University of Bath

UNIVERSITY OF BATH

PHD

Use of accelerometry to predict energy expenditure in military tasks

Horner, Fleur

Award date:
2012

Awarding institution:
University of Bath

Link to publication

UNIVERSITY OF BATH

DEPARTMENT FOR HEALTH

USE OF ACCELEROMETRY TO PREDICT ENERGY EXPENDITURE IN MILITARY TASKS

BY

FLEUR ELIZABETH HORNER

THESIS FOR THE DOCTOR OF PHILOSOPHY

This thesis has been completed as a requirement for the degree of
Doctor of Philosophy, University of Bath

November 2012

ABSTRACT
UNIVERSITY OF BATH - DEPARTMENT FOR HEALTH
Thesis for the <u>Doctor of Philosophy</u>
USE OF ACCELEROMETRY TO PREDICT ENERGY EXPENDITURE IN MILITARY TASKS
Fleur Elizabeth Horner

The overarching aim of this thesis was to enhance the prediction of physical activity energy expenditure (PAEE) in military personnel; specifically, improving accuracy and minimising obtrusiveness.

The first experimental chapter provided a thorough assessment of the reliability and validity of the 3DNX accelerometer. Within unit reliability (CV_{intra}) physical activity counts (PAC) was 0.0-8.9% in all axes in a mechanical setting. Between unit reliability (CV_{inter}) did not exceed 4.5%. The relationship between PAC and acceleration was $r^2 = 0.99$ and standard error of the estimate (SEE) of 6 counts$\cdot$5s^{-1}. During treadmill exercise, the relationship between $\dot{V}O_2$ and PAC was linear (walking, $r^2 = 0.65$, SEE = 1.42 ml$\cdot$kg$^{-1}\cdot$min^{-1}; running, $r^2 = 0.62$, SEE = 3.63 ml$\cdot$kg$^{-1}\cdot$min^{-1}). 3DNX PAC output was valid and reliable when subjected to a physiologically relevant range of mechanically generated accelerations and yielded a linear relationship with $\dot{V}O_2$ during treadmill walking and running.

Chapter 7 investigated the effect of anatomical placement on PAC in order to find the most suitable wear location. Hip and back placements returned similar reliability (CV_{intra} = 3.0% and 2.8% respectively). Hip PAC were higher ($p < 0.01$) for walking with no differences observed for running. Indices of adiposity were related to hip PAC. Regression analysis revealed hip and back PAC as significant predictors of $\dot{V}O_2$. Back PAC was the least variable placement. Supraspinale skinfold thickness explained 15% additional variance in $\dot{V}O_2$ to PAC and reduced SEE.

In Chapter 8, three available devices were compared to doubly labelled water (DLW) for the prediction of free living PAEE using a user-oriented approach. All devices underestimated PAEE. Actiheart-derived PAEE was not different from DLW. However, the wide absolute limits of agreement (LoA) indicated large individual error which was attributed to the use of group rather than individual calibration. 3DNX and GT3X PAEE predictions were different from DLW however LoA were narrower indicating the possibility of applying a correction factor in future.

Chapter 9 was an amalgamation of ten independent cohorts in an attempt to produce a military-specific multivariate model for the prediction of energy expenditure (EE). Stringent data reduction techniques were applied to a highly compliant dataset. Allometric models showed PAC, height and body mass were related to total energy expenditure (TEE) ($p < 0.01$). For models predicting TEE, PAC explained 4 % of the variance. For models predicting PAEE, PAC accounted for 6 % of the variance. The small amount of variance explained by PAC was likely due to the inability of accelerometers to detect EE as a result of day-to-day military activities such as load carriage. Such small portions of explained variance indicate that traditional accelerometry techniques are inadequate for use in military populations.

In Chapter 10, an alternative approach to characterising military-specific activities was explored due to the minor contribution of PAC to PAEE prediction in Chapter 9. Accelerometer raw signal (100 Hz) was used to develop a classification model which aimed to discriminate load carriage (LC) from unloaded ambulation during an occupationally relevant protocol (2-hours, 6.4km$\cdot$hr^{-1}, 25kg load). Fast Fourier transformation showed differences in the frequency distribution of the signal between conditions; caused by differences in gait parameters. Load carriage was detected in 97.2% of 1-minute samples with reduced classification accuracy during the last 30 minutes. Fatigue was suggested as a cause of misclassification; indicated by an upwards drift in $\dot{V}O_2$ and RPE across time.

In conclusion, accelerometer PAC is a weak contributor to the prediction of energy expenditure in military populations. Accuracy could be improved by detection of load bearing activities which is feasible given the advancement in technology and analysis techniques. New technologies such as optical interferometrics could be integrated into existing military equipment to detect heartbeat and respiration; providing data regarding the physiological strain of training and operations.

ACKNOWLEDGEMENTS

I would firstly like to thank my other half Dave for his ongoing support, I know the intricacies of accelerometry are not always thrilling but he has a good poker face! I would also like to thank my family and friends for their never ending enthusiasm.

I could not have completed this project without the help of my colleagues at Optimal Performance Ltd. Their guidance coupled with their assistance during long periods of data collection; particularly for Chapter 9, was invaluable. Thank you in particular to Mark and Gill for their financial support which allowed me to balance my studies with OPL work.

I am especially grateful to three undergraduate students for their assistance with data collection at various points during this project. Jo Slade for her prowess with the skinfold callipers in Chapter 7; Owen Tomlinson and Anthony Gomm for enduring some rather early mornings and temperamental equipment in Chapter 10. I would also like to thank Bec Toone for her constant efforts to maintain the laboratory equipment and Tom Nightingale for supervising data collection when I had an unexpected trip abroad.

Thank you to Alan Nevill for his statistical advice in Chapter 9 and to Anthony Wright and the Elsie Widdowson Laboratory for conducting the doubly labelled water analysis throughout this project.

There are too many participants who were involved in this project to thank by name but I am grateful to all who volunteered. A special mention is warranted for those who completed the extended load carry in their spare time without even the hint of a moan!

Finally, and most importantly I would like to thank my supervisor James Bilzon. It cannot have been an easy task guiding me through this process, yet he has helped me successfully balance my studies and employment. Among other things, his upbeat nature was infectious and on leaving our meetings I would always feel a renewed enthusiasm towards whatever challenge I was struggling with at the time.

LIST OF PUBLICATIONS

The following abstracts were presented at the American College of Sports Medicine conference:

> Horner, F.E,, Bilzon, J.L.J., & Rayson MP, (2010).The mechanical reliability and validity of the 3DNX^{TM} accelerometer. *Med Sci Sports Exerc*, *42(5)*, 476.

> Horner, F.E., Rayson, M.P., Richmond, V.L., & Carter, J.M. (2009). Effect of monitor placement on 3DNX^{TM} accelerometer output during treadmill exercise. *Med Sci Sports Exerc*, *41(5)*, 157-158.

The following abstract was presented at the Second International Congress on Soldiers' Physical Performance:

> Horner F.E., Bilzon, J.L.J., Blacker, S., Rayson, M.P., Carter, J., Richmond V.L., Wilkinson. D., Wright, A., & Nevill, A. (2011, May). *Validity of the 3DNX accelerometer to predict free-living human energy expenditure in military populations: A meta-analysis*. Poster presented at the Second International Congress on Soldiers' Physical Performance, Jyvaskyla, Finland.

The following manuscripts were published or pending publication in peer-reviewed journals:

> Horner, F.E., Rayson, M.P., & Bilzon, J.L.J. (2011). Reliability and validity of the 3DNX accelerometer during mechanical and human treadmill exercise testing International. *International Journal of Obesity*, *35*, S88-S97

> Horner, F.E., Slade, J., & Bilzon, J.L.J. (2012). The Effect of Anatomical Placement and Trunk Adiposity on the Reliability and Validity of Triaxial Accelerometer Output during Treadmill Exercise. *Journal of Physical Activity and Health*; *In press.*

> Horner. F.E., Bilzon, J.L.J., Rayson, M.P., Blacker, S.D., Richmond, V.L., Wilkinson, D.M., Carter, J., Wright, A., & Nevill, A.M. (2012). Development of an Accelerometer-Based Multivariate Model to Predict Free-Living Energy Expenditure in a Large Military Cohort. *Journal of Sports Sciences*; *In press.*

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

EE	Energy expenditure
PAEE	Physical activity energy expenditure
PAC	Physical activity counts
TEE	Total energy expenditure
BMR	Basal metabolic rate
DIT	Dietary induced thermogenesis
RMR	Resting metabolic rate
$\dot{V}O_2$	Rate of oxygen uptake
$\dot{V}CO_2$	Rate of carbon dioxide production
HR	Heart rate
SEE	Standard error of the estimate
MET	Metabolic equivalent
FFT	Fast fourier transformation
LoA	Limits of agreement
TBW	Total body water
OLP	Ordinary least product
IAA	Integral of the modulus
RMS	Root mean squared
SVM	Signal vector magnitude
MAST	Multi axis shaker table
CV	Coeffient of variation
ICC	Intraclass correlation coeffient
Skf	Skinfold
ISAK	International Society for the advancement of kinanthropometry
BMI	Body mass index
MSFT	Multi stage fitness test
LC	Load carriage
L	Loaded
U	Unloaded

1   INTRODUCTION

In an age where diet, health and fitness articles are regularly featured in the media there cannot be many members of the public who have not thought about how to effectively manage their weight at some point during their lifetime. Although for the general population this often means striving to maintain a healthy weight; for populations such as serving military personnel, satisfying the energy balance equation is a constant challenge due to the rigors of military life. A sustained negative imbalance between energy intake and energy expenditure (EE) could lead to impaired occupational functionality for soldiers on operations.

The desire to accurately assess physical activity energy expenditure (PAEE) is an important research goal regardless of whether the focus is health or performance-related. The association between physical inactivity and obesity-related ill-health is now well established (Pate *et al.*, 1995) and has given rise to a sustained academic focus on physical activity measurements, promotion and prescription. Indeed, the majority of energy balance literature has been generated by health researchers attempting to understand the complex physical, environmental, behavioural and psychological mechanisms behind why so many are unable to maintain a healthy weight or indeed, why a given dietary or physical activity intervention succeeds or fails.

The increased operational activity of the British military has meant that there is a renewed interest in quantifying the activity profile and energy demands of operations in different theatres, so that physical training and military field exercises can best match the physical demands experienced by soldiers in operational theatres. Beyond informing military training programmes, the successful quantification of energy expended during operations would better inform military feeding strategies. Energy deficits have been associated with reduced physical performance parameters. Guezennec et al. (1994) observed a 15% reduction in time to exhaustion during a cycling in a low-energy-intake group following five days of military exercise. No performance decrements were seen in the high-energy-intake group. More chronically, decrements in maximal dynamic lift strength have been observed during UK military exercise following a period of eight weeks (Fortes *et al.*, 2011). In a similar 8-week study, carried out at the same training establishment as Fortes et al. (2011), Richmond et al. (2010) discovered a mismatch between energy intake and EE

amounting to a deficit of approximately 644 kcal·day^{-1} across the 8-week observation period. Further, Nindl et al. (2007) found a 20 % reduction in maximal lift strength following a prolonged (8-week) caloric deficit of approximately 1000 kcal·day^{-1} during US army ranger training. Although the link between prolonged (8-week) energy deficit and reductions in physical performance may be implied, there are no conclusive data showing that performance is maintained in those who remain in energy balance over the same period. If decrements in role-related performance tests following training are related prolonged energy deficits then this may translate to similarly impaired performance on operations if feeding strategies in this environment are insufficient also. At present there are little empirical data available documenting the energy demands of operations upon which to form evidence-based feeding strategies.

While the deployment of a research team, to directly assess the energy demands of such operations is possible, on a large scale this is impractical, potentially dangerous and intrusive for the soldiers concerned. Movement sensing devices may offer a practical compromise whereby EE can be predicted unobtrusively without the need for deployment of a large civilian research team. Currently however, movement sensing technologies such as accelerometers have shortfalls in their application to military research as they were designed to characterise the physical activity of the general population rather than the unique activities associated with day-to-day soldiering (e.g. load-carriage).

The overall aim of this thesis is to enhance the prediction of PAEE in military personnel; specifically improving accuracy and minimising obtrusiveness.

The experimental chapters in this thesis have investigated the use of accelerometry in military populations with particular reference to advancing the prediction of PAEE. Initial pilot chapters (Studies 1 and 2) were conducted to produce a conversion factor to enable the inclusion of data collected using an earlier version of the 3DNX in chapter 9, and to evaluate the different internal processing techniques used to calculate an accelerometer count (Physical Activity Counts, PAC). The reliability and validity of the 3DNX accelerometer was then scrutinised using a controlled mechanical device, followed by human participants conducting treadmill exercise in the laboratory (Chapter 6). The most suitable anatomical attachment site for the accelerometer was then chosen by establishing the least variable accelerometer output from an incremental treadmill protocol (Chapter 7). Chapter 8 was a small-scale comparison of three commercially available devices in a free-

living environment to evaluate their prediction accuracy in relation to PAEE. A large-scale multi-cohort study was then conducted within a range of military populations using accelerometry and the doubly labelled water technique. Data were combined with similar historical data to create a multivariate model for the prediction of PAEE from PAC (Chapter 9). Due to the weak relationships revealed between PAC and PAEE in Chapter 9, a novel signal processing technique was employed to identify the unique signal features associated with load carriage; a common military activity for which PAEE is frequently underestimated by conventional accelerometers (Chapter 10). In Chapter 11 the results from the various experimental chapters are summarised, limitations highlighted, potential applications discussed and future research directions presented.

2 REVIEW OF THE LITERATURE: ESTIMATING ENERGY EXPENDITURE IN MILITARY POPULATIONS

The measurement of human EE is not a new concept. In the 1800s Lavoisier and Seguin discovered that larger people consumed more oxygen than their smaller counterparts and that standing up and moving around required more oxygen than sitting down. It was during this series of experiments that the method of indirect calorimetry was established; a method still used as a benchmark for quantifying human EE to this day (Ainslie *et al.*, 2003). The technology associated with indirect calorimetry has become markedly more sophisticated since the 18th century work of Lavoisier and Seguin however the requirement of the participant to be either confined in a chamber or even linked to a portable gas analysis system makes it near impossible to capture the complexity of activities of daily living (Speakman, 1998). Multiple attempts have been made to find an unobtrusive method of quantifying the energy demands of free-living activity. In this review of the literature, the components of EE will be explained followed by an outline of the nature of military training in order to provide context for the remainder of the review. A number of criterion measures of EE will then be introduced and evaluated, followed by a critical appraisal of contemporary methodologies employed to predict physical activity EE. Finally, recent advancements in the field of accelerometry and future directions will be discussed.

2.1 COMPONENTS OF HUMAN ENERGY EXPENDITURE

For the purpose of this review it is necessary to first distinguish between physical activity and EE. Total energy expenditure (TEE) is made up of three components; basal metabolic rate (BMR), dietary induced thermogenesis (DIT) and physical activity energy expenditure (PAEE). Basal metabolic rate, commonly recognised as the minimum amount of energy required to sustain life, is usually the largest component of TEE and is influenced mainly by body size and more specifically fat-free mass (Bluck, 2008; Kleiber, 1947; Schofield, 1985). Basal metabolic rate is the component that explains the largest proportion of TEE, except in those who are regularly participate in high intensity and long duration activities (De Lorenzo *et al.*, 1999). Often it is not possible to measure BMR due to equipment limitations as an accurate estimate is only obtained when a participant is able to sleep in a specially designed respiration chamber (i.e. a calorimeter), where oxygen (O_2) uptake and carbon dioxide (CO_2) production can be accurately measured. The literature tends to use

BMR and resting metabolic rate (RMR) terminology interchangeably however in reality, as most studies have not measured BMR overnight in a respiration chamber, the resting component of TEE will be referred to as RMR from here forward.

The energetic cost of utilisation of dietary fat, carbohydrate and protein is commonly termed dietary induced thermogenesis (DIT). Dietary induced thermogenesis is normally calculated by dividing the increase in EE above basal fasting level by the energy content of the food ingested (Westerterp *et al.*, 2008). Although a relatively small portion of TEE, DIT can vary depending upon the amount of dietary protein consumed habitually by the study population (Carlson, 1997). It is widely accepted that DIT accounts for approximately 10 % of TEE based on a standard western diet (Westerterp, 1999).

Physical activity energy expenditure is the increase in EE above resting levels that is due to body movement produced by skeletal muscle contraction (Westerterp, 2009). It is usually expressed independent of inter-individual differences in resting metabolic rate; the lowest level of energy expended at rest. At present it is difficult to accurately translate human movement into units of PAEE due to its varied and complex nature. This task is particularly challenging in free-living populations, often due to the lack of an appropriate criterion measure (Westerterp, 2009).

Following an introduction to common criterion measures, this review will evaluate the various validated methods for the prediction of TEE and PAEE, particularly in relation to potential military applications.

2.2   THE PHYSICAL DEMANDS OF MILITARY OCCUPATIONS

A career in the military is typically regarded as a physically demanding occupation. The unique military-specific activities that are completed habitually by recruits and trained soldiers necessitate that a practical and pragmatic approach be taken during this research project. The military aspect of this project will inform the study design of the experimental chapters and the feasibility of the experimental measures will be an ongoing consideration.

In the United Kingdom male and female recruits can begin a career in the military at the age of 16 as a regular soldier and begin officer training at 17 yrs 9 months. All regular soldiers (excluding infantry) complete the common military syllabus for recruits (CMS(R)) prior to their phase 2 training which includes training specific to their chosen trade i.e. engineer, nurse, chef.

The 14-week CMS(R) course comprises activities such as drill, physical training (PT), military exercise and adventurous training. Richmond et al. (2012) conducted a physical demands analysis of the course and reported mean daily physical activity levels (TEE/BMR) as 2.2 ± 0.2 and 2.3 ± 0.2 for males and female recruits respectively. These values are regarded as "high" (> 1.85) according to the classification scale proposed by Bouten et al (1996) for use in the general population.

Infantry soldiers complete the entirety of their training at the Infantry Training Centre (ITC) in Catterick. A physical demands analysis of the 24-week Combined Infantry Course (CIC) was conducted by Wilkinson et al (2008b). During this course, which is necessarily arduous, recruits completed a syllabus comprising basic infantry skills such as weapons training with the addition of individual and team events, prolonged periods of load carriage, a 10-mile speed march and a log race.   Mean daily TEE values were reported as approximately 4,700 kcal. Tharion et al. (2005) reported a similar mean value for energy expended by military personnel in the United States (US) (4,600 kcal); including data from military support staff based mainly in the garrison. However, the range of values was between 3109 and 7131 kcal per day. At the higher end of the spectrum, Marines completing the US Marine Corps Infantry Officer Course expended 5377 kcal per day and although lower than other values cited, recruits completing the US Army Ranger course sustained energy expenditures of ~4000 kcal per day for 65 days of continuous training.

Officer training in the UK consists of 44 weeks of training split into three terms of 14 weeks and two weeks of adventurous training (Harwood *et al.*, 1999), unfortunately there are no published data available relating to typical amounts of energy expended during this commissioning course.

Once military personnel have completed their respective training courses they enter a phase of pre deployment where their company is posted to a military base. Day to day activities include PT, personal administration, weapons training, sporting events and short duration military exercises in preparation for deployment on operations. When mean military EE values (~4600 kcal) are compared to those of civilians similar in age (Black, 1996), civilians appear to expend 38% less energy on a daily basis. It is clear that during periods of military exercise (where the majority of data are available), daily EE is likely to be much greater than for most civilian jobs. However, it is difficult to compare the demands of day to day military life as there are few data available for periods of time spent

in camp or on garrison when working days are shorter in duration and physical activity is less frequent and intense. Furthermore, a company comprises of a combination of trades and support staff, not solely infantry soldiers or marines which results in highly variable habitual activity. Demanding civilian occupations such as firefighting can also illicit EEs of ~4000 kcal per day during discreet arduous tasks such as wild fire suppression (Ruby *et al.*, 2002). Therefore, when age and job type are considered, TEE for military and civilian personnel can be comparable (Tharion*, et al.*, 2005).

What distinguishes military life from other occupations are the physical activities that are encountered on a daily basis. Activities cited in the literature as the most frequently occurring are walking, marching with a backpack, lifting and lowering loads, lifting and carrying loads, digging and running (Rayson *et al.*, 2000; Wyss *et al.*, 2010). This unique combination of habitual activities is not encountered by the general population on a day to day basis. Therefore, the quantification of EE requires careful consideration in terms of the experimental procedures and the instruments used.

2.3   CRITERION MEASURES OF ENERGY EXPENDITURE

2.3.1   Direct Calorimetry

Directly measuring EE involves quantifying heat production and heat loss associated with the combustion of fuel in the form or carbohydrate, protein or fat (Westerterp*, et al.*, 2008). Although accurate, direct calorimetry is uncommon in the exercise physiology literature as it requires a participant to remain unaccompanied in a room-sized calorimeter throughout the duration of the observation period. Therefore, as direct calorimetry is not an appropriate criterion measure for research undertaken in this thesis, this review will focus on more applicable measures; all of which are indirect.

2.3.2   Indirect Calorimetry

Indirect calorimetry measures respiratory gas exchange which is then used to calculate EE (Seale *et al.*, 1990). Although an indirect measure, Seale et al (1990) reported that direct and indirect calorimetry were equivalent when using a whole room calorimeter. Indirect calorimetry techniques are more adaptable than direct calorimetry. One of the most common indirect calorimetry methods is the Douglas bag technique which was first cited in 1911 (Douglas, 1911) and is described in detail by Hopker et al. (2012). Despite being

over 100 years old, this technique is still commonly referred to as the 'gold standard' for assessing oxygen uptake (metabolic rate) and EE in the laboratory (Carter *et al.*, 2002; Gladden *et al.*, 2012). Recently, Hopker et al. (2012) examined the reliability of repeated gas sampling and reported a coefficient of variation of 0.5% for both CO_2 and O_2 when best practice is adhered to.

During the Douglas bag sampling procedure the participant wears a mouthpiece and nose clip while a one-way valve ensures that only expired air is collected into a Douglas bag. The volume of expired air is measured and a sample obtained to measure O_2 and CO_2 concentrations. Volume of inspired air and subsequent oxygen uptake ($\dot{V}O_2$) is then calculated using the Haldane transformation (Haldane, 1912) and the Fick equation (Fick, 1870). Providing that steady-state values for $\dot{V}O_2$ and CO_2 production ($\dot{V}CO_2$) are known, the quantities of fat and carbohydrate oxidised per minute can be calculated and EE estimated (Jeukendrup *et al.*, 2005).

Although the Douglas bag technique is mainly confined to laboratory studies, online portable systems offer breath by breath pulmonary gas exchange measurements using equipment that comprises a face mask and a backpack-style harness weighing approximately 600g (Ainslie*, et al.*, 2003). Validation studies have reported significant differences between $\dot{V}O_2$ and $\dot{V}CO_2$ values measured by portable systems and Douglas bags during running (Duffield *et al.*, 2004) and cycling (McLaughlin *et al.*, 2001). The overestimation of the portable system reported by Duffield et al. (2004) appears to be consistent (indicated by intraclass correlation coefficients of between 0.7 and 0.9). Although the Douglas bag method is preferable in a laboratory setting, portable systems provide an appropriate measurement of respiratory gas exchange (Meyer *et al.*, 2001). Although such systems are being continuously updated, indirect calorimetry is not suitable for assessing EE in true free-living populations due to its acute interference with everyday activities and its limited deployment time.

2.3.3 Doubly Labelled Water (DLW)

The Doubly labelled water (DLW) method is a stable isotope-based technique that estimates whole-body CO_2 production and hence EE (Speakman, 1997). First reported for use in humans by Schoeller and van Santen (1982), the DLW technique is a method of assessing TEE in a free-living populations. Following a known oral dose of water enriched

with stable isotopes of hydrogen (^{2}H) and oxygen (^{18}O), the isotopes mix with the more abundant forms of the hydrogen (^{1}H) and oxygen (^{16}O) found in the endogenous body water pool. As energy is expended in the body, CO_2 and water are produced. Hydrogen is eliminated from the body via urination, sweating and other insensible losses. ^{18}O is also lost from the body via these routes with the addition of loss as CO_2 during exhalation. Consequently, as both isotopes are ingested simultaneously, their rate of elimination will differ, with the ^{18}O being eliminated at a faster rate than the ^{2}H (Speakman, 2005). Urine samples collected over a series of days (typically 7-14) throughout the period of observation therefore provide snapshots of the degree of 'label' remaining in the total body water pool (Bluck, 2008). The difference between the elimination of ^{18}O and ^{2}H reflects the rate at which CO_2 is produced. Coupled with an assumed mean respiratory quotient value of 0.85 based on a standard western diet, EE can be estimated (Westerterp, 1999).

The DLW technique is considered the 'gold standard' for measuring TEE, but it is not without its limitations. In field studies where the respiratory quotient is assumed rather than calculated from the food quotients of recorded dietary intake then predicted values are likely to be within approximately ±2 % of the true group TEE value (Black *et al.*, 1986). This error is increased if a researcher reports PAEE when RMR has been predicted rather than measured. If RMR is not measured then a standard error (Hopkins, 2000) of approximately 8% of PAEE is expected when using the Schofield RMR prediction equation in a population of young males (van der Ploeg *et al.*, 2001). The DLW technique is expensive to implement due to the high cost of the stable isotopes and associated analyses, particularly given the specialist equipment and expertise needed for the assessment of isotope ratios in specific body fluid pools via isotope ratio mass spectroscopy (Ainslie, *et al.*, 2003). Although the technique provides an accurate estimate of mean EE over an extended period of time (7-14 days), very little detail concerning frequency, duration or intensity of activity can be acquired (Plasqui *et al.*, 2007). Despite these limitations the technique still provides an appropriate criterion measure for the prediction of mean free-living EE in humans (Bouten, *et al.*, 1996).

2.4   PREDICTION OF ENERGY EXPENDITURE

Where it is not feasible to use either the DLW or Douglas bag technique, researchers have strived to develop user-friendly, objective methods of predicting EE. The development of prediction tools normally involves a validation study using one of the criterion measures

above. This section will provide a systematic review of the available prediction tools followed by an introduction to the more recent advances in technologies.

2.4.1   Questionnaires

Cheap and easy to administer, self-report activity questionnaires and diaries are the most common tool for assessing physical activity (Westerterp, 2009). Despite their continued wide spread use in large population studies, the reliability and validity of questionnaires to measure physical activity is limited (Shephard, 2003). This is demonstrated by Sarkin et al. (2000) who reported that the proportion of the United States population meeting fitness guidelines in the year 2000 ranged from 4% to 70% depending on the self-report method used.   Compared to DLW derived estimations, questionnaires generally show low correlations systematically underestimating in some cases where actual physical activity levels are high (Maddison *et al.*, 2007) and overestimating (Koebnick *et al.*, 2005) EE in others where actual physical activity levels are low. Questionnaires may be useful to characterise PAEE at the group level and rank groups in order of activity level but it is generally accepted that there is considerable error at the individual level (Bonnefoy *et al.*, 2001). There are no studies where self-reported physical activity has been validated in a military population but the indication from the existing literature is that questionnaires or activity diaries would not adequately characterise the energy demands of military training or operations.

2.4.2   Heart Rate

Heart rate (HR) is closely related to EE during exercise within a range of approximately 90-150 b·min^{-1} (Rennie *et al.*, 2001). It is a relatively easy method to administer and is used widely as a measure of physiological strain (Wilkinson*, et al.*, 2008b). At low levels of physical activity there is a weak relationship between HR and EE where increases in HR are not necessarily accompanied by similar increases in EE (Luke *et al.*, 1997). This is probably due to changes in stroke volume associated with change in posture; however, particularly at low levels of physical activity, HR is also affected by external factors such as ambient temperature, clothing, high humidity, altitude, dehydration and illness (Achten *et al.*, 2003). Such external factors can lead to changes in HR without associated changes in $\dot{V}O_2$ causing a dissociation in the nature of relationship between HR and EE (Rennie*, et al.*, 2001).

It is also difficult to estimate the EE associated with activity that is intermittent in nature using HR monitoring. Heart rate tends to lag behind changes in work rate meaning that an increase or decrease in work rate will not immediately result in a concurrent increase in HR to the level that would have been reached following steady state exercise at that same intensity (Achten*, et al.*, 2003).

When using HR to predict EE there are a number of ways to interpret the data. The most promising method is known as the flex HR method (Ainslie*, et al.*, 2003). First reported by Spurr et al. (1988) this method requires each individual to record both resting HR values and perform an incremental exercise test. This allows the development of an individual HR-$\dot{V}O_2$ calibration curve. The flex HR is found by averaging the highest HR observed at rest and the lowest HR observed during light activity (Livingstone, 1997). During the monitoring period, any HR values recorded below the flex HR assume that the participant is at rest and resting metabolic rate is used to determine EE for that time period. When any values above flex HR are recorded then the individual HR-$\dot{V}O_2$ curve is used to predict expended energy.

When comparing measured 24-hour EE and values predicted using flex HR there were no significant differences between the mean values (Spurr*, et al.*, 1988). Similarly when the flex HR method was compared to DLW measurements of EE the group mean estimate from flex HR was within 10% of measured values (Livingstone, 1997). However, when examined on an individual basis the results of both studies showed that the maximum deviations were +20 % to -15% of the 24-hour calorimetry values and +19 % to -17 % of the DLW determined values.

In summary, HR response is well correlated with physical activity EE but it can be affected by factors other than movement and requires individual calibration. The variation in HR values caused by altitude, ambient temperature and hydration coupled with the need for individual calibration means the HR flex method may not be appropriate for use in military populations.

2.4.3   Pedometry

Pedometers are worn at the hip and reliably count the number of steps accumulated during the day (Schneider *et al.*, 2004). Several pedometer models are commercially available (Schneider*, et al.*, 2004). If stride length is known then distance travelled via ambulation

can also be inferred (Plasqui, *et al.*, 2007). Pedometers have been shown, in the short term initially, to act as valuable exercise behaviour change tools (Tudor-Locke *et al.*, 2009), however they appear to show limited promise when used as a tool to predict EE. During low-moderate intensity scripted exercise, Thompson et al. (2006) found that step counts measured by pedometry did not provide a satisfactory estimate of criterion EE. Activities of daily living involve more complex movement patterns than both ambulation and scripted exercise; as such the applicability of pedometers to assess free-living activity is limited. Gardner and Poehlman (1998) assessed the validity of a hip-mounted pedometer against DLW in older adults and found a weak relationship between steps·day^{-1} and free-living PAEE ($r^2 = 0.38$ SEE = 124 kcal.day^{-1}) compared to the relationship of $r^2 = 0.70$ found between accelerometer counts and PAEE in the same population. Overall it appears that although pedometers can accurately measure ambulation and may be a valuable tool for motivating individuals to initially engage in exercise behaviour, step counts cannot provide detail regarding the frequency, duration and intensity of exercise and cannot be used to accurately predict EE.

2.5   ACCELEROMETRY

Movement sensors and more specifically body-borne accelerometer-based devices began to feature in physical activity research about thirty years ago (Montoye *et al.*, 1983; Wong *et al.*, 1981). Accelerometers can provide temporal information about the total amount, frequency, intensity and duration of physical activity (Westerterp, 2009). Two main types of accelerometer are in widespread use in physical activity research: uniaxial and increasingly, triaxial. Uniaxial accelerometers register movement in the vertical axis only and triaxial accelerometers register movement in the vertical, horizontal and mediolateral axes. A triaxial accelerometer 'count' is commonly derived in three ways. First, summing the integral of the modulus of acceleration from each axis (Bouten *et al.*, 1997), second, summing the integral of the route mean squared acceleration from each axis (Cook *et al.*, 2011; Schutz *et al.*, 2002) and third, summing the vector magnitude of acceleration from each axis (Esliger *et al.*, 2010; Sasaki *et al.*, 2011). The first approach is that used by the 3DNX accelerometer which is the unit used in the first three chapters of this research programme, but comparative analyses suggest there is little difference in the total counts derived (see Chapter 5).

Widespread use of accelerometry in population-based studies eliminates the extensive staffing required for observation methods (McKenzie, 2010) and avoids recall problems, and removes subjectivity and bias associated with self-report survey methods. However, the quality of information received from accelerometers is only as good as the reliability and validity of the devices themselves (Esliger *et al.*, 2006).

The reliability of commercially available accelerometers is mixed. Some units such as the widely-used Actigraph and Actical return coefficients of variation of below 10% intra-unit (Esliger, *et al.*, 2006). Whereas others such as the triaxial RT3 can vary up the 100% from day to day when subjected to the same testing (Esliger, *et al.*, 2006). The reliability of the 3DNX accelerometer, the device available for this project, will be thoroughly examined to form an important first stage of this research programme.

2.5.1   Using Accelerometry to Predict Energy Expenditure

Laboratory studies have shown close agreement between activity counts and EE for ambulation and other treadmill-based activities (Welk *et al.*, 2000). Despite this close agreement for both types of accelerometer in laboratory studies, it is still difficult to accurately translate accelerometry data into units of expended energy (Chu *et al.*, 2007). During free-living activities it appears that the greater sensitivity of the triaxial accelerometer to movement in different planes leads to better prediction of EE than uniaxial accelerometers in a free-living environment (Campbell *et al.*, 2002; Plasqui *et al.*, 2005; Welk, *et al.*, 2000). It is widely accepted that waist-mounted accelerometers tend to underestimate EE during free-living conditions, mainly due to the technology's limited ability to detect increased energy cost from load carriage, upper body movement and changes in surface or terrain (Hendelman *et al.*, 2000; Trost *et al.*, 2005). Nonetheless, it remains that triaxial accelerometry is a promising prospect in terms of objectively quantifying TEE and PAEE.

It is difficult to compare the individual contribution of accelerometer output to the estimation of EE between studies as few report PAC data alone. Most data are reported in combination with anothropometric variables. A summary of DLW validated studies are shown in Table 2.1.

Device	Study	Dependent Variable	Independent Variable	r^2	$\%r^2$
Actigraph	Masse et al. (2004)	TEE	BM, Counts·day^{-1}	38%	5%
Actigraph	Masse et al. (2004)	TEE	FFM, Counts·day^{-1}	43%	2%
3DNX	Carter et al. (2008)	TEE	FFM, Counts·day^{-1}	78%	35%
Tracmor	Plasqui et al. (2005)	TEE	Age, BM, H, Counts·day^{-1}	83%	19%
TracmorD	Bonomi et al. (2010)	TEE	BM, Counts·day^{-1}	54%	23%
Actigraph	Assah et al. (2009)	PAEE·kg^{-1}	Counts·day^{-1}	14%	14%
Actigraph	Masse et al. (2004)	PAEE	BM, Counts·day^{-1}	9%	5%
Actigraph	Masse et al. (2004)	PAEE	FFM, Counts·day^{-1}	12%	4%
Tritrac R3D	Jacobi et al.(2007)	PAEE	Counts·day^{-1}	N.S	N.S
Tracmor	Plasqui et al. (2005)	PAEE	Age, BM, H, Counts·day^{-1}	81%	33%
TracmorD	Bonomi et al. (2010)	PAEE	BM, Counts·day^{-1}	46%	16%

Table 2.1 A comparison of accelerometer validation studies using the DLW technique. $\%r^2$ refers to the variance in the dependent variable explained by Counts·day^{-1} alone.

Relatively few triaxial accelerometers have been validated against DLW. Those that have, often show large prediction error with low EEs commonly being over-predicted and high EEs under-predicted at the level of the individual. The early work of Bouten et al. (1996) reported that the output from the triaxial Tracmor explained 6 % of the variance in TEE measured by DLW with a standard error of the estimate of 18 %. Maddison et al. (2009) report an additional 4 % explained variance in TEE measured by DLW when RT3 PAC were included in a model containing gender, body fat (kg) and RMR. Bonomi et al. (2010) report the most favourable results from the triaxial Tracmor (Phillips New Wellness Solutions; www.directlife.phillips.com). Validated against measured TEE, the Tracmor returned an explained variance of 76% and standard error of the estimate of 0.9 MJ.day^{-1} (7.4 %) when combined with measured RMR. Carter et al. (2008) also reported similar figures with the triaxial 3DNX^{TM} (Biotel, Bristol, UK) explaining variance of up to 35 % using accelerometer counts alone and up to 78 % when accelerometer counts were combined with body composition variables. There is difficulty comparing studies using different devices due to the non-equivalence of PAC values (Welk *et al.*, 2012), and the inconsistency in reporting prediction error. To the author's knowledge, none of these

accelerometer-derived prediction equations have been validated against DLW in a truly independent cohort.

2.6 ADVANCES IN ENERGY EXPENDITURE PREDICTION USING ACCELEROMETER-BASED DEVICES

2.6.1 Combined sensors

In an attempt to address the shortcomings of accelerometry, researchers identified the potential for combining multiple sensor outputs and integrating accelerometry with physiological data such as HR in order to capture the physiological strain associated with behaviours that are undetectable by accelerometry alone (Lamonte *et al.*, 2001).

While other studies have investigated the integration of HR and movement sensing (Rennie*, et al.*, 2001; Su *et al.*, 2009); the Actiheart (CamnTech, Cambridge, UK) was the first device capable of integrating HR and accelerometry in a single unit (Brage *et al.*, 2005).The technical specifications of the Actiheart are described in chapter 3 and elsewhere (Brage*, et al.*, 2005). Depending on the intensity of the activity being performed, HR and accelerometer counts contribute in varying proportions, along with gender and HR above sleep, to the minute-by-minute calculation of expended energy via a branched equation. The development of the branched equation model used for the prediction of EE is outlined by Brage et al.(2004).

The Actiheart has shown encouraging results when used to predict free-living PAEE. Crouter et al (2008) aimed to extend the laboratory-based work of Thompson et al. (2006) by comparing PAEE values generated by the Actiheart versus those generated from a portable indirect calorimeter. Following a series of 18 scripted "free-living" activities, the authors reported that when using a group calibration method, PAEE predicted using the Actiheart was only significantly different to PAEE generated from the Cosmed K4b^2 in three activities. Since the study by Crouter et al. (2008) there have been few validation studies conducted in free-living populations (Spierer *et al.*, 2011), and only one discovered using the DLW method as a criterion measure. Assah et al. (2011) evaluated the validity and accuracy of the actiheart in sub-saharan adults. The mean bias between measured and predicted PAEE was not significant, however the 95% limits of agreement were large, particularly in rural populations when individual calibration was not performed (-84.1, +52.9 kJ·kg·d^{-1}). The authors attributed the wide limits of agreement to the increased

digging, lifting and load carrying associated with rural activities resulting in heart rate responses different to that of the group calibration equation. However, the premise of a combined HR and accelerometry device is that the HR component can account for such activities. Whether the Actiheart would be suitable for use within military populations remains to be established. However, due to the theoretical potential of integrating HR and accelerometry for characterising activities where increases in EE are not accompanied by comparative increases in movement (e.g. load carrying), the Actiheart will be considered alongside other methodologies for use in this research programme.

Other combined sensors validated in the literature includes the multi-sensor IDEEA (Minisun, CA) which was able to predict EE for steady state activities with 95% accuracy following a strict laboratory protocol (Zhang *et al.*, 2004). The multi-sensor ShakeNet (CSEM, Switzerland) also returned impressive prediction accuracy even within a free-living population. Rumo et al. (2011) compared ShakeNet-derived values of PAEE to those derived using the DLW technique. This system, which combines two triaxial accelerometers (upper arm and upper thigh) with HR, predicted PAEE with a standard error of 9%. The strength of this system was its ability to recognise seven common activities of daily living via unique signal features and apply individual regression equations to each epoch when these activities were detected. Sensors such as these are clearly more accurate than traditional waist-mounted accelerometers however have limited ecological validity as they are likely to be too burdensome for use in the field, particularly in the military.

2.6.2   Alternative Accelerometer processing techniques

The most recent accelerometry research has begun to recognise the value of the rich raw signal an accelerometer can capture. The increased memory of newer devices means they are capable of storing the raw, unfiltered acceleration pattern for extended periods (Intille *et al.*, 2012). A number of different processing techniques have been proposed in a move away from using the relationship between the arbitrary accelerometer 'count' and expended energy. Reporting accelerometer output in SI units (g or m·s^2) has been actively encouraged by world leaders in the accelerometry field (Freedson *et al.*, 2012). Reporting accelerometry-derived data in such a way improves accuracy, utility and allows comparison between devices more readily (Heil *et al.*, 2012; Intille*, et al.*, 2012).

The ability to capture raw accelerometer data has led some authors to embark on more sophisticated methods of predicting PAEE and even detecting the types of activity a person is performing (Bonomi *et al.*, 2009a). The principle being that the raw accelerometer signal contains useful information beyond the mean (or total) acceleration (Staudenmayer *et al.*, 2012). No universal standard exists for transforming raw activity monitor data into units of EE (Heil*, et al.*, 2012). Historically the majority of studies have used linear models to predict TEE and PAEE (Lyden *et al.*, 2010). This appears an over-simplistic approach as the relationship between activity counts and AEE is known to vary according to the type of activity (Midorikawa *et al.*, 2007). Thus, prediction models that account for the type of activity performed could result in more accurate estimates of EE (Bonomi*, et al.*, 2009a). Crouter et al. (2006) recognised the shortcomings of linear-style prediction models and proposed a two-step procedure. These authors observed that locomotion returned consecutive 1-second count values with a relatively low coefficient of variation as opposed to less rhythmic activities. This distinction allowed the authors to first classify each minute of counts as locomotion or not and then apply specifically calibrated equations to estimate metabolic equivalent (MET) values per minute (Staudenmayer*, et al.*, 2012). This approach reduced the root mean squared error associated with predicted MET values by 50% compared to simple linear regression methods employed most famously by Freedson et al. (1998). Rothney et al. (2010) validated the two-regression model in an independent free-living sample using the DLW technique as a criterion measure. At a population level the prediction of TEE was not significantly different from DLW-derived values with a mean difference ($\pm$SD) of 5.67 $\pm$ 16%. The range was -20.0 to 32.3% indicating large prediction error at the individual level. As the detailed raw accelerometry signal was not available to Crouter et al. (2006), the range of activities that could be identified was limited and may cause the large prediction error at the individual level. Regardless, this early methodology has been reported as an intuitive first step in advancing the analysis of accelerometer data (Staudenmayer*, et al.*, 2012).

2.6.3   Activity classification from accelerometer data

Exploiting the accelerometry signal beyond the mean (or total) acceleration per minute or even per second is a promising route for improving the prediction of PAEE in military populations. The activity profile of military personnel varies markedly from that of the general population however, methodologies such as artificial neural networks (Rothney *et*

al., 2007; Staudenmayer *et al.*, 2009; Trost *et al.*, 2012), decision trees (Karantonis *et al.*, 2006; Schutz*, et al.*, 2002) and bayes classifiers (Long *et al.*, 2009) may be adapted for use in this unique population.

To date, using accelerometry to predict PAEE within military populations has been challenging due to the amounts of load carriage and resistance exercise performed. For example, soldiers are expected to perform frequent assessments involving load carriage during training (Rayson*, et al.*, 2000), and repeated bouts of load carriage during operations (McCaig *et al.*, 1986); activities with associated energy demands that are widely recognised as problematic for accelerometers to fully capture (Hendelman*, et al.*, 2000).

Recently, Wyss and Mader (2010) proposed a method of identifying common military tasks through a combination of body-fixed sensors, such as the GT1M accelerometer (Actigraph, Fort Walton, FL), and HR in a similar approach to that of Rumo et al. (2011) introduced in section 2.6.1. Their algorithm aimed to identify some of the most frequently performed military tasks including walking, marching with a backpack, running, weapon handling and lifting and lowering loads. Although the classification algorithm assigned a fixed kilocalorie per minute value for each activity and did not allow for performing each task at variable intensities, predicted free-living PAEE was not significantly different to measured PAEE during a randomly chosen 90-minute snapshot of a daily military routine (Wyss *et al.*, 2011). Of note was the inability of the classification algorithm to detect marching with a backpack without the output from an additional sensor located on the backpack itself.

Although these two military-specific studies, took important steps forward in the characterisation of military routines, the use of multiple sensors increases user burden and may place restrictions on activities performed habitually (Long*, et al.*, 2009). Long et al. (2009) used signal features from a single, waist-mounted accelerometer-device to identify walking, running, driving, cycling and 'sports'. The authors trained a Bayesian classifier and cross-validated the model using a leave-one-out method. The classifier was reported as 79.5% ± 11.6% accurate following cross-validation. Although not military-specific, this study highlights the potential of using a single waist-mounted accelerometer to identify different activities. Transferring a similar methodology to military-specific activities could provide a solution to the burden of wearing multiple sensors. In a similar study, Wang et al (2009) also used a classification-style model, based on 19 salient features of a 50Hz

triaxial accelerometer signal, to determine four different modes of ambulation. In this study, an overall accuracy of 90.9% was achieved when a Gaussian Mixture Model classifier was used to determine inclined walking of 4.8 and 17.3% from declined walking at the same gradients. Most recently, Zhang et al. (2012) used a an accelerometer capturing raw signal (80 Hz) attached at the waist and the wrist to classify sedentary, household, walking and running activities in a laboratory setting. The fast fourier transform (FFT) and wavelet decomposition analysis were capable of discriminating the activities with up to 99% accuracy.

Loaded marching has been associated with a distinctive forward trunk lean (Majumdar *et al.*, 2010) and an increase in stance phase of stride, double support time and pelvic rotation (Birrell *et al.*, 2009). These kinematic changes that occur when an individual carries a military load could be reliably and significantly different enough from normal gait to confer with consistently different accelerometer signals when compared to normal ambulation.

Comparing the technical merits of different classifier methodologies is outside the scope of this review; however, Zhang et al (2012) compared fives types of classifier in their recent publication and found their classification accuracies of physical activities to be similar. What is apparent is that salient features of a raw signal generated by a single, waist-mounted accelerometer device can be used to identify a range of activities. It is feasible that a classification-style model could be trained to identify military-specific activities such as load carriage, given the distinctive kinematics associated with this activity. Accurate identification of load carriage in the controlled environment of a laboratory would be an important first step towards to a more accurate estimation of the frequency and duration of performing such activities in the field and ultimately, improved PAEE estimation during training and on operations.

2.7   SUMMARY OF THE LITERATURE

The increased operational activity of the British Army has led to a renewed interest in quantifying the physiological demands of operations (Section 2.1). A suitable tool for estimating EE would help inform feeding strategies as sustained energy imbalances and performance decrements have been reported. There is no shortage of studies aiming to improve the prediction accuracy of tools for estimating EE. Questionnaires, HR, pedometry and accelerometry have all been used in health research to quantify TEE and

PAEE with varying degrees of success although few of such tools have been validated in a military population (Section 2.5).

Accelerometry is widely used to quantify TEE and PAEE in health research however at best, PAC alone can only account for 35% of the variance in TEE and 33% of the variance in PAEE (Table 2.1). This is surprising given that theoretically, accelerometers measure the body accelerations associated with the active component of daily EE. There are little data available regarding the prediction of free living PAEE specific to military populations; however, the failure of Carter et al (2008) to find a significant relationship between PAC and PAEE suggests that the participants were performing activities that the devices were unable to fully detect.

Combined HR and accelerometry sensors were developed on the basis that the HR component of such devices can detect the strain associated with activities an accelerometer cannot register (Section 2.6.1). Mean data are often not significantly different to measured PAEE however 95% LOA can be wide, particularly if individual calibration has not been performed. Further investigation and comparison of the ability of stand-alone accelerometers versus combined accelerometry and HR devices to predict PAEE in military populations has not yet been conducted.

Recently, there has been a shift in the direction of accelerometry-based research which may address some of the difficulties associated with using traditional accelerometer methodologies in military populations. Alternative processing techniques have been discussed in section 2.6.2 however there are little data available that are military specific. Wyss and Mader (2010) developed a novel activity recognition technique using military participants. However, the multi-sensor device combined the output from HR and two further accelerometers. This limits the broader use of such a device due to its user burden.

The potential of the raw accelerometer signal from a single device (pre processing) has been recognised and the research focus advocated by subject matter experts is centred on using unique signal features to identify activities of daily living. Although in its infancy, successful recognition of activities of daily living using accelerometer signal features could be applied to military specific activities such as load carriage that traditional accelerometer methodologies cannot detect (Section 2.6.3).

3   GENERAL METHODS

This chapter introduces methods common to multiple studies and the procedures followed so that repetition in subsequent chapters is avoided.

3.1   DOUBLY LABELLED WATER (DLW)

The Doubly labelled water (DLW) method is a stable isotope-based technique that estimates whole-body CO_2 production and hence EE (Speakman, 1997). First reported by Schoeller and van Santen (1982), the DLW technique is a method of assessing TEE in a free-living individual. Following a known oral dose of water enriched with stable isotopes of hydrogen (^2H) and oxygen (^{18}O), the isotopes mix with the more abundant forms of the hydrogen (^1H) and oxygen (^{16}O) found in the endogenous body water pool. As energy is expended in the body, CO_2 and water are produced. Hydrogen is eliminated from the body via urination, sweating and other insensible losses. ^{18}O is also lost from the body via these routes with the addition of loss as CO_2 during exhalation. Consequently, if both isotopes are ingested simultaneously, their rate of elimination will differ with the ^{18}O being eliminated at a faster rate than the ^2H (Speakman, 2005). Urine samples collected over a series of days throughout the period of observation provide snapshots of the degree of 'label' remaining in the total body water (TBW) pool (Bluck, 2008). The difference between the elimination of ^{18}O and ^2H reflects the rate at which CO_2 is produced. Coupled with an assumed mean respiratory quotient value of 0.85 based on a standard western diet, EE can be estimated (Westerterp, 1999).

3.1.1   Kinetics of Elimination

A description of elimination process is given below. The information is adapted from the paper by Bluck (2008) as it describes the methodology employed by the laboratory that conducted the DLW analysis throughout this project.

In order to calculate TEE the rate of oxygen turnover in the TBW pool must first be calculated. The rate of loss of water (rH_2O) plus the rate of loss of carbon dioxide (rCO_2) is equal to the turnover rate of oxygen in the TBW pool ($k_O.N$) and is given in equation 3.1.

$$rH_2O + 2\ rCO_2 = k_o{\cdot}N \qquad \text{(Equation 3.1)}$$

The factor two is required since each mole of CO_2 contains equal amounts of oxygen to two moles of water. As the equation above describes a reaction rate, it follows that for every molecule of CO_2 expired, an individual will lose twice the amount of O_2 per unit of time than each water molecule. The difference between the two products (divided by 2) is thus equal to rCO_2 and is expressed in Equation 3.2

$$R_{CO_2} = \frac{N_O k_O - N_H k_H}{2} \qquad \text{(Equation 3.2)}$$

In theory both ingested ^2H and oxygen ^{18}O would equilibrate fully with body water. This is not the case due to the free exchange with other compounds in the body and a correction must be applied to the equations to represent non-ideal tracer behaviour (Equation 3.3).

$$R_{CO_2} = \frac{N_O k_O - N_H k_H - F_S(f_2 - f_1)}{2 f_3 + q(f_2 - f_1)} \qquad \text{(Equation 3.3)}$$

Where R_{CO_2} = CO_2 production rate

$k_H\ k_O$ = fractional constants of elimination for hydrogen and oxygen

N_H and N_O the associated pool sizes.

F_S = represents the skin losses that are fractionated (27.3 moles.day^{-1} for adults in a temperate climate)

q = fractionated respiratory losses as a fraction of CO_2 production (1.1)

$f_1 = 0.941$ fractionation factor for ^2H leaving the body as water vapour

$f_2 = 0.991$ fractionation factor for ^{18}O leaving the body as water vapour

$f_3 = 1.037$ fractionation factor for the exchange of ^{18}O between CO_2 and water

Rate of CO_2 production rate must finally be converted to EE (Equation 3.4). Some assumptions must be made in terms of the energy equivalent of CO_2 (substrate oxidation). This method assumes 12% of total energy intake comes from protein (Elia *et al.*, 1988).

$$TEE = \left(\frac{15.48}{RQ} + 5.550\right) R_{CO_2} \qquad \text{(Equation 3.4)}$$

Where RQ is taken as 0.85

Where 127.5kcal is used per mol CO_2 produced.

3.1.2   Isotope dosing and sampling procedure

Each participant provided baseline urine samples before ingesting a weighed oral dose of $^2H_2^{18}O$ (day 0). The doses were 80 mg·kg^{-1} deuterium oxide and 145 mg.kg^{-1} $H_2^{18}O$. Post-dose urine samples were collected daily for 7-10 days depending on the study and the time of day noted. Urine samples were subsequently frozen at -20 °C and later analysed in duplicate using isotope-ratio mass spectrometry (MRC, Human Nutrition Research, Elsie Widdowson Laboratory, Cambridge, UK). Using this dosing regimen, TEE can be measured with a coefficient of variation lower than 5 % (Bluck, 2008). A sample of drinking water was also collected to correct for the natural abundance of $^2H_2^{18}O$ at each study location.

3.1.3   Energy expenditure calculations

Energy expenditure was calculated, as previously described by Schoeller et al. (1982), from the slopes and intercepts of isotope disappearance curves, based on samples collected at the first two, middle and last two days of the study. For all subjects the respiratory quotient was assumed to have a daily average value of 0.85 and resting metabolic rate (RMR) was estimated using the Schofield equations (1985). Physical activity energy expenditure (PAEE) was calculated as (TEE x 0.9) – RMR, assuming that diet-induced thermogenesis was 10 % of TEE (Plasqui, *et al.*, 2005).

3.1.4   Limitations

Despite the advantages of the DLW technique for estimating free living EE, this criterion method inevitably contains some precision error and this must be appreciated when interpreting the findings of this project.

There are concerns regarding the accuracy of EE prediction at the individual level. It is suggested that imprecision of individual DLW measurements may occur because individuals vary in how closely they approximate the necessary assumptions that are made when performing the calculations outlined in 3.1.1 (Speakman, 2005).

For example, the assumption that fractionated water loss is 20% of total water loss may not be appropriate for the participants in Chapter 8 where ambient temperatures were in excess of 30 °C on a daily basis. Further error may be introduced should the background enrichments of O_2 and CO_2 change during the course of the observation period. Baseline samples of the drinking water consumed by participants were sent to the laboratory

conducting the DLW analysis as standard, however it was assumed that the background enrichment remained constant.

3.2 DOUGLAS BAG METHOD

One of the most common indirect calorimetry methods is the Douglas bag technique which was first cited in 1911 (Douglas, 1911). Despite being over 100 years old, this technique is still commonly referred to as the 'gold standard' for assessing oxygen uptake (metabolic rate) and EE in the laboratory (Carter*, et al.*, 2002; Gladden*, et al.*, 2012). Recently, Hopker et al. (2012) examined the reliability of repeated gas sampling and reported a coefficient of variation of 0.5% for both CO_2 and O_2 when best practice is adhered to.

During the Douglas bag sampling procedure the participant wears a mouthpiece and nose clip while a one-way valve ensures that only expired air is collected into a Douglas bag. The volume of expired air is measured and a sample obtained to measure O_2 and CO_2 concentrations. Volume of inspired air and subsequent $\dot{V}O_2$ is then calculated using the Haldane transformation (Haldane, 1912) and the Fick equation (Fick, 1870). Providing that steady-state values for $\dot{V}O_2$ and CO_2 production are known, the quantities of fat and carbohydrate oxidised per minute can be calculated and EE estimated (Jeukendrup*, et al.*, 2005).

3.2.1 Sampling Procedure

One-minute collections of expired gases were made using Douglas bags (Hans Rudolph, MO, USA). The Douglas bags were flushed with room air and fully evacuated prior to gas collection. Fractions of expired air (O_2 and CO_2) were assessed using paramagnetic and infrared analyzers, respectively (Series 1440 gas analyser, Servomex plc., Crowborough, UK) and volume of expired air determined using a dry gas meter (Harvard Apparatus Ltd., Edenbridge, UK). The temperatures of expired gases were measured with a digital thermometer (model C, Edale Instruments, UK). The gas analyser was calibrated using a two point calibration: O_2 and CO_2 were zeroed using 100 % nitrogen gas; O_2 and CO_2 were spanned to 16.0 and 5.0 % using a known gas mixture (16.0% O_2, 5.0% CO_2, 79.14% N).

The dry gas meter was checked regularly during laboratory testing periods. Room air was pumped through in 35l increments up to 175l using a 7l syringe (Model 4900, Hans Rudolph Inc., Kansas City, USA).

Volume of oxygen uptake was calculated using the Haldane transformation and EE was calculated according to the equations suggested in a review by Jeukendrup and Wallis (2005).

3.3   ACCELEROMETRY

3.3.1   3DNX

The 3DNX model v3 (BioTel Limited, Bristol, UK; www.biotel.co.uk) is sensitive to movements in three planes: X (anteroposterior), Y (mediolateral) and Z (vertical). The unit measures 54 x 54 x 18 mm, and weighs 70 g including a 3.6 v lithium battery (Saft Ltd., UK) with a life of ~21 days collecting data at 5-second epochs. The unit contains two ADXL321 biaxial micro-electro-mechanical (MEMS) sensors (Analog Devices Ltd., Surrey, UK) positioned orthogonally over a third uniaxial sensor to measure acceleration in three movement planes. The raw sample frequency of the 3DNX is 100 Hz with a low pass filter set at 0.2 Hz and a high pass filter set at 10Hz which ensures that most non-human movement, such as vibration, is not registered. The unit has a full dynamic range of ±18 g however the dynamic range is limited to ±5.5 g. A product specification sheet can be found in Appendix A (Chapter 13).

Following data capture, the raw AC signal is first converted to a DC signal, it is then filtered, amplified and the modulus of the absolute acceleration signal (IAA) for each sensing axis is integrated over one-second. One-second values are then summed depending on the epoch chosen by the user (5-60 seconds) resulting in the variables IAA_x, IAA_y and IAA_z. These variables are then summed to produce a single value for the three axes combined (IAA_{TOT}).

$$\mathbf{IAA_{TOT}} = \int_0^t |x(t)|dt + \int_0^t |y(t)|dt + \int_0^t |z(t)|dt \qquad \text{(Equation 3.5)}$$

The integral of the modulus of acceleration is described in further detail by Bouten et al. (1997). This method was compared with ten other processing methods and was found to have the strongest relationship with energy expended during common activities of daily living.

The frequency range of the 3DNX (0.2 – 10 Hz) is compared to other commercially available devices in Table 3.1. Most human movement occurs at frequencies between 0.3 and 3.5 Hz (Sun *et al.*, 1993). Therefore, the remaining portion of the frequency range

between 3.5 and 10 Hz is largely redundant and the unit may register accelerations resulting from sources other than human movement i.e. vibration (Chen *et al.*, 2005). This may impact on the interinstrument reliability of a device. Conversely, a frequency range too narrow and the acceleration resulting from some more vigorous movements such as running may be incorrectly removed (Fudge *et al.*, 2007).

Model	Dimensions: Length x Width x Height (mm)	Weight (g)	Pieziosensor Orientation	Dynamic Range (*g*)	Frequency Range (Hz)
Actigraph GT3X	38 x 37 x 18	27	Triaxial	± 3	0.2 - 2.5
RT3	71 x 56 x 28	65	Triaxial	0.5 - 2	2 - 10
3DNX	54 x 54 x 18	70	Triaxial	± 5.5	0.2 - 10

Table 3.1 Comparison of the technical specifications for three accelerometer devices

3.3.2 Actiheart

Developed in conjunction with Brage et al. (2005) the Actiheart (CamnTech, Cambridge, UK), was the first device to combine HR and acceleration components into a single wearable device. The Actiheart consists of two clips attaching to standard ECG electrodes (Red Dot 2771, 3M, London). The main body of the device contains an omni-directional accelerometer and memory capable of storing 21 days of data recording at 60-second epochs. The second component is clipped to a second adjacent electrode 10cm away. Heart rate (bpm) is generated from an ECG signal. The R wave is detected and the mean of the last 16 R-R intervals is taken as the beats per minute value; ignoring any values outside 25% of the mean. The accelerometer component has a sampling rate of 32Hz and a dynamic range of ±2.5g. In both chapters 8 and 10 the device is worn in the lower position (Figure 3.1.) however in chapter 10 the device is attached to a sports strap (polar).
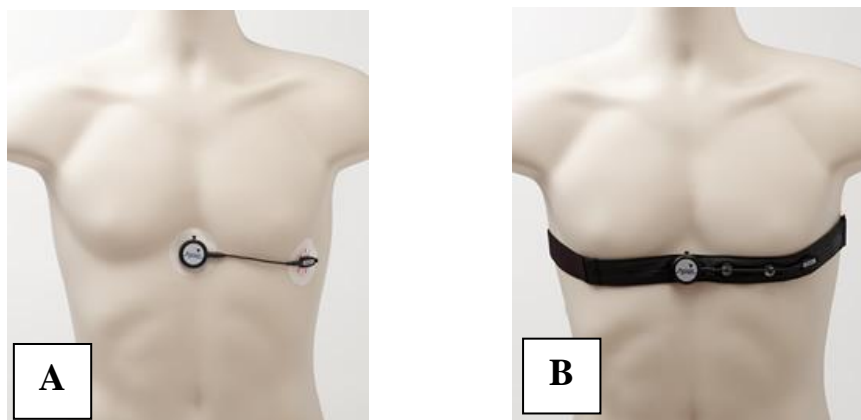
Figure 3.1 **A**: the Actiheart placed in the lower position, **B**: the Actiheart attached to the modified sports strap.

The branched equation model used by the Actiheart software (version 4) to predict EE using the group calibration (Brage *et al.*, 2007) is given in the diagram below (Figure 3.2). The 'activity value' was implemented as the lowest value observed during cycling. This prevents cycling from following the low intensity route in the branched model.

```
                                              > ┌──────────────────────────┐
                             ┌──────────────┐   │ (0.1 x ActEE) + (0.9 x   │
                             │ Transition   │   │         HREE)            │
          > 25 counts·min⁻¹  │ HRas         │   └──────────────────────────┘
    ┌──────────────┐         │[0.54 x SHR   │ < ┌──────────────────────────┐
    │              │─────────│ + 54.3]      │   │ (0.5 x ActEE )+ (0.5 x   │
    │ Activity     │         └──────────────┘   │        HREE)             │
    │ Value        │                            └──────────────────────────┘
    │[25 counts·   │         ┌──────────────┐ > ┌──────────────────────────┐
    │ min⁻¹]       │         │              │   │ (0.5 x ActEE )+ (0.5 x   │
    │              │─────────│ Flex HRas    │   │        HREE)             │
    └──────────────┘         │[0.05 x SHR   │   └──────────────────────────┘
          < 25 counts·min⁻¹  │ + 21.2]      │ < ┌──────────────────────────┐
                             └──────────────┘   │ (0.9 x ActEE) + (0.1 x   │
                                                │        HREE)             │
                                                └──────────────────────────┘
```
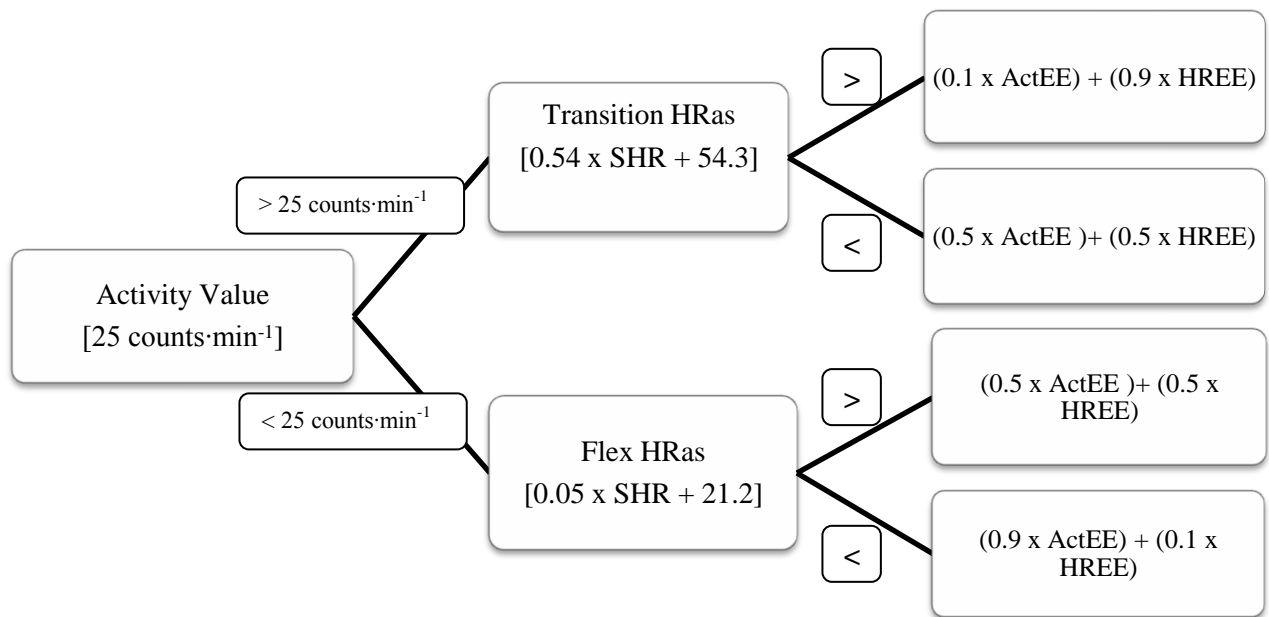
Figure 3.2 Where HRas: Heart rate above sleep: HREE is given by: $HREE = 5.5 * HRas + (1.6 * HRas * gender) - (7.8 * SHR * gender) + (338 * gender) - (4.7 * SHR + 207)$

4   PILOT WORK – COMPARISON OF 3DNX VERSION 2 AND 3DNX VERSION 3 ACCELEROMETER OUTPUT

4.1   INTRODUCTION

The 3DNX version 2 (v2) was last used during the gender streaming in army training OPL contract at ATC Pirbright in 2007 (Richmond, *et al.*, 2012). Since then the smaller version 3 (v3) has been used which has been subjected to both hardware and software upgrades. The 3DNX v2 is described by Carter et al. (2008) and the 3DNX v3 is described in Chapter 3. In order to compare 3DNX data from current work to that conducted before 2007, a conversion factor is needed to ensure all data can be combined. Both v2 and v3 accelerometers are triaxial and the hardware is similar therefore, providing the relationship between accelerometer output, in the form of physical activity counts (PAC), and acceleration is linear for both versions then a simple regression equation can be used to convert the PAC output from version 2 to version 3 or *vice versa*.

4.2   METHODS

4.2.1   Mechanical Testing

Five 3DNX v3 and four 3DNX v2 units were attached to a multi axis shaker table (MAST) using double-sided floor tape. The units were shaken for eight minutes at three accelerations (0.5g, 1g and 1.5g). The units were first positioned so the Z-axis of the accelerometer corresponded to the direction of the horizontal actuator. The units were then rotated 90º so that the X-axis was aligned with the horizontal actuator. The same protocol was then repeated. The units were not tested in the Y-axis as the design of the v2 units meant that they were too thin to attach firmly to the shaker in the upright position.

4.2.2   Human Testing

During the gender streaming for recruits study at ATR Pirbright in 2007 (Richmond, *et al.*, 2012), 12 participants wore both a v2 and v3 unit located at the small of the back in the same pouch. The units were returned following ten days of data collection. Total PAC were then divided by ten days to arrive at an average daily physical activity counts value

for both units. These data provide a second, free-living, comparison of the output from the v2 and v3 units.

All units were set to record at one-minute epochs and downloaded into an Excel spreadsheet. The relationship between v2 and v3 units was examined using ordinary least product (OLP) regression (Ludbrook, 1997) so that the equation generated is the same (once rearranged) regardless of whether converting v2 to v3 counts or *vice versa*. Standard error of the estimate (SEE) was calculated for both relationships.

4.3   RESULTS

4.3.1   Mechanical Testing

The relationship between v2 and v3 counts was linear ($r^2 = 1.00$, n = 206, SEE = 17 counts·min^{-1}). Figure 4.1 shows the relationship between v2 and v3 counts.
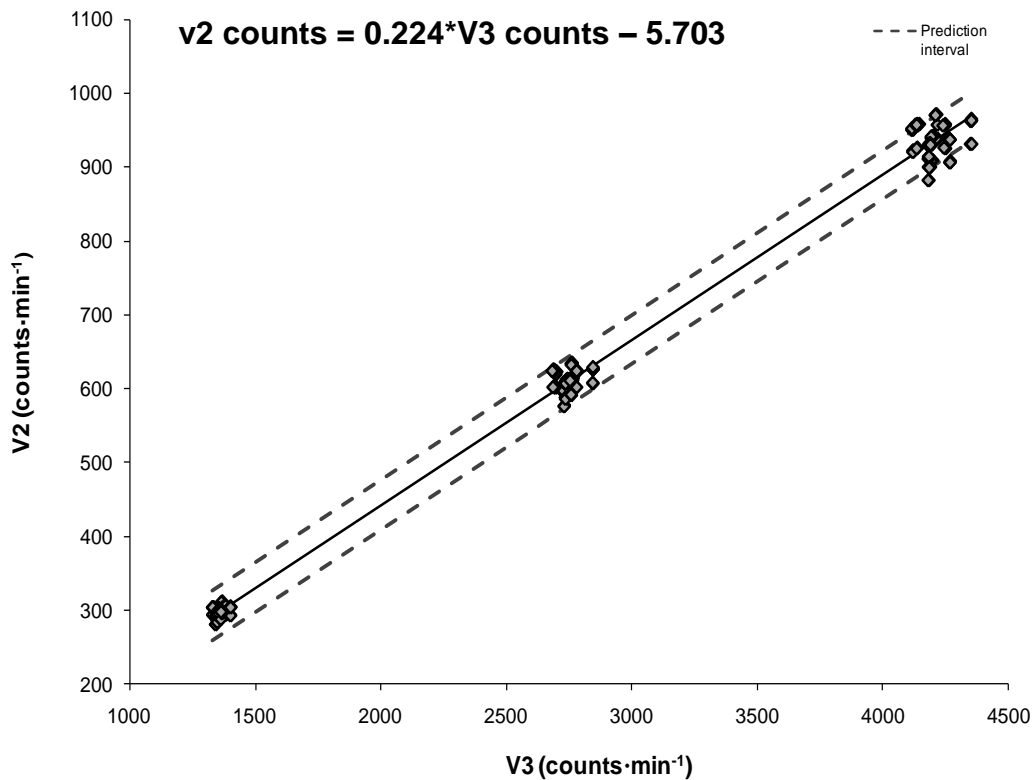
Figure 4.1. The relationship between v3 and v2 physical activity counts (counts·min^{-1}) from the mechanical shaker table at 0.5 *g*, 1 *g* and 1.5 *g*.

The relationship between v2 and v3 counts was linear ($r^2 = 0.99$, SEE = 5347 counts·day^{-1}) (Figure 4.2).
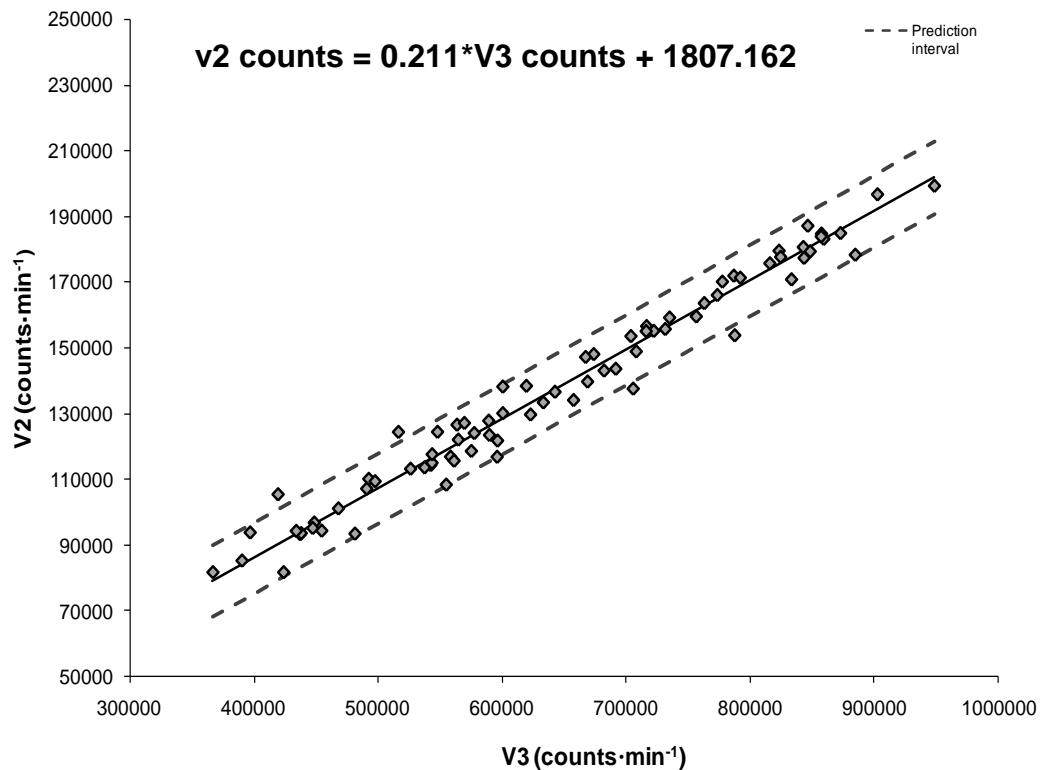
Figure 4.2 The relationship between v2 and v3 counts during free-living activity.

4.4   DISCUSSION

The relationships between v2 and v3 accelerometer counts were linear for both mechanical and free-living data. Version 2 counts can be successfully converted to version 3 counts using a simple OLP regression equation (Equation 4.1) given by:

$$v2 \text{ counts} = 0.224 * v3 \text{ counts} - 5.703 \qquad \text{(Equation 4.1)}$$

This equation allows current accelerometer data and pre 2007 data to be compared or combined following a simple mathematical conversion. We recommend using the conversion equation generated from the mechanical testing due to the tightly controlled nature of the experiment. Although the free living data provide a valuable visual comparison, external factors such as tightness of the belt and body fatness could have impacted on the relationship between v2 and v3 accelerometer outputs.

5   PILOT WORK – COMPARISON OF TWO ACCELEROMETER PROCESSING METHODS FOR THE PREDICTION OF OXYGEN UPTAKE

5.1   INTRODUCTION

Accelerometer devices are commonly used to characterise the relationship between physical activity and EE (Westerterp, 2009). Broadly speaking, such devices measure accelerations resulting from body movement and convert the direct current (DC) signals (following internal processing) into activity "counts". The linear relationship between "counts" and metabolic rate during ambulation is widely used to predict EE (Lyden *et al.*, 2011). Some manufacturers do not reveal the internal processing steps used to generate "counts" (Arvidsson *et al.*, 2007); however, as some manufacturers do publicise their processing steps, it becomes clear that there are a number of methods that can be used to calculate a physical activity "count" (PAC) from a raw acceleration signal (Chen*, et al.*, 2005). Three commonly cited methods for deriving PAC from raw acceleration signals are summing the integral of the modulus of acceleration (Bouten*, et al.*, 1997), the integral of the route mean squared acceleration (Cook*, et al.*, 2011; Schutz*, et al.*, 2002) and the vector magnitude of acceleration (Esliger*, et al.*, 2010; Sasaki*, et al.*, 2011) over a user-defined time period (epoch). At present no consensus has been reached regarding the most valid data processing method for the prediction of metabolic rate and subsequent EE. Furthermore, there is little transparency by manufacturers regarding the step by step method used to generate accelerometer count output (Maddison*, et al.*, 2009). The lack of a standard output hinders the comparison of accelerometer performance and may prevent researchers from choosing the most appropriate device for their specific application.

The integral of the modulus of acceleration is described in detail by Bouten et al. (1997). This method was compared with ten other processing methods and was found to have the strongest relationship with energy expended during common activities of daily living. Specifically, after a raw AC signal has been converted to a DC signal it is filtered, amplified and the modulus of the absolute acceleration signal (IAA) for each sensing axis is integrated over one-second. One-second values are then summed depending on the epoch chosen by the user (5-60 seconds) resulting in the variables IAA_x, IAA_y and IAA_z.

These variables are then summed to produce a single value for the three axes combined (IAA_{TOT}).

$$\mathbf{IAA_{TOT}} = \int_0^t |x(t)| \, dt + \int_0^t |y(t)| \, dt + \int_0^t |z(t)| \, dt \qquad \text{(Equation 5.1)}$$

This method has been utilised and discussed in several previous studies and was deemed a suitable and appropriate method for characterising an acceleration signal (Bouten*, et al.*, 1997; Karantonis*, et al.*, 2006; Mathie *et al.*, 2004; Montoye*, et al.*, 1983).

The second processing method for deriving PAC that is commonly cited in the literature is the root mean squared (RMS) acceleration (Schutz*, et al.*, 2002). The RMS acceleration is often included in more complex activity recognition studies where raw acceleration signals are interrogated for common features and subsequently utilised to produce models which can be trained to detect subsequent activity patterns (Atallah *et al.*, 2010; Voleno *et al.*, 2010). From a theoretical perspective this method is a more intuitive choice as it returns, from a mechanical standpoint, total acceleration over a given time period for all axes (RMS_{TOT}). The RMS processing method is defined as:

$$\mathbf{RMS_{TOT}} = \sqrt{\frac{(\sum_{n=1}^N (x_n - \bar{x})^2) + (\sum_{n=1}^N (y_n - \bar{y})^2) + (\sum_{n=1}^N (z_n - \bar{z})^2)}{N}} \qquad \text{(Equation 5.2)}$$

Where N = the sample rate of a given device

The third processing method that appears regularly in the literature is the signal vector magnitude (SVM) (Esliger*, et al.*, 2010; Karantonis*, et al.*, 2006; Powell *et al.*, 2004), which provides a measure of movement intensity derived from the raw accelerometer signal (Karantonis*, et al.*, 2006). The SVM method is defined by:

$$\mathbf{SVM} = \sqrt{x_i^2 + y_i^2 + z_i^2} \qquad \text{(Equation 5.3)}$$

Where x_i is the ith sample of the x-axis signal (similarly for y_i and z_i)

The processing method which returns the strongest relationship between PAC and rate of oxygen uptake and hence, EE during treadmill exercise has not yet been established. The 3DNX accelerometer (Biotel Ltd., Bristol, UK) derives PAC using the IAA method (3DNX_{IAA}). However, a modified unit version using the RMS method has been developed (3DNX_{RMS}) in order to objectively compare the two processing methods. The hardware specifications of the devices are identical. The only difference being the post-hoc signal

processing method (i.e. IAA vs. RMS methods). Unfortunately, although requested, the manufacturers declined to produce a 3DNX device using the SVM processing method, restricting this pilot study to a two-way comparison. Although the 3DNX samples at 100 Hz, these data are rectified, integrated over one second and gravity is subtracted before these values are stored. If data were stored in their raw format then any method of post processing would be possible, however in its current form the 3DNX does not have the capacity to store raw data.

The aim of this pilot study was to determine which PAC output, $3DNX_{IAA}$ or $3DNX_{RMS}$, produces the least random error in the prediction of rate of oxygen uptake ($\dot{V}O_2$) during treadmill exercise in humans.

5.2   METHODS

5.2.1   Accelerometry

Both the $3DNX_{IAA}$ and the $3DNX_{RMS}$ contain the same hardware; the 3DNX device is described in Chapter 3. The $3DNX_{IAA}$ unit was configured using the IAA_{TOT} approach and the $3DNX_{RMS}$ was configured using the RMS approach (described in Equation 5.1). Both accelerometers recorded PAC at 5-second epochs.

5.2.2   Exercise protocol

Eleven male, recreationally active participants (age 21 $\pm$ 3 yr; height 182 $\pm$ 5.4 cm; mass 78.0 $\pm$ 7.1 kg), completed an incremental treadmill protocol on a pre-programmed, motorized treadmill (Woodway ELG70, Waukesha, US). The protocol consisted of three 5-minute walking stages (4 km·h^{-1}, 5 km·h^{-1} and 6 km·h^{-1}) and three five-minute jogging stages (8 km·h^{-1}, 10 km·h^{-1} and 12 km·h^{-1}). Each participant wore one $3DNX_{IAA}$ unit and one $3DNX_{RMS}$ unit on an elastic belt, positioned side by side in the small of the back. The rate of oxygen uptake was measured during the last minute of all walking and jogging stages using the Douglas Bag technique described in chapter 3.

All accelerometer data were downloaded using dedicated software (Biotel, UK) and imported into a spreadsheet where five-second epoch accelerometer counts (PAC) were averaged across the duration of each stage. Both the first and the last two epochs of the jogging stages were discarded. This ensured that all data used for the subsequent analysis were not affected by a change of stage.

5.2.3   Statistical Analyses

A custom model analysis of covariance (ANCOVA) was performed firstly, to determine the homogeneity of regression slopes for walking and running data and secondly, to establish whether an interaction existed between $3DNX_{IAA}$ and $3DNX_{RMS}$. A two-way (2 x 6) mixed model analysis of variance (ANOVA) was performed to test for differences between unit version (two levels) and PAC for all treadmill speeds (6 levels). Paired sample t-tests were used to locate differences. Simple linear regression was then used to examine the strength of the relationship between physical activity counts and rate of $\dot{V}O_2$ for each unit version. Standard error of the estimate (SEE) was generated for all relationships. All data are presented as mean ± SD. Statistical significance was set a-priori at $p<0.05$.

5.3   RESULTS

Descriptive statistics for all treadmill speeds are shown in Table 5.1. Analysis of covariance revealed that the slopes and intercepts of the relationship between PAC and $\dot{V}O_2$ did not vary ($p=0.442$) according to exercise mode indicating that the assumption of homogeneity of regression slopes was not violated. All walking and running data were therefore combined in subsequent analyses.

Treadmill speed ($km \cdot hr^{-1}$)	Oxygen Uptake ($ml \cdot kg^{-1} \cdot min^{-1}$)	Physical Activity Counts (counts$\cdot 5s^{-1}$)	
		$3DNX_{IAA}$	$3DNX_{RMS}$
4	13.0 ±1.7	120 ±7	154 ±14
5	14.4 ±1.1	164 ±10	205 ±15
6	17.8 ±1.4	221 ±19	270 ±24
8	30.7 ±3.1	507 ±48	585 ±58
10	36.8 ±2.9	615 ±46	710 ±65
12	43.7 ±2.6	680 ±47	787 ±72

Table 5.1 Group values. $3DNX_{IAA}$, physical activity counts generated using equation 5.1; $3DNX_{RMS}$, physical activity counts generated using equation 5.2 (values displayed as mean ±SD).

A two-way mixed model ANOVA revealed a significant interaction between unit version and speed ($p=0.046$). Pairwise comparisons showed the $3DNX_{RMS}$ returning higher PAC than the $3DNX_{IAA}$ unit at all treadmill speeds ($p<0.05$). The interaction between unit

versions is shown in figure 5.1 where the difference between $3DNX_{IAA}$ PAC and $3DNX_{RMS}$ PAC is greater at 8, 10 and 12 km·hr^{-1}.
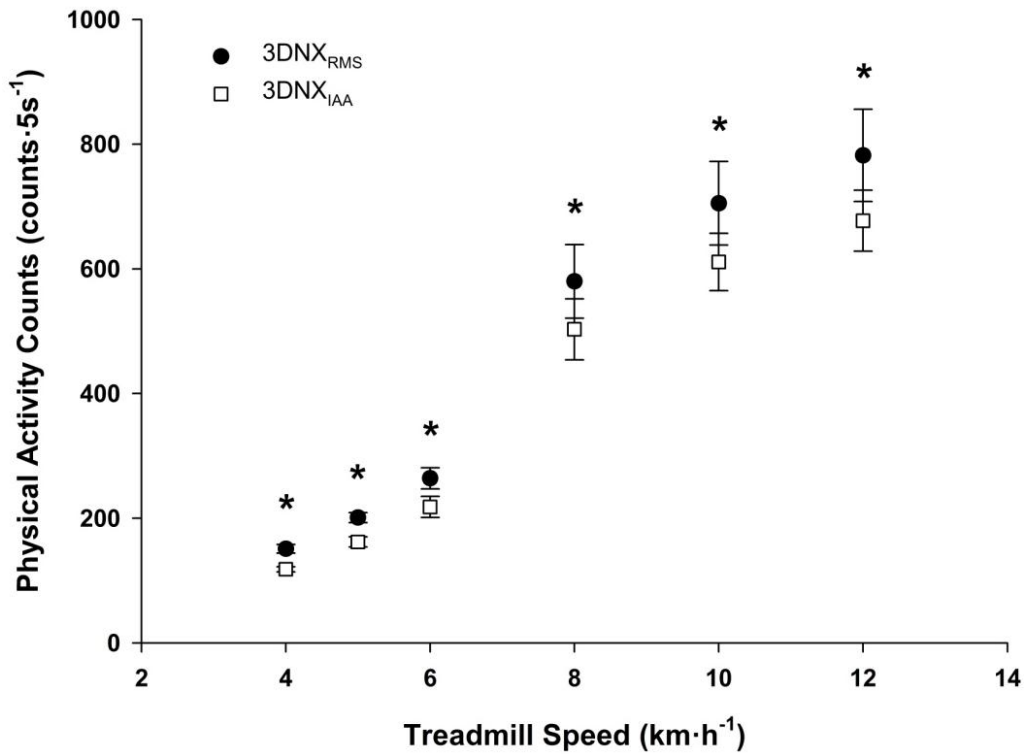
Figure 5.1 Mean PAC values for $3DNX_{IAA}$ and $3DNX_{RMS}$. * denotes a significant difference between integration methods at the $p<0.001$ level.

The relationship between PAC for both integration methods and $\dot{V}O_2$ is shown in Figure 5.2. The relationship between $3DNX_{IAA}$ PAC and $\dot{V}O_2$ was $r^2=0.94$ with a SEE of 2.89 ml.kg^{-1}.min^{-1} (11.1 %). The relationship between PAC generated by the $3DNX_{RMS}$ and $\dot{V}O_2$ was also strong ($r^2=0.94$) with a SEE of 2.93 ml.kg^{-1}.min^{-1} (11.3 %).
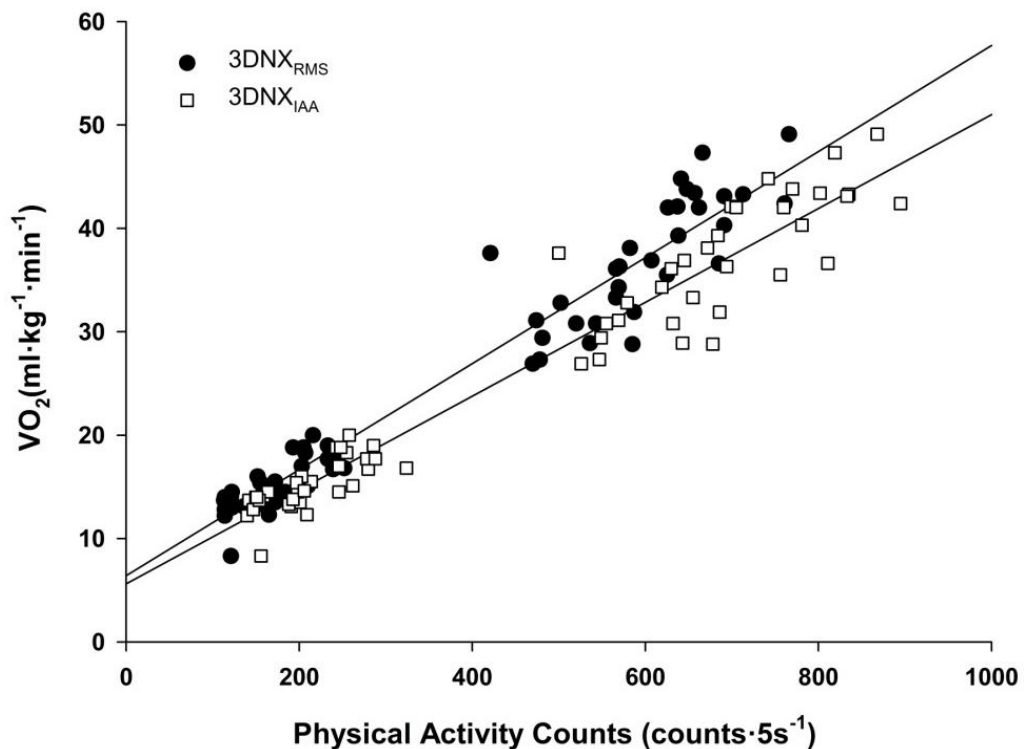
Figure 5.2 Relationships between PAC and oxygen uptake for the 3DNX$_{IAA}$ and 3DNX$_{RMS}$ units.

5.4   DISCUSSION

This short pilot study aimed to compare the validity of two methods used to integrate acceleration signals that are commonly cited in the accelerometer literature. The ability of each device to predict oxygen uptake was similar with the 3DNX$_{RMS}$ returning consistently higher count values but no greater variability indicated by similar SEEs of 11.1 % and 11.3% for the 3DNX$_{IAA}$ and 3DNX$_{RMS}$ units respectively.

The interaction effect observed between unit version and speed indicates that a different relationship exists between 3DNX$_{RMS}$ and oxygen uptake rather than simply generating consistently higher PAC than the 3DNX. The difference in PAC between the two units became larger with increasing speed with the prediction accuracy becoming weaker for both units. These findings show that both unit versions could not be used interchangeably and indicates that a simple linear correction factor would not be suitable to convert 3DNX$_{IAA}$ PAC to 3DNX$_{RMS}$ PAC. These data also indicate a logarithmic transformation of variables would be appropriate to attenuate the heteroscedasticity observed if these relationships were to be used in future studies (Nevill *et al.*, 1995b).

In order to fully establish which processing method is the most suitable for assessing human EE, a 3DNX unit modified using the SVM processing method should also be developed and compared to the current $3DNX_{IAA}$ model. Further comparison work should be undertaken during activities of daily living where movements are more complex. However, it does not appear that the $3DNX_{RMS}$ method is any more valid then the $3DNX_{IAA}$ method for predicting human metabolic rate during treadmill exercise, despite an arguably more intuitive theoretical basis. Future research will be conducted using the unmodified $3DNX_{IAA}$ unit.

6 RELIABILITY AND VALIDITY OF THE 3DNX ACCELEROMETER DURING MECHANICAL AND HUMAN TREADMILL EXERCISE TESTING

6.1 INTRODUCTION

The association between physical inactivity and obesity-related ill-health is now well established (Pate, *et al.*, 1995). To characterise the relationship between physical activity and health outcomes among different free-living populations, and to provide evidence-based recommendations, objective measures of physical activity and associated EE are required (Maddison, *et al.*, 2009).

The doubly labelled water (DLW) method is considered the gold standard for assessing total EE but its use is often limited to small experimental studies for practical and financial reasons (Plasqui, *et al.*, 2007). Furthermore, the method provides only a mean estimate of EE over a series of days, and no information regarding the pattern (i.e. frequency, intensity and duration) of physical activity and associated EE (Pober *et al.*, 2006).

In recent years, technical developments have allowed the expansion of physical activity assessment and body-borne accelerometers have emerged as valuable tools for objectively assessing frequency, duration and intensity of activity (Chen, *et al.*, 2005). However, it is still difficult to accurately translate accelerometry data into units of expended energy (Chu, *et al.*, 2007). Widespread use of accelerometry in population-based studies eliminates the extensive staffing required for observation methods (McKenzie, 2010) and avoids recall problems and bias associated with self-report survey methods. However, the quality of information received from accelerometers is only as good as the reliability and validity of the devices themselves (Esliger, *et al.*, 2006).

Few commercially available triaxial accelerometers are available on the market today, one of which is the 3DNX (BioTel, Bristol, UK). Two laboratory studies have reported favourably on the prediction of metabolic rate (i.e. $\dot{V}O_2$) by the 3DNX model version 2 (v2) during treadmill exercise (Fudge, *et al.*, 2007; Richmond *et al.*, 2007). The capability of the 3DNX to predict EE in free-living adult and adolescent cohorts has also been examined using the DLW method as a criterion measure. Carter et al. (2008) reported that

35% of the variance in TEE in young adults was explained by the 3DNX output; this increased to 78% when combined with anthropometric variables.

The 3DNX model version 3 (v3) is smaller and lighter than the v2. Minor software changes mean that as v3 emerges on the market it is important that its validity and reliability be assessed as the quality of information received from accelerometers is only as good as the reliability and validity of the devices themselves (Esliger*, et al.*, 2006). Current accelerometer reliability research comprises studies using mechanical devices and/or human applications, either in the laboratory or under free-living conditions. The use of a mechanical set-up allows the precise control of the amplitude and the frequency of the oscillation, which are important variables in relation to accelerometer output. Due to the precise control of these experimental conditions, any variability observed can be attributed solely to the accelerometer (Metcalf *et al.*, 2002). The information gathered from mechanical apparatus aids the interpretation of results from human studies as if the measurement error inherent to the accelerometer is low then any variability observed in human reliability data can be attributed to other sources such as the position the accelerometer is worn on the body or day-to-day variation (Esliger*, et al.*, 2006).

The aims of this study are fourfold. The objectives are to 1) assess intra- and inter-unit reliability during mechanical testing in the physiological range of human movement. 2) Assess the technical validity, or more specifically, the individual and combined effects of acceleration and frequency of oscillation on 3DNX output during mechanical testing. 3) Assess the intra- and inter-unit reliability during human treadmill exercise.4) Assess the predictive validity of the 3DNX against metabolic rate during treadmill exercise.

6.2   METHODS

6.2.1   Accelerometer

The triaxial 3DNX model v3 (BioTel Limited, Bristol, UK) is described in detail in Chapter 3. It is sensitive to movements in three planes: X (anteroposterior), Y (mediolateral) and Z (vertical). Accelerometers were set to record at five-second epochs

6.2.2   **Multi Axis Shaker Table (MAST)**

The MAST (MAST-9720, Instron Structural Testing Systems Ltd., UK) is powered by three vertical, one horizontal and two lateral hydraulic actuators. Only the horizontal actuator was used during this testing protocol (Figure 6.1).
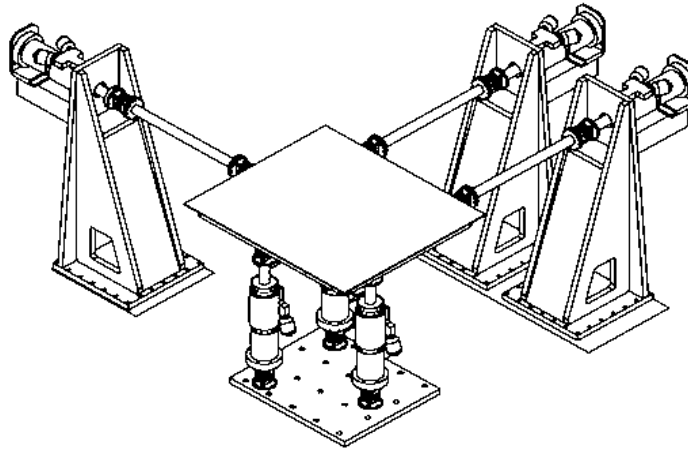
Figure 6.1 Schematic of the MAST 9720

The accelerometers were attached securely to the aluminium alloy mounting plate using double-sided floor tape (DS Scrim 306/250, Tape Range distributors Ltd, UK) shown in Figure 6.2. The accelerometers were rotated before each test in order that horizontal motion of the MAST corresponded to movement along the X-, Y- and Z-axes of the 3DNX.
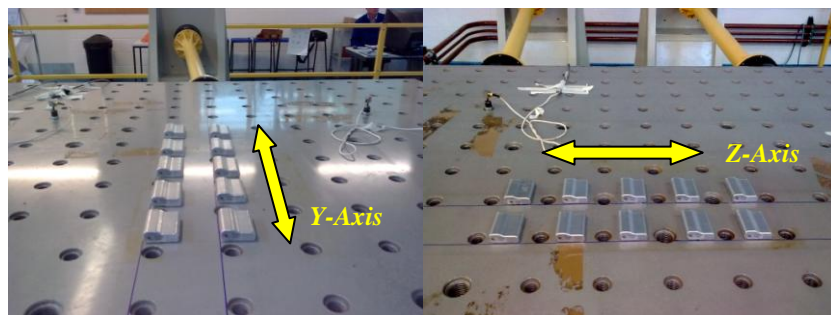
Figure 6.2. The 3DNX units configured in the Y and Z measurement axes.

6.2.3   **Experiment 1 – Technical Validity**

The MAST testing conditions were chosen to replicate a range of physiologically relevant movements which typically occur within a ±6*g* range at waist level (Bhattacharya *et al.*, 1980) and have frequency components which tend to fall between 0.3 and 3.5Hz (Sun*, et*

al., 1993). The protocol was pre-programmed using the MAST's dedicated software (RS Replay, Instron Structural Testing Systems Ltd., High Wycombe, UK).

In total, ten units were subjected to the testing schedule (Table 6.1) in all three axes. This was repeated on two occasions separated by no more than five-minutes. Initially one trial run was completed without logging the data to ensure the MAST was functioning at an optimum operating temperature. One unit was excluded following tests as the battery casing worked loose causing failure mid-test. The maximum displacement amplitude of the MAST (60 mm) limited the testing conditions to a maximum acceleration of 1.5 g. The MAST rig is calibrated regularly to an accuracy of 0.1 g.

By keeping acceleration constant at 4.90m·s^{-2} in conditions 4, 5 and 6 while changing amplitude and frequency, an independent assessment of the effect of frequency on count magnitude was conducted. All raw accelerometer data were downloaded using dedicated software (Biotel Limited, Bristol, UK) and imported into a spreadsheet where 5-second epoch accelerometer counts were averaged over each stage. As each condition was one minute long, this was typically the mean of ten values after excluding the first and last epochs of each stage.

Condition	Duration (min)	Frequency (Hz)	Acceleration (g)	Acceleration (m·s^{-2})
1	1	0.5	0.08	0.79
2	1	0.75	0.17	1.67
3	1	1	0.29	2.85
4	1	1.5	0.5	4.90
5	1	2	0.5	4.90
6	1	2.5	0.5	4.90
7	1	2	1.0	9.81
8	1	2.5	1.0	9.81
9	1	4	1.0	9.81
10	1	2.5	1.5	14.72
11	1	4	1.5	14.72

Table 6.1 MAST testing protocol

6.2.4   Experiment 2 – Human Validity

Ethical approval was obtained from the University of Birmingham research ethics committee through chair's action as a similar protocol had previously been approved prior to the start of the current project.

five male and six female, recreationally-active participants (mean ± s: age 22 ± 3 yr; height 173 ± 9 cm; mass 71 ± 11 kg) completed two incremental treadmill protocols separated by no more than five days on a pre-programmed, motorised treadmill (Woodway ELG70). The protocol consisted of three five-minute walking stages (4, 5 and 6 $km \cdot h^{-1}$), three five-minute jogging stages (8, 10 and 12 $km \cdot h^{-1}$) and four 30-second sprints (14, 16, 18 and 20 $km \cdot h^{-1}$). Each participant wore two 3DNX units on an elastic belt, positioned firmly in the small of the back. The counts from each axis were stored every five seconds (5s). The rate of oxygen uptake was measured during the last minute of all walking and jogging stages using the Douglas Bag technique described in Chapter 3. All raw accelerometer data were downloaded using dedicated software (Biotel Limited, Bristol, UK) and imported into a spreadsheet where the five-second epoch accelerometer counts from X-, Y- and Z-axes were summed. Values were then averaged over each stage where both the first and the last two epochs of the jogging stages, and the first and last epochs of the sprinting stages, were discarded. This ensured that all data used for analysis were not affected by a change of movement velocity.

6.2.5   Statistical Analyses

Experiment 1 - Technical Reliability

For each unit in each condition and each axis (297 in total), the coefficient of variation (CV_{intra}) was calculated as a measure of intra-unit reliability. Bland and Altman plots and absolute limits of agreement (LoA) were generated for each axis combining data from all nine units.

Inter-unit reliability has been noted by previous investigators to be problematic when comparing multiple units (Metcalf*, et al.*, 2002). In this paper an approach similar to that of Brage et al. (2003) was used to determine the variability between accelerometers. First, the relative difference of the unit means to the overall mean (3DNX_All) was calculated for each condition and plotted for visual inspection (modified Bland and Altman plot). Second, coefficients of variation (CV_{inter}) and intraclass correlation coefficients were calculated using unit mean values and 3DNX_All.

Technical Validity

Technical validity was established using simple linear regression for each axis (n=198) and all axes combined (n=594). The correlation coefficients provide an expression of criterion-related validity. Standard error of the estimate (SEE) was calculated for all relationships.

Experiment 2 - Human Reliability

Mean ratio bias limits of agreement, coefficients of variation and intraclass correlation coefficients were generated both intra- and inter-unit.

Human Validity

Simple linear regression analysis was used to examine the relationship between oxygen uptake and accelerometer counts for both units combined and individual units. Standard error of the estimate was calculated for all relationships.

6.3   RESULTS

6.3.1   Experiment 1 - Technical reliability

Mean ($\pm$ SD) accelerometer counts for each of the nine units tested and all units combined (3DNX_All) are listed in Table 6.2, for the X-axis, as an example. Intra-unit reliability is expressed in Table 6.2 as CV (%). Intra-unit reliability was good with CV_{intra} for X-, Y-, and Z-axes ranging from 0.0-8.9%, 0.0-5.7% and 0.0-6.4%, respectively. The absolute bias $\pm$ 95% LoA were 0.00 (p>0.05) $\pm$ 2.90 counts·5 s^{-1}, -0.12 (p>0.05) $\pm$ 2.94 counts·5 s^{-1} and -0.23 (p>0.05) $\pm$ 3.86 counts·5 s^{-1} for X-, Y-, and Z-axes, respectively (Figure 6.3).

Individual units displayed low variability across all conditions with CV_{inter} remaining below 4.5%. CV_{inter} ranged from 1.4–3.1%, 2.3–4.0% and 0.9–4.5% for X-, Y- and Z-axes, respectively. Mean CV_{inter} for all units in all axes and all tests was 2.61%. Intraclass correlation coefficients were 1.0 for all stages in all axes. Inter-unit differences relative to 3DNX_All are shown in Figure 6.4 for the X-, Y- and Z-axes.

Acceleration (g)	f (Hz)	Amplitude (mm)	Test	Unit 1 Mean (SD)	CV	Unit 2 Mean (SD)	CV	Unit 3 Mean (SD)	CV	Unit 4 Mean (SD)	CV	Unit 5 Mean (SD)	CV	Unit 6 Mean (SD)	CV	Unit 7 Mean (SD)	CV	Unit 8 Mean (SD)	CV	Unit 9 Mean (SD)	CV	3DNX_All Mean (SD)
0.08	0.50	62	1	10 (0.0)	3.2	12 (0.5)	0.0	13 (1.2)	8.9	11 (0.7)	4.7	10 (0.0)	0.0	13 (0.8)	6.7	13 (0.5)	6.1	12 (0.7)	4.9	10 (0.0)	0.0	**12 (1.2)**
			2	11 (0.0)		12 (0.7)		14 (1.5)		10 (0.0)		10 (0.0)		13 (0.8)		13 (0.5)		10 (0.4)		10 (0.2)		**12 (1.3)**
0.17	0.75	62	1	30 (0.9)	1.6	30 (0.9)	2.2	30 (1.1)	2.2	30 (0.6)	1.2	31 (0.8)	1.2	31 (1.4)	2.0	31 (1.3)	1.3	30 (0.9)	1.0	30 (0.7)	0.7	**30 (0.2)**
			2	31 (0.5)		30 (0.7)		30 (0.7)		31 (0.9)		30 (0.7)		31 (1.0)		31 (0.9)		30 (0.9)		30 (0.7)		**30 (0.5)**
0.29	1.00	62	1	55 (0.0)	0.0	55 (0.0)	0.0	55 (0.3)	0.0	55 (0.0)	0.9	55 (0.0)	0.0	55 (0.0)	1.8	55 (0.0)	0.0	55 (0.0)	0.0	55 (0.0)	0.0	**55 (0.0)**
			2	55 (0.6)		55 (0.9)		55 (0.0)		56 (0.8)		55 (0.0)		57 (1.5)		55 (0.0)		55 (0.6)		55 (0.0)		**55 (0.8)**
0.5	1.5	55	1	115 (0.0)	0.5	115 (0.0)	0.4	112 (1.5)	0.8	113 (1.0)	0.9	117 (0.7)	0.4	115 (0.5)	0.4	114 (0.7)	0.3	114 (0.5)	0.4	116 (0.8)	0.8	**114 (1.6)**
			2	115 (0.8)		117 (0.5)		115 (0.5)		115 (0.5)		115 (0.0)		116 (0.7)		116 (0.3)		116 (0.4)		113 (0.9)		**115 (1.1)**
0.5	2	31	1	115 (0.3)	0.2	114 (0.4)	0.6	110 (0.3)	0.5	112 (1.0)	0.7	114 (1.0)	0.8	115 (0.0)	0.6	112 (1.0)	0.6	115 (0.4)	0.4	115 (0.3)	0.3	**113 (1.8)**
			2	110 (0.0)		113 (1.3)		112 (0.7)		114 (0.7)		114 (1.3)		111 (0.9)		114 (0.7)		111 (0.7)		110 (0.3)		**112 (1.8)**
0.5	2.5	20	1	115 (0.5)	0.4	115 (0.5)	0.4	114 (0.5)	0.5	114 (0.7)	0.5	114 (0.7)	0.5	115 (0.4)	0.4	112 (0.5)	0.6	113 (0.9)	0.5	113 (0.5)	0.4	**114 (0.9)**
			2	114 (0.4)		115 (0.5)		114 (0.5)		115 (0.5)		115 (0.5)		114 (0.5)		113 (0.7)		113 (1.0)		115 (0.5)		**114 (0.8)**
1.00	2	62	1	232 (0.3)	0.2	231 (0.3)	0.3	229 (0.7)	0.2	233 (0.0)	0.1	232 (0.4)	0.2	233 (0.4)	0.3	234 (0.4)	0.3	232 (0.3)	0.2	232 (0.5)	0.2	**232 (1.3)**
			2	231 (0.3)		231 (0.7)		230 (0.4)		234 (0.4)		233 (0.5)		232 (0.5)		235 (0.5)		231 (0.5)		231 (0.3)		**232 (1.6)**
1.00	2.5	40	1	233 (0.3)	0.3	235 (0.7)	0.3	233 (0.5)	0.2	233 (0.7)	0.2	232 (0.7)	0.2	233 (1.1)	0.6	231 (0.5)	0.2	231 (0.5)	0.2	231 (0.5)	0.2	**232 (1.3)**
			2	233 (0.4)		236 (0.5)		233 (0.7)		233 (0.0)		233 (0.8)		232 (1.1)		231 (0.0)		231 (0.3)		232 (0.3)		**233 (1.6)**
1.00	4	16	1	226 (0.5)	0.2	230 (0.8)	0.3	227 (0.5)	0.2	228 (0.7)	0.2	232 (0.5)	0.4	229 (0.4)	0.1	229 (0.4)	0.1	229 (0.7)	0.3	226 (0.5)	0.2	**228 (1.8)**
			2	227 (0.6)		229 (0.4)		226 (0.0)		227 (0.5)		230 (0.7)		229 (0.5)		229 (0.0)		230 (0.4)		227 (0.5)		**228 (1.4)**
1.50	2.5	60	1	352 (0.5)	0.1	360 (0.5)	0.1	353 (0.5)	0.1	353 (1.1)	0.2	357 (0.7)	0.2	354 (0.5)	0.2	357 (1.1)	0.2	354 (0.5)	0.2	356 (0.5)	0.1	**355 (2.5)**
			2	354 (0.5)		359 (0.9)		355 (0.0)		355 (0.0)		356 (0.4)		353 (0.7)		355 (0.4)		352 (0.7)		355 (0.3)		**355 (2.0)**
1.5	4	23	1	330 (0.7)	0.3	337 (0.5)	0.2	333 (0.4)	0.2	333 (0.0)	0.2	338 (0.4)	0.2	335 (0.5)	0.2	335 (0.9)	0.2	335 (0.7)	0.2	331 (0.5)	0.3	**334 (2.6)**
			2	331 (0.9)		336 (0.3)		334 (0.7)		330 (0.5)		338 (0.0)		336 (0.0)		333 (0.6)		335 (0.0)		332 (0.7)		**334 (2.6)**

Table 6.2 X-axis descriptive statistics. All values are average counts•5s^{-1}.
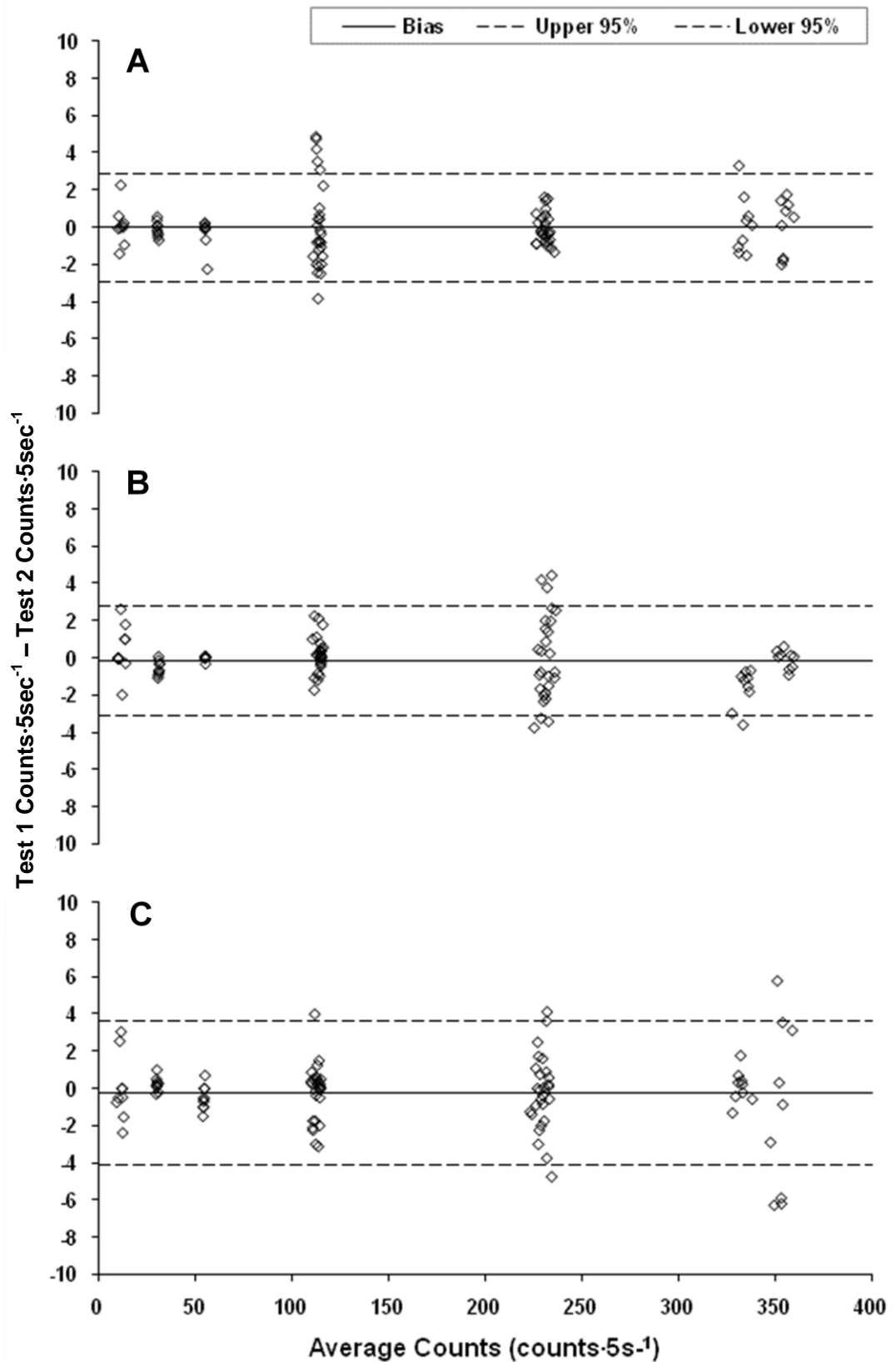
Figure 6.3. Figures 1.3A, 1.3B and 1.3C. Bland and Altman plots showing the absolute bias and 95% Limits of Agreement for the X- , Y and Z-axes.
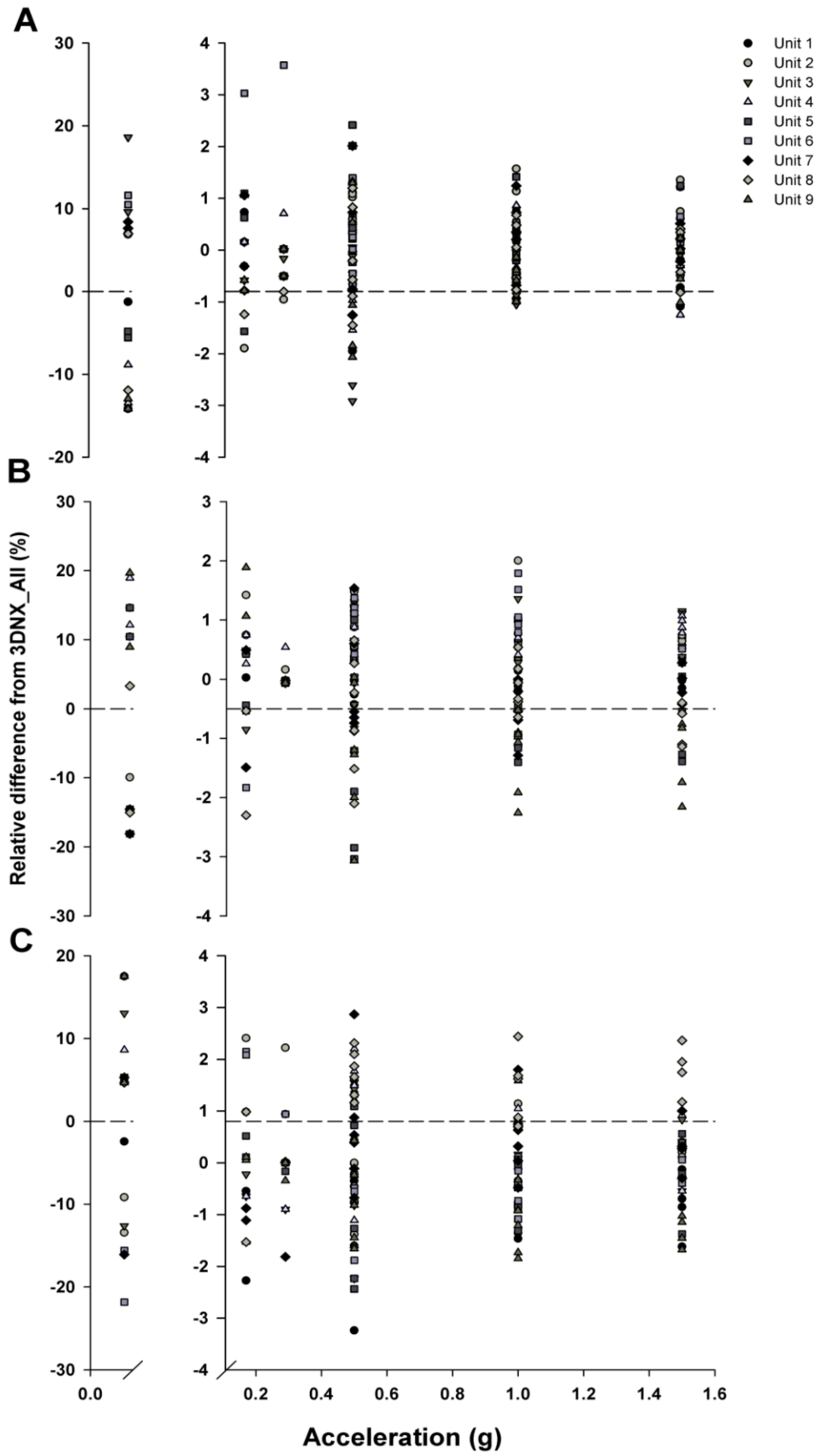
Figure 6.4. Figures 1.4A, 1.4B, and 1.4C - Modified Bland and Altman plots showing the relative difference between individual units and the overall mean (3DNX_All) for the X-, Y-, and Z-axes.

6.3.2   Technical Validity

Figure 6.5 shows a strong positive linear relationship observed when the average counts from conditions 1 - 11 from all axes in all units are displayed together (n=594). The relationship yielded an r^2 value of 0.99 and a standard error of the estimate (SEE) of 6 counts·5s^{-1}. When the data are reported by individual axis, r^2 values are all 0.99 and SEEs are 6 counts·5s^{-1}.
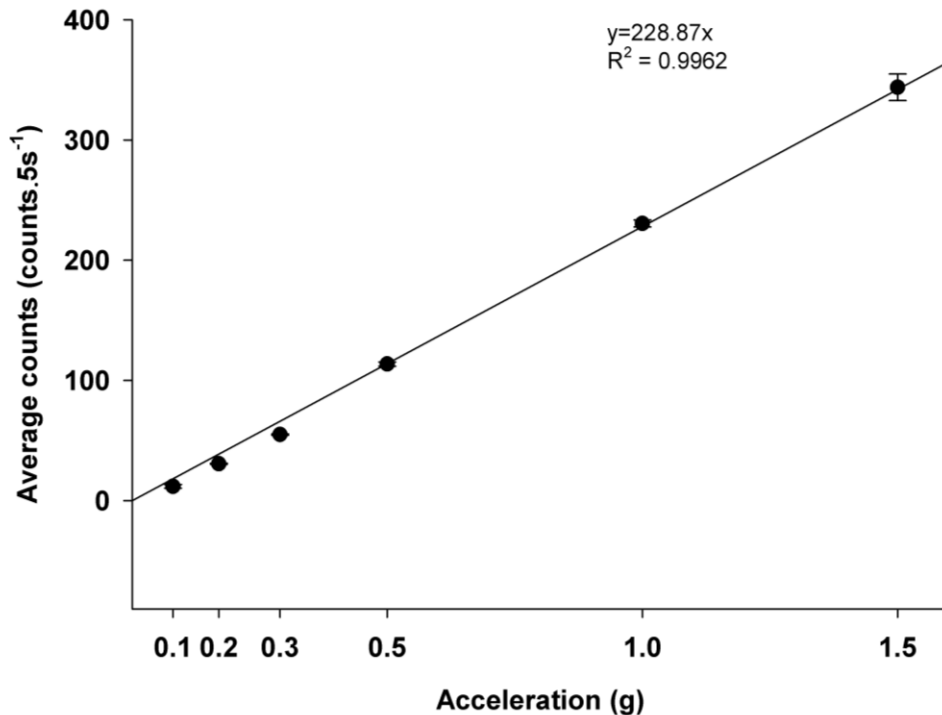
Figure 6.5 Relationship between count magnitude and acceleration (n = 594).

The results from holding acceleration constant at 0.5 g while changing the frequency of oscillation are displayed in Figure 6.6. The 3DNX_All accelerometer counts at 4.9 m·s^{-1} are 114 ± 2 counts·5s^{-1}. Figure 6.6 is a graph adapted from data reported by Esliger and Tremblay (2006) who used a mechanical rig to investigate the reliability and validity of three commercially-available accelerometers. The graph displays data from three models with the addition of the 3DNX data collected in this study, showing the independent effect of changing the frequency of oscillation on count magnitude. 3DNX values have been multiplied by 12 so that they correspond to counts·min^{-1}.
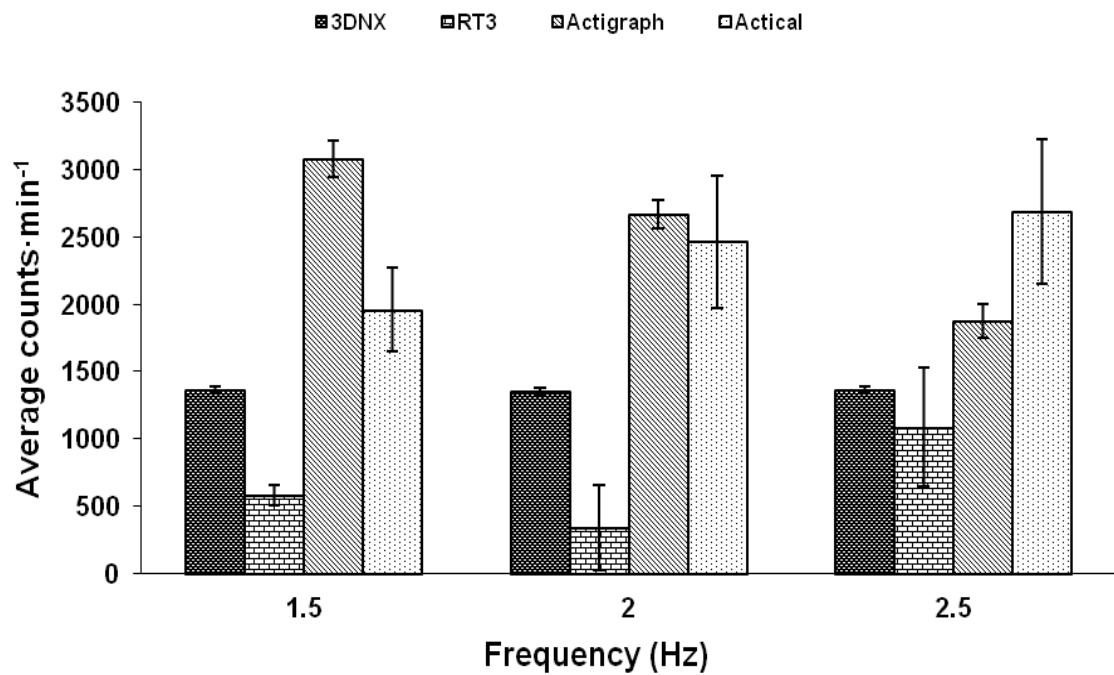
Figure 6.6 Between model comparison of frequency effects on count magnitude and variability (acceleration held constant at 4.9 m·s⁻¹). Data from the present study has been added to data reported by Esliger and Tremblay (2006) for comparison. It should be noted that although these data are expressed in the same units (counts·min⁻¹), each device generates a "count" using different processing techniques.

6.3.3   Human reliability

Table 6.3 shows the descriptive statistics and intra-unit reliability for both trials. The variability between units was highest during walking stages with coefficients of variation for inter-unit reliability ranging from 7.7 to 16.0 % for trial one and 7.7 to 16.2 % for trial two. Intraclass correlation coefficients between units ranged from 0.95 to 1.00 for trial one and 0.90 to 0.99 for trial two. Mean ratio bias ± 95 % limits of agreement for intra-unit (n=220) and inter-unit (n=220) reliability were -0.7 % (p>0.05) ± 12.4 % and 0.4 % (p>0.05) ± 5.6 %, respectively.

Treadmill speed $(Km \cdot hr^{-1})$	Trial	$\dot{V}O_2$ (ml·min⁻¹·kg⁻¹)		Unit 1 (counts·5s⁻¹)				Unit 2 (counts·5s⁻¹)			
		Mean	SD	Mean	SD	CV	ICC	Mean	SD	CV	ICC
4	1	12.3	1.3	126	17	16.0	0.91	122	16	16.0	0.91
	2	11.8	0.9	127	20			124	20		
5	1	14.3	0.9	171	23	15.1	0.91	165	23	16.2	0.89
	2	14.1	1.1	172	24			168	26		
6	1	17.4	0.9	235	29	13.0	0.87	227	29	14.0	0.90
	2	17.0	1.1	238	29			234	30		
8	1	30.4	2.2	538	40	8.0	0.56	542	39	8.3	0.71
	2	29.6	1.7	543	41			547	44		
10	1	36.8	2.2	640	53	7.7	0.56	637	48	7.7	0.64
	2	36.7	1.7	641	41			642	47		
12	1	43.7	2.5	706	60	7.8	0.65	700	56	7.7	0.74
	2	42.9	2.2	705	44			704	46		
14	1	54.6	0.3	751	64	8.2	0.77	740	61	8.3	0.73
	2			744	55			745	57		
16	1	61.6	0.3	802	79	9.3	0.78	792	76	9.4	0.83
	2			798	63			797	65		
18	1	68.6	0.3	860	96	10.1	0.68	848	88	10.0	0.77
	2			854	69			851	73		
20	1	75.6	0.3	930	107	10.9	0.80	927	112	11.8	0.78
	2			928	86			928	94		

Table 6.3 Descriptive statistics for units 1 and 2. Coefficient of variation (CV) and Intraclass correlation coefficient (ICC) values are intra-unit. Due to equipment limitations it was not possible to measure $\dot{V}O_2$ values for 14, 16, 18 and 20 km·hr⁻¹. Estimates have been generated using the equation by Hagan et al. (1980).

6.3.4   Experiment 2 - Human validity

The relationship between $\dot{V}O_2$ expressed as ml·kg⁻¹·min⁻¹ and average 3DNX counts from units 1 and 2 (Figure 6.7) was linear during walking (r^2=0.65, SEE=1.42 ml·kg⁻¹·min⁻¹) and running (r^2=0.62, SEE=3.63 ml·kg⁻¹·min⁻¹).

When relative $\dot{V}O_2$ values are converted to METs by dividing by 3.5, cut-scores can be established that correspond to ACSM guidelines for light (< 3 METs), moderate (3 − 6 METs) and vigorous (> 6 METs) exercise. These are < 106, 106–281 and > 281 counts·5s⁻¹ for light, moderate and vigorous zones, respectively.
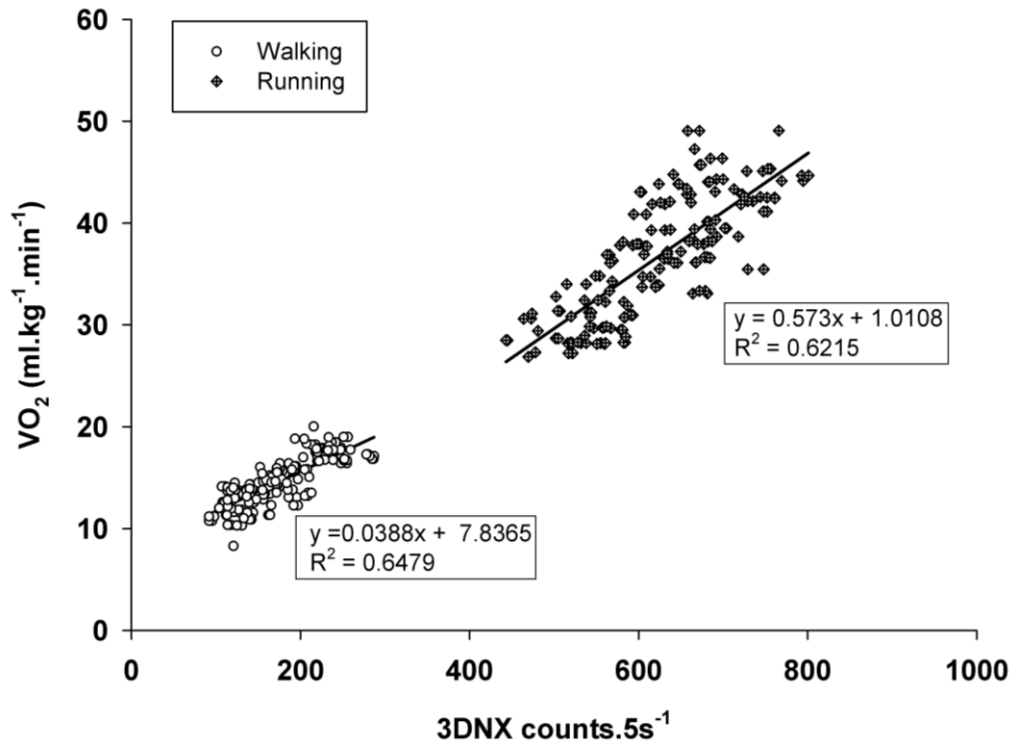
Figure 6.7 Relationship between average counts and oxygen uptake for walking and running (n = 264).

6.4   DISCUSSION

Body-born accelerometers are emerging as valuable tools for the prediction of EE. When attempting to characterise the relationship between physical activity and health in free-living populations it is imperative that the device chosen to measure physical activity is valid and reliable. This study aimed to test extensively the reliability and validity of 3DNX v3 triaxial accelerometer in a mechanical setting and during treadmill exercise. The unit yielded promising results in both.

6.4.1   Experiment 1 - Technical reliability

All nine units, regardless of measurement axis, demonstrated high intra- and inter-unit reliability. Inter-unit coefficients of variation (CV_{inter}) in the present study compared well to those reported for the $RT3^{TM}$ (Stayhealthy Inc, Monrovia, CA, USA) - another commercially-available triaxial accelerometer. Powell et al. (2003) reported inter-unit CVs ranging from 6.3 % to 38.5 % across two similar frequencies for the RT3 whereas CV_{inter} for the 3DNX did not rise above 4.5 % in any measurement axis. More recently, Esliger

and Tremblay (2006) reported slightly poorer values for the RT3 to those of Powell et al. (2003) citing a mean CV_{inter} of 42.9% in comparable testing conditions. The 3DNX displayed higher inter-unit variability at the lower accelerations due to variations from the mean of one or two counts equating to a larger percentage of the mean count value. Other units have displayed similar trends (Brage, *et al.*, 2005; Brage, *et al.*, 2003; Powell, *et al.*, 2003) but in real terms such low absolute variability at the lowest accelerations is unlikely to impact on the overall classification of physical activity. High inter-unit reliability is particularly desirable for studies attempting to compare across a population. If poor technical reliability is inherent between units then it becomes unclear whether any variability found within a human population can be attributed solely to variation in behaviour. Individual unit calibration is a potential solution but may be impractical on a large scale, indeed the literature differs in relation to whether individual calibration is worth the effort. Brage et al. (2003) noted that for the CSA 7164 (Computer & Science Applications Inc., MTI, Fort Walton Beach FL, USA), the magnitude of inter-unit variability translated *in vivo* to up to 20% difference in activity counts for walking and 40% for running. This would lead to a distortion of the classification of which subjects met the recommended daily amount of PAEE and which did not. In contrast, Moeller et al. (2008) concluded that calibration of the Actigraph 7164 using a mechanical rig was almost ineffectual when applied to free-living data due to other, more dominant sources of variation.

Intra-unit reliability becomes most important when trying to characterise changes in activity profiles over time during intervention studies. The 3DNX units performed well across all accelerations. CV_{intra} ranged from 0.0-8.9 %. These results are slightly poorer than the intra-unit reliability expressed as CV_{intra} reported by Esliger and Tremblay (2006) for the Actical (0.2-1.1 %), similar to the Actigraph 7164 (0.2-9.5 %) and better than the RT3 (36.2-106.9 %). Intra-unit variability is best observed in the Bland and Altman plot (Figure 6.3) with 95% LoA for all units in all axes within ±3.86 counts.5s^{-1} between test one and two.

It is unusual for CV_{intra} values to be higher than CV_{inter} values. Although the MAST performed a dummy run of the protocol to warm up it is possible that it was not fully warm at the start of the experimental protocol. The actual accelerations that the units were subjected to may have been slightly different towards the end of testing. Furthermore,

although care was taken to attach each unit as securely as possible to the MAST it is possible that the double-sided tape lost some adhesiveness by the final trial which may have allowed some movement during stages seven and eight. This could account for the heteroscedasticity, most easily observed in the Z-axis plot (Figure 6.3).

6.4.2   Technical Validity

The results from experiment one indicate that the 3DNX demonstrated good validity, expressed as a linear correlation with increasing acceleration. The reported linearity of 0.99 between 0.08 and 1.5 *g* is in accordance with most industry-standard accelerometers (Chen*, et al.*, 2005).

In addition to the data collected in this study, data reported by previous researchers in similar mechanical set-ups provide valuable sources of comparison between commercially-available devices. Outputs from the Actigraph 7164 and Actical (Esliger*, et al.*, 2006), RT3 (Powell*, et al.*, 2003) and the CSA 7164 (Brage*, et al.*, 2003) have all been shown to be dependent on movement frequency in a mechanical set-up, whereas 3DNX count output remained unchanged at 0.5 *g* when subjected to a range of frequencies and amplitudes. It should be noted that due to the limitations of the MAST, it was not possible to create testing conditions near to the upper band pass filter of the 3DNX. The comparison between the Actigraph, Actical and RT3 should not be taken at face value as frequencies of 1.5, 2 and 2.5 Hz are well within the frequency range of the 3DNX whereas they are either below or close to the lower limit of the RT3 band pass filter (2 Hz) and near to high band pass filter of the Actigraph and Actical. The salient discussion point is that users should be aware of the frequency and dynamic range of the devices they are using and whether they are appropriate for their chosen study population thereby avoiding loss of meaningful data.

The current study is one of only a few that have scrutinised the mechanical validity of commercially-available units; by contrast there are numerous human validation studies (Trost*, et al.*, 2005). We contend that all accelerometers should undergo a similar protocol to the one outlined in this paper to ensure that when human variability is removed, accelerometer counts are directly related to acceleration. As earlier studies have demonstrated, a unit can possess strong inter- and intra-unit reliability but may show disconcerting results when an independent assessment of frequency is conducted (Esliger*, et al.*, 2006).

6.4.3   Study 2 - Human Reliability

Following the results of the mechanical testing it can be inferred that any variability in the 3DNX observed in humans can be attributed mainly to biological variables.

Under laboratory conditions inter-unit reliability was acceptable across all treadmill speeds (ICC = 0.90 – 1.00, CV = 6.6 – 17.4%). The 3DNX compares well to the results reported by Krasnoff et al. (2008) (CV = 9.5 – 34.5%) for the triaxial RT3^{TM}. As observed in other accelerometer models, reliability of the 3DNX at the two slowest walking speeds was poorer than at the jogging speeds (Trost *et al.*, 1998). Although both units were placed in the small of the back (they were positioned side by side) and as such, may have been subjected to contralateral variability which is more apparent at slower speeds (Zijlstra, 1997).

Many population studies measure physical activity over a number of separate days. This makes the test-retest reliability of any one unit particularly important. In this study it was only practical to attach two units to each participant due to participant comfort and the difficulty of firmly fixing more than two devices. With coefficients of variation ranging from 7.7 - 16.2 % (mean 10.8 %) the 3DNX compares favourably to the results of Welk (2004) who examined the within-unit reliability of four commercially-available accelerometers during treadmill exercise. The results of this earlier study reported that for a standardised bout of treadmill activity, accelerometers yielded coefficients of variation from 16 – 31 % for each participant on separate days.

6.4.4   Human Validity

The 3DNX was shown to be a valid predictor of $\dot{V}O_2$ during incremental treadmill exercise over a wide range of walking and running speeds. 3DNX counts increased linearly with $\dot{V}O_2$ during walking and running displaying positive relationships in both exercise modes with low standard error. On first inspection, these correlations appear lower than findings elsewhere for other accelerometer units which typically report r^2 values of 0.85 - 0.92 under laboratory conditions (Welk*, et al.*, 2004). However, these values are derived from walking and running data combined leading to a larger range of values and hence a higher r^2 value. When walking and running data are combined in the present study an r^2 value of 0.95 is produced.

To date there have been no other studies reported which have examined the validity of the 3DNX model v3 during treadmill exercise and comparative data are lacking. However, the relationship between counts and $\dot{V}O_2$ from the present 3DNX v3 model can be compared to the 3DNX model v2 in a similar study conducted by Fudge et al. (2007) which yielded a higher r^2 values and larger SEEs (r^2=0.83, SEE=2.20 and r^2=0.76, SEE=4.88) for walking and running respectively. The higher r^2 values reported by Fudge et al. (2007) are likely due to the wider range of treadmill speeds (3-20km·hr^{-1}) where $\dot{V}O_2$ was calculated. The smaller standard error values of the present study indicate slightly improved predictive validity compared to the 3DNX model v2.

While the findings of the current study demonstrate promising capabilities of the 3DNX, the human reliability component of the study would have been more complete if a range of non-ambulatory movements had been included. Also an accurate estimate of the resting component of EE would have permitted a more specific analysis of the relationship between activity counts and PAEE. In addition, a more extensive analysis would have been possible had the raw signal been accessible. In order to fully compare the performance of accelerometer models, acceleration in m·s^{-2} rather than arbitrary "counts" should be compared to a criterion measure. Accelerometer manufacturers should be urged to release these data before applying proprietary algorithms.

In conclusion, the 3DNX is a reliable and valid device for measuring acceleration in a mechanical setting and during treadmill exercise.

7   EFFECT OF ANATOMICAL PLACEMENT AND TRUNK ADIPOSITY ON THE RELIABILITY AND VALIDITY OF TRIAXIAL ACCELEROMETER OUTPUT DURING TREADMILL EXERCISE

7.1   INTRODUCTION

Physical activity (PA) has many different dimensions (e.g. intensity, frequency, duration and type) and can be assessed using a variety of wearable sensors.

Accelerometer-based devices have emerged as a promising tool for assessing free-living human PA and energy expenditure (EE) (Chen, *et al.*, 2005). Accelerometers measure body accelerations in either one (uniaxial), two (biaxial) or three (triaxial) sensing axes. Outputs in the form of physical activity counts (PAC) can subsequently be used to estimate the intensity, duration and frequency of the motion (Chen, *et al.*, 2005). Physical activity counts can be translated into METs and then categorised into light, moderate and vigorous exercise intensity (Freedson, *et al.*, 1998; Vanhelst *et al.*, 2010). These zones are normally established by calibrating accelerometer output against an EE criterion, such as measured oxygen uptake ($\dot{V}O_2$) via, most commonly, linear regression equations during specific activities of varying intensities (Lyden, *et al.*, 2011).

When wearing an accelerometer, users are often instructed to attach devices at the hip in the mid-axillary line. This is true for the uniaxial ActiGraph GT1M (Fort Walton Beach, FL); the most studied commercial accelerometer device (Rothney *et al.*, 2008). In contrast, the triaxial 3DNX^{TM} (Biotel, Bristol, UK) is positioned in the small of the back in accordance with the manufacturer's instructions. It has been argued that the lower back placement is a more intuitive location due to its proximity to the whole body centre of gravity. Furthermore, it has been suggested that count output from hip placement could be differentially affected, depending on the type of activity performed as movements on one side are not always representative of the other (Yngve *et al.*, 2003; Zijlstra, 1997). The effect of placement on the sensors at either the lower back or the hip has previously been assessed (Bouten, *et al.*, 1997; Yngve, *et al.*, 2003) showing a difference in accelerometer output between the hip and back. However, neither study provided any empirical evidence

to explain the differences observed or investigated the level of variability in count output at each placement.

Sources of error in accelerometer outputs can originate from detecting factors other than the intended physical activity. For example, accelerations due to movement of the tissue over which the monitor is placed will affect PAC (Bouten, *et al.*, 1997). This could impact on the reliability and validity of the devices for predicting EE and possibly lead to the misclassification of PA intensity. Brage et al. (2003) suggested that individual characteristics such as body size, hip geometry and the amount of fat deposition on the hip may explain the large inter-instrument variability that they found in the uniaxial Computer Science Applications (CSA) monitor. Light et al. (1980) suggested that fixing accelerometers to soft tissue overestimates the amplitude of the accelerometer output compared with fixing the device to a bony landmark. From this, it could be inferred that thicker skinfold measurements beneath the sensor location may affect the validity of the device to predict metabolic rate because of oscillations that are additional and more variable than the acceleration experienced at the centre of mass. Feito et al (2011) investigated the effect of body mass index (BMI) on the accuracy of activity monitor step count function and found no main effect of BMI on counting accuracy; however, to our knowledge, there are no previously published studies specifically investigating the relationship between trunk adiposity and accelerometer PAC reliability and validity.

The aim of this study was to establish whether there was a difference in the relationship between PAC and $\dot{V}O_2$ (i.e. predictive validity) when accelerometers were positioned at the hip or lower back. A secondary aim of the study was to determine whether individual differences in trunk adiposity mediated the relationship between $\dot{V}O_2$ and PAC. Finally, intra-instrument reliability at each landmark was also assessed.

It was hypothesised that PAC would be higher and more variable at the hip and that there would be a difference in the PAC-$\dot{V}O_2$ relationship between placements. It was further hypothesised that measures of abdominal/hip adiposity would contribute to the prediction of $\dot{V}O_2$ from PAC at the hip.

7.2   METHODS

7.2.1   Accelerometry

The triaxial 3DNX[TM] model v3 (BioTel Limited, Bristol, UK; www.biotel.co.uk) is described in Chapter 3. Briefly, the device is sensitive to movements in three planes: X (anteroposterior), Y (mediolateral) and Z (vertical). The output from the three sensing axes are integrated in a process described by Bouten et al. (1997) to return a PAC value. The unit dimensions are 54 x 54 x 18 mm with a mass of 70 g including a 3.6 v lithium battery (Saft Ltd., UK). The unit contains two ADXL321 biaxial micro-electro-mechanical (MEMS) sensors (Analog Devices Ltd., Surrey, UK) positioned orthogonally to measure acceleration in three movement planes.

7.2.2   Exercise protocol

Following ethics approval from the University of Bath Research Ethics Approval Committee for Health (REACH), a convenience sample of 26 recreationally active, female participants (mean ± SD: age 20.4 ± 1.3 yr; mass 62.72 ± 6.9 kg; height 1.67 ± 0.05 m) reported to the laboratory on two occasions. No more than two weeks separated each visit to ensure that no significant changes in physical characteristics occurred. During the first visit, anthropometric measurements were taken followed by familiarisation on a motorised treadmill (Woodway ELG 70, Munich, Germany), consisting of walking and jogging. Stature was measured using a stadiometer and body mass was measured using traditional beam scales (Weylux, England). Lower limb length, waist circumference and hip circumference were measured using a fabric tape measure. Skinfold thickness (Skf) was measured using Harpenden skin callipers (British Indicators, West Sussex, UK) from eight marked locations on the right side of the body: biceps, triceps, abdominal, supraspinale, iliac crest, subscapular, anterior thigh and medial calf. All reported values are the mean of three measurements. All anthropometric measurement techniques and sites have been previously defined and described by the International Society for the Advancement of Kinanthropometry (ISAK) (Marfell-Jones *et al.*, 2006).

During the second visit participants were fitted with two 3DNX accelerometers fixed directly to the skin (Hyperfix self-adhesive dressing retention tape, Smith & Nephew Healthcare Ltd., UK) at the hip (midaxillary line) and the lower back (Figure 7.1). A 15 cm^2 patch of tape was used to minimise the risk of introducing error from movement of an

elasticated belt (the normal method of affixation). Participants performed a continuous pre-programmed treadmill protocol consisting of four, 4-minute stages (4, 6, 8 and 10 km·h^{-1}). Participants rested for 20-min before repeating the treadmill protocol; this was considered a suitable recovery period based on the submaximal nature of the exercise.

Unprocessed accelerometer data were downloaded using dedicated software (Biotel Limited, Bristol, UK) and imported into spreadsheets where 5-second epoch accelerometer counts were averaged over each stage. The first and last epochs of each stage were removed to ensure only steady state values were included in the analysis. All data are expressed as counts per five seconds (counts·5s^{-1}).
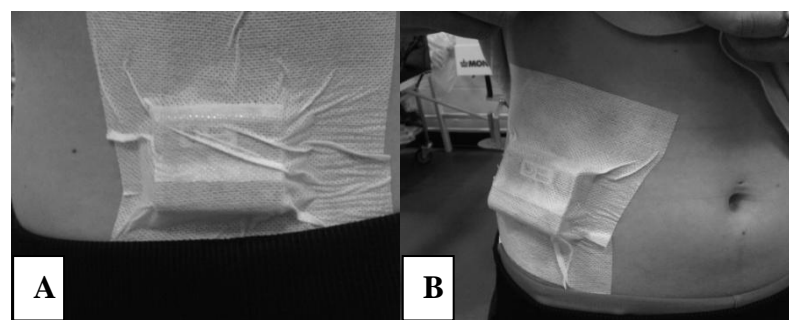
Figure 7.1 3DNX accelerometers placed at A) the lower back and B) the supra-illiac skinfold site.

Expired gases were collected during the last minute of each stage using the Douglas bag technique. Fractions of expired O_2 and CO_2 were assessed using paramagnetic and infrared analyzers, respectively (Servomex 1440, UK). The total volume of gas expired was determined using a dry gas meter (Harvard Apparatus, UK), and the temperatures of expired gases were measured with a digital thermometer (model C, Edale Instruments, UK). All values were corrected to reflect standard temperature and pressure. The analysers were calibrated before the test with gases of known composition and volume within the physiological range.

7.3   STATISTICAL ANALYSIS

Intra-instrument reliability

A custom model analysis of covariance (ANCOVA) was used to test for homogeneity of test one and test two regression slopes. The dependent variable was $\dot{V}O_2$ and the independent variables were PAC and test number (one or two). Where no significant differences were observed, data were subsequently clustered. Intra-instrument coefficients of variation (CV$_{intra}$) were used to compare the variability of the PAC between tests.

Placement and $\dot{V}O_2$-PAC relationship

A custom model ANCOVA was also used to test for an interaction effect of placement on the PAC-$\dot{V}O_2$ relationship with PAC and placement as the independent variables. Where an interaction was found, the data were split for subsequent analyses.

Exercise Mode and $\dot{V}O_2$ -PAC relationship

A custom model ANCOVA was used to test for homogeneity of regression lines for walking and running. This tested for a main effect and an interaction effect of different exercise mode (walking/running) on the PAC- $\dot{V}O_2$ relationship.

Speed and PAC

A 2-way mixed model ANOVA was performed with the four speeds (4, 6, 8 and 10 km·h^{-1}) as repeated measures and placement as the between-subjects variable. This ANOVA was used to test for significant differences in PAC between placements at different speeds. In the event of a main effect, paired t-tests with Bonferroni correction were used post-hoc to locate differences.

Adiposity and $\dot{V}O_2$-PAC relationship at the hip

An ANCOVA testing for main effects was performed using only hip data with PAC as the dependant variable and speed and individual body composition measures as independent variables. This ANCOVA was used to establish which, if any, body composition variables might explain any differences in PAC from the hip compared to the lower back.

Instrument Validity

Multiple linear regression analysis was used to compare the prediction of $\dot{V}O_2$ between placement, exercise mode and the contribution of individual anthropometric measurements. Significance was set *a priori* at p < 0.05.

Statistical analyses were conducted using the software package SPSS, version 18 (SPSS, Inc., Chicago, IL).

7.4   RESULTS

There were 24 completed trials from the 26 participants due to spurious accelerometer data from two trials. Group characteristics are shown in Table 7.1.

	Age (years)	Height (m)	Mass (kg)	BMI	Waist to hip ratio	Leg length (cm)	Abdominal (mm)	Supraspinale (mm)	Iliac crest (mm)
Mean	20.4	1.67	62.7	22.4	0.8	85.5	21.4	14.57	20.42
± SD	1.3	0.05	6.9	2.0	0.1	3.8	7.0	5.12	7.47
MAX	23.0	1.75	75.0	26.7	0.9	93.4	35.3	24.80	36.30
MIN	18.0	1.59	51.6	19.5	0.6	79.2	7.9	7.10	10.20

Table 7.1 Group characteristics (mean ± SD). BMI; body mass index

7.4.1   Intra-Instrument Reliability

Analyses revealed that the assumption of homogeneity was not violated between test 1 and test 2 since there was no significant main effect (p = 0.89 or interaction (p = 0.94) between tests on PAC. For all remaining analyses the data were combined. This demonstrates good intra-unit reliability which was also reflected by the mean CV_{intra} of 3.0 % at the hip (range of 0.0-12.4 %) and 2.8 % at the lower back (range = 0.02-17.41).

7.4.2   Inter-instrument Validity

There was a significant main effect (p = 0.03) and a significant interaction effect (p = 0.02) of placement on the prediction of $\dot{V}O_2$ from PAC. Therefore the assumption of homogeneity of regression slopes was violated and all remaining analyses were performed with the data split by placement. A second ANCOVA performed on data split by placement revealed no significant main (p = 0.26) or interaction effect (p = 0.21) of exercise mode on the $\dot{V}O_2$-PAC relationship (Figure 7.2), indicating that the slopes and intercepts of both relationships were similar for walking and running.
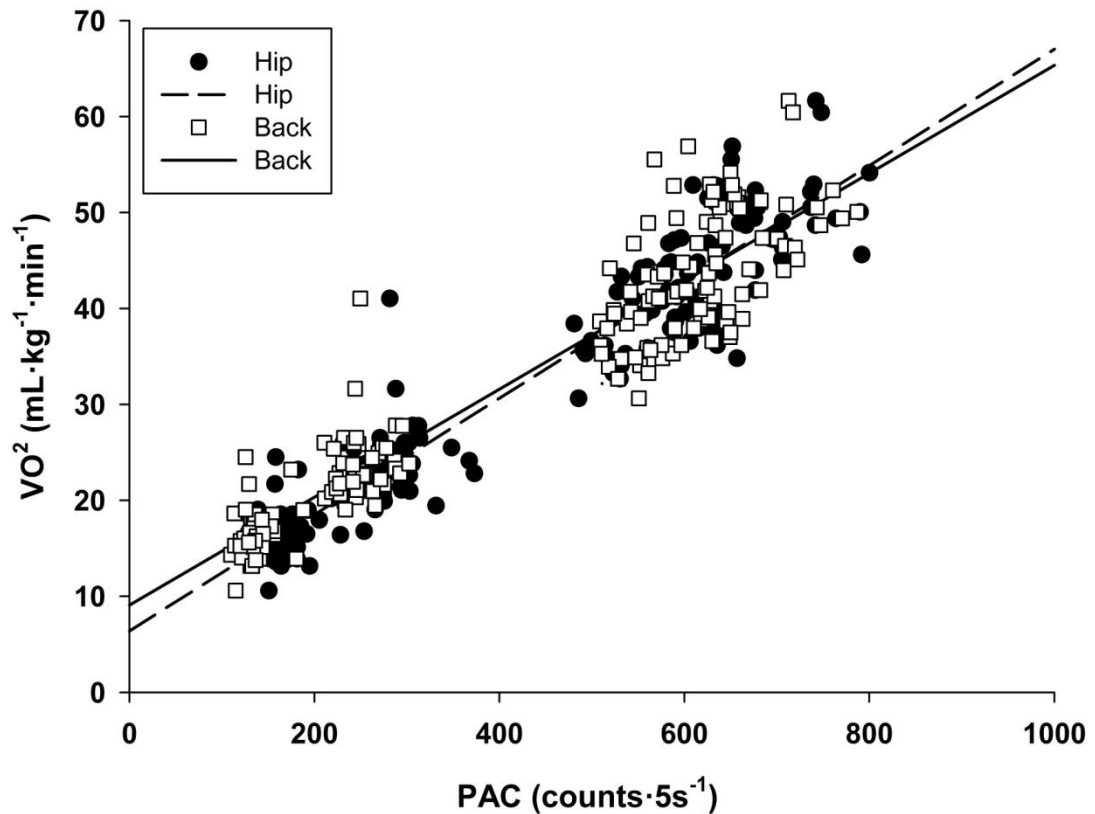
Figure 7.2 The relationship between physical activity counts and $\dot{V}O_2$ from accelerometers at the hip and back.

7.4.3   Placement and PAC

There was a significant main effect of speed ($p < 0.01$) and placement ($p < 0.01$) and an interaction effect of speed and placement ($p < 0.01$) on PAC. Post hoc analyses showed that the difference in PAC between the hip and back was found at speeds 4 and 6 km.h^{-1} ($p < 0.01$) (Figure 7.3). Physical activity counts were higher at the hip (4 km.h^{-1} = 165 ± 22.6, 6 km.h^{-1} = 281 ± 32.1 counts·5s^{-1}) than the back (4 km.h^{-1} = 136 ± 16.0, 6 km.h^{-1} = 249 ± 22.9 counts·5s^{-1}).
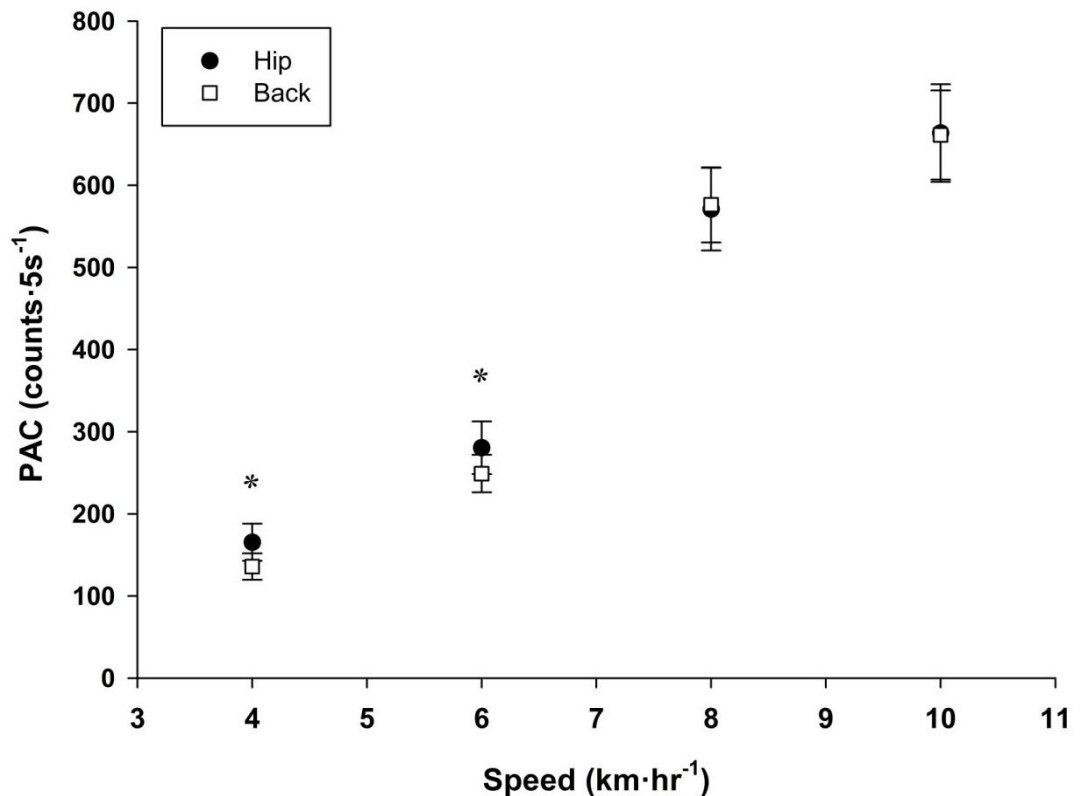
Figure 7.3 The relationship between PAC and speed at the hip and lower back.

* Significant differences in PAC between the hip and lower back ($p < 0.05$).

7.4.4   Body Composition and PAC at the hip

Since the 2-way mixed model ANOVA indicated a significant difference in accelerometer output between the hip and the back during walking, two separate ANCOVAs were performed for each placement using data from speeds 4 and 6 km·h[-1]. Various anthropometric measurements were offered into the analysis in an attempt to explain the differences between placement during walking (n = 96).

At the two walking speeds the supraspinale ($p < 0.01$), abdominal ($p < 0.01$) and iliac crest ($p < 0.01$) skinfold thicknesses were all significant predictors of PAC at the hip. Body mass index, waist to hip ratio and leg length were also significant predictors (p<0.01). None of these variables was a significant predictor of PAC at the lower back. Of all the variables, supraspinale explained the most variance in PAC additional to speed and was therefore retained in the regression equations.

Linear regression analysis between PAC (counts·5s^{-1}) and $\dot{V}O_2$ (ml·kg^{-1}·min^{-1}) for all speeds (n= 384) revealed that PAC alone explained 91.8 % of variance in $\dot{V}O_2$ (r^2 = 0.92) with a standard error of the estimate (SEE) = 3.80 ml·kg^{-1}·min^{-1} at the hip and 90.1 % of variance at the back with SEE = 4.18 ml·kg^{-1}·min^{-1}. When the hip data were split by exercise mode, the supraspinale skinfold thickness explained the most additional variance in $\dot{V}O_2$ other than PAC. Regression equations are shown in Table 7.2.

Source	Placement	Linear Regression equation	r^2	SEE
PAC	Hip	$\dot{V}O_{2\text{pred}} = (0.061*PAC) + 6.267$	0.92	3.80
PAC	Back	$\dot{V}O_{2\text{pred}} = (0.056*PAC) + 8.960$	0.90	4.18
PAC (walking data only)	Hip	$\dot{V}O_{2\text{pred}} = (0.052*PAC) + 8.205$	0.60	2.76
PAC (walking data only)	Back	$\dot{V}O_{2\text{pred}} = (0.059*PAC) + 8.589$	0.66	2.53
PAC (running data only)	Hip	$\dot{V}O_{2\text{pred}} = (0.065*PAC) + 3.889$	0.51	4.64
PAC (running data only)	Back	$\dot{V}O_{2\text{pred}} = (0.059*PAC) + 7.418$	0.34	5.38
PAC (walking data only), supraspinale	Hip	$\dot{V}O_{2\text{pred}} = (0.057*PAC) - (0.339*\text{supraspinale}) + 12.137$	0.75	2.18
PAC (running data only), supraspinale	Hip	$\dot{V}O_{2\text{pred}} = (0.067*PAC) - (0.425*\text{supraspinale}) + 8.669$	0.61	4.13

Table 7.2 Regression equations for the prediction of $\dot{V}O_2$

7.5   DISCUSSION

The purposes of this study were three-fold. Firstly, the aim was to establish any influence of placement on accelerometer output comparing hip and lower back placements. The second purpose was to determine whether intra-instrument reliability and instrument validity for the prediction of $\dot{V}O_2$ was influenced by placement. The final objective was to determine whether adiposity at the accelerometer attachment site, measured by skinfold thickness, explained any variation in accelerometer output or instrument validity. It was hypothesised that monitors placed at the hip would produce higher and more random PAC than those at the lower back and that any increases would be related to hip adiposity.

The 3DNX accelerometer demonstrated good intra-unit reliability indicated by a similar $\dot{V}O_2$-PAC relationship between tests and coefficients of variation below 3.0 % for both placements. However, physical activity counts were significantly higher at the hip, with

the differences identified at walking velocities of 4 and 6 km·h^{-1}. Interestingly, no differences were observed during running. Adiposity at the hip, determined by a range of variables, was related to accelerometer output at the hip but not at the back. Indeed, placement at the hip marginally increased the random error associated with predicting $\dot{V}O_2$ from PAC. Linear regressions for all data and split by exercise mode revealed good validity for the prediction of $\dot{V}O_2$ from monitors at both placements.

7.5.1   Intra instrument reliability

Intra-unit reliability is important for longitudinal studies assessing PA levels of a population over time. It is essential that any observed variance in PA or EE is due to changes in participant behaviour rather than random error associated with the device (Welk, 2005).

The 3DNX monitors demonstrated good intra-unit reliability indicated by the lack of an effect of trial on the relationship between PAC and $\dot{V}O_2$. This is in agreement with the results of study 3 assessing the intra-unit reliability of the 3DNX^{TM} over speeds ranging from 4-20 km·h^{-1} during treadmill exercise. The CV_{intra} in the present study ranged from 0.0-12.4 % at the hip and 0.0-17.4 % at the back where the mean values were 3.0 % and 2.8 %, respectively. These values are slightly lower than those reported in study 3 which found a range of 7.7–16.2 % and a mean of 10.8 % when the 3DNX was positioned at the lower back. The reduced variability in the current study is likely due to fixation of the device directly to the skin.

7.5.2   Instrument validity

The regression slopes for PAC and treadmill speed were not homogenous between placements suggesting that, in a laboratory setting, hip and lower back placements cannot be used interchangeably within studies. These findings are in accordance with those by Yngve et al. (2003) who concluded that the placement of a uniaxial accelerometer (hip versus lower back) had a significant effect on the classification of different exercise intensities. The use of a triaxial accelerometer may exacerbate the differences observed by Yngve et al. (2003) as the additional sensing axes could register rotation and horizontal displacement that uniaxial devices could not detect.

In contrast to the findings of the present study, Yngve et al. (2003) found that output from the hip was lower during walking and running compared to the lower back. Washburn & Laporte (1988) also compared hip and lower back placements and found no effect on PAC from a uniaxial monitor during normal and fast paced walking. Both of these studies used uniaxial devices. The equivocal findings between uni and triaxial monitors could be explained by the relative contributions from the horizontal acceleration component during different exercise modes which are not detected by uniaxial accelerometers; it could be that horizontal accelerations of the hip are greater than the lower back during walking (Kavanagh *et al.*, 2008).

Various indices of trunk adiposity were independently associated with PAC at the hip but not the lower back during walking. Bouten et al. (1997) suggested that output from body-mounted accelerometers could be affected by the motion of soft tissue under the sensor. In the present study, external vibrations and oscillations from loose attachment were mitigated by direct attachment to the skin. With increasing soft tissue at the monitor attachment site, there may have been increased movement compared to the centre of mass (Light*, et al.*, 1980). Measurements other than hip adiposity were independently related to PAC at the hip (BMI, leg length and waist:hip ratio) however, due to the co-linearity between many of these physiological variables, only the variable explaining most additional variance was included in regression analysis (supraspinale skinfold thickness).

7.5.3   Prediction of $\dot{V}O_2$

3DNX counts increased linearly with $\dot{V}O_2$ during walking and running. Regression analysis revealed strong, positive relationships between PAC and $\dot{V}O_2$ at all speeds for both hip and back placements. When all speeds were combined, r^2 and SEE values at the hip (r^2 = 0.92, SEE = 3.8 ml·kg^{-1}·min^{-1}) and back (r^2 = 0.90, SEE = 4.2 ml·kg^{-1}·min^{-1}) were similar. These values are similar to the r^2 values reported by Welk et al. (2004) who compared two different devices, the Biotrainer (r^2 = 0.88, SEE = 1.47 kcal·min^{-1}) and Actical (r^2 = 0.91, SEE = 1.24 kcal·min^{-1}). The marginally higher r^2 values in the present study are likely due to the wider range of speeds studied, naturally increasing the correlation coefficient value ((Hopkins, 2000)). Fudge et al. (2007) validated the 3DNX v2 during treadmill exercise. These authors reported r^2 and SEE values of 0.83 and 2.20 ml·kg^{-1}·min^{-1} and 0.76 and 4.88 ml·kg^{-1}·min^{-1} for walking and running, respectively. These

data were derived from treadmill velocities ranging from 3 – 20 km·h[-1], which could explain the increased linearity and similar SEE values. In Chapter 6, the 3DNX v3 was validated during treadmill exercise and returned r^2 and SEE values similar to those of the present study for walking (r^2 = 0.65, SEE = 1.4 ml·kg[-1]·min[-1]). Values for running were more favourable in the present study than the previous chapter (r^2 = 0.62, SEE = 3.6 ml·kg[-1]·min[-1]); however, in Chapter 6, a greater range of running velocities were included which would contribute to producing higher r^2 values.

It was necessary in the current study to split the data by mode of exercise for sub-analyses because of an interaction effect between placement and PAC during walking. When walking data only were further separated by placement, the $\dot{V}O_2$-PAC relationship using hip data only was r^2 = 0.60 (SEE = 2.76 ml·min[-1]·kg[-1]) and r^2 = 0.66 (SEE = 2.53 ml·min[-1]·kg[-1]) using back data only.   This regression analysis revealed that during walking, PAC at the back displayed a slightly stronger linear relationship with $\dot{V}O_2$ and lower SEE compared to the hip. This infers that, during walking, the lower back placement may be more suitable due to reduced random error. Walking is one of the most common movements performed as part of habitual PA (Wang *et al.*, 2010) consequently, it is often prescribed as part of rehabilitation and disease prevention interventions (Hakim *et al.*, 1998). It would therefore be beneficial to select the least variable accelerometer attachment site so that reliable estimates of physical activity and EE could be made. The opposite was found during running; the r^2 value at the back was the lowest observed in all regression analyses (r^2 = 0.34) and the SEE was the highest observed at 5.3 ml·kg[-1]·min[-1]. This implies that different attachment sites would be preferable depending on the activity selected. The participants involved in the current study were all recreationally active, young and with a narrow range of body size and trunk adiposity. For this reason, caution must be taken before applying these findings to a wider sample of females, and of course, to males.

When indices of adiposity were added to the regression analysis for walking, r^2 and SEE values were improved at the hip. Previous laboratory studies found the inclusion of individual characteristics improved the prediction of EE (Welk*, et al.*, 2004) but to the authors' knowledge, there are no published articles on the effect of trunk adiposity on accelerometer reliability and validity during treadmill walking or running. As such, comparison with the wider literature is impossible.

The regression equations presented in this study could be applied to a similar sample of young active females for the prediction of metabolic rate. It has been suggested that the use of individual calibration equations might improve the accuracy of prediction equations although this is impractical for large samples and applications (Welk, 2005). The finding that trunk adiposity affects PAC suggests that if the monitor were worn at the hip during walking, different calibration equations would be required for populations with different levels of trunk adiposity. This provides support for positioning accelerometer devices at the lower back during walking activities so that group calibration equations could be employed. Further investigation into the relationship between PAC and trunk adiposity is required in more variable populations. Furthermore, males tend to have less subcutaneous fat stores than females (Geer *et al.*, 2009) suggesting the relationship between PAC and body fatness may differ by gender.

In conclusion, the 3DNX is a valid and reliable tool for estimating $\dot{V}O_2$ during treadmill exercise in young active females. During walking, PAC are higher at the hip compared to the lower back and therefore these placements cannot be used interchangeably within studies. Overall, the increase in PAC at the hip, in this population, was largely consistent and therefore did not affect the validity of this monitor to predict $\dot{V}O_2$. Individual differences and variability in hip-derived PAC, at least during walking, appear to be explained by subcutaneous hip adiposity. A more objective measure of fat deposition at the hip obtained from dual x-ray absorptiometry (DEXA) may strengthen such associations. Overall, the lower back appears a more suitable location for wearing the 3DNX if anthropometric data are unavailable or individual calibration impractical. The effect of accelerometer placement on PAC in overweight, obese and mixed gender populations remains to be established.

8   A COMPARISON OF THREE ACCELEROMETER DEVICES FOR

THE PREDICTION OF FREE-LIVING ENERGY EXPENDITURE

8.1   INTRODUCTION

Body-borne accelerometers are now widely used for objectively assessing the frequency, duration and intensity of free-living physical activity (Chen*, et al.*, 2005), the most studied being the uniaxial Actigraph GT1M (Rothney*, et al.*, 2008). However, it is still difficult to accurately use the resulting dimensionless physical activity 'counts' (PAC); the standard output of accelerometers, to predict units of expended energy (Chu*, et al.*, 2007). Despite this difficulty, prediction of physical activity energy expenditure (PAEE) remains the end goal of many device manufacturers and researchers alike.

While it has its limitations, the doubly labelled water (DLW) method is considered the 'gold standard' for determination of free-living total energy expenditure (TEE) in humans (Ainslie*, et al.*, 2003). PAEE can be calculated by subtracting measured or predicted resting metabolic rate (RMR) from TEE (Plasqui*, et al.*, 2007). The use of DLW is often limited to small experimental studies due to resource and financial costs. Furthermore, it only provides a mean estimate of EE over a series of days (usually 7-14), rather than a detailed profile of TEE or PAEE over time. Although DLW has been used to describe differences in TEE over individual days (Myers *et al.*, 2008), the accuracy of these estimations is questionable (Campbell*, et al.*, 2002). Despite these limitations it remains a suitable reference technique for the evaluation of other methods of TEE and/or PAEE estimation (Bouten*, et al.*, 1997).

The most common approach to predict TEE or PAEE has been to develop equations in laboratory settings combining anthropometric variables with accelerometer PAC. However, algorithms and equations are often proprietary, meaning that the interpretation of the returned values may be limited and prediction errors are often high. The prediction error appears to arise from two main sources. Firstly, waist-mounted accelerometers tend to underestimate EE associated with upper body movement, load carriage and changes in surface or terrain (Hendelman*, et al.*, 2000; Trost*, et al.*, 2005). Secondly, the majority of devices employ linear models to predict TEE and PAEE (Lyden*, et al.*, 2010). This is an

over-simplistic approach as the relationship between activity counts and PAEE is known to vary according to the type of activity being performed (Midorikawa*, et al.*, 2007).

In an attempt to better capture the multi-directional nature of free-living activity, devices that measure acceleration in three dimensions, so called triaxial accelerometers, have been developed and many have been validated during treadmill exercise using protocols similar to that used in Study 3. Two such devices are the 3DNX (v3) (BioTel Ltd., Bristol, UK) and the Actigraph GT3X (ActiGraph, Pensacola, USA). In both these devices, post data collection prediction equations can be used convert arbitrary activity 'counts' to units of expended energy (Carter*, et al.*, 2008; Lyden*, et al.*, 2011). Indeed, triaxial accelerometers appear to offer greater prediction accuracy of PAEE than uniaxial accelerometers (de Graauw *et al.*, 2010; Plasqui*, et al.*, 2005; Welk*, et al.*, 2000). To our knowledge, no DLW validation studies have been performed using the GT3X. A recent DLW validation (Chapter 9) reported that 3DNX PAC and height explained 41 % of the variance in PAEE with a standard error of prediction of 32 %. Physical activity counts alone explained 6% of the variance in PAEE. This modest amount of explained variance indicates that accelerometry alone is not sufficient to accurately estimate PAEE even when using the more sophisticated triaxial devices. Using a combination of accelerometry and HR data has been used in an attempt to improve the accuracy of PAEE prediction. The ActiHeart (CamNtech Ltd., Cambridge, UK), uses a branched-equation modelling approach to ensure that the prediction of PAEE is more heavily weighted towards HR during moderate and intense activities and accelerometry during lower intensity activities (Ainslie*, et al.*, 2003).

Given that the physical characteristics of the population being studied plays such an important role when predicting of free living PAEE, a true comparison of the accuracy of available devices should be performed using the same population undertaking the same activity. Therefore, the aim of this study was to take a user-oriented approach to comparing the prediction accuracy of the GT3X, 3DNX and Actiheart against criterion values of PAEE derived from the DLW technique. The hypothesis was that there would be no significant difference between DLW derived PAEE and the HR integration method (i.e. ActiHeart), but that accelerometer-only estimates would be significantly different from DLW PAEE.

8.2 METHODS

8.2.1 General Approach

Ten young adult, Emirati nationals from a military preparation school in the United Arab Emirates volunteered to participate in this study, which was approved by the School of Sport and Exercise Sciences Ethics Committee at the University of Birmingham as part of an ongoing project pre-dating the start of this research programme. Participants provided written informed consent following verbal and written explanation of all procedures.

Stature (Leicester Stadiometer, Seca, Hamburg, Germany) and body mass (Alpha 770, Seca, Hamburg, Germany) were measured on the day prior to the start of the observation period. Following these measurements, participants completed a multistage fitness test to estimate maximum oxygen uptake (VO_2max) and determine maximum HR max (HRmax) (Ramsbottom *et al.*, 1988) .

Environmental temperature and relative humidity were measured using a wet bulb globe thermometer (QUESTemp°32 Thermal Environment Monitor, QuestTechnologies, Wisconsin)

During an observation period of ten days, participants total energy expenditure (TEE) was estimated using DLW (see Section 8.2.2). Each individual also wore a HR monitor (Team [2] system, Polar Electro Oy, Kempele, Finland) and three accelerometer devices (as per manufacturer's instructions); the 3DNX (in the small of the back), the GT3X on the (midaxillary line at the right hip) and the Actiheart fixed to the chest in the upper position (2006). Participants were fitted with the devices each morning on waking and instructed to wear the devices during the entire waking day. The devices were removed by experimenters at "lights out". The typical waking duration of a weekday was approximately 15 hours (05:45–21:30). However, participants regularly removed the devices for approximately two hours during the day while sleeping. Time spent sleeping was recorded by participants, verified by visual inspection of the data and removed from the total waking hours for that day. At the weekend, sleeping patterns varied between individuals but the duration of a waking day remained 15 ± 1.8 hours.

No formal information was gathered regarding daytime sleep patterns over the weekend, resulting in daytime accelerometer removal for sleeping during this period being included

as part of the waking day and classed as non-wear. Non wear time was identified in two ways. Firstly, a period of 20 minutes rolling zero counts was classed as accelerometer non wear. Secondly, due to the small participant numbers, data were visually inspected and non-wear was verified on a case by case basis. Maximum possible waking hours were calculated for the ten days and actual wear time (total waking hours − non wear time), expressed as a percentage of this maximum. All devices were downloaded using dedicated software and imported into Microsoft Excel for subsequent analysis.

8.2.2 Measurement of Energy Expenditure using DLW

The DLW method used has been described in full in Chapter 3. Briefly, each participant provided baseline urine samples before ingesting a weighed oral dose of $^2H_2{}^{18}O$ (day 0). The doses were 80 mg·kg^{-1} deuterium oxide and 145 mg.kg^{-1} $H_2{}^{18}O$. Post-dose urine samples were collected daily for the 10 subsequent days and the time of day the urine sample was produced was recorded by each participant. Urine samples were subsequently frozen at -20 °C and later analysed in duplicate using isotope-ratio mass spectrometry (MRC, Human Nutrition Research, Elsie Widdowson Laboratory, Cambridge, UK). Using this dosing regimen, EE can be estimated with a coefficient of variation lower than 5.0 % (Bluck, 2008). A sample of drinking water was also collected to correct for the natural abundance of $^2H_2{}^{18}O$ at the study location.

8.2.3 Prediction of Energy Expenditure using Accelerometer Devices

3DNX

The triaxial 3DNX (54 x 54 x 18 mm) (v3; BioTel Ltd, Bristol, UK) is described in Chapter 3. It is sensitive to movements in three planes: X (anteroposterior), Y (mediolateral) and Z (vertical). Accelerometers were set to record at one-minute epochs and activity counts were converted to PAEE using a prediction equation described in Chapter 9.

Actigraph

The Actigraph GT3X (38 x 37 x 18 mm) weighs 27 g and measures acceleration within a dynamic range of ± 3 g. The GT3X has a sample frequency of 30 Hz. The accelerometer signal is digitised and the accelerations summed each minute to provide an activity counts

value. Once downloaded the software offers the user three options for converting activity counts to PAEE. The first is the 'Work energy theorem' (GT3X_{WE}):

$$PAEE\ (kcal.min^{-1}) =\ counts * 0.0000191 * mass \qquad \text{(Equation 8.1)}$$

The 'Freedson' option (GT3X_{F}) (Freedson, *et al.*, 1998) is the second option where:

$$PAEE(kcal.min^{-1}) = scale * \big(0.00094 * counts + (0.1346 * mass - 7.37418)\big)$$
$$\text{(Equation 8.2)}$$

$$scale\ =\ \frac{epoch\ period}{60} \qquad \text{(Equation 8.3)}$$

Finally the 'combination' option where the Freedson equation (Freedson, *et al.*, 1998) was used to predict EE above 1951 counts.min[-1] and the work-energy theorem was used below that threshold (GT3X_{C}). The outputs from all three of these options have been included in subsequent analyses.

Actiheart

The Actiheart (CamnTech, Cambridge, UK) is described in chapter 3 and elsewhere (Brage, *et al.*, 2005). In short, the device returns a value for EE for each minute based on a combination of the relationships between HR, accelerometer counts and EE. The group calibration proposed by Brage (2007) was selected for all participants. Individual values for maximum HR were entered manually and sleeping HR was recorded as the lowest value from a 20-minute rolling average. Where HR data from the Actiheart were missing due to incomplete ECG data, values from the Polar HR monitor were integrated into existing files using bespoke Actiheart software (CamnTech, Cambridge, UK) altered by the manufacturer following a request by the authors. This provided an additional, more complete dataset to demonstrate the capability of the Actiheart had it worked according to the manufacturers' specifications (Actiheart_{Polar}).

8.2.4   Data reduction

Given that an entire waking day was approximately 15 hours long, the true amount of physical activity undertaken during that period may not have been fully captured on any day where an accelerometer was worn for less than 15 hours.

In the absence of a consensus regarding best practice for data reduction, the threshold at which a day was labelled complete or incomplete was carefully considered and a

population-specific approach was taken. A histogram was generated similar to that of Colley et al. (2010) to visualise the effect of manipulating the threshold at which a day was deemed viable. The value of 11 hours allowed the inclusion of the largest number of days while exceeding the recommended threshold of 10 hours recommended by Triano et al. (2008) Days when wear time was less than 11 hours (including removal for daytime sleeping), wear time was deemed insufficient to capture the majority of physical activity for that day and coded as "incomplete". No individuals registered more than three "incomplete" days during the 10-day monitoring period. Any days that did not meet the 11 hour threshold were excluded from the 10-day mean for each participant. It was decided that imputing missing days was not acceptable for such a small sample size. Overall wear time compliance based on the maximum possible waking hours (summation of time between waking and "lights out" for the entire observation period) was 85.0 ± 11.3 % before incomplete days were excluded. After excluding incomplete days the wear time compliance for the remaining days was $88.8 \pm 7.8\%$.

8.2.5   Statistical Analysis

Paired t-tests were used to compare between mean PAEE estimated by DLW and each of the accelerometer prediction equations. Limits of agreement (LoA) were calculated to show the level of agreement between methods (Bland *et al.*, 1986). Statistical significance was set at *a priori* <0.05.

8.3   RESULTS

Participant characteristics and estimated PAEE values are shown in Table 8.1. One Actiheart monitor was removed due to skin irritation and one 3DNX unit failed during the observation period, therefore the final dataset was reduced to n=8 for all comparisons. DLW PAEE was greater than that predicted by 3DNX (p=0.001), $GT3X_{WE}$ (p<0.001), $GT3X_F$ (p<0.001) and $GT3X_C$ (p<0.001). However, there were no differences between DLW PAEE and PAEE predicted by Actiheart (p=0.199) or $Actiheart_{Polar}$ (p=0.381).The bias and LoA between the predicted PAEE values from each of the accelerometer devices and those measured by DLW are shown in Figure 8.1

The absolute bias $\pm95\%$ LoA were -860 (p< 0.001) ±318 kcal·d^{-1}, -897 (p< 0.001) ±263 kcal·d^{-1} and -1047 (p< 0.001) ±304 kcal·d^{-1} for the $GT3X_{WE}$, $GT3X_F$ and $GT3X_C$

respectively. The absolute bias $\pm$ 95% LoA were -326 (p< 0.001) $\pm$335 kcal·d^{-1} for the 3DNX, -288 (p>0.05) $\pm$1060 kcal·d^{-1} and -215 (p>0.05) $\pm$1198 kcal·d^{-1} for the Actiheart and ActiheartPolar, respectively.

Environmental temperature and relative humidity during the observation period were 32.6 $\pm$ 2.2 °C and 54 $\pm$ 11% respectively.
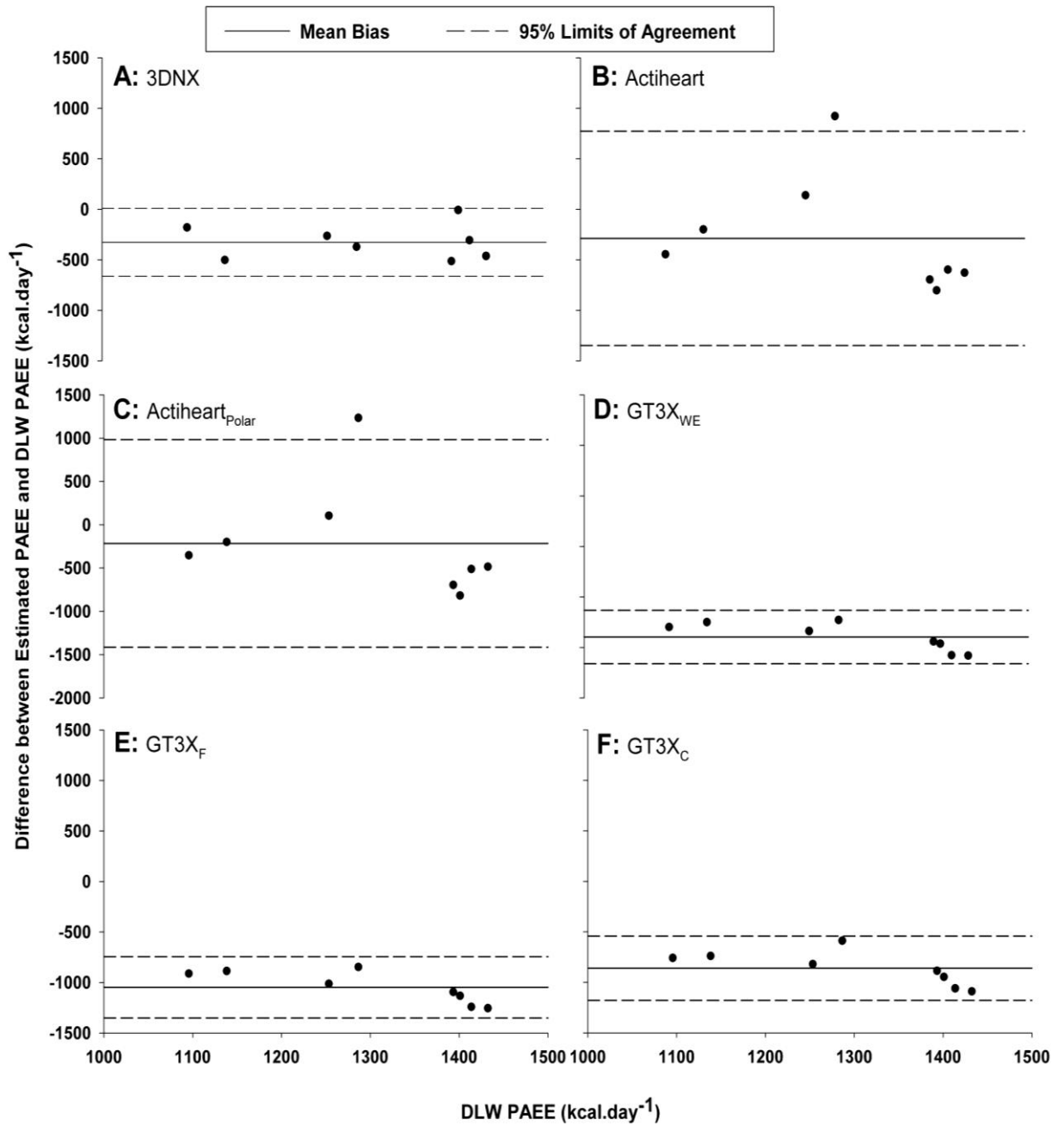
Figure 8.1. The relationships between PAEE estimated by the DLW technique and PAEE predicted by the 3DNX^{TM} (**A**), the Actiheart (**B**), Actiheart$_{Polar}$, (**C**), GT3X$_{WE}$ (**D**), GT3X$_F$ (**E**), and the GT3X$_C$ (**F**)

Participant	Height (cm)	Mass (kg)	Age (Yrs)	RMR (kcal·d^{-1})	Mean PAEE (kcal·d^{-1})						
					DLW	3DNX	Actiheart	Actiheart$_{Polar}$	GT3X$_{WE}$	GT3X$_F$	GT3X$_C$
1	172.3	65.7	16.3	1820	1253	990	1391	1358	415	241	434
2	174.6	99.5	18.1	2186	1287	915	2209	2520	558	439	699
3	177.4	63.6	18.8	1646	1401	1393	599	583	438	268	455
4	176.8	81.6	18.7	1917	1096	916	650	742	298	184	337
5	175.9	57.4	17.8	1690	1432	969	804	947	352	177	343
6	178.9	65.2	18.8	1670	1414	1107	816	902	336	171	354
7	161.5	61.8	18.3	1620	1138	635	939	939	389	252	400
8	171.0	74.3	18.0	1807	1393	879	698	698	452	300	507
Mean	173.6	71.1	18.1	1794	1302	976*	1013	1086	389*	240*	424*
± SD	± 5.5	± 13.7	± 0.8	± 188	± 131	± 215	± 543	± 624	± 86	± 84	± 115

Table 8.1. Participant characteristics and estimated physical activity energy expenditure (PAEE). RMR, predicted resting metabolic rate; Mean PAEE, estimated physical activity energy expenditure over the 10 day observation period; 3DNX, PAEE estimated using the equation generated in Chapter 9; Actiheart, PAEE estimated using the untouched Actiheart data; Actiheart$_{Polar}$, Actiheart PAEE estimation with missing HR values replaced; GT3X$_{WE}$, PAEE estimation using the 'work energy theorem' option; GT3X$_F$, PAEE estimation using the 'Freedson' option; GT3X$_C$, PAEE estimated using the 'combination' option

*symbol denotes that mean values were significantly lower than DLW (p≤0.01).

8.4   DISCUSSION

This study was the first to compare the validity of 3DNX, GT3X and Actiheart accelerometer devices for the prediction of PAEE in the same free living population. The practical user-oriented approach was designed to mimic that of a novice end user meaning that the equations used for the prediction of PAEE were either those embedded in the device software in the case of the GT3X and the Actiheart, or the only available generic prediction equation in the case of the 3DNX. The main finding of this study was that all three devices under-predicted PAEE compared to the DLW technique regardless of prediction equation chosen.

A comparison of the findings of this study to previous research is challenging due to different evaluation criteria (e.g. doubly labelled water, indirect calorimetry etc.) and different durations of wear time. Furthermore, the GT3X; the newest of the Actigraph models and the 3DNX are comparatively new devices, so at present there are no published independent free-living validation studies.

The Actiheart and the subsequent Actiheart$_{Polar}$ derivative were the only devices that returned PAEE values that were not statistically different from DLW. This pattern continued when the absolute bias was calculated with both Actiheart derivatives returning a non-significant mean difference. This may be slightly misleading as the 95% limits of agreement were the largest for this device indicating the greatest random error at an individual level. The GT3X underestimated PAEE to the largest extent, regardless of which prediction equation option was chosen by the user. Overall these findings are similar to those summarised by de Graauw et al. (2010) whereby accelerometry consistently underestimated PAEE with large limits of agreement. These authors also observed that accelerometry underestimated more vigorous activity to a greater degree. The substantially larger underestimation by the GT3X may be in part due to the dynamic acceleration measurement range of the unit. In the present study it is likely that participants were, at times, performing vigorous activities above the 3g and 2.5Hz thresholds that are filtered out by the manufacturers' settings of the GT3X (John *et al.*, 2012). This has previously been reported during incremental treadmill exercise for the GT1M; the predecessor of the GT3X (Fudge*, et al.*, 2007) and the GT3X (Brychta *et al.*, 2010) where ambulation above 9.66 km·hr^{-1} no longer resulted in increased counts.

The overall underestimation of PAEE estimated by accelerometry is unsurprising given that the activities undertaken by the subject population were military in nature and often included resistance exercise such as push ups and sit ups. It is well known that waist-mounted accelerometers cannot reasonably quantify the EE associated with such movements (Trost*, et al.*, 2005). In these scenarios the Actiheart should display clear advantages over the 3DNX and the GT3X as HR data should better reflect the true metabolic cost of such exercise. Despite the improved suitability of the Actiheart to this population, its estimation of PAEE was not as accurate as that reported by Crouter et al. (2008) where the combined HR and activity algorithm provided similar estimates of PAEE at both group and individual level.

The underestimation of PAEE across all devices could also be partly explained by the ~11% of missing data. Wear time compliance after excluding incomplete days was 88.8 ±7.8 % leaving a proportion of the waking day where the devices were not worn. During this time any physical activity performed and its resulting EE would not have been captured by the devices.

The large limits of agreement, and indeed the non significant mean bias, returned by the Actiheart were mainly due to a large overestimation of PAEE for participant 2. The data for participant 2 were scrutinised and no reason could be found to remove the participant from the analysis. It appears that the large overestimation is a result of the participant's relatively high body mass (99.5kg) and body mass index ($32kg{\cdot}m^2$), and low sleeping heart rate (48 bpm). When this parameter was manually altered to correspond to the group mean (71.1kg), the estimated PAEE value reduces to a value similar to the DLW PAEE group mean. It would seem that the group calibration (Brage*, et al.*, 2007) used in the Actiheart prediction model may not be suitable for use in physically active populations which include overweight or obese participants. This observation should be substantiated in a larger population with a wider range of body sizes and PAEE levels. Furthermore, the effect of high seasonal temperatures (32.6 °C) could have caused the dissociation of the HR-$\dot{V}O_2$ relationship. The participants were however; all local and as such were acclimatised to the environment.

There were a number of limitations to the current study which was driven by the desire to evaluate the three devices from a user perspective conducting a field study. The first methodological limitation was the large amount of missing HR data from the Actiheart. It

would not be feasible for the majority of users to integrate their own HR data; it was done in the present study so that the device could be evaluated as if it had worked perfectly. The Actiheart was positioned by a researcher each day according to the user instructions; thus, no more could have been done to improve the quality of the HR data recorded by the device. The imputation of missing HR data for six out of eight participants made minor improvements to the prediction accuracy of the device over 10 days (Figure 8.1 B and C), but further investigation in a larger sample and a greater range of PAEE values would be valuable to find out to what extent missing HR data affects predicted PAEE. In addition, although sleeping HR was recorded over night, it was not repeated on additional nights introducing the potential for error in the Actiheart predictions as heart rate is subject to large random variation across a period of several days (Cipryan *et al.*, 2012).

Unfortunately due to the nature of the assumptions made during the DLW analysis, there are likely to be some inaccuracies introduced to the criterion measure of PAEE. For example, the factor used to correct for the labelled isotopes leaving the body via sweat and water vapour in breath (fractionation) may not be appropriate for the UAE where summer temperatures can be in excess of 50°C. Furthermore, the assumption of a respiratory quotient of 0.85 may not be appropriate for a non-western diet. Despite these limitations, it is unlikely that a large portion of PAEE underestimation is attributable to error in the criterion measure.

It was not possible in the present study to measure RMR which was instead estimated using the equations of Schofield (1985) incorporating age, gender, body mass and height. Such an equation is widely used despite concerns that it may not be appropriate for all populations and individuals. At present there are many different published equations for the prediction of RMR and others have suggested that the Schofield equation can lead to inaccurate predictions particularly in non-western populations (Hofsteenge *et al.*, 2010). Introducing inaccuracies such as predicted RMR into the criterion measure of PAEE is a likely contributor to the poor prediction of PAEE. Despite this limitation, Carter et al. (2008) contend that predicting RMR may be just as appropriate for validation of prediction equations in a field-based setting as accurate measures of RMR are rarely possible. The present study should be repeated on a larger cohort with a wider range of PAEE values and measured RMR to gain a better insight into the performance of the devices for potential use in larger epidemiological studies. In the case of the 3DNX, the prediction equation

used in this study was developed in a military-style setting, albeit in a more active population, so it could be argued that the current validation was not truly independent for that unit.

In summary, the Actiheart appears the most accurate tool for the prediction of free-living PAEE due, in part, to the integration of HR data. However, in practise the quantity of missing data leads to inaccuracies. These inaccuracies could be exacerbated if participants were left to position their own device. The 3DNX and the GT3X both underestimated PAEE, the 3DNX to a much lesser extent. Although both the mean values for 3DNX and GT3X units were statistically different to that of DLW values, the narrower 95% limits of agreement compared to the Actiheart suggest that these units could still be more suitable for intervention studies. Any changes in group mean values would be detected due to the consistent nature of the prediction error both pre and post intervention. In conclusion, we contend that based on our observations, users should be aware of the large limits of agreement associated with the Actiheart caused by the large overestimation of PAEE for one participant. Although Actiheart mean values were not significantly different from DLW PAEE, this was likely due to the large over estimation for one participant. Future users should determine what level of individual and group error is acceptable before its deployment. Despite similar findings showing poor prediction accuracy in other published research (de Graauw*, et al.*, 2010; Ekelund *et al.*, 2001; Maddison*, et al.*, 2009), accelerometers and linear prediction equations continue to be used by the research community partly due to their ease of application (Rothney*, et al.*, 2007). Therefore, comparison studies of this nature provide valuable information for researchers designing studies where more complex processing techniques may not be feasible. Exploring alternative but user-friendly approaches to establishing relationships between actual PAEE and that predicted using accelerometer counts and other variables is likely to be of benefit.

9 DEVELOPMENT OF AN ACCELEROMETER-BASED MULTIVARIATE MODEL TO PREDICT FREE-LIVING ENERGY EXPENDITURE IN A LARGE MILITARY COHORT

9.1 INTRODUCTION

The doubly labelled water (DLW) method is considered the gold standard for determination of free-living total energy expenditure (TEE) in humans (Ainslie*, et al.*, 2003). Physical activity energy expenditure (PAEE) can be calculated by subtracting measured or predicted resting metabolic rate (RMR) from TEE (Plasqui*, et al.*, 2007). The use of DLW is often limited to small experimental studies due to resources and financial costs. Furthermore, it only provides a mean estimate of EE over a series of days (usually 7-14), and no pattern of activity over time can be determined (Campbell*, et al.*, 2002). Despite these limitations it remains a suitable reference technique for the evaluation of other methods of physical activity measurement where estimating EE over an extended period of time is the end goal (Bouten*, et al.*, 1997).

Body-borne accelerometers are now widely used for objectively assessing frequency, duration and intensity of activity (Chen*, et al.*, 2005). It is still difficult however, to accurately use the dimensionless physical activity counts (PAC) to predict units of expended energy (Chu*, et al.*, 2007). Many laboratory-derived prediction equations are developed using a series of scripted activities such as walking, stepping, sweeping and deskwork (Chen *et al.*, 2003). However, such equations may not accurately translate to activities of daily living. An alternative and, arguably, more ecologically valid approach, is to derive a prediction equation from relevant free-living data in a population that is to be observed in future studies. During such a study, physical activity should ideally be accurately determined over a period of time representative of habitual activity (Bonomi*, et al.*, 2010). Past validation studies that have been conducted in a free-living cohort often use accelerometer data from periods substantially shorter than the DLW observation period (Colley*, et al.*, 2010). Trost et al. (2000) recommended that over a 7-day measurement period, 4-5 days of monitoring including a weekend day (i.e. 2-3 missing or incomplete days) returns sufficient reliability to characterise an individual's typical weekly activity profile. By including an incomplete dataset when developing a prediction model, a researcher assumes that the activity patterns observed on the selected days are consistent

for the remainder of the measurement period. This approach ignores potential differences in activity levels between days, making the estimated summary statistics subject to selection bias (Catellier *et al.*, 2005). This is of particular importance when developing prediction models in military cohorts where the waking day has been previously reported as ~17-h on camp and up to 24-h during military exercise when military activities are continued throughout the night (Wilkinson*, et al.*, 2008b). In this scenario an observation period encompassing both days in camp and days on exercise will lead to variable physical activity patterns which would unlikely be characterised by a subsample of 4-5 days of accelerometer data.

Two main types of accelerometer are in widespread use in physical activity research: uniaxial and triaxial. Laboratory studies have shown close agreement between both types of accelerometer for human ambulation (Welk*, et al.*, 2000). However, during free-living activities it has been suggested that triaxial accelerometry offers greater prediction accuracy of EE (Plasqui*, et al.*, 2005; Welk*, et al.*, 2000). Even so, it is widely accepted that waist-mounted accelerometers tend to underestimate EE associated with load carriage, upper body movement and changes in surface or terrain (Hendelman*, et al.*, 2000; Trost*, et al.*, 2005), mainly due to technological limitations in their ability to detect increased energy cost associated with such activities. Nonetheless, it appears that the greater sensitivity of the triaxial accelerometer to movement in different planes leads to better prediction of EE than uniaxial accelerometers in a free-living environment (Campbell*, et al.*, 2002).

The 3DNX (v3) is a small, lightweight triaxial accelerometer which has recently become commercially available. It has been shown to be valid and reliable in a mechanical setting and correlates well with oxygen consumption during treadmill exercise (Fudge*, et al.*, 2007). A recent study demonstrated the capability of the 3DNX (v2) to predict TEE in two free-living cohorts and reported explained variance of up to 35 % using accelerometer counts alone and up to 78 % when accelerometer counts were combined with body composition variables (Carter*, et al.*, 2008). However, the models presented by Carter et al. (2008) were divided into two groups; young adults and adolescents. When considered as independent cohorts, the young adult military-based prediction equations were developed using only twelve data points. Furthermore, a prediction equation developed in a single platoon of trainee guardsmen may not be applicable to other military groups, particularly female personnel.

This study aims to test the hypothesis that PAC are independently associated with TEE and PAEE in a large pooled dataset from a number of independent cohort studies. This will be achieved using a stringent data reduction method to produce a model which is based upon a large, highly compliant sample and developed using DLW as a gold standard criterion measure. The inclusion of a range of cohorts would make the resultant mathematical model more generic, at least within the military context.

To our knowledge, this will be the largest independent evaluation of the efficacy of an accelerometer to predict EE against the DLW method in multiple cohorts.

9.2   METHODS

9.2.1   General Approach

This study has added together a group of studies with common methodologies. A large and diverse group of participants have been combined for the first time using data collected prior to and during this project.

Over a six-year period from 2004 to 2010, ten data collection periods with military cohorts in the UK and UAE produced 288 individual datasets where average daily 3DNX PAC and average daily EE using the DLW technique were measured simultaneously. Data from studies 4, 6 and 7 (Table 1) have been reported in a different format elsewhere (Carter, *et al.*, 2008; Richmond, *et al.*, 2012) and were collected during a period of employment pre-dating this research programme. Only data from study 4 were collected by an independent team of researchers who have given their consent for these data to be included in the present study. Studies 1, 2, 3, 5, 8, 9 and 10 were completed during this research programme.

In all studies, written informed consent was provided and where necessary countersigned by a legal guardian following a full explanation of procedures. Ethical approval for each study was sought and obtained from the most appropriate research ethics committee, including the UK Ministry of Defence Research Ethics Committee and the School of Sport and Exercise Sciences Ethics Committee at the University of Birmingham. Ethical approval was granted prior to the commencement of this project as part of a larger body of work.

Despite small methodological differences between each study i.e. different monitoring period lengths, DLW and accelerometry techniques remained constant throughout.

9.2.2   Preliminary Measures

Stature (Leicester Stadiometer, Seca, Hamburg, Germany) and body mass (Alpha 770, Seca, Hamburg, Germany) were measured on the day prior to the start of each observation period. Body composition was calculated from total body water using the average Deuterium and 18-Oxygen dilution spaces from the DLW measurements. Total body water was corrected by 1.04 and 1.01 respectively to account for non-aqueous exchange of the two isotopes within the body. Fat free mass was calculated using a hydration factor of 73 % with fat mass as the difference between body mass and fat free mass.

9.2.3   Isotope dosing and sampling

The DLW method is described fully in chapter 3. Briefly, each participant provided baseline urine samples before ingesting a weighed oral dose of $^2H_2^{18}O$ (day 0). The doses were 80 mg.kg^{-1} deuterium oxide and 145 mg.kg^{-1} $H_2^{18}O$. Post-dose urine samples were collected daily for 7-10 days depending on the study (see below) and the time of day noted. Urine samples were subsequently frozen at -20 °C and later analysed in duplicate using isotope-ratio mass spectrometry (MRC, Human Nutrition Research, Elsie Widdowson Laboratory, Cambridge, UK). Using this dosing regime, TEE can be measured with a coefficient of variation lower than 5 % (Bluck, 2008). A sample of drinking water was also collected to correct for the natural abundance of $^2H_2^{18}O$ at each study location.

9.2.4   Accelerometry

The triaxial 3DNX model v3 (BioTel Ltd, Bristol, UK) is described fully in chapter 3.

In studies 5, 6, 7 and 8 an earlier version of the 3DNX accelerometer (v2) was used. For a full description of the 3DNX v2 see Carter et al. (2008). Slight differences in the unit software meant that a conversion factor was applied in order to standardise the activity counts between unit versions. This conversion factor was developed by in Chapter 4 by simultaneously shaking both versions on a multi-axis shaker table (MAST-9720, Instron Structural Testing Systems Ltd., UK) at a range of accelerations and producing an ordinary least product linear regression equation (Ludbrook, 1997). The relationship between v3 and v2 units is given by y = 0.224x − 5.703, where y is v2 counts and x is v3 counts. This

relationship was confirmed during unpublished observations in free-living human data where both versions of the accelerometer were worn simultaneously and showed close agreement once converted ($r^2 = 0.99$ SE = 3.9 %).
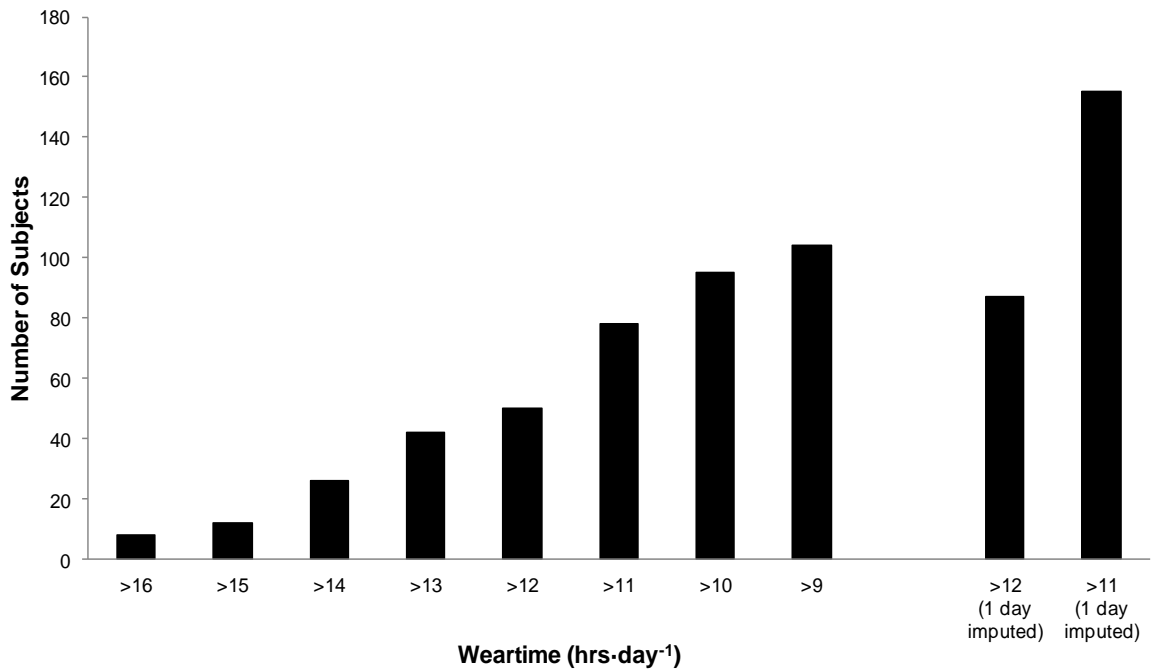
Figure 9.1 Histogram showing the effect of changing the threshold for a 'viable' day. The histogram also shows the additional viable datasets that were included when one "missing" day was imputed using a population mean for that day.

Participants were given the accelerometers on waking and instructed to wear the unit in the small of the back (attached to an elastic belt) during the whole of the waking day. Participants were asked to remove the device during water-based activities which amounted to two hours swimming during the 7-day observation period for studies 1 and 2 and exclusion of 1 entire day for study 8 due to a scheduled river crossing. A typical waking day was approximately 16 hours long (unless during military exercise where it could be up to 20 hours) and collection periods ranged from 7-10 days depending on the study protocol. Maximum possible waking hours were calculated for each study and actual wear time expressed as a percentage of the maximum. Data were downloaded on collection

of the unit using dedicated software (BioTel, UK) and imported into Microsoft Excel for further analysis.

Actual wear time and daily physical activity counts for each individual were calculated and the dataset was reduced in an approach similar to Chapter 8. A histogram was generated (Figure 9.1) and a period of 11 hours monitoring was chosen as a threshold that included as many participants as possible whilst capturing most of the waking day. Any days where the accelerometer was worn for less than 11 hours (including removal for water activities) were coded as "missing". An individual was excluded from the analysis if they had more than one "missing" day. Overall wear time compliance based on each cohort's maximum possible waking hours before imputation was 93.5 ± 5.6 % (i.e. participants wore the activity monitors for an average of 93.5 % of the time between the distribution of the monitors in the morning and their collection at night). For those individuals with no more than one "missing" day, a group mean value was imputed. Imputation is encouraged when incomplete accelerometer data are observed (Catellier, et al., 2005).The imputation method for this study was chosen due to the military nature of the subject population. Minimal between-participant variation was observed in the type and pattern of activities performed on a daily basis as groups often performed activities as a platoon. The inconsistent nature of physical activity patterns between days could have led to an artificial increase or decrease in the mean PAC had a day simply been excluded. In total, PAC for 30 individual days were imputed out of 1349 days measured (2.2%).

9.2.5   Statistical analysis

As EE, along with other physiological variables, is known to increase with body size (Nevill *et al.*, 2005), a multiplicative allometric modelling approach was used as detailed by Nevill and Holder (1995b). Where Y is the dependent variable, a is the intercept, X^b is the independent variable and $\in$ is the error associated with the model.

$$Y = a \cdot X^b \cdot \in \qquad \text{(Equation. 9.1)}$$

Linear regression assumes the error term associated with the model is constant. When the residual errors associated with the model are proportional to the size of the dependent variable, which is often the case with physiological variables (Nevill, *et al.*, 1995b), the data must be transformed. Both the dependent and independent variables were log-transformed to overcome the heteroscedasticity which was observed when plotting size-

related independent variables such as body mass and height against TEE and PAEE. Analysis of covariance (ANCOVA) was then used to identify any categorical differences (e.g. sex differences) by defining $\log_e$(TEE) and $\log_e$(PAEE) as dependent variables and using $\log_e$(PAC), $\log_e$(Mass) and $\log_e$(Height) as covariates. The log-linear model used to investigate the relationship between EE (*Z*) and PAC (*Y*), mass (*X*) and height (*W*) is given by

$$Z = a \cdot Y^{b}{}_1 \cdot X^{b}{}_2 \cdot W^{b}{}_3 \cdot \in \qquad\qquad \text{(Equation. 9.2)}$$

The coefficient of determination (r^2) was used to describe the association between PAC and dependent variables. The partial r^2 value is reported to indicate how much variance is explained by the PAC variable alone above that which is explained by the combination of other variables in the model. The standard error of the estimate (SEE) is reported for all models. All measured variables are reported as mean ± standard deviation (SD). All statistical analysis was conducted using SPSS (version 16.0, SPSS inc., Chicago, IL). The significance level was set *a priori* at $p < 0.05$.

9.3   RESULTS

Following application of the exclusion process, the final dataset was reduced to n = 155. Participant characteristics are presented by study in Table 9.1. Physical activity counts alone, expressed as $\log_e$(PAC), were associated with $\log_e$TEE ($r^2 = 0.08$, $P < 0.01$) and $\log_e$PAEE ($r^2 = 0.07$, $P < 0.01$).

General linear modelling using multiple log-linear regression revealed a difference ($p < 0.01$) between the male and female models for height *a* and *b* parameters. Therefore, two separate models were developed for the prediction of TEE and PAEE, allowing the intercepts and height exponents to vary with gender. In all models PAC, height and body mass were related to EE ($p < 0.01$).

Study	Study duration (days)	Study Description	N (M/F)	Age (years)	Body Mass (kg)	Height (m)	RMR (kcal·day^{-1})	TEE (kcal·day^{-1})	PAEE (kcal ·day^{-1})	PAC - no imputation (counts·day^{-1})	PAC (counts·day^{-1})
1	7	GCC officer cadets	29 (29/0)	19.2 ± 1.9	70.0 ± 14.4	1.69 ± 0.06	1769 ± 239	3270 ± 489	1174 ± 340	645002 ± 65940	657379 ± 66437
2	7	GCC officer cadets	15 (15/0)	20.5 ± 1.9	67.2 ±7.7	1.72 ± 0.03	1701 ± 116	3278 ± 476	1249 ± 356	572908 ± 51000	587763 ± 64524
3	7	GCC officer cadets	6 (6/0)	21.1 ± 1.8	68.3 ± 9.0	1.71 ± 0.06	1716 ± 136	3010 ± 442	993 ± 324	460208 ± 74993	510268 ± 80513
4	10	Advanced field training	6 (6/0)	29.5 ± 4.1	72.2 ± 2.7	1.74 ± 0.06	1748 ± 30	4946 ± 224	2704 ± 216	645789 ± 56385	667570 ± 58585
5	10	GCC military college	6 (6/0)	16.0 ± 1.3	67.3 ± 21.7	1.66 ± 0.02	1806 ± 293	3233 ± 451	1104 ± 220	477007 ± 102580	476570 ± 111492
6	10	Recruits in training	21 (8/13)	19.4 ± 2.4	64.0 ± 10.8	1.71 ± 0.10	1571 ± 249	3426 ± 760	1512 ± 496	604635 ± 109088	578904 ± 109242
7	10	Recruits in training	9 (4/5)	18.7 ± 1.9	64.3 ± 10.1	1.70 ± 0.10	1601 ± 249	3612 ± 428	1650 ± 245	592349 ± 106714	610079 ± 107619
8	10	Infantry training	10 (10/0)	19 ± 2.0	71.0 ± 9.2	1.78 ± 0.06	1806 ± 148	4502 ± 378	2426 ± 251	657190 ± 87025	669529 ± 61372
9	10	Recruits in training	27 (14/13)	21.8 ± 4.4	67.9 ± 12.4	1.71 ± 0.10	1621 ± 285	3580 ± 704	1600 ± 387	536976 ± 59464	536342 ± 55144
10	10	Recruits in training	20 (10/10)	22.6 ± 4.4	68.5 ± 12.9	1.73 ± 0.10	1639 ± 291	3625 ± 771	1623 ± 470	522238 ± 78769	536040 ± 62304
Overall			**155 (111/41)**	**20.6 ± 3.9**	**67.9 ± 12.0**	**1.71 ± 0.10**	**1679 ± 246**	**3553 ± 725**	**1531 ± 546**	**531107 ± 125880**	**586388 ± 93457**

Table 9.1 Physical characteristics of participants (±SD). GCC, Gulf Cooperation Council; RMR, predicted resting metabolic rate; TEE, total energy expenditure; PAEE, physical activity energy expenditure; PAC, 3DNX output

9.3.1 Prediction of TEE

The power function model relating TEE to PAC, mass and height for males is given by

$$TEE(kcal) = 1.67 \cdot PAC^{0.257} \cdot Mass^{0.434} \cdot Height^{2.322}$$ (Equation. 9.3)

The power function model relating TEE to PAC, mass and height for females is given by

$$TEE(kcal) = 2.666 \cdot PAC^{0.257} \cdot Mass^{0.434} \cdot Height^{0.364}$$ (Equation. 9.4)

These models displayed a strong positive relationship with TEE (r^2 = 0.65, SEE = 462 kcal.day^{-1} (13.0 %)). Physical activity counts contributed significantly to the prediction of TEE explaining 4 % of the total variance.
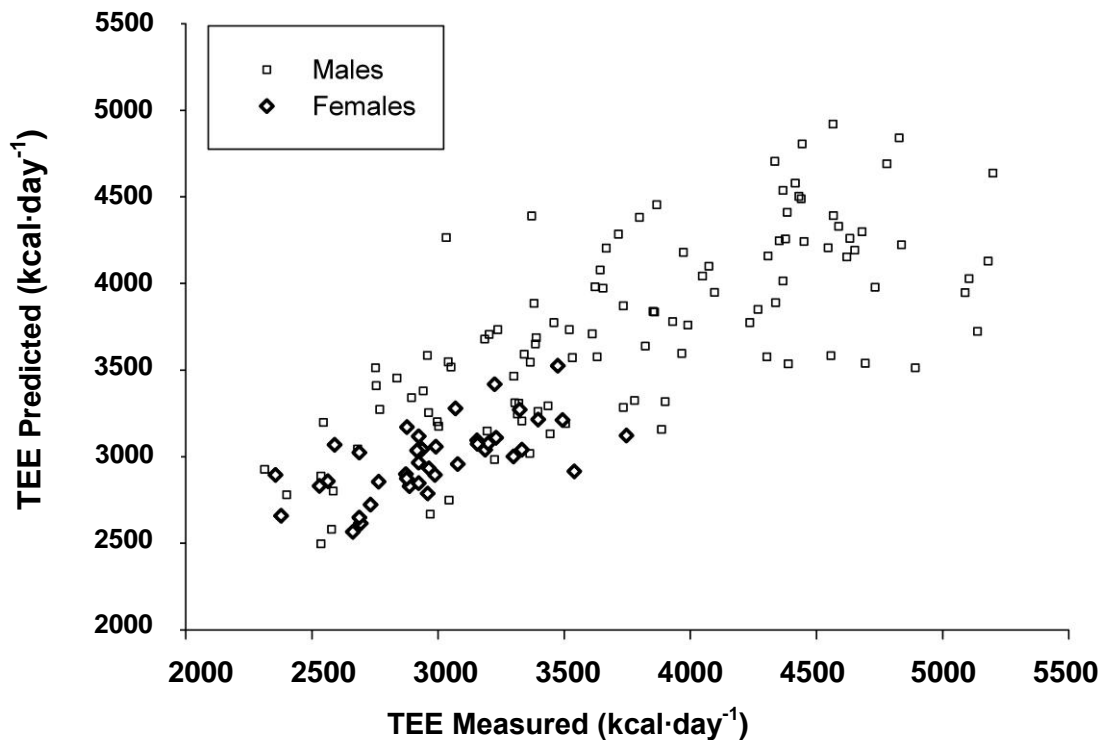
Figure 9.2 Regression plot of the prediction models of total energy expenditure (TEE) based on 3DNX output (PAC), body mass, height and gender Regression plot of the prediction models of TEE based on 3DNX output (PAC), body mass, height and gender

9.3.2   Prediction of PAEE

Body mass did not make a significant contribution to the prediction of PAEE and was removed from the model, hence the power function model relating PAEE to PAC and height for males is given by

$$PAEE = -3.755 \cdot PAC^{0.590} \cdot Height^{5.856}$$ (Equation 9.5)

The power function model relating PAEE (Z) to PAC (Y) and height (X) for females is given by

$$PAEE(kcal) = -1.227 \cdot PAC^{0.590} \cdot Height^{1.217}$$ (Equation. 9.6)

These models showed a moderate positive relationship with PAEE (r^2 = 0.41, SE = 490 kcal.day^{-1} (32.0 %)). Physical activity counts contributed significantly to the prediction of PAEE explaining 6 % of the total variance.
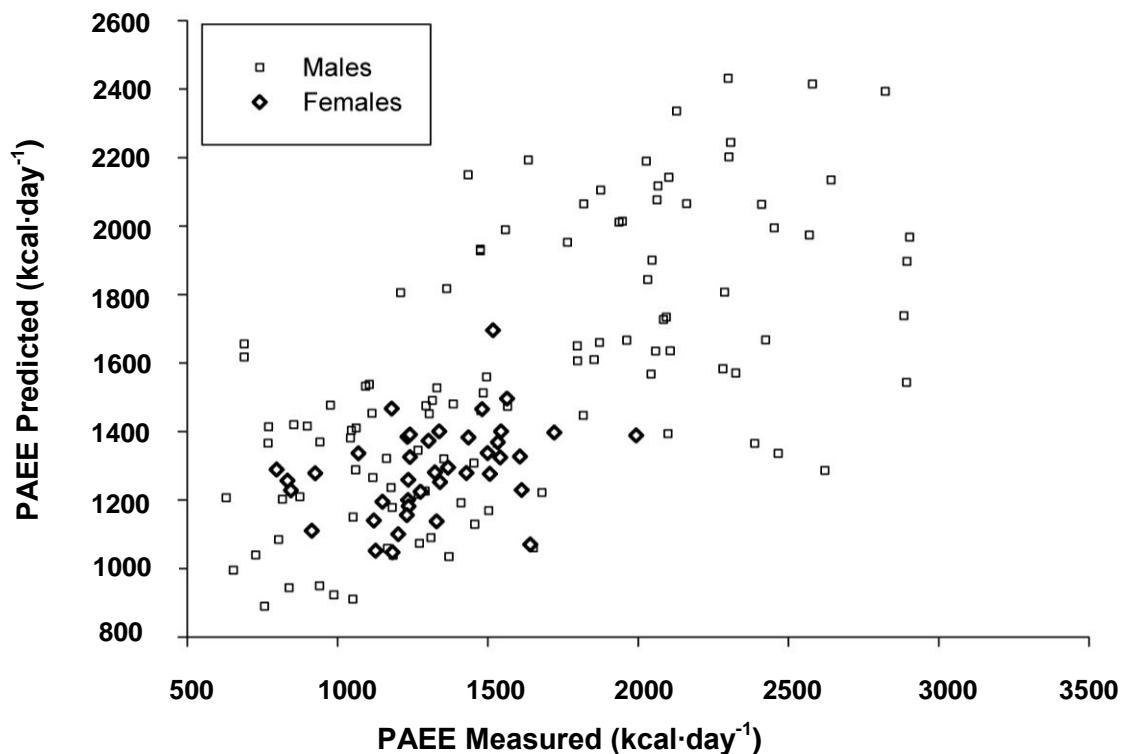
Figure 9.3 Regression plot of the prediction models of physical activity energy expenditure (PAEE) based on 3DNX output, height and gender

9.4   DISCUSSION

Triaxial accelerometry has emerged as a useful tool for assessing physical activity and its associated EE in large scale epidemiological studies (Vanhelst *et al.*, 2012). In order to translate physical activity counts into more meaningful units of expended energy, statistical models must be developed using an appropriate criterion measure such as DLW. In the present analysis, ten independent cohort studies, using the same methodological approach, were combined with the aim of developing a generic model for the prediction of both TEE and PAEE in free-living military populations. Allometric modelling was used to provide a biologically plausible method of fitting the data. The relatively poor fit of the models in this study suggest that it may not be appropriate to use a generic prediction equation for different military populations and that careful consideration must therefore be given to generalising such models across other diverse sub-populations.

The prediction models developed showed that PAC measured by the 3DNX accelerometer significantly contributed to the explained variance in TEE and PAEE for both genders. As height was the only component that was different for males and females gender could not be used as a covariate in a single model. PAC accounted for between 4 % and 6 % of the variance in TEE and PAEE respectively for both male and female models. These small but significant contributions of PAC to the prediction of TEE and PAEE meant that greater accuracy could be achieved when including PAC as a variable in each model. All models were improved by the addition of fat free mass as a covariate in exploratory analyses. However, as the majority of researchers would not be able to measure fat free mass via the DLW technique, simple anthropometric measurements were used throughout. Height is a known surrogate estimate of lean body mass (Nevill *et al.*, 1995a) and contributes more to the prediction of TEE in males than in females. This is unsurprising given that predicted RMR, which is largely dependent on body size (Ainslie*, et al.*, 2003), is 23 % lower for females than males.

It is difficult to compare the individual contribution of accelerometry to the estimation of EE between other studies in the extant literature as few report these data. Most data are reported in combination with anothropometric variables. The results of the present study compare favourably to those of Mâsse et al. (2004) where the uniaxial accelerometer's counts·day^{-1} explained 2 % of the variance in TEE when combined with body mass. The early work of Bouten et al. (1996) reported that the output from the triaxial Tracmor

explained 6 % of the variance in TEE measured by DLW with a standard error of the estimate of 18 %. The results from Bouten's (1996) work are largely similar to the present study with the exception of a slightly lower standard error demonstrated here. The 3DNX returns similar results to the triaxial RT3. Maddison et al. (2009) report an additional 4 % explained variance in TEE measured by DLW when RT3 PAC were included in a model containing gender, body fat (kg) and RMR. More recently, Bonomi et al. (2010) compared TEE measured by DLW and PAC measured by the Tracmor and reported explained variance of 16 % (SE = 7.4 %) when combined with body mass. Furthermore, Carter et al. (2008) reported explained variance of 27 % (SE = 7.0%) when 3DNX (v2) combined with height was compared to TEE measured by DLW in young adults. These two studies report a higher degree of predictive validity, with lower prediction error, compared to the present study. Although it is not readily apparent why the present study reports less favourable results for the predictive validity of 3DNX PAC, differences in study populations and the very high EEs (up to 5199 kcal.day^{-1}) reported for some subjects may have been accumulated during activities which the accelerometer failed to quantify sufficiently. This is likely comprised of the frequent load carriage and uneven terrain encountered as part of military exercise (Knapik *et al.*, 1996; Pandolf *et al.*, 1977). When combined with RMR, Tracmor PAC explained 76 % of the variance in TEE and when combined with body composition variables the 3DNX (v2) explained 78 % of the variance in TEE (Bonomi*, et al.*, 2010; Carter*, et al.*, 2008). In the present study, 3DNX (v3) output explained 65 % of the variance in TEE when combined with body composition and study variables with an SE of 13.0 %.

In the model to predict PAEE, PAC accounted for 6 % of the variance in PAEE (SEE = 32 %). In comparison to the results of Bouten (1996) and Plasqui et al. (2005) the 3DNX appears to poorly predict PAEE. Bouten (1996) reported 22 % of the variance in PAEE could be explained by Tracmor PAC alone (SEE = 12 %) and more recently Plasqui et al. (2005) reported an increase in the explained variance of PAEE measured by DLW from 48 to 81 % (SEE = 17 %) when Tracmor PAC was added to a model containing body composition variables. The poorer prediction of PAEE than TEE may be in part due to the military populations studied. As well as frequent load carriage and ambulation on uneven terrain, the majority of military training is conducted in groups (i.e. Platoons), where recruits largely complete the same physical activity, particularly during transits and drill. It is unsurprising therefore that mean PAC·day^{-1} does not accurately predict PAEE when

there is such little variation in PAC between subjects in a cohort. For example, when the standard deviation of PAC for each group is expressed as a percentage of the mean, the variation ranges between 8.8 and 23.4 % (overall = 15.9 %). In comparison to other studies using the DLW technique this overall variation in PAC is markedly lower than that reported by Bonomi et al. (2010) (28.6 %), Plasqui et al. (2005) (29.1 %) and Maddison et al. (2009) (42.0 %).

The reduction of variance in TEE and PAEE explained by 3DNX (v3) compared to the 3DNX (v2) in the recent Carter et al. (2008) paper is not easily accounted for but is most likely due to the combination of a number of different populations. It seems that there may have been different sources of unexplained variance for each population leading to a cumulative reduction in the common explained variance. This indicates that a generic equation for predicting EE in different free-living populations may not be appropriate particularly when researchers seek to combine cohorts with different physiological characteristics and physical activity patterns. This is particularly pertinent when the prediction of EE available to the user is derived from a manufacturer's proprietary equation. Indeed, some accelerometer models such as the RT3 do not publish details of their proprietary prediction equation (Maddison, *et al.*, 2009), and consequently limits the user's knowledge as to whether the instrument is suitable for predicting EE within their chosen population and/or physical activities. We contend that some form of calibration should be undertaken, any proprietary equations should be refined and specific prediction equations should be developed for each free-living population studied.

9.4.1 Limitations

The data reduction approach that was taken in this study was driven by the desire to retain as many participants while maintaining a stringent approach to wear time compliance. There is no consensus in the literature regarding a best practice method for deciding a threshold for viable datasets and as such there may have been a better way to conduct this process. We used an absolute value of 11 hours as a threshold for classifying a day of wear as valid. On reflection it may have been better to use a percentage value due to the variation in the waking day between populations. For example, soldiers on exercise could be active for up to 18 out of 24 hours in which case an individual registering 11 hours for this day would be retained when a large proportion of physical activity could be missed. Equally, 11 hours could be close to 100% of waking hours for a weekend day in the UAE.

The imputation of one "missing" day i.e. less than 11 hours wear time using a population mean for that day enabled the retention of almost twice the number of participants. This imputation procedure did not markedly alter the mean PAC values for each population (shown in Table 9.1).

Unfortunately due to the nature of the assumptions made during the DLW analysis, there are likely to be some inaccuracies introduced to the criterion measure of TEE and PAEE. For example, the factor used to correct for the labelled isotopes leaving the body via sweating and water vapour in breath (fractionation) may not be appropriate for the UAE where summer temperatures can be in excess of 50°C. Furthermore, the assumption of a respiratory quotient of 0.85 may not be appropriate for a non-western diet. Despite these limitations the DLW technique is still an appropriate criterion as it enables participants to be truly free living; negating a source of error inherent in other available measures such as portable indirect calorimeters.

The field-based setting of the studies combined in this analysis meant it was not feasible to directly measure RMR. As a result, RMR was estimated using the equations of Schofield (1985) incorporating age, gender, body mass and height. Such an equation is widely used despite concerns that it may not be appropriate for all populations however; the population used to develop the equation contained a number of conscripts making it more relevant to this study population than other equations. At present there are many different published equations for the prediction of RMR and according to a recent study by Hofsteenge et al. (2010) the Schofield equation can lead to inaccurate predictions particularly in non-western populations. In addition, the Schofield equation does not include a measure of fat free mass which may lead to an under prediction of RMR in athletic individuals such as military personnel (De Lorenzo*, et al.*, 1999; Thompson *et al.*, 1996). Introducing inaccuracies such as predicted RMR into the criterion measure of PAEE a likely cause of reduced explained variance compared to other validation studies where RMR has been measured. Despite this limitation, Carter et al. (2008) contend that it may be just as appropriate to validate prediction equations in a field-based setting as accurate measures of RMR are rarely possible.

9.4.2   Summary and Conclusions

This is the first large-scale study where the DLW technique has been used as a criterion measure to assess the accuracy of an accelerometry-derived multivariate model to predict

free-living TEE and PAEE within a military population. We have used a highly compliant subject sample and provided a transparent and open account of the methodological issues surrounding the pooling of different populations. The results of this study demonstrate that 3DNX (v3) does improve the estimation of free-living EE in sub-groups of military populations when combined with anthropometric variables. However, the accurate prediction of individual free-living EE requires careful consideration of multiple variables such as anthropometric measurements as well as the type of activities undertaken. It is for researchers to decide whether such small portions of explained variance are of sufficient value to warrant the additional effort and expense of instrumenting similar populations with accelerometers like the 3DNX in future.

We contend that in its present form, triaxial accelerometry is not suitable for predicting energy expenditure in military populations. For such populations, where day-to-day soldiering includes large amounts of resistance-style activity such as load carriage; a more accurate estimate of PAEE could be achieved if time spent performing such activities could be identified and appropriate relationships with $\dot{V}O_2$ attributed. The recent shift in accelerometer-based research towards using the rich, raw accelerometer signal to identify activities of daily-living (Heil*, et al.*, 2012) could be a suited to identifying common soldiering activities that traditional accelerometry methodologies cannot, at present, fully detect. The exploitation of raw accelerometry signals for the identification common military tasks such as load carriage warrants further investigation as it appears that only small gains in prediction accuracy can be made using traditional analysis techniques regardless of how 'high quality' a dataset may be.

10   CLASSIFICATION OF LOAD CARRIAGE USING A SINGLE WAIST-MOUNTED ACCELEROMETER: A FEASIBILITY STUDY

10.1 INTRODUCTION

Load carriage (LC) is a common task for military personnel and is regarded as a key component of day-to-day soldiering in both training and operations (Knapik *et al.*, 2004). A review of US military personnel currently serving in Afghanistan noted that infantrymen were, on average, regularly carrying 'fighting loads' of approximately 29kg, representing, on average, 35% of their body weight (Dean, 2004). This is comparable to the LC task that forms the final assessment during British Army Infantry training, which requires recruits to walk 12.8 km in two hours carrying a 25 kg load depending on specific job role (Rayson*, et al.*, 2000).

Interest in quantifying the energy demands of operations has increased in recent years due to extensive deployment of British forces. Energy expenditure (EE) is one component that can contribute to characterising the physical demands of operations. Gathering this information would assist in developing physical training programmes and military field exercises that best match the demands experienced by soldiers in theatre.

Beyond informing military training programmes, the successful quantification of energy expended during training and operations would better inform military feeding strategies to help maintain health and physical performance (Henning *et al.*, 2011). Energy deficits are commonly reported during military training and have been associated with reduced physical performance parameters. Guezennec et al. (1994) observed a 15% reduction in time to exhaustion during cycling in a low-energy-intake group following five days of military exercise. No performance decrements were seen in the high-energy-intake group. More chronically, Nindl et al. (2007) found a 20 % reduction in maximal lift strength following a prolonged (8-week) caloric deficit of approximately 1000 kcal·day^{-1} during US army ranger training. Although the link between prolonged (8-week) energy deficit and reductions in physical performance may be implied, there are no conclusive data showing that performance is maintained in those who remain in energy balance over the same period. If decrements in role-related performance tests following training are related to prolonged energy deficits, then this may translate to similarly impaired performance on operations if feeding strategies in this environment are also insufficient.

Although it has been reported that training and field exercises elicit a high level of energy expenditure (EE), (upwards of 5000 kcal·day^{-1}) (Tharion*, et al.*, 2005), there are little empirical data available documenting the energy demands of operations upon which to form evidence-based feeding strategies.

While the deployment of a research team to measure the energy demands of such operations is possible, on a large scale this is impractical, potentially dangerous and intrusive for the soldiers concerned. Movement sensing devices may offer a practical compromise whereby EE can be predicted unobtrusively in a large number of personnel without the need for deployment of a large civilian research team. Currently however, movement sensing technologies such as accelerometers have shortfalls in their application to military research. This was demonstrated in chapter 9 where accelerometer counts (PAC) explained only 4% of the variance in PAEE in military populations. A likely reason for the small contribution of PAC was that participants were often carrying loads as part of military training so the increased EE was not matched by increased displacement of the centre of mass or indeed PAC. The ability to detect when military personnel are conducting LC activities could capture the additional EE resulting from LC compared to walking unloaded (Knapik*, et al.*, 2004). At present single accelerometer-based devices are largely incapable of detecting this additional EE.

Variation in activity-specific relationships between PAC and EE is not a problem unique to military populations. Attempts have been made in the health-related research domain to account for the variation in daily activities using a number of methods as it is widely recognised that the relationship between EE and PAC varies depending on the type of activity performed (Bonomi *et al.*, 2009b). One example of such an approach is the work of Rumo et al. (2011) who integrated the raw signal features from two triaxial accelerometers (hip and thigh) with HR to identify seven common activities of daily living so that different regression equations could be assigned to each activity.

Wyss and Mader (2010) used a multi-sensor approach (two accelerometers and a HR monitor) in an attempt to recognise specific military activities conducted during training and assign specific regression equations to each. Although this research took important steps forward in the characterisation of military routines, and predicted EE was not different from that measured by indirect calorimetry, the use of multiple sensors increases user burden and may place restrictions on activities performed habitually (Long*, et al.*, 2009).

Combined sensors are a strong theoretical prospect for the quantification of EE associated with military activities; particularly, integrated biosensors which are capable of detecting increased physiological strain without a concomitant increase in centre of mass displacement. The SenseWear Pro armband (BodyMedia, Pittsburgh, PA) integrates accelerometry with a variety of heat-realted sensors (galvanic skin response, heat flux, skin temperature and near body temperature) to arrive at a minute-by-minute kilocalorie value (via a proprietary algorithm). Although this device has been shown to accurately predict EE during low-moderate intensity activity (St-Onge *et al.*, 2007), the SenseWear Pro underestimated energy expenditure at high intensities (Drenowatz *et al.*, 2011). Although little detail is available regarding the product specifications of this device it is likely that high intensity activities register accelerations outside of the frequency and dynamic range of the device, contributing to the underestimation of EE. In its current form, the SenseWear Pro is not suitable for use in military populations.

A further consideration regarding the use of combined sensors in military populations is the feasibility of deploying such a device. The Actiheart is included in this study for comparison however it is not acceptable for military applications due to its location on the chest. Anecdotal evidence from personal correspondence suggests that a device worn under the body armour would create a point of pressure limiting the effectiveness of the protective equipment.

Elsewhere, researchers have begun to recognise the value of the rich raw signal an accelerometer can capture. The increased memory of newer devices means they are capable of storing raw, unfiltered acceleration signals for extended periods (Intille, *et al.*, 2012). In the health-research domain, recognition of activities from the unique accelerometer signal features they generate has been identified as a research priority (Freedson, *et al.*, 2012; Heil, *et al.*, 2012) and has been achieved with 99% accuracy in a laboratory setting (Zhang, *et al.*, 2012).

The kinematic changes that have been associated with LC include a distinctive forward trunk lean (Majumdar, *et al.*, 2010) and an increase in stance phase of stride, double support time and pelvic rotation (Birrell, *et al.*, 2009). These kinematic changes that occur when an individual carries a military load could be reliably and significantly different enough from normal gait to confer with consistently different accelerometer signals when compared to normal ambulation.

The potential of using raw accelerometer signals to recognise military specific activities such as LC has yet to be investigated yet the use of a small unobtrusive device located outside of the body armour is attractive from an applied perspective. Therefore, the primary aim of this study

was to investigate the feasibility of using raw accelerometer signal to build a model with the ability to distinguish loaded (L) from unloaded (U) ambulation in a sample of experienced load carriers. Load carriage for longer durations (i.e. >15 mins), which is typical of military LC tasks (Knapik, *et al.*, 2004) is known to result in a gradual drift in EE over time (Blacker *et al.*, 2011), neuromuscular fatigue (Blacker *et al.*, 2010b) and change in gait (Attwells, 2008). Therefore, the secondary aim of this study was to evaluate the classification accuracy of the model during prolonged LC.

10.2 METHODS

10.2.1 Participants

Following approval from the University of Bath Research Ethics Approval Committee Health (REACH), 12 male participants (mean ± SD: age 23.8 ± 5.2 years; height 1.77 ± 0.07 m; mass 76.8 ± 7.6 kg; $\dot{V}O_{2\max}$ 59.4 ± 4.1 ml·kg⁻¹·min⁻¹) visited the laboratory on two occasions separated by no less than three days. Participants were a combination of military reservists and experienced mountaineers. All participants provided a self-reported account of their years of LC experience.

Prior to the beginning of the study all procedures were verbally explained to participants and a questionnaire was completed to ensure the good general health of the participants; particularly the absence of musculoskeletal injuries associated with LC.

10.2.2 Preliminary Measures

Height was determined using a stadiometer (Holtain Ltd., Crymych, UK) and nude body mass was determined using Seca 770 digital scales (Seca, Hamburg, Germany).

Participant's maximal oxygen uptake ($\dot{V}O_{2\max}$) was estimated from self-reported 2.4 km run time; a regular military fitness test, using the equation:

$$\text{Predicted } \dot{V}O_{2\max} \ (\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}) = \ 3.5 \ + \left(\frac{483}{x}\right) \qquad \text{Equation 10.1}$$

Where x is the time to run 2.4 km expressed as a decimal in minutes (American College of Sports Medicine. *et al.*, 2010).

Resting metabolic rate (RMR) was measured during the first laboratory visit, prior to the L protocol. Participants were asked to abstain from vigorous intensity exercise, caffeine consumption and alcohol ingestion for a period of 24 hours prior to visiting the laboratory and maintain a food diary for the same period. In addition, participants arrived at the laboratory in a fasted state to avoid any dietary influence upon RMR. The measurement of RMR was conducted in accordance with best practice recommendations of Compher et al. (2006). In summary, following a 10-minute rest period, expired air was collected from the participant via the Douglas bag method for five minute periods until a steady state was achieved with less than a 10% variation in calculated RMR between collection periods.

10.2.3 Experimental Protocol

Participants visited the laboratory on two occasions where they walked on a treadmill for two-hours, either unloaded (U) or carrying a 25 kg load (L) separated by no less than 72 hours to allow for adequate recovery(Blacker, *et al.*, 2010b).

During each trial, participants marched on a treadmill (Woodway ELG 70, Steinackerstraße, Munich, Germany) at a constant speed of 6.4 km·h^{-1} (1.78 m·s^{-1}) and 0 % gradient. This speed replicated the minimum constant speed required to pass the British Combat Fitness Test (BCFT); 12.8 km in two hours (Rayson, *et al.*, 2000). For both L and U trials, participants wore, when possible, military issue C95 combat dress (trousers, combat belt and lightweight shirt) and combat boots. If this was not possible, walking boots, lightweight walking trousers and shirt were worn.

During the L trial, each participant carried the same military issue Bergen (rucksack) containing evenly-spaced gravel bags and padding totalling 24 kg. A simulated rifle with a mass of 4.3 kg was also carried providing a total external load of 28.3 kg.

10.2.4 Experimental Measures

Expired gas was collected at 15 minute intervals, with samples collected for 60 seconds using the Douglas bag method. Oxygen uptake ($\dot{V}O_2$) and energy expenditure (EE) were calculated according to Jeukendrup and Wallis (2005). The protocol is outlined in Chapter 3.

Core body temperature was recorded at 5-minute intervals for safety reasons throughout the duration of each trial using an ingestible temperature sensor (HQ Inc., Palmetto, FL, USA), transmitting to a hand held recorder. A core temperature threshold of 39.5 °C was set as the

upper limit of safe working conditions. If any participant reached a core temperature of 39.5 °C the trial would have been stopped and the individual actively cooled.

Ratings of Perceived Exertion (RPE) were obtained on a 15 point scale (Borg, 1982) following each gas collection. Water was provided *ad libitum* throughout the trial and kept at a constant 37.5 °C using a digital water bath (Grant Instruments, Cambridge, UK). Warming the participants' drinking water attenuated any interference of ingested water temperature with temperature pill in the gut (Wilkinson *et al.*, 2008a).

10.2.5 Accelerometry

Two body-borne, accelerometer-based movement sensors were worn during each trial. The 3DNX (Biotel Ltd., Bristol, UK) triaxial accelerometer, and the ActiHeart (v4.2, CamnTech, Cambridge, UK) combined HR and accelerometer device. Both devices aim to predict EE and are described in detail in Chapter 3 (3.3.1 and 3.3.2 respectively).

The Actiheart was attached to a modified material strap (Polar sports strap, Kempele, Finland) and worn around the chest. Sleeping HR was entered as the lowest value observed during the measurement of RMR. Data were downloaded using bespoke software (Version 4) and exported to a spreadsheet where EE was calculated for each 15-minute sample interval. The Actiheart was calibrated using the group calibration described by Brage et al. (2007).

The 3DNX was worn at the lower back; attached to the participants' combat belt. All accelerometer data were downloaded after each trial using dedicated software (Biotel Limited, Bristol, UK) and imported into a spreadsheet where the 5-second epoch accelerometer counts (PAC) from X-, Y- and Z-axes were summed. The mean PAC value was calculated for each 15-minute sample interval.

Physical activity counts were used to predict $\dot{V}O_2$ (ml·kg^{-1}·min^{-1}) using the walking equation outlined in Chapter 6 (Figure 6.7). The predicted relative $\dot{V}O_2$ value (ml·kg^{-1}·min^{-1}) was converted to an absolute value by multiplying by body mass and dividing by 1000 to convert to absolute $\dot{V}O_2$ (l·min^{-1}). Absolute $\dot{V}O_2$ (l·min^{-1}) was converted to an estimate of EE (kcal·min^{-1}) using the ACSM assumption that for every litre of oxygen consumed, 5 kcal of energy is expended (Glass *et al.*, 2007).

10.2.6 Pattern recognition

Three high-resolution triaxial accelerometer-devices (Geneactiv, Activinsights, Cambridge, UK) were worn by each participant. The GENEA, the predecessor of the Geneactiv is described in detail elsewhere (Esliger, *et al.*, 2010). Briefly, the Geneactiv is a small (43 x 40 x 13 mm), and lightweight (16 g) unit. The device has an adjustable sampling frequency of between 10-100Hz and is capable of storing up to seven days of data sampled at 100 Hz. One sensor was worn in the small of the back and one at each hip (midaxillary line). Each Geneactiv was attached directly to the skin (Hyperfix self-adhesive dressing retention tape, Smith & Nephew Healthcare Ltd., UK) and set to record the raw accelerometer signal at 100Hz. Each unit was oriented so that the X-axis corresponded to vertical movement (for participant comfort).

The accelerometer data from the vertical axis only was exported to MATLAB where spurious data were removed using a Butterworth filter (anomalies coinciding with adjustment of clothing or Bergen).

An initial Fast Fourier Transform (FFT) was conducted and an initial visual inspection of the plot (Figure 10.1) indicated differences between unloaded and loaded conditions in the frequency range of >5Hz. The analysis also revealed that some individuals recorded greater amplitudes than others. However, these differences were consistent across the whole frequency range allowing the signal to be normalised to account for individual variation. In order to normalise the data between individuals, the frequency range >5Hz was integrated and expressed as a percentage of the integral across the whole 0-50Hz frequency range examined.

Using the first 15 minutes of data from all participants, normal distribution curves (assuming normality) were fitted to the loaded and unloaded data (Figure 10.2). Only the first 15 minutes were used to create the model thereby ensuring that the data were not contaminated by possible fatigue-related changes in gait. The x intersect (0.1) was derived; any values greater than 0.1 were classified as unloaded and any values less than 0.1 were classified as loaded. This classification model was applied to all subsequent minutes of data for all participants for each 15-minute sample interval. For clarity, Figure 10.3 is a flow diagram of the analysis procedure.

10.2.7 Biomechanical analysis

A subsample of 6 participants (mean ± SD; age: 21.5 ± 0.6 years; height: 1.82 ± 0.05 m; mass: 78.1 ± 6.0 kg, LC experience 3.3 ± 3.1 years) were instrumented with 10 reflective markers (CODA motion analysis system, Version CX1-6.30B) sampling at 200 Hz to capture objectively any changes in gait patterns between conditions.

Markers were attached to participant's clothing and footwear at the fifth metatarsal-phalangeal joint, the calcaneous, the malleolus, the femoral condyle, the greater trochanter, the mid-thigh (placed directly in the middle of the markers at the femoral condyle and greater trochanter) and the greater tubercle. Further markers were placed medially on the left foot at the first metatarsal-phalangeal joint and calcaneous. All markers were placed on the right side of the body.

Data were sampled for a period of 12 seconds after the first 13 minutes of walking. Data collection was scheduled in the minute preceding expired gas collection to avoid any potential gait disturbances associated with the collection.

Data were filtered and interpolated to estimate coordinate data at time points where markers were obscured. A range of spatiotemporal and joint angle kinematic variables were calculated for a total of six consecutive strides from each trial.
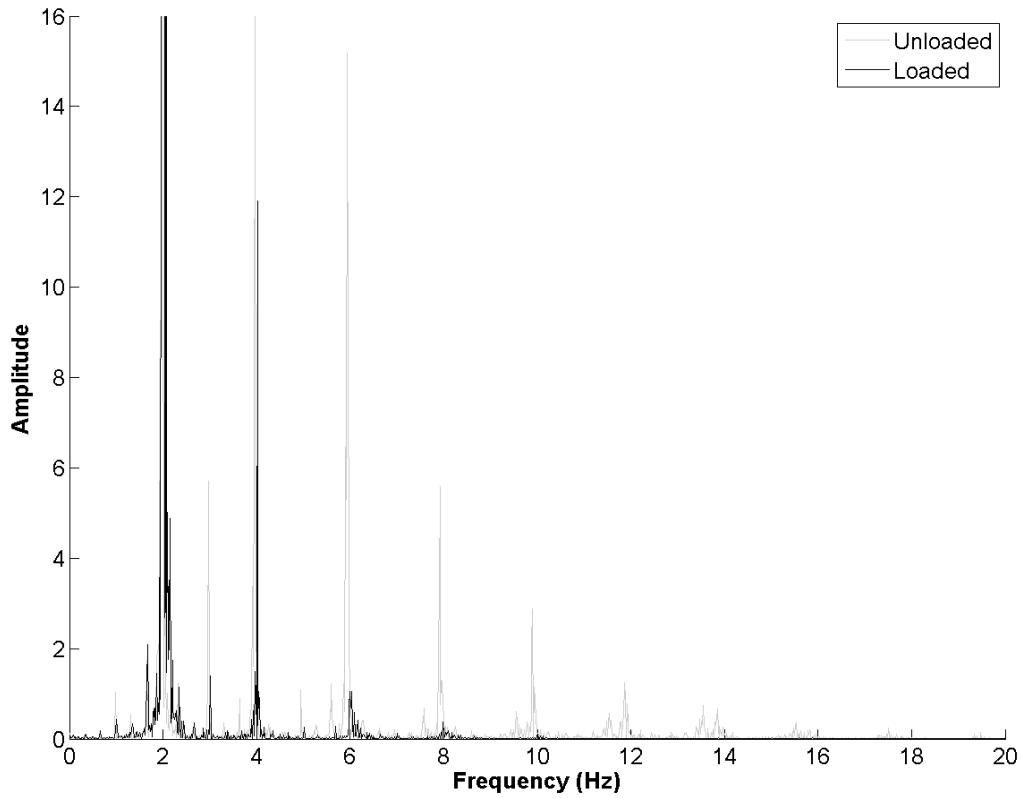
Figure 10.1 Example Fast Fourier Transformation plot for one participant, showing the differences in the signal frequency distribution between loaded and unloaded trials
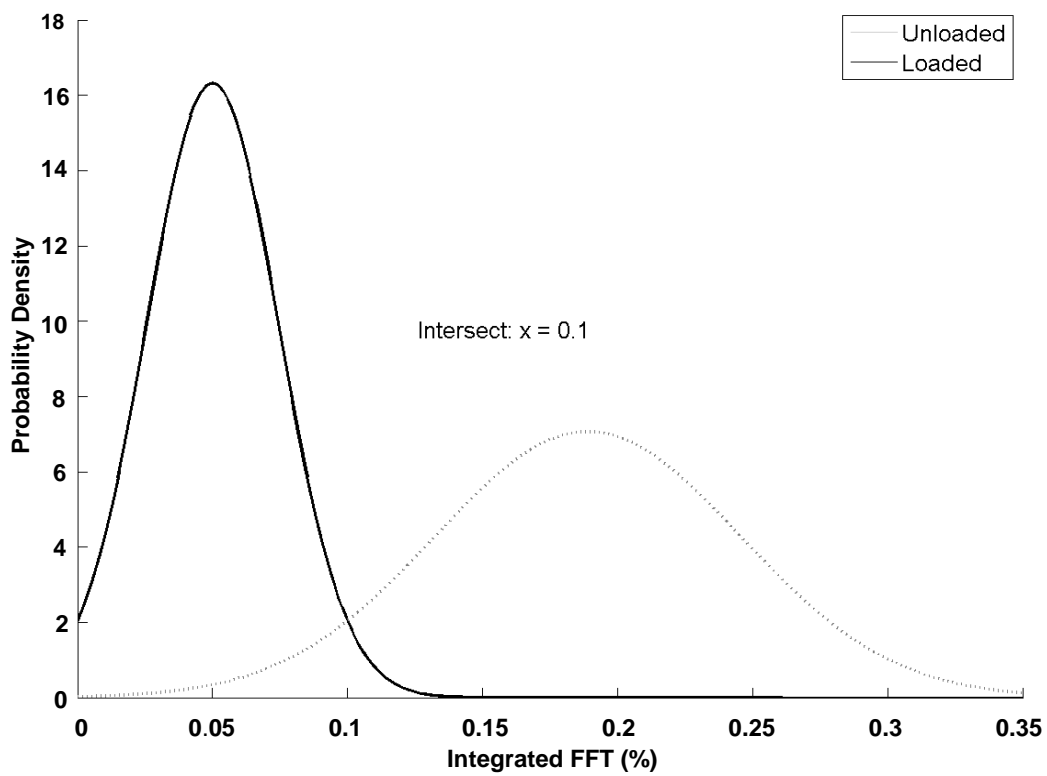
Figure 10.2 Normal distribution curves for all data (minutes 0-15). FFT, fast fourier transform.

```
┌─────────────────────────────────────────────────────────────┐
│                     Data collection                          │
│                     [2 hrs, 100Hz]                           │
└─────────────────────────────────────────────────────────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│                     Data reduction                           │
│                [cropping to 2hr .CSV files]                  │
└─────────────────────────────────────────────────────────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│                     Data Export                              │
│                     [MATLAB]                                 │
└─────────────────────────────────────────────────────────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│         Removal of spurious data [Butterworth filter]        │
└─────────────────────────────────────────────────────────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│                     FFT analysis                             │
│              [visual inspection of group data]               │
└─────────────────────────────────────────────────────────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│        Integration of frequencies >5Hz for each participant  │
└─────────────────────────────────────────────────────────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│                     Normalisation                            │
│  [Integrated FFT % = (Integral of Frequency range >5Hz/      │
│       Integral of 0-50Hz frequency range) * 100]             │
└─────────────────────────────────────────────────────────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│              Normal distribution curves fitted               │
│           [first 15 minutes of normalised data]              │
└─────────────────────────────────────────────────────────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│            x-axis intersect identification (figure 1)        │
└─────────────────────────────────────────────────────────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│               Classification model developed                 │
│   [if value > 0.1, unloaded; if value < 0.1, loaded]         │
└─────────────────────────────────────────────────────────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│                     Validation                               │
│       [model applied to all subsequent 1-minute segments]    │
└─────────────────────────────────────────────────────────────┘
                              ↓
┌─────────────────────────────────────────────────────────────┐
│                     Evaluation                               │
│    [accuracy confusion matrix, 15-minute intervals           │
│                    (% accuracy)]                             │
└─────────────────────────────────────────────────────────────┘
```

Figure 10.3 Flow diagram of the model building and evaluation process for load carriage recognition

10.2.8 Statistical Analysis

Physiological measures

Two-way (2x8) repeated measures analysis of variance (ANOVA) tests were performed to determine within-participant main effects for trial (two levels) and time (eight levels), as well as interaction effects for a range of physiological variables (e.g. $\dot{V}O_2$ and RPE). Where significant differences were identified, students' paired t-tests were performed post hoc to identify where differences occurred between conditions or between specific time points. A Bonferroni adjustment was used to minimise type 1 error if significant differences were found at more than three time points.

Pearson bivariate correlations were used to determine the relationship between measured and predicted EE and students' paired t-tests were used to assess the agreement between measured and predicted values.

Pattern Recognition

A classification-accuracy confusion matrix was generated to assess the percentage of correct classifications at each time interval. Box and whiskers plots were generated for each 15-minute sample interval to show the differences between normalised FFT values for loaded and unloaded conditions across time. The extent to which the whiskers overlap indicates where normalised FFT values were similar for both conditions causing potential misclassification.

Biomechanical analysis

Paired t-tests were used to determine which variables were different between loaded and unloaded conditions.

Statistical analysis was performed using SPSS for Windows v.18 (SPSS Inc., Chicago, IL, USA) and Statistical significance was set *a priori* at $p < 0.05$ for all analyses. Data are presented as mean ± SD.
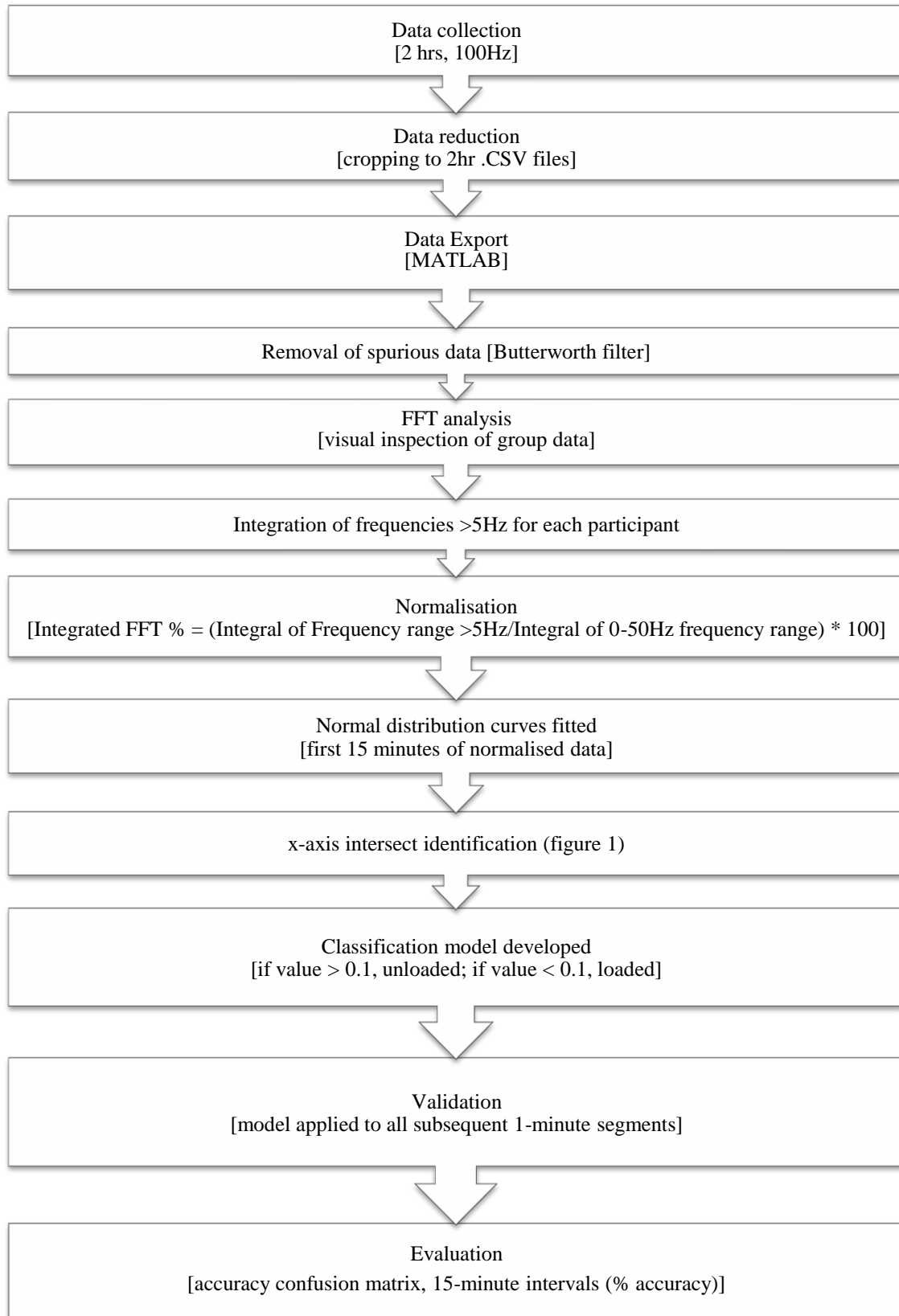
10.3 RESULTS

Two participants were unable to complete the L condition due to injury and were removed from the analysis. Participant characteristics are detailed in Table 10.1 for the ten subjects who completed both trials. The mass of the Bergen and weapon carried in the loaded condition equated to a mean of 35.4 ± 2.8 % of participants' body mass. When participants' clothing was

included, the mean total load increased to 31.6 ± 0.6 kg, representing 39.9 ± 3.0 % of participants' body mass.

Participant	Age (yrs)	Height (m)	1.5 mile run (mm:ss)	Body Mass (kg)	RMR (kcal.day^{-1})	LC experience (yrs)	Predicted $\dot{V}O_{2max}$ (ml·kg^{-1}·min^{-1})
1	34	1.71	10:30	87.5	2060	4.0	49.50
2	24	1.78	07:50	72.5	1656	5.0	65.19
3	18	1.84	09:14	72.1	1940	0.5	55.83
4	22	1.84	09:50	83.9	2126	1.0	52.64
5	35	1.77	09:38	83.4	2124	9.0	53.66
6	22	1.83	11:00	84.1	1624	6.2	47.41
7	21	1.85	09:30	74.2	2180	7.0	54.34
8	21	1.81	08:35	82.4	2100	0.6	59.79
9	23	1.73	08:30	71.8	1872	2.0	60.32
10	23	1.79	08:30	77.3	2246	1.5	60.32
Mean ± SD	**24.3 ± 4.6**	**1.80 ± 0.05**	**09:19 ± 00:59**	**78.9 ± 6.0**	**1992 ± 215**	**3.7 ± 3.0**	**55.9 ± 5.5**

Table 10.1 Participant characteristics. RMR: resting metabolic rate. LC: Load carriage

10.3.1 Oxygen Uptake and RPE

There was a main effect of trial on $\dot{V}O_2$ (L vs. U) (p<0.001) and a main effect of time (p<0.001). Post hoc analysis showed differences at all eight time points (p=0.001-0.019). There was also an interaction effect for $\dot{V}O_2$ (p<0.001). During the L trial, $\dot{V}O_2$ increased across time with a difference found between 15 and 120 minutes. Furthermore, there was a main effect of trial (p<0.001) and a main effect for time (p<0.001) on RPE. Post hoc analysis showed differences between trials at seven out of eight time points. The difference in RPE between L and U trials increased across time with a difference between 15 minutes and 90, 105 and 120 minutes respectively (p<0.001), for the loaded condition (Figure 10.4).

10.3.2 Energy Expenditure

The relationships between measured EE and EE predicted by accelerometer-based devices are shown in Figure 10.5 and Figure 10.6. The Actiheart failed to record HR on five occasions due to internal faults; for these participants' Actiheart data were removed from the analysis for both loaded and unloaded conditions. The 3DNX unit was removed on five occasions during the loaded condition due to participant discomfort; these participants' data were removed from the analysis for both loaded and unloaded conditions.
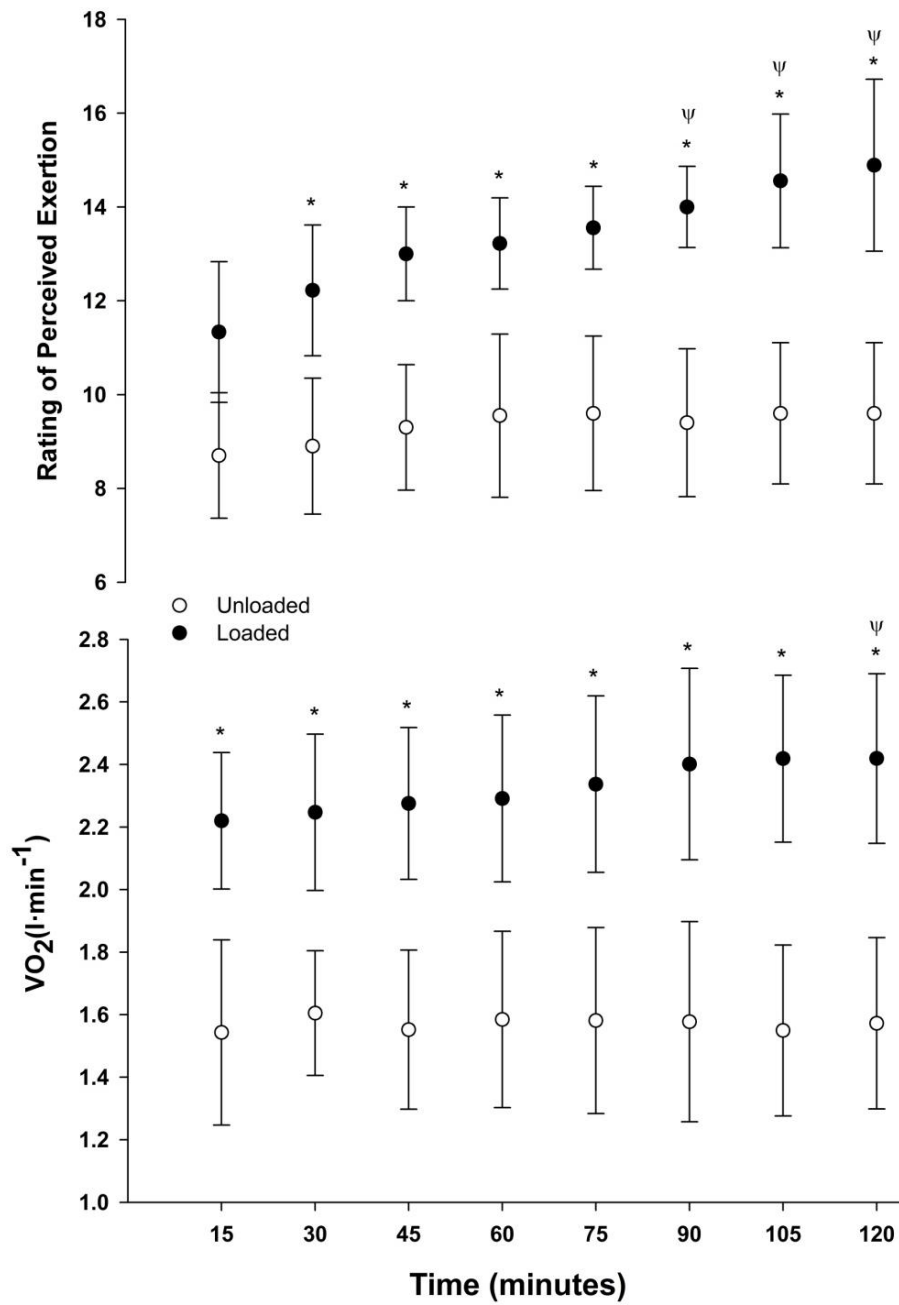
Figure 10.4 Mean ($\pm$SD) difference in absolute $\dot{V}O_2$ between loaded (L) and unloaded (U) conditions. * denotes a significant difference between conditions. ψ denotes a significant difference between time points.
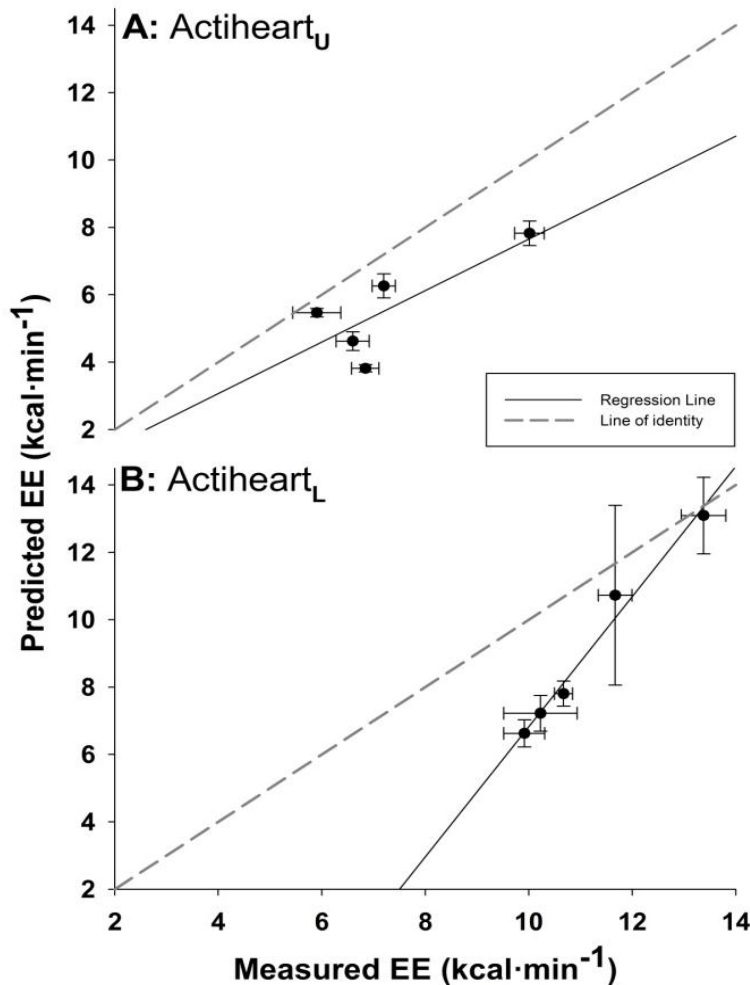
Figure 10.5 The relationship between measured and predicted EE for the Actiheart. n=5 (subjects 1, 2, 5, 9 and 10). Error bars represent the standard deviation of both measured and predicted EE across all time points. L, loaded; U, unloaded.

The relationship between measured and predicted EE by the Actiheart for the unloaded trial (Actiheart$_U$) was not significant (p = 0.06, r^2 = 0.61, SEE = 1.14 kcal·min^{-1}). Although the Actiheart significantly underpredicted EE (p=0.02), the relationship between measured and predicted EE for the loaded condition (Actiheart$_L$) was significant (p < 0.01, r^2 = 0.98, SEE = 0.233 kcal·min^{-1}).

The relationship between measured and predicted EE by the 3DNX for the unloaded trial (3DNX$_U$) was not significant (p = 0.06, r^2 = 0.65, SEE = 0.48 kcal·min^{-1}). However, although the 3DNX significantly underpredicted EE (p< 0.001), the relationship between measured and predicted EE for the loaded condition (3DNX$_L$) was significant (p = 0.01, R^2 = 0.92, SEE = 0.35

kcal·min^{-1}). Mean PAC (counts·5s^{-1}) for 3DNX$_U$ were not different from mean PAC for 3DNX$_L$ (p=0.07).
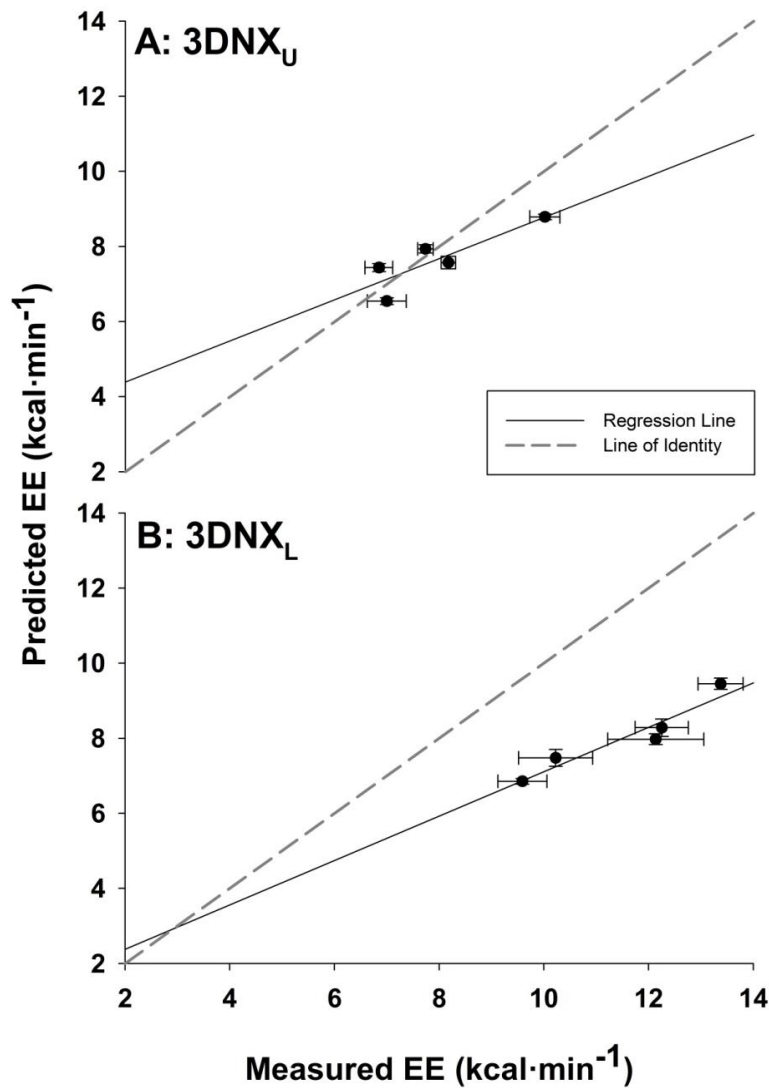
Figure 10.6. The relationship between measured and predicted EE for the 3DNX. n=5 (subjects 1, 4, 6, 7 and 10). Error bars represent the standard deviation of both measured and predicted EE across all time points.

10.3.3 Load carriage classification

The distributions of loaded and unloaded integrated FFT data are shown in Figure 10.7. The classification accuracy of the model is given in Table 10.2. The overall accuracy of the model was 99.5%.
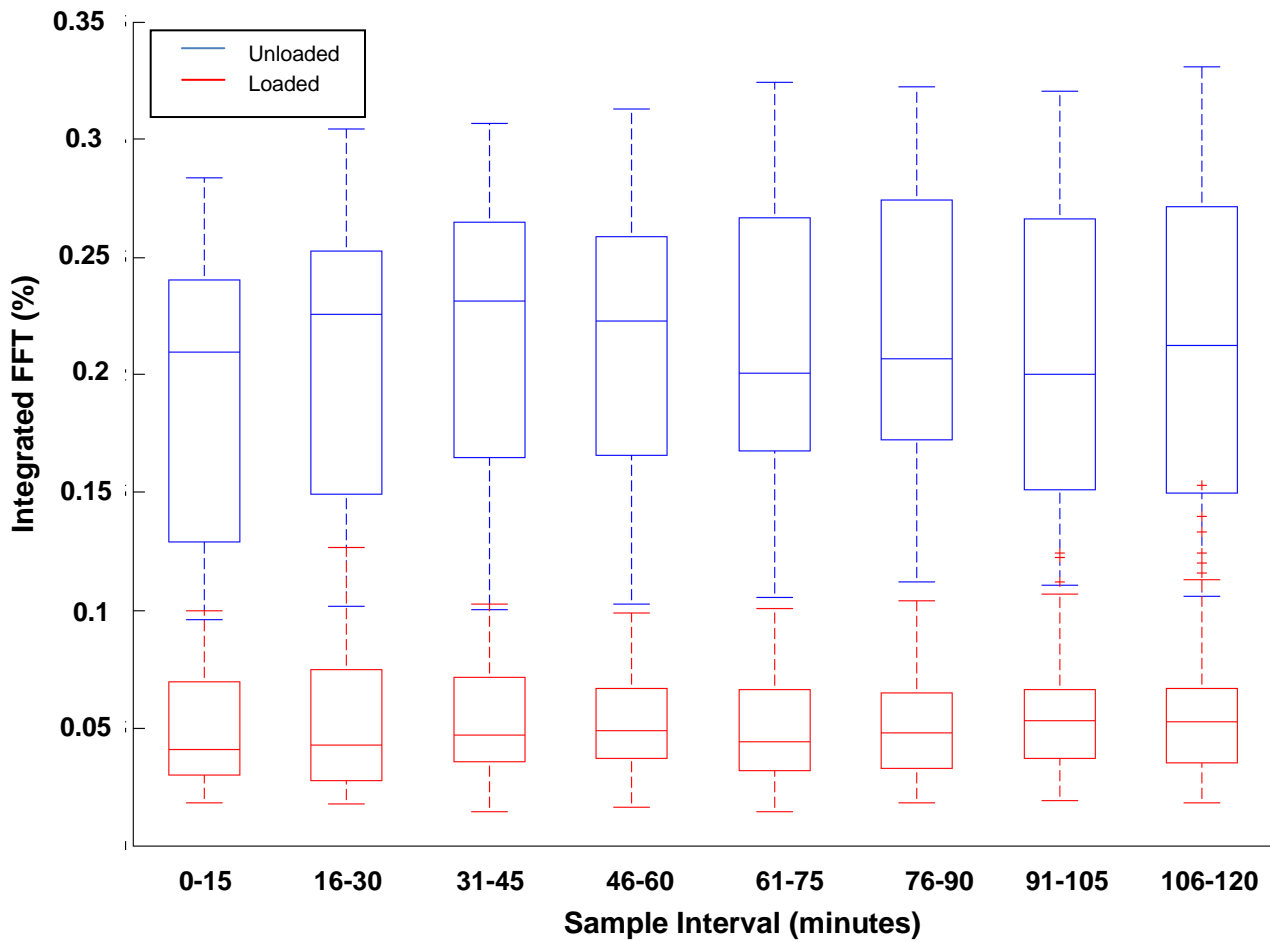
Figure 10.7 Box and whiskers plots showing the distribution of integrated FFT values with increasing time. The tops and bottoms of each "box" are the 25th and 75th percentiles of the samples respectively, with the line in the box indicating the sample median. Whiskers denote the range of the data (excluding outliers). Outliers are indicated by crosses and represent values which fall outside the 25th or 75th percentiles plus (or minus) 1.5 times the interquartile range.

			Time (minutes)							
			0-15	16-30	31-45	46-60	61-75	76-90	91-105	106-120
Condition	Unloaded	No. of samples	140	149	150	150	149	148	149	150
		Accuracy (%)	**98.6**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
	Loaded	No. of samples	138	150	145	132	137	148	149	132
		Accuracy (%)	**100**	**98**	**99.3**	**100**	**99.3**	**98**	**93.3**	**89.4**

Table 10.2 Classification accuracy of the load carriage recognition model. Where the number of samples is less than 150 this indicates removal of spurious samples by the Butterworth filter.

10.3.4 **Biomechanical analysis**

Two spatiotemporal parameters were different between conditions at minute 13. Load carriage decreased flight phase duration (p = 0.025) and increased double support time by 5.9 % (p = 0.001). Load carriage also effected seven kinematic changes which are shown in Table 10.3.

Variable	Unloaded (mean± SD)	Loaded (mean ± SD)	p - value
Mean trunk angle	-2.80° (± 4.21°)	-18.45° (± 2.68°)	< 0.001
Maximum hip extension	187.50° (± 6.51°)	178.64° (± 3.88°)	0.023
Maximum hip flexion	160.81° (± 5.95°)	138.15° (± 3.73°)	0.001
Hip range of motion	26.69° (± 3.56°)	40.49° (± 5.14°)	0.020
Maximum trunk extension	4.86° (± 4.27°)	-14.86° (± 2.46°)	< 0.001
Maximum trunk flexion	-10.36° (± 5.00°)	-22.64° (± 3.02°)	0.002
Trunk range of motion	15.22° (± 2.96°)	7.78° (± 1.93°)	0.019

Table 10.3. Changes in joint angle kinematics between loaded and unloaded conditions

10.4 DISCUSSION

Load carriage forms an integral part of day-to-day soldiering activities (Knapik, *et al.*, 2004). The ability to recognise the additional energy expended due to carrying loads using unobtrusive, wearable movement sensors could improve the prediction of daily EE in military populations. Current methods of processing accelerometer signals may be unsuitable when day-to-day routines contain resistance-style activities.

This is the first study to demonstrate the feasibility of using accelerometer signal characteristics to detect LC which could lead to improved prediction of EE. We hypothesised that differences in gait associated with carrying a military-style load would be reflected in the accelerometer signal and could be detected using FFT analysis. A secondary aim of the study was to build a classification model and assess its accuracy during a prolonged L trial.

10.4.1 **Prediction of EE using PAC**

The initial phase of this study investigated the ability of traditional accelerometer-based devices to predict EE associated with treadmill ambulation with (L) and without (U) load.

$\dot{V}O_2$ was greater for L compared to U walking and increased during L between 15 and 120 minutes. This $\dot{V}O_2$ drift has previously been observed during LC (Blacker, *et al.*, 2010b; Patton

et al., 1991). $\dot{V}O_2$ drift during LC has been suggested to be caused by a change in mechanical efficiency due to fatigue (Patton*, et al.*, 1991) or a change in substrate oxidation (Blacker*, et al.*, 2011). The increase in energy cost during prolonged LC suggests that an equation aiming to predict EE should include a variable capturing the minutes elapsed since LC commenced.

When $\dot{V}O_2$ was expressed as EE, there were no relationships between predicted and measured EE for either the $3DNX_U$ or the $Actiheart_U$. This is not wholly unexpected for $3DNX_U$ given that the trial was completed at a constant treadmill speed reflecting the pace required to pass the British Army backpack fitness test. Physical activity counts and subsequent predictions did not vary greatly between individuals across time (shown by the narrow Y error bars in Figure 10.6 A). The prediction equation generated in chapter 6 which was used to calculate EE from 3DNX physical activity counts did not include anthropometric variables such as body mass as covairates. Body size; which is highly related to EE (Nevill*, et al.*, 2005), is likely to explain the most variance in EE between individuals when all participants are walking at the same constant speed.

For the $Actiheart_U$ the branched equation model used to predict EE does include body mass as a covariate, however no individual HR calibration was performed prior to the main trials and true sleeping HR was not recorded which could explain the weak relationship as the HR $\dot{V}O_2$ relationship is known to vary at an individual level (Ainslie*, et al.*, 2003).

For both devices there were relationships between predicted and measured EE in the L condition. This may be because higher EEs were associated with less efficient movers (i.e. higher PAC) leading to more detectable variability between individuals. However, predicted EE was lower than measured EE. This consistent under-prediction demonstrates the difficulties traditional accelerometer devices face when trying to account for load-bearing activities. If LC could be detected, and the external load known, then this could be added to body mass to create an alternative prediction equation for use in a two-regression model similar to that of Crouter et al. (2006).

No other research has specifically investigated using a single accelerometer and the raw signal to identify LC tasks. Wyss and Mader (2010) used body acceleration (Actigraph GT1M recording data at 2-second epochs), HR and backpack acceleration to develop a decision tree to classify six of the most common military-specific activities performed by Swiss Army infantry recruits in training. This method showed the classifier's inability to discriminate 'marching with a

backpack' from 'walking' without using the data from a sensor placed on the backpack. This finding supports that of the present study where PAC from the unloaded condition were not different from the mean PAC values generated during the loaded condition despite differences in EE between the two conditions.

These data should be interpreted with care as only a sub-sample of data from five participants could be used for each device and participants were not the same for both device sub-samples. Despite the small sample size, this sub analysis indicates that traditional accelerometry is unlikely to be a suitable method for predicting EE when LC is performed. Of note is the both the failure rate of the Actiheart heart rate component (50%) and the removal of the 3DNX from the combat belt due to discomfort (50%). Chapter 7 suggested that the lower back was the most suitable location for positioning the 3DNX and Chapter 9 deployed the device during periods of military exercise without issue however this may need to be reconsidered to maximise compliance if the unit is to be used during periods of LC.

10.4.2 Load carriage classification

The model created to classify loaded from unloaded ambulation returned promising accuracy. The differences seen in the spatiotemporal and kinematic variables translated to differences in the raw acceleration output generated in the vertical axis. The addition of load led to an increase in double support time and decreased flight phase duration which have been previously observed during LC (Birrell*, et al.*, 2009). The increase in the time both feet spent in contact with the ground during one stride could have conferred with a different dominant frequency content of the vertical acceleration signal compared to unloaded ambulation (Figure 10.1).

The model validation showed that it was able to correctly classify unloaded ambulation with 99.8% accuracy for 1186 samples during treadmill walking at 6.4 km·hr^{-1}. When the model was validated in the loaded condition accuracy decreased across time. This was indicated by the overlap in the whiskers, and the outliers in the last two stages on the box and whiskers plot. Loaded ambulation was misclassified as unloaded ambulation on 24 of 281 occasions in the last 30 minutes of the loaded trial. A direct measure of fatigue was not collected in this study; however there was an increase in RPE across time during the loaded trial. In a similar L protocol (120 minutes, walking at 6.4 km·hr^{-1}, carrying 25 kg), fatigue of the lower limb, trunk and shoulder muscle groups has been demonstrated (Blacker *et al.*, 2010a). These findings suggest

that fatigue-related changes in gait may have been a contributor to the classification error towards the end of the trial.

There are few data available detailing the spatiotemporal and kinematic changes associated with prolonged LC therefore the link between fatigue, gait characteristics and reduced classification accuracy cannot be fully established. A study conducted by Frykman et al. (1994) manually digitised the video captured during paced walking pre and post a 20 km load carry. The authors observed significantly increased forward trunk inclination, and decreased flight phase, stride frequency and stride length following the load carry. Although this protocol used different methodology to the present study it does suggest that fatigue causes changes in gait characteristics. Attwells et al. (2008) also observed an increased trunk lean following a two-hour military load carry and increases in stride frequency that were approaching significance.

10.4.3 Limitations

The main limitation of this study was the small sample size. In trying to maintain the ecological validity of the trial by recruiting experienced load carriers (from the Territorial Army and Marine reservists where possible), recruitment of subjects was difficult. When fitting normal distribution curves to the first 15 minutes of data (Figure 10.2), normality was assumed despite observing a degree of skewness. It was decided that assuming that the data were non-normally distributed would be as inaccurate for a small sample as assuming that the data would subsequently become normally distributed with the addition of more participants.

Unfortunately the signal from the accelerometer and the CODA system were not synchronised and the biomechanical analysis was only performed on 6 participants. Although the biomechanical analysis provided evidence showing the differences in gait between loaded and unloaded ambulation, temporal aspects of the accelerometer signal could not be objectively identified (e.g. foot strike, stance phase, toe off). Secondly, it was not possible to verify the changes in gait which caused the changes in accelerometer signal across the 120 minutes. CODA markers fell off the participants' clothing at varying time points which prevented the collection of biomechanical data across time. An alternative method for attaching markers should be employed in future. The small-scale biomechanical aspect of this study was included for descriptive purposes; interpretations should be made with care.

While the simplicity of the raw signal analysis could be regarded from a user perspective as a strength of this study, it is likely that FFT analysis alone is not sufficient to distinguish LC in a

more variable free-living scenario as the energy content of the signal could be contaminated by other motion artefacts. Wang et al. (2010) included FFT in addition to other signal features during their classification of inclined versus level walking. This could be a possible future direction if the detection of LC is of interest to the military.

10.4.4 Summary and Conclusions

Health and military-specific researchers alike have recognised the potential for assigning individual prediction equations to predict EE following the classification of discrete activities (Crouter*, et al.*, 2006; Wyss*, et al.*, 2011). The findings of the present study indicate that EE is underestimated using current accelerometer-based technologies and could be improved by the differentiation of loaded from unloaded ambulation and the assigning of separate prediction equations accordingly. The use of FFT is a valuable first step in the development of more sophisticated accelerometer signal processing techniques however it is unlikely to be sensitive enough as a standalone methodology for detecting more varied and complex movements.

In conclusion, the model designed to discriminate L from U ambulation classified LC correctly in 97.2% of cases across two hours with an overall accuracy of 99.5 % for both conditions. Prolonged LC has been shown to illicit an upwards drift in $\dot{V}O_2$ and associated EE. The classification of loaded ambulation is less accurate at 105 and 120 minutes compared to 15 minutes and this is likely due to the effect of fatigue on gait parameters. Prediction of EE may be improved by using the accelerometer raw signal to distinguish LC from unloaded ambulation however, more complex modelling of the relationship between gait characteristics, $\dot{V}O_2$ and fatigue over time could further improve LC classification accuracy. Future research investigating the feasibility of classifying LC with the addition of uneven terrain, variable speeds, variable loads and more participants would make any subsequent model more directly applicable to military scenarios. Indeed, the ability to detect large differences in EE associated with progressively increasing military loads is of particular importance.

11   GENERAL DISCUSSION

A simple online search (www. http://www.ncbi.nlm.nih.gov/pubmed/) using accelerometry as a keyword returns 663 results since 2009 alone; the year this research programme began. It is clear that accelerometry has emerged as a popular tool for objectively quantifying physical activity yet a challenge that still faces the research community is the conversion of accelerometer output to EE. It has been shown widely that, in a laboratory setting, there are strong relationships between accelerometer output and EE. However, when moving from scripted activities to free-living scenarios, there does not appear to be an accelerometer-based device to date that can account for the varying relationships between day-to-day activities and EE; certainly not without the addition of either individually calibrated and integrated HR monitors or a network of multiple sensors.

The overarching aim of this thesis was to improve the prediction of EE in military personnel using an accelerometer-based device. From the first three experimental chapters it became clear that in order to improve prediction accuracy, a methodology accounting for the energy expended due to resistance style activities (i.e. load carriage) would be central to reducing prediction error.

11.1   SUMMARY

The first experimental chapter (Chapter 6) provided a thorough assessment of the reliability and validity of the 3DNX accelerometer. Care was taken in this thesis to use an instrument which was shown to be valid and reliable in a mechanical setting and during incremental human treadmill exercise so that any error in subsequent studies could be attributed to sources other than the accelerometer itself.

The second experimental chapter investigated the effect of accelerometer placement (hip or lower back) on the reliability and validity of the 3DNX output. This chapter provided evidence for using the lower back as the anatomical placement of preference for future studies. The results showed that in a sample of female participants with a narrow range of skinfold thickness, accelerometer output was higher at the hip than at the back during treadmill exercise with suprailliac skinfold thickness mediating the relationship.

The third experimental chapter (Chapter 8) sought to compare three commercially available devices for the prediction of free-living EE in a military population. Although the

sample size was small, the results indicated that the Actigraph GT3X was not a suitable device for use in military populations as it underestimated PAEE to the greatest degree. The Actiheart did not clearly outperform the 3DNX in terms of prediction accuracy; this was attributed to the use of a group rather than individual HR calibration.

Chapter 9 combined the data from ten independent military cohorts with the aim of developing a multivariate model to predict EE. This was the first large scale study using the doubly labelled water technique as a criterion measure. Accelerometer output explained small but meaningful portions of variance in TEE and PAEE however; the particularly poor overall prediction accuracy for PAEE indicated that, where possible, equations should be derived specific to the population being studied. Furthermore, results indicated that there was a large portion of energy expended due to activity that was neither explained by anthropometric variables nor detected by accelerometer output.

It was suggested that day-to-day military activities such as load carriage would be difficult for traditional accelerometer processing methodologies to detect due to the additional energy cost of carrying a load not conferring with a similar increase in whole body movement. Therefore, Chapter 10 used an alternative accelerometer device (capable of storing raw accelerometer signal) to differentiate between loaded and unloaded treadmill ambulation based on the signal characteristics associated with each. A model was developed to classify load carriage which correctly identified the activity in 97.2% of cases when validated across a two hour time period. The model was less accurate in the last 30 minutes of the trial which was attributed to the effect of fatigue on gait parameters.

11.2 PREDICTING ENERGY EXPENDITURE USING ACCELEROMETER DEVICES

The initial laboratory study conducted in Chapter 6 showed a strong positive relationship between physical activity counts (PAC) and $\dot{V}O_2$ during incremental treadmill exercise. Different equations were generated for walking and running as the slopes and intercepts of the regression lines were different. It was recognised that using equations derived during treadmill walking and running could not be applied to free-living populations. Chapters 8 and 9 sought first to directly compare accelerometer devices and second to develop a prediction equation in a large, highly compliant sample. However due to the applied nature of this thesis, some compromises were made in order to maintain the 'ecological validity' of the research.

11.2.1 Prediction of RMR

One of the main limitations of the field-based studies (Chapters 8 and 9) was the lack of a direct measurement of resting metabolic rate (RMR). It was not possible for participants in these studies to visit a laboratory therefore the prediction equation developed by Schofield (1985) was used to predict resting metabolic rate for each individual. It is recommended that prediction equations for RMR are developed within a similar population to that studied. In the majority of cases this is not possible so authors must choose the most appropriate equation for their population. The inclusion of conscripts in the population used to develop the Schofield equation meant that in the absence of a prediction equation developed exclusively in military personnel, it was deemed the most appropriate choice. Comparison of the mean measured RMR value for participants in Chapter 10 and the mean predicted value for male participants in Chapter 9 indicates that the Schofield equation may underestimate RMR, for military personnel, by approximately 200 kcal·day^{-1}. This is unsurprising as it is likely that military personnel would have a larger percentage of fat free mass than the majority of the population due to the nature of the occupation. It is known that fat free mass has approximately 6.5 times higher resting EE than adipose tissue (124.3 vs. 18.8 kJ·kg^{-1}·day^{-1}) (Elia, 1992) therefore, development of an RMR prediction equation for military personnel would be of benefit for future field studies where direct measurement of RMR is not practical.

11.2.2 Calibration of Accelerometer Devices

Throughout this thesis, no attempts were made to individually calibrate the accelerometer devices. It is known that the relationship between HR and work rate varies between individuals (Ainslie*, et al.*, 2003) and as such the Actiheart device offers a step-test calibration function to try and capture this variability. In keeping with the applied nature of this thesis, it was deemed that performing individual calibrations in larger, population-wide military studies would not be acceptable so a "black box" approach was taken to mimic that of a novice user. It seems that a simple, field-based calibration procedure could have improved the prediction accuracy achieved by both the Actiheart and the 3DNX.

Since the completion of this study, a field-based calibration for the 3DNX has been proposed as a practical method of normalising physical activity counts (PAC) for

individual differences. The premise being that the total counts accrued by each in the first six stages of the multi-stage fitness test (MSFT) acted as a calibration factor. The total PAC accumulated each day were then divided by the calibration factor to produce adjusted values (PAC_{MSFT}). When this method was employed by Blacker et al. (2012), the relationship between PAC and PAEE (doubly labelled water-derived), r^2 improved from 0.44 to 0.72. Furthermore, PAC_{MSFT} significantly contributed to the explained variance in a multivariate model ($r^2 = 0.68$, SEE = 7.67%) for the prediction of PAEE where unadjusted PAC did not. Using a similar field-based calibration could also be suitable for the Actiheart due to the incremental nature of the MSFT and should be investigated further.

11.3 CLASSIFICATION OF LOAD CARRIAGE

Using the raw signal from the Geneactiv accelerometer to build a model capable of discriminating loaded from unloaded ambulation was successful (Chapter 10). The model was built using simple processing steps and showed clear differences in the frequency distribution of loaded and unloaded signals.

The main limitation of this study was the small sample size. When fitting normal distribution curves to the first 15 minutes of data (Figure 10.2), normality was assumed and curves fitted accordingly despite a degree of skewness observed. It was decided that it would be as inaccurate to assume skewness as it would be to assume that if more participants were added, the distribution would become normal.

Further limitations of the study were the failure to synchronise the real time of the accelerometers to the CODA markers and the failure to collect biomechanical data throughout the duration of the trial. Although the biomechanical analysis provided evidence showing the differences in gait between loaded and unloaded ambulation, temporal aspects of the accelerometer signal could not be objectively identified (e.g. foot strike, stance phase, toe off). Secondly, it was not possible to verify the changes in gait that caused the changes in accelerometer signal across time. CODA markers fell off the participants' clothing at varying time points which prevented the collection of biomechanical data across time. An alternative method for attaching markers must be employed in future.

If more complex models are to be built to account for changes in speed, load and terrain, these additional data would aid the interpretation of accelerometer signals and subsequent model development.

11.4   FUTURE RESEARCH DIRECTIONS

The final experimental chapter in this thesis was designed to investigate the feasibility of discriminating L from U ambulation in a shift away from traditional accelerometer methodology. The protocol required participants to exercise at a set speed, carrying a set load, for a set duration. This provided a relevant occupational framework within which to assess the accuracy of the classifier. However, it would be of benefit to continue the research by varying load carried and speed marched, extending the duration of the trial and eventually, altering the terrain covered (i.e. free-living). It is probable that the noise introduced to the accelerometer signal would make a more complex model-building process necessary. Wang et al (2009) used a more complex approach to classifying four different gradients of incline using a down-selection of 13 unique signal features including both statistical (e.g. Standard deviation of the 'left swing' segments in the anteroposterior plane), and 'energy' (e.g. frequency spectrum energy in the band 0.5-4.5 Hz for 'left swing' segments in the anteroposterior plane) features. This type of analysis would require substantial input from an engineering subject matter expert however, this type of collaboration is advocated by leaders in the field of accelerometry research (Heil*, et al.*, 2012).

The ability to detect LC would benefit the military in a number of ways. If the detection of LC improved the accuracy of predicting EE during military field exercises and operations then the army would have evidence on which to base their feeding strategies and ultimately better fulfil their duty of care to personnel. Furthermore, simply detecting the number of minutes per day spent carrying loads regardless of energy expended would be of use to the British Army. Modelling and manipulating risk factors for injury has long been a priority for the military in order to increase operational effectiveness and reduce the demands on associated medical care provision (Wilkinson *et al.*, 2011). Load carriage is known to be associated with injury (Birrell*, et al.*, 2009; Birrell *et al.*, 2007) and as such, information regarding the frequency and duration of LC would be a useful variable to include in future injury models.

Although the Actiheart may not be suitable for use within military populations due to its position on the chest, there are a number of technologies currently under development that could allow the integration of movement and biosignals in an unobtrusive manner. For example, textile-based smart systems exist such as the MagIC that incorporates seismocardiogram technology into a vest (Castiglioni *et al.*, 2007). However at present; the micro accelerations generated by the ejection of blood from the heart are masked by the major accelerations associated with locomotion (Di Rienzo *et al.*, 2011). A different technique called optical interferometrics, which is mainly used for unobtrusively detecting vital signs in resting or unconscious individuals, is also a prospect for quantifying heart rate. An optical cable can be in direct or indirect contact with the human body is capable of detecting the mechanical and acoustic activity of cardiac muscle through distortions and concomitant changes in cable length (Sprager *et al.*, 2012). However, this technique has not yet been adapted for use during human movement.

In terms of military application, the technology most advanced in its potential application is the integration of pressure sensors into insoles. Saito et al. (2011) have recently developed a system comprising eight small rubberised sensors that can fit into an individual's own shoe without affecting normal gait. With further development (the current data storage limits wear time to approximately 20 hours), a sensor such as this has the potential to both detect LC as well as variation in the mass being carried and changes in stride frequency. Integration of a pressure sensor into military issue boots warrants further investigation.

11.5 CONCLUSION

The research presented in this thesis has shown that in order to characterise EE in military populations, the identification of military specific activities such as load carriage is both necessary and feasible.

Evaluation in a mechanical set-up showed that physical activity counts generated by the 3DNX accelerometer were positively related to acceleration. Physical activity counts in turn were positively related to $\dot{V}O_2$ during walking and running but a single regression equation characterising both activities it not appropriate. A further laboratory study found that increases and variability in hip-derived physical activity counts, at least during walking, were explained by subcutaneous hip adiposity. The lower back is recommended

as the most suitable attachment site as it tends to have less fatty deposition in the area and avoids the need for knowing more detailed anthropometric measurements.

Field-based studies have shown that of three commercially available accelerometer devices, the mean Actiheart-derived predictions of physical activity energy expenditure were not different to doubly labelled water derived values; however, Actiheart had the widest 95% limits of agreement of all three devices. Users should determine what level of individual and group error is acceptable before its use on a case by case basis. A large field-based study found that accelerometry improved the estimation of free-living EE in sub-groups of military populations when combined with anthropometric variables. However, the contribution of accelerometry was small which was likely due to accelerometers failing to detect the additional energy cost of military-style activities such as load carriage.

The final laboratory based study found that a model designed to discriminate L from U ambulation during a prolonged treadmill protocol returned good overall classification accuracy. The accuracy of the classifier was reduced in the last 30 minutes of the trial and this was attributed to the effects of fatigue on gait parameters and the resulting accelerometer signal.

12 REFERENCES

Achten, J., & Jeukendrup, A. E. (2003). Heart rate monitoring: applications and limitations. *Sports Medicine, 33*(7), 517-538.

Ainslie, P., Reilly, T., & Westerterp, K. (2003). Estimating human energy expenditure: a review of techniques with particular reference to doubly labelled water. *Sports Medicine, 33*(9), 683-698.

American College of Sports Medicine., Thompson, W. R., Gordon, N. F., & Pescatello, L. S. (2010). *ACSM's guidelines for exercise testing and prescription* (8th ed.). Philadelphia: Lippincott Williams & Wilkins.

Arvidsson, D., Slinde, F., Larsson, S., & Hulthen, L. (2007). Energy cost of physical activities in children: validation of SenseWear Armband. *Medicine and Science in Sports and Exercise, 39*(11), 2076-2084.

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

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

Atallah, L., Leong, J. J., Lo, B., & Yang, G. Z. (2010). Energy Expenditure Prediction Using a Miniaturised Ear-Worn Sensor. *Medicine and Science in Sports and Exercise*.

Attwells, R. L. (2008). *Military load carriage: the effect of increased load, gender and load carriage duration on gait and posture.* PhD, Loughborough University.

Bhattacharya, A., McCutcheon, E. P., Shvartz, E., & Greenleaf, J. E. (1980). Body acceleration distribution and O2 uptake in humans during running and jumping. *Journal of Applied Physiology, 49*(5), 881-887.

Birrell, S. A., & Haslam, R. A. (2009). The effect of military load carriage on 3-D lower limb kinematics and spatiotemporal parameters. *Ergonomics, 52*(10), 1298-1304.

Birrell, S. A., & Hooper, R. H. (2007). Initial subjective load carriage injury data collected with interviews and questionnaires. *Military Medicine, 172*(3), 306-311.

Black, A. E. (1996). Physical activity levels from a meta-analysis of doubly labeled water studies for validating energy intake as measured by dietary assessment. *Nutrition Reviews, 54*(6), 170-174.

Black, A. E., Prentice, A. M., & Coward, W. A. (1986). Use of food quotients to predict respiratory quotients for the doubly-labelled water method of measuring energy expenditure. *Human Nutrition. Clinical Nutrition, 40*(5), 381-391.

Blacker, S., Horner, F. E., Wilkinson, D., Linnane, D., Brown, P. I., & Rayson, M. (2012). Improving estimation of energy expenditure with accelerometers during military training using a field-based calibration procedure. *Medicine and Science in Sports and Exercise, 44*(5), S174.

Blacker, S. D., Fallowfield, J. L., Bilzon, J. L., & Willems, M. E. (2010a). Neuromuscular function following prolonged load carriage on level and downhill gradients. *Aviat Space Environ Med, 81*(8), 745-753.

Blacker, S. D., Williams, N. C., Fallowfield, J. L., Bilzon, J., & Willems, M. E. (2010b). Carbohydrate vs protein supplementation for recovery of neuromuscular function following prolonged load carriage. *J Int Soc Sports Nutr, 7*, 2.

Blacker, S. D., Williams, N. C., Fallowfield, J. L., & Willems, M. E. (2011). The effect of a carbohydrate beverage on the physiological responses during prolonged load carriage. *European Journal of Applied Physiology and Occupational Physiology*.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 1*(8476), 307-310.

Bluck, L. (2008). Doubly labelled water for the measurement of total energy expenditure in man − progress and applications in the last decade. *Nutrition Bulletin, 33*, 80-90.

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

Bonomi, A. G., Goris, A. H., Yin, B., & Westerterp, K. R. (2009a). Detection of type, duration, and intensity of physical activity using an accelerometer. *Medicine and Science in Sports and Exercise, 41*(9), 1770-1777.

Bonomi, A. G., Plasqui, G., Goris, A. H., & Westerterp, K. R. (2009b). Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer. *Journal of Applied Physiology, 107*(3), 655-661.

Bonomi, A. G., Plasqui, G., Goris, A. H., & Westerterp, K. R. (2010). Estimation of Free-Living Energy Expenditure Using a Novel Activity Monitor Designed to Minimize Obtrusiveness. *Obesity (Silver Spring)*.

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

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

Bouten, C. V., Verboeket-van de Venne, W. P., Westerterp, K. R., Verduin, M., & Janssen, J. D. (1996). Daily physical activity assessment: comparison between movement registration and doubly labeled water. *Journal of Applied Physiology, 81*(2), 1019-1026.

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

Brage, S., Brage, N., Franks, P. W., Ekelund, U., & Wareham, N. J. (2005). Reliability and validity of the combined heart rate and movement sensor Actiheart. *European Journal of Clinical Nutrition, 59*(4), 561-570.

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

Brage, S., Ekelund, U., Brage, N., Hennings, M. A., Froberg, K., Franks, P. W., et al. (2007). Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity. *Journal of Applied Physiology, 103*(2), 682-692.

Brage, S., Wedderkopp, N., Franks, P. W., Andersen, L. B., & Froberg, K. (2003). Reexamination of validity and reliability of the CSA monitor in walking and running. *Medicine and Science in Sports and Exercise, 35*(8), 1447-1454.

Brychta, R. J., Klonoff, A., Klonoff, M., McMahon, G. C., Starrs-Zorn, S. J., Clawson, M. C., et al. (2010). *Influence of Actigraph Filter Settings on Detecting Low and High Intensity Movements.* Paper presented at the International Congress on Physical Activity and Public Health, Toronto.

Campbell, K. L., Crocker, P. R., & McKenzie, D. C. (2002). Field evaluation of energy expenditure in women using Tritrac accelerometers. *Medicine and Science in Sports and Exercise, 34*(10), 1667-1674.

Carlson, G. L. (1997). Nutrient induced thermogenesis. *Baillieres Clin Endocrinol Metab, 11*(4), 603-615.

Carter, J., & Jeukendrup, A. E. (2002). Validity and reliability of three commercially available breath-by-breath respiratory systems. *European Journal of Applied Physiology and Occupational Physiology, 86*(5), 435-441.

Carter, J., Wilkinson, D., Blacker, S., Rayson, M., Bilzon, J., Izard, R., et al. (2008). An investigation of a novel three-dimensional activity monitor to predict free-living energy expenditure. *Journal of Sports Sciences, 26*(6), 553-561.

Castiglioni, P., Faini, A., Parati, G., & Di Rienzo, M. (2007). Wearable seismocardiography. *Conf Proc IEEE Eng Med Biol Soc, 2007*, 3954-3957.

Catellier, D. J., Hannan, P. J., Murray, D. M., Addy, C. L., Conway, T. L., Yang, S., et al. (2005). Imputation of missing data when measuring physical activity by accelerometry. *Medicine and Science in Sports and Exercise, 37*(11 Suppl), S555-562.

Chen, K. Y., Acra, S. A., Majchrzak, K., Donahue, C. L., Baker, L., Clemens, L., et al. (2003). Predicting energy expenditure of physical activity using hip- and wrist-worn accelerometers. *Diabetes Technol Ther, 5*(6), 1023-1033.

Chen, K. Y., & Bassett, D. R., Jr. (2005). The technology of accelerometry-based activity monitors: current and future. *Medicine and Science in Sports and Exercise, 37*(11 Suppl), S490-500.

Chu, E. Y., McManus, A. M., & Yu, C. C. (2007). Calibration of the RT3 accelerometer for ambulation and nonambulation in children. *Medicine and Science in Sports and Exercise, 39*(11), 2085-2091.

Cipryan, L., & Litschmannova, M. (2012). Intra-day and inter-day reliability of heart rate variability measurement. *Journal of Sports Sciences*.

Colley, R., Gorber, S. C., & Tremblay, M. S. (2010). Quality control and data reduction procedures for accelerometry-derived measures of physical activity. *Health Reports, 21*(1), 63-69.

Compher, C., Frankenfield, D., Keim, N., & Roth-Yousey, L. (2006). Best practice methods to apply to measurement of resting metabolic rate in adults: a systematic review. *Journal of the American Dietetic Association, 106*(6), 881-903.

Cook, D. P., Mileva, K. N., James, D. C., Zaidell, L. N., Goss, V. G., & Bowtell, J. L. (2011). Triaxial modulation of the acceleration induced in the lower extremity during whole-body vibration training: a pilot study. *J Strength Cond Res, 25*(2), 298-308.

Crouter, S. E., Churilla, J. R., & Bassett, D. R., Jr. (2008). Accuracy of the Actiheart for the assessment of energy expenditure in adults. *European Journal of Clinical Nutrition, 62*(6), 704-711.

Crouter, S. E., Clowers, K. G., & Bassett, D. R., Jr. (2006). A novel method for using accelerometer data to predict energy expenditure. *Journal of Applied Physiology, 100*(4), 1324-1331.

de Graauw, S. M., de Groot, J. F., van Brussel, M., Streur, M. F., & Takken, T. (2010). Review of prediction models to estimate activity-related energy expenditure in children and adolescents. *Int J Pediatr, 2010*, 489304.

De Lorenzo, A., Bertini, I., Candeloro, N., Piccinelli, R., Innocente, I., & Brancati, A. (1999). A new predictive equation to calculate resting metabolic rate in athletes. *Journal of Sports Medicine and Physical Fi   ness, 39*(3), 213-219.

Dean, C. E. (2004). The modern warrior's combat load. Dismounted operations in Afhanistan. April-May 2003. Fort Levenworth, KS: Army Centre for lessons learned.

Di Rienzo, M., Meriggi, P., Rizzo, F., Vaini, E., Faini, A., Merati, G., et al. (2011). A wearable system for the seismocardiogram assessment in daily life conditions. *Conf Proc IEEE Eng Med Biol Soc, 2011*, 4263-4266.

Douglas, C. G. (1911). A method for determining respiratory exchange in man. *Journal of Physiology, 42 (suppl)*, 17-18.

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

Duffield, R., Dawson, B., Pinnington, H. C., & Wong, P. (2004). Accuracy and reliability of a Cosmed K4b2 portable gas analysis system. *Journal of Science and Medicine in Sport, 7*(1), 11-22.

Ekelund, U., Sjostrom, M., Yngve, A., Poortvliet, E., Nilsson, A., Froberg, K., et al. (2001). Physical activity assessed by activity monitor and doubly labeled water in children. *Medicine and Science in Sports and Exercise, 33*(2), 275-281.

Elia, M. (1992). Organ and tissue contribution to metabolic rate. In J. M. Kinney & H. N. Tucker (Eds.), *Energy Metabolism: Tissue Determinants and Cellular Corollaries* (pp. 61-79). New York: Raven Press.

Elia, M., Folmer, P., Schlatmann, A., Goren, A., & Austin, S. (1988). Carbohydrate, fat, and protein metabolism in muscle and in the whole body after mixed meal ingestion. *Metabolism, 37*(6), 542-551.

Esliger, D. W., Rowlands, A. V., Hurst, T. L., Catt, M., Murray, P., & Eston, R. G. (2010). Validation of the GENEA Accelerometer. *Medicine and Science in Sports and Exercise*.

Esliger, D. W., & Tremblay, M. S. (2006). Technical reliability assessment of three accelerometer models in a mechanical setup. *Medicine and Science in Sports and Exercise, 38*(12), 2173-2181.

Feito, Y., Bassett, D. R., Thompson, D. L., & Tyo, B. M. (2011). Effects of Body Mass Index on Step Count Accuracy of Physical Activity Monitors. *J Phys Act Health*.

Fick, A. (1870). Ueber die Messung des Blutquantums in den Hertzventrikeln. *Verh Phys Med Ges., 2*, 169-170.

Fortes, M. B., Diment, B. C., Greeves, J. P., Casey, A., Izard, R., & Walsh, N. P. (2011). Effects of a daily mixed nutritional supplement on physical performance, body composition, and circulating anabolic hormones during 8 weeks of arduous military training. *Appl Physiol Nutr Metab, 36*(6), 967-975.

Freedson, P., Bowles, H. R., Troiano, R., & Haskell, W. (2012). Assessment of physical activity using wearable monitors: recommendations for monitor calibration and use in the field. *Medicine and Science in Sports and Exercise, 44*(1 Suppl 1), S1-4.

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

Frykman. (1994). Backpack vs. front-back pack: Differential effects of fatigue on loaded walking posture. *Medicine and Science in Sports and Exercise, 26*(5), S140.

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

Gardner, A. W., & Poehlman, E. T. (1998). Assessment of free-living daily physical activity in older claudicants: validation against the doubly labeled water technique.

Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 53(4), M275-280.

Geer, E. B., & Shen, W. (2009). Gender differences in insulin resistance, body composition, and energy balance. *Gend Med, 6 Suppl 1*, 60-75.

Gladden, L. B., Yates, J. W., & Howley, E. T. (2012). Who needs a bag? *Medicine and Science in Sports and Exercise, 44*(2), 288-289.

Glass, S., & Dwyer, G. B. (2007). *ACSM'S metabolic calculations handbook*. Philadelphia: Lippincott Williams & Wilkins.

Guezennec, C. Y., Satabin, P., Legrand, H., & Bigard, A. X. (1994). Physical performance and metabolic changes induced by combined prolonged exercise and different energy intakes in humans. *Eur J Appl Physiol Occup Physiol, 68*(6), 525-530.

Hagan, R. D., Strathman, T., Strathman, L., & Gettman, L. R. (1980). Oxygen uptake and energy expenditure during horizontal treadmill running. *Journal of Applied Physiology, 49*(4), 571-575.

Hakim, A. A., Petrovitch, H., Burchfiel, C. M., Ross, G. W., Rodriguez, B. L., White, L. R., et al. (1998). Effects of walking on mortality among nonsmoking retired men. *N Engl J Med, 338*(2), 94-99.

Haldane, J. S. (1912). *Methods of Air Analysis*. London: Griffin.

Harwood, G. E., Rayson, M. P., & Nevill, A. M. (1999). Fitness, performance, and risk of injury in British Army officer cadets. *Military Medicine, 164*(6), 428-434.

Heil, D. P., Brage, S., & Rothney, M. P. (2012). Modeling physical activity outcomes from wearable monitors. *Medicine and Science in Sports and Exercise, 44*(1 Suppl 1), S50-60.

Hendelman, D., Miller, K., Baggett, C., Debold, E., & Freedson, P. (2000). Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Medicine and Science in Sports and Exercise, 32*(9 Suppl), S442-449.

Henning, P. C., Park, B. S., & Kim, J. S. (2011). Physiological decrements during sustained military operational stress. *Military Medicine, 176*(9), 991-997.

Hofsteenge, G. H., Chinapaw, M. J., Delemarre-van de Waal, H. A., & Weijs, P. J. (2010). Validation of predictive equations for resting energy expenditure in obese adolescents. *American Journal of Clinical Nutrition, 91*(5), 1244-1254.

Hopker, J. G., Jobson, S. A., Gregson, H. C., Coleman, D., & Passfield, L. (2012). Reliability of cycling gross efficiency using the Douglas bag method. *Medicine and Science in Sports and Exercise, 44*(2), 290-296.

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine, 30*(1), 1-15.

Intille, S. S., Lester, J., Sallis, J. F., & Duncan, G. (2012). New horizons in sensor development. *Medicine and Science in Sports and Exercise, 44*(1 Suppl 1), S24-31.

Jacobi, D., Perrin, A. E., Grosman, N., Dore, M. F., Normand, S., Oppert, J. M., et al. (2007). Physical activity-related energy expenditure with the RT3 and TriTrac accelerometers in overweight adults. *Obesity (Silver Spring), 15*(4), 950-956.

Jeukendrup, A. E., & Wallis, G. A. (2005). Measurement of substrate oxidation during exercise by means of gas exchange measurements. *International Journal of Sports Medicine, 26 Suppl 1*, S28-37.

John, D., & Freedson, P. (2012). ActiGraph and Actical Physical Activity Monitors: A Peek under the Hood. *Medicine and Science in Sports and Exercise, 44*(1 Suppl 1), S86-89.

Karantonis, D. M., Narayanan, M. R., Mathie, M., Lovell, N. H., & Celler, B. G. (2006). Implementation of a real-time human movement classifier using a triaxial

accelerometer for ambulatory monitoring. *IEEE Trans Inf Technol Biomed, 10*(1), 156-167.

Kavanagh, J. J., & Menz, H. B. (2008). Accelerometry: a technique for quantifying movement patterns during walking. *Gait Posture, 28*(1), 1-15.

Kleiber, M. (1947). Body size and metabolic rate. *Physiol Rev, 27*(4), 511-541.

Knapik, J., Harman, E., & Reynolds, K. (1996). Load carriage using packs: a review of physiological, biomechanical and medical aspects. *Appl Ergon, 27*(3), 207-216.

Knapik, J. J., Reynolds, K. L., & Harman, E. (2004). Soldier load carriage: historical, physiological, biomechanical, and medical aspects. *Military Medicine, 169*(1), 45-56.

Koebnick, C., Wagner, K., Thielecke, F., Moeseneder, J., Hoehne, A., Franke, A., et al. (2005). Validation of a simplified physical activity record by doubly labeled water technique. *Int J Obes (Lond), 29*(3), 302-309.

Krasnoff, J. B., Kohn, M. A., Choy, F. K., Doyle, J., Johansen, K., & Painter, P. L. (2008). Interunit and intraunit reliability of the RT3 triaxial accelerometer. *J Phys Act Health, 5*(4), 527-538.

Lamonte, M. J., & Ainsworth, B. E. (2001). Quantifying energy expenditure and physical activity in the context of dose response. *Medicine and Science in Sports and Exercise, 33*(6 Suppl), S370-378; discussion S419-320.

Light, L. H., McLellan, G. E., & Klenerman, L. (1980). Skeletal transients on heel strike in normal walking with different footwear. *Journal of Biomechanics, 13*(6), 477-480.

Livingstone, M. B. (1997). Heart-rate monitoring: the answer for assessing energy expenditure and physical activity in population studies? *British Journal of Nutrition, 78*(6), 869-871.

Long, X., Yin, B., & Aarts, R. M. (2009). Single-accelerometer-based daily physical activity classification. *Conf Proc IEEE Eng Med Biol Soc, 2009*, 6107-6110.

Ludbrook, J. (1997). Comparing methods of measurements. *Clinical and Experimental Pharmacology and Physiology, 24*(2), 193-203.

Luke, A., Maki, K. C., Barkey, N., Cooper, R., & McGee, D. (1997). Simultaneous monitoring of heart rate and motion to assess energy expenditure. *Medicine and Science in Sports and Exercise, 29*(1), 144-148.

Lyden, K., Kozey, S. L., Staudenmeyer, J. W., & Freedson, P. S. (2010). A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. *Eur J Appl Physiol*.

Lyden, K., Kozey, S. L., Staudenmeyer, J. W., & Freedson, P. S. (2011). A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. *European Journal of Applied Physiology and Occupational Physiology, 111*(2), 187-201.

Maddison, R., Jiang, Y., Hoorn, S. V., Mhurchu, C. N., Lawes, C. M., Rodgers, A., et al. (2009). Estimating energy expenditure with the RT3 triaxial accelerometer. *Research Quarterly for Exercise and Sport, 80*(2), 249-256.

Maddison, R., Ni Mhurchu, C., Jiang, Y., Vander Hoorn, S., Rodgers, A., Lawes, C. M., et al. (2007). International Physical Activity Questionnaire (IPAQ) and New Zealand Physical Activity Questionnaire (NZPAQ): a doubly labelled water validation. *Int J Behav Nutr Phys Act, 4*, 62.

Majumdar, D., & Pal, M. S. (2010). Effects of military load carriage on kinematics of gait. *Ergonomics, 53*(6), 782-791.

Marfell-Jones, M., Olds, T., Stewart, A., & Carter, L. (2006). International Standards for Anthropometric Assessment *International Society for the Advancement of Kinanthropometry (ISAK)*.

Masse, L. C., Fulton, J. E., Watson, K. L., Mahar, M. T., Meyers, M. C., & Wong, W. W. (2004). Influence of body composition on physical activity validation studies using doubly labeled water. *Journal of Applied Physiology, 96*(4), 1357-1364.

Mathie, M. J., Coster, A. C., Lovell, N. H., & Celler, B. G. (2004). Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. *Physiol Meas, 25*(2), R1-20.

McCaig, R. H., & Gooderson, C. Y. (1986). Ergonomic and physiological aspects of military operations in a cold wet climate. *Ergonomics, 29*(7), 849-857.

McKenzie, K. L. (2010). 2009 C. H. McCloy Lecture. Seeing is believing: observing physical activity and its contexts. *Res Q Exerc Sport, 81*(2), 113-122.

McLaughlin, J. E., King, G. A., Howley, E. T., Bassett, D. R., Jr., & Ainsworth, B. E. (2001). Validation of the COSMED K4 b2 portable metabolic system. *International Journal of Sports Medicine, 22*(4), 280-284.

Metcalf, B. S., Curnow, J. S., Evans, C., Voss, L. D., & Wilkin, T. J. (2002). Technical reliability of the CSA activity monitor: The EarlyBird Study. *Medicine and Science in Sports and Exercise, 34*(9), 1533-1537.

Meyer, T., Georg, T., Becker, C., & Kindermann, W. (2001). Reliability of gas exchange measurements from two different spiroergometry systems. *International Journal of Sports Medicine, 22*(8), 593-597.

Midorikawa, T., Tanaka, S., Kaneko, K., Koizumi, K., Ishikawa-Takata, K., Futami, J., et al. (2007). Evaluation of low-intensity physical activity by triaxial accelerometry. *Obesity (Silver Spring), 15*(12), 3031-3038.

Moeller, N. C., Korsholm, L., Kristensen, P. L., Andersen, L. B., Wedderkopp, N., & Froberg, K. (2008). Unit-specific calibration of Actigraph accelerometers in a mechanical setup - is it worth the effort? The effect on random output variation caused by technical inter-instrument variability in the laboratory and in the field. *BMC Med Res Methodol, 8*, 19.

Montoye, H. J., Washburn, R., Servais, S., Ertl, A., Webster, J. G., & Nagle, F. J. (1983). Estimation of energy expenditure by a portable accelerometer. *Medicine and Science in Sports and Exercise, 15*(5), 403-407.

Myers, S. D., Leamon, S. M., Nevola, V. R., & Llewellyn, M. G. (2008). Energy expenditure during a single-handed transatlantic yacht race. *British Journal of Sports Medicine, 42*(4), 285-288.

Nevill, A. M., Bate, S., & Holder, R. L. (2005). Modeling physiological and anthropometric variables known to vary with body size and other confounding variables. *American Journal of Physical Anthropology, Suppl 41*, 141-153.

Nevill, A. M., & Holder, R. L. (1995a). Body mass index: a measure of fatness or leanness? *British Journal of Nutrition, 73*(4), 507-516.

Nevill, A. M., & Holder, R. L. (1995b). Scaling, normalizing, and per ratio standards: an allometric modeling approach. *Journal of Applied Physiology, 79*(3), 1027-1031.

Nindl, B. C., Barnes, B. R., Alemany, J. A., Frykman, P. N., Shippee, R. L., & Friedl, K. E. (2007). Physiological consequences of U.S. Army Ranger training. *Medicine and Science in Sports and Exercise, 39*(8), 1380-1387.

Pandolf, K. B., Givoni, B., & Goldman, R. F. (1977). Predicting energy expenditure with loads while standing or walking very slowly. *Journal of Applied Physiology, 43*(4), 577-581.

Pate, R. R., Pratt, M., Blair, S. N., Haskell, W. L., Macera, C. A., Bouchard, C., et al. (1995). Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *JAMA, 273*(5), 402-407.

Patton, J. F., Kaszuba, J., Mello, R. P., & Reynolds, K. L. (1991). Physiological responses to prolonged treadmill walking with external loads. *Eur J Appl Physiol Occup Physiol, 63*(2), 89-93.

Plasqui, G., Joosen, A. M., Kester, A. D., Goris, A. H., & Westerterp, K. R. (2005). Measuring free-living energy expenditure and physical activity with triaxial accelerometry. *Obesity Research, 13*(8), 1363-1369.

Plasqui, G., & Westerterp, K. R. (2007). Physical activity assessment with accelerometers: an evaluation against doubly labeled water. *Obesity (Silver Spring), 15*(10), 2371-2379.

Pober, D. M., Staudenmayer, J., Raphael, C., & Freedson, P. S. (2006). Development of novel techniques to classify physical activity mode using accelerometers. *Medicine and Science in Sports and Exercise, 38*(9), 1626-1634.

Powell, S. M., Jones, D. I., & Rowlands, A. V. (2003). Technical variability of the RT3 accelerometer. *Medicine and Science in Sports and Exercise, 35*(10), 1773-1778.

Powell, S. M., & Rowlands, A. V. (2004). Intermonitor variability of the RT3 accelerometer during typical physical activities. *Medicine and Science in Sports and Exercise, 36*(2), 324-330.

Rajkovic, M. G., Rumora, L., & Barisic, K. (2011). The paraoxonase 1, 2 and 3 in humans. *Biochem Med (Zagreb), 21*(2), 122-130.

Ramsbottom, R., Brewer, J., & Williams, C. (1988). A progressive shuttle run test to estimate maximal oxygen uptake. *British Journal of Sports Medicine, 22*(4), 141-144.

Rayson, M., Holliman, D., & Belyavin, A. (2000). Development of physical selection procedures for the British Army. Phase 2: relationship between physical performance tests and criterion tasks. *Ergonomics, 43*(1), 73-105.

Rennie, K. L., Hennings, S. J., Mitchell, J., & Wareham, N. J. (2001). Estimating energy expenditure by heart-rate monitoring without individual calibration. *Medicine and Science in Sports and Exercise, 33*(6), 939-945.

Richmond, V., Wilkinson, D., Carter, J., Blacker, S., & M, R. (2007). *Validity of the 3DNX^{TM} accelerometer for predicting oxygen uptake.* Paper presented at the International conference on contemporary ergonomics.

Richmond, V. L., Carter, J. M., Wilkinson, D. M., Homer, F. E., Rayson, M. P., Wright, A., et al. (2012). Comparison of the physical demands of single-sex training for male and female recruits in the British Army. *Military Medicine, 177*(6), 709-715.

Richmond, V. L., Horner, F. E., Wilkinson, D. M., Rayson, M. P., Wright, A., Cook, A. C., et al. (2010). Energy balance during the British Army Section Commanders' Battle Course. *Medicine & Science in Sports & Exercise, 42*(5), 287.

Rothney, M. P., Apker, G. A., Song, Y., & Chen, K. Y. (2008). Comparing the performance of three generations of ActiGraph accelerometers. *Journal of Applied Physiology, 105*(4), 1091-1097.

Rothney, M. P., Brychta, R. J., Meade, N. N., Chen, K. Y., & Buchowski, M. S. (2010). Validation of the ActiGraph two-regression model for predicting energy expenditure. *Medicine and Science in Sports and Exercise, 42*(9), 1785-1792.

Rothney, M. P., Neumann, M., Beziat, A., & Chen, K. Y. (2007). An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *Journal of Applied Physiology, 103*(4), 1419-1427.

Ruby, B. C., Shriver, T. C., Zderic, T. W., Sharkey, B. J., Burks, C., & Tysk, S. (2002). Total energy expenditure during arduous wildfire suppression. *Medicine and Science in Sports and Exercise, 34*(6), 1048-1054.

Rumo, M., Amft, O., Troster, G., & Mader, U. (2011). A stepwise validation of a wearable system for estimating energy expenditure in field-based research. *Physiol Meas, 32*(12), 1983-2001.

Saito, M., Nakajima, K., Takano, C., Ohta, Y., Sugimoto, C., Ezoe, R., et al. (2011). An in-shoe device to measure plantar pressure during daily human activity. *Medical Engineering and Physics, 33*(5), 638-645.

Sarkin, J. A., Nichols, J. F., Sallis, J. F., & Calfas, K. J. (2000). Self-report measures and scoring protocols affect prevalence estimates of meeting physical activity guidelines. *Medicine and Science in Sports and Exercise, 32*(1), 149-156.

Sasaki, J. E., John, D., & Freedson, P. S. (2011). Validation and comparison of ActiGraph activity monitors. *Journal of Science and Medicine in Sport, 14*(5), 411-416.

Schneider, P. L., Crouter, S. E., & Bassett, D. R. (2004). Pedometer measures of free-living physical activity: comparison of 13 models. *Medicine and Science in Sports and Exercise, 36*(2), 331-335.

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

Schofield, W. N. (1985). Predicting basal metabolic rate, new standards and review of previous work. *Human Nutrition. Clinical Nutrition, 39 Suppl 1*, 5-41.

Schutz, Y., Weinsier, S., Terrier, P., & Durrer, D. (2002). A new accelerometric method to assess the daily walking practice. *International Journal of Obesity and Related Metabolic Disorders, 26*(1), 111-118.

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

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

Speakman, J. (1997). *Doubly -Labelled Water: Theory and Practice*: Chapman and Hall, London.

Speakman, J. R. (1998). The history and theory of the doubly labeled water technique. *American Journal of Clinical Nutrition, 68*(4), 932S-938S.

Speakman, J. R. (2005). The role of technology in the past and future development of the doubly labelled water method. *Isotopes Environ Health Stud, 41*(4), 335-343.

Spierer, D. K., Hagins, M., Rundle, A., & Pappas, E. (2011). A comparison of energy expenditure estimates from the Actiheart and Actical physical activity monitors during low intensity activities, walking, and jogging. *European Journal of Applied Physiology and Occupational Physiology, 111*(4), 659-667.

Sprager, S., & Zazula, D. (2012). Heartbeat and respiration detection from optical interferometric signals by using a multimethod approach. *IEEE Transactions on Biomedical Engineering, 59*(10), 2922-2929.

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

St-Onge, M., Mignault, D., Allison, D. B., & Rabasa-Lhoret, R. (2007). Evaluation of a portable device to measure daily energy expenditure in free-living adults. *American Journal of Clinical Nutrition, 85*(3), 742-749.

Staudenmayer, J., Pober, D., Crouter, S., Bassett, D., & Freedson, P. (2009). An artificial neural network to estimate physical activity energy expenditure and identify

physical activity type from an accelerometer. *Journal of Applied Physiology, 107*(4), 1300-1307.

Staudenmayer, J., Zhu, W., & Catellier, D. J. (2012). Statistical considerations in the analysis of accelerometry-based activity monitor data. *Medicine and Science in Sports and Exercise, 44*(1 Suppl 1), S61-67.

Su, S. W., Celler, B. G., Savkin, A. V., Nguyen, H. T., Cheng, T. M., Guo, Y., et al. (2009). Transient and steady state estimation of human oxygen uptake based on noninvasive portable sensor measurements. *Medical and Biological Engineering and Computing, 47*(10), 1111-1117.

Sun, M., & Hill, J. O. (1993). A method for measuring mechanical work and work efficiency during human activities. *Journal of Biomechanics, 26*(3), 229-241.

Tharion, W. J., Lieberman, H. R., Montain, S. J., Young, A. J., Baker-Fulco, C. J., Delany, J. P., et al. (2005). Energy requirements of military personnel. *Appetite, 44*(1), 47-65.

Thompson, D., Batterham, A. M., Bock, S., Robson, C., & Stokes, K. (2006). Assessment of low-to-moderate intensity physical activity thermogenesis in young adults using synchronized heart rate and accelerometry with branched-equation modeling. *Journal of Nutrition, 136*(4), 1037-1042.

Thompson, J., & Manore, M. M. (1996). Predicted and measured resting metabolic rate of male and female endurance athletes. *Journal of the American Dietetic Association, 96*(1), 30-34.

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

Trost, S. G., McIver, K. L., & Pate, R. R. (2005). Conducting accelerometer-based activity assessments in field-based research. *Medicine and Science in Sports and Exercise, 37*(11 Suppl), S531-543.

Trost, S. G., Pate, R. R., Freedson, P. S., Sallis, J. F., & Taylor, W. C. (2000). Using objective physical activity measures with youth: how many days of monitoring are needed? *Medicine and Science in Sports and Exercise, 32*(2), 426-431.

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

Trost, S. G., Wong, W. K., Pfeiffer, K. A., & Zheng, Y. (2012). Artificial Neural Networks to Predict Activity Type and Energy Expenditure in Youth. *Medicine and Science in Sports and Exercise*.

Tudor-Locke, C., & Lutes, L. (2009). Why do pedometers work?: a reflection upon the factors related to successfully increasing physical activity. *Sports Medicine, 39*(12), 981-993.

van der Ploeg, G. E., Gunn, S. M., Withers, R. T., Modra, A. C., Keeves, J. P., & Chatterton, B. E. (2001). Predicting the resting metabolic rate of young Australian males. *European Journal of Clinical Nutrition, 55*(3), 145-152.

Vanhelst, J., Beghin, L., Duhamel, A., Bergman, P., Sjotrom, M., & Gottrand, F. (2012). Comparison of uniaxial and triaxial accelerometry in the assessment of physical activity among adolescents under free-living conditions: the HELENA study. *BMC Med Res Methodol, 12*(1), 26.

Vanhelst, J., Beghin, L., Rasoamanana, P., Theunynck, D., Meskini, T., Iliescu, C., et al. (2010). Calibration of the RT3 accelerometer for various patterns of physical activity in children and adolescents. *Journal of Sports Sciences, 28*(4), 381-387.

Voleno, M., Redmond, S. J., Cerutti, S., & Lovell, N. H. (2010). Energy expenditure estimation using triaxial accelerometry and barometric pressure measurement. *Conf Proc IEEE Eng Med Biol Soc, 2010*, 5185-5188.

Wang, N., Ambikairajah, E., Redmond, S. J., Celler, B. G., & Lovell, N. H. (2009). Classification of walking patterns on inclined surfaces from accelerometry data. *Digital Signal Processing, 2009 16th International Conference on*, 1-4.

Wang, N., Redmond, S. J., Ambikairajah, E., Celler, B. G., & Lovell, N. H. (2010). Can triaxial accelerometry accurately recognize inclined walking terrains? *IEEE Transactions on Biomedical Engineering, 57*(10), 2506-2516.

Washburn, R. A., & LaPorte, R. E. (1988). Assessment of walking behavior: effect of speed and monitor position on two objective physical activity monitors. *Res Q Exerc Sport, 59*(1), 83-85.

Welk, G. J. (2005). Principles of design and analyses for the calibration of accelerometry-based activity monitors. *Medicine and Science in Sports and Exercise, 37*(11 Suppl), S501-511.

Welk, G. J., Blair, S. N., Wood, K., Jones, S., & Thompson, R. W. (2000). A comparative evaluation of three accelerometry-based physical activity monitors. *Medicine and Science in Sports and Exercise, 32*(9 Suppl), S489-497.

Welk, G. J., McClain, J., & Ainsworth, B. E. (2012). Protocols for evaluating equivalency of accelerometry-based activity monitors. *Medicine and Science in Sports and Exercise, 44*(1 Suppl 1), S39-49.

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

Westerterp, K., & Schols, A. M. (2008). Basics in clinical nutrition: Energy metabolism. *The European e-journal of Clinical Nutrition and Metabolism, 3*, e281-e284.

Westerterp, K. R. (1999). Body composition, water turnover and energy turnover assessment with labelled water. *Proc Nutr Soc, 58*(4), 945-951.

Westerterp, K. R. (2009). Assessment of physical activity: a critical appraisal. *European Journal of Applied Physiology and Occupational Physiology, 105*(6), 823-828.

Wilkinson, D. M., Blacker, S. D., Richmond, V. L., Horner, F. E., Rayson, M. P., Spiess, A., et al. (2011). Injuries and injury risk factors among British army infantry soldiers during predeployment training. *Inj Prev, 17*(6), 381-387.

Wilkinson, D. M., Carter, J. M., Richmond, V. L., Blacker, S. D., & Rayson, M. P. (2008a). The effect of cool water ingestion on gastrointestinal pill temperature. *Medicine and Science in Sports and Exercise, 40*(3), 523-528.

Wilkinson, D. M., Rayson, M. P., & Bilzon, J. L. (2008b). A physical demands analysis of the 24-week British Army Parachute Regiment recruit training syllabus. *Ergonomics, 51*(5), 649-662.

Wong, T. C., Webster, J. G., Montoye, H. J., & Washburn, R. (1981). Portable accelerometer device for measuring human energy expenditure. *IEEE Transactions on Biomedical Engineering, 28*(6), 467-471.

Wyss, T., & Mader, U. (2010). Recognition of military-specific physical activities with body-fixed sensors. *Military Medicine, 175*(11), 858-864.

Wyss, T., & Mader, U. (2011). Energy expenditure estimation during daily military routine with body-fixed sensors. *Military Medicine, 176*(5), 494-499.

Yngve, A., Nilsson, A., Sjostrom, M., & Ekelund, U. (2003). Effect of monitor placement and of activity setting on the MTI accelerometer output. *Medicine and Science in Sports and Exercise, 35*(2), 320-326.

Zhang, K., Pi-Sunyer, F. X., & Boozer, C. N. (2004). Improving energy expenditure estimation for physical activity. *Medicine and Science in Sports and Exercise, 36*(5), 883-889.

Zhang, S., Rowlands, A. V., Murray, P., & Hurst, T. L. (2012). Physical activity classification using the GENEA wrist-worn accelerometer. *Medicine and Science in Sports and Exercise, 44*(4), 742-748.

Zijlstra, W., Hof, AL. (1997). Displacement of the pelvis during human walking: experimental data and model predictions. *Gait and Posture, 6*, 249-262.

13   APPENDIX A

13.1   3DNX PRODUCT SPECIFICATIONS

Manufacturer   : Biotel Ltd. Bristol, UK
Measurement range   : full scale range +/- 18 *g*
Frequency response   : 0.2 – 10 Hz
Power supply   : 3.6 Volt
Current consumption : 2 mA
Battery model   : 3.6 V AA Lithium (Saft (128-2764 or 128-2765))
Battery lifetime   : 6-7 weeks (60 second sampling interval)
Memory size   : 8 MB (ca. 705 days storage time / 60 seconds sampling interval)
Communication   : RS232, 921k600, 8 bits, 1 stopbit, No Parity
Dimensions   : 58 x 46 x 17 mm
Weight   : 75 g (Including battery)

Part	# of Axes	Range (*g*)	Sensitivity (mV/g)	Sensitivity Accuracy (%)	Output	Band Width (kHz)	Noise Density (μg/rtHz)	Voltage Supply (V)	Supply Current (mA)
ADXL321	2	±18	57	±10	Analog	2.5	320	2.4 to 6	0.5

Table 13.1 Technical specifications of the accelerometer used by the 3DNX

13.1.1 Dynamic Range

The ADXL321 has a measurement range of +/- 18 *g*. The electronics do not use the full range. The sensitivity of the sensor is 57 mVolt/*g*.

For all positive *g* values 3DNX has a range of 1.25 V, amplified four times between the sensor and the AD converter. The 3DNX has an actual measurement range of 1.25 / (4 x 0.057) = +5.5 *g*.

For negative *g* values the 3DNX has the same range, which results in a full scale range of ± 5.5 *g*.