only to be sufficient to locate the corresponding accelerometer output in the dataset. Additionally, the participants were not stringently guided on the route – they had a small amount of freedom to walk wide or narrow on corners and paths. The combination of these issues is likely to mean that the measured average speeds for these walking trials contain a certain amount of error, though this may be mitigated by the length of the route.

There were sixteen subjects who had performed the protocols for both the classification study and the walking speed estimation study. Of these, eleven were in the obese BMI range and five were in the normal BMI range. Three tests were performed according to which dataset (as collected in the walking speed estimation study) was used to train the algorithm – these were as follows: combined treadmill walking and overground (laboratory) walking; treadmill walking only; and overground (laboratory) walking only. Results had been similar for the three speed estimation models derived in chapter 4, therefore model 1 was arbitrarily chosen for these tests.

Participants had been asked to walk at their preferred speed throughout the campus route. However, it is likely that their walking speeds did not remain constant. For this reason, the speed estimation algorithm returned an average walking speed estimate for a two minute period of representative walking data which was extracted from the middle of the walking trial data. Two minutes was chosen to ensure that the extracted data contained only walking – participants had taken different amounts of time to complete the course, but all had taken more than two minutes. Bland-Altman plots were used to analyse the agreement between the estimated average walking speed and the measured average speed.

7.4.1 Results

Bland-Altman plots are presented below for the three walking speed estimation tests. The two measurements being compared are the average walking speed of the timed walk over 293m, and the estimate for average speed returned by the speed prediction algorithm of the two-minute sample period of walking. The first figure (Figure 34) shows the results of training the speed prediction algorithm using accelerometer data for combined treadmill and laboratory-measured walking and applying this to outdoor walking data. The second and third figures show the results of training the algorithm using treadmill alone (Figure 35), and laboratory-measured walking alone (Figure 36) respectively. The mean of the differences and the values at +/- 1.96 standard deviations from this mean are indicated on each plot.

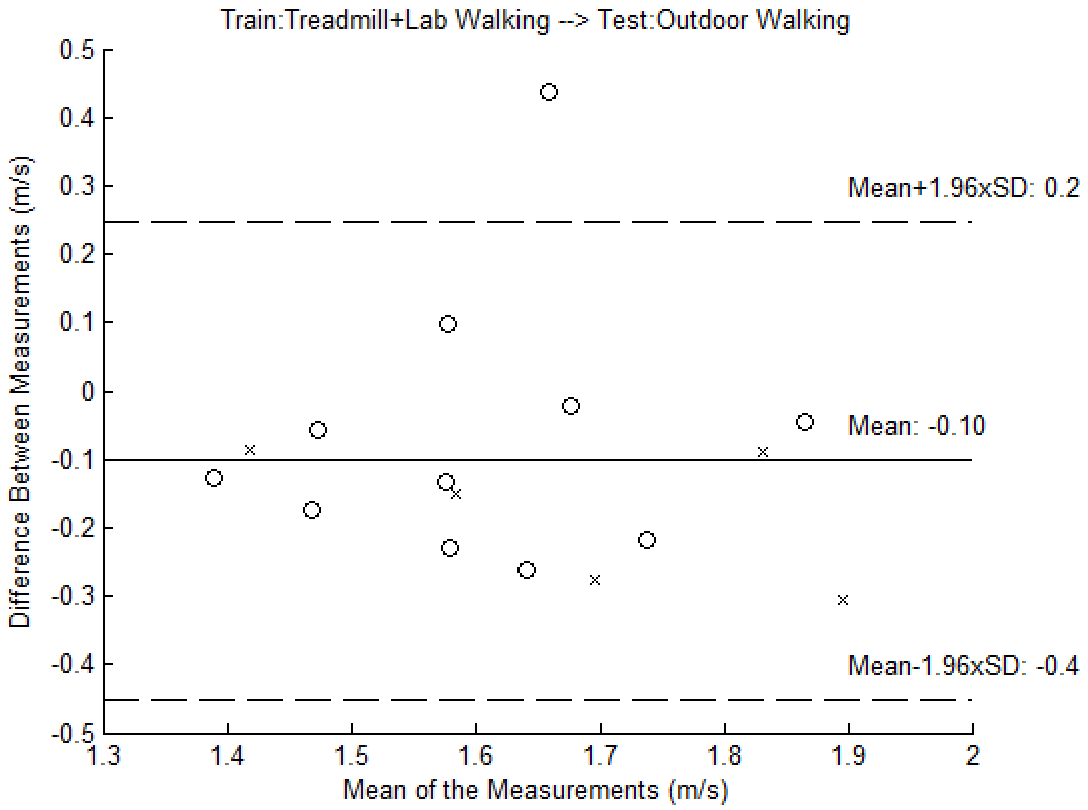

Figure 34: Bland-Altman diagram of the walking speed results when the speed prediction algorithm was trained using combined treadmill and overground (laboratory) walking data. (Obese subjects: O, Normal BMI subjects: x).

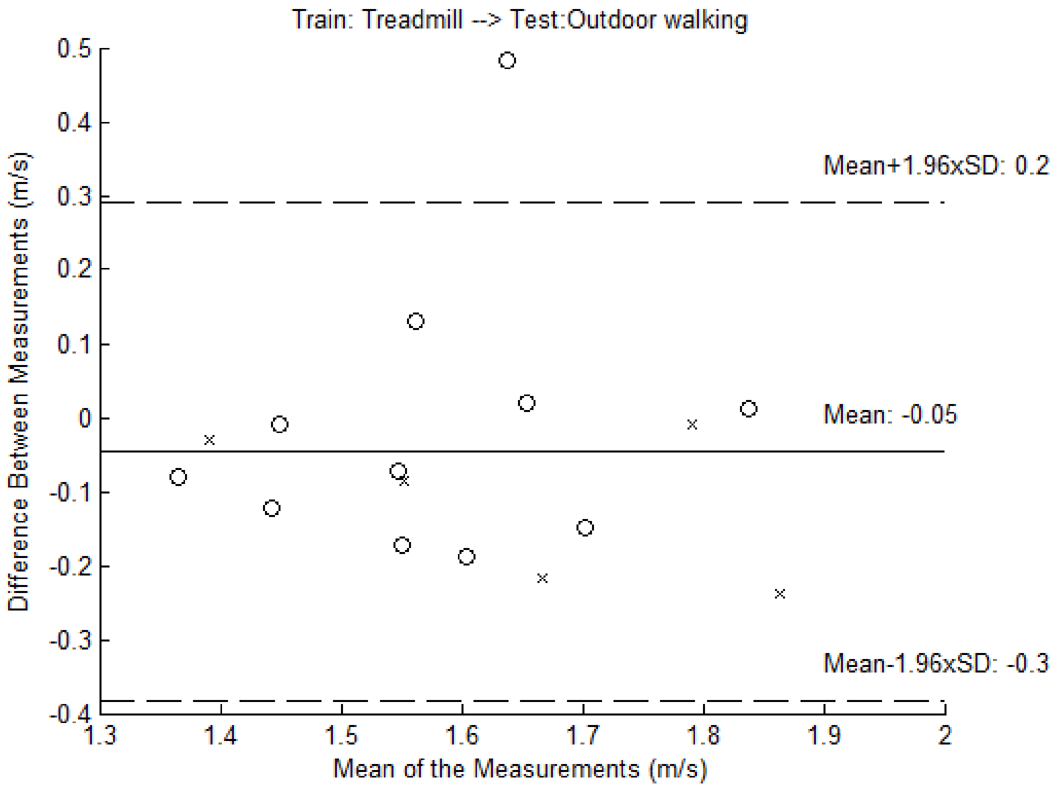

Figure 35: Bland-Altman diagram of the walking speed results when the speed prediction algorithm was trained using treadmill data alone. (Obese subjects: O, Normal BMI subjects: x).

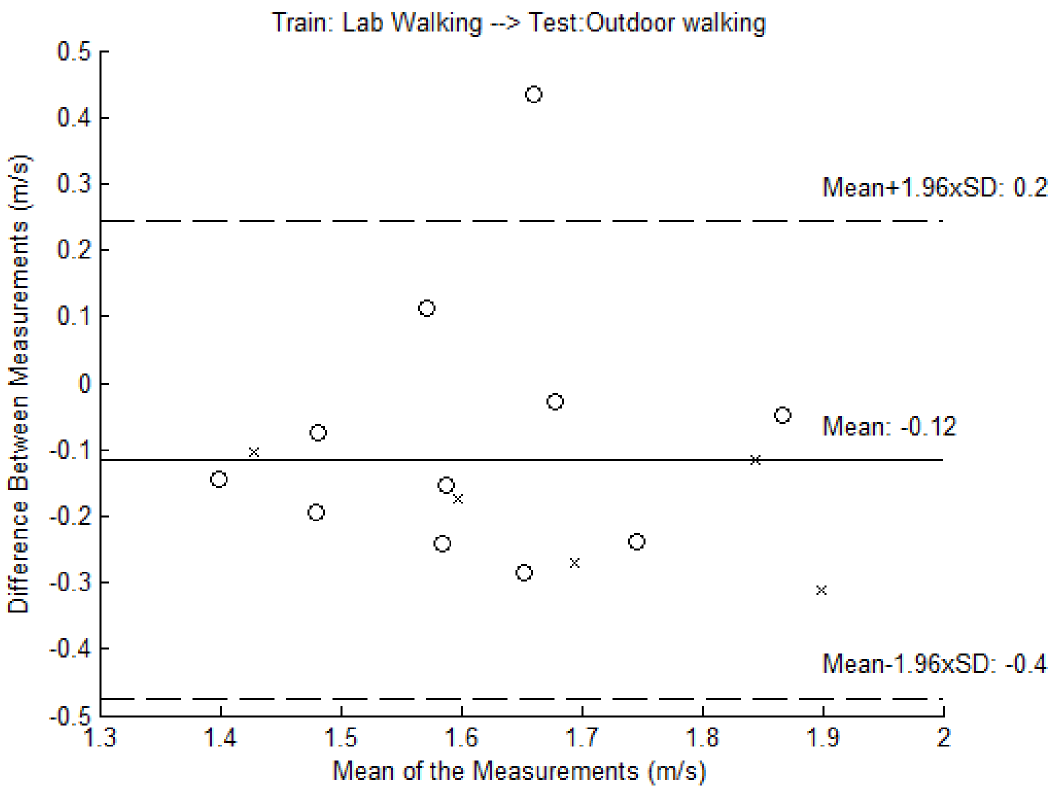

Figure 36: Bland-Altman diagram of the walking speed results when the speed prediction algorithm was trained using overground (laboratory) walking data alone. (Obese subjects: O, Normal BMI subjects: x).

References

1. World Health Organization. 10 Facts on Obesity. 2013 [cited 2014 18 March]; Available from: http://www.who.int/features/factfiles/obesity/en/.
2. Health and Social Care Information Centre. Health Survey for England - 2012. Summary of key findings. 2012 [cited 2014 1 July 2014]; Available from: http://www.hscic.gov.uk/catalogue/PUB13218/HSE2012-Sum-bklet.pdf.
3. McCormick B, Stone I. Economic costs of obesity and the case for government intervention. Obesity Reviews. 2007;8:161-4.
4. Butland B, Jebb S, Kopelman P, McPherson K, Thomas S, Mardell J, et al. Tackling Obesities: Future Choices: Project Report: Department of Innovation Universities and Skills; 2007.
5. Flegal KM, Carroll MD, Ogden CL, Curtin LR. Prevalence and trends in obesity among US adults, 1999-2008. Jama. 2010;303(3):235-41.
6. Finkelstein EA, Trogdon JG, Cohen JW, Dietz W. Annual medical spending attributable to obesity: payer-and service-specific estimates. Health affairs. 2009;28(5):w822-w31.
7. Brownson RC, Boehmer TK, Luke DA. Declining rates of physical activity in the United States: what are the contributors? Annu Rev Public Health. 2005;26:421-43.
8. Rolls BJ. The supersizing of America: portion size and the obesity epidemic. Nutrition Today. 2003;38(2):42-53.
9. National Health Service. Obesity - Causes 2012 [cited 2014 20 March 2014]; Available from: http://www.nhs.uk/Conditions/Obesity/Pages/Causes.aspx.
10. Miller WC, Koceja D, Hamilton E. A meta-analysis of the past 25 years of weight loss research using diet, exercise or diet plus exercise intervention. International Journal of Obesity. 1997;21(10):941-7.
11. Wu T, Gao X, Chen M, Van Dam R. Long-term effectiveness of diet-plus-exercise interventions vs. diet-only interventions for weight loss: a meta-analysis. Obesity Reviews. 2009;10(3):313-23.
12. Curioni C, Lourenco P. Long-term weight loss after diet and exercise: a systematic review. International Journal of Obesity. 2005;29(10):1168-74.
13. Welk GJ, Blair SN. Physical Activity Protects against the Health Risks of Obesity. President's Council on Physical Fitness and Sports Research Digest. 2000.
14. Haskell WL, Lee I, Pate RR, Powell KE, Blair SN, Franklin BA, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Medicine and Science in Sports and Exercise. 2007;39(8):1423.
15. King A, Tribble D. The role of exercise in weight regulation in nonathletes. Sports Medicine (Auckland, NZ). 1991;11(5):331.
16. Shaw K, Gennat H, O'Rourke P, Del Mar C. Exercise for overweight or obesity. Cochrane Database Syst Rev. 2006;4.
17. Tremblay A, Després J-P, Leblanc C, Craig CL, Ferris B, Stephens T, et al. Effect of intensity of physical activity on body fatness and fat distribution. The American journal of clinical nutrition. 1990;51(2):153-7.
18. National Institute for Health and Clinical Excellence. CG43 Obesity. Guidance on the prevention, identification, assessment and management of overweight. 2006 [cited 2014 1 July 2014]; Available from: http:guidance.nice.org.uk/CG43.

19. U.S. Department of Health and Human Services. 2008 Physical Activity Guidelines for Americans. 2008 [cited 2014 1 July 2014]; Available from: http://www.health.gov/paguidelines/pdf/paguide.pdf.
20. Janssen I, Fortier A, Hudson R, Ross R. Effects of an energy-restrictive diet with or without exercise on abdominal fat, intermuscular fat, and metabolic risk factors in obese women. Diabetes care. 2002;25(3):431-8.
21. Kiernan M, King AC, Stefanick ML, Killen JD. Men Gain Additional Psychological Benefits by Adding Exercise to a Weight-Loss Program. Obesity Research. 2001;9(12):770-7.
22. Nieman DC, Nehlsen-Cannarella SL, Henson D, Koch AJ, Butterworth DE, Fagoaga OR, et al. Immune response to exercise training and/or energy restriction in obese women. Medicine and Science in Sports and Exercise. 1998;30(5):679-86.
23. Svendsen OL, Hassager C, Christiansen C. Effect of an energy-restrictive diet, with or without exercise, on lean tissue mass, resting metabolic rate, cardiovascular risk factors, and bone in overweight postmenopausal women. The American journal of medicine. 1993;95(2):131-40.
24. Ross R, Rissanen J, Pedwell H, Clifford J, Shragge P. Influence of diet and exercise on skeletal muscle and visceral adipose tissue in men. Journal of Applied Physiology. 1996;81(6):2445-55.
25. Thong FS, Hudson R, Ross R, Janssen I, Graham TE. Plasma leptin in moderately obese men: independent effects of weight loss and aerobic exercise. American Journal of Physiology-Endocrinology and Metabolism. 2000;279(2):E307-E13.
26. Wadden TA, Vogt RA, Andersen RE, Bartlett SJ, Foster GD, Kuehnel RH, et al. Exercise in the treatment of obesity: effects of four interventions on body composition, resting energy expenditure, appetite, and mood. Journal of consulting and clinical psychology. 1997;65(2):269.
27. Whatley J, Gillespie W, Honig J, Walsh M, Blackburn A, Blackburn G. Does the amount of endurance exercise in combination with weight training and a very-low-energy diet affect resting metabolic rate and body composition? The American journal of clinical nutrition. 1994;59(5):1088-92.
28. Pritchard JE, Nowson CA, Wark JD. A worksite program for overweight middle-aged men achieves lesser weight loss with exercise than with dietary change. Journal of the American Dietetic Association. 1997;97(1):37-42.
29. Stefanick ML, Mackey S, Sheehan M, Ellsworth N, Haskell WL, Wood PD. Effects of diet and exercise in men and postmenopausal women with low levels of HDL cholesterol and high levels of LDL cholesterol. New England Journal of Medicine. 1998;339(1):12-20.
30. Wood PD, Stefanick ML, Dreon DM, Frey-Hewitt B, Garay SC, Williams PT, et al. Changes in plasma lipids and lipoproteins in overweight men during weight loss through dieting as compared with exercise. New England Journal of Medicine. 1988;319(18):1173-9.
31. Gordon NF, Scott CB, Levine BD. Comparison of single versus multiple lifestyle interventions: are the antihypertensive effects of exercise training and diet-induced weight loss additive? The American journal of cardiology. 1997;79(6):763-7.
32. Van Aggel-Leijssen DP, Saris WH, Hul GB, Van Baak MA. Short-term effects of weight loss with or without low-intensity exercise training on fat metabolism in obese men. The American journal of clinical nutrition. 2001;73(3):523-31.
33. Wing RR, Venditti E, Jakicic JM, Polley BA, Lang W. Lifestyle intervention in overweight individuals with a family history of diabetes. Diabetes care. 1998;21(3):350-9.
34. Anderssen S, Hjermann I, Urdal P, Torjesen P, Holme I. Improved carbohydrate metabolism after physical training and dietary intervention in individuals with the

'atherothrombogenic syndrome'. Oslo Diet and Exercise Study (ODES). A randomized trial. Journal of internal medicine. 1996;240(4):203-9.
35. Schwartz RS. The independent effects of dietary weight loss and aerobic training on high density lipoproteins and apolipoprotein AI concentrations in obese men. Metabolism. 1987;36(2):165-71.
36. Hellenius M, de Faire U, Berglund B, Hamsten A, Krakau I. Diet and exercise are equally effective in reducing risk for cardiovascular disease. Results of a randomized controlled study in men with slightly to moderately raised cardiovascular risk factors. Atherosclerosis. 1993;103(1):81-91.
37. Cox KL, Puddey IB, Morton AR, Burke V, Beilin LJ, McAleer M. Exercise and weight control in sedentary overweight men: effects on clinic and ambulatory blood pressure. Journal of hypertension. 1996;14(6):779-90.
38. Wood PD, Stefanick ML, Williams PT, Haskell WL. The effects on plasma lipoproteins of a prudent weight-reducing diet, with or without exercise, in overweight men and women. New England Journal of Medicine. 1991;325(7):461-6.
39. Cox KL, Burke V, Morton AR, Beilin LJ, Puddey IB. Independent and additive effects of energy restriction and exercise on glucose and insulin concentrations in sedentary overweight men. The American journal of clinical nutrition. 2004;80(2):308-16.
40. Jakicic JM, Marcus BH, Gallagher KI, Napolitano M, Lang W. Effect of exercise duration and intensity on weight loss in overweight, sedentary women: a randomized trial. Jama. 2003;290(10):1323-30.
41. Neumark-Sztainer D, Kaufmann N, Berry E. Physical activity within a community-based weight control program: program evaluation and predictors of success. Public health reviews. 1994;23(3):237-51.
42. Van Aggel-Leijssen DP, Saris WH, Wagenmakers AJ, Senden JM, Van Baak MA. Effect of exercise training at different intensities on fat metabolism of obese men. Journal of Applied Physiology. 2002;92(3):1300-9.
43. Raz I, Hauser E, Bursztyn M. Moderate exercise improves glucose metabolism in uncontrolled elderly patients with non-insulin-dependent diabetes mellitus. Israel journal of medical sciences. 1994;30(10):766-70.
44. Gillett PA, Eisenman PA. The effect of intensity controlled aerobic dance exercise on aerobic capacity of middle-aged, overweight women. Research in nursing & health. 1987;10(6):383-90.
45. Manning JM, Dooly-Manning CR, White K, Kampa I, Silas S, Kesselhaut M, et al. Effects of a resistive training program on lipoprotein--lipid levels in obese women. Medicine and Science in Sports and Exercise. 1991;23(11):1222-6.
46. Wallace MB, Mills BD, Browning CL. Effects of cross-training on markers of insulin resistance/hyperinsulinemia. Medicine and Science in Sports and Exercise. 1997;29(9):1170-5.
47. Andersen RE, Wadden TA, Bartlett SJ, Zemel B, Verde TJ, Franckowiak SC. Effects of lifestyle activity vs structured aerobic exercise in obese women: a randomized trial. Jama. 1999;281(4):335-40.
48. Wirth A, Diehm C, Hanel W, Welte J, Vogel I. Training-induced changes in serum lipids, fat tolerance, and adipose tissue metabolism in patients with hypertriglyceridemia. Atherosclerosis. 1985;54(3):263-71.
49. Van Aggel-Leijssen DP, Saris WH, Wagenmakers AJ, Hul GB, Van Baak MA. The Effect of Low-Intensity Exercise Training on Fat Metabolism of Obese Women. Obesity Research. 2001;9(2):86-96.

50. Ainsworth BE, Jacobs DR, Jr., Leon AS. Validity and reliability of self-reported physical activity status: the Lipid Research Clinics questionnaire. Med Sci Sports Exerc. 1993 Jan;25(1):92-8.
51. Washburn RA, Montoye HJ. The assessment of physical activity by questionnaire. Am J Epidemiol. 1986 Apr;123(4):563-76.
52. Welk G. Physical activity assessments for health-related research: Human Kinetics Publishers; 2002.
53. Lichtman SW, Pisarska K, Berman ER, Pestone M, Dowling H, Offenbacher E, et al. Discrepancy between self-reported and actual caloric intake and exercise in obese subjects. New England Journal of Medicine. 1992;327(27):1893-8.
54. Jakicic JM, Polley BA, Wing RR. Accuracy of self-reported exercise and the relationship with weight loss in overweight women. Medicine and Science in Sports and Exercise. 1998;30(4):634-8.
55. Swain DP, Leutholtz BC. Exercise prescription: a case study approach to the ACSM guidelines: Human Kinetics; 2007.
56. Aittasalo M, Miilunpalo S, Suni J. The effectiveness of physical activity counseling in a work-site setting. A randomized, controlled trial. Patient Educ Couns. 2004 Nov;55(2):193-202.
57. Schoeller D, Van Santen E. Measurement of energy expenditure in humans by doubly labeled water method. Journal of Applied Physiology. 1982;53(4):955-9.
58. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Jr., Tudor-Locke C, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. Med Sci Sports Exerc. 2011 Aug;43(8):1575-81.
59. Edholm O. The assessment of habitual activity in health disease. Proc Beitosolen Symp Oslo. 1966:187-97.
60. Paffenbarger RS, Wing AL, Hyde RT. Physical activity as an index of heart attack risk in college alumni. American journal of epidemiology. 1978;108(3):161-75.
61. Taylor HL, Jacobs Jr DR, Schucker B, Knudsen J, Leon AS, Debacker G. A questionnaire for the assessment of leisure time physical activities. Journal of chronic diseases. 1978;31(12):741-55.
62. Ainsworth BE, Leon AS, Richardson MT, Jacobs DR, Paffenbarger Jr R. Accuracy of the college alumnus physical activity questionnaire. Journal of clinical epidemiology. 1993;46(12):1403-11.
63. Jacobs Jr DR, Ainsworth BE, Hartman TJ, Leon AS. A simultaneous evaluation of 10 commonly used physical activity questionnaires. Medicine and Science in Sports and Exercise. 1993;25(1):81-91.
64. Boon RM, Hamlin MJ, Steel GD, Ross JJ. Validation of the New Zealand physical activity questionnaire (NZPAQ-LF) and the international physical activity questionnaire (IPAQ-LF) with accelerometry. British journal of sports medicine. 2010;44(10):741-6.
65. Kalkwarf H, Haas J, Belko A, Roach R, Roe D. Accuracy of heart-rate monitoring and activity diaries for estimating energy expenditure. The American journal of clinical nutrition. 1989;49(1):37-43.
66. American College of Sports Medicine. ACSM'S metabolic calculations handbook: Lippincott Williams & Wilkins; 2006.
67. American College of Sports Medicine. ACSM's Resources for the Health Fitness Specialist: Lippincott Williams & Wilkins; 2013.
68. Gottschlich MM. The science and practice of nutrition support: a case-based core curriculum: Kendall Hunt; 2001.

69. Bouten CVC, Koekkoek KTM, Verduin M, Kodde R, Janssen JD. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. IEEE Trans Biomed Eng. 1997 Mar;44(3):136-47.
70. Lim M, Du H, Su C, Jin W. A micromachined piezoresistive accelerometer with high sensitivity: design and modelling. Microelectronic Engineering. 1999;49(3):263-72.
71. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc. 1998 May;30(5):777-81.
72. Jakicic JM, Winters C, Lagally K, Ho J, Robertson RJ, Wing RR. The accuracy of the TriTrac-R3D accelerometer to estimate energy expenditure. Med Sci Sports Exerc. 1999 May;31(5):747-54.
73. Swartz AM, Strath SJ, Bassett DR, Jr., O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. Med Sci Sports Exerc. 2000 Sep;32(9 Suppl):S450-6.
74. Terrier P, Aminian K, Schutz Y. Can accelerometry accurately predict the energy cost of uphill/downhill walking? Ergonomics. 2001 Jan 15;44(1):48-62.
75. Lopez-Alarcon M, Merrifield J, Fields DA, Hilario-Hailey T, Franklin FA, Shewchuk RM, et al. Ability of the actiwatch accelerometer to predict free-living energy expenditure in young children. Obes Res. 2004 Nov;12(11):1859-65.
76. Brooks AG, Gunn SM, Withers RT, Gore CJ, Plummer JL. Predicting walking METs and energy expenditure from speed or accelerometry. Med Sci Sports Exerc. 2005 Jul;37(7):1216-23.
77. Crouter SE, Churilla JR, Bassett DR, Jr. Estimating energy expenditure using accelerometers. Eur J Appl Physiol. 2006 Dec;98(6):601-12.
78. Assah FK, Ekelund U, Brage S, Corder K, Wright A, Mbanya JC, et al. Predicting physical activity energy expenditure using accelerometry in adults from sub-Sahara Africa. Obesity (Silver Spring). 2009 Aug;17(8):1588-95.
79. Foerster F, Fahrenberg J. Motion pattern and posture: correctly assessed by calibrated accelerometers. Behav Res Methods Instrum Comput. 2000 Aug;32(3):450-7.
80. Allen FR, Ambikairajah E, Lovell NH, Celler BG. Classification of a known sequence of motions and postures from accelerometry data using adapted Gaussian mixture models. Physiol Meas. 2006 Oct;27(10):935-51.
81. Juang C-F, Chang C-M. Human body posture classification by a neural fuzzy network and home care system application. Ieee Transactions on Systems Man and Cybernetics Part a-Systems and Humans. 2007 Nov;37(6):984-94.
82. Lee SH, Park HD, Hong SY, Lee KJ, Kim YH, editors. A study on the activity classification using a triaxial accelerometer. 25th Annual Conf of the IEEE Engineering in Medicine and Biology Society (Cancun); 2003.
83. Mathie MJ, Celler BG, Lovell NH, Coster AC. Classification of basic daily movements using a triaxial accelerometer. Med Biol Eng Comput. 2004 Sep;42(5):679-87.
84. Parkka J, Ermes M, Korpipaa P, Mantyjarvi J, Peltola J, Korhonen I. Activity classification using realistic data from wearable sensors. IEEE Trans Inf Technol Biomed. 2006 Jan;10(1):119-28.
85. Bonomi AG, Goris AH, Yin B, Westerterp KR. Detection of type, duration, and intensity of physical activity using an accelerometer. Med Sci Sports Exerc. 2009 Sep;41(9):1770-7.
86. Mannini A, Sabatini AM. Machine learning methods for classifying human physical activity from on-body accelerometers. Sensors (Basel). 2010;10(2):1154-75.
87. Oshima Y, Kawaguchi K, Tanaka S, Ohkawara K, Hikihara Y, Ishikawa-Takata K, et al. Classifying household and locomotive activities using a triaxial accelerometer. Gait Posture. 2010 Mar;31(3):370-4.

88. Mannini A, Sabatini AM. Accelerometry-based classification of human activities using Markov modeling. Comput Intell Neurosci. 2011;2011:647858.
89. Edgar SR, Fulk GD, Sazonov ES. Recognition of household and athletic activities using SmartShoe. Conf Proc IEEE Eng Med Biol Soc. 2012;2012:6382-5.
90. Zijlstra W, Hof AL. Assessment of spatio-temporal gait parameters from trunk accelerations during human walking. Gait Posture. 2003 Oct;18(2):1-10.
91. Brandes M, Zijlstra W, Heikens S, van Lummel R, Rosenbaum D. Accelerometry based assessment of gait parameters in children. Gait & Posture. 2006 Dec;24(4):482-6.
92. Lee S-W, Mase K, Kogure K, editors. Detection of spatio-temporal gait parameters by using wearable motion sensors. Engineering in Medicine and Biology Society, 2005 IEEE-EMBS 2005 27th Annual International Conference of the; 2006: IEEE.
93. Kavanagh JJ, Menz HB. Accelerometry: A technique for quantifying movement patterns during walking. Gait Posture. 2008 Jan 4;28:1-15.
94. Bugane F, Benedetti MG, Casadio G, Attala S, Biagi F, Manca M, et al. Estimation of spatial-temporal gait parameters in level walking based on a single accelerometer: Validation on normal subjects by standard gait analysis. Computer Methods and Programs in Biomedicine. 2012 Oct;108(1):129-37.
95. Lindemann U, Hock A, Stuber M, Keck W, Becker C. Evaluation of a fall detector based on accelerometers: a pilot study. Medical & Biological Engineering & Computing. 2005 Sep;43(5):548-51.
96. Kangas M, Konttila A, Winblad I, Jamsa T, Ieee. Determination of simple thresholds for accelerometry-based parameters for fall detection. 2007 Annual International Conference of the Ieee Engineering in Medicine and Biology Society, Vols 1-162007. p. 1367-70.
97. Bourke AK, O'Donovan KJ, Nelson J, Olaighin GM, Ieee. Fall-detection through vertical velocity thresholding using a tri-axial accelerometer characterized using an optical motion-capture system. 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vols 1-82008. p. 2832-5.
98. Kang DW, Choi JS, Lee JW, Chung SC, Park SJ, Tack GR. Real-time elderly activity monitoring system based on a tri-axial accelerometer. Disabil Rehabil Assist Technol. 2010 Jul;5(4):247-53.
99. Janssen W, Kulcu DG, Horemans H, Stam HJ, Bussmann J. Sensitivity of accelerometry to assess balance control during sit-to-stand movement. Neural Systems and Rehabilitation Engineering, IEEE Transactions on. 2008;16(5):479-84.
100. Mayagoitia RE, Lötters JC, Veltink PH, Hermens H. Standing balance evaluation using a triaxial accelerometer. Gait & Posture. 2002;16(1):55-9.
101. De Vries SI, Garre FG, Engbers LH, Hildebrandt VH, Van Buuren S. Evaluation of neural networks to identify types of activity using accelerometers. Med Sci Sports Exerc. 2011 Jan;43(1):101-7.
102. Lyden K, Kozey SL, Staudenmeyer JW, Freedson PS. A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. Eur J Appl Physiol. 2010 Feb;111(2):187-201.
103. Rothney MP, Neumann M, Beziat A, Chen KY. An artificial neural network model of energy expenditure using nonintegrated acceleration signals. J Appl Physiol. 2007 Oct;103(4):1419-27.
104. Staudenmayer J, Pober D, Crouter S, Bassett D, Freedson P. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. J Appl Physiol. 2009 Oct;107(4):1300-7.
105. Hänggi JM, Phillips LR, Rowlands AV. Validation of the GT3X ActiGraph in children and comparison with the GT1M ActiGraph. Journal of Science and Medicine in Sport. 2013;16(1):40-4.

106. Robusto KM, Trost SG. Comparison of three generations of ActiGraph™ activity monitors in children and adolescents. J Sports Sci. 2012;30(13):1429-35.
107. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. Journal of Science and Medicine in Sport. 2011;14(5):411-6.
108. Vanhelst J, Mikulovic J, Bui-Xuan G, Dieu O, Blondeau T, Fardy P, et al. Comparison of two ActiGraph accelerometer generations in the assessment of physical activity in free living conditions. BMC research notes. 2012;5(1):187.
109. van Hees VT, Renström F, Wright A, Gradmark A, Catt M, Chen KY, et al. Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. PLoS One. 2011;6(7):e22922.
110. Baek J, Lee G, Park W, Yun BJ, editors. Accelerometer signal processing for user activity detection. Knowledge-Based Intelligent Information and Engineering Systems, Pt 3, Proceedings; 2004.
111. Bao L, Intille SS. Activity recognition from user-annotated acceleration data. Pervasive Computing, Lecture Notes in Computer Science. 2004;3001:1-17.
112. He Z, Jin L, editors. Activity recognition from acceleration data based on discrete consine transform and SVM. Systems, Man and Cybernetics, 2009 SMC 2009 IEEE International Conference on; 2009: IEEE.
113. Huynh T, Schiele B, editors. Towards less supervision in activity recognition from wearable sensors. Proc of the 10th IEEE International Symposium on Wearable Computers (Montreaux); 2006.
114. Chen KY, Sun M. Improving energy expenditure estimation by using a triaxial accelerometer. J Appl Physiol. 1997 Dec;83(6):2112-22.
115. Montoye HJ, Washburn R, Servais S, Ertl A, Webster JG, Nagle FJ. Estimation of energy expenditure by a portable accelerometer. Med Sci Sports Exerc. 1983;15(5):403-7.
116. Bouten CV, Westerterp KR, Verduin M, Janssen JD. Assessment of energy expenditure for physical activity using a triaxial accelerometer. Med Sci Sports Exerc. 1994 Dec;26(12):1516-23.
117. Ermes M, Parkka J, Mantyjarvi J, Korhonen I. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. IEEE Trans Inf Techn Biomed. 2008 Jan;12(1):20-6.
118. Vathsangam H, Emken BA, Spruijt-Metz D, Sukhatme GS, editors. Toward free-living walking speed estimation using Gaussian process-based regression with on-body accelerometers and gyroscopes. Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2010 4th International Conference on-NO PERMISSIONS; 2010: IEEE.
119. Spyropoulos P, Pisciotta JC, Pavlou KN, Cairns MA, Simon SR. BIOMECHANICAL GAIT ANALYSIS IN OBESE MEN. Archives of physical medicine and rehabilitation. 1991 Dec;72(13):1065-70.
120. Lai PPK, Leung AKL, Li ANM, Zhang M. Three-dimensional gait analysis of obese adults. Clinical Biomechanics. 2008 2008;23:S2-S6.
121. de Souza SAF, Faintuch J, Valezi AC, Sant'Anna AF, Gama-Rodrigues JJ, Fonseca ICD, et al. Gait cinematic analysis in morbidly obese patients. Obesity Surgery. 2005 Oct;15(9):1238-42.
122. da Silva-Hamu TCD, Formiga CKMR, Gervasio FM, Ribeiro DM, Christofoletti G, de Franca Barros J. The impact of obesity in the kinematic parameters of gait in young women. International journal of general medicine. 2013 2013 Jun;6:507-13.
123. Peyrot N, Thivel D, Isacco L, Morin J-B, Duche P, Belli A. Do mechanical gait parameters explain the higher metabolic cost of walking in obese adolescents? Journal of Applied Physiology. 2009 Jun;106(6):1763-70.

124. Haight DJ, Lerner ZF, Board WJ, Browning RC. A Comparison of Slow, Uphill and Fast, Level Walking on Lower Extremity Biomechanics and Tibiofemoral Joint Loading in Obese and Nonobese Adults. Journal of Orthopaedic Research. [Article]. 2014 Feb;32(2):324-30.
125. Messier SP, Davies AB, Moore DT, Davis SE, Pack RJ, Kazmar SC. Severe obesity: effects on foot mechanics during walking. Foot & Ankle International. 1994;15(1):29-34.
126. Mignardot JB, Olivier I, Promayon E, Nougier V. Origins of Balance Disorders during a Daily Living Movement in Obese: Can Biomechanical Factors Explain Everything? Plos One. [Article]. 2013 Apr;8(4).
127. Galli M, Crivellini M, Sibella F, Montesano A, Bertocco P, Parisio C. Sit-to-stand movement analysis in obese subjects. International Journal of Obesity. [Article]. 2000 Nov;24(11):1488-92.
128. Sibella F, Galli M, Romei M, Montesano A, Crivellini M. Biomechanical analysis of sit-to-stand movement in normal and obese subjects. Clinical Biomechanics. [Article]. 2003 Oct;18(8):745-50.
129. Runhaar J, Koes BW, Clockaerts S, Bierma-Zeinstra SMA. A systematic review on changed biomechanics of lower extremities in obese individuals: a possible role in development of osteoarthritis. Obesity Reviews. [Review]. 2011 Dec;12(12):1071-82.
130. Jeffery RW, Wing RR, Thorson C, Burton LR. Use of personal trainers and financial incentives to increase exercise in a behavioral weight-loss program. Journal of consulting and clinical psychology. 1998;66(5):777.
131. Leutholtz BC, Keyser RE, Heusner WW, Wendt VE, Rosen L. Exercise training and severe caloric restriction: effect on lean body mass in the obese. Arch Phys Med Rehabil. 1995;76(1):65-70.
132. Stensel D, Brooke-Wavell K, Hardman A, Jones P, Norgan N. The influence of a 1-year programme of brisk walking on endurance fitness and body composition in previously sedentary men aged 42–59 years. European journal of applied physiology and occupational physiology. 1994;68(6):531-7.
133. Irwin ML, Yasui Y, Ulrich CM, Bowen D, Rudolph RE, Schwartz RS, et al. Effect of exercise on total and intra-abdominal body fat in postmenopausal women: a randomized controlled trial. Jama. 2003;289(3):323-30.
134. Cox KL, Burke V, Morton AR, Beilin LJ, Puddey IB. The independent and combined effects of 16 weeks of vigorous exercise and energy restriction on body mass and composition in free-living overweight men—a randomized controlled trial. Metabolism. 2003;52(1):107-15.
135. Duda RO, Hart PE, Stork DG. Pattern classification, 2nd Ed. New York: John Wiley & Sons; 2001.
136. Bishop CM. Pattern recognition and machine learning: springer New York; 2006.
137. Ghahramani Z. Unsupervised learning. Advanced Lectures on Machine Learning: Springer; 2004. p. 72-112.
138. Hinton GE, Sejnowski TJ. Unsupervised learning: foundations of neural computation: MIT press; 1999.
139. Gyllensten IC, Bonomi AG. Identifying types of physical activity with a single accelerometer: evaluating laboratory-trained algorithms in daily life. IEEE Trans Biomed Eng. 2011 Sep;58(9):2656-63.
140. Pober DM, Staudenmayer J, Raphael C, Freedson PS. Development of novel techniques to classify physical activity mode using accelerometers. Med Sci Sports Exerc. 2006 Sep;38(9):1626-34.

141. Huynh T, Schiele B, editors. Unsupervised Discovery of Structure in Activity Data Using Multiple Eigenspaces. Second International Workshop on Location- and Context-Awareness (Dublin); 2006.
142. Nguyen A, Moore D, McCowan I, editors. Unsupervised clustering of free-living human activities using ambulatory accelerometry. 29th Annual Conf of the IEEE Engineering in Medicine and Biology Society (Lyon); 2007.
143. Krause A, Siewiorek DP, Smailagic A, Farringdon J. Unsupervised, dynamic identification of physiological and activity context in wearable computing. 2003.
144. Ravi N, Dandekar N, Mysore P, Littman M, editors. Activity recognition from accelerometer data. Proceedings of the 7th Innovative Applications of Artificial Intelligence Conference (California); 2005.
145. Mathie MJ, Coster AC, Lovell NH, Celler BG. Detection of daily physical activities using a triaxial accelerometer. Med Biol Eng Comput. 2003 May;41(3):296-301.
146. Lester J, Choudhury T, Borriello G. A practical approach to recognizing physical activities. Pervasive Computing, Lecture Notes in Computer Science. 2006;3968:1-16.
147. Huynh T, Schiele B, editors. Analyzing features for activity recognition. Proc Conf Smart objects and ambient intelligence: innovative context-aware services: usages and technologies (Grenoble); 2005.
148. Lyons G, Culhane K, Hilton D, Grace P, Lyons D. A description of an accelerometer-based mobility monitoring technique. Medical Engineering & Physics. 2005;27(6):497-504.
149. Walker JS. A primer on wavelets and their scientific applications: CRC press; 1999.
150. Graps A. An introduction to wavelets. Computational Science & Engineering, IEEE. 1995;2(2):50-61.
151. Rioul O, Vetterli M. Wavelets and signal processing. IEEE Signal Processing Magazine. 1991;8(LCAV-ARTICLE-1991-005):14-38.
152. Sekine M, Tamura T, Togawa T, Fukui Y. Classification of waist-acceleration signals in a continuous walking record. Med Eng Phys. 2000 May;22(4):285-91.
153. Wang N, Ambikairajah E, Lovell NH, Celler BG. Accelerometry based classification of walking patterns using time-frequency analysis. Conf Proc IEEE Eng Med Biol Soc. 2007;2007:4899-902.
154. Nyan MN, Tay FE, Seah KH, Sitoh YY. Classification of gait patterns in the time-frequency domain. J Biomech. 2006;39(14):2647-56.
155. Palmerini L, Rocchi L, Mellone S, Valzania F, Chiari L. Feature selection for accelerometer-based posture analysis in Parkinson's disease. IEEE Trans Inf Technol Biomed. 2011 May;15(3):481-90.
156. Janecek A, Gansterer WN, Demel M, Ecker G. On the Relationship Between Feature Selection and Classification Accuracy. Journal of Machine Learning Research-Proceedings Track. 2008;4:90-105.
157. Long X, Yin B, Aarts RM. Single-accelerometer-based daily physical activity classification. Conf Proc IEEE Eng Med Biol Soc. 2009;2009:6107-10.
158. Molina LC, Belanche L, Nebot À, editors. Feature selection algorithms: A survey and experimental evaluation. Data Mining, 2002 ICDM 2003 Proceedings 2002 IEEE International Conference on; 2002: IEEE.
159. Preece SJ, Goulermas JY, Kenney LPJ, Howard D, Meijer K, Crompton R. Activity identification using body-mounted sensors—a review of classification techniques. Physiological measurement. 2009;30:R1.
160. Foerster F, Smeja M, Fahrenberg J. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. Comp Hum Behav. 1999 Sep;15(5):571-83.

161. Foerster F, Fahrenberg J. Motion pattern and posture: correctly assessed by calibrated accelerometers. Behavior research methods, instruments, & computers. 2000;32(3):450-7.
162. Bussmann JB, Martens WL, Tulen JH, Schasfoort FC, van den Berg-Emons HJ, Stam HJ. Measuring daily behavior using ambulatory accelerometry: The Activity Monitor. Behav Res Methods Instrum Comput. 2001 Aug;33(3):349-56.
163. The Pennsylvania State University. Quadratic Discriminant Analysis (QDA). 2014 [cited 2014 6 November 2014]; Available from: https://onlinecourses.science.psu.edu/stat557/node/43.
164. Goulermas JY, Findlow AH, Nester CJ, Howard D, Bowker P. Automated design of robust discriminant analysis classifier for foot pressure lesions using kinematic data. Biomedical Engineering, IEEE Transactions on. 2005;52(9):1549-62.
165. The Pennsylvania State University. Linear Discriminant Analysis. 2014 [cited 2014 6 November 2014]; Available from: https://onlinecourses.science.psu.edu/stat557/node/37.
166. Culhane KM, Lyons GM, Hilton D, Grace PA, Lyons D. Long-term mobility monitoring of older adults using accelerometers in a clinical environment. Clin Rehabil. 2004 May;18(3):335-43.
167. Maurer U, Rowe A, Smailagic A, Siewiorek D. Location and activity recognition using eWatch: A wearable sensor platform. Ambient Intelligence in Everday Life, Lecture Notes in Computer Science. 2006;3864:86-102.
168. Krause A, Ihmig M, Rankin E, Leong D, Gupta S, Siewiorek D, et al., editors. Trading off prediction accuracy and power consumption for context-aware wearable computing. Proc of the 9th IEEE International Symposium on Wearable Computers; 2005.
169. Salarian A, Russmann H, Vingerhoets FJ, Burkhard PR, Aminian K. Ambulatory monitoring of physical activities in patients with Parkinson's disease. Biomedical Engineering, IEEE Transactions on. 2007;54(12):2296-9.
170. Lee S, Mase K. Activity and location recognition using wearable sensors. IEEE Perv Comp. 2002;1(3):25-32.
171. Lester J, Choudhury T, Kern N, Borriello G, Hannaford B, editors. A hybrid discriminative/generative approach for modeling human activities. 19th International Joint Conference on Artificial Intelligence (Edinburgh); 2005.
172. Maurer U, Smailagic A, Siewiorek DP, Deisher M, editors. Activity recognition and monitoring using multiple sensors on different body positions. Wearable and Implantable Body Sensor Networks, 2006 BSN 2006 International Workshop on; 2006: IEEE.
173. Ruch N, Rumo M, Mader U. Recognition of activities in children by two uniaxial accelerometers in free-living conditions. Eur J Appl Physiol. 2011 Aug;111(8):1917-27.
174. Bussmann HB, Reuvekamp PJ, Veltink PH, Martens WL, Stam HJ. Validity and reliability of measurements obtained with an "activity monitor" in people with and without a transtibial amputation. Phys Ther. 1998 Sep;78(9):989-98.
175. Godfrey A, Bourke AK, Olaighin GM, van de Ven P, Nelson J. Activity classification using a single chest mounted tri-axial accelerometer. Med Eng Phys. 2011 Nov;33(9):1127-35.
176. Pirttikangas P, Fujinami K, Nakajima T. Feature Selection and Activity Recognition from Wearable Sensors. Ubiquitous Computing Systems, Lecture Notes in Computer Science. 2006;4239:516-27.
177. Van Laerhoven K, Schmidt A, Gellersen H-W, editors. Multi-sensor context aware clothing. Wearable Computers, 2002(ISWC 2002) Proceedings Sixth International Symposium on; 2002: IEEE.
178. Olguın DO, Pentland AS. Human activity recognition: Accuracy across common locations for wearable sensors. 2006.

179. Heinz EA, Kunze KS, Sulistyo S, Junker H, Lukowicz P, Troster G. Experimental Evaluation of Variations in Primary Features Used for Accelerometric Context Recognition. Ambient Intelligence, Lecture Notes in Computer Science. 2003;2875:252-63.
180. Oudre L, Jakubowicz J, Bianchi P, Simon C. Classification of periodic activities using the Wasserstein distance. IEEE Trans Biomed Eng. 2012 Jun;59(6):1610-9.
181. Zhang K, Werner P, Sun M, Pi-Sunyer FX, Boozer CN. Measurement of human daily physical activity. Obes Res. 2003 Jan;11(1):33-40.
182. Preece SJ, Goulermas JY, Kenney LPJ, Howard D. A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. IEEE Trans Biomed Eng. 2008;56:871-9.
183. Pate RR, Pratt M, Blair SN, Haskell WL, Macera CA, Bouchard C, et al. Physical activity and public health: a recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. Jama. 1995;273(5):402-7.
184. Asikainen TM, Miilunpalo S, Oja P, Rinne M, Pasanen M, Vuori I. Walking trials in postmenopausal women: effect of one vs two daily bouts on aerobic fitness. Scand J Med Sci Sports. 2002;12(2):99-105.
185. Kobayashi Y, Hosoi T, Takeuchi T, AOKI S. Benefits of a convenient, self-regulated 6-month walking program in sedentary, middle-aged women. 体力科學. 2001;50(3):313-23.
186. Murphy MH, Nevill AM, Murtagh EM, Holder RL. The effect of walking on fitness, fatness and resting blood pressure: a meta-analysis of randomised, controlled trials. Prev Med. 2007;44(5):377-85.
187. Morris JN, Hardman AE. Walking to health. Sports Medicine. 1997;23(5):306-32.
188. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. Medicine and Science in Sports and Exercise. 2000;32(9; SUPP/1):S498-S504.
189. Schutz Y, Chambaz A. Could a satellite-based navigation system (GPS) be used to assess the physical activity of individuals on earth? European journal of clinical nutrition. 1997;51(5):338-9.
190. Le Faucheur A, Abraham P, Jaquinandi V, Bouye P, Saumet JL, Noury-Desvaux B. Study of human outdoor walking with a low-cost GPS and simple spreadsheet analysis. Medicine and Science in Sports and Exercise. [Article]. 2007 Sep;39(9):1570-8.
191. Li Q, Young M, Naing V, Donelan JM, Ieee. Walking Speed and Slope Estimation Using Shank-mounted Inertial Measurement Units. 2009 Ieee 11th International Conference on Rehabilitation Robotics, Vols 1 and 2. New York: Ieee; 2009. p. 978-83.
192. Salarian A, Russmann H, Vingerhoets FJ, Dehollaini C, Blanc Y, Burkhard PR, et al. Gait assessment in Parkinson's disease: toward an ambulatory system for long-term monitoring. Biomedical Engineering, IEEE Transactions on. 2004;51(8):1434-43.
193. Li Q, Young M, Naing V, Donelan JM. Walking speed estimation using a shank-mounted inertial measurement unit. Journal of Biomechanics. [Article]. 2010 May;43(8):1640-3.
194. Laudanski A, Yang SZ, Li QG, Ieee. A Concurrent Comparison of Inertia Sensor-Based Walking Speed Estimation Methods. 2011 Annual International Conference of the Ieee Engineering in Medicine and Biology Society. New York: Ieee; 2011. p. 3484-7.
195. Miyazaki S. Long-term unrestrained measurement of stride length and walking velocity utilizing a piezoelectric gyroscope. Biomedical Engineering, IEEE Transactions on. 1997;44(8):753-9.
196. Aminian K, Najafi B, Bula C, Leyvraz PF, Robert P. Spatio-temporal parameters of gait measured by an ambulatory system using miniature gyroscopes. J Biomech. 2002 May;35(5):689-99.

197. Sagawa K, Inooka H, Satoh Y, editors. Non-restricted measurement of walking distance. Systems, Man, and Cybernetics, 2000 IEEE International Conference on; 2000: IEEE.
198. Aminian K, Robert P, Jequier E, Schutz Y. Estimation of speed and incline of walking using neural network. Instrumentation and Measurement, IEEE Transactions on. 1995;44(3):743-6.
199. Panagiota A, Layal S, Stefan H, editors. Assessment of Human Gait Speed and Energy Expenditure Using a Single Triaxial Accelerometer. Wearable and Implantable Body Sensor Networks (BSN), 2012 Ninth International Conference on; 2012: IEEE.
200. Schutz Y, Weinsier S, Terrier P, Durrer D. A new accelerometric method to assess the daily walking practice. Int J Obes Relat Metab Disord. 2002 Jan;26(1):111-8.
201. Zijlstra W. Assessment of spatio-temporal parameters during unconstrained walking. Eur J Appl Physiol. 2004 Jun;92(1-2):39-44.
202. GURFINKEL VS, OSEVETS M. Dynamics of the vertical posture in man. Biophysics 1972;17:496-506.
203. Kuo AD. The six determinants of gait and the inverted pendulum analogy: A dynamic walking perspective. Human movement science. 2007;26(4):617-56.
204. Song Y, Shin S, Kim S, Lee D, Lee KH. Speed estimation from a tri-axial accelerometer using neural networks. Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Conference. 2007 2007;2007:3224-7.
205. Lobet S, Detrembleur C, Massaad F, Hermans C. Three-dimensional gait analysis can shed new light on walking in patients with haemophilia. The Scientific World Journal. 2013;2013.
206. Bishop E, Li Q, Ieee. Walking speed estimation using shank-mounted accelerometers. In: Rakotondrabe M, Ivan IA, editors. 2010 Ieee International Conference on Robotics and Automation2010. p. 5096-101.
207. Barnett A, Cerin E. Individual calibration for estimating free-living walking speed using the MTI monitor. Med Sci Sports Exerc. 2006 Apr;38(4):761-7.
208. Yeoh W-S, Pek I, Yong Y-H, Chen X, Waluyo AB, Ieee. Ambulatory Monitoring of Human Posture and Walking Speed Using Wearable Accelerometer Sensors. 2008 30th Annual International Conference of the Ieee Engineering in Medicine and Biology Society, Vols 1-82008. p. 5184-7.
209. He Z, Zhang W. Estimation of Walking Speed Using Accelerometer and Artificial Neural Networks. In: Yu YX, Yu ZT, Zhao JY, editors. Computer Science for Environmental Engineering and Ecoinformatics, Pt 22011. p. 42-7.
210. Mannini A, Sabatini AM, editors. On-line classification of human activity and estimation of walk-run speed from acceleration data using support vector machines. Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE; 2011: IEEE.
211. Catenacci VA, Grunwald GK, Ingebrigtsen JP, Jakicic JM, McDermott MD, Phelan S, et al. Physical Activity Patterns Using Accelerometry in the National Weight Control Registry. Obesity. [Article]. 2011 Jun;19(6):1163-70.
212. Chen KY, Bassett DR, Jr. The technology of accelerometry-based activity monitors: current and future. Med Sci Sports Exerc. 2005 Nov;37(11 Suppl):S490-500.
213. Nichols JF, Morgan CG, Sarkin JA, Sallis JF, Calfas KJ. Validity, reliability, and calibration of the Tritrac accelerometer as a measure of physical activity. Medicince and Science in Sport and Exercise. 1999 Jun;31(6):908-12.
214. Heil DP. Predicting activity energy expenditure using the Actical® activity monitor. Research Quarterly for Exercise and Sport. 2006;77(1):64-80.

215. Yngve A, Nilsson A, Sjostrom M, Ekelund U. Effect of monitor placement and of activity setting on the MTI accelerometer output. Medicine and Science in Sports and Exercise. 2003;35(2):320-6.
216. Leenders N, Nelson T, Sherman W. Ability of different physical activity monitors to detect movement during treadmill walking. Int J Sports Med. 2003;24:43-50.
217. Hendelman D, Miller K, Baggett C, Debold E, Freedson P. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. Med Sci Sports Exerc. 2000 Sep;32(9 Suppl):S442-9.
218. Nichols JF, Morgan CG, Chabot LE, Sallis JF, Calfas KJ. Assessment of physical activity with the Computer Science and Applications, Inc., accelerometer: laboratory versus field validation. Research Quarterly for Exercise and Sport. 2000;71(1):36-43.
219. Foster RC, Lanningham-Foster LM, Manohar C, McCrady SK, Nysse LJ, Kaufman KR, et al. Precision and accuracy of an ankle-worn accelerometer-based pedometer in step counting and energy expenditure. Prev Med. 2005;41(3):778-83.
220. Leenders N, Sherman WM, Nagaraja H, Kien CL. Evaluation of methods to assess physical activity in free-living conditions. Medicine and Science in Sports and Exercise. 2001;33(7):1233-40.
221. Schoeller DA. Recent advances from application of doubly labeled water to measurement of human energy expenditure. The Journal of nutrition. 1999;129(10):1765-8.
222. Glass S, Dwyer GB, Medicine ACoS. ACSM's metabolic calculations handbook: Lippincott Williams & Wilkins; 2007.
223. Weir JdV. New methods for calculating metabolic rate with special reference to protein metabolism. The Journal of physiology. 1949;109(1-2):1-9.
224. Brown SP, Miller WC, Eason JM. Exercise physiology: basis of human movement in health and disease: Lippincott Williams & Wilkins; 2006.
225. Sesso HD, Paffenbarger RS, Lee I-M. Physical activity and coronary heart disease in men the Harvard Alumni Health Study. Circulation. 2000;102(9):975-80.
226. Waxman M, Stunkard AJ. Caloric intake and expenditure of obese boys. The Journal of pediatrics. 1980;96(2):187-93.
227. Bowman SA, Gortmaker SL, Ebbeling CB, Pereira MA, Ludwig DS. Effects of fast-food consumption on energy intake and diet quality among children in a national household survey. Pediatrics. 2004;113(1):112-8.
228. Gropper S, Smith J. Advanced nutrition and human metabolism: Cengage Learning; 2012.
229. Jabde PV. Text Book Of General Physiology: Discovery Publishing House; 2005.
230. Ruggiero C, Ferrucci L. The endeavor of high maintenance homeostasis: resting metabolic rate and the legacy of longevity. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences. 2006;61(5):466-73.
231. Johnstone AM, Murison SD, Duncan JS, Rance KA, Speakman JR. Factors influencing variation in basal metabolic rate include fat-free mass, fat mass, age, and circulating thyroxine but not sex, circulating leptin, or triiodothyronine. The American journal of clinical nutrition. 2005;82(5):941-8.
232. Puyau MR, Adolph AL, Vohra FA, Zakeri I, Butte NF. Prediction of activity energy expenditure using accelerometers in children. Medicine and Science in Sports and Exercise. 2004;36(9):1625-31.
233. Su S, Wang L, Celler B, Ambikairajah E, Savkin A. Estimation of walking energy expenditure by using support vector regression. Conf Proc IEEE Eng Med Biol Soc. 2005;4:3526-9.

234. Hunter GR, Weinsier RL, Gower BA, Wetzstein C. Age-related decrease in resting energy expenditure in sedentary white women: effects of regional differences in lean and fat mass. The American journal of clinical nutrition. 2001;73(2):333-7.
235. Donahoo WT, Levine JA, Melanson EL. Variability in energy expenditure and its components. Current Opinion in Clinical Nutrition & Metabolic Care. 2004;7(6):599-605.
236. Vaughan L, Zurlo F, Ravussin E. Aging and energy expenditure. The American journal of clinical nutrition. 1991;53(4):821-5.
237. Weyer C, Snitker S, Rising R, Bogardus C, Ravussin E. Determinants of energy expenditure and fuel utilization in man: effects of body composition, age, sex, ethnicity and glucose tolerance in 916 subjects. International Journal of Obesity. 1999 Jul;23(7):715-22.
238. Brazeau A-S, Suppere C, Strychar I, Belisle V, Demers S-P, Rabasa-Lhoret R. Accuracy of Energy Expenditure Estimation by Activity Monitors Differs with Ethnicity. Int J Sports Med. 2014(EFirst).
239. Rising R, Harper IT, Fontvielle AM, Ferraro RT, Spraul M, Ravussin E. Determinants of total daily energy expenditure: variability in physical activity. The American journal of clinical nutrition. 1994;59(4):800-4.
240. Waters RL, Mulroy S. The energy expenditure of normal and pathologic gait. Gait & Posture. 1999;9(3):207-31.
241. Preece SJ, Kenney LPJ, Goulermas JY, Howard D. Does Accelerometer Placement Affect Metabolic Energy Expenditure Estimation in Normal-Weight and Obese Subjects? International Conference on Ambulatory Monitoring of Physical Activity and Movement 2008. 2008;Rotterdam:193.
242. Bassett DR, Jr., Ainsworth BE, Swartz AM, Strath SJ, O'Brien WL, King GA. Validity of four motion sensors in measuring moderate intensity physical activity. Med Sci Sports Exerc. 2000 Sep;32(9 Suppl):S471-80.
243. Crouter SE, Clowers KG, Bassett DR, Jr. A novel method for using accelerometer data to predict energy expenditure. J Appl Physiol. 2006 Apr;100(4):1324-31.
244. Freedson PS, Lyden K, Kozey-Keadle S, Staudenmayer J. Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: validation on an independent sample. Journal of Applied Physiology. 2011;111(6):1804-12.
245. Brandes M, Van Hees VT, Hannöver V, Brage S. Estimating energy expenditure from raw accelerometry in three types of locomotion. Medicine and Science in Sports and Exercise. 2012;44(11):2235-42.
246. Brage S, Brage N, Franks PW, Ekelund U, Wong M-Y, Andersen LB, et al. Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. Journal of Applied Physiology. 2004;96(1):343-51.
247. Strath SJ, Bassett Jr DR, Swartz A, Thompson DL. Simultaneous heart rate-motion sensor technique to estimate energy expenditure. Medicine and Science in Sports and Exercise. 2001;33(12):2118-23.
248. Strath SJ, Brage S, Ekelund U. Integration of physiological and accelerometer data to improve physical activity assessment. Medicine and Science in Sports and Exercise. 2005;37(11 Suppl):S563-71.
249. Preece SJ, Goulermas JY, Kenney LP, Howard D. A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. IEEE Trans Biomed Eng. 2009 Mar;56(3):871-9.
250. PAL Technologies Ltd. Palstickies (tm). [cited 2014 22 November]; Available from: http://www.paltech.plus.com/stickies%20sa.htm.
251. Nyquist H. Certain topics in telegraph transmission theory. American Institute of Electrical Engineers, Transactions of the. 1928;47(2):617-44.

252. Whiting HTA. Human motor actions: Bernstein reassessed: Elsevier; 1983.
253. Choudhury T, Consolvo S, Harrison B, Hightower J, LaMarca A, LeGrand L, et al. The mobile sensing platform: An embedded activity recognition system. Pervasive Computing, IEEE. 2008;7(2):32-41.
254. Consolvo S, McDonald DW, Toscos T, Chen MY, Froehlich J, Harrison B, et al., editors. Activity sensing in the wild: a field trial of ubifit garden. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2008: ACM.
255. Khan AM, Kalkbrenner G, Lawo M. Recognizing Physical Training Exercises Using the Axivity Device. ICT meets Medicine and Health, Bremen, Germany. 2013.
256. Wu JK, Dong L, Xiao W, editors. Real-time physical activity classification and tracking using wearble sensors. Information, Communications & Signal Processing, 2007 6th International Conference on; 2007: IEEE.
257. Nyan M, Tay F, Seah K, Sitoh Y. Classification of gait patterns in the time–frequency domain. Journal of Biomechanics. 2006;39(14):2647-56.
258. Mantyjarvi J, Himberg J, Seppanen T, editors. Recognizing human motion with multiple acceleration sensors. IEEE International Conference on Systems, Man, and Cybernetics (Tucson); 2001.
259. Veltink PH, Bussmann HB, de Vries W, Martens WL, Van Lummel RC. Detection of static and dynamic activities using uniaxial accelerometers. IEEE Trans Rehabil Eng. 1996 Dec;4(4):375-85.
260. van Laerhoven K, Cakmakci O, editors. What shall we teach our pants? Proc 4th Int Symposium on Wearable Computers; 2000.
261. Najafi B, Aminian K, Paraschiv-Ionescu A, Loew F, Bula CJ, Robert P. Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly. IEEE Trans Biomed Eng. 2003 Jun;50(6):711-23.
262. Song K-T, Wang Y-Q, editors. Remote activity monitoring of the elderly using a two-axis accelerometer. Proceedings of the CACS Automatic Control Conference; 2005.
263. Preece SJ, Goulermas JY, Kenney LPJ, Howard D. A Comparison of Feature Extraction Methods for the Classification of Dynamic Activities From Accelerometer Data. Ieee Transactions on Biomedical Engineering. 2009 Mar;56(3):871-9.
264. Aminian K, Robert P, Jequier E, Schutz Y. Incline, speed, and distance assessment during unconstrained walking. Med Sci Sports Exerc. 1995 Feb;27(2):226-34.
265. van Hees VT, Fang Z, Langford J, Assah F, Mohammad A, da Silva IC, et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. Journal of Applied Physiology. 2014;117(7):738-44.
266. Bao L. Physical activity recognition from acceleration data under semi-naturalistic conditions: MASSACHUSETTS INSTITUTE OF TECHNOLOGY; 2003.
267. Rose J, Gamble JG, Adams JM. Human walking: Lippincott Williams & Wilkins Philadelphia; 2006.
268. He Y, Li Y. Physical Activity Recognition Utilizing the Built-In Kinematic Sensors of a Smartphone. International Journal of Distributed Sensor Networks. 2013;2013.
269. Palmerini L, Rocchi L, Mellone S, Valzania F, Chiari L. Feature selection for accelerometer-based posture analysis in Parkinson's disease. Information Technology in Biomedicine, IEEE Transactions on. 2011;15(3):481-90.
270. United States. Department of Health. Physical activity and health: a report of the Surgeon General: DIANE Publishing; 1996.
271. Browning RC, Baker EA, Herron JA, Kram R. Effects of obesity and sex on the energetic cost and preferred speed of walking. Journal of Applied Physiology. 2006 Feb;100(2):390-8.

272. Abdel-Malek K, Arora J. Human Motion Simulation: Predictive Dynamics: Academic Press; 2013.
273. Herren R, Sparti A, Aminian K, Schutz Y. The prediction of speed and incline in outdoor running in humans using accelerometry. Med Sci Sports Exerc. 1999 Jul;31(7):1053-9.
274. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. Canadian medical association journal. 2006;174(6):801-9.
275. Wang GC, Jain CL. Regression analysis: modeling & forecasting: Institute of Business Forec; 2003.
276. MATHWORKS. MATLAB Documentation: stepwisefit. 2014 [cited 2014 22 November]; Available from: http://uk.mathworks.com/help/stats/stepwisefit.html.
277. Kunz I, Schorr U, Klaus S, Sharma AM. Resting metabolic rate and substrate use in obesity hypertension. Hypertension. 2000;36(1):26-32.
278. Brock DW, Tompkins CL, Fisher G, Hunter GR. Influence of resting energy expenditure on blood pressure is independent of body mass and a marker of sympathetic tone. Metabolism. 2012;61(2):237-41.
279. Snodgrass JJ, Leonard WR, Sorensen MV, Tarskaia LA, Mosher M. The influence of basal metabolic rate on blood pressure among indigenous Siberians. American journal of physical anthropology. 2008;137(2):145-55.
280. Harris JA, Benedict FG. A biometric study of human basal metabolism. Proceedings of the National Academy of Sciences of the United States of America. 1918;4(12):370.
281. Horner F, Bilzon JL, Rayson M, Blacker S, Richmond V, Carter J, et al. Development of an accelerometer-based multivariate model to predict free-living energy expenditure in a large military cohort. J Sports Sci. 2013;31(4):354-60.
282. Van der Ploeg G, Withers R. Predicting the resting metabolic rate of 30-60-year-old Australian males. European journal of clinical nutrition. 2002;56(8):701-8.
283. Zarrugh M, Todd F, Ralston H. Optimization of energy expenditure during level walking. European journal of applied physiology and occupational physiology. 1974;33(4):293-306.
284. Jackson AS, Pollock ML. Generalized equations for predicting body density of men. British journal of nutrition. 1978;40(03):497-504.
285. Capodaglio P, Faintuch J, Liuzzi A. Disabling Obesity: From Determinants to Health Care Models: Springer; 2013.
286. Creutzberg E, Schols A, Bothmer-Quaedvlieg F, Wouters E. Prevalence of an elevated resting energy expenditure in patients with chronic obstructive pulmonary disease in relation to body composition and lung function. European journal of clinical nutrition. 1998;52(6):396-401.
287. Rennie KL, McCarthy N, Yazdgerdi S, Marmot M, Brunner E. Association of the metabolic syndrome with both vigorous and moderate physical activity. International journal of epidemiology. 2003;32(4):600-6.
288. Van Pelt RE, Jones PP, Davy KP, DeSouza CA, Tanaka H, Davy BM, et al. Regular Exercise and the Age-Related Decline in Resting Metabolic Rate in Women 1. The Journal of Clinical Endocrinology & Metabolism. 1997;82(10):3208-12.
289. Barasi M. Human nutrition: a health perspective: CRC Press; 2003.
290. Shimokata H, Kuzuya F. [Aging, basal metabolic rate, and nutrition]. Nihon Ronen Igakkai zasshi Japanese journal of geriatrics. 1993;30(7):572-6.
291. Ravussin E, Bogardus C. RELATIONSHIP OF GENETICS, AGE, AND PHYSICAL-FITNESS TO DAILY ENERGY-EXPENDITURE AND FUEL UTILIZATION. American Journal of Clinical Nutrition. [Article]. 1989 May;49(5):968-75.

292. Arizona State University. Compendium of Physical Activities: Unit Conversions. 2011 [cited 2014 21 Nov 2014]; Available from: https://sites.google.com/site/compendiumofphysicalactivities/help/unit-conversions.
293. Howe CA, Staudenmayer JW, Freedson PS. Accelerometer prediction of energy expenditure: vector magnitude versus vertical axis. Med Sci Sports Exerc. 2009 Dec;41(12):2199-206.
294. McKinsey Global Institute. Overcoming obesity:
An initial economic analysis. 2014 [cited 2014 26 November]; Available from: http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights/Economic%20Studies/How%20the%20world%20could%20better%20fight%20obesity/MGI%20Obesity_Full%20report_November%202014.ashx.