

**The relationship between psychological well-being and physical activity:
The impact of measurement**

Submitted by **Lisa Rachel Stephanie Phillips** to the University of Exeter
As a thesis for the degree of
Doctor of Philosophy in Sport and Health Sciences
In May 2013

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature:

Abstract

Psychological well-being occurs when there is an absence of mental disorders and presence of positive states. Given the increasing prevalence of mental disorders, which are thought to have their roots in childhood, improving psychological well-being in children is currently an important area of research. Physical activity has been proposed as a method by which negative states can be reduced and positive states increased, thereby increasing children's overall psychological well-being and in turn helping to protect against a decline into clinical disorders. Research focusing on physical activity and psychological well-being has mainly used self-reported measurements to assess physical activity, a method which leads to considerable non-differential misclassification that in turn will attenuate associations between physical activity and psychological wellbeing. Few studies have employed more precise, objective measures such as accelerometry. Despite providing a more precise measure of physical activity, a number of limitations are present with the use of accelerometry, specifically concerning the data reduction processes. Various decisions made when handling accelerometer data can result in misclassification of time spent in different intensities of physical activity and can introduce selection bias. The present thesis aims to address how the decisions made during data reduction can affect estimates of physical activity prevalence and alter the observed relationships between physical activity and psychological well-being in children.

The first study of this thesis assessed the misclassification of activity intensities occurring as a result of the use of various accelerometer cut-points and the resulting variation in relationships between physical activity and psychological well-being that occurs. Results showed that the use of different cut-points to determine physical activity intensity alters the magnitude of the relationship between physical activity and psychological well-being; relationships were attenuated, with some becoming non-significant. The second study addressed the issue of children's compliance with wear time requirements over multiple time points; compliance with wear time decreased over time, whilst inclusion and exclusion rules based on minimum wear times introduced selection bias. The use of more lenient wear time criteria, to reduce selection bias, introduced misclassification of physical activity intensities. Further,

longitudinal relationships between physical activity and psychological well-being differed depending upon the wear time criteria employed. The third study aimed to address whether compliance, and in turn selection bias would systematically differ between groups of a trial of a physical activity intervention, and whether this would alter the results of the intervention itself. Results showed that compliance varied across trial condition, that selection bias with groups was different for each condition and that non-compliance hindered the exploration of the mediating effect of physical activity on psychological well-being. Study four involved the validation and calibration of a new wrist worn, waterproof physical activity monitor more compatible with 24 hour wear, thus potentially overcoming the compliance problems noted in the earlier studies. Results showed good concurrent and criterion validity, with high classification accuracy for the cut-points created. The final study assessed the acceptability and compliance with 24 hour wear in children and allowed a detailed examination of the underestimation of time spent in PA intensities that occurs from capturing shorter and different periods of the day. Results showed large misclassification with 10 hour capture periods relative to complete observation, with time in activity intensities varying across different periods of the day.

The results of this thesis demonstrate that substantial selection bias and misclassification of time in activity intensities can be introduced through the decisions made during the processing of raw accelerometry data. Furthermore, this error alters the relationships between physical activity and psychological well-being. The results indicate that the true relationship between physical activity and psychological well-being in children may still be unknown, with researchers reporting relationships and effects only relevant to the measurement methods and data reduction processes they have employed. A method of overcoming selection bias and reducing misclassification is through 24 hour wear, which through the design of new accelerometers is now possible. Future studies should use monitors compatible with and acceptable for complete observation. This would result in more precise estimates of time spent in physical activity intensities and less selection bias. Both of these improvements would greatly increase our understanding of the relationship between physical activity and psychological well-being in children.

Acknowledgments

Firstly, I would like to thank my supervisors, Associate Professor Melvyn Hillsdon and Dr. Alex Rowlands for their ever present support and their eternal patience during the completion of this thesis. Their continued guidance and advice throughout this process have been greatly appreciated. Thank you both, for your dedication to this thesis. Special thanks to Melvyn, who joined the supervisory team half way my doctoral programme, and since, has shared knowledge and expertise and provided an insurmountable amount of support and patience, even with the most simple of questions.

I would also like to thank Gaynor Parfitt for her supervision and encouragement throughout the early stages of my research. I am grateful to David Childs and Joss Langford for their technical support, and to Brad Metcalf, who acted as a sounding board, provided advice and played the role of devil's advocate well during discussions in the final stages of this thesis. I would also like to convey my appreciation to my ever understanding employers, Dr. Jennifer Lloyd and Associate Professor Katrina Wyatt, who afforded me great flexibility in working hours during the final writing period, accompanied by huge encouragement.

My PhD research could not have taken place without the cooperation and enthusiasm of the children who kindly participated in the studies. Alongside the participants, I would like to express my gratitude to the schools and parents, without whom, I would be unable to undertake my research. I am immensely grateful to all involved! In relation to this, I thank Dr Charlotte Hamlyn-Williams, Johanna Hänggi, Anastasia Shvedko, Lorraine McNulty, Caitlin Vulgattio and Rebecca Calver, to whom I am indebted, for their help and hard work during the data collection process.

Thanks to Charlotte, Lou, Izzy and Jodi for being a sounding board for ideas, whilst providing support, encouragement and a calming influence when I needed it most. I also would like to thank Andy Price for putting up with me even through the most stressful periods; for remaining positive, keeping me sane and providing much needed laughter, relaxation and encouragement, accompanied by lots of food and cups of tea!

Finally, I would like to thank my parents for their unwavering support, not only during the last four years, but throughout my entire education. Thank you for always believing in me, you've made this all possible! Special acknowledgement is given to my dad, Stephen Phillips, who sadly passed away prior to the completion of this thesis; your wisdom and guidance are with me every day.

Dedication

I would like to dedicate this thesis to my parents, Sandra and Steve Phillips,
who have provided unconditional love and support always.

I hope I've made you proud!

List of contents

	Page
Abstract	2
Acknowledgments	4
Dedication	6
List of Tables	11
List of Figures	14
Definitions and abbreviations	15
Chapter 1: Defining children’s psychological well-being	16
1.1 Conceptualising psychological well-being.....	16
1.2 Mental disorders.....	16
1.3 Positive states.....	19
1.4 Increasing psychological well-being	22
1.5 Summary.....	24
Chapter 2: A review of the psychological well-being and physical activity literature	27
2.1. Physical activity and psychological well-being.....	27
2.2. Physical activity and psychological well-being in children and adolescents	29
2.3 Proposed direction of relationships.....	43
2.4 Longitudinal research	43
2.5 Physical activity intervention trials in children and adolescents	49
2.6 Meta-analyses.....	55
2.7 Dose – response	58
2.8 Discussion of the PA and PWB literature.....	59
Chapter 3: Methodology	63
3.1 Psychological well-being measurement	63
3.1.1 Self-report.....	65
3.1.2 Interviews.....	67
3.1.3 Choice of methods	68
3.2 Physical activity	69
3.2.1 Definition and classification	69
3.3 Measurement of physical activity	69
3.3.1 The importance of accurate physical activity measurement.....	73
3.3.2 Measures of energy expenditure.....	75
3.3.3 Measures of behaviour	78
3.3.4 Measures of physical activity via movement sensors.....	83
3.4 Sources of misclassification in the association between PA and PWB	94
3.4.1 Misclassification through random and systematic error.....	95
3.4.2 Misclassification through accelerometer data...	98
3.4.3 Cut-point misclassification	101
3.4.4 Selection bias and loss to follow up bias in PA	104
3.5 Choice of methods	106

3.5.1	Validity, feasibility and reliability	106
3.6	Method choice summary.....	111
3.7	Aim of thesis.....	111
3.7.1	Hypothesis.....	112
Chapter 4:	Associations between children’s psychological well-being and physical activity intensity by accelerometer defined intensity levels	113
4.1	Introduction.....	113
4.2	Method	119
4.2.1	Participants	119
4.2.2	Protocol.....	119
4.2.3	Psychological well-being	120
4.2.4	Data reduction.....	123
4.2.5	Data analysis	124
4.3	Results.....	125
4.3.1	Descriptives.....	125
4.3.2	ANOVAs	125
4.3.3	Correlations	127
4.4	Discussion	131
4.4.1	Time in activity intensities	131
4.4.2	PA and PWB relationships	132
4.4.3	Limitations and future directions	138
4.5	Conclusion	139
Chapter 5:	Children’s compliance with wear time criteria across multiple time points	140
5.1	Introduction	140
5.1.1	Wear time compliance	140
5.2.	Method	144
5.2.1	Protocol	144
5.2.2	Participants	144
5.2.3	Physical activity	144
5.2.4	Psychological well-being	145
5.2.4	Data reduction and analysis.....	145
5.3	Results	146
5.3.1	Compliance	146
5.3.2	Selection bias arising from compliance and non-compliance	148
5.3.3	Classification of time in activity intensities depending upon wear time	151
5.5.4	Cross sectional and longitudinal associations by wear time criteria	153
5.4.	Discussion	157
5.4.1	Compliance	157
5.4.2	Wear time compliance and selection bias	159
5.4.3	Wear time compliance and misclassification ...	161
5.4.4	Wear time criteria and relationships between PA and PWB	162
5.5	Conclusion	164

Chapter 6: Wear time compliance across a physical activity intervention	166
6.1 Introduction.....	166
6.1.1 Introduction.....	166
6.1.2 Intervention rationale	167
6.2. Method	169
6.2.1 Participants	169
6.2.2 Protocol	170
6.2.3 Measures	171
6.2.4 Data reduction and analysis	171
6.3 Results	172
6.3.1 Loss to follow up	172
6.3.2 Effects of the intervention	173
6.3.3 Wear time compliance	173
6.3.4 Selection bias	177
6.3.5 Misclassification of time in activity intensities ..	182
6.3.6 Intervention effects upon PA	185
6.4 Discussion.....	185
6.4.1 Intervention effects	186
6.4.2 Wear time compliance	187
6.4.3 Limitations	193
6.5. Conclusion	194
Chapter 7: Validation and calibration of the GENE accelerometer for assessment of children’s physical activity	195
7.1 Introduction	195
7.2 Methods	196
7.2.1 Participants	196
7.2.2 Protocol	197
7.2.3 Data analysis	198
7.3. Results	200
7.3.1 Descriptives	200
7.3.2 ANOVAs.....	202
7.3.3 Validity	202
7.3.4 Regression analysis	203
7.3.5 Receiver operator characteristic curve analysis.....	203
7.5 Discussion	204
7.6 Conclusion	207
Chapter 8: 24 hour physical activity monitoring: An examination of misclassification and the impact upon relationships with psychological well-being	208
8.1 Introduction	208
8.1.1 Bias and misclassification	209
8.2 Method	211
8.2.1 Participants	211
8.2.2 Protocol	211
8.2.3 Measures	211
8.2.4 Physical activity measurement	213
8.2.5 Data analysis	213
8.3 Results	216
8.3.1 Compliance	216

8.3.2	Time in activity intensities	219
8.3.3	Relationships between PA and PWB	221
8.4	Discussion	223
8.4.1	Compliance	223
8.4.2	Time in Physical activity intensities	225
8.4.3	Physical activity and psychological well-being	228
8.4.4	Limitation and strengths	229
8.5	Conclusion	229
Chapter 9:	General discussion	231
9.1	The effect of misclassification due to accelerometer cut-points	232
9.2	The effect of selection bias and misclassification as a result of wear time requirements	234
9.3	Validation and calibration of a wrist-worn accelerometer	238
9.4	Misclassification through shorter capture periods	240
9.5	Overall summary.....	242
9.6	Strengths and limitations	243
9.7	Future directions	247
9.8	Practical implications.....	251
9.9	Conclusion	253
	References	255
	Appendix	
1.	Ethics certificates	294
2.	Informed consent	297
3.	Questionnaires	322
4.	Further analysis	331
5.	Gantt chart	336
6.	Screen view for GENEActiv Macro.....	337

List of tables

		Page
3.1	Adult and child Activity MET values from Ridley and Olds (2008).	72
3.2	Classification of physical activity intensities by relative energy expenditure (adapted from ACSM, 2006).	73
3.3	Summary of physical activity assessment methods.	109
4.1	Published cut-points for activity intensities established with ActiGraph models 7164 and GT1M.	124
4.2	Means \pm SD for psychological well-being variables.	125
4.3	Zero-order and partial relationships between cut-point defined sedentary behaviour and psychological well-being variables.	127
4.4	Zero-order and partial correlations between cut-point defined light activity and psychological well-being variables.	128
4.5	Zero-order and partial correlations between cut-point defined moderate activity and psychological well-being variables.	129
4.6	Zero-order and partial correlations for cut-point defined vigorous activity and psychological well-being variables.	130
4.7	Zero-order and partial correlations for cut-point defined MVPA and psychological well-being variables.	130
5.1.	Percentage of children achieving different wear times for varying combinations of days and hours per day.	147
5.2.	Number of participants complying with wear time criteria from total N with percentages shown in brackets, across multiple time points.	148
5.3	Means (\pm SD) of anthropometric and PWB measures for compliers and non-compliers with two minimum wear time criteria at baseline.	149
5.4	Means (\pm SD) of anthropometric and PWB measures for compliers and non-compliers with two minimum wear time criteria at mid-point.	150
5.5.	Means (\pm SD) of anthropometric and PWB measures for compliers and non-compliers with two minimum wear time criteria at end-point.	151

5.6.	Mean \pm SD time (minutes) accumulated in each activity intensity across three time points for 10 hours minimum wear for 3 weekdays and 1 weekend day criteria.	152
5.7.	Mean \pm SD time accumulated in each activity intensity across three time points for a minimum 8 hours wear for 3 weekdays and 1 weekend day criteria.	152
5.8	Correlations between baseline PA and follow up measures of PWB for different wear criteria.	155
5.9	Regression analysis for baseline PA and PSW for 10 hour and 8 hour wear time criteria.	156
6.1	Number of children with complete PWB questionnaires (n) in the total sample for each condition across time points (N).	173
6.2	Number of accelerometer files for each intervention condition and each time point, with number (n) and percentage (%) of participants complying and not complying with 10 hour wear time criteria across time points.	175
6.3	Number of accelerometer files for each intervention condition and each time point, with number (n) and percentage (%) of participants complying and not complying with 8 hour wear time criteria across time points.	176
6.4	Means \pm SD for compliers and non-compliers with two wear time criteria for intervention criteria at baseline.	179
6.5	Mean (\pm SD) time accumulated in different activity intensities for compliers with different wear time criteria by condition at each time point.	174
7.1	Means and standard deviations for ActiGraph (counts-second), GENE A (g-second) output and METs.	201
7.2	Criterion and concurrent validity of accelerometer output for each group.	203
7.3	Sensitivity, specificity and area under the curve and resultant cut-points for each GENE A monitor.	204
8.1	Thresholds for classification of normal / abnormal SDQ values (adapted from Goodman, 1998).	212
8.2	Number and percentage of children complying with different wear time criteria at each day, along with cumulative percentages for the inclusion of more lenient 'day' criteria.	218

8.3	Minutes difference in time accumulated in activity intensities relative to 24-hour monitoring.	220
8.4	Means and standard deviations and the percentage of children classed as normal, borderline and abnormal for SDQ subscales of participants achieving full wear time (N = 78).	221
8.5	Correlation matrix for 1 second epochs for 10 hour PA capture periods and 24 hours for 7 days with PWB variables (N =78).	222

List of Figures

		Page
1.1	Multidimensional hierarchy of self-esteem.	21
3.1.	Trade-off between feasibility and validity of physical activity monitors, modified from Esliger and Tremblay (2007).	107
4.1	Significant differences between times spent in each activity intensity when classified by different cut-points.	126
5.1	Compliance with different wear-time criteria across measurement points.	148
8.1	The number of children recruited, those eligible to be assessed for compliance and subsequently the number of children achieving 24 hour wear for 7 days and ≥ 10 hours for 3week and 1 weekend day.	216
8.2	Mean minutes per day spent in activity intensities for different periods of wear time in participants achieving complete observation.	220

Definitions

Physical activity:

Throughout this thesis physical activity is defined as '*any bodily movement produced by skeletal muscle that results in energy expenditure above resting level*' (Caspersen et al., 1985, p. 128).

Significant / significance:

The term significant is used throughout this thesis to indicate statistical significance, determined as $p < 0.05$.

Abbreviations

PA	Physical activity
MVPA	Moderate to vigorous physical activity
LTPA	Leisure time physical activity
PWB	Psychological well-being
GSW	Global self-worth
PSW	Physical self-worth
BMI	Body mass index
MET	Metabolic equivalent term

CHAPTER ONE

Children's psychological well-being

The following chapter introduces the concept of psychological well-being. The prevalence of common mental disorders in children will be discussed, along with the importance of including positive states when addressing well-being. Finally, the association between physical activity and psychological well-being will be considered.

1.1. Conceptualising Psychological well-being

The World Health Organisation (WHO) has defined psychological well-being (PWB) or mental health as the

“state in which an individual realizes his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community” (WHO, 2001).

This definition highlights the importance of considering not only the absence of mental disorders but also the presence of positive state (Massé et al., 1998). It can be interpreted as encompassing both positive and negative affective feelings alongside personal cognitions about life satisfaction and human potential that many researchers believe are needed in order to achieve psychological well-being (e.g. Shah & Mark, 2004).

1.2. Mental disorders

In order to achieve PWB, there should be an absence of mental disorders. In 2001, it was estimated that around 450 million people worldwide suffered from mental health disorders (WHO, 2001), which can be categorised into three main groups; neurosis, psychoses and personality disorders (Cooper & Bebbington, 2006). Neurotic disorders include some of the more commonly experienced conditions such as depression, anxiety, obsessive compulsive disorders, phobias and panic attacks. Disorders of this nature are characterised by marked emotional distress and are thought to interfere with daily functioning (Deverill &

King, 2009). A survey of psychiatric morbidity among adults in England showed that one in six adults (around 15%) suffered from a common mental disorder in 2007 (McManus, Meltzer, Brugha, Bebbington & Jenkins, 2009) with mixed anxiety and depression proving to be the most prevalent, followed by generalised anxiety disorder and depressive episodes. Higher prevalence in females was reported for all of these common disorders.

Mental disorders occur in children and adolescents, with around 20% of children experiencing a mental health problem within any given year (Mental Health Foundation, 2005) and approximately 10% of children aged 5 – 16 years suffering at any given time (Green, McGinnity, Meltzer, Ford, Goodman, 2005). A 2004 report stated approximately 4% of children suffered from neurotic disorders (e.g. anxiety and depression; Maughan, Brock & Ladva, 2004), though other common disorders included conduct and hyperkinetic disorders such as oppositional defiant disorder (Sutton-Hamilton & Armando, 2008) and Attention Deficit Hyperactivity Disorder (ADHD) (Taylor et al., 2004). Interestingly, unlike the adult population, males, rather than females were more likely to develop mental disorders (10% of males, 6% of females aged 5 – 10 years and 13% of males and 10% of females aged 11 – 15 years). Though appearing to conflict with adult sex prevalence, the child statistics are influenced by the types of disorders being measured. Males appear to suffer more from conduct and hyperkinetic disorders, whereas females report more neurotic disorders. A follow up report in 2008 showed that around 30 per cent of the children reporting neurotic disorders in 2005 had persistent disorders at follow up. Of those children not reporting emotional disorders in 2005, 3% had gone on to develop some type of emotional disorder in 2008 (Parry-Langdon, 2008).

Although the prevalence of common mental disorders appears relatively low, addressing this issue seems pertinent as it is thought that adult disorders have their roots in earlier life. Both depression and anxiety are thought to be recurrent, lifelong disorders (Paluska & Schwenk, 2000; Harrington, 2002), with evidence that depression in youth can predict more severe depression in later life (Motl, Birnbaum, Kubik & Dishman, 2004). For example, depressed adolescents were found to have 4.91 times the odds of experiencing depression 10 years later (Naicker et al., 2013). Similarly, most adults who suffer from

anxiety report experiencing anxious feelings in earlier life (Maughan, Brock & Ladva, 2004). This is particularly important given the projection by the WHO that, by 2020, depression will be second only to cardiovascular disease as the world's leading cause of disability and death (Murray & Lopez, 1997).

As common mental disorders are apparent in children, defining these concepts in terms of behaviours and symptoms may lead to a better understanding of their effect upon children's lives. Generalized anxiety is characterized by feelings of worry, apprehension and self-doubt (Taylor, 2000) relating to a number of events or situations (Brown, O'Leary & Barlow, 2001). Examples of worry provoking situations include school work, body weight, making friends and bad things happening to others (Green et al. 2005). Additionally, anxiety may be accompanied by irritability, fatigue, muscle tension, restlessness, sleep disturbances and low levels of concentration (DSM-IV; American Psychiatric Association, 1994). Other anxiety disorders suffered by children include separation anxiety (concerns about separation from caregivers; Scaini, Ogliari, Eley, Zavos & Battaglia, 2012), Obsessive Compulsive Disorder (irrational thoughts, leading to compulsive routines in order to reduce anxiety; Jacob & Storch, 2013) and specific and social phobias (characterised by fear of object or situations and large / new social encounters respectively; Green et al., 2005).

Depression is often seen alongside anxiety (Singleton et al., 2001; Kessler, Chiu, Demler & Walters, 2005), but can also present individually in a child. It is characterised by sadness and loss of ability to derive pleasure from previous interests, alongside irritability and may be associated with feelings of worthlessness, alterations in sleep or appetite and slowing of speech (Belmaker & Agam, 2008; Sowislo & Orth, 2013).

Although only a small percentage of children will experience chronic or acute mental disorders, the remainder of the population may experience episodes of mild depression and anxiety such as being worried or fed up. To prevent acute psychological health issues developing into chronic problems that may persist into adulthood, it is of paramount importance to optimise their overall psychological well-being. In relation to this notion, it is essential to reiterate that psychological well-being does not just arise from a lack of mental health

disorders but also involves the presence of positive states (Massé et al., 1998) and the realisation of human potential (Waterman, 1993). A number of positive states are considered important contributors to one's psychological well-being, including self-esteem (Fox, 1997), positive affect, mood and emotion (Biddle, 2000) and quality of life (Deiner, 2009). In a 'healthy' child population, psychological well-being can be increased by not only reducing any worry or sadness children may feel, but also by improving their positive states. Doing so could directly impact on psychological health; increasing positive states should lead to a higher level of PWB. Yet in addition, increasing positive states may have a protective effect from developing clinical disorders, increasing overall well-being may act as a buffer for any negative feelings (i.e. anxiety or depression) the children experience (Mann, Hosmann, Schaalma & De Vries, 2004), and prevent a decline into a clinical disorder. One potential explanation for the buffering effect may be that the impact of a stressful life event upon psychological well-being is mediated by how one copes with that event (Lazarus & Folkman, 1984). So children with better PWB are able to employ different coping mechanisms, and are potentially more resilient to negative life events (Ebata & Moos, 1991) and prevent a decline into a clinical disorder. Folkman & Moskowitz, (2004) noted that avoidance coping has been associated with poorer mental health, whilst greater approach/problem focused coping had been related to high psychological well-being (Ebata & Moos, 1991).

1.3. Positive states

The concept of psychological well-being includes a number of positive states and cognitions. One highly researched area of PWB appears to be the concept of self-esteem or self-worth; defined as how much one values themselves as a person (Harter & Whitesell, 2003). It is thought to encompass the evaluative and affective components of self-concept (a description of the self; Biddle & Mutrie, 2008) (McAuley et al., 2000). It has been noted that the consideration of self-esteem should be an important focus in mental health promotion (Mann et al. 2004), as it is thought to be an indicator of emotional stability and is associated with the ability to adjust to life's demands (Fox, 2000). Furthermore, Fox (2000) noted that high self-esteem is linked to qualities such as leadership

(Beaumeister, Campbell, Krueger & Vohs, 2003), resilience to stress and recovery from disease (Mann et al., 2004).

Low self-esteem, characterised by being sensitive to criticism, avoidance of people and situations that may threaten self-esteem and hiding inner thoughts and feeling (Sowislo & Orth, 2013), has been linked to maladaptive behaviours in adults such as alcohol use and risky sexual behaviour (Wild, Fisher, Bhana & Lombard, 2004) along with smoking (Carvajal, Wiatrek, Evans, Knee & Nash, 2000), and eating problems (McGee & Williams, 2000) in adolescents. More importantly, it is linked to a higher prevalence of mental disorders including anxiety (Lee & Hankin, 2009) and depression (Goodman & Whitaker, 2002; Dishman et al., 2006). A recent meta-analysis demonstrated that low self-esteem was a contributing factor of depression and had a reciprocal relationship with anxiety (Sowislo & Orth, 2013), providing evidence for the notion that self-esteem may act as a marker for recovery from mental illness (Fox, 1999).

The construct of self-esteem has been described as multi-dimensional and hierarchical (Harter, 1985; Fox & Corbin, 1989; Marsh, 1997). At the apex of the concept is global self-esteem or self-worth (GSW) that reflects peoples' positive and negative evaluations of themselves as a totality or in general (Rosenberg, Schooler, Schoenbach, & Rosenberg, 1995). This global perception is made up of how one views themselves in different domains of life, which become increasingly specific throughout the hierarchical model (Whitehead, 1995). The domains contributing to global self-worth may include academic self-worth or scholastic competence, social self-worth, emotional self-worth, behaviour conduct, physical appearance and physical self-worth (Harter, 1985; Hagger, Asçi, Lindwall, 2004) depending upon the population. Each domain is further made up of self-perceptions in sub-domains of these areas. For example, physical self-worth (PSW), which relates to one's feelings and perceptions of happiness, satisfaction, and confidence in the physical self (Edwards, Ngcobo, Edwards & Palavar, 2005), can be influenced by perceptions of sports competence, body attractiveness, physical strength and physical conditioning (Fox & Corbin, 1989; Whitehead, 1995). This hierarchy is demonstrated below in figure 1.1.

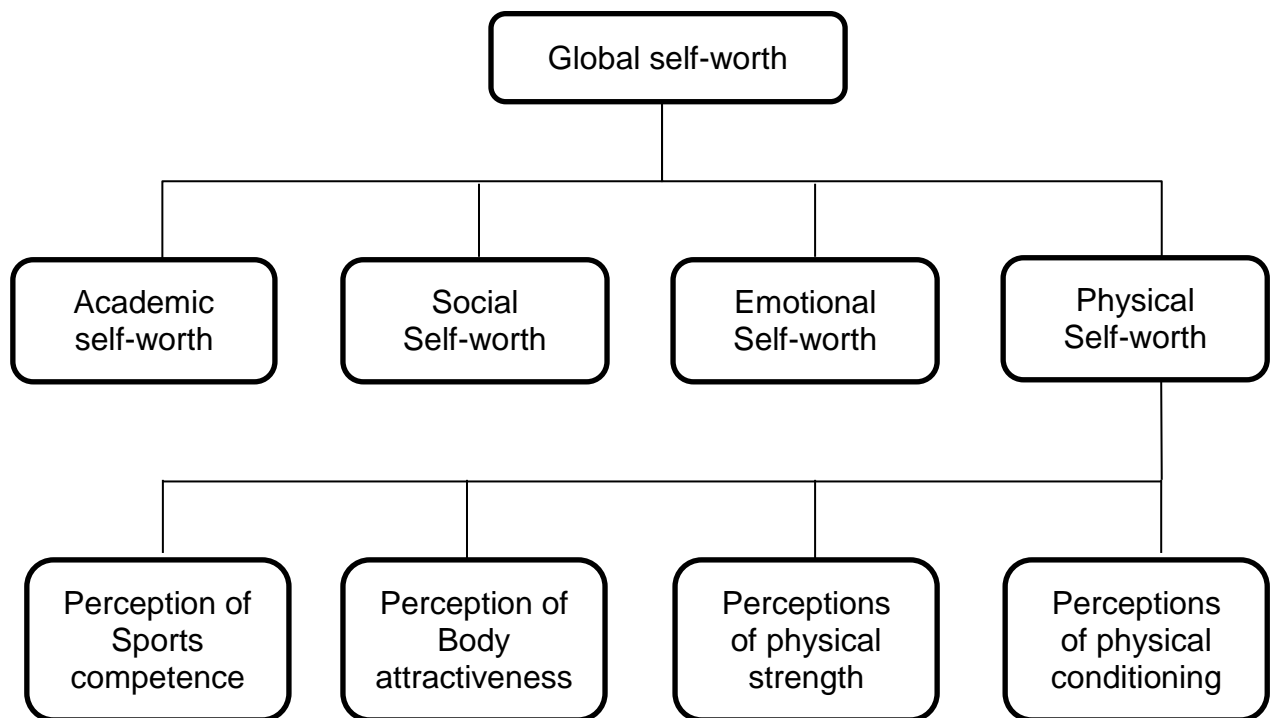


Figure 1.1 Multidimensional hierarchy of self-esteem

The sub-domain levels of the self-esteem hierarchy are deemed to be the more specific and changeable areas of the self (Hagger et al., 2005), while progression through the hierarchical structure leads to more general and stable constructs that are resistant to change (Harter & Whitesell, 2003). It is important then to target the domain and sub-domain levels of self-worth when addressing psychological well-being. In relation to this, PSW is an important area to improve global self-worth, due to its unique position in the self-esteem hierarchy; the body provides the public interface between the individual and their surroundings (Fox, 2000; Lindewall, Asci, Palmeira, Fox & Hagger, 2011) and can therefore be of great importance to one's GSW. Furthermore, PSW has been seen to have its own relationships with psychological well-being (Fox, 1999; Biddle & Asare, 2011).

Whilst generally it appears that self-esteem can be beneficial to overall PWB, it has been noted that high self-esteem can sometimes promote maladaptive feelings and behaviours (Chamberlain & Haaga, 2001), causing people to react badly feedback and criticism (Seery, Blacovich, Weisuch & Vick, 2004; Crocker & Park, 2004), and potentially cause antisocial behaviour and violence in reaction to ego-threats (Beaumeister, Smart & Boden 1996).

Additionally, with high self-esteem, the possibility to develop feelings of worthlessness or low self-esteem through poor performance in important areas of life is ever present (Chamberlain & Hagga, 2001; Crocker & Knight, 2005). Instead, it is thought that achieving a feeling of unconditional self-acceptance, whereby an individual accepts themselves, unconditionally, regardless of any evaluation or comparison to others (Ellis, 1994), would be more beneficial to PWB. Unconditional self-acceptance is reported to arise when the individual can accept that they exist, they are neither worthy nor unworthy and that they see no reason to be different to who they are (Nichols, 2012). While evaluation of one's performance may still take place, those with high unconditional self-acceptance will not reflect those evaluations upon themselves as a person. This theory differs substantially from that of self-esteem, which suggests that perceptions of competence, in important areas of life, judged against a criterion, contribute to one's overall feeling of worth (GSW).

Though unconditional self-acceptance would more than likely lead to high PWB as demonstrated by Macinnes (2006), the reality of not reflecting evaluations of one's performance in life domains upon one's self, seems an implausible task, as many people define themselves by certain areas of their life. Additionally, both Macinnes (2006) and Chamberlain & Haaga, (2001) noted that self-esteem is more highly related to depression in adults than unconditional self-acceptance; while multiple studies in adolescents have also reported links between low self-esteem and depression (Goodman & Whitaker, 2002). Indicating that self-esteem may be a more viable construct to address when targeting psychological well-being.

1.4. Increasing psychological well-being

The reduction of negative affect or mental disorders is often achieved through medication or psychotherapy; although these methods may encourage recovery of mental illness which in turn allow people to fulfil some of the requirements of well-being (e.g. cope with normal stressors and contribute to the community) they do not necessarily increase important positive states. Furthermore, as noted above, for a 'healthy' child population, who may occasionally experience worry or feel melancholic, medication is unnecessary, yet these maladaptive

states need to be reduced and positive states encouraged so as to increase psychological well-being. One method of doing so is through physical activity.

1.4.1. Mechanisms through which PA may impact PWB

Physical activity may impact PWB through a variety of mechanisms including psychological, psychosocial, pharmacological or physiological (Scully et al., 1998). Psychological mechanisms can include aspects such as distraction from unpleasant stimuli, although this may only be apparent with acute exercise (Paluska & Schwenk, 2000), and the social interaction hypothesis, whereby social interaction and support can aid the increase in psychological well-being (McAuley & Rudolph, 1995; Fox, 2000; Guindon & Cappeliez, 2010). Monshauer, Have, Van Poppel, Kemper & Vollebergh (2012) recently reported the mediating effects of social interaction between PA and psychological health relationships, but noted that perceptions of body weight can also mediate the relationship. In line with this McClure et al. (2010) found obesity to be the largest, modifiable risk factor associated with low self-esteem in adolescents. The results of Monshauer et al., (2012) and McClure et al., (2010) seem to imply that reducing obesity and modifying body image perceptions, through physical activity, may increase self-esteem and in turn psychological well-being. Yet Wardle and Cooke (2005) noted that being overweight or obese does not inevitably result in lower PWB through poor body image and self-esteem, rather, they reported both normal weight and overweight individuals showed dissatisfaction with their body image and that many obese children do not have low self-esteem. The combination of these results therefore indicates that the evaluation of the self against one's own criteria is what impacts self-esteem, rather than the actual weight itself.

Finally, the 'mastery hypothesis' has been proposed as a psychological mechanism, whereby PWB may be increased via mastery of a particular skill, in turn increasing one's coping self-efficacy (Craft & Perna, 2004), aiding those with negative affective states (e.g. depression). Furthermore, Morgan and Sonstroem (1989) and later Fox (1998) proposed self-efficacy as a mechanism by which self-esteem is increased through physical activity in the Exercise and Self-esteem model (EXSEM). It has been proposed that low self-esteem develops from perceptions of discrepancies in competence between the

perceived ideal and the perceived real self (Rosenberg, 1979; Higgins, 1987; Harter, 1993) in aspects that an individual values and views as important (James, 1890; Harter, 1999; Lindewall et al., 2011). Increasing competence through mastery experiences (an antecedent of self-efficacy; Bandura, 1977) will reduce discrepancies and increase one's perceptions of themselves.

Physiological mechanisms have also been proposed, for example, McAuley et al., (1997, 2000) found that in adults, changes in maximal oxygen uptake (during maximal graded exercise test) significantly predicted changes in physical self-perceptions. Additionally, physical activity may result in an increase in brain monoamines (e.g. dopamine, serotonin and noradrenaline) and endorphins, reducing negative affective states and promoting positive moods (Craft & Perna, 2004). Although these mechanisms have been proposed, there seems to be little in the way of evidence to support one or the other. Researchers have therefore noted that it is unlikely that one mechanism can fully explain or account for the relationship between physical activity and psychological well-being, rather, a multi-disciplinary approach should be taken when attempting to understand the concept (McAuley & Rudolph, 1995).

1.5. Summary

There is a need to increase psychological well-being in children in order to prevent the continuation or development of acute mental disorders in later life. By reducing negative affective states and optimising positive states, the child will be able to experience heightened levels of well-being and possess a mechanism by which to buffer against any negative emotions experienced that may lead to a decline into mental illness; a strategy for increasing PWB is through physical activity. The following chapter reviews the literature concerning psychological well-being and physical activity, addressing both the maladaptive and positive states in both clinical and healthy child populations.

1.6. Thesis Preface

The following provides an overview of the order in which the studies were uptaken and the rationale for the layout of the chapters.

1.6.1. Chronological framework of studies

This thesis consists of five studies, data for which were collected between March 2010 and April 2012. The first study to be undertaken was the calibration and validation of the GENE accelerometer presented in chapter 7. Recruitment of participants for this study was initiated in January 2010, and continued throughout the duration of the data collection period which covered the period between March and July in 2010. Following this, data analysis was undertaken, with chapter 7 being written up and accepted for an oral presentation at the 27th Paediatric Work Physiology conference in September 2011. Additionally, a version of chapter 7 was submitted for publication to the Journal of Science and Medicine in Sport in September 2011. The article was published ahead of print in September 2012.

Whilst undertaking the write up for chapter 7, recruitment of participants was undertaken for the studies presented in chapters four, five and six. Participants were recruited during November and December 2010, whilst data collection was undertaken between January and April 2011. Analysis of data for chapter four was conducted between August 2011 and December 2011. Analysis of data for chapter five and six was undertaken between October 2011 and February 2012. Each study was subsequently written up as a chapter for this thesis.

Finally, recruitment of participants for chapter eight was undertaken during November and December 2011, with data collection being undertaken between January and April 2012. Data analysis for chapter eight took place between May and October 2012, and was subsequently written up for inclusion within the present thesis. Write up of all chapters was a continuous process that took place at various time points throughout the period of study. Appendix 5. provides a Gantt chart detailing the timings for completion of stages of each study presented within this thesis.

1.6.2. Rationale for sequence of studies and chapter organisation

The original focus of this thesis was to measure the relationships between PA and PWB in children using an objective exposure measure to overcome the issues associated with self-reported measures that have previously been used

within the PA and PWB literature. The aim was to use the GENEActiv accelerometer as the PA measure, hence the validation and calibration of this monitor was the first study to be undertaken. Upon completion of this study, it was noted that the casing for the monitor was not fully waterproof, and that further developments of the mounting and surround of the accelerometer were needed in order to obtain a fully waterproof monitor that could be worn for extended periods of time. As a result of this, the ActiGraph accelerometer was employed as the exposure measures in the studies presented in chapters four, five and six. However during the data reduction process it became clear that the decisions made would have a large influence on the relationships and effects observed. The impact of data reduction decisions upon the results appeared to have been overlooked in other studies previously employing accelerometry both within the PA-PWB literature and in the wider literature. These findings resulted in a change of focus within the thesis; the emphasis was placed upon exploring the impact that data reduction decisions used within accelerometry can have upon the relationships observed between PA and PWB.

Once the impact of data reduction processes had been explored, it became apparent that the use of a wrist-worn, waterproof device could provide a means of overcoming some of these problems. The development of the new GENEActiv, and the availability of cut-points for the use of this monitor (chapter 7), meant that the monitor could now be used to explore the possibility of complete observation (24-hour wear) with children and as a result could provide greater insight into the effects of data reduction decisions upon the observed relationships. Hence, the chapters are presented as a logical progression as opposed to the chronological order of data collection.

CHAPTER TWO

A REVIEW OF THE PSYCHOLOGICAL WELL-BEING AND PHYSICAL ACTIVITY LITERATURE

The following chapter will outline the findings of previously published literature examining the relationship between PA and PWB in children. Studies utilising cross-sectional and longitudinal designs will be presented first, followed by a review of experimental studies attempting to improve PWB through changing PA. Specific areas of PWB (depression, anxiety and self-esteem) will be addressed; however literature relating to overall well-being will also be included. A brief introduction pertaining to the relationship between PA and PWB in adult populations will be provided, though the main focus will remain on childhood populations. The chapter will conclude with a critique of the literature, specifically focusing on the measurement methods used to assess the relationships.

2.1. Physical activity and psychological well-being

Over the last few decades, researchers have assessed the impact that PA can have upon aspects of PWB in adult populations. Overall, findings suggest that PA improves depressive symptoms for those with clinical depression (Donaghy & Taylor, 2010). Specifically, PA has been shown to be inversely related to concurrent levels of depression (Harris, Cronkite & Moos, 2006), can be an effective treatment in sufferers of mild to moderate depression (Mead et al., 2009), and have effects comparable to those of anti-depressants medication in participants with major depressive disorder (Blumenthal, Michael, & Babyak; 2007). Conversely, recent studies have shown that increases in PA have no beneficial effects over and above the effect of usual care in depressed patients (Chalder et al., 2012).

A meta-analysis supports these findings, showing that PA in the form of exercise was effective in comparison to receiving no treatment (Standard Mean Difference -0.67, 95% CI: -0.90 to -0.43), yet was no more effective than cognitive therapy or antidepressant medication (Rimer et al., 2012). It has been

noted that PA may be more effective for sufferers of more severe depression (Paluska & Schwenk, 2000). Importantly, a decrease in depression scores has been related to more PA within the general population (Steptoe et al., 1997; Torres, Sampsel, Ronis, Neighbors and Gretebeck, 2013), with other studies noting a protective effect against the development of depression (e.g. Strawbridge, Deleger, Roberts, & Kaplan, 2002).

Beneficial effects of PA upon anxiety disorders have also been reported; a review of the anxiety literature (Taylor, 2000) reported that cross sectional studies (e.g. Aldana, Sutton, Jacobson and Quirk, 1996) showed an inverse relationship between physical activity and anxiety, with those who were less active reporting higher levels of anxiety. Similarly, it was noted that exercise training has been successful in reducing trait anxiety in both anxious and non-anxious individuals (Long & Van Stavel, 1995; Taylor, 2000); both single bouts and longer periods of PA (in the form of exercise) produce positive moderate reductions in state and trait anxiety (Fox, 1999). A meta – analysis looking at the effect of PA interventions upon anxiety in adults reported an overall effect of PA interventions upon anxiety (Number of studies (k) = 15; effect size (ES) = 0.219), reporting a beneficial effect of PA in healthy adults (Conn, 2010).

Thus, research into anxiety and depression shows that PA is associated with a reduction in symptoms in adults who are already suffering from anxiety and/or depression as well as lower prevalence and future risk of anxiety and depression. However, as noted in chapter one, the absence of disorders is not enough to ensure PWB; there must also be a presence of positive states (Masse et al., 1998). A positive link between PA and self-esteem has been established within the adult literature (Scully, Kremer, Meade, Graham & Dudgeon, 1998). Spence, McGannon and Poon's (2005) meta-analysis of 113 studies noted a small overall effect size of 0.23 (95% CI: 0.18 - 0.28), with changes in physical fitness found to be a significant moderator of this effect; lifestyle (ES = 0.36) and exercise (ES = 0.26) interventions were found to be more effective than skills based programmes (ES = -0.03).

Overall the literature demonstrates a positive relationship between PA and PWB in adults. The research presented above by no means represents an exhaustive review of the adult literature, rather, it is provided to introduce the

reader to the apparent relationships. The findings reported in children's literature appear to echo those found within studies of adults, however considerably less research has been undertaken in children.

2.2. Physical activity and psychological well-being in children and adolescents

The following review details the findings from studies in children. It is important to note at this stage that definitions and measurement of physical activity vary considerably between studies, an issue that will be considered later in this review; the following is intended to provide an overview of the relationships between PA and PWB. Studies were obtained through searches of computer databases including PubMed, web of science and PsycINFO; key words used in the searches included, child, youth, adolescent, psychological well-being, depression, anxiety, self-esteem, emotional distress, physical self-worth, sports competence, self-perceptions physical activity, exercise, moderate activity, vigorous activity, sedentary, intervention, trial, review. Studies were excluded if participants were over 18 years of age and if areas of well-being beyond emotional distress and self-esteem were the main focus of the study. For example, studies pertaining to social identity, cognitive functioning, hyperkinetic disorders and psychosis, or studies assessing mental health disorders arising from accompanying medical conditions, (e.g. arthritis) were excluded. In addition to the initial searches, further searches were performed for references noted in review articles that had not been identified in the initial search.

2.2.1. Cross-sectional relationships between physical activity, depression and anxiety

The use of a cross sectional study designs provide fundamental evidence on which to base future longitudinal and experimental studies, and can therefore be used in the first instance to establish associations between variables. Despite the benefits, this type of design is at greater risk of threats to internal validity (Ho, Peterson & Masoudi, 2008), which can confound the results observed and cannot address causality. The studies outlined below present

evidence pertaining to the relationships between PA and aspects of PWB undertaken using a cross-sectional study design.

Early research noted that adolescents who reported higher levels of PA in terms of sports participation and vigorous intensity activity had a lower risk for emotional distress (Steptoe & Butler, 1996). Subsequent, studies reported similar relationships. For example, Tomson, Pangrazi, Friedman & Hutchison (2003) reported significant differences in depression scores between self-reported 'active' and 'inactive' boys, with lower depression scores for boys reporting more sports participation outside of school. Brosnahan, Steffen, Lytle, Patterson and Boostrom (2004) found that adolescents participating in 3 – 5 physical education sessions per week were less likely to report feeling sad than those partaking in fewer sessions. While the findings of Kantomaa, Tammelin, Ebeling & Taanila (2008) showed that active boys (reporting ≥ 4 hours per week of moderate to vigorous physical activity [MVPA]) had significantly lower levels of anxiety and depressive symptoms than boys who were moderately active (2 – 3 hours MVPA per week) or inactive (≤ 1 hours MVPA per week), indicating more emotional problems for those undertaking less MVPA. These results are supported by those of Hong et al. (2009) who noted that pupils undertaking more time in PA, have lower odds ratios for depressive symptoms. Additionally, Adeniyi, Okafor and Adeniyi (2011) reported a strong inverse relationship between PA and depression ($r = - .82, p < 0.001$), with 67% of the variance in depression accounted for by PA; this finding could be considered highly significant with a low p value and a large sample size ($n = 1100$), meaning that the study was highly powered.

In general, available evidence supports the notion that being more physically active, in a variety of ways is correlated with fewer depressive symptoms. The studies presented above were undertaken in large samples ($n=1100$ to 7002) across a variety of populations. Despite this, some of the studies have weakness regarding the measurement of variables; for example, Brosnahan et al. (2004) assessed feelings of sadness using one question requiring recall for the previous year, indicating a low measurement precision, which may result in an underestimation or overestimation of the outcome variable.

Contrary to the findings presented above, some studies have reported an absence of associations. For example in a study of 4211 Canadian adolescents, Allison et al. (2005) found no association between self-reported vigorous intensity PA and depression, anxiety or psychological distress after controlling for age, sex and socio-economic status (SES), but found that vigorous activity was negatively related to problems in social functioning (i.e. more pro-social behaviour). The absence of relationships reported by Allison et al., (2005) is supported by the results of Johnson et al. (2008), who investigated the relationship between PA and depression in girls using both objective measure (accelerometers) and self-reported methods. The authors reported no significant associations between MVPA and depression with either measurement method. Similarly, Hume et al. (2011), using objective measurement methods (accelerometers) in 115 children aged 14 (± 0.61) years, reported no cross-sectional relationship between MVPA or vigorous PA and depressive symptoms.

Studies addressing the relationship between PA, depression and anxiety in children appear equivocal; the majority of studies using self-reported methods appear to demonstrate a strong inverse relationship, yet this seems attenuated with the use of more objective measures. Although the use of objective measurement can result in decreased sample size, the attenuation of relationships with objective measures was apparent even with the use of a large sample size ($n = 1397$; Johnson et al., 2008), which further indicates the equivocal nature of the PA and depression relationship. This attenuation may occur as the use of self-reported methods could lead to aspects such as social desirability and common method variance. These issues will be discussed further in chapter three (p. 97).

It should also be noted that compared to depression, few studies have examined the relationship between physical activity and anxiety, yet studies exploring the relationship between PA and multiple aspects of PWB also include anxiety and will be reviewed later.

2.2.2. Relationships between physical activity and self esteem

Cross-sectional studies have reported a significant association between physical activity and self-esteem (Fox, 2000). In particular, child based studies have focused on the sub-domains of physical self-worth rather than the global aspects of self-esteem; this is in accordance with the recommendations of Spence et al. (2005). For example, Hagger, Ashford and Stambulova, (1998) found that physical self-perceptions, in particular physical conditioning, physical strength and sports competence sub-domains, differed between high and low active children. Similarly to this, Crocker, Eklund and Kowalski (2000), using structural equation modelling, showed that in children aged 10 – 14 years, perceptions of the physical self (all sub domains of PSW) were significantly related to self-reported vigorous PA, accounting for 27% – 29% of the variance in PA. Though relationships appear to be present, studies tend to report only weak to moderate relationships, based on Cohen's effect sizes for correlation coefficients (Cohen, 1992). For example, Raudsepp, Liblik & Hannus, (2002) examined the link between children's' physical self-perceptions, self-reported MVPA and physical fitness in 11 – 14 year old children (n = 134 males; n = 119 females), correlational analysis demonstrated significant relationships between all sub domain physical self-perceptions for both males (r = .17 to .37), females (r = .17 to .30) and all participants (r = .25 to .28). Furthermore, multiple regression analysis revealed that perceptions of physical strength, sports competence and PSW accounted for 21% of the variance in MVPA and 28% of the variance in physical fitness for males. Similarly, perceptions of body attractiveness, PSW and sports competence accounted for 14% and 23% of the variance in MVPA and fitness, respectively, in females. These findings support those reported by Crocker et al. (2000).

The studies presented above allow an examination of PA and self worth relationships by employing domain specific PSW measures, a strength of the studies. Despite this, the findings are limited by the nature of the relationships addressed; specifically, the authors did not address whether any of the variance in PSW or its subdomains were accounted for by PA, yet they support the notion that relationships exist between the two variables. Additional studies

addressing the relationship between PA and self-esteem are noted below, included within those studies that address multiple aspects of PWB.

2.2.3. Physical activity and overall psychological well-being

The studies noted above examine the relationship between PA and individual aspects of PWB. A number of studies, however, address the relationship between PA and multiple areas of PWB, as well as addressing well-being as a single entity. For example, Goldfield et al. (2011) administered self-reported measures of depression, anxiety, body esteem and PA to children and adolescents between the ages of 12 and 18 years ($n = 1259$). Partial correlations (controlling for body mass index [BMI]) showed weak but significant associations for moderate intensity activity (defined as non-exhaustive activity, but causes light sweating) with depression ($r = -.07$), anxiety ($r = -.08$), and feelings about appearance ($r = .09$). Vigorous intensity activity (defined as causing heavy sweating and rapid heartbeats) however, was more strongly associated with depression, anxiety and body esteem ($r = -.10$; $r = -.16$; $r = .14$ respectively). The authors also noted an effect of sex, in males depression remained significantly related to moderate intensity activity, while no relationships with moderate intensity remained significant in females. For vigorous activity, the relationship with anxiety became non-significant in males, while the same was true for depression in females. The findings of this study are interesting, as associations are weak yet are reported as statistically significant with p values below 0.05 and 0.001 for different variables. A potential explanation for this is the large sample size ($n = 1259$) used within this study. The findings, that PA is related to aspects of PWB, should therefore be treated with caution.

Assessing similar areas of PWB, Parfitt and Eston (2005) examined the relationship between habitual PA of children (mean age = 10.4 ± 0.4 years), measured with pedometers and PWB in terms of depression, anxiety and GSW. Correlational analysis revealed significant relationships for all variables ($r = -0.48$; $r = -0.60$; $r = 0.66$ respectively). Yet when undertaking partial correlations, controlling for the other psychological variables (e.g. controlling for anxiety and GSW when assessing the relationship between PA and depression), only the relationship between PA and GSW remained significant. Additionally, children

accumulating more than 12000 steps per day had better psychological profiles than those accumulating less than 9200 steps per day. The results of Parfitt and Eston (2005) are not comparable to those of Goldfield et al. (2011) as Parfitt and Eston did not consider the intensity of activity, a limitation of the methods employed; rather relationships were developed between total activity and PWB. The findings appear to partially support the finding of previous studies using self-report, showing significant relationships between PA, depression and anxiety (e.g. Kantomaa et al., 2008; Hong et al., 2009; Adeniyi et al., 2011). Despite the more objective methods of PA measurement employed within this study, which overcome some of the problems associated with self-report methods, the findings are limited due to the use of a global measure of self-worth rather than the use of a domain specific measure.

In an attempt to establish a more precise estimate of the association between PA and PWB, Parfitt, Pavey and Rowlands (2009) employed accelerometry rather than self-reports or pedometer methods to assess the relationship between specific activity intensities and PWB variables. The authors reported that higher levels of anxiety and depression were observed in children who accumulated high levels of very light activity ($r = .384$ & $r = .345$ respectively $p < .001$); these children also reported low levels of global and physical self-esteem ($r = -.400$, $r = -.388$ respectively, $p < .001$). Alternatively, those children partaking in high levels of high intensity activity (vigorous) reported lower levels of anxiety ($r = -.310$, $p < .001$), a trend towards lower depression ($r = -.177$; not significant) and higher levels of self-esteem in the subdomain level ($r = .324$ to $.325$, $p < .05$). It is of note that, after controlling for body fatness, relationships were attenuated, although the majority remained significant; relationships between, perceptions of body attractiveness and time in very light and light intensity activity, and GSW and PSW relationships with light intensity were attenuated. Additionally, the authors reported that children spending a modest time in very light intensity activity had a better psychological profile than those spending less time in low intensity exercise. Furthermore, the profiles of those spending a modest time in low intensity activity corresponded to individuals who spent a high amount of time in vigorous activity. This finding led the researchers to suggest that a combination and balance of both intensities would be most beneficial. The results provide support for the self-reported relationships noted

previously, though differ from the relationships noted by Goldfield et al. (2011), which suggested a significant link between vigorous activity and depression; despite similar magnitude and direction, significance of the relationship differed between the two studies. This may be a result of large differences in sample sizes apparent between the two studies, resulting in a weak relationship ($r = -.10$) reported as statistically significant with a larger sample size (Goldfield et al. 2011), as previously noted.

The exploration of relationships between multiple aspects of PWB within a single study allows for the examination of potential mediating relationships between well-being variables. It has been suggested that low self-esteem is present in many mental illnesses (Fox, 1999) with some researchers finding strong negative correlations between self-esteem and depression (Watson, Suls & Haig, 2002), with levels of self-esteem able to predict depression (Goodman & Whitaker, 2002). Researchers have therefore hypothesised that changes in self-esteem mediate the relationship between PA and depression. Dishman et al. (2006), assessed the proposed mediating relationship between self-reported PA, sports participation, self-esteem and depression in adolescent girls ($n = 1250$). The results demonstrated that physical self-worth mediated the relationship between PA/sport participations with self-esteem, which in turn was negatively associated with depression symptoms.

More recently, McPhie and Rawana (2012) attempted to examine the link between self – reported PA and depression, with self-esteem as the mediator in early and late adolescence. Results from the mediation analysis suggested that the relationship between frequency of PA and depressive symptoms were fully mediated by global self-esteem in young adolescent boys (mean age 14.7 ± 1.4 years) and late adolescent boys and girls (mean age = 16.2 ± 1.3 years), while controlling for age, ethnicity, parent education and BMI. Interestingly, no significant relationship and hence no mediation was observed for young adolescent girls; the authors suggested the absence of a significant relationship may indicate that PA is a less salient factor of PWB among girls of this age.

The results of the above studies further highlight the complex nature of the relationships between physical activity and PWB, and provide supporting evidence for Parfitt and Eston (2005) who noted that only the relationship

between PA and GSW remained significant once other PWB variables were accounted for.

2.2.4. Physical activity, sedentary behaviour and psychological well-being

Alongside the examination of specific PA intensities, researchers have begun to assess the relationship between sedentary time with PWB. It is thought that sedentary behaviour could present a unique relationship with PWB, separate to that of PA (Webb, Benjamin, Gammon, McKee & Biddle, 2013). Sedentary behaviour is usually defined as activities with an energy expenditure of less than 1.5 times resting energy expenditure. (see chapter three, p. 70 for definitions).

Using self-reported and proxy measures (parents' reports of children's PA; see chapter three, p.73 for more detail) of PA, Ussher, Owen, Cook, Whincup (2007), reported that PA and sedentary behaviour were independently associated with psychological difficulties in adolescents aged 13 – 16 years. Specifically, lower levels of PA and more time spent in sedentary behaviours resulted in higher levels of psychological difficulties. However, with the proxy reports of PA, this relationship was only significant in girls. Adjustment for confounders, including social class, smoking status, alcohol intake, number of parents and predicted GCSE exam passes, resulted in only minor variation in the magnitude of relationships, with no change in significance. For the fully adjusted models, PA and sedentary behaviour were predictive of total psychological difficulties in boys, whereas only PA was predictive in girls.

Similar results were found by Hamer, Stamatakis and Mishra (2009), who assessed the relationship between psychological distress, PA and sedentary time in children aged 4 – 12 years. PA and sedentary time were measured through parental reports of frequency of sports or active play (≥ 15 minutes) and time per week spent watching television, respectively. PA and sedentary time were independently related to psychological difficulties after adjustment for multiple confounding variables; more time spent watching television and low PA levels were associated with higher difficulty scores, as was noted by Ussher et al. (2007). Additionally, Hamer et al. (2009) reported that high sedentary time and low PA were found to interact, to increase psychological distress; being

grouped into the highest television viewing times (>163 minutes per day) and lowest PA group (≤ 6 sessions per week) was associated with increased odds of abnormal psychological difficulties (Odds Ratio [OR]: 1.45, 95% CI: 0.94 – 2.25; OR: 1.64, 95% CI: 1.02 – 2.61, respectively). The authors noted that PA could potentially reduce the risk of psychological distress associated with television viewing. The results of both Ussher et al. (2007) and Hamer et al. (2007) are supported by those of Cao et al. (2011) who's findings showed that a combination of higher sedentary time and insufficient vigorous PA resulted in a 12% increase in depression and a 9% increase in anxiety in Chinese children.

Page, Cooper, Griew and Jago, (2010), used an objective measure (accelerometry) to assess sedentary time and PA in children age 10 – 11 years and investigated relationships with measures of psychological difficulties. Additionally, the authors included self-report measures of sedentary behaviour to provide context for the sedentary time. Results appeared somewhat equivocal; self-reported sedentary behaviours (television viewing and computer use) were positively associated with more psychological difficulties, yet objectively measured sedentary time was associated with lower psychological difficulties (i.e. more positive well-being). MVPA was related to lower difficulties only when objectively measured sedentary time had been accounted for; the authors noted that this may be due to the potentially neutralising effect of the relationships between MVPA and the individual subscales that contributed to the total difficulties score. Risk of psychological difficulties was highest in children who spent a total of ≥ 2 hours viewing televisions viewing and ≥ 2 hours using a computer. Additionally, it was noted that children who met the recommended PA guidelines (≥ 60 minutes MVPA; Department of Health, 2011) but exceeded television or computer use guidelines were at increased risk of psychological distress. The disparity between the associations detected using difference measures of sedentary behaviour highlighted the confounding effect of measurement methods when assessing PA and health relationships. The inverse relationship that was apparent between objectively measured sedentary time and psychological distress contrasts with previous research employing subjective, self-reported measurement methods, but provides support for the results reported by Johnson et al. (2008), who also found an inverse association

between objectively measured sedentary behaviour and depression in girls aged 11 – 12 years.

Similar findings have also been noted by Nihill, Lubans and Plontnikoff (2013), who in a study of 357 girls (mean age = 13.2 ± 0.5 years) showed significant inverse associations between self-reported sedentary behaviours (e.g. computer use, watching DVDs, total screen time) and global self-esteem after adjustment for confounders. However, accelerometry assessed sedentary time showed no significant relationships with self-esteem. The two studies noted above highlight the potential impact of measurement method upon observed associations, with relationships apparent when both constructs are measured through self-report. This may be a result of common method variance, whereby constructs appear related due to the use of the same measurement method (e.g. Self-report). The concept will be discussed further in chapter three (p. 97).

Recently, Webb et al. (2013) explored the relationships between self-reported MVPA, sedentary behaviour and aspects of self-worth, including GSW, PSW and its subdomains. Pearsons correlations showed significant positive relationships between MVPA and three aspects of physical self-perceptions; physical conditioning ($r = .198$), physical strength ($r = .228$) and sports competence ($r = .148$). Sedentary behaviour was significantly and inversely related with perceptions of conditioning, strength, sports competence and PSW ($r = -.135$ to $-.233$). Furthermore, the authors noted that MVPA fully mediated the relationship between sedentary behaviour and perceptions of physical strength, while controlling for sedentary behaviour resulted in non-significant relationships between MVPA, physical conditioning and sports competence. These results not only support previous findings demonstrating a relationship between PA and self-esteem variables, but also provide further support for the notion that MVPA and sedentary behaviour have both independent and interactive effects upon children's PWB, yet the magnitude of the relationships provided is weak.

The above studies indicate that both physical activity and time spent sedentary are independently related to psychological difficulties such as

depression. Again results are seen to vary depending upon the measurement method used to assess activity levels.

2.2.5. Confounding variables of PA and PWB relationships

Many of the studies discussed above include or account for potential covariates that may impact the relationships observed (e.g. Raudsupp et al., 2002; Allison et al., 2005; Ussher et al., 2007; Parfitt et al., 2009. Goldfield et al., 2011). Two of the most widely used covariates are BMI, as an indicator of adiposity (Cole, Faith, Pietrobelli & Heo, 2005) and sex differences.

BMI and adiposity have been related to both physical activity (Collings et al., 2013; Kreuser, Kromeyer-Hauschild, Gollhofer, Korsten-Reck & Rottger, 2013) and PWB variables (Sjoberg, Nilsson & Leppert, 2005; Wardle & Cooke, 2005). Wardle and Cooke (2005) noted that self-esteem and well-being may be seriously compromised in children who are obese, though this is not inevitable. The authors note that self-esteem was generally lower in clinical samples (those seeking treatment for obesity), than usually reported in the community (including obese and normal weight children). These results are supported by those of Raudstorp et al. (2005) who noted negative relationships between BMI and all but one physical self-perception (physical strength) in Swedish children aged 10 to 14 years. Additionally, Cuddihy, Michaud-Tompson, Erica & Johnston (2006), reported negative relationships between BMI and physical self-perceptions ($r = -.144$ to $-.325$) in a sample of 294 Australian adolescent females.

In relation to maladaptive aspects of well-being, Sjoberg et al. (2005) showed that BMI was associated with more depressive symptoms, and that adolescents with the highest BMI suffered more often with major depression than those with lower BMI values. More recently, Halfon, Larson & Slusser (2013) reported increased odds of depression for obese children (OR:1.41, 95% CI: 1.04 – 1.93) children aged 10 – 17 years, though a significant risk was not apparent for anxiety.

An example of the effect that controlling for indicators of adiposity or not has on the relationships noted above is reported by Parfitt et al. (2009). The authors

noted that when controlling for body fatness, relationships became attenuated, despite most relationships remaining significant, all relationships between light activity and PWB variables became non-significant. It is important to note however, that Parfitt et al. (2009) used bioelectrical impedance to gain measures of body fat percentage, while most studies use BMI. Although BMI is noted to be a good indicator of fatness, the impact that controlling for BMI may have upon relationships could differ to those described above for body fatness.

Recently, Fairclough, Boddy, Ridgers and Stratton (2012) examined the associations between objectively measured PA, physical self-perceptions and weight status among children aged 10 – 11 years. Results showed differences between normal weight and overweight/obese children's physical self-perceptions; normal weight children demonstrated higher perceptions of physical conditioning, body attractiveness and physical self-worth than overweight children. Additionally, vigorous physical activity was significantly associated with normal weight status, which was associated with physical self-perceptions of physical conditioning and body attractiveness.

The studies discussed above indicate that body fatness may confound the results observed when examining the relationship between PA levels and PWB in children due to its associations with both variables. If body fatness is not controlled for, the resultant relationships are most likely to be over-estimated.

In addition to BMI, many studies have noted differential relationships according to sex. Boys undertake more physical activity than girls (Sallis, Prochaska & Taylor, 2000; Pate et al., 2002; Riddoch et al., 2007; Van der Horst, Paw, Twisk & Van Mechelen, 2007), while females may experience emotional symptoms such as depression and anxiety more than males (Green et al., 2005). Further, some self-worth sub-domains may be more important depending upon sex (e.g. Raudsepp et al., 2002). Due to these potential differences some studies chose to use only single sex participants (e.g. Dishman et al., 2006; Johnson et al., 2008), while others chose to analyse results separately for each sex (e.g. Ussher et al., 2007; McPhie & Rawana, 2012).

Examples of sex differences are presented in Goldfield et al. (2011), who showed that, despite a significant relationship between moderate and vigorous activity and PWB for boys and girls analysed together, some relationships were no longer significant for boys and girls analysed separately. For example, depression was related to moderate activity for the whole group and in males, but not females. For vigorous activity, the relationship with anxiety became non-significant in males, while the same was true for depression in females. Similarly, the variance accounted for in PA by physical self-perceptions differed for males and females in Raudsepp and colleagues' (2002) study; perceptions of physical strength, sports competence and PSW accounted for 21% of the variance in MVPA, in males, while perceptions of body attractiveness, PSW and sports competence accounted for 14% of the variance in MVPA in females.

Both body fatness and sex represent important factors that may both confound and moderate the observed relationships between PA and PWB, and should be accounted for whenever possible.

2.2.6. Discussion of cross-sectional research

Overall, the results of cross-sectional studies appear equivocal, though a trend towards PA having a positive relationship with PWB in children and adolescents emerges. The studies addressed here vary in the populations used, and the sample sizes employed. Of particular interest is that the results appear to vary slightly depending upon the measurement method employed to assess PA. It seems that studies employing self-report methods report relationships of a larger magnitude than those employing objective measures, this is particularly apparent in studies focusing upon emotional difficulties, such as depression; in some cases, not only is the relationship attenuated, but a reversal in direction occurs. For example, self-reported sedentary time shows relationships with emotional problems and psychological difficulties (Ussher et al., 2007; Hammer et al., 2009; Page et al., 2010; Nihill et al., 2013) and self-reported PA shows an inverse association with emotional problems such as depression (e.g. Brosnahan et al., 2004; Hong et al., 2009) yet objective measurements report either no significant relationships between PA, sedentary and emotional problems (Hume et al., 2011) or report relationships occurring in the opposing direction (Johnson et al., 2008; Page et al., 2010). However, these

differences cannot solely be attributed to PA measurement, rather, inconsistencies with both outcome and exposure measures employed mean that the true relationship between PA and PWB is unclear. In addition Sterne and Davey-Smith (2001) argue that results should be interpreted with regard to the context and type of study and that although $p < 0.05$ provides a 'significant' finding, the value of that finding should take into consideration the power of the study and potential bias. The authors note that increased sample size and measurement precision can increase the power of a study (Sterne & Davey-Smith, 2001), resulting in a more confident interpretation of the results. Those studies using a cross-sectional design and employing self-reported methods tend to have large sample sizes compared to those using a more objective measurement method. Yet whilst this may increase the power of the studies, it may also go some way to explaining the equivocal nature of the results (e.g. goldfield et al. 2011 and Parfitt et al., 2009). Whilst those studies using more precise measurement may sacrifice the power provided by larger samples, the results may be less subject to the bias associated with self-reported PA measures. Making the findings more robust.

The relationships between PA and aspects of self-esteem appear more consistent across measurement methods, with only variations in the sub-domains associated with PA apparent between self-report and objective methods. For example, relationships were apparent between all sub-domains and self-reported MVPA in the study by Raudsepp et al. (2002) whereas Parfitt et al. (2009) reported no sub-domains were associated with moderate intensities and only perceptions of sports competence and physical conditioning showed associations with vigorous activity. It should be noted that Parfitt and colleagues did not report associations with MVPA, therefore the results presented serve as an indication of differences rather than a demonstration of differences between like for like intensities.

Despite the ambiguous findings, researchers have attempted to further examine the relationship between PA and PWB through longitudinal studies, in an attempt to unravel the direction of the relationships that cannot be determined through cross-sectional studies.

2.3. Proposed direction of relationships.

It has been proposed that bi-directional relationships may be apparent between maladaptive states such as depression and PA. Birkeland, Torsheim and Wold (2009) noted that PA may have a protective effect against depressive symptoms. Alternatively, the inhibition hypothesis states that a depressed mood inhibits a person's capability to be physically active (Birkeland et al., 2009). Biddle and Mutrie (2008) have noted two similar arguments for the self-esteem and PA relationship; firstly, the motivational approach proposes that high self-esteem or PSW motivates people to partake in physical activities, which in turn maintains or increases their feelings of self-worth (Sonstroem, 1997a). Alternatively the personal development approach proposes that engagement in PA results in positive or negative experience and promotes mastery of skills and tasks that result in an increase or decrease in self-esteem (Sonstroem, 1997b). In order to assess these directional relationships, researchers have made use of longitudinal study designs.

2.4. Longitudinal research

The use of longitudinal research designs can progress the understanding of the relationship between PA and PWB; by measuring the exposure prior to the outcome variable this design is useful in attempting to determine the sequence of events (Ho et al., 2008). Despite these benefits, it is not possible to determine causality, and the risk of confounding variables is increased (Ho et al., 2008) due to the time between baseline and follow up measurement points.

One of the earliest studies to examine the prospective relationships between PA and areas of PWB was undertaken by Motl et al. (2004), who examined the relationship between changes in self-reported PA and changes in depressive symptoms over a two year period in 4594 adolescents. Using latent growth modelling, results showed that not only were higher PA scores associated with decreased depression in the initial assessment, but that changes in PA that occurred over time were associated with changes in depression scores. Specifically, a one standard deviation unit increase in leisure time PA (LTPA) was associated with 0.25 standard deviation decrease in depression. Despite

the effect being small compared to pharmacological treatments, the change was significant.

Following this, a number of other studies attempted to provide evidence for the longitudinal relationship between PA and PWB variables. For example, Wiles et al. (2008) undertook a longitudinal study assessing the relationship between self-reported PA (frequency of sporting activities lasting >20minutes undertaken in the previous week) and emotional problems one year later in children (mean age at baseline = 12.9 years). Findings showed that children achieving 1hour daily activity at baseline had fewer emotional problems at follow up. In addition to this, when adjusting for baseline emotional symptoms, age, deprivation and BMI, children with one hour daily activity at baseline scored on average .29 units lower on emotional symptoms at one year follow up compared to those not achieving a minimum of one hour per day. Though small, this effect equated to a 9% reduction in emotional symptoms from baseline. In agreement with these findings, Neissaar and Raudsepp's (2011) longitudinal study assessing the relationship between changes in PA and depression over a two year period, showed that higher baseline PA was related to lower depressive symptoms at follow up. Furthermore, changes in LTPA across time were inversely associated with changes in depressive symptoms in 181 girls (aged 11-12 years).

Additionally, Sund, Larsson and Wichstrøm (2011), examined the longitudinal relationship between self-reported vigorous PA, sedentary time, depressive symptoms and emotional stress in adolescents aged 12 – 15 years over a 1 year period. Findings showed both high sedentary time and low vigorous activities at baseline predicted the chance of reporting higher depressive symptoms at follow up. The authors also noted that vigorous activity was associated with a reduction in depressive symptoms when stress levels were high, concluding that vigorous PA has a potential stress buffering effect. In contrast, Hume et al. (2011), using objective measurement, and Rethon et al. (2010) using self-reported measures showed no significant prospective relationships between physical activity and depressive symptoms.

Employing a different method to the prospective studies previously noted, Jacka et al. (2011) attempted to assess the longitudinal relationships

retrospectively. Participants (aged 20–97 years) were asked to recall childhood, (≤ 15 years of age) PA levels, and self-report lifetime prevalence rates of depression. Logistical regression analysis showed that lower childhood PA levels were associated with an increased risk of depression in adulthood when sex, age and adult PA were controlled (OR = 1.35, 95% CI: 1.01 – 1.78, $p = .04$). However, childhood depression was not accounted for in this model. The results of this study are dubious, as the assessment of PA through retrospective recall can result in inaccurate estimations, which will influence the observed relationships. In addition, although hard to obtain in a retrospective study, the use of recall to establish childhood depression could have altered the results observed, as childhood depression is thought to continue into adulthood (Paluska & Schwenk, 2000; Naicker et al., 2013). Therefore, failing to control for childhood levels of depression creates limited confidence within the results reported.

The mostly consistent results of the studies reported in this section provide an indication of the longitudinal nature of the relationships between PA and PWB. However the majority of studies discussed above only track changes over a short period of time (e.g. 2 years). Birkeland et al. (2009) attempted to assess the directional relationship between PA and depression using a 10 year longitudinal study design. Nine hundred and twenty four, 13 year old children, provided self-report LTPA and depressive symptoms at baseline, and follow up points at one, two, three, five six, eight and 10 years. Multivariate latent curve modelling showed that depression and PA covary, but that adding causality in either direction did not improve the fit of the model, providing inconclusive results. The authors concluded that no prospective relationships were apparent between LTPA baseline (13 years of age) and depression at follow up (23 years of age). The disparity between the results of Birkeland et al. (2009) and previous longitudinal studies may arise from the extended length of the measurement; PA may predict depression in the short term, but may be unrelated in the long term. Conversely, Burnet et al. (2013) recently undertook a 10 year prospective study with children (aged 12 – 13 years at baseline). Using latent growth curve modelling, results showed a significant relationship between past involvement in team sports and depressive symptoms in early adulthood.

However there was no association between past MVPA and depression in early adulthood.

Compared to the cross-sectional studies, the results of the longitudinal relationships are more consistent, even though there is still some uncertainty about stability over time. Interestingly, there appears to be consistency across measurement methods, though few studies employ objective measurement (e.g. Hume et al., 2009). As mentioned in the previous section, differences in measurement instruments and methods makes comparison between studies challenging.

The longitudinal relationship has not just been assessed in terms of PA and maladaptive components of PWB; Schmalz, Deane, Birch & Davison (2007) examined the temporal relationship between PA and GSW in girls using individual growth models whilst controlling for BMI and other confounders (socioeconomic status & age). Participants' PA and GSW were assessed through self-reported methods at ages 9, 11 and 13 years. The authors noted that PA accounted for 13% of the variance in the trajectory of self-esteem. Notably, PA at age 9 had a greater effect upon GSW at age 11 than PA at age 11 had upon GSW at age 13. The authors noted that this may have been due to social constraints perceived by girls transitioning to adolescence, though the ability to explore this further is hindered by the use of a GSW measure. A more in-depth view of the impact of PA upon self worth at different ages may have been obtained with the administration of a sub-domain self worth scale. Specifically, if social constraints are placed upon girls entering adolescence, PA may be more salient to girls PSW and sub-domains than to their GSW.

The Results of Schmalz et al (2007) also showed that the effects of PA upon GSW increased as BMI increased; at one standard deviation (SD) above the BMI mean, a one unit increase in PA at age 9, was associated with a .13 unit increase in self-esteem at age 11, while at two SD above the mean BMI, a one unit increase in PA increased GSW at age 11 by .20 units. The authors suggested that this may be due to potential weight loss and social benefits available with PA, which may be more available to those with high BMI z-scores.

Beyond these findings, Schmalz and colleagues (2007) noted that self-esteem had predicted none of the variance in future PA and concluded that the direction of the relationship was more in the direction of PA predicting self-esteem, providing support for the personal development hypothesis (Sonstroem, 1997b). This is in contrast to the results of cross-sectional studies previously discussed (Crocker et al., 2000; Raudsepp et al., 2002) that noted self-esteem accounted for some of the variance in PA. The contradictory findings however are likely due to the measurement points included in each study. While Crocker and Raudsepp only examined the variance predicted in sub-domains of self-esteem at a single time point, this predictive ability may not track across time. Additionally, while GSW was examined by Schmalz et al. both Crocker et al. and Raudsepp et al. examined the sub-domains of self-worth. Relationships with sub-domains may be stronger than relationships with GSW (Spence et al., 2005).

In line with this, Inchley, Kirby and Currie, (2011) examined the changes in physical self-perceptions (specifically perceptions of sports competence and PSW) alongside GSW and self-reported PA in children transitioning from primary to secondary school. PA was associated with high levels of perceptions of sports competence at all time points (last year of primary, second and fourth year of secondary school) for both boys and girls, supporting the findings of previous cross-sectional studies (e.g. Raudsepp et al., 2002). Additionally, PA was associated with high PSW at time points two and three in boys and all time points in girls, while GSW was only related to PA in girls at time point one and three. Alongside baseline levels of PA, perceptions of sports competence predicted PA in boys at measurement points two and three. While in girls, higher perceptions of sports competence increased the odds of being active at time points one and two. Additionally, PSW predicted activity at time point two and three. GSW did not predict PA for either sex or any time point; agreeing with the statement made by Spence et al. (2005) that examining sub-domains may provide more of an effect. It appears that the results of Inchley et al. (2011) support the motivation hypothesis of the self-esteem, PA relationship. More recently, Birkeland, Melkevik, Holsen & Wold (2012) examined the trajectories of GSW from early adolescence to early adulthood; the authors reported that frequency of physical activity (self-reported) at age 13, was significantly related

to GSW at age 14 ,15 ,18 ,21 and 23 years (r values between .010 - .140). Despite being significant ($p < 0.05$), r values reported between PA at age 13 and GSW at follow up are weak, with the strength of associations fluctuating slightly between follow up measurements.

In an examination of both aspects of the bi-directional relationship between PA and PSW, Raudsepp, Neissaar and Kull (2013) assessed what was termed the reciprocal effects model (REM; similar to the motivational approach), using self-reported PA measures in girls (mean age = 12.3 years) at three time points (approximately 12 months apart). Their findings supported the REM, showing higher levels of previous PSW lead to high subsequent levels of PA at future time points, and higher initial levels of PA lead to higher levels of PSW at subsequent time points. Further, the authors noted that this study provided more definitive evidence than previous studies employing cross-sectional methodology demonstrating the PWB benefits of PA.

2.4.1. Summary of longitudinal relationships

Results of longitudinal research generally demonstrate a positive relationship between PA and PWB in children, showing reduced negative states and increased positive states. The effect of different measurement methods did not appear to effect relationships, however only one longitudinal relationship noted above employed an objective measurement. Rather, variations in self reporting methods, and time between baseline and follow up measures, may result in some contrasting relationships.

The studies included within this section are generally well powered, with large sample sizes used, which may explain the statistical significance of small changes observed in some studies (e.g. Wiles et al., 2008), though Neissaar & Raudsepp (2011) reported longitudinal relationships in a smaller sample ($n = 181$). The time between baseline and follow up may also impact upon the results observed, for example, Birkeland et al. (2009) showed no prospective relationships, yet studies employing shorter follow up periods tend to show a prospective link. This in turn may influence the results observed, as relationships may not track across longer time periods, where there is greater

risk of bias from confounding variables. The lack of studies using long term follow up is a limitation of the PA and PWB findings.

Additionally, longitudinal analysis provides limited information pertaining to causality; therefore it is still unclear whether the manipulation of time spent in physical activity can affect PWB. In order to fully understand the causal link between PA and PWB, intervention trials are necessary.

2.5. Physical activity experimental trials in children and adolescents

Experimental trials are used to determine causality through the manipulation of the exposure measure. Ho et al. (2008) have suggested that internal validity and strength of evidence is increased with the use of experimental designs, in particular when RCT are undertaken. In respect of this, the authors proposed a hierarchy of evidence evaluation, whereby RCTs reside at the top of the hierarchy, followed in descending order by prospective and retrospective cohort studies, case-control, cross sectional and finally case series designs (Ho et al., 2008).

Despite a trend in the evidence pointing to a relationship between physical activity and PWB variables in children and adolescents, relatively few studies have attempted to manipulate physical activity in order to assess changes in PWB. A selection of studies employing experimental designs are outlined below.

An early study undertaken by Norris, Carroll & Cochrane (1992) recruited 60 adolescent participants to a 10 week PA intervention program, with an additional 20 pupils acting as the control group. Participants in the intervention arm were randomised to a high intensity (n=22), moderate intensity (n = 19) or flexibility (n = 19) training group. Each group undertook two supervised training session per week, lasting 25 – 30 minutes for the duration of the intervention. The high intensity group conducted aerobic exercise at 70 – 75% of maximal heart rate, while the moderate intensity group maintained 50 – 60% of their maximal heart rate. Results demonstrated a significant reduction in anxiety levels for the high intensity group compared to the moderate group after completion of the intervention, while the flexibility group showed a significant

increase in depression. The authors summarised that higher intensity exercise was more beneficial for PWB in healthy adolescents. Despite these findings, a number of limitations are present that may impact the observed results. Firstly, all three intervention groups were recruited from one school, whilst the control group was recruited from a different school, though the authors attempted to control for baseline differences between groups by including baseline measures as covariates within analysis. In addition, although the authors acknowledge that there may have been differences between groups for participants taking part in activities outside of the intervention, this was not measured or controlled for, meaning that systematic differences may have been apparent between groups, particularly as the control group was recruited from a different school. Furthermore, those participants undertaking the flexibility training were aware of the other interventions, therefore any increase in depression may have been a result of not being included within the moderate / high intensity training groups.

More recently, Bonhauser et al. (2005) employed a quasi-experimental school based intervention in which adolescents of low socio-economic status in one school were randomly assigned to take part in three 90 minute sessions ($n = 98$) or one 90 minute session ($n = 100$) of PA per week for one school year in an attempt to reduce anxiety and depression and increase GSW. Those undertaking more frequent PA showed a 13.7% reduction in reported anxiety and an increase of 2.3% in self-esteem compared to 2.8% and 0.1% (respectively) for the comparison group. These results remained significant after controlling for BMI, blood pressure, out of school PA and academic performance. Interestingly, the intervention group increased maximum oxygen uptake by 8.5% compared to an improvement of 1.8% in the low activity group, demonstrating a potential mechanism through which changes may occur. No intervention effects were apparent for depression. The results reported by Bonhauser et al. (2005) demonstrated an overall effect of the intervention, which is in contrast to other intervention studies. For example, Schneider, Dunton & Cooper (2008), attempted to increase PSW through a nine month, non-randomised school based intervention. Sedentary girls assigned to the intervention group undertook 60 minutes of PA on 4 days a week and 60 minutes on the remaining week day was dedicated to lectures or discussions focusing on the benefits of PA. A wide range of activities were undertaken as part of the

intervention, including dance, yoga, swimming and Tae Bo, amongst others. Alternatively, the control group continued with the normal physical education classes at school. PA (via self-report), physiological (Fitness, BMI, Body fat) and psychological variables (GSW, PSW, & PSW sub-domains) were assessed at three time points over an academic year (approximately 9 months). Intention to treat analysis showed that fitness and time spent in vigorous PA significantly increased for the intervention group, while no similar changes were apparent in the control group. Interestingly, no time by group interactions were obtained in relation to psychological variables, rather both the intervention and the control group increased in perceptions of sports competence, appearance, physical strength, PSW and GSW. Changes in GSW were greater for those who had increased cardiovascular fitness in the intervention group, which does agree with the findings of Bonhauser et al. (2005).

Difference between the intervention effects noted in the above studies may have occurred due to the different sample demographics (e.g. low SES). In addition, large differences are apparent between intervention activities including duration, intensity and frequencies of PA. In addition to this, Bonhauser et al. used a randomised design in which classes within one school were assigned to an intervention or control group. Schneider et al. employed a non-randomised design in which schools were assigned to either intervention or control, from which girls volunteered to take part in the study; the voluntary nature of this study could provide an explanation as to why both intervention and control groups increased in self worth. The differences between these studies limit the comparability and make it difficult to draw conclusions regarding the dose of PA that may benefit PWB.

The results of Schnider et al. (2008) do however provide partial support for those reported by Huang, Norman, Zabanski, Calfas and Patrick (2007), who examined changes in self-esteem and body image in adolescents undertaking a computer supported, home based intervention to increase PA and alter nutritional behaviour. The study was part of a one year RCT designed to increase PA, reduce sedentary and improve dietary behaviours. The intervention group received informational documents and counselling sessions to try and increase PA. The authors noted no effect for self-esteem in girls, yet

boys in the intervention and control group reported improvements in self-esteem across time. The study did not however report whether physical activity had increased in the intervention group, therefore the interpretation of these results is limited, as the effectiveness of the manipulation is unknown.

The studies noted above have examined the intervention effects of PA upon various aspects of PWB, in 'healthy' populations. In contrast to this, Petty, Davis, Tkacz, Young-Hyman and Waller, (2009) examined the mediating effect of self-worth upon the PA and depression relationship in 207 overweight, sedentary children, aged 7 – 11 years. Participants were randomly assigned to either a low dose (20 minutes per day) or a high dose (40 minutes per day) of intermittent aerobic activity or a control group for a 13 week intervention. Post intervention, those assigned to the high dose group had significantly lower depression, higher GSW, PSW and perceptions of physical appearance than the control group whilst controlling for baseline measures and change in BMI z score. No differences were apparent between the low dose and control and the low dose and high dose groups. The authors noted that this was indicative of a dose response effect, as differences between high and low groups approached significance. However, the increase in GSW was only apparent for white children. In terms of the mediation, improvements in perceptions of physical appearance partially mediate the effect of the interventions upon depressive symptoms. GSW was also found to mediate the relationship.

Additionally, Daley, Copeland, Wright, Roalfe and Wales (2006) assessed the impact of a PA intervention upon depression, and physical self - perceptions in an obese sample of adolescents. Participants aged 11 to 16 years with a BMI SDS greater than 3.5 were randomised to receive either exercise therapy, exercise placebo or a control. Both exercise groups undertook 30 minutes of exercise, three times per week for eight weeks. For the therapy group, this consisted of intermittent aerobic activities undertaken at a moderate intensity (40 – 59% heart rate reserve). The exercise placebo, undertook light body-conditioning / stretching, exercises, while the control continued with their daily routine as normal. Significant differences were apparent in children's PSW scores post intervention, 14 weeks and 28 weeks after baseline (mean differences: 0.21, 0.26 and 0.23 respectively). Improvements equated to 6.5%

for PSW in the exercise therapy group. Further improvements were apparent for GSW and perceptions of physical strength, which equated to 12.3% and 7.8% respectively for exercise therapy. Self-reported PA also increased for the exercise therapy group, though no differences were apparent between groups for depression post intervention. The authors concluded that exercise was a successful in increasing self-esteem in obese youth.

In an examination of the effects of PA upon another clinical population, Dopp, Mooney, Armitage & King, (2012) employed a 12 week exercise intervention in adolescents aged 13 – 17 years who had previously been diagnosed with depression. Participants were volunteers, who met the criteria of having depressive symptoms and low physical activity. The intervention involved three supervised exercise sessions per week for the first two weeks; this was subsequently reduced to one supervised session and two unsupervised sessions for the remainder of the program. Results showed an overall intervention effect for reduction in depressive symptoms (ES = 2.0) from pre to post intervention. Additionally, self-reported PA increased across the intervention (ES = 1.6), continuous increases were apparent at three months follow up (mean score = 1.46 (baseline); 2.10 (post intervention); 2.21 (3 month follow up)), at which point, depression scores had continued to decrease (mean score = 48.9 (baseline); 28.5 (post intervention); 25.9 (3 month follow up)). Intervention effects cannot be attributed solely to the exercise itself, as no control group was included in the study, therefore other confounding variables not measured within the study may have resulted in the reduction of depression. In addition, the use of self-reported measures post intervention may explain the continued increase in PA, as aspects of social desirability may have influenced participant response, though during the intervention, accelerometers were used to verify independent activity sessions. In addition to this, participants were receiving a mixture of anti-depressive treatments including psychotherapy and medication at various stages throughout the study, which may have influenced the findings. Finally, the intervention undertaken appeared to vary for half of the participants; the authors noted an expected aerobic workout for the first six participants, yet when this was not achieved, the remaining participants were set target HRs to meet during their supervised sessions. Therefore the

interpretation of the effectiveness of the intervention becomes confusing, especially in terms of the intensity required to produce these effects.

The majority of studies already discussed employ aerobic exercise programs to improve PWB. Lau, Yu, Lee & Sung (2004) however examined the effect of a six week resistance training and nutrition programme. As part of an obesity treatment trial, participants were randomised to receive either a resistance training plus nutrition program or a control receiving just the nutrition strand of the trial. The resistance training consisted of one hour of exercise, performed at 70 – 80% of the individual's one repetition maximum on multiple resistance machines. The nutrition program consisted of educational sessions and behaviour modification, and recommended diets based on individual energy requirements and activity levels. A decrease in anxiety and an increase in depression were apparent for the exercise group post intervention, however these differences were non-significant. The results support those of Daley et al. (2006) who reported no intervention effects for depression in an obese sample, however disagree with the finding of Petty et al. (2009). These differences may be due to the use of clinical samples in Daley et al. (2006) and Lau et al. (2004) as opposed to a school sample employed within the Petty et al (2009) study.

2.5.1. Summary of experimental designs

As with the cross-sectional and longitudinal studies, the experimental studies show inconsistent findings as to whether changes in PA can increase PWB. In general the findings seem to suggest that PA benefits aspects of self-esteem, with only a few studies demonstrating an effect upon depression and anxiety. Importantly, the studies presented above provide information on a host of populations; however the use of both healthy populations and clinical samples makes comparison between studies difficult. Sample size also varies largely between studies, ranging from $n = 13$ to $n = 207$.

Those studies using a randomised study design produce stronger evidence than those who choose not to or cannot randomise. Those with non-randomised group designs may have systemic differences between groups prior to the intervention, which in turn may alter the observed result, and reduce the strength of the findings reported. In addition, the strength of evidence obtained

from studies with no comparison group is limited, as the effects cannot be attributed solely to the intervention (e.g. Dopp et al., 2012).

The length of intervention periods also varies dramatically between studies; ranging from eight weeks (Daley et al., 2006) to 1 year (e.g. Huang et al., 2007), again limiting the conclusions that can be drawn relating to the length of time needed to affect PWB, with both short and longer studies demonstrating equivocal findings regarding the effects of intervention programme. In comparison to the longitudinal studies, follow up times are short and it is possible that changes in PWB take longer than a few weeks or months.

In addition to the potential limitations outlined above, the treatment of loss to follow up and subsequent analysis is not consistent across trials; whilst the majority of studies performed an intention – to – treat analysis, a few studies conducted completers only analysis which may over-estimate the observed results (e.g. Norris et al., 1990) whilst others failed to report the use of either analysis (e.g. Lau et al. 2004).

Finally, studies controlled for different covariates (e.g. BMI, SES, and age) and examined different modes and doses of activity; combined with the above limitations it is difficult to draw conclusions regarding the effect of PA upon PWB. Despite this, a number of meta-analytical reviews have been conducted in an attempt to summarise the reported effects of PA and PWB.

2.6. Meta-analyses

Meta-analysis are employed when attempting to summarise data pertaining to a specific topic, this methods integrates the findings of several studies in order to establish greater power than can be achieved through individual studies (Eggerm, Davey-Smith & Phillips, 1997). Although there are obvious advantages of this type of analysis, differences between studies included within the meta-analysis (heterogeneity) may limit the validity of the findings (Ho et al., 2008).

Meta-analytical reviews of early intervention studies showed a significant effect of PA upon depression (North, McCullagh & Tran, 1990; Calfas & Taylor, 1994; Craft & Landers, 1998), anxiety (Petruzzello et al., 1991; Calfas & Taylor,

1994) and self-esteem (Greuber, 1986; Calfas & Taylor 1994) in both adults and children. More recently, a Cochrane review conducted by Larun, Nordheim, Ekeland, Hagen and Heian (2009) assessed 16 studies to establish the effect of PA upon anxiety and depression in children, the majority of which were published between 1983 and 2005 (only 2 were published post 2000). The included studies varied widely in terms of intervention period (6 to 40 weeks), type of activity (including walking, running, cycling, weightlifting) and comparison groups (no treatment, low intensity PA and psychotherapy) as well as measurement methods of depression and anxiety. When compared to no intervention, exercise showed a non-significant trend towards reduced anxiety (ES = -0.48, 95% CI: -.97 to - 0.01) and a significant effect for depression (ES = -0.66, 95% CI: -1.25 to -0.08) but the authors concluded that there was little difference between moderate and vigorous activity intensities. Additionally it was noted that despite small benefits apparent for the general population, there was a lack of sufficient evidence to examine benefits in clinical populations. Despite these findings, large heterogeneity values were reported for both anxiety and depression analysis ($\text{Chi}^2 = 20.93$, $\text{df} = 5$ ($P = 0.00083$); $I^2 = 76\%$; $\text{Chi}^2 = 19.98$, $\text{df} = 4$ ($P = 0.00050$); $I^2 = 80\%$, respectively), indicating large between study differences that limit the strength of conclusions that can be drawn.

An updated review of the depression and PA literature has recently been undertaken by Brown, Pearson, Braithwaite, Brown and Biddle (2013). The meta-analysis included nine studies, some of which have been described above (e.g. Norris et al., 1992; Bonhauser et al., 2005; Daley et al., 2006 and Petty et al., 2009). Overall, results revealed a significant but small effect of physical activity upon depression (ES = -0.26, 95% CI: -0.43 – 0.08), though the effect size appears much smaller than those reported by Larun et al. (2009). Again however, the heterogeneity score for this analysis was high ($Q = 19.84$, $I^2 = 59.68$). The authors therefore undertook sub-group analysis based on the moderator variables such as methodology, study design, gender, outcome measure and study quality etc.). Of note, those studies with a higher quality produced a larger ES ($k = 4$, ES = -0.39, 95% CI: -0.61 to -0.16) than those of a lower quality ($k = 5$, ES = -.145, 95% CI: -0.363 to 0.074). These results

demonstrate the impact the studies with low quality can have upon the findings of a meta-analysis.

Ekeland, Heian and Hagen's (2005) meta-analysis of the effect of PA upon self-esteem demonstrates an overall mean difference ($k = 8$) of 0.49 (95% CI: 0.16 – 0.81) in favour of exercise improving self-esteem compared to no intervention. Additionally, when exercise was included as part of a more comprehensive intervention ($k = 5$), average mean difference was 0.51 (95% CI: 0.15 to 0.88). Both meta-analyses revealed around 10% difference between the treatment and control conditions, demonstrating a positive effect for exercise in improving self-esteem. However, a number of limitations are apparent with this meta-analysis; only studies relating to GSW were included in the meta – analysis, resulting in a small number of included studies and restricting the conclusion that can be drawn.

Including both maladaptive and positive aspects of PWB, Ahn and Fedewa (2011) performed an updated meta-analysis, which differentiated between randomised control trials (RCT) and non-RCT. In RCTs, interventions significantly reduced depression ($k = 14$, $ES = -0.41$, $SE = 0.13$), anxiety ($k = 16$, $ES = -.35$, $SE = .18$) and increased self-esteem ($k = 26$, $ES = .29$, $SE = .08$). However in non RCTs, PA interventions only increased self-esteem ($k = 16$, $ES = .78$, $SE = .28$). Interestingly, Ahn and Fedewa (2011) found that equal effects were apparent for children deemed overweight / obese and normal weight in both RCTs and non-RCTs. This finding may potentially overcome the issues noted previously for the confounding effect of weight status.

2.6.1. Summary of meta-analyses

Despite the inconsistent findings presented throughout this chapter, meta-analytical reviews support the beneficial effects of PA, usually in the form of exercise, upon PWB of children and adolescents. Effects are generally small and whether they are sustained over time is unclear. In addition, the findings are generally from heterogeneous samples with large I^2 values reported. I^2 ; the total variation in treatment effect estimates due to between study differences (Higgins & Tompson, 2002), is generally moderate to high (> 50%; Higgins, Thompon, Deeks & Altman, 2003). In addition, a number of analyses include

low quality studies which can influence the observed effects (Ho et al., 2008). Despite these limitations, the use of meta-analysis can provide an overview of the available evidence pertaining to PA and PWB in children. The dose of activity needed in order to obtain these benefits however is still unknown.

2.7. Dose – response

Government guidelines currently recommend that children achieve at least 60 minutes of MVPA every day, with vigorous activities incorporated on at least three days of the week, and a reduction in the time spent sedentary in order to obtain health benefits (Department of Health, 2011). Evidence from the literature presented above suggest a dose-response effect with higher frequency (e.g. Bonhauser et al., 2005; Schneider et al., 2008) and longer durations (e.g. Petty et al., 2009) of PA resulting in greater PWB benefits. Ahn & Fedewa (2011) however show an effect for activities performed once (RCT ES = -.57; Non-RCT ES = -.81) or twice a week (RCT ES = -.37), with the largest effect for durations of more than 33 hours over the course of the intervention for RCTs (ES = -0.55) but for shorter time in non-RCT (< 20hours; ES = -1.84). The results therefore still appear inconsistent regarding duration and frequency, though evidence relating to intensity is less clear, with only a few studies examining the importance of intensity (e.g. Norris et al., 1992). The adult literature is also unclear. A literature review of 67 studies including observational and experimental designs reported that higher intensities of PA may be more effective than lower intensities for depression risk reduction (Teychenne, Ball & Salmon, 2008), yet the exact intensity is unknown. Dunn, Trivedi, Kampert, Clark & Chambliss, (2005) suggested that moderate intensity activity, equivalent to the public health dose ($17.5 \text{ kcal}\cdot\text{kg}^{-1}\cdot\text{week}^{-1}$) showed the greatest effect, yet others suggested that vigorous activity was the most beneficial (E.g. Wise, Adams – Campbell, Palmer and Rosenberg, 2006). Furthermore, Lindwall, Rennemark, Halling, Berglund and Hassmen (2007) reported differential effects of intensity for elderly men and women suggesting sex differences may be an important factor when considering the effect of activity intensities on PWB. The majority of these studies have focused upon the appropriate intensity to reduce depression, while little attention has been focused on other areas of PWB, though Osei – Tutu and Campagna (1998)

found that moderate activity had the most beneficial effect upon anxiety. Yet the intensity required for self-esteem is also unclear.

Within the children's literature, a number of cross-sectional studies indicate that vigorous or MVPA is related to better PWB, however those studies using self-report instruments may not be capturing true estimates of moderate or vigorous physical activity. The few studies that have employed objective measures indicate a balance between time in higher intensities and low activity intensities (Parfitt et al., 2009) and less time in sedentary behaviours (Page et al., 2010) is related to better PWB. Ahn & Fedewa reported that 'intense' PA significantly reduced psychological problems in RCTs ($k = 26$, $\bar{d} = -.27$), though larger effect sizes were apparent for RCTs that did not report intensity ($k = 52$, $\bar{d} = -.41$). In non-RCTs, only moderate intensity showed a significant decrease in psychological problems, however few studies were included ($k = 3$, $\bar{d} = -1.89$). Thus, the dose-repose nature of PA for PWB is unclear, with more research using accurate assessment of activity intensity needed before conclusions can be drawn.

2.8. Discussion of the PA and PWB literature

The children's literature presented above indicates a general trend towards a positive relationship between PA and PWB and the effect of PA on PWB variables, though inconsistent findings have been noted throughout. There appear to be many gaps in the literature preventing a clearer understanding of the nature of the relationship between PA and PWB in children (Hamer et al., 2009). The majority of research has been undertaken with adolescent populations, yet prevalence rates of mental health disorders in children that can track into later life indicate that more research should focus on PA and PWB during childhood (Neisser & Raudsepp, 2011). Furthermore, it has been noted that symptoms of depression are occurring at increasingly younger ages (Weissman & Klerman, 1992). Additionally, children are thought to develop awareness of their competence in life domains from the age of 8 (Harter, 1982), indicating that bolstering self-esteem from this age may help to buffer against stress, depression and improve overall PWB.

There also seems to be limited research addressing anxiety compared to depression and self – esteem. However, this may be due to publication bias, with studies failing to find significant relationships not being published. As depression and anxiety often accompany each other it makes sense to research them together.

Finally, and most importantly, at the present time, it appears that limited conclusions can be drawn about the relationship between PA and PWB due to measurement issues arising from the studies focusing on these constructs (Parfitt & Eston, 2005). This may not only impact on the relationships observed but hinders the progression of research concerned with dose-response relationships and the effects of different intensities. As previously mentioned, a number of studies are limited due to the nature of the exposure measure employed. The most frequently used measurement methods appear to be self-report, which are subject to recall bias (Biddle et al., 2011), social desirability (Crown & Marlow, 1984), and common method variance (Padsakoff et al. 2003) due to aspects such as general affectivity (Watson & Clark, 1984). These issues will be thoroughly discussed in chapter three (p. 97). Each of these methodological issues will impact the relationship reported between PA and PWB, potentially overestimating the magnitude of the relationship. These problems may account for the apparent variations in the cross-sectional studies focusing on depression, as there was a suggestion that relationships may be attenuated with the use of objective measures.

Furthermore, studies using self-reported measures capture different behaviours depending on the phrasing of the questions. For example, Brosnahan et al. (2004) assessed moderate activity with the question '*On how many of the past seven days did you participate in physical activity for at least 30 minutes that did not make you sweat or breath hard, such as fast walking, slow bicycling, skating, pushing a lawn mower or mopping floors*' (p. 819), while Kantomaa et al. (2008) assessed MVPA through the question '*How much do you participate in brisk physical activity outside school hours?*', defining brisk as causing some sweating and becoming out of breath. This demonstrates how different studies measure different behaviours, yet claim to be assessing the same construct e.g. moderate intensity physical activity. A number of studies

employ validated self-report questionnaires such as the 3-day physical activity recall questionnaire (3 day PAR; Pate, Ross, Dowda, Trost & Sirard, 2003), which may provide more accurate estimations of physical activity than single item measures. Interestingly, some studies have employed proxy-measures of PA whereby parents or teachers indicate PA levels of children. Errors associated with this method are also discussed in chapter three; however the effects different responders have upon the relationships observed can be seen in the results of Tomson et al. (2003). The authors examined the relationship between PA and depressive symptoms in 933 children aged 12; PA was measured using two proxy reports – a teacher completed one and a parent completed one, alongside a self-reported children's measure. The parents and teachers were asked to rate children as active or inactive, while children were asked whether they play sports outside of school. Significant differences for depression were apparent between active and inactive boys, with active boys having fewer depression symptoms for parent proxy reports. However significant differences for depression were not apparent between active and inactive boys when using the teacher PA reports, yet self-reported PA agreed with the results from the parent report. For girls on the other hand, differences between active and inactive children for levels of depression were only significant for the teacher PA reports. These results were reported 10 years ago, yet the impact that different measurement methods can have upon the noted relationships between PA and PWB is still overlooked.

The use of different measures of self-report and the associated misclassification hinders the progression of research into the exact nature of the relationship between the frequency, intensity, duration and volume of specific physical activities and PWB. Some studies have attempted to overcome the problems associated with self-report by using more precise objective measurement methods such as accelerometers. Whilst these methods provide a more accurate estimate of PA and time spent in certain intensities, few studies have utilised these measures. Those that have, failed to take into account the multitude of methodological issues that accompany accelerometer use (Crocker et al., 2006). The methodological issues, such as cut-point classification, epoch length, wear-time requirements and compliance, discussed in chapter three (p. 85 to 94), can result in misclassification of time in activity

intensities and different biases being introduced, both of which may alter the relationship between PA and PWB.

The present thesis aims to explore how methodological considerations of objective measures can impact the relationships observed between PA and PWB variables in children, and explores methods that may overcome some of the apparent problems. The following chapter presents a discussion of the strengths and weaknesses of different measurement methods available for the assessment of PWB and PA in children.

CHAPTER THREE

METHODOLOGY

The focus of this chapter will be on methods for measuring the construct of psychological well-being (PWB) and physical activity behaviour. The measurement of PWB will be addressed in terms of methods available, advantages and disadvantages and the subsequent selection of measures for this thesis. Following this, physical activity will be considered; definitions of physical activity and related constructs will be provided prior to addressing the measurement methods available along with the pros and cons of each method. The discussion will then progress into the potential to introduce misclassification and bias through various mechanisms before finally discussing the selection of the most appropriate measures for the population of interest.

3.1. Psychological well-being measurement

As noted in previous chapters, the term psychological well-being alludes to a multitude of concepts and has been defined in a number of ways ranging from two theoretical stand points; a hedonic view focusing on happiness and affect, and a eudaimonic view focusing on psychological functioning and the realisation of human potential (Ryan & Deci, 2001).

Early research from the hedonic stand point stated that *“an individual will be high in psychological well-being in the degree to which he has an excess of positive over negative affect and will be low in well-being in the degree to which negative affect is predominant over positive”* (Bradburn, 1969. p.9). Other researchers however have noted that well-being includes feelings and cognitions relating to life satisfaction (Park, 2004; Deci & Ryan, 2008; Winefield, Gill, Taylor & Pilkington, 2012); Deiner (2009) conceptualised well-being as, experiencing positive levels of pleasant mood states, whilst having low levels of negative emotions and a cognitive appraisal that one’s life is good.

Definitions from a eudaimonic view have focused on concepts such as autonomy; relationships and self-acceptance (Ryff, 1989) while some researchers have combined the two views; noting that well-being arises from

personal development and fulfilment, making a contribution to the community alongside feeling happy and satisfied (Shah & Mark, 2004). The WHO definition also encompasses a number of these points; they define well-being as a “*state in which an individual realizes his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community*” (WHO, 2001).

Recently, Dodge, Daly, Huyton and Sanders (2012) discussed the difficulties in defining such a construct; the authors noted that many of the previous definitions were in fact descriptions, focusing on certain dimensions of well-being. Following this the authors developed their own definition in which well-being is “*the balance point between an individual’s resource pool and the challenges faced*” (p.230) where both resources and challenges can be psychological, social or physical.

Though it is widely believed that the construct of well-being is multi-dimensional (Deiner, Scollon & Lucas, 2003), it appears to be a complex structure which proves difficult to define (Pollard & Lee, 2003) and has many measurable elements (Seligman, 2011). The disparity between definitions has resulted in well-being being measured in multiple ways; it can be examined in terms of a single construct (i.e. well-being), as individual aspects that contribute to well-being as a whole (e.g. depression), or through a combination of individual aspects (e.g. anxiety, depression & self-worth) that can provide an indication of overall PWB. Despite the vast amount of known constructs and combinations of constructs, the methods available to measure PWB, are limited to either self-report / proxy-report measures or via structured interview. The appropriateness of these measurement methods differ depending upon the population of interest (e.g. clinical / non clinical; adult / child). For example, a questionnaire may not be adequate to diagnose clinical depression (e.g. Hall, Hern & Fallowfield, 1999), although it may be used to screen for the presence of depressive symptoms (e.g. Silverston, 1994; Lowe et al., 2004), and indicate whether further evaluation of a participant is needed. Alternatively, a structured interview, based on the diagnosis and statistics manual for mental illness (DSM-IV; American Psychiatric Association), would allow for a clinical diagnosis to be made.

3.1.1. Self-report

Self-report methods for assessing PWB come in the form of questionnaires that measure either the frequency, or intensity of thoughts and feelings (Warr, 2012) in relation to the target variables. Questionnaires that assess overall well-being or individual constructs are available for both adult and child populations. Examples of self-reported measures that assess overall well-being include the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS; Tennant et al., 2007); which assesses subject's thoughts and feeling of most areas of well-being (Stuart-Brown & Janmohamed, 2008). Participants are asked to respond to 14 items on a likert scale of 1 – 5 about the frequency to which they experience the feelings described by each item. For example, participants answering item 6 '*I've been dealing with problems well*', would rate whether they felt this '*none of the time*', '*barely*', '*Some of the time*', '*often*' or '*all of the time*'. Item scores are summed and a rating between 14 and 70 is obtained, with higher scores representing better PWB. Other overall measures, with similar rating scales include, the World Health Organisation (WHO) – Five well-being scale (WHO-5; Bech, 1998; 2001); the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) and the Scales of Psychological Well-Being (SPWB; Ryff, 1989b), although many more exist. Each of these questionnaires assesses various aspects of well-being to gain a composite score of global PWB.

Alternatively, inventories examining single constructs are also available, which can be used in conjunction with other single measures to provide a general view of PWB. For example, the Beck Depression Inventory (Beck, Ward, Mendelson, Mock & Erbaugh, 1961) measures the intensity of depression symptoms (e.g. self-dislike, crying, irritability and loss of appetite etc.) on a zero to three scale in adults. Modifications to this questionnaire have enabled a child specific version; the Child Depression Inventory (CDI; Kovacs & Beck, 1977) to be developed. The items on the modified inventories usually follow the same theme as the original, but adapt the language to be more suitable for children, and remove items that may be inappropriate for the target age. Additionally, with respect to the CDI, the number of choices for each item

has been reduced, with items being scored on a 0-2 scale. Similar questionnaires are available for other individual constructs such as anxiety (The Worry Scale [WS; Wisocki, Handen, & Morse, 1986], The State-trait Anxiety Inventory [STAI; Spielberger et al., 1970]), self-concept (the physical self-description questionnaire [PSDQ; Marsh, Rihards, Johnson, Roche & Tremayne, 1994]), and mindfulness (Mindful Attention Awareness Scale [MAAS; Brown & Ryan, 2003]) along with numerous others for both adult and child populations.

Each version of a questionnaire must demonstrate adequate psychometric properties, primarily validity and reliability. Validity refers to whether a test measures what it is intended to measure (Field, 2009) and can be assessed in many ways depending on the type of validity to be established.

- Content validity; do the items represent the construct of interest?
- Construct validity; how well a measure reflects the underlying construct
- Criterion validity; how well a measure correlates with a 'gold standard' criterion measure.
- Concurrent validity; how well a measure correlates with another instrument that measures the same construct when measured at the same time

(Hubley & Zumbo, 1996)

In questionnaires, one of the most common ways to assess validity is through correlation analysis of the outcome with established questionnaires assessing similar constructs (Ratney & Jones 2007), providing an indication of concurrent validity. Alternatively, confirmatory factor analysis can be employed to determine the construct validity of a measure.

Reliability refers to the consistency of the method to reproduce similar results under the same circumstances (Vincent, 2005). With questionnaires this is usually assessed through internal consistency (Cronbach's alpha) and test-retest reliability, with an $\alpha \geq .70$ demonstrating high reliability (Nunnally & Bernstein, 1994; Streiner, 2003). Reliability can be influenced by many factors,

including the characteristics of the test, the subject being measured, and the person administering the measurement (Kohl, Fulton & Caspersen, 2000). Self-report measures may therefore be susceptible to factors influencing reliability, including social desirability and children's cognitive abilities which will be discussed in detail later in this chapter.

In an attempt to combat these problems, Goodman and colleagues (2000) suggested that questionnaires should be obtained from multiple informants, e.g. teachers and parents, to strengthen the information gained from individual informants and to give context to any potential issues that may arise. Although this may aid with clinical diagnosis of psychological disorders, within research settings it seems an unrealistic target, especially when using teachers, who are responsible for large numbers of children, as informants.

Self-reported measurement of psychological well-being is cheap and easy to administer, especially for research purposes that may include assessing large groups of children at one time. Furthermore, it is easy to assess both positive aspects of well-being along with maladaptive aspects, to gain an indication of well-being overall. Alternatively, structured interview techniques can be employed to assess some areas of PWB.

3.1.2. Interviews

Interview techniques are typically used in clinical environments to gain accurate and reliable diagnosis of psychological / psychiatric disorders (Segal, Hersen & Van Hasselt, 1994), however they can also be employed in research settings, providing training has been undertaken. Interviews such as the Structured Clinical Interview for the diagnosis and statistics manual (DSM-III) (SCID; Spitzer, Williams, Gibbon, First, 1992; First, Spitzer, Gibbon, Williams, 1995) provides a set format for clinicians and trained researchers to follow in order to diagnose psychological disorders. These types of interviews can be undertaken either face to face or via telephone (Rohde, Lewinsohn & Seeley, 1997). Although the SCID provides structured formats for diagnosis of all disorders contained within the DSM, alternative structured interviews are available for use by clinicians and researchers that assess individual disorders,

for example, the Structured Interview for Disorders of Extreme Stress (SIDES; Pelcovitz et al., 1997). As with questionnaires, each interview, is assessed for validity and reliability, and as with both methods, it is important to employ an instrument that is technically sound, and assesses the aspect of well-being of interest (Warr, 2012).

3.1.3. Choice of methods

Interviews are inappropriate for large scale research, due to the time and interviewer training required. Furthermore interviews are traditionally only used for assessing psychological disorders rather than the presence of positive states, which are fundamental to psychological well-being. Due to these drawbacks, throughout the following thesis PWB will be assessed using self-report questionnaires, specifically designed for children. Within the PWB and PA literature, researchers, have either assessed single aspects of well-being, such as depression (e.g. Motl et al., 2004) or have employed a series of individual construct measurements to gain a composite score of well-being, including maladaptive and positives state assessment (e.g. Parfitt et al., 2009), though others utilise a measure of overall well-being (e.g. Steptoe & Butler, 1996; Page et al., 2010)

During the first three studies of this thesis, in line with previous PA and PWB literature (Parfitt & Eston, 2005; Parfitt et al., 2009), three measures were used to assess children's PWB. In order to gain an indication of levels of depression, the Child Depression Inventory (CDI; Kovacs & Beck, 1977) was employed. The State-Trait Anxiety Inventory for Children (STAIC; Spielberger, 1973) was administered to assess anxiety and finally, a combination of the child and youth Physical Self-Perception Profile (PSPP-CY; Whitehead, 1995), and measures of physical (Whitehead & Corbin, 1991) and GSW (Harter, 1985) were administered. Difficulties, such as a high percentage of spoilt data, occurring through missing, extra or mixed-up observations were encountered with the administration of these questionnaires; consequently a singular measure of psychological well-being, namely the strengths and difficulties questionnaire (Goodman, Meltzer & Bailey, 1998), was administered to children in the final study.

3.2 Physical activity

3.2.1. Definition and classification

The terms physical activity, exercise and fitness are often used synonymously in everyday life. Though the concepts possess a number of common elements (Caspersen, Powell & Christenson, 1985), the fundamental aspects differ somewhat. Physical activity (PA) has been defined as '*any bodily movement produced by skeletal muscle that results in energy expenditure above resting level*' (Caspersen et al., 1985, p. 128). This definition encompasses everything one does in daily life and represents the most variable aspect of total energy expenditure (TEE), which is comprised of resting energy expenditure (REE), the thermic effect of food, and energy expenditure through physical activity (PAEE) (Montoye, Kemper, Saris & Washburn, 1996).

Exercise, which can subsequently be viewed as a subset of physical activity, is usually a planned and structured activity that is used as a means to an end; for example, one may take an exercise class to lose weight, improve a certain aspect of their health or train for a sport. Conversely, physical fitness is comprised of attributes relating to people's ability to perform physical activity, such as cardio-respiratory fitness, muscle strength and flexibility (Thompson et al., 2003) and can be increased through physical activity and exercise. Physical activity will be the focus throughout the following thesis.

3.3. Measurement of physical activity

Although physical activity can be comprised of multiple types of behaviour (e.g. washing the floor, running, digging in the garden, fishing etc.), each can be classified in terms of frequency of occurrence, duration of each occurrence and the intensity at which it is performed (Butte, Ekelund & Westerterp, 2012). Frequency is typically expressed as the number of days per week an activity occurs but can also include the number of times a day an activity occurs. Duration is measured in units of time and can be expressed as seconds or minutes per occasion, minutes per day or hours per week. Frequency and duration are often combined to estimate the total time spent per day or per week, performing a given activity (e.g. sum of the number bouts of walking in

the same day multiplied by the duration of each bout) or the sum of multiple activities performed over a fixed time period such as a week.

The measurement of intensity is more complex; it can be measured in absolute and relative terms and is an indicator of the physiological demand of an activity. Absolute intensity gives an overall estimation of intensity for activities without consideration of individual characteristics such as age and physical fitness (Welk, 2002). On the other hand, relative intensity refers to the percentage of maximum aerobic power used during exercise and is usually expressed as a percentage of maximum physiological intensity (e.g. % of heart rate max / reserve and / or % of $\dot{V}O_2$ reserve) (Thompson et al., 2003).

Energy expenditure can be expressed in terms of metabolic equivalents (METs), representing the amount of energy being expended as a multiple of the energy cost of rest, with 1 MET considered to represent the resting level of energy expenditure (Welk, 2005). MET values are normally estimated from measuring oxygen consumption, via indirect calorimetry, in laboratory settings as this is directly proportional to energy expenditure during physical activity (Wilmore & Costill, 2005). Determination of MET values can be based on measured resting, active and maximal oxygen consumption or through estimations of oxygen consumptions based on a standard resting metabolic values. A standard MET equates to $3.5 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ in terms of oxygen consumption, with activities using $7 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ equating to 2 METs (Welk, 2005) and so on. METs can be used to group activities by their intensity level, commonly used groups are; sedentary, light, moderate and vigorous (e.g. Puyau, Adolph, Vohra, & Butte, 2002; Treuth et al., 2004), the usual MET thresholds for these intensity are < 1.5 (sedentary), $1.5 - < 3$ (light), $3 - 6$ (moderate), > 6 (Vigorous) (Treuth et al., 2004).

Estimating physical activity intensity based on METs is not without its problems. The standard $3.5 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ of oxygen consumption being equal to 1 MET is usually applied to an adult population; however oxygen consumption at rest may be higher in some adults (e.g. women, older men and unfit populations) (Sheppard, 2001). This has been demonstrated by Byrne et al (2005) who reported small differences between males of different ages with the

same BMI (e.g. 20 years = $3.16 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$; 40 years = $3.08 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$) and males and females of the same age and BMI (e.g. $2.35 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ and $2.17 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$, respectively). Furthermore, the authors reported that the average measured resting $\dot{V}O_2$ of a 20 year old male with a BMI of $20 \text{ kg}\cdot\text{m}^{-2}$ was $3.16 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ whereas the measured resting value for a 20 year old male with a BMI of $40 \text{ kg}\cdot\text{m}^{-2}$ was $2.42 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$. Additionally, there was a mean difference of 32.6% between the resting $\dot{V}O_2$ values of 20 year old females with $20 \text{ kg}\cdot\text{m}^{-2}$ and $40 \text{ kg}\cdot\text{m}^{-2}$ BMI values. Overall the authors noted that the average, resting value differed from the standard $3.5 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ of oxygen consumption by up to 35% in a large heterogeneous sample.

Moreover, the resting oxygen consumption for children also differs from that of adults (Puyau et al., 2002); children are likely to have higher rates of energy expenditure expressed per kg of body mass than adults when performing similar tasks, therefore an activity given a value of 2 METs could equate to $7 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ for an adult but $9 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ for a child (Ridley & Olds, 2008). Compendium tables estimating absolute adult and child MET values for a variety of activities have been developed (e.g. Ainsworth et al., 2000; Ridley & Olds 2008; Ainsworth et al., 2011). These tables allow the energy cost of physical activities to be estimated via observation or self-report rather than indirect calorimetry. Examples of activities included in the compendium are given below, along with a demonstration of the different METs for adults and children performing the same activities, examples in this table are based on adult 1 MET values of $3.5 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ and mean predicted child values of $4.59 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ for 1 MET (Ridley & Olds, 2008).

Table 3.1: Adult and child Activity MET values from Ridley and Olds (2008)

Activity level	Activity	Adult METs	Child METs
Sedentary	Television Viewing (lying down)	1.0	1.1
	Colouring (crayoning)	1.5	1.4
Light	Washing Dishes	2.3	1.9
	Playing the cello	2.4	1.7
Moderate	Playground games	5.0	4.9
	Shovelling	5.0	4.3
Vigorous	Skipping (Rope jumping)	8.0	8.3
	Swimming Front crawl (1m·s ⁻¹)	8.0	9.9

Alternatively, relative energy expenditure, based on percentages of oxygen uptake reserve ($\dot{V}O_{2\text{ res}}$) or heart rate reserve (HRR) / max (HRmax), allows for consideration of individual characteristics, such as fitness and age when assigning physical activity intensities. HRR and $\dot{V}O_{2\text{ res}}$ are calculated in much the same way, with the resting values for HR and $\dot{V}O_2$ being subtracted from the maximum values and calculating the percentage of difference between the two. (Epstein et al., 2001; Swain, Leutholtz, King, Haas & Branch, 1998). The American College of Sports Medicine (ACSM; 2006) assigned the relative energy expenditure gained from these values to the standardised descriptions of intensities, as outlined in table 3.2.

Table 3.2 Classification of PA intensities by relative energy expenditure (Adapted from ACSM, 2006, p. 4)

Intensity	% VO₂ reserve % HR reserve	% HR max
Light	20 – 39	50 – 63
Moderate	40 – 59	64 – 76
Vigorous	60 – 84	77 – 93
Very hard	≥ 85	≥ 94

Expressing intensity relative to maximum capacity has numerous benefits in terms of assessing physical activity intensity, for example the use of relative intensity can aid comparison between groups, and in turn ensure that individuals are obtaining the same physiological stress levels. The use of absolute values, e.g. where an absolute value 4 METs given to an activity, may actually be requiring substantially different physical effort depending upon the person performing the activity (Lee, Sesso, Oguma & Paffenbarger, 2003). Further, the use of relative intensity may be more important when prescribing exercise intensities to improve cardio-respiratory fitness (Hui & Chan, 2006) and in targeting specific health outcomes rather than monitoring physical activity levels.

3.3.1. The importance of accurate physical activity measurement

The ability to accurately measure physical activity is important for a number of reasons, which have been outlined by Bauman, Phongsavan, Schoeppe & Owen (2006), and include the surveillance of population PA prevalence across the life span and the identification of population trends (Welk, 2002). Physical activity measurement can also be used to screen people for low physical activity and the need for intervention (Rennie & Wareham, 1998). Furthermore, measurement of PA is needed to assess the effectiveness of interventions designed to increase PA or decrease sedentary behaviours, understand correlates and determinants of PA; including biological, psychological, social, and environmental aspects (Armstrong & Welsman, 1997). Finally, accurate measurement is important to understand the protective and therapeutic effects

of physical activity on both physical and psychological health (Rennie & Wareham, 1998).

Several ways of measuring physical activity exist, all of which possess advantageous and disadvantageous characteristics. The suitability of different methods depends upon a multitude of factors, including purpose of study, age of participants, sample size, desired outcome and budget (Loprinzi & Cardinal, 2011).

Until recently, measures of physical activity in epidemiological, surveillance and intervention studies have relied on imprecise self-report of physical activity. Whilst this has been adequate for demonstrating overall associations between physical activity and health outcomes it has prevented advances in knowledge of issues of dose-response (Rennie & Wareham, 1998) and the differential effect of different types of physical activity on specific health outcomes. The development of more precise objective measures of physical activity has revealed that the true association between physical activity and health has been underestimated, population prevalence has been overestimated and the effectiveness of interventions to increase physical activity has been attenuated toward the null (Basterfield et al., 2008). The advantages of more precise measurement of physical activity include:

- Increased specificity of physical activity exposures for specific health outcomes
- More accurate estimates of effect sizes
- Greater understanding of dose response issues
- Greater ability to detect temporal trends in physical activity
- Greater sensitivity to the effects of interventions
- Reduced sample size requirements due to more precise exposure and outcome measures leading to improved cost effectiveness

(Rennie & Wareham, 1998)

Children's physical activity is notoriously hard to measure; in 1995, Bailey and colleagues demonstrated the ephemeral nature of children's activity, reporting that around 95% of activity bouts last less than 15 seconds, with

low/moderate and vigorous activity bouts lasting an average of 6 seconds and 3 seconds respectively. This highly transitory activity pattern makes capturing activities, especially those of higher intensities, difficult and highlights one of the major issues facing researchers. Therefore, ensuring the most appropriate method of measurement is crucial when attempting to establish relationships with health variables.

A number of methods for measuring energy expenditure and physical activity in children are discussed below along with the benefits and costs of each method.

3.3.2. Measures of Energy expenditure

3.3.2.1. Doubly labelled water

The use of doubly labelled water (DLW) provides researchers with precise measurement of TEE over a period of time by using biochemical markers to reflect the rate of metabolism in the body (Starling, 2002). This measurement method involves participants ingesting isotopes of hydrogen (^2H ; deuterium) and oxygen (^{18}O) as a liquid solution ($^2\text{H}_2^{18}\text{O}$) (Loprinzi & Cardinal, 2011). The hydrogen isotope is then depleted through water loss, while the oxygen is depleted through both water loss and respiration, consequently the difference in the depletion rates of the isotopes in water (measured via urine samples) provides a measure of carbon dioxide production (Speakman, 1998) from which TEE can be calculated (Schoeller, 1999).

DLW provides an accurate measure of TEE (Sirard & Pate, 2001); however, the method possesses several limitations. Primarily, isotopes are expensive and difficult to obtain (Dale, Welk & Matthews, 2002), there may be a high participant burden if urine samples are to be frequently collected (Kohl et al., 2000) and what is more, physical activity energy expenditure cannot be separated from TEE. However, by combining DLW with indirect calorimetry (discussed below) to gain measures of resting metabolic rate, accurate assessments of physical activity energy expenditure can be made (Starling, 2002). Nevertheless, the combination of the two techniques provides an overall estimate of the energy expended through physical activity rather than a

measure of patterns of physical activity behaviour in terms of frequency, duration and intensity (Sirard & Pate, 2001). This may be more or less important depending on the aim of the research. For example, in obesity studies, the relevant measure of physical activity is energy expenditure whereas in intervention studies it may be more important to know what type of physical activity has changed.

3.3.2.2. Indirect calorimeter

Indirect calorimetry measures energy expenditure by assessing oxygen consumption and carbon dioxide production (Simonson & DeFronzo, 1990). This can be undertaken through the use of Douglas bags or online gas analysis systems. The Douglas bag method (Douglas, 1911), involves collecting expired air and analysing the contents for O₂ and CO₂ levels (Simonson & DeFronzo, 1990). This is considered the gold standard of gas analysis and is often used as a criterion measure for other gas analysis systems (Atkinson, Davidson & Nevill, 2005). However, the use of Douglas bags is impractical for activities requiring freedom of movement and can only collect gas for a limited time period (Rosdahl, Gullstrand, Salier – Erkison, Johansson & Schantz, 2010). The use of online gas analysis systems (e.g. Cortex Metalyzer; Biophysik, Leipzig, Germany) allows breath-by-breath analysis of O₂ and CO₂ concentrations (Carter & Jeukendrup, 2002) over prolonged periods of time, allowing for an accurate estimate of energy expenditure during activities. Additionally, portable gas analysers such as the Cosmed K4B2 (Cosmed, Rome, Italy) and the Oxycon Mobile (Carefusion, Germany 234 GmbH, Hoechberg, Germany) permit researchers greater freedom in the type of activities that can be assessed. Indirect calorimetry can be used to estimate the energy cost of specific physical activities and as mentioned above can be used in conjunction with DLW to provide estimates of physical activity energy expenditure separately from TEE, by obtaining levels of resting energy expenditure.

Despite the obvious benefits of indirect calorimetry, the ability to measure physical activity in habitual and free-living settings is limited (Dale et al., 2002). Specifically, the portable gas analysers requires a participant to wear a chest harness, housing the analysis equipment on the front and a battery pack on the

back, along with a head harness to secure a face mask in position; the amount and size of this equipment would intrude on everyday activities. Additionally, battery packs have to be frequently exchanged in order to prevent data loss. Both of these issues make indirect calorimetry unsuitable for extended periods of measurement outside of the laboratory and therefore impractical for capturing habitual activity in everyday living environments.

Due to the accuracy afforded by this method, indirect calorimetry is often used as the criterion standard when validating the energy cost of physical activity derived from other objective physical activity measures such as heart rate monitors (e.g. Beghin et al., 2000), pedometers (e.g. Eston, Rowlands & Ingledeu, 1998) and accelerometers (e.g. Treuth et al., 2004; Evenson et al., 2008; Pulsford et al., 2011; Phillips, Parfitt & Rowlands, 2013), which will be discussed later.

3.3.2.3. Heart rate monitors

The use of heart rate monitors to measure energy expenditure is based on the early work by Berggren & Christensen (1950) who noted a direct linear relationship between heart rate (HR) and O_2 . Modern heart rate monitors consist of a lightweight transmitter attached to the chest using electrodes. A computerised telemetry system then transmits minute to minute changes in heart rate to a receiver worn on the wrist (Armstrong & Welsman, 1997). The data can subsequently be converted into estimates of energy expenditure through the aforementioned linear relationship (Livingstone, 1997). Acquired HR data can also be used to estimate the pattern of physical activity in terms of frequency, intensity and duration (Janz, 2002). Studies have shown HR measures to be in excellent agreement with ECG recording (Chen, Janz, Zhu, Brychta, 2012) and to be a valid method for assessing energy expenditure and physical activity patterns in children (Sirard & Pate, 2001). However, at lower intensity activities, HR can be confounded by factors such as emotional state, current climate and caffeine intake which may result in a non-linear relationship (Vanhees et al., 2005), therefore making the prediction of EE at these low intensities difficult (Armstrong, 1998).

It should be noted that although patterns of PA can be estimated, HR only gives an indication of the stress placed on the cardio-respiratory system as a result of activities, and does not measure the movement itself (Vanhees et al., 2005). What is more, a lag may be apparent between onset of activity and the heart rate response (Rowlands & Eston, 2007), causing the assessment of activity patterning to be problematic (Corder, Ekelund, Steele, Wareham, Brage, 2008). Further disadvantages of using HR measurement with children include more practical problems such as the potential loss of data caused by disruption to the transmitters through tampering (Livingstone, 1997).

3.3.3. Measures of Behaviour

3.3.3.1. Direct observation

One method of examining children's physical activity is by directly observing their behaviour. This technique, involves trained researchers observing individual children (in real time or via video tape) for a specific period of time and recording their activity level based on pre-determined classification categories (Kohl et al., 2000; Sirard & Pate, 2001, Dollman et al., 2009; Loprinzi & Cardinal, 2011).

Several observation systems have been developed for this type of measurement, including the children's activity ratings scale (CARS; Puhl, Greaves, Hoyt & Baranowski, 1990). Using this system, researchers record participants' activity at the start of each minute and record any subsequent change in activity level during that minute, though, an activity level can only be recorded once per minute. Activities are classified between levels 1 and 5 as follows: level 1 – stationary or motionless; level 2 – Stationary with limb or trunk movements; level 3 – translocation at a slow / easy pace; level 4 – translocation at a moderate pace; level 5 – translocation at a fast pace (Puhl et al., 1990). Level 1 and 2 represent 'sedentary behaviours', level 3, light activity, level 4, moderate and level 5, vigorous activity.

Direct observations methods have also attempted to contextualise physical activity. Early systems such as the Behaviours of Eating and Activity for Child

Health Evaluation System (BEACHES; McKenzie et al., 1991) and the System for Observing Play and Leisure Activity in Youth (SOPLAY; McKenzie, Marshall, Sallis & Conway, 2000), attempted this by including information about the environment in which the activity was taking place, e.g. Indoors / outdoors (McIver, Brown, Pfeiffer, Dowda & Pate, 2009). However, more recent systems (Observational system for recording physical activity in children – preschool and home versions; OSRAC-P; Brown et al., 2006 and OSRAC-H; McIver et al., 2009) also include information about the type of activity, location, activity initiators, group compositions and indoor / outdoor context (McIver et al., 2009). Although these new systems allow highly specific and detailed information to be obtained about children's physical activity, the amount of information gathered is reduced due to the sampling frequency. The researcher is required to observe a child for 5 seconds, and record the highest level activity undertaken in those 5 seconds (using a similar method to that described above in the CARS system). The observer then has 25 seconds to record the relevant information about this activity, before undertaking another 5 second observation, resulting in 60 observations every 30 minutes. Considering the well documented spontaneous and transitory nature of children's physical activity (e.g. Bailey et al., 1995; Eston et al., 1998; Hand, Parker & Larkin, 2006; Baquet, Stratton, Van Praagh, Berthoin, 2007) the use of these methods may result in researchers missing changes in activities during the recording period which may lead to misclassification.

The use of direct observation systems has been found to be both a valid and reliable technique for measuring children's PA behaviour (McKenzie, 2002). The method can be highly beneficial for acquiring very specific physical activity data on small groups of children in specific environments (Dollman et al., 2009) especially for studies that aim to assess more contextual data (Bailey et al., 1995). However, the drawbacks of employing this method are numerous. Primarily, the monetary and time costs of employing a measurement method such as this means that it is not feasible for use with large sample sizes (Dale et al., 2002). Furthermore, reactivity to researcher's presence may occur, causing children to deviate from their usual activity (Kohl et al., 2000). Finally, the ability to obtain information about habitual physical activity over a number of days and places is unrealistic due to the time constraints and intrusion into children's

lives. Though the development of a wearable digital camera such as the SenseCam (Microsoft Research, Microsoft, Cambridge, UK) may overcome some issues of measuring habitual PA through direct observation. The SenseCam automatically captures and stores photographic images upon the detection of a change in the environment (Chapman, Love, Burgess & Lahav, 2010) and has been shown to be feasible for multiple days of data collection in an adolescent sample (Kelly et al., 2012). In addition, the SenseCam can also be used for verifying self-reported PA (Kelly et al., 2011; 2012).

3.3.3.2. Self-reported behaviour

Self-report methods of physical activity measurement include questionnaires, activity logs and diaries, interviews and proxy reports (Sallis & Saelens, 2000). A recent review by Biddle, Gorely, Pearson and Bull (2011) of the available self-report measures for use with children noted that there are large numbers of instruments available for use with young people that possess varying degrees of formatting. For example, children may be asked to complete recall questionnaires relating to the frequency that they performed certain activities during the previous week, from a checklist of activities (E.g. Physical Activity Questionnaire for Children [PAQ-C; Crocker, Bailey, Faulkner, Kowalski, & McGrath, 1997; Kowalski, Crocker, & Faulkner, 1997). The PAQ-C was developed to assess general activity levels, with frequency given on a 1 to 5 Likert scale that is averaged to form a composite score of physical activity level (1 being *low physical activity* and 5 being *high physical activity*; Kowalski, Crocker & Donen, 2004). Alternatively, children may be asked to complete physical activity diaries or logs in real time, whereby activities undertaken are recorded throughout the week. For example, participants may be asked to record their dominant physical activity every 15 minutes for 7 days (Bouchard et al., 1983; Henry, Lightowler & Al-Hourani, 2004). Diaries can also be completed retrospectively, with for example the 3-day physical activity recall instrument (Pate et al., 2003). Here, participants are asked to recall activities undertaken in the last 3 days in 30-minute time blocks and rate the relative intensities of each activity (Pate et al., 2003). Due to the variety of self-report methods, the resultant outcomes differ widely, ranging from estimates of total energy expenditure (e.g. Computerised activity recall; McMurray, Harrell,

Bradley, Webb & Goodman, 1998), time spent in various MET levels of activities (e.g. 3-day PAR; Pate et al., 2003), to crude categorisation of low and high physical activity (e.g. PAQ-C; Crocker et al., 1997).

3.3.3.2.1. Advantages and disadvantages of self-report

One of the main benefits of using self-report measures is the low cost, which makes the measures ideal for population surveillance (Nusser et al., 2012). Furthermore, information relating to type and context of physical activity can be obtained (Biddle et al., 2011). Nevertheless, there are a number of disadvantages in using self-reported physical activity in children. Primarily, children's cognitive abilities may hinder their recall; they are more likely to recall more 'retrievable' data concerning large bouts of physical activity (Biddle et al., 2011), rather than the short sporadic bouts that occur naturally as part of their daily routines (Loprinzi & Cardinal, 2011). Additionally, social desirability may occur, causing children to exaggerate their activity levels, resulting in an overestimation of activity (Dollman et al., 2009). Errors in recall and the effects of social desirability may lead to considerable misclassification of physical activity in children (discussed further below). Assuming the misclassification is random across population sub groups and exposures of interest, and ranking children according to physical activity levels is required rather than an absolute estimate of physical activity, then the misclassification may be acceptable when weighed against the advantages. As a result of these well documented disadvantages, many researchers have suggested that self-report measures should not be employed in children younger than 10-years of age (Montoye, Kemper, Saris & Washburn, 1996; Dollman et al., 2009; Loprinzi & Cardinal, 2011), as limited cognitive ability and recall problems may be particularly pronounced with this age group (Telford, Salmon, Jolley, Crawford, 2004).

3.3.3.2.2. Proxy-report

A method often used to overcome some of the issues noted above, is the use of proxy measures/reports (Waters, Stewart-brown, Fitzpatrick, 2003). These measures are conducted in much the same manner as the self-report questionnaires, except those in close proximity to the participant (i.e. teachers, parents, peers) report their perceptions of the child's PA levels through

retrospective questionnaires and checklists (Burdette, Whitaker & Daniels, 2004). For example, Chaumeton, Duncan, Duncan & Strycker (2011), asked parents to answer questions such as “*On how many of the past 7 days did your child exercise or take part in hard physical activities that made them sweat and breath hard for at least 20 minutes without stopping*” (p.211.) This particular question was designed to measure vigorous intensity PA behaviours. Alternatively, some proxy questionnaires attempt to quantify correlates of PA rather than the behaviour itself (e.g. time spent outdoors, Burdette et al., 2004) which, in turn also act as a proxy measure of PA. For example, Tremblay and Williams (2003) attempted to examine the relationship between Canadian children’s PA, sedentary behaviours and BMI, by asking parents questions such as “*In the last 12 months, outside of school hours, how often has [Name of child] taken part in any sports that involved coaching and instruction*” (p.1101). Responses were given on a 5 point scale relating to frequency (e.g. almost never, once a week etc.). The authors did not attempt to determine time in PA levels through this measure; rather, they used the reported frequency to sport club attendance in the subsequent analysis, similarly to Burdette et al. (2004).

Although the use of proxy measures overcomes the potential disadvantages of self-report with younger populations (Corder et al., 2008), the method is limited by the extent to which respondents observe PA behaviours (Chaumeton et al., 2011). Additionally, parents/carers may themselves be biased by the social desirability of describing their children in the best possible light. Telford et al. (2004) found low correlations between self-report and parental proxy reports of children when assessing the frequency and duration of children’s moderate and vigorous physical activity, with activities such as physical education classes (frequency $r = .02$; duration $r = .05$), household chores (frequency $r = .16$; duration $r = .3$, $p < .01$) and walking for exercise (frequency $r = .16$; duration $r = .23$, $p < .01$) sometimes showing non-significant relationships. Higher correlations were apparent for organised sports ($r = .38$ to $.84$, p ’s < 0.05). These results further demonstrate the potential problems of using proxy-measures to estimate children’s physical activity, though it is not apparent which measure provided the closest estimate of PA.

Both self-report and proxy-report of physical activity can also be administered through interview techniques that take place either face-to face or via the telephone (Warren et al., 2010). Generally, interviews follow the same format as the reported scales but can be explained in greater detail by researchers, with the potential to obtain large amounts of information. For example, the 7 day physical activity recall (7 day PAR; Sallis et al., 1985) allows the interviewer to guide the participant through the previous weeks' activities, prompting the subject when recall is difficult. Information can be gained on the frequency, duration and intensity of sleep and time in MVPA, and estimates of energy expenditure can be obtained. These types of interviews can be employed with children and adults, following a standard partially scripted interview technique, in an attempt to reduce bias (Sallis et al., 1985). However, the potential for bias still exists in terms of social desirability and recall difficulties with sporadic activities.

3.3.4 Measures of physical activity via movement sensors

The challenges of accurately measuring children's habitual physical activity with the methods discussed above and the development of new technologies has led to the increased popularity of movement sensors such as pedometers and accelerometers (Eston et al., 1998). As physical activity is defined as movement of the body resulting in energy expenditure (Caspersen et al., 1985), measuring bodily movement directly, with motion sensors, should lead to more precise estimates of physical activity (Sirard & Pate, 2001).

3.3.4.1. Pedometers

One method of measuring body movement is with small mechanical and electronic devices called pedometers; these sensors are usually worn at the hip, attached to a waistband or belt (Morgan, Pangrazi & Beighle, 2003). Pedometers contain spring-suspended mechanical lever arms positioned horizontally, that move up and down as each step is taken which acts to open and close an electrical circuit (Basset & Strath, 2002; McNamara, Hudson & Taylor, 2010). Essentially, pedometers measure vertical accelerations of the body, produced when undertaking activity such as walking (Sequeira,

Rickenbach, Wietlisbach, Tullen & Schutz, 1995). The circuit that is created when the mechanical arm moves results in steps being recorded and displayed on the pedometer display (Basset & Strath, 2002). These step counts can be recorded for a number of days, dependent upon the memory capacity of the monitor used, or can be noted down by the participant / parent each day and reset for the following day. Tentative guidelines now exist for interpreting the output in terms of meeting the 60-minute per day MVPA guidelines for children (Department of health, 2011); A range from 10,000 to 15,000 steps-day⁻¹ has been proposed to approximate the guidelines dependent upon age and sex (Tudor-Locke et al., 2011). There are multiple commercially available pedometers with varying degrees of published validity and reliability (Tudor-Locke, McClain, Hart & Sisson, 2009).

Benefits for employing pedometers to measure's children physical activity include their reasonably low cost meaning that it is feasible to use them in large scale studies (McNamara et al., 2010). They have been shown to be valid and reliable for use with children (e.g. Jago et al., 2006), they possess a standard output unit (steps-day⁻¹) (Beets, Bornstein, Beighle, Cardinal & Morgan, 2010) and recent advances in technology and research means the output can be tentatively interpreted in terms of meeting PA guidelines for adults (Marshall et al., 2009); if the monitor is equipped with a time sampling mechanism accumulating 3000 steps or more in 30 minutes equates to the MVPA guidelines for adults. Despite these benefits, the lack of detailed information concerning intensity and timings, available with other physical activity measures, means that further interpretation of data is limited. Additionally, there is confusion about the length of monitoring time needed to obtain accurate estimates of physical activity habits (Tudor-Locke et al., 2009). Studies have also reported reactivity to wearing the device (Tudor-Locke et al., 2009), as participants can monitor their own physical activity levels (Lubans, Morgan, Tudor-Locke, 2009), however, this issue is contested elsewhere (Ozdoba, Corbin & Masurier, 2004). Finally, the impact of children's stride length, as a function of stature (Jago et al., 2006), and gait upon step counts recorded may mean that monitors need to be individually calibrated (Dollman et al., 2009). A number of practical issues also exist for the use of pedometers; removal is necessary for some activities including water based activities due to lack of

waterproofing and contact sports due to the potential damage to the monitor and risk of injury to participants. Furthermore, pedometers are unable to capture certain activities such as upper body movement and cycling due to the wear position.

3.3.4.2. Accelerometers

Like pedometers, accelerometers are movement sensors that measure accelerations of the body. Traditionally, these devices contain a piezoelectric acceleration sensor, that deforms under acceleration (Chen & Bassett, 2005), producing a voltage charge proportional to the extent of the deformation, that in turn is recorded as acceleration (Mijer, Westerterp, Verhoeven, Koper & Hoor, 1991). The direction in which the piezoelectric sensor is most sensitive (in relation to its mounted position) is the vertical plain, best suited to recoding ambulatory movements when worn on the hip (Chen & Bassett, 2005). Newer devices have employed piezoelectric and piezoresistive integrated chip sensors that change resistance once mechanical deformation has occurred. These devices are sensitive to both gravitational and motion accelerations (Chen et al., 2012), which enables researchers to obtain information about changes in posture and orientation of the body (Yang & Hsu, 2010), with respect to the position of the sensors (Knight et al., 2007). A device measuring accelerations in one plane is referred to as uniaxial accelerometer; whereas tri-axial devices measure accelerations in three planes; vertical & antero-posterior and medio-lateral (VanHelst et al., 2012).

Acceleration is the change in speed in relation to time, which is measured by meter per second squared (m/s^2 or $m \cdot s^{-2}$) and conventionally expressed as a gravitational unit of acceleration (g), with 1 g being equal to $9.81 m \cdot s^{-2}$ (Chen & Bassett, 2005). In commonly used accelerometers (e.g. ActiGraph GT1M, GT3X [ActiGraph, Pensacola, Florida, USA], the RT3 [stayhealthy.com, Monrovia, California, USA] and the Actical [Mini Mitter Co., Inc., Bend, OR]), acceleration output from the sensors is passed through a high and low band filter to eliminate any recorded motion outside the range of human movement (Welk, 2002). Subsequently, raw acceleration data is converted from an analogue to a digital signal (Chen & Bassett, 2005) with the resultant output

converted from an SI unit to 'counts', a somewhat arbitrary unit (Welk, McClain & Ainsworth, 2012). Once this conversion has taken place, the 'count', data is summed over a pre-determined time period (epoch) and stored for analysis (McClain, Abraham, Brusseau & Tudor-locke 2008). The stored data can be used to facilitate the application of some biological, physiological or behavioural meaning to the movement. Although both uniaxial and triaxial monitors are available, there is uncertainty as to whether one is superior to the other in terms of estimating physical activity level (Welk & Corbin, 1995; Hendleman, Miller, Debold & Freedson, 2000; Ott, Pate, Trost, Ward & Saunders, 2000; Butte et al., 2012), additionally, the output from triaxial monitors is often collated into a single vector magnitude and analysed in the same manner as uniaxial monitors.

3.3.4.3 Data reduction with accelerometry

Most accelerometers can record data in epochs between one and 60 seconds, however the period for which these epochs can be recorded depends upon the memory of the device itself. Previously, studies have employed 60 second epochs (Rowlands, Powell, Humphries & Eston, 2006), a sampling period required in order to capture multiple days of data, yet short epochs are recommended for research with children (Ward, Evenson, Vaughn, Rodgers & Troiano, 2005) based on the highly transitory nature of children's activity (Bailey et al., 1995). It is important to note that, the memory limitations that led to the use of large epochs have largely been solved; despite this, studies still employ larger epochs than necessary (e.g. Huberty et al., 2011; Van Cauwenberge et al., 2011). The resultant counts per epoch, gained from accelerometry, can be assigned some physiological or behavioural meaning (Freedson et al., 2005), following a series of data reduction processes.

Firstly, a quality control check should be undertaken to assess whether the monitors have been worn for an appropriate amount of time. The process involves determining when accelerometers are being worn (wear time) and when they may have been removed (non-wear time). In some instances removal may occur due to water based activities, contact sports or children's preference (Crocker, Holowachuk & Kowalski, 2001). Non-wear time is usually classified as periods of time with continuous '0' counts, sometimes allowing for

interruptions of one to two minutes of counts up to 100, (Colley, Gorber & Tremblay, 2010) to account for actions such as the moving of the monitor. Previous studies have defined non-wear time as being between 10 minutes (Riddoch et al., 2004) and 180 minutes (Van Coevering et al., 2005) of consecutive '0's, though recent studies have recommended either 20 (Esliger et al., 2005) or 30 (Rowlands et al., 2010) minutes when collecting data with children. Children must achieve a minimum amount of wear time in order for a measured day to be classed as 'valid'. The purpose of this is to try to ensure that all activities undertaken are captured. If wear time is low it is possible that some activity will not be recorded leading to misclassification. Some studies have set their criteria at as little as 3 hours a day (McLure, Summerbell & Reilly, 2009) whereas others have deemed eight (Janz et al., 2009) or 10 hours of monitoring per day is necessary (Anderson et al., 2006; Riddoch et al., 2007; Troiano et al., 2008). Colley et al. (2010) noted that valid day wear time must be high enough to ensure that days of non-wear are eliminated, but low enough to prevent large numbers of days from being eliminated, which may create sample bias. However, an additional consideration must be made about the section of the day recorded. Recording time must be adequate enough to ensure that misclassification does not occur as a consequence of only capturing certain sections of the day (e.g. school time / after school hours), as physical activity during school hours may be restricted (Dale, Corbin & Dale, 2000).

It is recommended that seven valid days of monitoring is achieved, however, a minimum of four days can be used in analysis to gain reliable estimates of weekly habitual activity (Troost et al., 2000). Furthermore, studies have emphasised the importance of including weekend days within the data collection due to activity differences being apparent between weekday and weekend activities (Troost et al., 2000; Trost, McIver & Pate, 2005; Rowlands 2007; Yildirim et al., 2011). Wear time requirements are generally accepted as standard within the children's physical activity literature, with the majority of studies measuring activity for seven days, and including in analysis participants that have at least three week and one weekend valid days of data with 10 hours per day of wear time. Some studies have used fewer days (e.g. McLure et al., 2009; Page et al., 2010) in an attempt to maximise participant numbers, however in doing so, they risk misclassification for the reasons highlighted

above. Recently, Heil et al., (2012) have proposed a seven step algorithm to act as a guide for researchers when attempting to undertake the data reduction process, including the consideration of epochs, wear time, data transformation and interpretation of data within the relevant steps.

Once the data reduction process has occurred, estimates of intensity, frequency and duration of physical activity can take place, by assigning each epoch of counts an intensity level. This is undertaken using pre-determined 'thresholds' or 'cut-points' that detail high and low boundaries of accelerometer counts, between which activity intensities are estimated to lie (Freedson et al., 2005). Time accumulated between thresholds of accelerometer counts can be determined and reported as time spent in sedentary, light, moderate, vigorous and MVPA intensities.

3.3.4.4. Interpreting accelerometer data

Classification of physical activity intensity cut-points comes from calibrating the output with physiological or biological variables. Traditionally, researchers ask participants to undertake a laboratory based protocol whilst accelerometry and indirect calorimetry measures are taken. These protocols have ranged from treadmill – based (e.g. Trost et al., 1998; Freedson et al., 2005) to free living activities (e.g. Mattocks et al., 2007) and some have included both treadmill and daily living activities (e.g. Rowlands, Thomas, Eston, Topping, 2003; Treuth et al. 2004; Phillips et al., 2013). Accelerometer output can be calibrated to assess for energy expenditure (Kcal/min) by the use of regression equations, whereby the relationship between activity counts and intensities of activities are established (Welk, 2002) and can subsequently be used to determine the boundaries of accelerometer counts between which activities intensities are occurring. For example, Mattocks et al. (2007) created a prediction equation for energy expenditure from counts per minute, age and sex by using a random intercept model.

Equation 3.1. Mattocks et al. (2007) energy expenditure regression equation.

Energy expenditure ($\text{kJ}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$) =

$$-0.933 + 0.000098 \text{ counts}\cdot\text{min}^{-1} + 0.091 \text{ age}(\text{yrs}) - 0.0422 \text{ sex}(\text{male}0; \text{female}1)$$

In order to develop intensity cut-points from this model, $\dot{V}O_2$ replaced EE as the outcome variable in the equation. $\dot{V}O_2$ ($\text{ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$) values equating to MET values of 3 (moderate) and 6 (vigorous), based on group mean resting values, were established and the corresponding counts per minute determined. For example, the resultant cut-points for moderate intensity activity (3 METS; $17.1 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$) was 2306 counts $\cdot\text{min}^{-1}$, with 6130 counts $\cdot\text{min}^{-1}$ as the lower bound for vigorous (6 METS; $34.1 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$).

More recently, activity thresholds for classifying activity intensity into binary variables, e.g. MVPA or not MVPA have been derived through receiver operator characteristic (ROC) curve analysis, a method conventionally used in signal detection and medical trials for determining the sensitivity and specificity of medical tests (Greiner, Pfeiffer & Smith, 2000). The ROC can determine the accuracy of tests in discriminating between the occurrence and non-occurrence of a binary variable. In the case of accelerometer cut-points, the ROC curve can assess all possible count thresholds between two activity intensities (Jago, Zakeri, Baranowski & Watson, 2007) so as to determine the optimal count threshold for discrimination between intensities (e.g. between moderate and vigorous activities). Using this method, activities undertaken during the calibration protocol are assigned an intensity (e.g. sedentary, light, moderate or vigorous) based on a criterion measure, for example METs. Subsequently, the corresponding accelerometer counts for each activity can be coded into binary indicator variables (0 or 1) based on intensity (sedentary versus > sedentary, less than moderate activities versus moderate to vigorous activity, and vigorous activities versus < vigorous activities).

The ROC curve provides an Area Under the Curve (AUC) value; this represents the discrimination accuracy of the accelerometer counts to distinguish between the occurrence and non-occurrence of intensity. An AUC of <0.70 demonstrates poor accuracy (Mckintosh, Fairclough, Stratton & Ridgers, 2012) with an AUC of 1 indicating perfect classification accuracy (Pulsford et al., 2011). Values of sensitivity (correctly identifying at or above the intensity threshold, i.e. the true positive rate) and specificity (correctly excluding activities below the threshold for intensity, i.e. the true negative rate) (Jago et al., 2007)

of different cut-points along the ROC curve are obtained. The cut-point that maximises both sensitivity and specificity is usually selected as the best threshold to use in determining the intensity of an activity.

This type of analysis for calibration studies is now being employed more frequently and offers an alternative to the traditional method of linear regression. The use of the ROC curve allows for a selection of cut-points that minimise the potential for error by maximising the true positive and true negative values, whilst limiting the false positive and false negative occurrence. Additionally, the ROC avoids the problem of non-linear relationships between intensities and counts at higher intensity activities (Jago et al., 2007), which may impact upon the traditional method of cut-point determination. However the ROC cannot be employed to determine energy expenditure at different counts, but purely offers the discrimination between intensities.

Along with accelerometer specific cut-points, the establishment of population specific cut-points has been encouraged (Romanzini, Petroski & Reichert, 2011), therefore multiple calibration studies have been undertaken with different devices in both adult and child populations. In an attempt to enable comparison of studies using different accelerometers, researchers (Straker & Campbell, 2012; Paul, Kramer, MoshFegh, Baer & Rumpler, 2007) have developed regression equations to translate total activity and minute by minute counts from the Actical and ActiGraph (vertical axis & vector magnitude) for comparison. These studies result in thresholds being available which equate Actical to ActiGraph counts and vice versa. E.g. Straker and Campbell reported that 100 counts·min⁻¹ on the ActiGraph was equal to 91 Actical counts·min⁻¹, which both represent the same sedentary threshold. Similarly, 100 Actical counts·min⁻¹ equalled 110 counts·min⁻¹ on the ActiGraph, again representing a sedentary threshold.

Although the use of accelerometer thresholds allows for estimation of absolute intensity, there may be substantial error associated with this method; through the use of cut-points, all activities are categories within MET boundaries (e.g. moderate; 3 – 6 METS), however, as noted previously in section 3.3 of this chapter, activities with absolute MET values may actually require greater

physical effort when performed by people within the same population, depending upon their fitness level.

If we use an example provided in table 3.1., shovelling has an absolute MET value of 5.0 for adults and is classed as moderate intensity, regardless of whom it is performed by, however this may require a much greater physical effort from an elderly woman, than it would for a young man (Lee et al., 2003), resulting in different relative EE but defined as the same absolute EE. Therefore it is possible for a wide range of relative intensities to be classified as the same absolute value within a population sample.

Attempts have also been made to determine the actual activity being undertaken through pattern recognition technology applied to accelerometer output. One monitor capable of such classifying activities is the Intelligent Device for Estimating Energy Expenditure and Activity (IDEEA; Minisun, Fresno, CA). This device monitors movement through five sensors attached to the chest, thighs and feet of participants (Zhang, Werner, Sun, Pi-Sunyer, & Boozer, 2003). It is possible to successfully discriminate 32 activities, including sitting, standing, lying, walking and running at different speeds and climbing stairs. Despite providing accurate assessment of physical activity, the sensors need to be attached to the main body of the monitor via wires, which may prove too intrusive for long term activity assessment. The generation of new unobtrusive high frequency sampling accelerometers has the potential to facilitate the use of pattern recognition in habitual physical activity assessment and increase the accuracy of estimates of time spent in different physical activity behaviours such as walking, running and household chores. (Bonomi, Goris, Yin & Westerterp, 2009; Zhang, Rowlands, Murray & Hurst, 2012).

3.3.4.5. Advantages and disadvantages of accelerometry

Accelerometers allow for the direct assessment of movement with minimal burden to the user (Rowlands et al., 2004), while their small size also makes them practical for use with children (Freedson et al., 2005). The rise in popularity of accelerometers has been noted by Rowlands (2007), who reported that the number of publications relating to physical activity and accelerometers

use had risen dramatically between 2001 and 2005. A recent PubMed search for 'accelerometers' revealed 545 studies were published in 2011, 594 studies in 2012, and by 30th April, 130 studies had been published relating to accelerometer use in 2013. The inference of this seems to be that numbers will continue to increase as the benefits of the discrete devices overcome the disadvantages apparent with other assessment methods, and technology continues to advance. However, the use of accelerometers is not entirely free from drawbacks; for example, each commercially available accelerometer will have its own A/D converter, with differences in sampling frequencies and signal filters, meaning that the output from one accelerometer is not comparable to those from a different device; with the exceptions of the ActiGraph GT3X+ and the GENEActiv accelerometers, where output is provided in the raw format pre filtering. Furthermore, plateau effects have been observed in ActiGraph 'count' output over certain movement speeds (Brage, Wedderkopp, Franks, Andersen & Froberg, 2003; Rowlands et al., 2007). Irrespective of the type of accelerometer and the output obtained, maximising participant compliance and differentiating between wear and non-wear is a priority.

3.3.4.6. Wear time compliance

Although wear time recommendations exist within the literature, a major problem faced by researchers is children's lack of compliance with these wear-time requirements. Crocker et al. (2001) examined the feasibility of wearing the hip mounted Tritrac (Professional products, a division of Reining International, Madison, WI) for a period of 7 days with 79 children (mean age = 11 years \pm 1.4). Results showed that children wore the monitor (established by a minimum vector magnitude count of 1 per 15 minutes) for a mean of 76 hours \pm 19.75, assuming 8 hours of sleep, maximum wear would be 112 minutes, yet the actual wear ranged from 33 to 112.8 hours. Reasons provided for monitor removal included forgetfulness, discomfort, water based activities, embarrassment, travel and removal requested by external authorities, which appeared to be the most prevalent reason for removal in the Crocker et al. (2001) study.

Compliance was also examined by Van Coevering and colleagues (2005) with regard to the hip mounted ActiGraph. From a sample of 255 children, only

50% complied with the wear time criteria of wearing between 8am to 9pm weekday and midday to 9pm on weekends (non-wear = 180 minutes consecutive '0's) for seven days. 86% of participants had 4 valid days; whereas 92% had three valid days. More recently, Colley et al. (2010) demonstrated that for children aged 6 to 11 wearing an ActiGraph at the right hip, 86 % of males and 88% of females had 4 or more valid days, when wear-time was classed as >10 hours and non-wear as 60 minutes of '0' counts; whereas only 35% males and 42% females recorded seven valid days of data. Similar results were also demonstrated by Audrey, Bell, Hughes and Campbell (2012), with only 75% of 822 participants meeting their inclusion criteria (≥ 600 minutes for ≥ 3 days) when wearing an ActiGraph mounted at the hip.

Lack of compliance with accelerometer wear is a major problem facing researchers as it results in missing data that may lead to bias and misclassification. Two strategies have been adopted to deal with this. The first has focused on methods to improve compliance and the second is based on statistical modelling. Sirard and Slater (2009) assessed the impact of methods to increase compliance of 87 students aged 14 – 18 years. Pupils were assigned to either a compliance strategy group, including the use of phone reminders, journal logs of wear and non - wear times, a monetary reward, or a control group. Results showed that the monetary incentive (contingent on the number of day with >10 hours wear) resulted in the highest percentage of participants completing 7 valid days (53.9%). This was also apparent for participants accumulating >4 valid days, with 96.2% in the monetary group, 85% in the journal group, 71.5% in the phone reminder group and 70% in the control group. Providing participants with a financial incentive to complete sufficient days of accelerometer wear may prove unrealistic for many researchers and therefore attempting to employ less costly methods may be more appealing and may also increase compliance rates (Matthews, Hagströmer, Pober & Bowles, 2012) beyond the compliance of participants receiving no incentives.

The placement of monitors and the emergence of new technologies may also impact on compliance and monitor wear time advice. The traditional placement of monitors at the hip was due to movement being captured near the subject's

centre of gravity (Heil et al. 2012). However placement on the wrist may increase compliance to the wear time protocol (Rowlands & Stiles, 2012). Furthermore, the development of waterproof monitors such as the GENEActiv (ActivInsights Ltd, Kimbolton, Cambridge, UK) and the ActiGraph GT3X+ mean that monitors can be worn for 24 hours a day. It is hypothesised that these qualities largely negate the need to remove the monitor and thus may lead to greater compliance during assessment of habitual activity in both children and adults. In essence 24-hour wear removes the amount of non-wear due to forgetting to put the monitor back on following removal. Wear-time compliance will be examined further in chapter eight of this thesis.

In a different approach, some studies have attempted to reduce missing data caused by non-compliance through statistical methods of imputation. Imputation utilises the observed activity values to predict activity when missing values occur (Catellier et al., 2005). Multiple methods of imputation have been examined, with the simplest involving using the mean counts from either the existing group or remaining individual data of the same day (e.g. Monday) or the same type of day (e.g. weekend) to fill in the missing data points (Kang, Rowe, Barreira, Robinson & Mahar, 2009). Kang et al. (2009) reported a decrease in variance when using the group information approach, but found that using the individual approach negated the drop in variance somewhat. Though a variety of imputation methods exist, very few studies have employed them and therefore their value remains relatively uncertain.

3.4. Sources of misclassification in the association between physical activity and psychological well-being

It is possible when measuring either physical activity or psychological well-being for misclassification to occur. In terms of physical activity, misclassification occurs when time spent in a given activity intensity is either over- or under- estimated; whilst with the measurement of psychological wellbeing, misclassification may occur, again either under- or over- reporting in response to questionnaires. The following section explores the potential sources of misclassification through the use of inappropriate exposure measures, participant response and decision processing prior to analysis.

3.4.1. Misclassification through random and systematic error

The employment of an inappropriate or imprecise exposure measure for the objectives of the research may result in non-differential or random error (Fosgate, 2006) and in turn cause misclassification to occur (Blair, Stuart, Lubin & Forastiere, 2007). In other words, the measures chosen must reflect the aim of the study in order to avoid potential error; if the aim of a study was to gain an indication of pattern of activity over a week, measurement through self – report methods would provide imprecise estimation due to a lack of sufficient information required to accurately assess the desired outcome. Estimates of time spent in different intensities over varying periods of the day from measures such as this would lack adequate validity for conclusions to be drawn, and misclassification of the outcome variable would occur. If this type of error occurs at random it can alter associations between physical activity and various health outcomes. For example, Brown, Kreiger, Darlington and Sloan (2001) examined the concept of misclassification with regard to exposures of total caffeine intake or coffee consumption as a surrogate exposure measure upon odds ratio estimations for various diseases at multiple levels of exposure. The results showed that using a surrogate (coffee consumption) resulted in high misclassification rates and underestimation of caffeine intake, which was noted to be a contributing factor to the lack of associations between caffeine intake and various diseases; the non-differential error masked the true associations. Additionally, employing an inappropriate exposure measure may also lead to misclassification; for example, self-reported physical activity is not recommended for young children (Loprinzi & Cardinall, 2011) for the reasons discussed earlier in section 3.3.3.2, therefore children's physical activity assessed via self-report methods would lead to a misclassification of activity levels. Misclassification due to the employment of inappropriate and imprecise exposure measures may be occurring within the PA literature, as opposing trends between PA and markers of health have recently been noted (Katzmaryk & Tremblay, 2007; Basterfield et al., 2008). Specifically, Basterfield and Colleagues note that self-reported MVPA has shown a steady increase, yet increases in obesity (Stamatakis, Zaninott, Falaschetti, Mindell & Head, 2010; McCarthy, Ellis & Cole, 2003; Reilly, 2006) and proxy measures of inactivity

(e.g. number of TVs and computers) also have increased (Katzmaryk & Tremblay, 2007). If PA is over-reported the association apparent with health variables may be underestimated; a simple example of this is, if a sample of children report high levels of MVPA, but the same children present with high levels of adiposity, it may be determined that PA had a weak relationship with obesity in children. Whereas in actual fact the sample were doing much less activity than they had reported, and the relationship was stronger.

Sometimes the result of misclassification is obvious. If differing cut points are used to determine MVPA then it logically follows that the resulting prevalence estimates will also vary. Sometimes the exact nature of the misclassification is more difficult to detect, especially when two different measures have been used in the same sample. Basterfield and Colleagues (2008) asked parents to report children's physical activity using the Health Survey for England proxy report, whilst children wore an accelerometer for seven days. Results of the questionnaire data led to an estimate that 83% of boys and 56% of girls achieved the recommended 60 minutes MVPA per day, however only 3% of boys and 2% of girls achieved the guideline levels when accelerometers were used as the measure. Sliotmaker, Schuit, Chinapaw, Seidell & Mechelen, (2009) also examined the discrepancy between prevalence estimates from self-report and from accelerometry; they found that that adolescents reported a mean of 596 minutes-week⁻¹ more time in moderate and 178 minutes-week⁻¹ more time in vigorous activity than was measured through accelerometry. In both of these examples it is not possible to say what the 'true' values of physical activity are, just that the two methods differ.

As stated above misclassification of physical activity can lead to errors in estimating physical activity prevalence in population surveillance (Basterfield et al., 2008) but can also alter the association between physical activity and various outcomes (Baranowski, Mâsse, Ragan, Welk, 2008). If physical activity is an exposure measure and a measure of health is the outcome, non-differential (or random) misclassification will attenuate the association between exposure and outcome towards the null (Flegal, Keyl & Nieto, 1991; Brenner & Blettner, 1993; Gustafson & Le Nhu, 2002). In line with this, Straker & Campbell (2012) noted that the likelihood of detecting associations and the strength of

those associations increases with the use of activity monitors because of their additional precision. A recent study conducted by Celis-Morales et al. (2012) noted that adults' self-reported physical activity revealed only a few associations with metabolic and vascular risk factors of disease, whereas accelerometry measured PA in the same population revealed more associations. Martinez-Gomez et al. (2012) also reported that, within the same population of adolescents, only objectively measured vigorous activity was correlated to one inflammatory marker of arterogenesis (the build-up of fatty material in the arterial walls) with no relationships found between self-reported PA and inflammatory markers.

These studies demonstrate how physical activity research outcomes depend upon precise measurement, with lack of precision introducing considerable uncertainty. Non-random systematic misclassification of physical activity may also occur, where sub populations may systematically under or overestimate their physical activity. For example, Watkinson et al. (2010) reported that inactive individuals over estimate the amount of activity they do. Furthermore, the authors noted that individuals with low BMI scores interpret their anthropometric measure as an indicator of adequate activity, with the reverse being true for those with higher BMI score; those with higher BMI may view this as a result of inadequate activity levels. These over- and under- estimations that result in systematic error can be avoided if appropriate exposure measures are employed.

Systematic error can occur with the measurement of physical activity and psychological well-being due to aspects such as social desirability (Crown & Marlowe, 1964), where respondents tend to over-rate traits deemed desirable and in turn underestimate less desirable traits (Klesges et al., 2004). For both questionnaires and interviews measuring PWB, if the respondent is aware of the purpose, they may control their answers to portray an appropriate image of themselves (Phillips, 1973; Niemi, 1993). This may also be true in PA measurement. Adams et al. (2005) explored the concept of social desirability in PA self-report methods compared to the criterion of DLW and accelerometry; the authors noted that social desirability lead participants to overestimate both their physical activity energy expenditure and the duration spent in light and

moderate activity intensity on 7 day physical activity recall questionnaires compared to the criterion measures. In addition, a respondent's affectivity (Watson & Clark, 1984); a participant's natural disposition towards a positive or negative affect, may influence their responses (Podsakoff, Mackenzie, Lee, Podsakoff, 2003). For example if a person's disposition was towards negative affect they may over report their negative feelings and under report their activity resulting in an overestimate of the true relationship between inactivity and negative feelings.

Common method variance (CMV; Podsakoff et al. 2003) may also result in systematic error; CMV refers to the amount of "*spurious covariance shared among variables due to a common method used in collecting data*" (Malhotra, Kim, Patil, 2006. p1865.) In other words, it may be that relationships occur partially due to similar methods being employed to assess the constructs of interest, which in turn can lead to misleading conclusions (Campbell & Fiske, 1959). Specific to the present thesis, researchers may expect measures of PWB to be correlated with measures of PA, however if the measures share a common method (e.g. self-report), systematic error may be introduced, inflating or masking the true nature of relationships (Doty & Glick, 1998; Meade et al. 2007) resulting in a type 1 error if inflation occurs, or a type 2 error if the relationship is masked. An example would be if self-reports of PWB tended to underestimate PWB and self-reports of physical activity tended to overestimate physical activity then the relationship between the two variables would be underestimated.

As discussed at the beginning of the chapter, psychological well-being measurement must take place through self-report methods, therefore in order to avoid the potential confounding effects of CMV, physical activity should be measured through non-subjective methods, such as accelerometry.

3.4.2. Misclassification in accelerometer data.

Although the above examples of misclassification occur due to measurement method choices made prior to data collection, misclassification can also become apparent through choices formulated after data collection, within the

analysis. One of the main issues facing researchers whose chosen PA measurement method is accelerometry, is the interpretation of the output from these monitors and the lack of consensus on an appropriate algorithm to be employed when reducing the data to a suitable format (Masse et al., 2005; Heil et al., 2012).

As previously discussed, decisions about appropriate epochs, non-wear time and valid days must be made prior to any analysis of accelerometer data. Choices within these categories can easily lead to an over- or under- estimation of time spent at different intensity categories. For example, when assigning physical activity intensity levels (sedentary, light, moderate and vigorous) to epochs in order to assess the accumulation of time at different intensities (Trost et al., 2005), use of a 60 second epoch has been shown to under-estimate time in higher intensity activities (Trost et al., 2005; Rowlands, 2007) in children. Nilsson, Ekelund, Yngve & Sjöström (2002), reported that time in higher activity intensities significantly differed when epochs of 5, 10, 20, 30, 40, and 60 seconds were employed. Specifically, participants recorded 37.1 and 11.7 minutes of high and very high intensity activity, respectively for 5 second epochs, but only 9.9 minutes and 1.3 minutes for high and very high intensity activity, respectively, when a 60 second epoch was employed. These results have been echoed more recently by Vale, Santos, Silva, Soares-Mirander and Mota (2009), with differences of 9.33, 7.49 and 16.41 minutes being shown for moderate, vigorous and MVPA respectively when 5 second and 60 second epochs were utilised. However, underestimation of activity may still occur with 5 second epochs, as the finding of Bailey et al., (1995) show that high activity bouts last as little as 3 seconds. With the recording capabilities of new accelerometers, shorter epochs should be employed so as to capture all relevant activity and classify activity accordingly.

The determination of wear- and non-wear time, through set periods of consecutive '0' counts, may also lead to misclassification of sedentary behaviour. These set periods of '0' counts may not only occur during monitor removal but also during static activities such as sitting. It is important therefore to distinguish between prolonged periods of non-movement and monitor removal, to prevent the misclassification of non-wear as activities involving little

or no movement, resulting in an over-estimation of time in sedentary behaviours (Rowlands et al., 2010). Studies examining the most biologically plausible period of consecutive '0' counts, have differed in their conclusions; Esliger, Copeland, Barnes and Tremblay (2005) reported that, mean time for motionless activity in children was 17.5 minutes, concluding that ≥ 20 minutes of '0' counts was adequate in determining non-wear. However, Rowlands et al. (2010) examined the occurrences of supposed monitor removals when 10, 20, 30 and 60 minutes of non-wear time were applied to accelerometer data obtained during prolonged motionless activity; they determined that a criteria of 30 minutes of '0' counts was the least likely to misclassify inactivity as monitor removal.

Further misclassification is introduced as physical activity may be underestimated when shorter wear time periods are used as the valid day criterion; Catellier et al. (2005) reported that average time in MVPA differed depending upon the wear time requirements, showing that for 12 h wear time, average time in MVPA was 159 minutes, whereas for 10 h and 8 h, time in MVPA was 147 and 148 respectively, demonstrating that wear time requirements may result in missed activity and under-estimation of time spent in activity levels throughout the day. More recently, Herrmann, Barreira, Kang and Ainsworth (2012) used a semi-simulation approach to explore time accumulated in different activity intensities with different accelerometer wear time. Using a 14 hour wear day as a reference, samples of 10, 11, 12 and 13 hours wear were simulated. The authors noted that time in intensity increased with longer wear time, with the most notable differences in sedentary and light activity. Time spent sedentary was 456.9 minutes for the reference day, which decreased to 421.8, 389, 355.6 and 322.2 minutes with subsequent reductions in wear time. Smaller differences were apparent for moderate and vigorous activities, showing approximately 2 minute reductions in moderate activity between wear times, with reductions between 0.1 and 0.3 minutes for vigorous activity.

Following this, it is important to note that the potential for misclassification introduced by wear time requirements may also occur by failing to specify which hours of the day should be captured. For example, activity during the school day may differ from the activity undertaken after school; therefore by only

examining a small section of the day (e.g. 8 or 10 hours) time in activity intensities may be misclassified. In line with this, Jago, Anderson, Baranowski and Watson (2005) noted that children's physical activity pattern varies throughout the day, with MVPA accumulated more during the late afternoon (3pm – 7pm) while more sedentary activity seemed to be accumulated between early (6am – 3pm) and late (7pm - 12pm) segments of the day. Alternatively, comparisons may be made between those whose wear time is predominantly school day (e.g. 7am – 3pm), and those wear time is accumulated after school (e.g. 3pm – 11pm). For example, Pulsford, Griew, Page, Cooper and Hillsdon, (in prep) reported sedentary minutes for segments of the day before (7am – 8.59am), during (9am – 3pm) and after school (3pm – 11pm), recorded mean sedentary time was 37, 218, and 194 minutes respectively; by using these reported times combined with the above example of separate wear time periods (e.g. 7am – 3pm and 3pm – 11pm) the accumulation of sedentary time would be 255 minutes and 194 minutes respectively, demonstrating that determination of activity levels varies depending on the time of day captured. Importantly, neither of the estimated sedentary times above represents the extent of sedentary minutes accumulated if the monitor was to be worn from 7am – 11pm (449 minutes). Although this example demonstrates difference in sedentary time, similar results may be occurring for time in other activity intensities.

3.4.3. Cut-point misclassification

One of the most common sources of misclassification apparent in accelerometer measured physical activity comes from the application of intensity cut-points. One of the most widely used accelerometers; the ActiGraph (ActiGraph, Pensacola, FL) has been the subject of multiple calibrations within similar populations (e.g. children). The availability of multiple thresholds for use with children presents researchers with a conundrum about the most appropriate thresholds to employ, whilst employment of various cut-points may lead to the misclassification of physical activity levels in children.

The impact of employing different thresholds has been examined by relatively few studies, however, those that have, found large differences in time accumulated in PA intensities. For example, McLure et al. (2009) examined the

portion of children achieving the recommended 60 minutes MVPA per day (Department of Health, 2011) after the application of two sets of cut-points. When using the Freedson/Trost (Freedson et al., 2005, Trost, 2002) cut-points for children aged 9-10 (≥ 1100 counts \cdot min $^{-1}$), mean time in MVPA was 126.0 min \cdot day $^{-1}$, with 97% of participants achieving the physical activity guidelines. However when the Puyau, Adolph, Vohra and Butte (2002) cut-points were applied (≥ 3200 counts \cdot min $^{-1}$), mean time in MVPA decreased to 28.9 min \cdot day $^{-1}$, with the percentage of children meeting physical activity recommendations decreasing to 7%.

This problem was also examined by Stone et al. (2009), who applied three different activity intensity cut-points to ActiGraph data; sample specific (MVPA > 2910 counts \cdot min $^{-1}$), Mattocks et al. (2007) (> 3581 counts \cdot min $^{-1}$), and individualised thresholds developed using Ekelund, Aman and Westerterp's (2003) ArteACC method. Results showed that 8.5%, 48.9% and 100%, of children met the recommendations when sample specific, Mattocks et al. (2007) and ArteAcc, cut-points were applied respectively. Additionally, Richardson, Stewart-Brown, Wilcock, Oldfield and Thorogood (2011) reported differences of 44 minutes of accumulated MVPA between Freedson & Puyau cut-points. More recently, Guinhouya, Samouda & de Beaufort (2013) undertook a literature review of 35 studies of objectively measured physical activity level of children across Europe, reporting results for various cut-points; Using the Freedson age and sex specific equation, between 78 and 100% of children met the recommended level of MVPA, while studies using cut-points of around 2000 counts \cdot min $^{-1}$ reported 36 – 87% of children met recommendations. For the use of cut-points > 3000 counts \cdot min $^{-1}$, 3 to 9 % achieved sufficient MVPA while only 1% of children met MVPA recommendations with a cut-point of > 4000. Similarly in adolescents, the use of different cut-points resulted in between 4% and 100% of adolescents reported as meeting the MVPA guidelines. The findings from these studies highlight the importance of cut-point application upon estimates of physical activity levels in children; however, the number of thresholds examined in each study is relatively small compared to the number of cut-points available within the literature.

Recently, researchers have attempted to establish consensus within the literature as to which published thresholds may be most appropriate for accurately determining children's physical activity levels. Trost, Loprinzi, Moore and Pfeiffer (2011) compared the classification accuracy of a series of regularly used, published cut-points (Evenson, Cattellier, Gill, Ondrak & McMurray, 2008; Freedson et al., 2005; Trost, 2002; Mattocks et al., 2007; Puyau et al., 2002; Treuth et al., 2004) to a criterion measure (indirect calorimetry) for sedentary, light, moderate and vigorous physical activity intensities via ROC analysis. Across all intensities, Evenson et al. (2008) and Freedson / Trost cut-points reported consistently high sensitivity, specificity and area under the curve. The authors therefore, concluded that these cut-points were the most accurate method of assessing children's time in activity intensities and should be employed in future research in an attempt to establish comparability between studies. Following this, Kim, Beets and Welk (2012) conducted a systematic review of the cut-point literature, they reported that the Freedson and Evenson cut-points gave estimates ranging from 64–124 minutes per day and 47–61 minutes per day respectively, and subsequently noted that the difference between the lower-bound thresholds for MVPA is where possible misclassification of PA occurs.

The possibility of misclassification does not occur solely between the light and moderate boundaries, the discrepancies between cut-points for other intensities may also result in this misclassification. Despite this, researchers are compelled to continue to employ cut-point methods to define accumulated time in physical activity intensities for the immediate future. Thus the potential for misclassification of children's physical activity levels remains high. The matter becomes yet more confounded as researchers, perplexed as to the most suitable cut-points to use, create their own sample specific cut-points, adding to the plethora already available within the literature (e.g. Mckintosh et al., 2012).

The misclassification of physical activity has recently been examined with regard to the impact it may have upon relationships between physical activity and health related variables. Only a few studies have explored this concept in relation to physical health. For example, both Stone et al. (2009) and Atkin et al. (2013) have shown that the magnitude and significance of relationships between specific activity intensities and health variables differed depending

upon the ActiGraph cut-points employed. Similar results have also been noted with other accelerometer models. Bailey, Boddy, Savory, Denton and Kerr (2013) showed that in children aged 10 to 14 years, the use of different cut-points developed for the RT3 resulted in attenuation of relationships between PA and cardio metabolic risk factors. For example, a relationship between light PA and body fat percentage was apparent in girls with the use of the Rowlands et al., (2004) cut points ($r = .303$, $p = .044$), yet this relationship was attenuated with the employment of the Vanhelst et al. (2010) and Chu & McManus (2007) cut-points ($r = .055$ & $r = .283$, p 's > 0.05), indicating that misclassification not only alters time reported in activity intensities but also impacts on the observed relationships with health. The impact of misclassification upon health variables will be the focus of chapter four of this thesis.

A potential method of overcoming the misclassification due to cut-point method is to establish a more appropriate way to analyse the data. Newer accelerometers such as the GENEActiv and the ActiGraph GT3X+ allow examination of the raw acceleration data, rather than counts. Furthermore, pattern recognition technologies are continuing to be developed, which remove the need for examination of counts and classification of intensities. Yet, until these methods are available, the cut point method will continue to be used in physical activity research.

3.4.4. Selection bias and loss to follow up bias in physical activity

Selection bias can occur due to a number of reasons; Alonso et al. (2006) listed the following as sources of selection bias a) unwillingness to take part b) exclusion from the study due to missing information and c) through attrition or loss to follow up. Although some children or parents may be unwilling for their children to take part, the final two issues listed above can prove particularly pertinent.

As discussed previously, for accelerometer measured physical activity, children need to achieve a certain amount of wear-time in order for their data to be representative of 'habitual physical activity', the definition of non-wear time (periods of consecutive '0's) during the data reduction process, necessary with

accelerometry, will have a direct impact upon whether participants achieve the criterion level at which, a day is deemed 'valid' or not. This process therefore has the potential to introduce selection bias to a study through the inclusion or exclusion of participants who do or do not meet the desired criteria. Colley et al. (2010) addressed this issue by examining the mean wear hours per day when non-wear periods defined as 10, 20, 30 or 60 minutes of consecutive zeros were used in the data reduction process. The results showed that the mean wear time was 8.5, 9.7, 10.6 and 12.0 hours respectively. This difference in means indicates the potential for misclassification (as previously noted), however in turn the misclassification of wear time has a direct influence on the number of participants meeting the requirements for 'valid days', leading to selection bias being introduced to a sample. The same authors also reported an increase in the percentage of participants with a 'valid day' of data when the criterion for number of hours wear time was decreased (non-wear was classified as 60 minutes); when >14 hours were required, only 40% achieved the wear time, whereas at >10, >8 and >6 hours, 79%, 83% and 87% of participants achieved the set criterion. The above data reduction methods present opportunities for bias, through the removal of participants failing to achieve the required wear time criteria. The resulting sample may have different physical activity levels than those excluded and may also vary on other important covariates (Kang et al., 2009). Hypothetically, it could be that children achieving the wear time criteria are those that are more physically active, as they may be more inclined to adhere to wear time requirements than those who are more sedentary, and are aware that doing less activity may be an undesirable trait (e.g. social desirability). Furthermore, excluded participants may have failed to achieve recommended wear time due to prolonged sedentary behaviours that have been misclassified as non-wear. The data reduction system therefore introduces more potential bias into the study by potentially excluding some groups from analysis.

Bias in measurement of both physical activity and psychological well-being measurement may occur through participants being lost to follow up, whereby participants not completing a study may have different health responses to those who remain part of the research (Gail, 2005); more specifically the outcome(s) of interest, in this case either physical activity or psychological well-

being, may differ between those participants who remain in a study until it's termination and those that are lost to follow up (Sica, 2006). It may be, that children achieve increasingly fewer 'valid' accelerometer wear time days across multiple time points, with those undertaking less physical activity increasingly likely to fail to adhere to wear time.

3.5. Choice of methods

The measurement methods discussed above provide researchers with a great deal of choice when deciding upon an appropriate method to assess children's physical activity levels. Although advantages and disadvantages have been presented for each method in turn, along with the potential for methods to be susceptible to bias and finally, the possibility that PA frequency, duration and intensity may be misclassified by the use of inappropriate measurement methods, further consideration must be given to the trade-off between feasibility, validity and the reliability of methods. A discussion of these considerations occurs below, along with a summary table (table 3.2) outlining the choices available for researchers when measuring physical activity in children.

3.5.1. Validity, feasibility and reliability

The trade-off between the feasibility of employing methods and the validity of the methods themselves is an important issue that must be considered when assessing physical activity in any population. Validity refers to the extent to which an instrument or tool measures what it is designed to measure (Vincent, 2005), whereas feasibility is concerned with the practicality of employing measures. Figure 3.1 shows an adapted version of the figure presented by Esliger and Tremblay (2007) which demonstrates the trade-off that may occur when various measurement methods are employed. As can be seen in the figure, the most feasible method for measuring physical activity occurs in the form of self-report; however this offers limited validity, especially in children, due to the disadvantages discussed previously.

Alternatively, indirect calorimetry offers highly valid and reliable physical activity data, but is not a feasible method for measuring habitual physical activity. The figure suggests that objective measurements, such as motion sensors, provide the most suitable tradeoffs between feasibility and validity when it comes to acceptable physical activity measurement methods. The decision then comes down to which measurement method is more advantageous in children

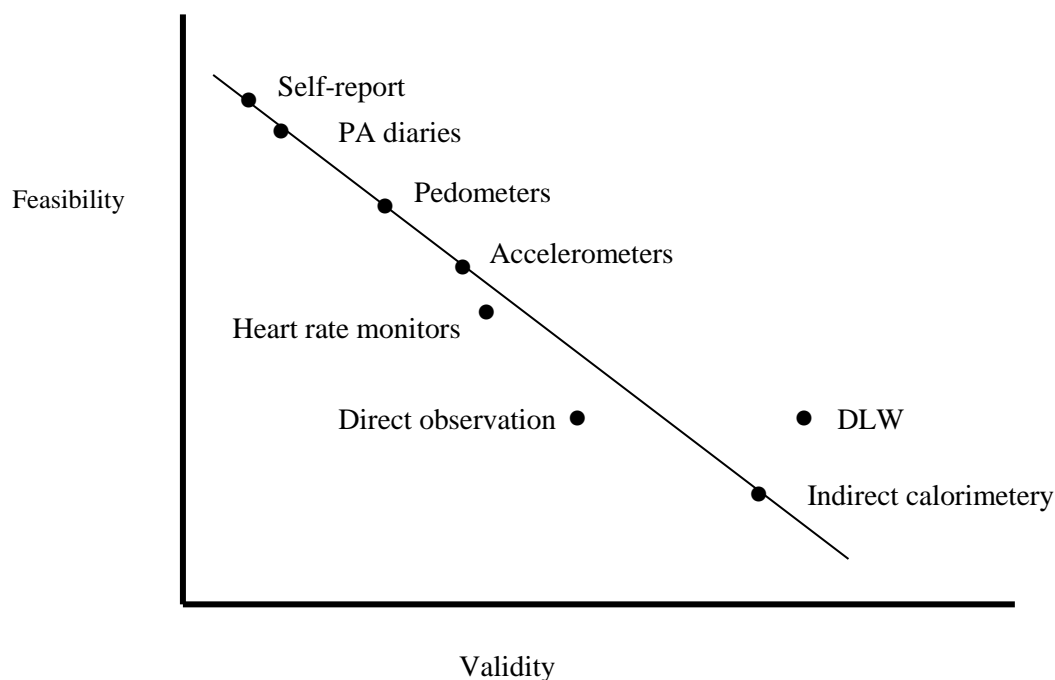


Figure 3.1. Trade-off between feasibility and validity of physical activity monitors, modified from Esliger and Tremblay (2007)

Also of consideration is the reliability of a measurement method; reliability refers to the consistency of the method to reproduce similar results under the same circumstances (Vincent, 2005). It can be influenced by many factors, including the characteristics of the test, the subject being measured, and the person taking the measurement (Kohl et al., 2000). In addition, the lack of stability in the measured behaviour itself (PA) may cause a reduction in reliability scores assessed over two time points (Kohl et al., 2000); this final point may be of particular importance when assessing physical activity, which may fluctuate depending on seasonality, school holidays and whether

measurement concerns week or weekend day activities. (Troost et al., 2000; Rowlands, 2007; Tucker & Gilland, 2007; Verbestel et al., 2011).

With respect to the measurement methods discussed above, most have demonstrated adequate reliability (e.g. Carter & Jekendrup, 2002; Goran, Pehlman & Danforth, 1994; Brown et al., 2006; Bauman & Merom, 2002; Tryon et al., 1991; Treuth et al., 2004; Esliger et al., 2011). However, it would seem that self-reported PA may be more effected by factors influencing reliability, despite this, reviews of the literature note that various self-report measures have demonstrated good test-retest reliability (Bauman & Merom, 2002; Brown, Trost, Bauman, Mummery, Owen, 2004). Another measure potentially susceptible to influential factors is direct observation. Here, in order for the observation system to be reliable agreement between raters must be high, in addition, reliability varies depending upon the complexity of the system (McKenzie, 2002), indicating that some systems may be unreliable depending upon the observers and the system itself. Evidence exists for the high reliability of each method of PA measurement; however, a balance must be achieved between reliability, validity and feasibility, in order for successful physical activity measurement to be achieved. A decision table is presented below to summarise these factors for each method.

Table 3.3. Summary of PA measurement methods

Method	Measurement	Outcomes	Research application	Sample Populations	Validity	Feasibility	Bias
Indirect calorimetry	Oxygen consumption	Energy expenditure	Assessing EE	Small numbers of Adults	High	Low	N/A
			Lab-based protocols	Small Numbers of children	High	Low	N/A
		Cardio-respiratory Fitness	Calibration of other measures				
Doubly labelled water	Isotope depletion	Total energy expenditure	Assessing TEE	Small numbers of Adults	High	Low	N/A
				Small Numbers of children	High	Low	N/A
Heart rate	Heart rate	Beat·min ⁻¹ Activity patterns: Frequency, Intensity, Duration EE through regression	Assessing patterns of MVPA Prediction of EE Lab & Field based research	All populations	Medium	Medium	Susceptible to influence from other sources

Accelerometers	Accelerations (m/s ² or g)	Activity patterns: Frequency, Intensity, Duration	Assessing time in sedentary, light, moderate & Vigorous	All populations	Medium	High	Data reduction methods cause potential bias & misclassification
	Counts		Estimation of EE				
Direct observation	Behaviour	Activity patterns Frequency, Intensity Duration	Assessing PA during specific time frames (e.g. PE lessons)	All populations although may be too intrusive for adults	High	Low	Raters must agree
		Estimated energy expenditure	Assessing context for PA behaviours				Potential for researcher bias during collection periods.
		Context of behaviour					
Self – report	Perceptions of behaviour	Estimates Frequency, Duration Intensity	Large, survey studies	Adults	Low	High	Social desirability
		Estimates energy expenditure	Population surveillance	Adolescents	Low	High	Recall bias
			Not advisable for children under 10 years	Very low	High	CMB Affectivity	

3.6. Method choice summary

The potential for misclassification in physical activity measurement is substantial if inappropriate measures are chosen. Previous literature within the psychological well-being and physical activity domain have mostly chosen to employ self-report measures to assess both physical activity behaviour and PWB, allowing for the possibility of common method bias and misclassification that may lead to an under- or over-estimate of the true association. With consideration to the methods discussed above, the use of an objective measure of physical activity, such as accelerometers, seems to provide the most appropriate solution in children. As PWB must be assessed via self-reported measures, the use of accelerometers overcomes the issue of common method bias, whilst maintaining a balance between feasibility and validity, and in turn allowing for greater precision in the exposure measure. Furthermore, accelerometers are particularly suited to research with children due to their small size, and ease of employment (Freedson et al., 2005; Rowlands et al., 2004). However, accelerometry is not without its disadvantages; the potential for misclassification of intensities is high due to data analysis techniques, wear time compliance and the use of absolute estimates of intensity. Though some methods have employed accelerometry to examine the relationship between PA and PWB, they have failed to take into account the multitude of methodological issues that accompany accelerometer use (Crocker et al., 2006) and result in misclassification of time in activity intensities and different bias being introduced, both of which may alter the relationship between PA and PWB. In order to gain a greater understanding of the relationships between the two constructs, further consideration should be given to how misclassification and bias can impact upon relationships observed.

3.7. Aim of thesis:

The aim of the following thesis therefore, is to address the impact of physical activity measurement upon the relationship between physical activity and children's psychological well-being. Specifically, this thesis will address how the decisions made during the accelerometer data reduction process results in misclassification of activity intensities and selection bias, and how this error can impact upon the observed relationships of PWB and PA. This will be achieved

through a number of studies that assess individual aspects of misclassification and bias upon this relationship. The first study (chapter four) will address the impact of cut-point classification upon PA and PWB relationships. This will be followed by an examination of compliance with popular wear time criteria over three time-points in study two (chapter five) and the assessment of the effect that selection bias and misclassification introduced through compliance with wear time requirements have upon the longitudinal relationship. The third study (chapter six) looks to examine whether compliance and selection bias systematically differ between trial arms of an intervention. Following this, the fourth study (chapter seven) aims to validate and calibrate a new accelerometer, which is hypothesised to increase compliance. Finally, study five (chapter eight) addresses whether 24-hour monitoring is acceptable in children, and whether misclassification alters the relationships observed between PA and PWB, independently to the effect of selection bias.

3.7.1 Hypothesis:

Based on the aims of this thesis and the previous literature it was hypothesised that the magnitude and significance of the observed relationship between children's physical activity and psychological well-being will vary as aspects of misclassification and bias relating to physical activity measurement are introduced to the relationship. Specific hypothesis are provided within each chapter.

CHAPTER FOUR

Associations between children's psychological well-being and physical activity intensity by accelerometer defined intensity levels

The aim of the following chapter is to examine how the direction and magnitude of the relationship between PA intensities and PWB is modified by the use of different accelerometer cut-points to define physical activity intensity. A synopsis of published cut-points will be provided along with a discussion of studies that have examined the impact of cut-point classification on reported relationships between PA and health variables. Hypotheses for the present study are then provided, followed by the methods of data collection and analysis. Results are presented and subsequently discussed in terms of the differences in time in activity intensities for each set of cut points and how reported relationships between PA intensity and health may differ depending upon the cut-points employed.

4.1. Introduction

The links between PA levels and PWB were highlighted in chapter two of this thesis. A major critique of this literature was that the reliance on self-report methods to measure PA may have led to misclassification of PA intensities which in turn may mask the true relationship between PA and PWB. Although some studies have attempted to use more objective measurement methods (E.g. Johnson et al., 2008; Parfitt et al., 2009; Hume et al., 2011) the potential for misclassification of PA intensity still exists. In order for research in this area to proceed, sources of misclassification with objective measures and the possible effect on observed relationships need to be examined.

As outlined in chapter three, a common source of misclassification when dealing with accelerometer measured physical activity is the employment of different cut-points of count data to estimate the intensity of PA. In April 2013, I undertook a literature search and found approximately 25 calibration studies in

children designed to identify cut points associated with predetermined intensities of PA for one of the most widely used accelerometers, the ActiGraph (CSA, 7164, GT1M, GT3X & GT3X+). The individual studies reported the creation of cut-points from various models of the ActiGraph for use with a variety of child populations, resulting in a multitude of different cut-points for determining PA intensity in children (e.g. Pate et al., 2006; Pulsford et al., 2011; Reilly et al., 2003; Sirard et al., 2005). Importantly, many more studies not in the search, may have undertaken their own calibration studies, the results of which are embedded in the text of publications not clearly identified as calibration studies. Therefore, the published calibration studies may not fully portray the extent of the cut-points currently in use.

The availability of so many published cut-points creates a dilemma for researchers deciding how to classify activity intensities from accelerometer cut points. Further, the lack of consensus within the published literature encourages some researchers to create their own, sample specific cut-points, adding to the plethora already available. The lack of a standardised methodology for undertaking calibration studies and the inconsistent use of cut-points makes comparison between studies problematic (Stone et al., 2009; Atkin et al., 2013). The choice of cut-point alters estimates of time spent in different PA intensities and as time spent in different PA intensities is important in the relationship between physical activity and health, the choice of cut-point is critical to fully understanding the health benefits of physical activity at different intensities.

Although the problems outlined above are well known (Intille, Lester, Sallis & Duncan, 2012), very little research has examined the impact that PA intensity misclassification by choice of accelerometer cut-point has upon relationships with health. A handful of studies (e.g. Reilly et al., 2008; McClure et al., 2009; Stone et al., 2009; Loprinzi et al., 2012) examining the use of various cut-points upon time spent in PA intensity levels have reported large discrepancies, particularly for time in MVPA. These discrepancies may occur due to wide disparities between the light / moderate intensity boundaries of published cut-points, allowing for easy misclassification of activity intensities (Kim et al., 2012). In other words the same physical activity can be classified as light intensity when using one set of cut-points and as moderate intensity when using

another. Discrepancies between published intensity boundaries are not limited to the light/moderate threshold, but are apparent between cut-points at all intensity boundaries to a certain extent, however the light/moderate boundary appears to have the more substantial differences.

For example, if a child performed an activity, (e.g. playground games) that produced an average oxygen consumption of $21.9 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$, the metabolic cost of the activity would be 3.7 METS (based on Harrell et al., 2005 resting MET values), categorising it as moderate intensity. Consider that the same activity resulted in an accelerometer output of $2398 \text{ counts}\cdot\text{min}^{-1}$, using the cut-points outlined in table 4.1, this activity would be correctly classified as moderate activity by the Evenson et al. (2008) cut-points, yet misclassified as light activity by the other cut-points points (Puyau et al., 2002; Treuth et al., 2004; Mattocks et al., 2007; Stone et al., 2009). Similarly, if a child averaged an oxygen consumption of $43 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ whilst running, equating to 7.3 METS (vigorous activity intensity), and produced an average accelerometer output of $5190 \text{ count}\cdot\text{min}^{-1}$, the activity would be correctly identified as vigorous intensity by the Evenson et al. (2008) and Stone et al. (2009) cut-points yet misclassified as moderate activity by the remaining sets of cut-points.

The potential for misclassification is highly important for establishing relationships between PA and health variables; if the true intensity was light but was coded as moderate, through choice of cut-point, then it is likely that the association between time spent in moderate intensity PA and the health variable would be underestimated, potentially leading to the false conclusions that no association was apparent (type 2 error).

Recent attempts to establish a consensus as to the most suitable activity intensity cut-points to apply in a child population have examined the classification accuracy of commonly used cut-points against a criterion measure. Originally, Trost et al. (2011) compared five cut-points to indirect calorimetry during both lab-based and free living activities performed for five minutes each. Results showed high classification accuracy of the use of the combined Freedson/Trost equation and the Evenson et al. (2008) cut-points, yet the authors concluded that the Evenson et al. (2008) cut-points should be employed in future studies, due to their higher classification accuracy across all

age groups. Subsequently, Alhassan et al. (2012) examined the accuracy of four cut-point regression equations, again using lab-based treadmill and free living activities. The authors noted that for overall activities (combined treadmill and free-living) only the Freedson equation accurately estimated children's energy expenditure, while Treuth et al. (2004), Puyau et al. (2002) and Trost et al. (1998) equations generally underestimated the energy cost of activities all activity combined. Despite this, Alhassan and colleagues reported that the Trost equation correctly classified the intensity of all of the activities performed. This difference between EE estimation and classification accuracy may occur as intensities are defined as falling between metabolic equivalent boundaries (e.g. 3 METs – 6 METs), therefore actual EE may be underestimated, yet the estimation by the equation may still lie between the MET boundaries. Importantly, both studies reported the Puyau cut-points to have the lowest classification accuracy of PA intensities.

These studies begin to address the most accurate method of accelerometer data reduction and interpretation. Further, cut-points with poor classification accuracy, which may lead to misclassification of PA intensities, are highlighted. Alhassan's assessment was based on cut-points developed from regression equations with a primary focus on accurate estimation of energy expenditure, hence, despite the recommendations of their use (Trost et al., 2011), the Evenson et al (2008) cut-points, developed via ROC curve analysis, could not be included to further assess their classification accuracy. The Evenson cut-points have yet to be compared in terms of time in PA intensities during habitual physical activity, which would be difficult to measure due to the restrictions associated with criterion measures, as discussed in chapter three. Nevertheless, the greater classification accuracy of Evenson et al. (2008) may provide a more accurate estimate of PA and health relationships due to the lower likelihood of misclassification of PA intensity. Whether the relationships observed between PA and health differ with the application of the cut-points compared to others are yet to be observed.

In an attempt to investigate the effect of cut-point choice upon PA and physiological health relationships (waist circumference, blood pressure and $\dot{V}O_2$ peak) in children, Stone et al. (2009) applied previously published,

(Ekelund et al., 2003; Mattocks et al., 2007) and sample specific cut-points to ActiGraph count data. The authors found that relationships between moderate intensity PA / vigorous intensity PA and waist circumference / $\dot{V}O_2$ peak did not differ between sample specific and Mattocks et al. (2007) cut-point application. However the relationship between moderate intensity and waist circumference became non-significant with the application of the Ekelund et al. (2003) individual - specific cut-points (ArteACC). Additionally, they reported non-significant relationships of similar magnitude between activity intensity and blood pressure for all three sets of cut-points. The authors concluded that the relationships between PA intensities and health variables did not vary depending upon the cut-points employed, despite the attenuation of one relationship.

More recently, Atkin et al. (2013) examined the effect of cut-point choice on the relationship between sedentary time, adiposity and metabolic risk; applying cut-points of < 100 (Matthews et al., 2008), < 500 (Sardinha et al., 2008), < 800 (Puyau et al., 2002) and < 1100 counts·min⁻¹ (Reilly et al., 2003) to define sedentary behaviour. The authors reported that cut-point use, demonstrated no effect on the relationship between sedentary time and adiposity, but that the magnitude of the relationship between metabolic risk and sedentary time was moderated by choice of cut-points; stronger relationships were apparent with the use of higher cut-points. The strongest relationship was found between <1100 counts·min⁻¹ defined sedentary time and metabolic risk. These findings are interesting, as such high cut-points are more likely to encompass misclassified light activity rather than sedentary behaviour. Indications of this can be seen in Fischer et al. (2012) who, after comparing accelerometer output of observed sedentary behaviour to published thresholds reported that <100 counts·min⁻¹ was the most appropriate cut-point for classifying sedentary behaviour. The results of Atkin and colleagues highlight the effect that cut-point choice can have upon relationships and the consequences that can occur from the use of inappropriate choices; the fact that the observed relationship became stronger as higher cut-points were employed may lead to a lead to a type 1 error, with the conclusion that a relationship exists between sedentary behaviour and metabolic risk, whereas the stronger relationships are more likely to be occurring between light intensity and metabolic risk. This is something the

authors acknowledge, noting that a combination of time in sedentary and large amount of time in light activity may have adverse associations with health factors in children. In relation to PWB, Page et al. (2010) noted that the relationships between sedentary time and fewer difficulties, and greater MVPA related to lower difficulties (accounting for sedentary) were not a function of cut-point, however the authors only examined this by employing the Puyau et al. cut-points.

The studies discussed above present equivocal evidence pertaining to the impact of cut-points upon the relationship between PA and health variables, with cut-point application appearing to impact upon some relationships and not others, for Stone et al. (2009) this was demonstrated by the attenuation of only one relationship with the application of one cut-point. For the Atkin et al. (2013), though only one relationship was affected, the magnitude continued to increase with the application of continually higher cut-points. A limitation of both studies are that few cut-points are examined in each, meaning that studies looking at the same relationships may find different results by employing separate cut-points to those examined in these studies. In particular, Stone et al. (2009) did not consider the Evenson et al. (2008) cut-points recommended in the previously described comparison study (Troost et al., 2011), which differ considerably from those that were employed (Stone et al., 2009 & Mattocks et al., 2007; see table 4.1.) and therefore may affect the relationships observed.

Additionally, despite appropriate theoretical reason to examine only one or two intensities (E.g. sedentary or MVPA), as the previously described studies did, the examination of the effect of cut-point classification on relationships across all intensities is warranted in order to assess whether cut-point classification has greater impact for relationships at one particular intensity. Finally, the impact of cut-point application upon the detection of relationships between PA and other health variables, such as PWB, are yet to be fully examined.

The aim of the present study was to assess whether reported relationships between children's PA intensity and PWB differ as a function of the cut-points employed. Despite results from previous research being equivocal, it follows that if cut-points differ in their classification accuracy (Troost et al., 2011) and

result in varying quantities of time reported in PA intensities (McLure et al., 2009), then relationships detected between PA intensity and health would vary depending on the cut-points employed. Consequently, two hypotheses were created. First, that time recorded in activity intensities will vary depending upon cut-points employed and secondly, that detected relationships between PA intensity and PWB will differ when alternative cut-points are used.

4.2. Method

4.2.1. Participants

A total of 82 children from the Devon area were recruited for the study, 49 ($n = 12$ males, $n = 37$ females) of which provided adequate data to be included in the final analysis. Participants were between the ages of 9 and 11 years ($\bar{X} = 9.8 \pm 0.7$), had a mean (\pm SD) height of 140.8 cm (± 7.6), mass of 37.2 kg (± 7.2) and BMI Z-score of .53 ($\pm .71$). The institutional ethics committee granted approval for the study (7th December, 2010; see appendix 1.1, p. 294). Prior to data collection, parental informed written consent (see appendix 2.1, p. 297) and the children's assent were obtained (See appendix 2.2 p. 302).

4.2.2. Protocol

Upon commencement of the study, height (cm) and mass (kg) were obtained from each child and were subsequently used alongside their date of birth and date of visit to calculate BMI Z-score, based on the US CDC 2000 growth scores (National Centre for Chronic Disease Prevention and Health Promotion; 2000). BMI Z-score is thought to be an appropriate way to assess adiposity when examined cross-sectionally (Cole, Faith, Pietrobelli & Heo, 2005) and allows for comparison between children of different ages and stages of maturation.

An ActiGraph GT1M or GT3X was attached to an elasticised belt, and positioned at the right hip, on the anterior axillary line. These ActiGraph models measure 38 mm x 37 mm x 18 mm (L x W x H) and weigh 27g. The GT1M can measure data on up to two axes, (vertical and antero-posterior), while the GT3X measures in up to three axes (as GT1M plus medio-lateral) in epochs ranging from one to 60 seconds, both devices measure accelerations between 0.05 and

2.5g at a frequency of 30 Hz (Sasaki, John & Freedson, 2011). Children were asked to wear the monitors for a period of seven days, with removal required for sleeping and water based activities. Participants were instructed to return the monitor to the correct position as soon as possible once activities requiring removal ceased. Stickers were used to indicate the correct orientation of the monitors to aid children in repositioning the devices. Data was collected on the vertical axis for both GT1M and GT3X models; different devices were used due to limited availability of both models. Output from the vertical axis of the newer ActiGraph GT3X has been found to be comparable to that of the GT1M (Sasaki, et al., 2011; Hänggi, Phillips & Rowlands, 2013). Monitors were set to record at 2 second epochs in order for sufficient storage of seven days of data, which was the highest frequency attainable for the required measurement period with the limited memory capacity of the GT1M model.

4.2.3. Psychological Well-being

Three measures of PWB were administered in order to obtain levels of both positive and negative aspects of well-being. These were the Child Depression Inventory (CDI; Kovacs & Beck, 1977), the State–Trait Anxiety Inventory for Children (STAIC; Spielberger, 1973) and the Child and Youth Physical Self-Perception Profile (PSPP-CY; Whitehead, 1995), which also included a physical self-worth subscale (Whitehead & Corbin, 1991) and a measure of global self-worth (Harter, 1985).

4.2.3.1. Child Depression Inventory (CDI)

The CDI (Kovacs & Beck, 1977; see appendix 3.1 p. 322), an adaption of the Beck Depression Inventory (Beck, Ward, Mendelson, Mock & Erbaugh, 1961) is a 27 item instrument designed to assess levels of depression in children. The instrument presents participants with a choice of three statements for each item; children are then required to choose which statement most describes their thoughts and feelings during the previous two weeks. For example, item eight on the CDI asks children to choose from the following statements. ‘All bad things are my fault’, ‘Many bad things are my fault and ‘Bad things are usually not my fault’. Items are scored on a 0 – 2 scale, with total score ranging from 0 – 54, higher scores on this scale represent higher levels of depression. Many

items are reverse scored and therefore need to be transformed prior to analysis. The questionnaire is phrased in an appropriate manner for children between the ages of eight and 13 years. One item pertaining to suicide was deemed unsuitable for children of this age and therefore removed, reducing the total score to 52.

The psychometric properties of the CDI were examined by Saylor, Finch, Spirito and Bennett (1984), reporting high internal consistency ($\alpha = .94$) and high split-half correlations for odd/even split and first/second half split ($r = .61$ & $.73$, respectively). Other studies have also reported high internal consistency (Cronbach's α ranging from $.84$ to $.86$; Smucker, Craighead, Craighead and Green, 1986; Parfitt & Eston, 2005) and test – retest reliability (ranging from $.49$ to $.77$; Smucker et al. 1986). Cronbach's alpha score for the present study was $.86$.

4.2.3.2. State -Trait Anxiety Inventory for Children (STAIC)

The STAIC (Spielberger, 1973) can be used to assess both state and trait levels of anxiety in children. For the present study, trait rather than state anxiety was assessed (see appendix 3.2 p.325) as the state subscale measures fluctuating, momentary anxiety (Finch, Montgomery & Deardorff, 1974) which can be affected by a number of variables, whereas trait anxiety is a more stable construct, which takes time to change. To remain consistent across all inventories, participants were asked to record their thoughts and feelings with regard to the past two weeks.

The trait anxiety scale of the STAIC consists of 20 items with participants asked to indicate the frequency that they experienced each item on a 3 point – likert scale. For example, item 7 on the instrument 'I get upset at home', is responded to as; 'hardly ever', 'sometimes' or 'often'. The resultant score for the inventory ranges from 20 – 60, with higher scores relating to higher levels of trait anxiety.

High Cronbach's alpha values for the STAIC-trait have previously been reported (α ranging from $.76$ to $.89$; Spielberger et al., 1973; Parfitt & Eston., 2005; Parfitt et al., 2009) when employed with child and adolescent populations.

Cronbach's α reported for the STAIC in the present study was .77, demonstrating adequate internal consistency.

4.2.3.3. Child and youth physical self-perception profile (PSPP – CY)

The final measure used to assess children's PWB was the child and youth physical self-perception profile (Whitehead, 1995). In addition to this, a physical self-worth subscale (Whitehead & Corbin, 1991) and Harter's (1985) global self-worth subscale were administered. This resulted in a 36 item inventory (See appendix 3.3 p. 326). The questionnaire presents participants with two options for each item, known as a structured alternative format, in an attempt to reduce any social desirability and ensure that either choice is equally acceptable (Harter, 1985; Whitehead, 1995). Children are first asked to choose between two statements, deciding which is more accurate for them, then asked to choose whether this statement is 'sort of true' or 'really true' for them. For example:

Really true for me	Sort of true for me			Sort of true for me	Really true for me	
<input type="checkbox"/>	<input type="checkbox"/>	Some kids think that they have stronger muscles than other kids their age	BUT	Other kids feel that they have weaker muscles than other kids their age	<input type="checkbox"/>	<input type="checkbox"/>

Items are then scored on a scale of 1 to 4, where 4 signifies the higher level of self-judgement and 1 represents the lower judgement of self. For the example above from left to right would be 4, 3, 2, 1. Due to counterbalancing of the questions, some items are scored 1, 2, 3, 4.

A combination of measures was used in order to enable assessment of each level of the self-worth hierarchical model. All sub-scales consisted of 6 items. At the apex of the model is one's global self-worth, which assesses the extent to which a child is happy with themselves as a person (Harter, 1985). An example of the alternative statement states for this sub-scale is 'Some kids like the kind of person they are' BUT 'Other kids often wish they were someone else'. At the domain level, the physical self-worth scale developed by Whitehead and Corbin

(1991) was used to assess participants overall feelings of their physical self, for example, '*Some kids don't feel very confident about themselves physically*' BUT '*Other kids feel good about themselves physically*'. The Whitehead (1995) scale measures children's perceptions of the physical self-worth sub-domains; perceptions of sport competence, perceptions of physical strength, perceptions of body attractiveness and perceptions of condition / stamina. Examples of the items included in these subscales are given below.

Sport competence: 'Some kids do very well at all kinds of sport' BUT 'Other kids don't feel that they are good when it comes to sports'.

Physical strength: 'When strong muscles are needed, some kids are the first to step forward' BUT 'Other kids are the last to step forward when strong muscles are needed'.

Body Attractiveness: 'Some kids are pleased with the appearance of their bodies' BUT 'Other kids wished that their bodies looked in better shape physically'.

Physical condition: 'When it comes to activities like running, some kids are able to keep going' BUT 'other kids soon have to quit and take a rest'.

All subscales have previously demonstrated high reliability (Harter, 1985; Whitehead, 1995; Parfitt et al., 2009); alpha values in the current studies ranged from .68 to .80.

4.2.4 Data reduction

Participants were excluded from the final analysis if they failed to achieve the minimum recommended wear time of three weekdays and one weekend day of PA monitoring (Troost, 2002). To be classed as a valid day, the participant must have achieved at least 480 minutes (eight hours) of wear time over the day. Non-wear time was defined as 30 minutes consecutive zero counts. These criteria excluded 25 children, resulting in 69.5% of participants achieving \geq four days of accelerometer wear; of the 25 children excluded, 2 data files could not be extracted, 1 monitor was not returned, and 3 children did not provide data due to absences during the measurement week. Compliance with wear time from available files was therefore 75%. Additionally, one participant was

excluded based upon a technical error with an accelerometer; one participant was excluded from the analysis as they failed to provide adequate data to calculate their BMI Z-score. Finally, six subsequent participants provided insufficient PWB data.

4.2.5. Data Analysis

Accelerometer data was processed using Kinesoft software (Version 3.3.62; Kinesoft). A series of cut-points, consistent with those used in Trost et al.'s (2011) comparison study, with the addition of Stone et al. (2009) cut-points, were applied to the data in order to assess the impact of intensity classification upon the relationship between PA intensity and PWB. Table 1 details the specific cut-points by author. These widely applied cut-points do not all use the same MET definition of moderate intensity activity. Puyau and Freedson use a 3 MET definition, whereas Mattocks provide cut-points for a 3 and 4 MET definition, for the present study only the 4 MET Mattocks values have been employed, representing the choices made in the wider literature, where the 4 MET cut-point is often employed (e.g. Tobias, Steer, Mattocks, Riddoch & Ness, 2007) and the 3 MET Mattocks cut-point seldom used.

Table 4.1. Published cut-points for activity intensities established with ActiGraph models 7164 and GT1M.

Author	Activity intensity cut-points			
	Sedentary	Light	Moderate	Vigorous
Evenson et al. 2008	≤ 100	> 100	≥ 2296	≥ 4012
Stone et al. 2009	≤ 300	> 300	≥ 2910	≥ 5010
Freedson / Trost ¹				
Mattocks et al. 2007	≤ 100	> 100	≥ 3581	≥ 6130
Puyau et al. 2002	< 800	≥ 800	≥ 3200	≥ 8200
Treuth et al. 2004	≤ 100	> 100	≥ 3000	> 5200

Time spent in activity intensities was computed for each set of cut-points presented in table 4.1. Repeated measures ANOVAs were then used to assess

¹. Freedson / Trost equation. METs = 2.757 + (0.0015 x counts per minute) – (0.08957 x age) – 0.000038 x counts per minute x age)

whether differences in time spent in each activity were present depending upon the cut point employed. Pairwise comparisons using a Bonferroni correction were undertaken to establish where any differences lay. In accordance with the findings of Parfitt et al. (2009) zero-order and partial correlations, controlling for BMI Z-score, were undertaken to assess the relationships between time spent at each activity intensity (categorised by the above cut-points) and PWB. Any relationships observed between a psychological variable and multiple cut-points were tested for significant differences using Meng, Rosenthal & Rubin's (1992) method of comparing correlated correlation coefficients (see appendix 4.1. p 331 for worked example).

4.3 Results

4.3.1. Descriptives

Means and standard deviations for psychological well-being, are presented in table 4.2.

Table 4.2. Means \pm SD for PWB variables. Possible ranges of scores for each scale are given in brackets.

Psychological variables	Means \pm SD
Depression (CDI; 0 -52)	6.8 \pm 5.6
Trait anxiety (STAIC; 20 - 60)	30.8 \pm 6.5
Self-perceptions	
Global self-worth (GSW; 1 - 4)	3.4 \pm 0.6
Physical self-worth (PSW; 1 - 4)	3.1 \pm 0.5
Sport competence (1 - 4)	3.0 \pm 0.6
Attractive Body (1 - 4)	2.9 \pm 0.6
Physical Strength (1 - 4)	2.7 \pm 0.5
Physical conditioning (1 - 4)	3.1 \pm 0.5

4.3.2. ANOVAs

A series of repeated measures ANOVAs revealed significant differences for time spent in activity intensities dependent upon the cut-points applied. Differences were apparent for time in sedentary ($F_{(5,240)} = 1328.83$, $p < 0.001$), light activity intensity ($F_{GG(1.631,78.30)} = 1094.01$, $p < 0.001$), moderate activity intensity ($F_{GG(1.413, 67.82)} = 801.42$, $p < 0.001$), and vigorous activity intensity

($F_{GG(1.1226, 58.868)} = 531.50, p < 0.001$) along with time in MVPA ($F_{GG(1.1226, 54.028)} = 1268.37, p < 0.001$). Figure 4.1 depicts where differences lie for each activity intensity.

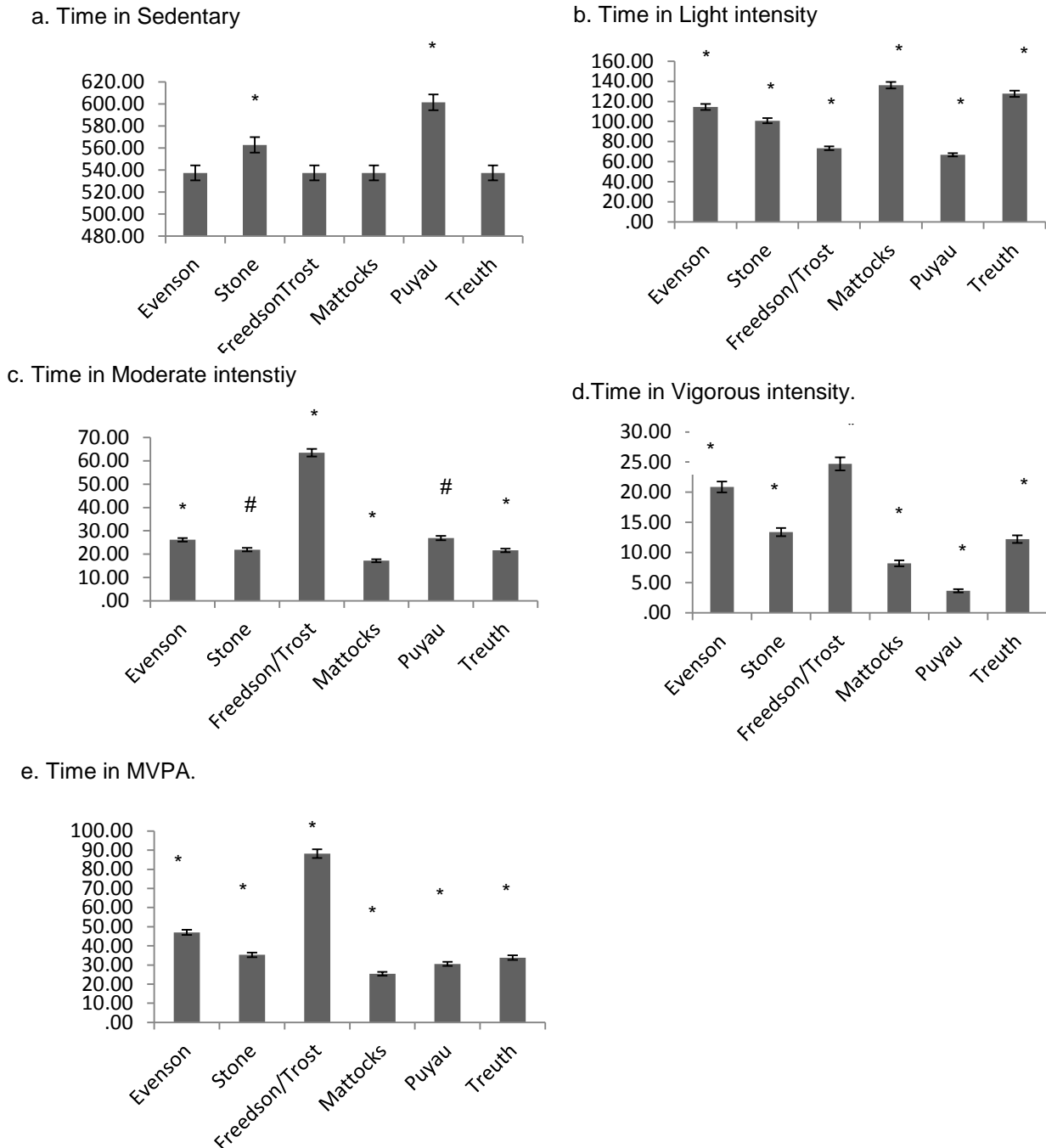


Figure 4.1. Significant differences between times spent in each activity intensity when classified by different cut-points * Significantly different from all other cut-points; # significantly different from all other cut-points except others marked by #

4.3.3. Correlations

No significant relationships were apparent between mean counts per day ($\bar{X}=313254 \pm 61279$) and psychological well-being variables. However, the relationships between time spent in different physical activity intensities and psychological variables (depression, anxiety and physical self-perceptions) differed slightly depending on the cut-points employed. Zero-order and partial correlations, controlling for BMI Z-score are presented in tables 4.3 – 4.7. No significant relationships were apparent when using Puyau cut-points to define PA intensity.

Table 4.3. Zero-order and partial correlations between cut-point defined sedentary behaviour and PWB variables (n = 49)

*p < 0.05

Cut-points		CDI	STAIC	SC	AB	PS	PC	PSW	GSW ²
EV ³	Zero-order	-.304*	-.075	.130	.215	.141	.057	.237	.221
	Partial	-.305*	-.076	.130	.215	.145	.058	.222	.237
ST	Zero-order	-.291*	-.072	.103	.204	.129	.047	.217	.211
	Partial	-.293*	-.075	.103	.203	.135	.047	.213	.217
F/T	Zero-order	-.304*	-.075	.130	.215	.141	.057	.237	.211
	Partial	-.305*	-.076	.130	.215	.145	.058	.222	.237
MT	Zero-order	-.304*	-.075	.130	.215	.141	.057	.237	.221
	Partial	-.305*	-.076	.130	.215	.145	.058	.222	.237
PU	Zero-order	-.261	-.063	.057	.177	.108	.028	.181	.192
	Partial	-.263	-.066	.057	.177	.114	.028	.194	.181
TR	Zero-order	-.304*	-.075	.130	.215	.141	.057	.237	.221
	Partial	-.305*	-.076	.130	.215	.145	.058	.222	.237

The relationship between PU cut-point defined sedentary and depression approached significance for both zero order and partial correlations (p = .07 & .071 respectively).

² CDI, depression; STAIC, Anxiety; SC, sports competences; AB, Attractive body; PS, Physical Strength; PC, physical conditioning; PSW, physical self-worth; GSW, Global self -worth

³ Ev, Evenson cut-points ; ST, Stone cut-points; F/T, Freedson / Trost equation; MT, Mattocks cut-points; PU, Puyau cut-points; TR, Treuth cut-points

Table 4.4. Zero-order and partial correlations between cut-point defined light activity and PWB variables (n = 49)

Cut-points		CDI	STAIC	SC	AB	PS	PC	PSW	GSW
EV	Zero-order	.180	.063	-.300*	-.150	-.135	-.128	-.204	-.068
	Partial	.178	.058	-.299*	-.151	-.131	-.128	-.204	-.065
ST	Zero-order	.205	.063	-.285*	-.148	-.121	-.105	-.189	-.033
	Partial	.204	.062	-.285*	-.149	-.122	-.105	-.189	-.032
F/T	Zero-order	.117	.019	-.236	-.081	-.108	-.069	-.168	-.067
	Partial	.114	.011	-.234	-.082	-.101	-.070	-.168	-.063
MT	Zero-order	.174	.052	-.262	-.112	-.091	-.073	-.161	-.015
	Partial	.172	.046	-.261	-.112	-.084	-.073	-.161	-.012
PU	Zero-order	.232	.069	-.244	-.129	-.094	-.060	-.153	.023
	Partial	.233	.071	-.245	-.129	-.098	-.060	-.153	.023
TR	Zero-order	.181	.054	-.283*	-.133	-.115	-.101	-.184	-.037
	Partial	.179	.049	-.282	-.133	-.110	-.101	-.184	-.034

*p < 0.05

The relationship between Mattock's cut-point defined light activity and sports competence approached significance for both zero-order ($p=0.069$) and partial correlations ($p=0.073$). Furthermore, although only the zero-order correlation between light activity and sports competence was significant when using the Treuth cut-points, the partial correlation approached significance ($p=0.053$).

Table 4.5. Zero-order and partial correlations between cut-point defined moderate activity and PWB variables (n = 49)

Cut points	CDI	STAIC	SC	AB	PS	PC	PSW	GSW
EV Zero-order	.045	-.012	.084	.146	.207	.256	.154	.234
Partial	.043	-.018	.086	.145	.223	.256	.154	.239
ST Zero-order	.036	.030	.088	.134	.266	.264	.160	.147
Partial	.034	.024	.090	.133	.284	.264	.161	.151
FR Zero-order	.208	.066	-.215	-.098	-.031	-.022	-.099	.063
Partial	.208	.064	-.214	-.098	-.028	-.022	-.099	.066
MT Zero-order	.077	.048	.020	.022	.208	.178	.071	.020
Partial	.074	.042	.022	.022	.226	.178	.072	.024
PU Zero-order	.041	.014	.064	.024	.214	.208	.077	.042
Partial	.034	.000	.069	.023	.243	.208	.078	.051
TR Zero-order	.038	.036	.077	.120	.262	.254	.151	.129
Partial	.035	.030	.080	.120	.281	.254	.152	.134

*p < 0.05

For relationships between moderate activity and psychological well-being variables, a number of relationships approached significance; relationships with perceptions of physical strength approached significance for zero-order and partial correlations when the Stone (p=0.065 & p=0.051 respectively) and Treuth (p=0.069 & 0.053 respectively) cut-points were used. Similar findings were apparent between perceptions of physical conditioning and moderate activity defined by Evenson (Zero-order, p= 0.076; partial, p= 0.079) and Stone (Zero-order, p=0.066; Partial, p = 0.069) cut-points.

Table 4.6. Zero-order and partial correlations for cut-point defined vigorous activity and PWB variables. (n = 49)

Cut-points	CDI	STAIC	SC	AB	PS	PC	PSW	GSW
EV Zero-order	-.033	-.025	.080	-.037	.181	.177	.005	-.019
Partial	-.045	-.051	.089	-.039	.223	.187	.007	-.005
ST Zero-order	-.079	-.058	.102	-.074	.129	.149	-.038	-.042
Partial	-.095	-.092	.113	-.076	.179	.150	-.036	-.025
FR Zero-order	-.011	.016	.026	-.061	.158	.124	-.012	-.018
Partial	-.023	-.008	.033	-.062	.197	.124	-.011	-.005
MT Zero-order	-.158	-.095	.162	-.050	.141	.171	-.029	-.018
Partial	-.178	-.134	.176	-.053	.197	.173	-.026	.002
PU Zero-order	-.237	-.106	.190	.010	.179	.201	-.005	.022
Partial	-.260	-.148	.205	.008	.239	.203	-.003	.043
TR Zero-order	-.093	-.067	.115	-.071	.125	.149	-.036	-.040
Partial	-.110	-.102	.126	-.074	.176	.150	-.034	-.023

*p < 0.05

Table 4.7. Zero-order and partial correlations for cut-point defined MVPA and PWB variables. (n = 49)

Cut-points	CDI	STAIC	SC	AB	PS	PC	PWS	GSW
EV Zero-order	.003	-.023	.100	.056	.235	.259	.088	.117
Partial	-.006	-.044	.106	.055	.271	.260	.090	.129
ST Zero-order	-.021	-.014	.112	.043	.239	.249	.080	.069
Partial	-.032	-.036	.120	.042	.279	.250	.082	.082
FR Zero-order	.142	.054	-.140	-.098	.053	.043	-.076	.037
Partial	.137	.042	-.136	-.099	.073	.043	-.075	.044
MT Zero-order	-.025	-.014	.092	-.010	.207	.202	.034	.005
Partial	-.036	-.036	.099	-.011	.246	.202	.036	.017
PU Zero-order	-.026	-.015	.103	.023	.227	.228	.064	.041
Partial	-.037	-.038	.111	.022	.267	.229	.066	.054
TR Zero-order	-.025	-.013	.111	.039	.236	.244	.078	.062
Partial	-.036	-.035	.119	.038	.276	.245	.080	.074

*p < 0.05

Relationships between cut-point defined time in MVPA and perception of physical strength approached significance when controlling for BMI Z-score when Evenson (p = .062), Stone (p=.055), Puyau (p=.066) and Treuth (p=.058) cut-points were employed.

Comparison between significant relationships occurring between a PWB variables and PA defined by multiple cut-points revealed no significant differences between relationships dependent on which cut-points were employed ($p > 0.05$).

4.4. Discussion

The aim of the present study was to assess whether relationships between physical activity intensities and psychological well-being varied with the employment of different accelerometer cut-points. Significant differences were observed for time spent in activity intensities when defined by various published cut-points. Those relationships apparent between time in activity intensities and psychological well-being across multiple cut-points did not differ from each other. Interestingly though, the application of some cut-points resulted in the attenuation of several relationships, with some becoming non-significant with the application of certain cut-points.

4.4.1. Time in activity intensities

Prior to examining the PA and PWB relationships, time in sedentary, light, moderate, vigorous and MVPA intensities were calculated for each set of cut points. It was expected that time in intensities would vary depending upon the cut-points employed, results showed this to be the case. Time in sedentary activity revealed only two significant differences; appearing when the Stone et al. (2009) and Puyau et al. (2002) cut-points were applied; these were the only cut-points to deviate from the standard $<100 \text{ counts}\cdot\text{min}^{-1}$, a threshold that appears to be widely accepted as the most appropriate when classifying sedentary activity (Troost et al., 2011; Yildirim et al., 2011). This point was also noted by Fischer et al. (2012), who examined the classification of observed sedentary behaviours using three of the cut-points employed in the present study (<100 , <300 & < 800) and the Reilly cut points (<1100). The authors reported that for observed behaviours, accelerometer output in the 75th percentile of $\text{counts}\cdot\text{min}^{-1}$ only exceeded the lowest cut-points during short periods in sedentary activity, yet did not exceed the <300 threshold. Furthermore, they noted that using the higher cut-points, for example ($<800 \text{ count}\cdot\text{min}^{-1}$) would lead to an over estimation of time in sedentary, as none of

the observed behaviours exceeded 300 counts·min⁻¹. Though during a free play activity, counts in the 25th percentile generally fell below 800 counts·min⁻¹, demonstrating that intensities above those classed as sedentary would be captured with the use of this high cut-point. For the remaining intensities in the present study, significant differences were apparent between all cut-points, with the exception being at moderate intensity, for which no differences were found between ST and PU cut-points.

As anticipated, the cut-points with the highest lower boundary for any given intensity resulted in the lowest time accrued in that intensity; for example in light activity, the PU threshold was highest (≥ 800 counts·min⁻¹), consequently the time reported in light activity whilst using these cut-points was the lowest (67 minutes); for moderate activity MT cut-point was highest (≥ 3581 counts·min⁻¹) resulting in 17.9 minutes accumulated at this intensity; while PU again resulted in the lowest time in vigorous activity (3 minutes) due to the highest lower boundary (> 8200 counts·min⁻¹). For MVPA, Mattocks cut-points (as above) revealed the shortest time (25 minutes). These results are similar to those found by Loprinzi et al, (2012), who noted that lowest estimates of MVPA were achieved when employing MTs cut-points, with the highest recorded time in MVPA apparent with the application of the FR equation, as was apparent in the present study.

These results serve to reinforce the findings of previous studies (Reilly et al., 2008; McClure et al., 2009; Stone et al., 2009; Fischer et al., 2012), whilst the examination of a wider variety of cut-points allows the issue to be more thoroughly examined and further highlights the impact that cut-point application has on classification of PA. Interestingly, on average, participants failed to achieve the recommended guidelines of ≥ 60 minutes daily of MVPA with the application of all cut-points except those derived from the Freedson equation.

4.4.2. PA and PWB relationships

4.4.2.1 Magnitude and significance

Correlational analysis revealed relationships between time in sedentary behaviour and measures of depression, and between time spent in light activity

and perceptions of sport competence for multiple cut-points; both of which were weak and in a negative direction. For the sedentary and depression correlation, the strength of relationship remained consistent across the majority of cut-points, which was to be expected as four of the cut-points had the same sedentary threshold ($<100 \text{ counts}\cdot\text{min}^{-1}$). When the slightly higher ST sedentary cut-points were employed, the magnitude of the relationship marginally decreased, still, the correlation did not differ significantly from those observed with the four other cut-points. Additionally, with the application of the PU cut-points, this relationship became non-significant, this is not surprising as the high sedentary/ light boundary apparent in the PU cut-points allows for greater misclassification of light activities as sedentary, diluting the relationship observed with the application of all other cut-points.

Similar results were also seen for the significant light/sport competence relationship, with no significant differences in the strength of relationships occurring between the cut-points. As with the results discussed above, these significant relationships between light and sports competence were not apparent across all cut-points; in fact, relationships were attenuated and became non-significant with the application of F/T, MT and PU cut-points. Once again, the activities included within the light activity boundaries for these cut-points may explain the attenuation of the relationships. MT has the highest upper-bound threshold for light activity, potentially allowing moderate activity to be misclassified as light. Additionally, PU not only has a high lower bound threshold ($800 \text{ counts}\cdot\text{min}^{-1}$) but also a high upper bound threshold ($3200 \text{ counts}\cdot\text{min}^{-1}$), therefore capturing less light activity and potentially more moderate activity.

When examining the non-significant relationships between activity intensity and PWB variables, large differences in magnitude can be seen. For example, time in MVPA classified using the ST revealed a non-significant relationship with perceptions of physical strength ($r = .239$, $p = .055$), whereas the same relationship using F/T to define MVPA revealed an r value of $.053$. Additionally, the direction of some relationships was also found to change with the application of various cut-points; the relationships between MVPA and sport competence became negative with the use of the F/T equation.

Relationships between PA intensities and PWB variables were shown to alter in magnitude, and become non-significant with the application of different cut-points; it is thought that these differences can be attributed to how intensities are classified and the extent to which those classifications are accurate. As determination of intensity depends on the counts·min⁻¹ at the upper and lower boundaries, those sets of cut-points with wide ranging boundaries will capture a broader variety of activity, classifying them into one category (e.g. light), potentially grouping more intense activities into a lower level. Whereas those with narrow range of counts·min⁻¹ may be classifying activities of lower intensity as higher than they actually are. An example from the present study has already been noted above with the PU upper and lower bound cut-points for light activity, which could have resulted in the non-significant relationship apparent between light and sports competence.

The potential for this misclassification is thought to occur mostly between the light and moderate boundaries (Kim et al. 2012), which may explain why significant relationships became non-significant with the application of some cut-points and why other relationships varied in their magnitude (regardless of significance); the inclusion of more intense activities through high upper limit boundaries has the potential to dilute the relationships.

Trost et al. (2011) examined the classification accuracy of five of the six cut-points examined in the present study, the authors found that for light activity, the PU cut-points had an extremely low sensitivity and area under the curve (AUC) values (11.8 & 0.43, respectively) demonstrating a poor classification accuracy; evidence for this was also found by Alhassan et al. (2012). The sensitivity, specificity and AUC for the MT light cut-points were somewhat higher (58.8, 68.4 & 0.64 respectively). Furthermore, the authors noted that both PU and MT also had poor classification accuracy for moderate activity. The aspects of the light-moderate boundaries described in the Trost et al. (2011) study may serve as an explanation as to why significant relationships between light activity and perceptions of sports competence were not observed when these cut-points (PU & MT) were employed in the present study. Similar misclassification potential may explain the variation in magnitude of non-significant relationships between PA and PWB variables across cut-points. Interestingly, Trost et al.

(2011) also reported good classification accuracy for the EV and F/T cut-points, which exhibited different relationships in the present study, with the F/T defined light activity and perceptions of sport competence being non – significant, yet the same relationship defined by EV demonstrated the strongest relationship. For light activity Trost et al. (2011) reported slightly better classification accuracy for EV than F/T (AUC = .70 & .69 respectively) indicating that the accuracy of the EV cut-points allows for the appropriate detection of the light / sports competence relationship.

Findings from the present study provide partial support for findings of Stone et al. (2009), who examined the impact of cut-points upon PA and physiological health related variables. The authors reported no difference in relationships across multiple cut-points, yet one relationship became non-significant with the application of one set of cut-points (ArteAcc; Ekelund et al. 2005), concluding that cut-points did not have an effect upon relationships. The findings of the present study echo those of Stone et al. (2009) yet can be concluded in a different manner; relationships were only apparent with the application of certain cut-points and that strength of relationships varied depending upon cut-point application demonstrating the high importance of cut-points when establishing relationships with health variables and disagreeing with the conclusion of Stone et al. (2009). The findings of the present study, do however support those of Atkin et al (2013) concerning sedentary behaviour and metabolic risk, discussed earlier, yet the present study highlights the importance of classification accuracy to a greater extent than Atkin and colleagues. Yet in the present study, relationships were attenuated with the application of high cut-points rather than strengthened. These finding may indicate that relationships in the present study were diluted with misclassification of activity, whereas the misclassification in the Atkin study may have been so great that rather than diluting effects of one intensity, the results actually represented a completely separate relationship.

4.4.2.2 Direction of relationships

An interesting finding from the present study was the direction of the relationship between time spent sedentary and depression. Time in sedentary was negatively related to levels of depression; the less time children spend sedentary, the more likely they are to be depressed. These findings are contrary

to the finding of previous studies that have used both subjective and objective measures of PA. Parfitt et al., (2009) found that very light activity (classed as < 1.9 METS) had a positive relationship with depression ($r = .345$). In the present study, sedentary activity was classed as < 1.5 METS, therefore encompassing the majority of Parfitt et al. (2009) classification of very light activity. It follows then, that the magnitude and direction of the relationships should be similar, especially considering the higher sedentary boundaries used with two cut-points, meaning that light activity would likely be included when these cut-points were used. Though other studies using self-reported sedentary time have also demonstrated positive relationships with depression (e.g. Sund, et al., 2011), discord within the literatures is apparent. Page et al. (2010) reported that higher levels of television viewing and computer use (popular sedentary activities) were related to greater psychological difficulties in children, yet the authors also noted that objectively measured sedentary time was inversely related to psychological difficulties. Furthermore Johnson et al (2008) reported an inverse relationship between objectively measured sedentary time and depressive symptoms; the authors attributed this to a statistical anomaly.

It is important to note that the context of sedentary behaviour is unknown, it may be that some sedentary activities, (e.g., reading) have a positive impact on PWB and others, (e.g., television viewing) a negative impact, a concept which has yet to be examined (Page et al., 2010). Additionally, certain sedentary activities are often attractive to children (Parfitt et al., 2009) and may increase levels of PWB, offering a potential explanation for the results found in the present study.

The relationship between time spent in light activity and perceptions of sport competence was also found to be significant, this time, in the expected negative direction; the more time children spent in light activity, the lower their perceptions of their sport competence. The results of the present study expand upon those of previous literature using objective measurement (Parfitt et al., 2009), which found a weak, non-significant relationship between these variables. These relationships may occur as perceptions of sports competence will be influenced through participation in sports activities which predominantly fall into a higher intensity categories, so if children are spend large amount of

time in light activity, it may be indicative of lack of involvement with sporting activities, therefore developing low levels of sport competence; this notion is speculative however and findings from the present study do not provide the necessary evidence to draw this conclusion, especially considering that no relationships are apparent between the psychological construct and moderate, vigorous or MVPA intensities. Findings from previous studies however demonstrate a weak to moderate, yet significant relationship between objectively measured vigorous physical activity and perceptions of sports competence (Parfitt et al., 2009) and between self-reported MVPA and perceptions of sports competence (Raudsepp et al., 2002).

Following this, it should be noted that the present study revealed no significant relationships for moderate, vigorous or MVPA and psychological well-being variables; though relationships for both perceptions of strength and physical conditioning with moderate intensity activity defined by a number of cut-points approached significance. These findings are in accordance with Parfitt et al. (2009) who also failed to demonstrate relationships between moderate activity and PWB variables, though the authors did find a relationship between vigorous activity and other PWB variables, namely perceptions of physical conditioning, sports competence and anxiety. However the authors did not examine the relationships with MVPA. For the present study, those relationships approaching significance may have become significant with the inclusion of more participants, yet this may have further demonstrated the influence of cut-point application on finding significant relationships with health variables.

4.4.2.3 Partial correlations

It was expected that relationships would alter once BMI Z-score was controlled for, due to its relationship with both psychological health (Duncan, Al-Nakeeb, Nevill & Jones, 2006) and physical activity (Rowlands, et al., 2000). However, contrary to previous findings of (Parfitt et al., 2009), controlling for BMI Z-score did not significantly alter the relationships observed, with some relationships remaining the same as the zero-order correlations. Only one relationship became non-significant when controlling for BMI z-score; Truth defined light activity and sports competence, the relationship was only

attenuated by .001, resulting in a p values of .053. Based on previous findings of Fairclough et al. (2012), it would be expected that partialling out the effects of BMI would substantially reduce the relationship between PA and specific self-perceptions, as normal weight children reported more positive associations with self-perceptions than overweight and obese children. The results of the present study demonstrate that BMI Z-score has little influence upon the relationships between PA intensities and PWB and will therefore not be considered with future studies of this thesis.

4.4.3. Limitations and Future direction.

The present study highlights the impact that misclassification can have upon observed relationships and provides evidence that the application of some accelerometer cut-points may result in relationships becoming diluted and non-significant. However, a major limitation to be addressed is that of low compliance with accelerometer wear time, only 75% of participants provided adequate data for analysis. Although low compliance is expected, levels in the present study are similar to those previously reported (e.g. Van Coevering et al, 2005; Sirrad & Slater, 2009), yet a shorter wear time criteria was applied. Low compliance may have resulted in smaller magnitude of relationships and non-significant relationships with higher PA intensities; according to Cohen (1992), to find a medium effect size, ($r = .30$) at an α of .05 and a power of .80, 85 participants are needed; although initial recruitment only fell a few participants short of this, the low compliance meant that the study was highly underpowered.

Additionally, a substantial number of published cut-points remain unexamined in terms of the relationships that may occur between PA and health variables, though it is doubtful that the intensity boundaries of these unexamined cut-points vary dramatically from those examined in the present study. It should be sufficient therefore to use the present study as an example of how classification and potential misclassification of activity intensities through cut-points may impact upon relationships. Future studies should aim to employ cut-points with high classification accuracy such as the Evenson cut-points recommended by Trost et al. (2011) in an attempt to prevent potential misclassification and dilution of PA and health relationships.

4.5 Conclusion

The present study provides evidence that the application of some cut-points can impact upon the relationships observed between PA intensities and PWB, and has highlighted the importance of considering the classification accuracy when choosing which set of cut-points to employ. For the next study, those cut-points reported to have the highest classification accuracy (Trost et al., 2011), the Evenson cut-points, will be employed so as to prevent potential misclassification. Additionally, these cut-points also demonstrated the strongest relationships between PA and PWB.

Traditionally, intervention studies would follow cross sectional research to advance knowledge into the area, however the low compliance rates observed in the present study indicate that before interventions can successfully take place, the feasibility of measuring PA across multiple time points and the implication of selection bias and misclassification of PA that accompany low compliance need to be addressed, this will be the aim of the next chapter.

Chapter Five

Children's compliance with wear time criteria across multiple time points

The aim of this chapter is to explore children's compliance with various levels of accelerometer wear time as well as two popular accelerometer wear-time criteria across multiple time points. The potential selection bias and misclassification of time spent in PA of different intensities that may be introduced as a result of variations in wear-time will also be examined. Further to this, assessment of variations in the associations between PA and PWB dependent upon wear time criteria will be assessed. The chapter will begin by reiterating the problems of compliance and potential for bias and misclassification raised earlier in the thesis, followed by a description of the methods used and a discussion of the observed results.

5.1. Introduction

Previous research has reported relationships between objectively measured PA and PWB constructs in children (Parfitt et al., 2009; Page et al., 2010; Chapter four of this thesis). The cross-sectional nature of these studies, using accelerometers was warranted prior to undertaking longitudinal or experimental research, as previous studies have employed self-reported methods, which are known to be subject to misclassification of exposure through systematic and random error (see chapter three), resulting in an over- or under- estimation of the true relationship with health variables. Yet cross-sectional studies cannot provide information about the causal relationship between PA and PWB (Sund et al., 2011), to do so, an intervention should be undertaken. However, low compliance with accelerometer wear-time, observed during cross-sectional research, may hinder this logical progression as the percentage of non-compliance may increase over time and with multiple measurements.

5.1.1. Wear time compliance

In chapter three of the thesis, the issue of non-compliance with required wear-time was raised; studies showed that compliance with various wear time

criteria ranges from 50% - 86% (Van Covering et al., 2005; Audrey et al., 2012; Colley et al., 2010). It is important to note that wear time requirements vary between studies; for example in the studies reported above, Audrey et al. (2012) required ≥ 600 minutes for ≥ 3 days, whereas four or more days of 10 hours were required for Colley et al. (2010). Some studies require as little as 6 hours over three days (e.g. King et al., 2011). Decisions about wear time criteria can introduce selection bias to a study through exclusion of children not complying with wear-time, who may be less active than those who do comply and may differ on other important covariates as well as the outcome of interest (Kang et al., 2009).

The compliance rates discussed above were observed during cross-sectional studies, requiring participants to wear PA monitors for 7 consecutive days at one time point. It is hypothesised that when participants are required to undertake measurement over multiple time-points, wear-time compliance will decrease further. Evidence of this can be seen in previous intervention studies attempting to change PA to impact upon health variables such as obesity. These studies often report large variations in the amount of accelerometer data available for analysis at different measurement time points. For example, Magnusson, Sigurgeirsson, Sveninsson and Johannsson (2011) reported that out of 151 intervention participants, and 170 control participants, only 76 (50%) and 49 (29%) participants from each group respectively, had accelerometer data at all three time points. Fitzgibbon et al. (2011) reported that valid accelerometer data was only available for 323 out of 504 (64%) study participants at baseline. Some data files were unavailable due to monitors malfunctioning or being lost ($n = 59$), but the majority of participants with unavailable data had failed to achieve valid wear-time ($n = 122$); representing a compliance rate of 73%. The same study reported a drop in compliance to 65.9% post intervention (14 weeks from baseline).

Researchers can choose from multiple methods to attempt to increase the number of participants achieving the required wear time, yet each of these methods presents additional problems. As these have been discussed in chapter three, only a brief overview will be provided here. Firstly, the use of reminders via phone message or journal logs have been suggested, but other, perhaps more appealing techniques, such as monetary incentives seem to have

the largest impact upon wear time compliance (Sirard & Slater, 2009), though the use of financial reward may not always be feasible.

Other techniques include less stringent wear time requirements. For example, a reduction in the number of days, or hours per day required for the accelerometer data to be classed as 'valid', allows a greater number of participants to be included, increasing statistical power, and potentially decreasing the chance of selection bias. However, by deciding to employ less stringent wear time criteria, there is a greater possibility of introducing misclassification of PA levels. Specifically, the time in activity intensities is likely to be under-estimated by failing to record large portions of the day, or alternatively, capturing different portions of the day for different children. These factors could in turn impact upon observed relationships between PA and PWB. For a full discussion on selection bias and misclassification caused by compliance with wear – time criteria, see chapter three (p. 98 - 106).

Imputation methods to replace missing values resulting from non-wear may also be used to maximise the number of participants included in analysis. This allows researchers to undertake an intention to treat analysis (e.g. Magnusson et al., 2011), as opposed to a per-protocol analysis (e.g. Fitzgibbon et al., 2011), where only participants with valid data across all required time points are included (Bubbar & Kreder, 2006). Intention to treat analysis can be used in an attempt to eliminate the selection bias caused by only including study completers who comply with wear time criteria at all time points. A completer's only analysis tends to over report the effects of an intervention whereas imputation of baseline values is likely to underestimate the true effects of an intervention (Montori & Guyatt, 2001). Montori & Guyatt (2001) noted that in order to reduce the possibility of underestimation resulting from imputation of missing values, efforts should be made to maximise adherence through the study protocol. This may be feasible with regard to the actual intervention and through the incentivising methods described by Sirard and Slater (2009). In addition, specific methods of imputation for accelerometer data, beyond simply carrying forward baseline values, may maximise available data and lead to better estimates of the effects of interventions. However, this is an understudied area (Lee, 2013).

Recently, Audrey et al. (2012) specifically addressed wear-time compliance across multiple time-points; results showed a drop in adolescents' compliance with wear-time criteria (≥ 10 hours for ≥ 3 days) from baseline (75%) to follow up at 10 weeks (56%), with reasons for non-wear-time including worries about losing or breaking the device, concerns about appearance, discomfort and simply forgetting. Though this study highlights some of the reasons for a decrease in compliance over multiple time points, other aspects remain un-assessed. Firstly, only pre and post-trial measures were undertaken, yet many trials also include a mid-point measure, which may be used in imputation methods such as 'last observation carried forward', where missing values are replaced with the last available measurement (Molnar, Hutton & Fergusson, 2008). Secondly, a monetary incentive for returning the monitors was provided. Although rewards were not given for wear-time compliance, the monetary incentive for monitor return may have improved compliance. Additionally, the study focused upon adolescents', whose reason for non-wear may differ from those of younger children (e.g. social acceptance & appearance; Crocker et al., 2001; Perry et al., 2010; Audrey et al., 2012).

Studies reporting compliance do so only in relation to their own chosen wear-time criteria. Assessing compliance rates for at least two commonly used criteria at multiple time points may allow for a better understanding of compliance, the potential bias and misclassification that may be introduced across a range of wear times. The aims of the following study were to a) report rates of compliance with a wide range of wear time criteria; b) assess children's compliance with two widely used sets of minimum wear-time criteria (≥ 8 hours & ≥ 10 hours for 3 week and 1 weekend day) across three time-points; c) examine differences in the classification of time in activity intensities that can occur with the application of different wear time criteria; d) examine the selection bias that may ensue from excluding participants based on different wear-time criteria and finally e) assess the impact that wear time criteria has upon the longitudinal association between PA and PWB.

5.2. Method

5.2.1. Protocol

Data collection for the present study was part of an eight week non-randomised, clustered controlled trial designed to increase PA at different intensities in order to improve PWB in children (full details in chapter six). Three schools in Devon agreed to take part in the trial. Baseline PA, PWB and anthropometric measures were obtained during week one. Schools were subsequently assigned to one of two interventions or a control group. Measures of PA and PWB were collected mid-way through the intervention (week four) and during the week immediately following cessation of the trial (week nine).

5.2.2. Participants

Participants were 82 apparently healthy year five pupils ($n = 48$ females, $n = 34$ males) from three primary schools within the Devon, UK area. Children were aged between 9 and 11 years, ($\bar{x} = 9.7 \pm 0.7$ y) had a mean height of 140.6 cm (± 7.2), mass of 36.5 kg (± 6.8). Participants BMI percentiles ranged from 9 to 98% ($\bar{x} = 63.9 \pm 23.5$); 4.8% of participants were obese ($>95^{\text{th}}$ percentile) and 13.4% were overweight ($>85^{\text{th}}$ and $<95^{\text{th}}$ percentile) The Ethics Committee at the University of Exeter granted approval for the study (7th December 2010). Prior to data collection, parental informed written consent and the children's assent were obtained (See appendix 2.1 and 2.2, p. 297 - 306).

5.2.3. Physical activity

Habitual PA was measured with 43 ActiGraph GT1M and 39 GT3X monitors, descriptions of which have been provided earlier in the thesis. Both the GT1M and the GT3X have been found to be valid measures of children's PA. Furthermore, the vertical axis of the newer ActiGraph GT3X has been found to be comparable to that of the GT1M (Saski et al., 2011; Hanggi et al., 2012). Monitors were positioned at the right hip, on the anterior axillary line. Children were instructed to wear the monitors for a period of seven days, removing them only for sleeping and water based activities. Participants were asked to return the monitor to the correct position upon cessation of activities requiring removal. The appropriate orientation for the monitors was indicated by stickers so as to

aid the repositioning of the devices. Monitors were set to record data on the vertical axis at epochs of 2 seconds.

5.2.4. Psychological well-being

PWB was assessed through the same measures as described in chapter four; the Child Depression Inventory (CDI; Kovacs & Beck, 1977), the State-Trait Anxiety Inventory for Children (STAIC; Spielberger, 1973) and Child and Youth Physical self-Perception Profile (PSPP-CY; Whitehead, 1995) including a physical self-worth subscale (Whitehead & Corbin, 1991) and Harter's (1985) global self-worth subscale. All measures were administered in the same manner as previously described in chapter four.

5.2.4. Data reduction and analysis

Pupils from each intervention group were combined to assess compliance with the different wear-time criteria outlined in the aims of this chapter. Participants were then assigned to a 'compliers' or 'non-compliers' group with respect to each wear time. For all criteria, non-wear-time was defined as 30 minutes of consecutive '0' values (Rowlands et al., 2007), with the Evenson et al. (2008) intensity cut-points applied during the data reduction process, based on the findings from Trost et al. (2011) and those from chapter four of this thesis. Compliance with each criterion was assessed at all three time points. Time recorded in each activity intensity at baseline, mid and end time points, for all wear-time criteria are described.

Independent T-tests were used to assess differences in PWB and anthropometric measures between compliers and non-compliers for both minimum wear-time criteria at each time point. Pearsons' correlation coefficients were used to assess the association between baseline PWB and follow up PWB (from end point measurement) and the association between time in PA intensities at baseline and follow up levels of PWB. Finally a series of regression models were fitted to examine the relationships between baseline PA and follow up PWB, one for each wear time criteria. In each regression, baseline PWB was entered into the model first, so as to assess any variance in follow up PWB independently accounted for by baseline PA.

5.3. Results

Of the 82 participants recruited, accelerometer data were available for 76 children at baseline, 73 children at midpoint and 73 children at endpoint. At baseline, two data files could not be extracted, one monitor was not returned, and three children did not provide data due to absences during the baseline measurement week. For the midpoint measures, two children dropped out of the study, three monitors were not returned, and four data files could not be downloaded. At end point, in addition to the two drop outs at midpoint, one monitor failed to record more than two days, two data files could not be extracted and four monitors were not returned. Compliance rates and subsequent analysis are undertaken only for participants with available accelerometer data (n = 76 baseline & n = 73 mid-point & endpoint).

5.3.1. Compliance

Average wear time for the week, across all participants was 67.3 hours \pm 20.9 at baseline, 57.6 hours \pm 23.5 at mid-point and 48.0 \pm 29.8 at end-point. A breakdown of the proportion of children achieving different wear times across the measurement week is shown in table 5.1. The table shows a decrease in the proportion of children wearing the monitor as hours per day and days per week increase with a trend of fewer children complying at any level for each of the two follow up time points. At baseline children complied more with fewer hours and all days (45%) compared to longer hours and just one day (24%) suggesting a greater willingness to wear the accelerometers for a little time each day rather than for longer periods for fewer days. This is somewhat true at the mid-point but had reversed at follow up (15% versus 21%). At any time point few children provided many days of data at >14 hours per day.

Table 5.1. Percentage of children achieving different wear times for varying combinations of days and hours per day

	1 day	2 days	3 days	4 days	5 days	6 days	7 days
Baseline							
>6 hours	100	96.1	93.4	89.5	78.9	72.4	44.7
>8 hours	98.7	92.1	88.2	84.2	75.0	64.5	31.6
>10 hours	89.5	85.5	78.9	69.7	48.7	35.5	7.9
>12 hours	78.9	68.4	52.6	34.2	14.5	2.6	1.3
>14 hours	23.7	10.5	6.6	3.9	1.3	0.0	0.00
Mid-point							
>6 hours	97.3	91.8	87.7	76.7	64.4	45.2	24.7
>8 hours	95.9	87.7	82.2	72.6	56.2	39.7	19.2
>10 hours	87.7	84.9	69.9	54.8	39.7	23.3	4.1
>12 hours	72.6	54.8	36.9	19.2	10.9	5.5	0.00
>14 hours	16.4	8.2	5.5	2.7	2.7	1.4	0.00
End-point							
>6 hours	87.7	80.8	71.2	63.0	49.3	34.3	15.1
>8 hours	82.2	76.7	65.7	53.4	42.5	27.4	9.6
>10 hours	71.3	64.4	50.7	39.7	32.9	20.5	5.5
>12 hours	58.9	42.5	34.3	24.7	8.2	5.5	0.00
>14 hours	20.6	13.7	9.6	4.1	2.7	2.7	0.00

Table 5.1. provides an overview of the percentage compliance over a combination of days and hours, however from this point forward, comparisons will be made based on children complying with two popular wear time criteria; ≥ 10 hours for three week and one weekend day, and \geq eight hours for three week and one weekend day.

Figure 5.1 depicts the percentage of participants with available data who complied with the different wear-time criteria at each measurement point. The specific number of children and percentages are shown in table 5.2. The percentage of children complying with wear time requirements decreased across time points, regardless of the wear time criteria employed. As expected, criteria of \geq eight hours for \geq four days showed the highest rate of compliance across all time points compared to the more stringent criteria. Baseline compliance was highest; with a consistent decrease across subsequent time points for both wear time criteria.

When comparing the requirement of a weekend day into the 4 day wear, a lower percentage of children comply in both the 8 and 10 hour criteria than if any 4 days are included, this trend is apparent across all time points.

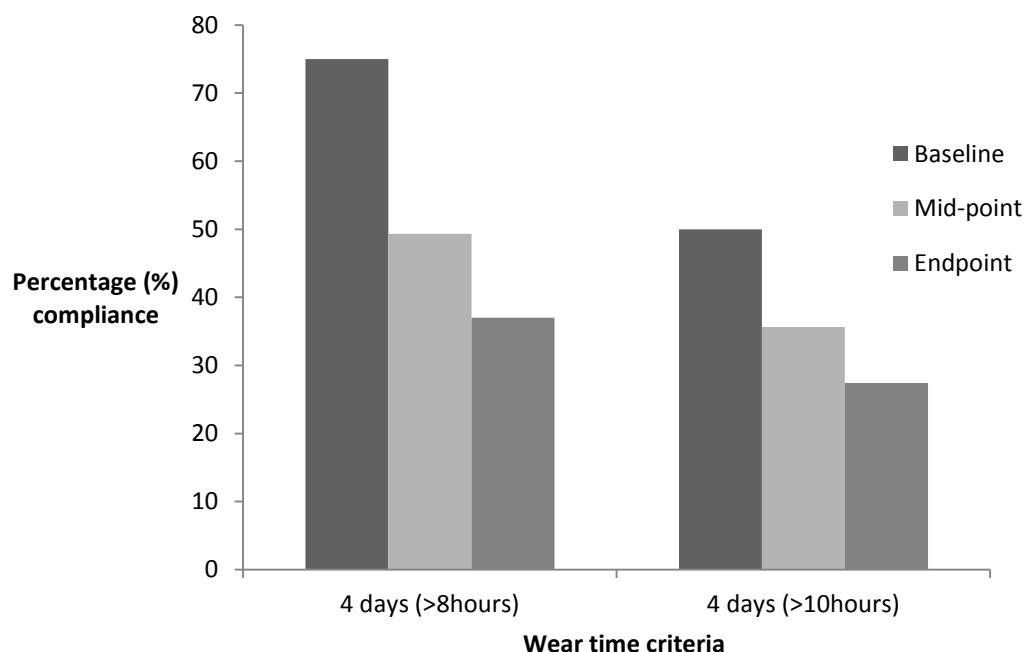


Figure 5.1. Compliance with different wear-time criteria across measurement points

Table 5.2. Number of participants complying with wear time criteria from total N with percentages shown in brackets, across multiple time points.

	Baseline n/N (%)	Mid-point n/N (%)	End-point n/N (%)
4 days (>8hours)	57/76 (75)	36/73 (49.3)	27/73 (36.9)
4 days (>10hours)	38/76 (50)	26/73 (35.6)	20/73 (27.4)

5.3.2. Selection bias arising from compliance and non-compliance

Tables 5.3 to 5.5. show the means and standard deviations of anthropometric and PWB variables of children complying and not complying with the two minimum wear time criteria. At baseline, independent t-tests revealed that participants who met the 10 hours wear criteria at baseline, had lower

depression scores compared to non-compliers ($t_{(70)} = -2.065$, $p = 0.43$); while compliers also showed a trend towards lower levels of anxiety ($p = .057$). For the eight hours wear-time criteria for three weekdays and one weekend day, compliers scored significantly lower for depression than those who did not comply ($t_{(70)} = -2.049$, $p = .044$). No other variables differed significantly at baseline, though the difference in PSW between compliers and non-compliers approached significance ($p = .052$) with compliers reporting higher PSW.

Table 5.3. Means (\pm SD) of anthropometric and PWB measures for compliers and non-compliers with two minimum wear time criteria at baseline

Baseline	≥ 10 hours of wear		≥ 8 hours of wear	
	Compliers	Non-compliers	Compliers	Non-compliers
Age (yrs)	9.6 \pm 0.6	9.8 \pm 0.7	9.8 \pm 0.7	9.6 \pm 0.7
Height (cm)	140.3 \pm 8.0	140.8 \pm 6.7	140.7 \pm 7.3	140.1 \pm 7.5
Body mass (kg)	36.8 \pm 7.4	36.3 \pm 6.5	36.9 \pm 7.0	35.5 \pm 6.6
BMI percentile	66.2 \pm 21.5	61.7 \pm 25.6	64.5 \pm 23.0	62.0 \pm 26.0
Depression (CDI)	6.4 \pm 5.1*	9.1 \pm 5.9	7.0 \pm 5.8*	10.0 \pm 4.6
Anxiety (STAIC)	29.7 \pm 5.9	32.9 \pm 8.1	30.5 \pm 6.3	33.7 \pm 9.2
Self-perceptions				
Sports competence	3.0 \pm 0.6	3.1 \pm 0.6	3.0 \pm 0.6	3.1 \pm 0.6
Body attractiveness	2.9 \pm 0.6	2.7 \pm 0.7	2.8 \pm 0.6	2.6 \pm 0.6
Physical strength	2.8 \pm 0.5	2.6 \pm 0.6	2.7 \pm 0.5	2.6 \pm 0.7
Physical conditioning	3.1 \pm 0.4	3.1 \pm 0.6	3.1 \pm 0.5	2.9 \pm 0.6
Physical self-worth	3.1 \pm 0.5	3.0 \pm 0.5	3.1 \pm 0.5	2.8 \pm 0.5
Global self-worth	3.4 \pm 0.5	3.1 \pm 0.6	3.3 \pm 0.6	3.1 \pm 0.6

* $p < 0.05$ for difference between compliers and non-compliers

Table 5.4. shows the differences between compliers and non-compliers for both minimum wear time criteria taken at the mid-point of the intervention for the 73 children who provided data . For the 10 hour wear criteria, compliers had a lower body mass than non-compliers ($t_{(70)} = -2.02$, $p = .047$), while differences in perceptions of sports competence ($t_{(59)} = -.2334$, $p = .023$), perceptions of body attractiveness ($t_{(59)} = -2.601$, $p = .012$), and perceptions of physical strength ($t_{(59)} = -2.758$, $p = .008$) were also apparent, with compliers reporting lower scores in

each self-perception sub-domain. Similar results were apparent for participants who complied with the minimum 8 hours wear time criteria at mid-point. Those complying with wear time were significantly younger ($t_{(66)} = -.2191$, $p = .032$) shorter ($t_{(70)} = -2.147$, $p = .035$), had lower body mass ($t_{(62.297)} = -3.394$, $p = .001$) and had a lower mean BMI percentile ($t_{(58.720)} = -2.613$, $p = .011$). Additionally, compliers with the 8 hours minimum wear had significantly lower perceptions of their sports competence ($t_{(59)} = -2.535$, $p = .014$) and physical strength ($t_{(59)} = -1.924$, $p = .007$).

No significant differences in anthropometric or PWB measures were apparent between compliers and non-compliers for either wear time criteria when measures were obtained after the trial, as shown in table 5.5.

Table 5.4. Means (\pm SD) of anthropometric and PWB measures for compliers and non-compliers with two minimum wear time criteria at mid-point

Mid-point	≥ 10 hour wear		≥ 8 hours wear	
	Compliers	Non-compliers	Compliers	Non-compliers
Age (yrs)	9.5 \pm 0.6	9.8 \pm 0.7	9.6 \pm 0.6*	9.9 \pm 0.8
Height (cm)	139.5 \pm 6.8	142.3 \pm 7.0	139.5 \pm 7.0*	143.0 \pm 6.8
Body mass (kg)	34.9 \pm 5.6*	38.3 \pm 7.5	34.4 \pm 5.0*	39.6 \pm 7.6
BMI percentile	60.5 \pm 25.0	66.5 \pm 23.0	56.8 \pm 25.4*	71.4 \pm 19.9
Depression (CDI)	7.3 \pm 6.1	6.5 \pm 6.1	7.3 \pm 6.0	6.3 \pm 6.2
Anxiety (STAIC)	29.9 \pm 7.8	30.6 \pm 8.4	30.7 \pm 7.7	29.97 \pm 8.7
Self-perceptions				
Sports competence	2.9 \pm 0.6*	3.3 \pm 0.6	2.9 \pm 0.6*	3.3 \pm 0.5
Body attractiveness	2.6 \pm 0.5*	3.0 \pm 0.6	2.7 \pm 0.6	3.0 \pm 0.6
Physical strength	2.5 \pm 0.6*	2.2 \pm 0.6	2.6 \pm 0.5*	2.9 \pm 0.6
Physical conditioning	3.0 \pm 0.5	3.2 \pm 0.6	3.0 \pm 0.5	3.2 \pm 0.6
Physical self-worth	2.9 \pm 0.5	3.1 \pm 0.5	2.9 \pm 0.5	3.1 \pm 0.6
Global self-worth	3.4 \pm 0.5	3.4 \pm 0.6	3.4 \pm 0.6	3.4 \pm 0.6

* $p < 0.05$

Table 5.5. Means (\pm SD) of anthropometric and PWB measures for compliers and non-compliers with two minimum wear time criteria at end-point

End-point	≥ 10 hour wear		≥ 8 hour wear	
	Compliers	Non-compliers	Compliers	Non-compliers
Age (yrs)	9.7 \pm 0.6	9.7 \pm 0.7	9.7 \pm 0.7	9.7 \pm 0.7
Height (cm)	140.6 \pm 6.6	142.7 \pm 7.8	140.9 \pm 7.1	142.9 \pm 7.8
Body mass (kg)	36.6 \pm 4.8	38.2 \pm 7.7	37.1 \pm 4.8	38.2 \pm 8.2
BMI percentile	69.5 \pm 20.4	65.0 \pm 22.2	70.4 \pm 19.3	63.4 \pm 22.9
Depression (CDI)	4.4 \pm 5.2	6.6 \pm 5.9	5.5 \pm 5.6	6.4 \pm 5.9
Anxiety (STAIC)	27.5 \pm 8.5	28.4 \pm 7.4	28.2 \pm 8.1	28.1 \pm 7.5
Self-perceptions				
Sports competence	3.2 \pm 0.6	3.2 \pm 0.7	3.3 \pm 0.9	3.1 \pm 0.5
Body attractiveness	2.8 \pm 0.6	3.0 \pm 0.6	2.8 \pm 0.5	3.0 \pm 0.6
Physical Strength	2.9 \pm 0.6	2.9 \pm 0.6	2.8 \pm 0.6	2.9 \pm 0.6
Physical Conditioning	3.2 \pm 0.6	3.0 \pm 0.6	3.1 \pm 0.6	3.1 \pm 0.6
Physical self-worth	3.0 \pm 0.5	3.1 \pm 0.6	3.0 \pm 0.5	3.1 \pm 0.6
Global self-worth	3.2 \pm 0.6	3.4 \pm 0.5	3.2 \pm 0.6	3.4 \pm 0.5

5.3.3. Classification of time in activity intensities depending upon wear-time at each measurement period

Mean time recorded per day in each activity intensity was assessed in those participants complying with both 8 and 10 hour wear time criteria. The data shows that at each measurement point, the ≥ 8 hour wear criteria compliers recorded shorter time periods in each activity intensity than those recorded by the ≥ 10 hour compliers. Differences in time accumulated appear larger for sedentary and light activity (30 to 42 minutes between wear time criteria for sedentary and 2 to 9 minutes for light) compared to the other intensities (table 5.6 and 5.7).

Table 5.6. Mean \pm SD time (minutes) accumulated in each activity intensity across three time points for 10 hours minimum wear for 3 weekdays and 1 weekend day criteria

10 hour wear			
	Baseline	Mid-point	End-point
	n = 38	n = 26	n = 20
Sedentary	574.7 \pm 59.7	582.7 \pm 83.0	602.6 \pm 119.4
Light	123.2 \pm 19.9	126.7 \pm 18.8	134.4 \pm 27.3
Moderate	27.8 \pm 6.1	28.0 \pm 4.8	34.7 \pm 10.8
Vigorous	23.3 \pm 8.7	25.2 \pm 8.8	27.1 \pm 11.9
MVPA	51.2 \pm 12.8	53.3 \pm 11.7	61.7 \pm 20.1

Table 5.7. Mean \pm SD time accumulated in each activity intensity across three time points for a minimum 8 hours wear for 3 weekdays and 1 weekend day criteria

8 Hour wear			
	Baseline	Mid-point	End-point
	n = 57	n = 36	n = 28
Sedentary	542.6 \pm 57.8	549.2 \pm 86.4	559.7 \pm 118.8
Light	114.1 \pm 20.2	120.2 \pm 20.6	132.0 \pm 24.3
Moderate	26.3 \pm 5.3	26.5 \pm 5.0	34.7 \pm 9.3
Vigorous	21.8 \pm 7.1	23.6 \pm 7.9	26.5 \pm 10.3
MVPA	48.0 \pm 10.5	49.3 \pm 10.3	61.2 \pm 17.4

5.3.4. Cross sectional and longitudinal associations between PA and PWB by wear time criteria

Baseline PA was only associated with one baseline measure of PWB, for the 10 hour wear criteria; perceptions of physical conditioning were positively associated with time in moderate activity ($r = .338, p = .047$). Using the 8 hour criteria, the relationship was attenuated ($r = .266, p = .052$). No other relationships between baseline PA and PWB were significant for either wear time criteria. The two different wear time criteria slightly altered the association between PWB at baseline and follow up. Notably, the association between the two measures of sports competence was non-significant when the 10 hour wear criteria was used ($r = .325, p = .061$) but became significant with the 8 hour wear criteria ($r = .399, p = .004$). Relationships for baseline and follow up variables in all other PWB domains were significant for both wear time criteria despite small variations in magnitude.

Correlations assessing the relationship between baseline PA and follow up PWB showed significant associations between a number of domains of PWB for both 10 and 8 hour wear time criteria (table 5.8). Associations present for both criteria were seen between baseline time in moderate intensity and physical self-perceptions (perceptions of physical conditioning, physical strength and PSW) and between time in MVPA at baseline and perceptions of physical strength and PSW. The significant relationships apparent in the 10 hour criteria were attenuated when the 8 hour criteria was applied, with one relationship becoming non-significant (time in vigorous at baseline and PSW). By contrast, the relationship between time in vigorous intensity at baseline and perceptions of physical conditioning was stronger, becoming significant with the application of the 8 hour wear time.

Regression analyses were used to further examine the association between baseline PA and follow up PWB. Baseline PWB was also entered into the models to assess the independent association of baseline PA. For each regression, multi-collinearity was assessed, in each model, the variance inflation factor (VIF) was below 10 and tolerance levels were above 0.2, indicating no collinearity (Field, 2009). Of the relationships assessed, for both 10 and 8 hour criteria, baseline PA only predicted perceptions of physical self-

worth once baseline measures of PSW had been controlled for. Table 5.9. below provides detail of the regression analysis (non-significant regression models in appendix 4.3, p. 333). As can be seen from the results, time in moderate, vigorous and MVPA at baseline predicted, 14.4%, 10.9% and 16.4% of the variance in PSW respectively at follow up when using the ≥ 10 hour wear time criteria. Alternatively, time in moderate and MVPA predicted 5.8% and 6.4% of the variance in follow up PSW respectively after controlling for baseline measures when applying the ≥ 8 hour criteria.

Table 5.8. Correlations between baseline PA and follow up measures of PWB for different wear criteria

	Depression	Anxiety	Sports competence	Body attractiveness	Physical strength	Physical conditioning	PSW	GSW
>10 hours for >4 days								
Time in sedentary	-.023	-.082	-.113	-.081	.053	-.091	.022	-.159
Time in Light	.235	-.017	.211	-.042	-.187	-.168	.054	.131
Time in Moderate	-.001	-.122	.178	.224	.357*	.353*	.443*	.182
Time in Vigorous	.024	-.107	.108	-.075	.281	.222	.343*	.093
Time in MVPA	.016	-.131	.157	.050	.360*	.317	.443*	.149
>8 hours for >4 days								
Time in sedentary	-.240	-.237	.088	.036	.180	.122	.066	.046
Time in Light	.208	.114	.191	-.146	-.108	-.077	-.122	-.032
Time in Moderate	-.047	-.050	.166	.232	.273*	.336*	.347*	.163
Time in Vigorous	.042	-.058	.139	-.093	.235	.292*	.217	.061
Time in MVPA	.006	-.064	.177	.051	.295*	.365*	.320*	.122

* p < 0.05

Table 5.9. Regression analysis for baseline PA and PSW for 10 hour and 8 hour wear time criteria.

	$R^2\Delta$	SEE	B	t	p
>10 hours for >4 days					
Time in Moderate ^a	.144	.409	.035	2.873	.007
Time in Vigorous ^a	.109	.423	.020	2.416	.022
Time in MVPA ^a	.164	.401	.017	3.120	.004
>8 hours for >4 days					
Time in Moderate ^a	.058	.408	.025	2.299	.026
Time in MVPA ^a	.064	.406	.013	2.412	.020

^amodel adjusted for baseline PSW

5.3.4.1. Regression equations

Regression equations were developed for the prediction of follow up PSW from time in moderate ($Y = .206 + .618$ (baseline PSW) + .035 (Baseline time in moderate)); time in vigorous ($Y = .618 + .657$ (baseline PSW) + .020 (baseline time in vigorous)); and time in MVPA ($Y = .265 + .635$ (baseline PSW) + .017 (Baseline time in MVPA) for the ≥ 10 hours for ≥ 4 days criteria.

For the ≥ 8 hour for ≥ 4 days criteria, the regression equations were as follows. For time in moderate, $Y = .387 + .685$ (baseline PSW) + .025 (baseline time in moderate), and for time in MVPA, $Y = .393 + .699$ (baseline PSW) + .013 (baseline time in MVPA).

5.4. Discussion

In this chapter children's compliance with various wear time combinations of days per week and hours per day were examined. In addition, the effect of applying two commonly used wear time standards on values of exposure and outcome measures, as well as other covariates, was examined at three different time points along with cross-sectional correlations between measures. Finally, the effect of wear time on the longitudinal association between baseline PA and PWB was examined. The main findings show that compliance decreases with stricter wear time criteria for both days and hours per day. Further, compliance reduces with repeated measures across time points, regardless of the criteria used, though stricter criteria resulted in even lower compliance levels. Selection bias was apparent with the use of both wear time criteria at baseline and mid-point measurement, while misclassification of time in PA intensities was apparent across all time points in all intensities. Importantly, longitudinal relationships between PA and psychological variables differed depending upon the wear time criteria employed.

5.4.1. Compliance

As was expected, the percentage of children complying with increasing hours per day and increasing days per week reduced substantially at each time point. Across time, compliance with at least 8 hours or 10 hours wear per day for the minimum of three weekdays and one weekend day also decreased with each subsequent time point. At baseline 75% and 50% of children complied with the 8 and 10 hours wear times respectively and by the third time point, compliance had decreased to 36.9% and 27.4%. In summary, children's wear-time decreased over multiple time points, with stricter wear-time criteria resulting in lower levels of compliance. Not only did compliance decrease but the average weekly wear time decreased substantially, with a drop in accrued wear time of approximately 10 hours for each subsequent measurement point.

Though a decrease in compliance was expected with subsequent time points, the extent to which non-compliance occurred at all-time points was greater than anticipated. Van Covering and colleagues (2005) reported that full wear-time (7

days; 13 hours for weekdays and 9 hours at weekends) was acquired from 50% of participants at a singular time point, yet in the present study, neither the 8 nor 10 hour criteria resulted in compliance as high as this, even at baseline. Furthermore, studies reporting reductions in compliance across time points, have not reported such dramatic decreases as those apparent in the present study. Fitzgibbon et al. (2011) reported post intervention compliance of 65.9%, while Audrey et al. (2012) reported slightly lower compliance of 56%, much higher rates than seen at post – intervention for the present study (≥ 10 hours = 27.4%; ≥ 8 hours = 36.9%, this may be due to the additional time point included in the present study. Compliance rates at mid-point (≥ 10 hours = 35.6%; ≥ 8 hours = 49.3%) are closer to those reported in earlier studies with follow up time points. Interestingly, Audrey and colleagues (2011) required more wear-time per day for fewer days (10 hours for ≥ 3 days) compared to the Fitzgibbon study (8 hours ≥ 4 days), yet reported lower levels of compliance. It may be reasonable to suggest then that more lenient wear criteria of hours rather than days would result in higher proportions of participants included in a study. Results in table 5.1 support this idea. Further, a higher proportion of children achieved Fitzgibbon's criteria than Audrey's criteria at each time point.

The results of the present study are in agreement with those reported by Colley et al. (2010), who showed that participant numbers increased with more lenient wear-time criteria; the authors reported that 79.5% of participants recorded >10 hours wear time and 83% of participants recorded >8 hours wear time, however it was unclear as to the number of days of data, that these values were in respect to. Colley et al. (2010) also reported the number of valid days achieved by children over one week's measurement period; showing that 86% of males and 88% of females aged 6 – 11 years achieved four days of monitoring for a minimum of 10 hours per day. The present study reports lower compliance rates with this at all time points, but extends the findings of Colley and colleagues by reporting the proportions of children complying with multiple hour and day criteria. This allows for comparison of compliance with other studies who employ different wear time criteria, such as those reported by Page et al. (2010), who employed a wear time criteria of ≥ 3 day for ≥ 10 hours and saw compliance rates of around 77%; similar to the portion of children meeting that criteria in the present study at baseline. Similarly King et al. (2011) used a

criterion of ≥ 3 day for ≥ 6 hours, and reported compliance of 94.7%, again similar to those in this study. The results reported in this chapter, along with those reported by others suggest that children find it easier to comply with shorter hours for more days rather than stricter hours for fewer days.

5.4.2 Wear time compliance and selection bias

Non-compliers with both ≥ 8 hours per day and ≥ 10 hours per day (for three week and one weekend day) reported higher depression scores compared to compliers at baseline. This suggests that in general, non-compliance excludes from analysis, children with higher levels of depression. The inclusion of only those children with lower depression scores may therefore reduce the potential to show improvements in depression over time or as a result of changes in PA. It should also be noted that compliers with the 10 hour criteria reported lower depression scores than compliers with the 8 hours; though the differences were small (differences were not formally tested).

At mid-point, in the subset of participants with any accelerometer data, participants meeting the 10 hours and the 8 hours wear-time criteria had a lower mean body mass than those who did not comply, while those meeting the 8 hour wear time criteria were significantly younger and shorter than those not achieving ≥ 8 hours for ≥ 4 days and had lower BMI percentiles than non-compliers, though the mean BMI percentile of non-compliers (>8 hours = 71.4) did not classify them as overweight ($> 85^{\text{th}}$ percentile) or obese ($> 95^{\text{th}}$ percentile). Once again the exclusion of participants with higher BMI percentiles / weight could introduce bias as these variables could be indicative of lower PA levels. For example, Dorsey, Herrin and Krumholz (2011), reported that normal weight children undertook significantly more MVPA than overweight and obese children ($\bar{X} = 142.9$ and 119.7 minutes per day respectively). This was not assessed in the present study, as only those with valid PA data could be used, and analysis shows that the compliers already report lower BMI percentiles, hence, further analysis would be undertaken using a bias, unrepresentative sample.

Differences were also observed in physical self-perception sub-domains at mid-point for the 10 hour and 8 hour wear criteria. Interestingly, non-compliers with the 10 hour criteria reported higher levels of perceptions of sports

competence, body attractiveness and physical strength. Non-compliers with the 8 hours wear time also reported higher levels of perceived sports competence and perceptions of physical strength, but not body attractiveness. It may be that children with higher self-perceptions in the physical domain undertake more sporting activities which required monitor removal due to safety reasons, requirement from authority figures or undertaking water based activities; reasons that have been reported for removal in previous studies (Crocker et al. 2001; Audrey et al. 2012). The exclusion of these participants from analysis could result in an under-estimation of self-perceptions and potentially an under-estimation of PA. Finally, it should be noted that no differences in PWB were observed between compliers and non-compliers in either wear-time conditions at end point, indicating that selection bias was only apparent for the first two measurement points.

Differences between compliers and non-compliers occurred for different variables depending upon the wear time criteria employed and the time point assessed. It could be that as more children failed to comply with wear-time, the differences between groups became smaller (e.g. for depression). Or that undertaking the PA intervention (described in chapter 6) resulted in a greater awareness of body size, hence the differences at mid-point. If this was the case, it may be expected that differences would occur between children assigned to different intervention groups and this possibility will be examined in the following chapter. Furthermore, variables differing between compliers and non-compliers between wear time criteria employed and different time points show that different bias is introduced depending on which criteria is employed at each time point, which in turn will have a different impact upon the results of a study depending upon the question asked. Combining children from three different arms of a trial may explain some of the differences between compliers and non-compliers at each time point. If compliance at each time point is systematically different across trial arms and the intervention is working then this would alter the values of the outcome above and beyond those due to non-compliance. Again, this will be fully examined in the next chapter.

5.4.3 Wear time compliance and misclassification

The employment of different wear time criteria effected the recorded time at each intensity. The results show less time accrued in each intensity at all time points when using the more lenient criteria of ≥ 8 hours. Differences were more substantial for sedentary and light activity intensity. The difference between wear time criteria for time spent sedentary increased from baseline (30 minutes) to end-point (43 minutes), whereas differences in light activity decreased (from 9 minutes to 2 minutes).

The results of the present study partially support those of Catellier et al. (2005) who reported that average time in MVPA for 12 h wear-time, was 159 minutes, while for 10 hour and 8 hour criteria, time in MVPA was 147 and 148 minutes, respectively. The present study shows that time in intensities differs depending upon criteria, with differences in MVPA between 10 hour and eight hour wear time criteria across time points only slightly larger than those noted by Catellier et al. (2005). It is unclear from the results of the present study, whether time accumulated differs between ≥ 10 hours and \geq eight hours due to the selection bias introduced, as discussed previously, or due to the capture of shorter periods of the day which may miss periods when PA is being undertaken. To address this problem fully, time in activity intensities needs to be examined in a within subject design, to eliminate the impact of selection bias upon this area of misclassification; this will be examined in chapter eight. Recently, Herrmann et al. (2012) attempted to examine this concept with an adult population, by simulating data; specifically by cropping data accumulating to different wear times (e.g. 10 hours, 12 hours etc.) from a day of 14 hour wear based on normal wear patterns, essentially creating new data sets from a reference day of 14 hours PA monitoring. The authors subsequently noted that estimated time in each activity intensity reduced with a decrease in the number of hours in the simulation data, reporting differences of around 135 minutes per day in sedentary time, and 95, 7 and 0.5 minutes per day difference in light, moderate and vigorous activities between 14 hour and 10 hours wear time.

For the present study though, it is likely that a combination of both selection bias and differences in observation periods are responsible for the differing time accrued in activity intensities.

5.4.4. Wear time criteria and relationships between PA and PWB

In the present study, longitudinal analysis revealed that baseline PA was related to PWB at follow up for a number of self-perception variables. However, regression analysis revealed that some of these associations were due to the confounding effect of baseline PWB. When baseline PWB was accounted for, baseline PA was only related to follow up perceptions of physical self-worth. These results demonstrate the importance of baseline measures in the tracking of PWB over time. The relationships observed seem to provide partial support for the published literature, but appear to disagree with previous findings in the present thesis. Correlations at baseline provided only one significant relationship for 10 hour wear time, while no-relationships were apparent with the employment of the 8 hour wear criteria, contrary to what was observed in chapter four of this thesis. The relationship at baseline between perception of physical conditioning and time spent in moderate activity somewhat agrees with wider literature; Raudespp et al. (2002) demonstrated a significant relationship between moderate activity and perceptions of physical strength, while more recently Webb et al. (2013) demonstrated a positive relationship between MVPA and physical conditioning, though both of these relationships were observed with self-reported MVPA. Additionally, both Raudsepp et al. (2002) and Webb et al. (2013) along with others (e.g. Parfitt et al., 2009; Goldfield et al. 2011; Chapter four of this thesis) have noted relationships between PA intensities and other psychological variables measured within this study. The results of the cross-sectional baseline relationships using the 8 hours wear time, agree with those of Hume et al. (2011) who found no cross-sectional relationships between PA and PWB.

The results of the longitudinal relationships indicate that PA performed at higher intensities can predict future physical self-perceptions, with more time spent in activity indicating higher PSW 8 weeks later. The results support the findings of previous research. For example Raudsepp et al. (2013) noted that higher initial levels of PA lead to higher levels of PSW at subsequent time points. Though studies have not specifically addressed the longitudinal relationships between objectively measured activity intensities and PSW, it seems likely that higher intensities would have a more positive impact upon PSW than lower intensities (light and sedentary). Indeed, sedentary behaviour

(self-reported) has been reported to have a negative relationship with PSW (Webb et al. 2013). It has also been noted that the relationship between PA and PSW are mediated by the sub-domain perceptions of the physical self (Dishman et al., 2006); a relationship which is indicated by the hierarchical nature of self-esteem (see figure 1.1, p.21) Though a mediation analysis was not undertaken in the present study, the results of correlation and regression analysis indicate that this may not be the case, as baseline PA was not related to the subdomains of the physical self once baseline levels of perception had been accounted for.

More important to the present study however is the apparent difference in the level of variance in PWB accounted for in the two different models based on the two wear time criteria. Notably, table 5.8 shows the attenuation of baseline PA and follow up PWB relationships from the 10 hour wear to the 8 hour wear criteria, although other relationships appear to become stronger. The fact that relationships are not all reduced in the same manner suggests that selection bias may be the cause of this, rather than purely a misclassification of time in activity intensities. Firstly, the use of the 8 hour wear criteria underestimated all activities (to varying degrees) compared to the 10 hour wear, therefore it would be expected that all relationships would be underestimated with the 8 hours wear time, however this is not the case. Also, as mentioned earlier, the fact that the children were drawn from three arms of a trial may add an additional level of bias if loss to follow up was systematically different across trial arms. Further, the majority of significant relationships occurred between higher activity intensities, which show only small variations in time accrued between wear-time criteria employed. The same may be true for the results of the regression models, a much larger portion of the variance was accounted for in PSW by moderate and MVPA when employing the 10 hour wear compared to the 8 hour wear. The differences reflect the impact that wear time criteria may have upon the magnitude of the relationship observed between PA and PWB. Specifically, the results demonstrate that if compliance with strict wear time is low, efforts to increase participant numbers by reducing wear time may result in a potential underestimation of the true relationship. Again this may be a result of selection bias rather than misclassification for the reason noted above, though it may be argued that selection bias is decreased with more lenient wear time due to the

inclusion of more participants. If this is true, a case could be made for the 10 hour wear time over-estimating the relationship due to the participants included, and it is less likely to be due to misclassification as larger portions of the day are captured. In the present study however, more selection bias was introduced with the more lenient wear time.

5.4.5. Limitations

The present study begins to address the effect of selection bias and misclassification, introduced through necessary data reduction processes, upon the distribution of values for exposure and outcome measures as well as the relationship between PA and PWB. However, there are a number of limitations that are worth considering. Primarily, the examination of misclassification by the choice of wear time criteria was undertaken with two separate samples, meaning that the cause of any misclassification could not be attributed to either selection bias or wear time differences. To test the independent effect of wear time on misclassification requires a single population with 'extended' wear times from which shorter wear times can be modelled.

Secondly, compliance was lower than expected for all time points and may in part be due to the lack of incentives suggested by Sirard and Slater (2009). This restricted the potential to compare high and low compliers and also restricted the sample for which longitudinal analysis could be undertaken.

Finally, data for the present study was obtained during a PA intervention, in which subjects were assigned to different activity groups. For the present analysis all groups were collapsed into one, however it may be that non-compliance was higher in some trial groups, particularly the control group, more than others and this could have affected the values and associations observed between compliers and non-compliers defined here. The examination of compliance across intervention groups will be undertaken in chapter six.

5.5. Conclusion

The present study highlights the issue of non-compliance with various wear time criteria at specific time points and across time. Irrespective of wear time criteria employed, compliance decreases with repeated measures. Imposing strict criteria, to improve precision, introduces selection bias which in itself can

alter observed associations. Additionally, the results reveal that attempts to overcome non-compliance and selection bias by reducing the wear time criteria, can result in an underestimation of time in activity intensities and attenuate relationships between PA and PWB variables. Furthermore, some selection bias was still evident even when employing more lenient wear time criteria suggesting that alternative methods to reduce participant numbers excluded from analysis should be used.

The present study could not separate the misclassification of PA intensity due to variations in wear time from that due to selection bias. Also, the limitations apparent within this study need to be addressed. Following chapters will address these issues. In the next chapter, selection bias and misclassification between different arms of the trial, referred to at the beginning of the chapter, will be examined along with the subsequent effect upon intervention results.

CHAPTER SIX

WEAR TIME COMPLIANCE ACROSS A PHYSICAL ACTIVITY INTERVENTION

The following chapter will attempt to extend the investigation of compliance with required wear time to assess differences in compliance between intervention conditions. Selection bias and misclassification occurring between and within conditions as a result of wear time compliance will be assessed.

6.1. Introduction

Previous chapters within this thesis have discussed the issue of non-compliance with required accelerometer wear time, for a waist worn accelerometer, showing similar compliance rates (75%) to other studies (e.g. Audrey et al., 2012) for single time point measurement (chapter four; baseline measures in chapter five), though this was only achieved with lenient wear time criteria (>8 hours for three week and one weekend day). Non-compliance increased with repeated measures at subsequent time points, reducing to 38% for the lenient criteria at the final measurement time point. The point has been raised throughout this thesis that selection bias may be greater with the employment of stricter wear time criteria; as was apparent in chapter five of this thesis. Non-compliers had significantly higher levels of depression at baseline measurement point, while anthropometric and well-being differences were apparent between compliers and non-compliers at mid-point measurement. Furthermore, efforts to reduce selection bias through the reduction of required wear time resulted in misclassification of time spent in activity intensities. Specifically, underestimation occurred across all activity intensities, but was apparent to a greater extent in lower intensities (sedentary and light). Similar results have previously been reported with an adult population. Herrmann et al. (2012) reported differences of around 135 minutes per day in sedentary time, and 95, 7 and 0.5 minutes per day difference in light, moderate and vigorous activities between 14 hour and 10 hours wear time. Additionally, chapter five demonstrated large differences in the results of the longitudinal relationship between PA and PWB depending upon the wear time criteria employed,

however it was unclear whether this was due to misclassification or selection bias.

A problem in all trials of behavioural interventions is loss to follow up, especially if study completers differ in exposure, outcome or confounder, compared to those lost. Low compliance with accelerometer wear time exacerbates this problem as people may complete all other aspects of a trial but get treated as lost because they did not meet the minimum wear time criteria to be included. Randomisation prior to intervention is designed to remove any bias created by between condition differences (Vincent, 2005), though recently Gaus and Muche (2013) noted that the assumptions of randomisation only holds true if no drop outs occur or there is no major violation in study protocol. Non-compliance with wear time, especially if differential by trial arm, demonstrates a violation in protocol, suggesting that randomised trials should take into account differences between compliers and non-compliers. Researchers should also be aware that the intervention itself may introduce selection bias. For example, lower compliance with required wear time may be higher in the control condition as reported by Magnusson et al. (2011) who showed that 50% of 151 intervention participants had accelerometer data for all three time points, whereas only 29% of 170 control participants complied on all measurement occasions.

6.1.2. Intervention rationale

The intervention in the present study was designed to increase PWB constructs by modifying children's PA levels. The results aim to add to the findings of cross-sectional studies and from interventions using self-reported PA measures; a review of this literature is provided in chapter two of this thesis. Studies using child populations have reported higher levels of anxiety and depression in children accumulating high levels of light intensity activity; these children also reported low levels of self-esteem, compared to children undertaking less low intensity activity (Parfitt et al., 2009). Chapter four of the present thesis provided some support for these findings; showing that more time spent in light intensity activity was associated with lower levels of PWB; specifically in the perceptions of sports competence. However opposing results were observed concerning time spent sedentary, which showed an inverse

relationship with depression, supporting the findings of others (Johnson et al., 2008; Page et al., 2010)

At the other end of the scale, vigorous intensity activity has been associated with lower levels of anxiety and depression and higher levels of self-esteem (Parfitt et al., 2009). Similarly, Page et al. (2010) noted an inverse relationship between MVPA and emotional difficulties, an area that can be related to the emotional disorders of anxiety and depression. Similar results have been noted with self-reported PA, for example Goldfield et al. (2011) reported negative associations between vigorous activity and depression and anxiety. Additionally, chapter 5 of the present study noted longitudinal relationships between higher intensity activities (moderate, vigorous and MVPA) and healthier PSW.

Intervention studies have demonstrated equivocal findings with regard to PA and PWB in children (e.g. Bronhauser et al., 2005; Schneider et al., 2008; Petty et al., 2009), though very few studies have examined the impact of time accumulated in different activity intensities on children's PWB (Biddle & Asare, 2011). A Cochrane report undertaken by Larun et al. (2009) noted moderate effects ($ES = -0.66$) of vigorous activity upon depression and non-significant trends ($ES = -0.48$) for anxiety. Overall, the findings from the children's literature suggest that a reduction of time spent in light activity and an increase in higher intensities such as MVPA would be beneficial for PWB. Results from the adult literature appear to correspond with this notion, showing that higher intensities may be more effective than lower intensities in aiding aspects of PWB (Teychenne, Ball and Salmon, 2008). Yet evidence is still inconclusive as to whether moderate (E.g. Dunn, Trivedi, Kampert, Clark & Chambliss, 2005; Osei – Tutu & Campagna, 1998) or vigorous intensity activity (E.g. Wise, Adams – Campbell, Palmer and Rosenberg, 2006) would be most beneficial.

In order to assess whether reducing light activity and increasing time in moderate or vigorous intensity activity can increase children's PWB, and whether wear time compliance affects associations, a school-based PA intervention was undertaken. Three schools were assigned to one of three conditions: moderate intensity or vigorous intensity intervention condition (matched energy expenditure) or a control condition. The assessment of the intervention's effectiveness can be undertaken for all those who provide

outcome (PWB) data. However, in order to assess whether changes in the outcome occur through the proposed processes, in this case, changing PA, participants are required to have both valid PA and PWB data.

The aims of the present study were to a) assess the effectiveness of the intervention upon PWB and PA; b) assess whether compliance with two popular wear time criteria across multiple time points would vary between intervention conditions; c) explore variations in selection bias and misclassification occurring across intervention conditions; d) examine whether differences in compliance between conditions would impact the results of the intervention.

Based on the aims of the study, and the previous literature, the following hypotheses were formulated. 1) Participants undertaking a PA intervention would have better PWB at end-point compared to the control condition. 2) Wear time compliance would decrease in all conditions at subsequent time points, for two wear time criteria, but a reduction in compliance would be larger for the control condition. 3) Selection bias would occur between compliers and non-compliers for all conditions across time and differentially between conditions. This bias would vary by wear time criteria and measurement point. 4) An underestimation of time in PA intensities, particularly sedentary and light intensity will occur with the use of the 8 hour wear time criteria for all intervention conditions. 5) Apparent differences in compliance between conditions and attempts to increase participant numbers will confound the effect of the PA intervention.

6.2. Methods

6.2.1. Participants

Participants were the same sample as reported in chapter five; 82 apparently healthy year five pupils ($n = 48$ females, $n = 34$ males) from three primary schools within the Devon, UK area. Children were aged between 9 and 11 years, ($\bar{x} = 9.7 \pm .7$ years) had a mean height of 140.6 cm (± 7.2), mass of 36.5 kg (± 6.8). Participants BMI percentiles ranged from 9 to 98% ($\bar{x} = 64.0 \pm 23.5$); 4.8% of participants were obese ($>95^{\text{th}}$ percentile) and 13.4% were overweight ($>85^{\text{th}}$ and $< 95^{\text{th}}$ percentile) The Ethics Committee at the University of Exeter granted approval for the study (7th December 2010). Prior to data collection,

parental informed written consent and the children's assent were obtained (See appendix 2.1 & 2.2, p. 297 - 306).

6.2.2. Protocol

The intervention was a non-randomised controlled trial, schools were invited to take part via email, the three schools that agreed to participate, were assigned to one of three conditions; a moderate intensity activity intervention (n = 29), vigorous intensity activity intervention (n = 25) and a control condition (n = 28). The two intervention conditions were designed to be matched for PA energy expenditure. Based on previous literature (Parfitt et al., 2009) an effect size of 1.06 was calculated, using the mean difference (SD) of 1.7 (1.6) in GSW between groups achieving different levels of light activity. To obtain power (β) of .80, with an α level of .5, it was determined that a sample size of 21 would be needed for each condition (Cohen, 1992). Information packs and consent forms were distributed to all year five children within the schools, four of which, from the school assigned to the vigorous condition chose not to take part.

Baseline PA, PWB and anthropometric measures were obtained during week one of the trial. Each condition then received a PA education session lasting 30 minutes, which consisted of an interactive presentation defining PA and outlining its main benefits, with discussion of how much PA should occur and ways in which their own PA could be increased. Classes assigned to the moderate and vigorous conditions then commenced an eight week PA program, while the control condition maintained normal curriculum. The duration was chosen in order to fit within one school term. Intervention conditions took part in three activity sessions a week. Moderate and vigorous conditions were matched for energy expenditure, in line with the recommendation of Teychenne et al. (2008), so as to isolate the effect of intensity upon children's PWB. Those assigned to the moderate condition completed three sessions of 30 minutes of activity per week, working at approximately 3 - 4 METs, whereas the vigorous intensity condition completed three 15 minute activity sessions per week at approximately 6 - 8 METS. PA, PWB and anthropometric measures were completed mid-way through the activity intervention, and upon cessation of the program.

6.2.3. Measures

6.2.3.1. Physical Activity

Habitual PA was measured with 43 ActiGraph GT1M and 39 GT3X monitors. The protocol for configuration and distribution of monitors has previously been described in chapter five.

6.2.3.2. Psychological Well-Being

PWB was assessed through the same measures as described in chapter four; the Child Depression Inventory (CDI; Kovacs & Beck, 1977), the State-Trait Anxiety Inventory for Children (STAIC; Spielberger, 1973) and Child and Youth Physical Self-Perception Profile (PSPP-CY; Whitehead, 1995) including a physical self-worth subscale (Whitehead & Corbin, 1991) and Harter's (1985) global self-worth subscale. All measures were administered in the same manner as previously described in chapter four.

6.2.4. Data reduction and analysis

To address all the aims of this study, a variety of analyses were undertaken, each of which are described below in relation to the aim they address.

6.2.4.1 Assessment of the effectiveness of intervention upon PWB

Primarily, analysis of covariance (ANCOVA) was employed to assess differences in PWB variables between conditions at the follow up measurement (week nine), whilst controlling for baseline PWB values. This was undertaken regardless of compliance with PA measurement.

6.2.4.2. Assessment of compliance with two wear time criteria

Pupils from each condition were assessed for compliance with two wear time criteria; a) three weekdays and one weekend day for a minimum of ten hours per day (≥ 10 hours ≥ 4 days), b) three weekdays and one weekend day for a minimum of eight hours per day (≥ 8 hours, ≥ 4 days). As in chapters 4 and 5, non-wear-time was defined as 30 minutes of consecutive '0' values (Rowlands et al., 2010), with the Evenson et al. (2008) cut points applied to determine time spent at each activity intensity. Compliance with each criterion was assessed at all three time points for each condition. Subsequently, Pearson's chi squared

tests (χ^2) were used to assess whether compliance with wear time requirements was related to the intervention conditions.

6.2.4.3. Exploring selection bias and misclassification

An examination of selection bias apparent within different conditions was then undertaken using independent t-tests to establish whether there were differences between compliers and non-compliers in each condition for baseline, mid-point and end-point measurement. One way analysis of variance (ANOVA) was used to assess differences in PWB and anthropometric measures between compliers in each condition, for both wear-time criteria. This was repeated for each time point.

In addition, mean time accumulated in each activity intensity for all intervention and control conditions across baseline and mid-point measurement was established for both wear times. Time in activity intensities was established for end-point measures but due to limited sample sizes with accelerometer data, means and standard deviations should be viewed with caution.

6.2.4.4. Examining mechanisms for intervention effects

Finally, as noted previously, the assessment of proposed mechanisms should be undertaken, whereby any intervention effects upon PA are assessed for those people who had valid PA data. Also, it should be established whether changes in PA can lead to changes in PWB. However for the present study this analysis could not be undertaken due to inadequate participant numbers at end-point across all conditions. This will be discussed further in the results and discussion section.

6.3. Results

6.3.1. Loss to follow up

Loss to follow up was apparent across all conditions, table 6.1. shows the numbers of participants in each condition at each time point (N), and the number of participants who had complete PWB data (n) at each time point. At baseline all but two participants in the moderate and vigorous condition completed the questionnaires, however full psychological profiles were only available for the numbers noted in the table due to incomplete data sets. For the

moderate condition, two children were absent. Additionally, some children had data for some constructs (e.g. depression), but not others (e.g. physical self-perceptions). In the vigorous condition, two children dropped out of the study by mid-point. An increase in the number of children completing questionnaires is seen at the mid-point in moderate and control conditions. However, only 56% of children in the vigorous condition had complete psychological profiles, a trend that did not continue to end-point.

Table 6.1. Number of children with complete PWB questionnaires (n) in the total sample for each condition across time points (N)

	Moderate (n/N)	Vigorous (n/N)	Control (n/N)
Baseline	26 / 29	22 / 25	24 / 28
Midpoint	27 / 29	13 / 23	26 / 28
End-point	27 / 29	20 / 23	26 / 28

6.3.2. Effect of the intervention

Multiple ANCOVAs exploring between condition differences at end-point while adjusting for baseline using completers only data indicated no intervention effects for any PWB measures ($p > 0.05$). Only one difference in physical self-worth approached significance ($p = 0.054$), with the moderate condition reporting the highest PSW ($\bar{x} = 3.3 \pm 0.6$), and the control condition the lowest ($\bar{x} = 2.9 \pm 0.5$).

6.3.3. Wear time compliance

Compliance with two accelerometer wear time criteria at each time point is shown below in tables 6.2 and 6.3. The number of data files available for analysis in each condition varied depending upon time point. Files were unavailable due to monitors not returned, failure to download, and absences during the week of data collection. This resulted in varying participant numbers in each condition. At baseline 76 accelerometer files were available (moderate $n = 26$; vigorous $n = 23$; control $n = 27$), for mid-point 73 files were available (moderate $n = 26$; vigorous $n = 21$; control $n = 26$). Seventy-three files were also available at end-point (moderate $n = 24$; vigorous $n = 21$; control $n = 28$). Numbers of files available are outlined in table 6.2. along with the number and

percentages of compliers and non-compliers with each of the two wear time criteria.

6.3.3.1. 10 hour compliance

For the 10 hour wear criterion, compliance at baseline was lowest in the moderate condition. Compliance declined across all conditions by mid-point, with the vigorous condition showing the largest reduction of around 50%. The chi-squared test showed no relationship between intervention conditions and compliance at any time point, indicating that compliance across conditions was similar at all time points when using the 10 hour wear criterion.

6.3.3.2. 8 hour compliance

For the 8 hour wear criterion, compliance with wear time was similar across conditions at baseline. However compliance in the moderate and vigorous condition declined substantially at mid-point, whereas compliance in the control condition remained high, with a reduction of only two participants. Chi squared statistic showed a significant association between condition and compliance at this time point ($X^2_{(2)} = 6.409$, $p = 0.041$). At the end-point, compliance in the vigorous and control conditions declined further, while compliance in the moderate condition showed a small increase with one extra participant complying. No other relationships were apparent between condition and compliance.

Table 6.2. Number of accelerometer files for each intervention condition and each time point, with number (n) and percentage (%) of participants complying and not complying with 10 hour wear time criteria across time points

>10 hours	Moderate			Vigorous			Control		
	N	Compliers n (%)	Non compliers n (%)	N	Compliers n (%)	Non compliers n (%)	N	Compliers n (%)	Non-compliers n(%)
Baseline	26	10 (38.5)	16 (61.5)	23	15 (65.2)	8 (34.8)	27	13 (48.1)	14 (51.9)
Mid-point	26	7 (26.9)	19 (73.1)	21	7 (33.3)	14 (66.7)	26	12 (46.1)	14 (53.9)
End-point	24	8 (33.3)	16 (66.7)	21	3 (14.3)	18 (85.7)	28	9 (32.1)	19 (67.9)

Table 6.3. Number of accelerometer files for each intervention condition and each time point, with number (n) and percentage (%) of participants complying and not complying with 8 hour wear time criteria across time points

>8 hours	Moderate			Vigorous			Control		
	N	Compliers n (%)	Non compliers n (%)	N	Compliers n (%)	Non compliers n (%)	N	Compliers n (%)	Non-compliers n(%)
Baseline	26	19 (73.1)	7 (26.9)	23	18 (78.3)	5 (21.7)	27	20 (74.1)	7 (25.9)
Mid-point*	26	10 (38.5)	16 (61.5)	21	8 (38.1)	13 (61.9)	26	18 (69.2)	8 (30.8)
End-point	24	11 (45.8)	13 (54.2)	21	6 (28.6)	15 (71.4)	28	11 (39.3)	17 (60.7)

*p < 0.05

6.3.4. Selection bias

Tables 6.4 show mean and standard deviations for compliers and non-compliers in each condition across baseline, mid-point and end-point. Significant differences are indicated for compliers and non-compliers within conditions and between compliers in each condition.

6.3.4.1. Selection bias at 10 hours for baseline

Independent t-tests revealed significant differences in baseline measures between compliers and non-compliers with the 10 hour wear criteria. Control condition compliers had lower depression ($t_{(24)} = -2.377, p = 0.026$) and anxiety ($t_{(19.775)} = -2.187, p = 0.041$), along with higher perceptions of physical strength ($t_{(22.921)} = 2.801, p = 0.010$) and perceptions of physical conditioning ($t_{(21.748)} = 2.122, p = 0.045$) than non-compliers. No differences were apparent for moderate or vigorous intervention conditions for 10 hour wear criteria. One way ANOVAs revealed no significant differences between compliers in each intervention for the 10 hour wear time criteria.

6.3.4.2. Selection bias at 8 hours for baseline

When employing the 8 hour wear time criteria, differences were apparent between compliers and non-compliers in the vigorous condition. Compliers reported lower levels of depression ($t_{(19)} = -2.180, p = 0.042$) and higher perceptions of body attractiveness ($t_{(19)} = 2.259, p = 0.036$). Only one difference was apparent for the control condition in the 8 hour wear; compliers reported higher perceptions of sports competence than non-compliers ($t_{(23)} = 2.153, p = 0.042$). Importantly, one way ANOVAs showed no differences between compliers in each arm of the trial for the 8 hour wear time criteria.

6.3.4.3. Selection bias at 10 hours for Mid-point measures

For the 10 hour wear criteria, differences between compliers and non-compliers at the mid-point measurement period were observed within the moderate condition. Compliers had significantly lower BMI percentiles than non-compliers ($t_{(23)} = -2.231, p = 0.036$). Differences were also apparent within the vigorous condition with compliers being significantly shorter ($t_{(18)} = -2.623, p =$

0.017) and having lower body mass ($t_{(17.598)} = -2.422$, $p = 0.026$) than non-compliers whilst also reporting higher levels of depression ($t_{(9)} = 2.292$, $p = 0.047$), along with lower perceptions of sports competence ($t_{(10)} = -2.809$, $p = 0.019$) and physical strength ($t_{(10)} = -3.844$, $p = 0.003$). In the control condition, compliers had lower perceptions of body attractiveness ($t_{(22)} = -2.548$, $p = 0.018$).

6.3.4.4. Selection bias at 8 hours for mid-point measures

During the application of the eight hour wear criteria, differences were apparent between compliers and non-compliers in the moderate condition for body mass ($t_{(24)} = -2.495$, $p = 0.020$) and BMI percentiles ($t_{(23)} = -2.529$, $p = 0.019$), for both variables, compliers were lower than non-compliers. In the vigorous condition, significant differences were apparent for depression ($t_{(9)} = 2.297$, $p = 0.047$), perceptions of sports competence ($t_{(10)} = -2.809$, $p = 0.019$) and perception of physical strength ($t_{(10)} = -3.844$, $p = 0.003$), with compliers reporting higher depression and lower self-perception scores than non-compliers. No differences were apparent between compliers and non-compliers in the control condition. As with the baseline measurement, ANOVA's revealed no significant differences between compliers in each condition, for either the 10 hour or the 8 hour wear criteria.

Table 6.4 Means \pm SD for compliers and non-compliers with two wear time criteria for intervention criteria at baseline

Baseline	Moderate		Vigorous		Control	
	Compliers	Non-compliers	Compliers	Non-compliers	Compliers	Non-compliers
> 10 hours						
Height	144.3 \pm 9.0	140.9 \pm 6.4	139.3 \pm 8.09	143.8 \pm 6.6	138.2 \pm 6.3	139.3 \pm 7.1
Mass	41.02 \pm 10.9	36.3 \pm 5.3	35.4 \pm 5.57	40.4 \pm 7.6	34.9 \pm 4.2	34.3 \pm 6.7
BMI percentiles	68.1 \pm 23.1	62.2 \pm 24.2	64.9 \pm 18.52	76.8 \pm 18.7	66.2 \pm 24.8	54.5 \pm 28.2
CDI	7.4 \pm 5.8	8.6 \pm 6.2	6.6 \pm 5.04	7.4 \pm 4.8	5.6 \pm 4.9*	10.8 \pm 6.1
STAIC	29.8 \pm 7.2	33.5 \pm 8.6	30.7 \pm 6.3	29.3 \pm 6.7	28.5 \pm 4.3*	34.0 \pm 8.3
Physical self-perceptions						
Sport competence	3.2 \pm 0.7	3.1 \pm 0.6	2.9 \pm 0.6	3.2 \pm 0.6	3.1 \pm 0.6	2.9 \pm 0.6
Body Attractiveness	2.8 \pm 0.7	2.9 \pm 0.8	2.9 \pm 0.5	2.7 \pm 0.7	2.8 \pm 0.6	2.5 \pm 0.4
Physical Strength	2.8 \pm 0.6	2.6 \pm 0.7	2.8 \pm 0.5	3.0 \pm 0.5	2.8 \pm 0.4*	2.4 \pm 0.4
Physical Condition	3.2 \pm 0.5	3.2 \pm 0.6	3.0 \pm 0.4	3.3 \pm 0.5	3.3 \pm 0.5*	2.7 \pm 0.7
Physical self-worth	3.2 \pm 0.5	3.0 \pm 0.7	3.1 \pm 0.4	3.1 \pm 0.4	3.1 \pm 0.5	2.9 \pm 0.4
Global self-worth	3.4 \pm 0.5	3.2 \pm 0.6	3.4 \pm 0.4	3.3 \pm 0.4	3.4 \pm 0.5	2.9 \pm 0.8
> 8 hours						
Height	143.8 \pm 6.9	137.8 \pm 7.9	139.4 \pm 7.5	146.8 \pm 6.9	138.8 \pm 7.0	138.7 \pm 5.7
Mass	39.5 \pm 8.2	34.3 \pm 7.0	36.7 \pm 6.3	38.7 \pm 8.4	34.5 \pm 5.7	34.9 \pm 5.3
BMI percentiles	67.4 \pm 19.3	55.7 \pm 33.7	68.8 \pm 19.2	66.3 \pm 21.2	58.3 \pm 28.2	65.6 \pm 23.0
CDI	7.7 \pm 6.2	9.3 \pm 5.5	5.6 \pm 4.9*	10.6 \pm 1.9	7.4 \pm 6.2	10.3 \pm 5.4
STAIC	31.7 \pm 7.03	33.0 \pm 11.3	29.5 \pm 6.4	33.5 \pm 5.8	30.3 \pm 5.7	34.6 \pm 9.9
Physical self-perceptions						
Sport competence	3.2 \pm 0.7	3.2 \pm 0.6	2.9 \pm 0.6	3.2 \pm 0.6	3.0 \pm 0.6*	2.9 \pm 0.7
Body Attractiveness	2.9 \pm 0.8	2.9 \pm 0.5	2.9 \pm 0.7*	2.3 \pm 0.8	2.7 \pm 0.5	2.5 \pm 0.4
Physical Strength	2.8 \pm 0.6	2.5 \pm 0.9	2.8 \pm 0.5	3.0 \pm 0.6	2.6 \pm 0.5	2.6 \pm 0.4
Physical Condition	3.2 \pm 0.6	3.3 \pm 0.5	3.1 \pm 0.5	3.0 \pm 0.4	3.2 \pm 0.6	2.6 \pm 0.7
Physical self-worth	3.2 \pm 0.5	2.7 \pm 0.8	3.1 \pm 0.5	2.9 \pm 0.2	3.0 \pm 0.5	2.9 \pm 0.2
Global self-worth	3.3 \pm 0.5	3.2 \pm 0.8	3.4 \pm 0.4	3.2 \pm 0.4	3.3 \pm 0.8	3.1 \pm 0.6

Mid-point	Moderate		Vigorous		Control	
	Compliers	Non-compliers	Compliers	Non-compliers	Compliers	Non-compliers
> 10 hours						
Height	139.7 ± 7.28	144.2 ± 7.3	135.9 ± 3.7*	143.7 ± 6.8	141.2 ± 7.6	138.2 ± 5.4
Mass	34.1 ± 6.58	40.9 ± 8.1	33.6 ± 3.3*	39.1 ± 6.9	35.9 ± 5.2	33.8 ± 5.4
BMI percentiles	50.0 ± 25.59*	72.2 ± 19.9	65.3 ± 18.4	69.3 ± 20.7	63.8 ± 28.4	56.3 ± 26.9
CDI	6.1 ± 4.38	8.5 ± 7.4	8.8 ± 4.4*	3.6 ± 3.1	7.6 ± 7.7	5.3 ± 4.3
STAIC	30.2 ± 6.11	32.2 ± 10.2	35.0 ± 9.4	26.6 ± 5.3	28.00 ± 7.9	30.8 ± 6.8
Physical self-perceptions						
Sport competence	2.9 ± 0.77	3.4 ± 0.5	2.5 ± 0.6*	3.5 ± 0.5	2.9 ± 0.5	2.9 ± 0.7
Body Attractiveness	2.6 ± 0.50	2.9 ± 0.6	2.9 ± 0.7	3.3 ± 0.4	2.5 ± 0.4*	3.0 ± 0.5
Physical Strength	2.5 ± 0.60	2.9 ± 0.6	2.1 ± 0.4*	3.2 ± 0.5	2.6 ± 0.6	2.8 ± 0.5
Physical Condition	3.2 ± 0.61	3.2 ± 0.7	3.0 ± 0.6	3.5 ± 0.5	2.9 ± 0.4	3.1 ± 0.5
Physical self-worth	3.1 ± 0.42	3.1 ± 0.6	2.8 ± 0.5	3.4 ± 0.6	2.9 ± 0.6	2.9 ± 0.5
Global self-worth	3.5 ± 0.56	3.3 ± 0.7	3.4 ± 0.4	3.7 ± 0.5	3.2 ± 0.6	3.5 ± 0.5
> 8 hours						
Height	140.6 ± 6.9	144.5 ± 7.5	138.2 ± 7.0	143.0 ± 6.6	139.5 ± 7.3	139.9 ± 4.9
Mass	34.5 ± 5.5*	41.9 ± 8.5	34.8 ± 4.4	38.9 ± 7.1	34.3 ± 5.2	35.9 ± 5.7
BMI percentiles	52.9 ± 23.9*	74.8 ± 18.9	65.0 ± 16.9	69.8 ± 21.5	55.5 ± 29.7	67.3 ± 21.2
CDI	6.3 ± 6.2	8.8 ± 7.1	8.8 ± 4.4*	3.6 ± 3.1	7.5 ± 6.4	3.8 ± 4.5
STAIC	30.9 ± 6.0	32.2 ± 10.9	32.0 ± 10.6	27.6 ± 5.0	30.2 ± 8.0	27.9 ± 5.5
Physical self-perceptions						
Sport competence	3.1 ± 0.7	3.4 ± 0.5	2.5 ± 0.6*	3.5 ± 0.5	2.9 ± 0.6	3.1 ± 0.7
Body Attractiveness	2.7 ± 0.6	2.9 ± 0.6	2.9 ± 0.7	3.3 ± 0.4	2.7 ± 0.5	3.0 ± 0.6
Physical Strength	2.6 ± 0.6	2.9 ± 0.7	2.1 ± 0.4*	3.2 ± 0.5	2.6 ± 0.5	2.9 ± 0.5
Physical Condition	3.2 ± 0.6	3.2 ± 0.6	3.0 ± 0.6	3.5 ± 0.5	2.9 ± 0.4	3.1 ± 0.7
Physical self-worth	3.1 ± 0.5	3.1 ± 0.6	2.8 ± 0.5	3.4 ± 0.6	2.9 ± 0.5	2.9 ± 0.5
Global self-worth	3.5 ± 0.7	3.2 ± 0.7	3.4 ± 0.4	3.7 ± 0.5	3.3 ± 0.6	3.5 ± 0.6

End point	Moderate		Vigorous		Control	
	Compliers	Non-compliers	Compliers	Non-compliers	Compliers	Non-compliers
> 10 hours						
Height	141.2 ± 5.8	143.6 ± 8.7	138.5 ± 4.1	144.7 ± 7.8	140.7 ± 8.3	139.7 ± 6.7
Mass	37.2 ± 5.1	39.8 ± 10.1	33.6 ± 5.3	39.8 ± 6.0	37.1 ± 4.5	34.9 ± 6.1
BMI percentiles	69.3 ± 23.7	65.5 ± 23.4	58.3 ± 21.2	71.4 ± 17.5	73.3 ± 18.4	58.4 ± 24.3
CDI	3.6 ± 4.5	7.2 ± 5.7	6.7 ± 7.4	5.5 ± 5.6	4.3 ± 5.4	7.2 ± 6.6
STAIC	25.6 ± 8.9 ^b	27.3 ± 7.8	40.0 ± 1.0 ^{*b}	27.6 ± 7.1	25.0 ± 5.7 ^b	30.1 ± 7.4
Physical self-perceptions						
Sport competence	3.2 ± 0.6	3.4 ± 0.5	2.3 ± 0.3 ^b	3.5 ± 0.9	3.4 ± 0.5 ^{*b}	2.81 ± 0.6
Body Attractiveness	2.8 ± 0.7	3.2 ± 0.7	2.6 ± 0.5	3.1 ± 0.4	2.8 ± 0.5	2.64 ± 0.5
Physical Strength	3.0 ± 0.7	3.1 ± 0.8	2.2 ± 0.4	2.9 ± 0.6	2.9 ± 0.4	2.70 ± 0.4
Physical Condition	3.2 ± 0.6	3.4 ± 0.6	2.5 ± 0.3	2.9 ± 0.6	3.3 ± 0.6 [*]	2.81 ± 0.5
Physical self-worth	3.2 ± 0.7	3.3 ± 0.7	2.6 ± 0.5	3.1 ± 0.5	2.9 ± 0.4	2.95 ± 0.5
Global self-worth	3.2 ± 0.8	3.4 ± 0.6	2.9 ± 0.1	3.3 ± 0.6	3.2 ± 0.4	3.31 ± 0.4
> 8 hours						
Height	141.2 ± 5.7	143.9 ± 9.4	140.5 ± 9.7	145.2 ± 6.1	140.5 ± 7.4	139.7 ± 7.1
Mass	37.0 ± 4.6	40.6 ± 11.0	36.5 ± 6.6	39.8 ± 5.9	37.4 ± 4.4	34.5 ± 6.1
BMI percentiles	68.0 ± 21.9	65.7 ± 24.8	67.3 ± 19.5	70.3 ± 18.3	74.3 ± 17.8 [*]	55.7 ± 24.0
CDI	5.2 ± 6.3	6.7 ± 4.9	7.7 ± 4.8	4.7 ± 6.02	4.6 ± 5.4	7.4 ± 6.7
STAIC	26.6 ± 8.8	26.9 ± 7.7	35.8 ± 7.1 ^{*b}	26.6 ± 6.7	25.6 ± 5.5	30.3 ± 7.8
Physical self-perceptions						
Sport competence	3.2 ± 0.6	3.5 ± 0.4	3.3 ± 1.7	3.3 ± 0.4	3.3 ± 0.5 [*]	2.8 ± 0.6
Body Attractiveness	2.8 ± 0.7	3.2 ± 0.8	2.8 ± 0.5	3.1 ± 0.4	2.8 ± 0.5	2.9 ± 0.5
Physical Strength	2.9 ± 0.7	3.1 ± 0.8	2.4 ± 0.4 [*]	3.0 ± 0.6	2.9 ± 0.4	2.7 ± 0.4
Physical Condition	3.2 ± 0.6	3.5 ± 0.5	2.7 ± 0.4	2.9 ± 0.7	3.2 ± 0.6	2.8 ± 0.5
Physical self-worth	3.2 ± 0.7	3.3 ± 0.7	2.8 ± 0.4	3.1 ± 0.6	3.0 ± 0.4	2.9 ± 0.5
Global self-worth	3.2 ± 0.8	3.5 ± 0.6	3.1 ± 0.4	3.3 ± 0.6	3.1 ± 0.4	3.3 ± 0.4

* p < 0.05 within condition differences; ^b p < 0.05 between condition differences (compliers only)

6.3.4.5. Selection bias for 10 hours at end-point measurement

At the end-point measurement, using the 10 hour criterion, compliers in the vigorous condition reported significantly more anxiety than non-compliers ($t_{(14.697)} = 6.222$, $p < 0.001$). Compliers in the control condition showed significantly higher perceptions of sports competence ($t_{(24)} = 2.672$, $p = 0.013$) and physical conditioning ($t_{(24)} = 2.277$, $p = 0.032$).

One way ANOVAs showed significant differences between compliers in each condition for anxiety ($F_{(2,17)} = 5.721$, $p = 0.013$), with compliers in the vigorous condition demonstrating more anxiety than those in the moderate or control. Furthermore, differences in perceptions of sports competence were also apparent ($F_{(2,17)} = 5.173$, $p = 0.018$), with the control condition reporting significantly higher perceptions of sports competence than the vigorous condition

6.3.4.6. Selection bias for 8 hour criteria at end-point

When applying the 8 hour criterion at end-point measurement, compliers in the vigorous condition reported significantly higher anxiety ($t_{(15)} = 2.676$, $p = 0.017$) and lower perceptions of physical strength ($t_{(16)} = -2.282$, $p = 0.041$) than non-compliers. In the control condition, compliers had significantly higher BMI percentiles ($t_{(24)} = 2.163$, $p = 0.041$) and higher perceptions of sports competence ($t_{(24)} = -2.529$, $p = 0.018$).

One way ANOVAs revealed significant difference for anxiety between compliers in each condition ($F_{(2,25)} = 4.273$, $p = 0.025$); compliers in the vigorous condition reported more anxiety than those in the control.

6.3.5. Misclassification of time in activity intensities

6.3.5.1. 10 hour criteria

The time accumulated in PA intensities (table 6.5) shows an increase in MVPA across time points for all conditions using the 10 hour wear time criterion, time spent in vigorous intensity activity increased for the vigorous and the control condition, but only minor increases were demonstrated in the moderate intervention condition. Time in moderate activity increased across all conditions. Time in light intensity activity increased from baseline to end-point in the

moderate and control conditions, but reduced in the vigorous activity condition, with a similar pattern apparent for sedentary time. Baseline sedentary time was significantly different between conditions ($F_{(2,34)} = 4.997$, $p = 0.013$), with the moderate condition accumulating significantly more sedentary time than the control condition, however, the assumption of homogeneity of variance was violated for this analysis, these results should therefore be taken lightly. At mid-point, significant differences were apparent between conditions for time accumulated in light activity intensity ($F_{(2,23)} = 4.184$, $p = 0.028$), the moderate intervention conditions recorded more time in light activity than the control condition. Differences between conditions were not tested at end-point for this wear time criteria due to small sample sizes.

6.3.5.2. 8 hour criteria

At baseline for the 8 hour wear criteria, time accumulated in light intensity activity was significantly different across conditions ($F_{(2,54)} = 4.018$, $p = 0.023$), with the vigorous condition accumulating more time in light activity than the moderate condition. For mid-point measures, significant differences were apparent for time in light ($F_{(2,33)} = 3.355$, $p = 0.045$), with the moderate intervention condition recording higher levels of light activity than the control. Between condition differences were also apparent for time in moderate intensity activity ($F_{(2,33)} = 8.325$, $p = 0.001$); differences were apparent for all conditions, with moderate intervention condition reporting the longest time and the control condition recording the shortest period. Differences for time in activity intensities were not assessed at end-point due to small sample sizes.

Time reported in all activity intensities was consistently underestimated with the use of the more lenient wear time criteria (8 hour) across all conditions and time points, with the largest differences apparent for time accumulated in sedentary activity, and smaller differences apparent for time in higher intensity activities (moderate and vigorous activity intensities). Furthermore, between condition differences varied depending upon the wear time criteria employed. Specifically, more differences were apparent with the use of the 8 hour wear than the 10 hour wear, demonstrating the occurrence of misclassification through the use of more lenient wear times.

Table 6.5. Mean (\pm SD) time accumulated in different activity intensities for compliers with different wear time criteria by condition at each time point ($p < 0.05$)

		N	Time in Sedentary	Time in Light	Time in Moderate	Time in Vigorous	Time in MVPA
10 hours							
Baseline	Moderate	10	620.09 \pm 98.28	122.15 \pm 12.17	30.89 \pm 6.77	27.42 \pm 13.46	58.32 \pm 18.31
	Vigorous	15	572.16 \pm 35.52	129.12 \pm 26.32	28.06 \pm 6.6	22.06 \pm 5.92	50.12 \pm 10.83
	Control	13	546.25 \pm 18.86	117.22 \pm 14.30	25.45 \pm 4.02	21.93 \pm 7.07	47.37 \pm 87.70
Midpoint	Moderate	7	636.13 \pm 114.99	140.04 \pm 13.5*	30.70 \pm 3.88	28.31 \pm 9.96	59.02 \pm 11.86
	Vigorous	7	558.52 \pm 46.04	129.57 \pm 24.44	28.49 \pm 5.67	23.54 \pm 4.68	52.04 \pm 8.99
	Control	12	565.52 \pm 69.23	117.33 \pm 12.80*	26.19 \pm 4.23	24.40 \pm 10.01	50.59 \pm 12.56
Endpoint	Moderate	8	636.50 \pm 121.75	136.48 \pm 25.43	36.51 \pm 10.01	28.0 \pm 16.02	64.51 \pm 24.37
	Vigorous	3	554.55 \pm 37.18	116.92 \pm 17.99	38.66 \pm 13.88	28.78 \pm 4.40	67.44 \pm 18.20
	Control	9	592.34 \pm 136.46	138.57 \pm 31.21	31.91 \pm 11.11	25.74 \pm 10.80	57.64 \pm 18.62
8 hours							
Baseline	Moderate	19	552.92 \pm 83.37	107.87 \pm 15.75*	27.46 \pm 5.45	23.05 \pm 8.78	50.51 \pm 12.54
	Vigorous	18	555.71 \pm 43.49	124.58 \pm 25.98*	27.23 \pm 6.57	21.33 \pm 6.65	48.56 \pm 11.75
	Control	20	521.05 \pm 28.61	110.54 \pm 14.46	24.22 \pm 2.99	20.95 \pm 5.78	45.18 \pm 6.29
Midpoint	Moderate	10	585.96 \pm 126.60	133.28 \pm 19.55*	31.05 \pm 3.36*	27.43 \pm 9.28	56.31 \pm 9.13
	Vigorous	8	549.22 \pm 44.66	118.50 \pm 26.55	25.84 \pm 5.87*	21.73 \pm 5.03	47.56 \pm 9.78
	Control	18	528.79 \pm 69.43	113.65 \pm 15.25*	24.33 \pm 3.78*	22.30 \pm 7.89	46.63 \pm 10.10
Endpoint	Moderate	11	583.15 \pm 123.02	135.04 \pm 23.60	38.34 \pm 8.69	28.23 \pm 13.22	66.59 \pm 20.14
	Vigorous	6	485.11 \pm 69.48	121.87 \pm 18.38	35.76 \pm 8.69	27.05 \pm 5.90	62.81 \pm 13.78
	Control	11	576.93 \pm 127.60	134.50 \pm 28.10	30.47 \pm 9.33	24.43 \pm 9.25	54.90 \pm 15.35

6.3.6. Effect of the intervention upon PA

No intervention effects were apparent for the examination of between condition differences for PWB, regardless of the availability of PA data. However only those with PA data can be assessed to examine the mechanisms, for example, did PA change across the intervention and did PWB increase for those that changed their PA? Additionally, the selection bias noted above may indicate that the intervention effects differ for those with PA data. Therefore in an attempt to examine whether the effects of the intervention were apparent in compliers, per protocol analysis should be undertaken exclusively with the compliers to the wear time criteria. This however was not possible within the present study as low numbers complying at end-point resulted in insufficient power.

6.4. Discussion

The aim of the present study was to assess the effectiveness of an intervention designed to increase PWB through changing PA levels in children and to assess variance in wear time compliance between three intervention conditions. Selection bias and misclassification of time in activity intensities across conditions and time points were explored. Finally, the study aimed to assess whether differences in compliance between conditions would impact the results of the intervention, through an exploration of the mechanisms used.

The main findings revealed no significant intervention effect upon PWB, though the effect of the intervention upon time in PA could not be assessed, therefore it is unclear as to whether the intervention achieved the proposed mechanism (changing PA) by which PWB could be increased. Compliance with wear time decreased across all conditions; though an association between compliance and intervention condition was only apparent at mid-point for the ≥ 8 hour wear criteria. Despite this, examination of differences between compliers and non-compliers revealed selection bias within conditions that differed across time points. No differences however were apparent between compliers of each condition, except at end-point measurement. Misclassification of times

accumulated in each activity intensity occurred in all conditions at each time point.

6.4.1 Intervention effects

The results of the present study indicated no between group differences post intervention, however the effects of the intervention upon PSW approached significance; the moderate condition reported the highest levels of PSW while the control condition reported the lowest levels. This trend is in the expected direction and builds upon the results observed in chapter five. In comparison to the wider literature, the fact that no intervention effect occurs provides partial support for Schneider et al. (2008) and Huang, et al., (2007), who showed no between group differences in self-esteem post intervention, though disagrees with the finding of meta-analyses that demonstrate positive effects of PA interventions upon depression (Larun et al., 2009; Ahn & Fedewa, 2011; Brown et al., 2013) and aspects of self-esteem (Ekeland et al., 2005 and Ahn & Fedewa, 2011). Additionally, the findings of cross-sectional studies would indicate that PWB should increase with the employment of higher intensity PA (Parfitt et al., 2009; Page et al., 2010; Goldfield et al., 2011; Chapter five of this thesis)

In order to assess whether the intervention was unsuccessful due to no change in PA, the effects of the intervention upon PA first needs to be established. However the low compliance levels in each condition using either the 10 hour or the 8 hour wear time criteria resulted in participant numbers that were too low to carry out a statistical test with sufficient power. Therefore, the reason for no effect of the intervention upon the proposed outcome (PWB) is presently unknown. Interestingly, loss to follow up only occurred in the vigorous condition, with two participants dropping out prior to the mid-point measurement. Despite this, complete psychological profiles were not available for the entirety of one condition at any time point. Ten participants in the vigorous condition failed to provide some or all PWB data at mid-point. This occurred through absences during the measurement period, participants failing to complete questionnaires and, despite this being the second measurement point some children had spoilt data through incorrect completion. Similarly, the moderate and control condition had missing or spoilt data, yet not to the same

extent. The vast amount of missing data however occurred during the mid-point measurement, which for the present study was not included in the final analysis, however for studies using imputation methods such as last observation carried forward, this may result in substantial numbers of children being imputed from baseline rather than mid-point measures and could result in an underestimation of any intervention effects (Montori & Guyatt, 2001).

6.4.2 Wear time compliance

The lack of understanding regarding the reason for the apparent failure of this intervention arises from the low compliance with the activity measurement protocol (i.e. wear time). Compliance decreased with subsequent time points across all conditions, with lower compliance apparent for the stricter wear time criteria. This supports the findings of Audrey et al. (2012) and the results presented in chapter five. There was a small increase in compliance in the moderate condition from mid-point to end-point, which was apparent for both 10 hours (26% to 33%) and 8 hours wear (38.5% to 45%). This increase was the result of one extra child complying with each criterion. For the 8 hours wear criteria, the vigorous activity condition consistently had the lowest compliance and generally had the largest drop in compliance between time points for both wear time criteria.

Chi-squared statistics showed that intervention condition was only related to compliance at mid-point for the 8 hours wear criteria, where those in the control condition demonstrated higher compliance than the two intervention conditions. Despite this, a trend occurs, as the control condition recorded a consistently higher compliance with the more lenient wear time (> 8 hours), along with the highest compliance at mid-point and end-point for the more stringent wear time criteria (>10 hours). This is contrary to what was hypothesised; it was expected that the control condition would experience the largest decline in compliance across subsequent wear, as reported by Magnusson et al. (2011). The intervention itself may have increased the likelihood of non-compliance; intervention conditions may have experienced an increase in social desirability across time points that may not have affected the control condition. Specifically, participants receiving the intervention maybe aware that their PA should be increasing, if they feel that this has not occurred, the participants may choose

not to comply. The control condition however has no preconceived expectations. This fits with the definition of social desirability provided in chapter three. Social desirability (Crown & Marlowe, 1964), may occur when participants over-rate traits deemed desirable and underestimate less desirable traits (Klesges et al., 2004), in other words, the participants control the outcome of the exposure measure in order to protect themselves from portraying a negative image (Phillips, 1973; Niemi, 1993).

6.4.2.1 Selection bias and wear time compliance

Selection bias differed between conditions, across time points and between wear time criteria. At baseline it appears that differences were only apparent between compliers and non-compliers with the 10 hour criteria in the control condition; with participants meeting the wear time requirements demonstrating better psychological profiles than non-compliers. Interestingly, the same differences for the control condition were not apparent with the use of the 8 hours wear criteria at baseline. Rather, control compliers with this wear criterion had higher perceptions of their sports competence. These results indicate a reduction in selection bias with the inclusion of more participants in the control condition.

Compliers and non-compliers did not differ in the moderate and vigorous intervention conditions for 10 hour wear at baseline. The application of the 8 hour wear criteria saw differences for the vigorous condition; compliers reported significantly higher perceptions of body attractiveness and lower levels of depression than non-compliers. These results demonstrate variations in selection bias according to the intervention conditions, though generally compliers at baseline report better PWB across all conditions. The results are representative of those reported in chapter five. It should be noted however that a reduction in wear time, aimed at overcoming low compliance rates should reduce selection bias through increased participant numbers, yet this is only partially true (e.g. selection bias in the control condition decreased), the results demonstrate that different bias is introduced along with misclassification, which again may impact on the results of the intervention.

Baseline differences in PWB for compliers and non-compliers in the present study are of particular interest, as the baseline measures are included as a

covariate in the ANCOVA. If the intervention effects depended only on those who complied with wear time requirements, so as to address the mechanisms, then a biased baseline sample would be used, with generally higher PWB than those who have not complied. This could result in a potential underestimation of the effects of the intervention on PWB. Although as noted previously, it was not possible to test this in the present study.

When assessing selection bias at the mid-point measurement period, a substantial number of differences occurred. In the control condition, compliers reported lower perceptions of body attractiveness than non-compliers for 10 hours wear, but no differences were apparent for the 8 hour wear. The compliers in the moderate intervention condition were found to have a lower mean BMI percentile compared to non-compliers; this was apparent for the application of both wear time criteria. Additionally, compliers for the moderate condition also had a lower body mass with the 8 hour wear period employed. Similar findings were apparent for the vigorous activity condition, with compliers being shorter and weighing less than non-compliers for the 10 hour wear period. Additionally, compliers demonstrated poorer psychological profiles (higher levels of depression, and lower perceptions of physical strength and sports competence) than non-compliers for both wear time criteria.

Differences between compliers and non-compliers at mid-point are interesting; as it appears that the intervention conditions (not control) experience the largest selection bias. This supports the chi-squared statistic presented earlier that notes associations between condition and compliance at this time point. Compliers in moderate and vigorous conditions appear to be smaller in body size than non-compliers; this may be due to the intervention raising awareness of body size and shape, resulting in selection bias through social desirability (Crown & Marlowe, 1964). Those children who are heavier, may be less inclined to wear a PA monitor for fear of showing low levels of PA (whether this is true or not). As previously noted, this raised awareness may not have occurred within the control condition and therefore may be the reason, not only for no anthropometric differences between compliers and non-compliers but also for the continued high compliance rates within this condition. Alongside anthropometric differences, compliers in the vigorous condition reported lower levels of PWB (as noted above). These results are unexpected, as it is thought

that those with higher levels of depression would comply less, as was observed for the whole sample in chapter five. Additionally, it should be noted that increased selection bias occurred in the 8 hour wear criteria for the moderate and vigorous conditions, which again is opposite to what was expected.

At the end-point measurement, compliers for the vigorous condition demonstrated significantly more anxiety, than non-compliers; this was apparent for both 10 hour and 8 hour wear criteria. Compliers with the 8 hour wear criteria in the vigorous condition also had lower perceptions of physical strength than non-compliers. No differences were apparent for the moderate activity condition. Yet, in the control condition, compliers with the 10 h criterion had higher perceptions of their sports competence and physical conditioning and compliers with the 8 h criterion had higher BMI percentiles and perceptions of their sports competence. The findings of the control condition at end-point seem to demonstrate an opposite trend to what was reported at the two previous time points and in other intervention conditions. At baseline and mid-point physical self-perceptions were lower in compliers than in non-compliers, this was not the case for the control condition at end-point, demonstrating that selection bias may differ dependent upon the time point and conditions examined. The reasons for which are not known; though social desirability was noted previously as a possible reason for smaller decreases in compliance with the control condition.

The selection bias, resulting from non-compliance, apparent across an activity intervention appears to be different for each of the intervention conditions and at each time point, indicating that intervention effects would vary according to level of non-compliance. If imputation methods were used for missing values then the last observation carried forward method would introduce bias if the time point of the last observation varied between groups. However, assessment of differences between compliers in each condition demonstrated no-differences for either wear time at baseline or mid-point, indicating that selection bias within conditions did not have an effect between conditions. At end-point however, using the 10 hour wear criteria, the vigorous condition reported higher anxiety than the moderate or control conditions. For the eight hour criteria, anxiety was higher in the vigorous than the control condition. The control condition reported higher perceptions of sports

competence than the vigorous condition with the 10 hour criteria. The occurrence of differences at end-point is indicative of the effect that selection bias could have on the results of an intervention. If only completers had been included in the present intervention, it could have been concluded that the intervention results in higher levels of anxiety in the vigorous condition than the other conditions. However this may purely have been because of the compliers that were included in the vigorous condition were higher in anxiety than the non-compliers, a bias that did not occur in the other conditions.

6.4.2.2. Misclassification between wear times

The results of the present study demonstrate that time in all activity intensities was under-estimated when using a more lenient wear time criteria. Time in sedentary and light appear to be the most underestimated intensities across all time points, as was demonstrated in chapter five. These results also agree with those highlighted in Herrmann et al. (2012), who reported larger underestimations in lower intensities.

At baseline, larger differences were apparent between 10 and 8 hour criteria for the moderate intervention conditions, with difference of 67 minutes in sedentary, and 15 minutes in light when classified by different wear time criteria. These are larger than the differences in time accrued in sedentary and light of 17 minutes and 5 minutes respectively for the vigorous condition and 25 minutes and 7 minutes for the control condition. These larger underestimations for the moderate condition also seem to occur at midpoint with time in sedentary underestimated by 51 minutes when using 8 hour criteria, compared to 9 minutes for vigorous and 37 minutes for control. Additionally, differences between conditions for time spent in activity intensities varied depending upon the wear time criteria employed. Specifically, no differences were apparent between conditions at baseline with the 10 hour criteria, yet with the eight hour criteria, participants in the vigorous condition accrued significantly more time in light intensities (125 minutes) than participants in the moderate condition (108 minutes). By examining the time accumulated in light intensity by these conditions at the 10 hour criteria, it is noted that the moderate condition reports much larger increases in light activity (14 minutes) than the vigorous condition (5 minutes). The fact that time spent in sedentary behaviour and light intensity

activity is underestimated more in one condition than the others indicates that this may be a result of the different selection bias within conditions previously noted.

At the mid-point measurements, the moderate intervention condition recorded significantly more light activity than the control condition (140 and 117 minutes respectively) for the 10 hour and 8 hour wear criteria (133 minutes and 114 minutes respectively). Differences between all conditions for time accumulated in moderate intensity activity were apparent with the use of the 8 hour wear criteria; though these differences were not apparent with the use of the stricter wear time criteria. This may be due to larger underestimations occurring in one condition more than the others, in this case, the vigorous condition showed the greatest underestimation (3 minutes) for the 8 hour criteria, while for the moderate condition, time accumulated in the 8 hour criteria was larger than the 10 hour criteria. Variance in the under- or over – estimation of time in activity intensities maybe due to the selection bias in individual conditions, yet misclassification through shorter wear time may also have occurred. Specifically, the mid-point measurement demonstrated a number of differences between compliers and non-compliers in each intervention condition; selection bias which varied with the use of different wear times.

The results of this study are important as they show how compliance varies between intervention conditions, introducing different selection bias within conditions at each time point. While attempts to overcome this through the use of a more lenient wear time criteria can result not only in different selection bias yet again, but also introduce misclassification through shorter monitoring periods. Additionally, the error associated with estimation of time in activity intensities varies depending upon the intervention condition, which may in turn be a result of selection bias. Although the impact of selection bias and misclassification cannot be assessed in terms of the effect on the intervention, due to small numbers of compliers, the large amount of selection bias apparent would suggest that variations in results may occur. When attempting to examine whether PWB changes as a results of changes in PA, only including compliers with either wear time would result in a bias sample at baseline and end–point, which would therefore alter the observed effect of the intervention.

6.4.3 Limitations

The present study furthered the findings of those in chapter five, by addressing the compliance with wear time criteria, and the resultant selection bias and misclassification between different arms of a trial. However, a number of limitations are worth considering; primarily, the investigation into whether the trial was unsuccessful due to not changing PA could not be undertaken due to small participant numbers. Therefore reasons for the lack of an intervention effect are not apparent. Furthermore, the effects of compliance, selection bias and misclassification upon the intervention could not be considered fully, again due to small participant numbers. Rather, the implication of bias and misclassification could only be inferred by the cross-sectional observations undertaken at each time point. The trial itself was not randomised, which may have controlled for between condition differences at baseline, some of which may be associated with compliance. However, by accounting for baseline scores when testing for between condition differences at end-point should have at least partially negated this problem. Additionally, randomisation does not control for compliance, therefore regardless of whether randomisation had occurred, compliance may still have varied by condition. It should be noted that a large number of participants failed to complete some or all of the PWB questionnaires for one of three reasons; the length of the battery of questionnaires, absence during measurement periods and large amounts of spoilt data obtained on the physical self-perceptions questions. Though this was mainly noted at the mid-point measurement period for the vigorous condition, it was apparent across all conditions. It is thought that the spoilt data occurred due to the structured alternative format employed in the PSPP-CY (Whitehead, 1995), which confused children and resulted in them providing multiple answers to the same question. Finally, an intention to treat analysis could have been used to overcome the low compliance rates that hindered the assessment of changes in PA, and whether changes in PWB occurred for those who had complied or not. However, imputing scores from baseline means that the sample is still limited to those who have complied at baseline, and with the large numbers of non-compliers, any intervention effect within this condition is likely to be attenuated, as intention to treat analysis underestimates intervention

effects to a greater extent when non-compliance is high (Bubbar & Kreder, 2006). Finally, because a single school was allocated to each of the intervention groups, it is not possible to separate out differences in compliance that were due to the intervention or due to the school. A larger cluster, randomised controlled trial could overcome this limitation.

6.5. Conclusion

The present study has highlighted the effect that low compliance can have upon an intervention study, with worsening compliance with wear time guidelines across each condition preventing a complete test of the effects of an intervention. Furthermore, the results show that compliance can be related to intervention conditions, with the between group differences in compliance varying by time point. The present study demonstrated that selection bias and misclassification vary depending upon the intervention conditions and the time-point examined, which may influence the results and subsequent interpretation of the effectiveness of an intervention. The differential non-compliance by group and time also means that intention to treat analysis, using imputation for missing values, could add to the problem of bias. This is particularly important when attempting to assess the mechanisms through which intervention effects can occur. The results of the present study indicate that researchers should not only take into account loss to follow up when designing an intervention, but should also attempt to power studies to account for losses through decreased compliance, if PA is the proposed mechanism through which changes are expected to occur.

This chapter has highlighted how non-compliance with wear time criteria in an intervention study can not only alter the assessment of effectiveness but also render it impossible to do. The next chapter validates and calibrates a new physical activity monitor, which, through design features has the potential to reduce levels of non-compliance, and in turn, selection bias and misclassification.

CHAPTER SEVEN

VALIDATION AND CALIBRATION OF THE GENEa ACCELEROMETER FOR ASSESSMENT OF CHILDREN'S PHYSICAL ACTIVITY⁴

The aim of the following study is to validate and calibrate a new waterproof, wrist worn accelerometer that is hypothesised to increase compliance with advised wear time. The chapter describes the GENEa accelerometer, highlighting how its use may overcome non-compliance, the validation protocol and calibration methods are then described, followed by the development of activity intensity cut-points and a discussion of the study.

7.1. Introduction

Non-compliance with accelerometer wear time, resulting in a significant loss of data is well documented (Colley et al., 2010; Crocker et al., 2001; Toriano et al., 2008), with detailed discussion of the issue presented in both chapters three, five and six of this thesis. The GENEa waveform triaxial accelerometer (Unilever Discover, Colworth, UK; manufactured and distributed by ActivInsights Ltd, Kimbolton, Cambridge, UK), is lightweight (16 g), small (L36xW30xH12mm) and collects data in three axes (vertical, anteroposterior & mediolateral) at a rate between 10 and 80 Hz; with the most recent model, the GeneActiv, capable of recording at up to 100Hz. It is designed to be worn on the wrist and is waterproof, two qualities that largely negate the need to remove the monitor and thus may lead to greater compliance during assessment of habitual activity in children and in turn reduce the possibility for bias and misclassification; aspects which will be addressed in chapter eight.

⁴ A version of this chapter has been published in the *Journal of Science and Medicine in Sport*. The full reference for the paper is: Phillips, L.R.S., Parfitt, G., & Rowlands, A.V., (2013). Calibration of the GENEa accelerometer for assessment of physical activity intensity in children. *Journal of Science and Medicine in Sport*, 16, (2), 124 – 128.

The GENE has been found to have high intra and inter-instrument reliability (coefficient of variation = 1.8% and 2.4% respectively), good criterion-referenced validity when compared to a Multi-axis shaking table (MAST; Instron Structural Testing Systems, Buckinghamshire, UK) ($r = 0.97$) and high concurrent validity with the ActiGraph GT1M ($r = 0.74$) (Esliger et al., 2011). Esliger and colleagues (2011) also found that, irrespective of whether the monitor was worn at the hip or the left wrist, the GENE could be used to distinguish between sedentary, light, moderate and vigorous intensity activities in adults. This study goes some way to providing support for the GENE worn at the wrist as a valid measure of physical activity, however, if this monitor is to be used in child based studies a separate validation and calibration needs to be undertaken within this specific population.

The purpose of the present study was to establish criterion validity of the GENE (via indirect calorimetry) and subsequently develop physical activity intensity cut-points for use with the GENE when assessing the intensity of children's physical activity. Furthermore, the secondary aim was to establish concurrent validity of the GENE, relative to the ActiGraph GT1M.

7.2 Methods

7.2.1. Participants

An opportunistic sample of 44 apparently healthy children ($n = 26$ females, $n = 18$ males) aged between 8 and 14 years, ($\bar{X} = 10.9 \pm 1.9$) from Devon, UK were recruited for the present study. Children were recruited through local schools, which were approached via email. Children had a mean height of 150.3 cm (± 13.0), mass of 41.8 kg (± 10.9), body mass index (BMI) of $18.3 \text{ kg}\cdot\text{m}^{-2}$ (± 2.8) and waist circumference of 64.0 cm (± 8.3). Participants peak $\dot{V}O_2$ ranged from $27.9 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ to $58.4 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ ($\bar{X} = 41.6 \pm 7.7$). The Ethics Committee at the University of Exeter granted approval for the study. Prior to data collection, parental informed written consent and the children's assent were obtained (see appendix 2.3 & 2.4, p. 207-315).

7.2.2. Protocol

Participants wore GENE monitors at three locations; one at each wrist, secured using a watch strap, and one at the right hip (positioned along the mid clavicle line). An ActiGraph GT1M was also worn adjacent to the hip mounted GENE. The GENE measuring ± 6 g, can record data at a rate of 80 Hz for up to 8 days, with the new GENEActiv capable of measuring ± 8 g at 100 Hz for up to 7 days; longer measurement periods are achievable with lower sampling frequency.

The GENE and ActiGraph were set to record at 80 Hz and 1 second epochs, respectively. Prior to each participant completing the protocol, all monitors were synchronised with Greenwich Mean Time (GMT). Throughout the testing procedure, $\dot{V}O_2$ and $\dot{V}CO_2$ were measured using the Cosmed K4b2 gas analyser (Cosmed, Rome, Italy). Participants wore a junior face mask, head net and harness. The K4b2 was calibrated with gases of known concentration, prior to commencing testing and on every day of data collection thereafter. The K4b2 has previously been shown to be a valid measure of oxygen uptake (McLaughlin, King, Howley, Bassett & Ainsworth, 2001). Upon completion of the protocol, each participant's accelerometer and calorimetry data were downloaded and stored on a computer.

Upon arrival at the laboratory, the participant's height (cm), seated height (cm), mass (kg), body fatness (assessed using bio-electrical impedance analysis (BIA) (Tanita, TBF-305 scales, Tanita UK Ltd. Middlesex, UK)), waist circumference (cm) and handedness were recorded. Participants were then familiarised with the equipment being used throughout the study, specifically the treadmill (PPS 55 MED- I; Woodway, Germany) and the Nintendo Wii (Nintendo, Windsor, UK).

After being briefed about the testing protocol, participants were fitted with the multiple accelerometers, a heart rate monitor (Polar Vantage NV; Polar, Finland) and the gas analyser (Cosmed K4b2; Cosmed, Rome, Italy). Each participant was then asked to perform a series of activities representative of various aspects of children's daily activity. These were, lying supine, seated

DVD viewing, active computer games (boxing) using a Nintendo Wii, slow walking, brisk walking, slow running and a medium run. All activities were performed for three minutes followed by a two minute rest, with the exception of lying supine which took place for 10 minutes. Using the protocol by Puyau et al (2002) as a guideline, walking and running speeds were set at; 4 km·h⁻¹ for slow walk, 6 km·h⁻¹ for brisk walk and between 8 - 12 km·h⁻¹ for running speeds (adjustable according to age & activity).

7.2.3. Data analysis

Using the GENE Post Processing software (version 1.2.1), the raw 80 Hz triaxial GENE data were summarized into a signal magnitude vector (gravity-subtracted) (SVMgs) expressed in 1 second epochs [see equation below] (Karatonis, Narayanan, Mathie, Lovell & Celler, 2006), as described by Esliger et al. (2011). The resulting SI units for this outcome variable are g · seconds.

Equation 7.1:
$$\text{SVMgs} = \sum \sqrt{x^2 + y^2 + z^2} - g$$

The correction for gravity was undertaken to focus the outcome variable on dynamic rather than static accelerations (Esliger et al., 2011) and enable comparison with the ActiGraph. Subsequently, criterion and concurrent validity were established and activity intensity cut-points created.

Previous research has debated whether age appropriate cut-points are necessary (Reilly et al., 2008; Trost et al., 2002); in order to assess this with the GENE, a series of age-group (group 1 = 8 – 10 years; group 2 = 11 – 12 years; group 3 = 13 – 14 years) X activity mixed model ANOVAs were undertaken to establish whether there were significant differences between age-groups for accelerometer output at each location. Where assumptions of sphericity were violated ($p < 0.05$) the Greenhouse-Geisser (GG) correction factor was applied to adjust the degrees of freedom. Post hoc analysis using a modified Tukey's (Stevens, 2002) was then undertaken to establish where any differences lay.

To establish criterion referenced validity, correlations between steady state $\dot{V}O_2$, taken from the final minute of each activity, and counts for each monitor (3 GENEAs & ActiGraph; taken from 15 seconds after start of activity until activity completion) were established for each participant. These correlations were then transformed into Fisher's Zr values, a mean taken and transformed back into a Pearson's r value. The same method was used to establish concurrent validity, with the individual correlations between each GENEa monitor and the ActiGraph GT1M. To determine whether validity differed across wearing site, correlations were assessed for significant differences (Meng, Rosenthal & Rubin, 2002), see appendix 4.2 (p. 332) for a worked example.

A series of hierarchical regression analyses were undertaken to assess whether height or age accounted for any variance in the METs over and above that accounted for by GENEa output at each wear location.

To establish cut-points for the GENEa monitors, the $\dot{V}O_2$'s for each activity were converted into METS, using age specific resting values (1 MET = 5.92 ml·kg⁻¹·min⁻¹ (8 – 12 year old boys / 8 – 11 year old girls) and 4.85 ml·kg⁻¹·min⁻¹ (13 – 15 year old boys / 12 – 14 year old girls). The activities were then coded into one of four intensity categories; sedentary (<1.5 METS), light (1.5 – 2.99 METS), moderate (3 – 5.99 METS) and vigorous (≥ 6). The accelerometer counts for activities were coded into binary indicator variables (0 or 1) based on intensity (sedentary versus > sedentary, less than moderate versus moderate to vigorous, and less than vigorous versus vigorous) in order for a Receiver Operator Characteristic (ROC) curve analysis to be carried out as described in chapter three. The cut-points were selected to maximise both sensitivity (correctly identifying at or above the intensity threshold) and specificity (correctly excluding activities below the threshold for intensity) (Jago et al., 2007) ROC analysis was undertaken using GraphPad Prism 5 software (Graphpad software, San Diego, USA).

7.3. Results

7.3.1. Descriptives

Table 1 shows the intensity of each activity performed by participants, given in METs, along with the corresponding accelerometer output. Increases in MET values coincided with an increase in accelerometer output, for both the GENEa and the ActiGraph. The exception to this concerns the computer game activity (Nintendo Wii) where the wrist worn GENEa monitors recorded a high number of counts while the MET values remained relatively low.

Table 7.1. Means and standard deviations for ActiGraph (counts-second), GENE (g-second) output and METs

		Lying		Sitting		DVD		Wii		Slow walk		Fast walk		Slow run		Medium run	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
GT1M	Group 1	0.11	0.21	0.18	0.69	0.07	0.25	7.06	10.17	27.57	11.59	60.28	16.55	92.64	27.4	101.35	26.27
	Group 2	0.01	0.04	0.009	0.03	0.005	0.013	2.63	7.65	32.53	6.19	68.27	8.33	123.77	20.32	126.47	16.44
	Group 3	0.14	0.38	0.09	0.22	0.02	0.06	0.17	0.25	28.67	7.52	63.96	10.63	125.14	25.77	132.29	26.19
	Overall	0.09	0.23	0.1	0.46	0.04	0.16	3.82	8.26	29.54	9.15	63.92	13.01	111.38	28.89	118.02	26.6
GENEA	Group 1	2.21	0.94	2.97	1.17	2.15	0.94	50.25	28.74	19.81	13.89	27.67	8.29	67.59	24.74	81.79	22.53
Right	Group 2	2.23	0.85	2.01	0.92	1.45	0.81	49.08	31.76	15.78	9.38	21.98	4.42	61.01	13.44	78.51	15.7
Wrist	Group 3	1.86	0.7	2.24	1.08	1.69	1.11	35.06	15.34	15.52	12.73	19.39	3.1	63.26	14.95	81.39	20.48
	Overall	2.13	0.85	2.46	1.13	1.79	0.97	46.05	27.41	17.36	12.12	23.66	6.94	64.26	19.01	80.55	19.44
GENEA	Group 1	2.16	0.94	2.99	1.06	1.98	1.1	48.84	29.65	20.41	11.97	29.18	7.14	68.66	21.53	84.57	18.65
Left	Group 2	1.93	0.66	2.38	0.97	2.85	4.15	38.74	30.17	19.63	21.92	24.62	10.73	58.41	19.79	75.07	22.99
Wrist	Group 3	1.79	0.73	2.62	0.88	1.73	1.38	31.43	15.11	14.84	11.25	20.89	6.88	67.37	28.13	87.64	36.62
	Overall	1.99	0.79	2.69	1	2.22	2.62	40.86	27.26	18.71	15.82	25.47	8.97	64.75	22.75	81.98	25.76
GENEA	Group 1	1.06	0.47	0.99	0.84	0.59	0.57	7.03	4.95	25.75	22.12	35.56	7.7	64.58	9.05	72.81	6.79
Hip	Group 2	1.12	0.71	0.57	0.31	0.57	0.61	6.15	5.41	16.37	6.27	31.52	10.19	56.23	12.08	62.24	16.52
	Group 3	1.1	0.37	0.75	0.51	0.59	0.62	4.61	2.17	12.08	3.28	24.95	5.22	51.89	12.47	63.16	13.16
	Overall	1.09	0.53	0.79	0.64	0.59	0.59	6.13	4.61	19.13	15.57	31.63	9.03	58.56	12.02	66.66	13.24
METs	Group 1	0.92	0.25	0.95	0.20	0.88	0.19	2.45	0.82	2.74	0.49	3.77	0.64	5.31	1.24	5.19	2.22
	Group 2	0.95	0.31	0.98	0.33	0.95	0.36	2.16	1.37	2.53	0.64	3.68	0.90	5.57	1.40	6.62	1.69
	Group 3	1.03	0.33	1.16	0.47	1.09	0.26	2.38	0.97	2.77	0.85	4.01	0.71	6.96	1.22	7.99	1.08
	Overall	0.96	0.29	1.02	0.34	0.96	0.28	2.34	1.06	2.68	0.64	3.8	0.75	5.81	1.44	6.38	2.1

7.3.2. ANOVAs

A series of age-group X activity mixed model ANOVAs revealed no significant differences for GENEa output between groups across activities for the wrist monitors, however there was a significant interaction for the GENEa hip monitor ($F_{GG(5.774, 115.488)} = 2.304, p = .041$). Similarly, a group X activity interaction was observed for MET output ($F_{GG(4.965, 101.778)} = 6.0, p < .001$). Post-hoc analysis revealed the youngest group's GENEa hip output was significantly higher than the oldest group's for all walking and running activities and the middle age group for slow walk and both running activities. For MET values, the eldest age group was significantly higher than the youngest group for both running activities. However differences between all groups were observed in the medium running activity, with the eldest group recording the highest values.

7.3.3. Validity

Both GENEa wrist monitors demonstrated good criterion validity (right: $r = .900, p < 0.01$; left: $r = .910, p < 0.01$), while the hip monitor showed excellent validity ($r = .965, p < 0.01$), performing to the same standard as the ActiGraph GT1M ($r = .970, p < 0.01$). The hip-mounted GENEa had the highest concurrent validity, relative to the ActiGraph ($r = .985, p < 0.01$), but both wrist monitors also demonstrated good concurrent validity (right: $r = .830, p < 0.01$; left: $r = .845, p < 0.01$). Overall the hip GENEa demonstrated significantly higher criterion and concurrent validity than the wrist-worn monitors ($p < 0.05$).

Due to the large age range of the children, criterion and concurrent validity of the GENEa monitors were also assessed separately across the three age ranges. Table 7.2 shows the validity values for each monitor in each age group. It can be seen that criterion validity fluctuates slightly across position and age-group, with validity of the GENEa positioned at the hip consistently higher than for the GENEa worn at either wrist across all groups, however validity at the hip location was only significantly higher than at the wrist for groups one and three ($p < 0.05$). Similar results were also found for concurrent validity, with small differences being shown between groups and across monitor position. The hip

mounted GENE A reported significantly higher concurrent validity relative to the wrist-worn GENE A, across all age groups ($p < 0.05$).

Table 7.2. Criterion and concurrent validity of accelerometer output for each group

Group	Criterion validity			Concurrent validity			
	Genea right wrist	Genea left wrist	Genea hip	GT1M	Genea right wrist	Genea left wrist	Genea hip
1	0.910**	0.910**	0.970**	0.970**	0.830**	0.860**	0.990**
2	0.890**	0.880**	0.965**	0.965**	0.795**	0.800**	0.985**
3	0.900**	0.925**	0.965**	0.975**	0.845**	0.870**	0.975**

** $p < 0.001$

7.3.4. Regression analysis

Hierarchical regression analyses (presented in appendix 4.4. p. 334) showed that no additional variance in METs was accounted for by height or age ($p > 0.05$) at any wear location over and above that accounted for by GENE A output. This supports the creation of cut-points for the whole sample rather than age- or height-specific cut-points.

7.3.5. Receiver operator characteristic curve analysis

Table 7.3 shows activity intensity cut-points, established via the ROC curve analysis, for the GENE A output when monitors were worn at the right wrist, left wrist and hip, along with sensitivity and specificity values and area under the curve (AUC). Cut-points are presented as $g \cdot \text{seconds}$.

ROC analyses showed that output from the GENE A monitors at each location were able to successfully discriminate between all intensity levels. However, the hip monitor gave a slightly more precise discrimination at each intensity level than the wrist monitors (AUC = 0.94 – 0.99, compared with AUC = 0.92 - 0.97).

With regard to the different intensities, sedentary behaviour was the easiest to classify, showing the largest area under the curve for each monitor location (0.97-0.99). The wrist GENEAs had lower sensitivity and specificity values for moderate intensity than the other intensities. However, this pattern was less evident in the hip GENEa monitor.

Table 7.3. Sensitivity, specificity and Area under the curve and resultant cut-points for each GENEa monitor.

Intensity	Sensitivity %	Specificity %	Area under the ROC curve (95% CI)	Cut-points (g · seconds)
Right wrist				
Sedentary	94.85	97.71	0.97(0.95 – 0.99)	< 6
Light	N/A	N/A	N/A	6 – 21
Moderate	82.43	83.25	0.92(0.89 – 0.94)	22 – 56
Vigorous	89.36	85.48	0.93(0.91 – 0.96)	> 56
Left wrist				
Sedentary	94.74	96.67	0.97(0.96 – 0.99)	< 7
Light	N/A	N/A	N/A	7 – 19
Moderate	88.11	84	0.93(0.90 – 0.95)	20 – 60
Vigorous	91.30	89.23	0.94(0.92 – 0.97)	> 60
Hip				
Sedentary	96	96.09	0.99(0.97 – 0.99)	< 3
Light	N/A	N/A	N/A	3 – 16
Moderate	88.54	88	0.95(0.93 – 0.97)	17 – 51
Vigorous	92	88.86	0.94(0.91 – 0.96)	> 51

7.5. Discussion

The aim of the present study was to establish validity for the GENEa waveform triaxial accelerometer as a measure of children's physical activity, and calibrate the output against energy expenditure to determine activity intensity cut-points for sedentary, light, moderate and vigorous physical activity in a child population.

The expected pattern of increasing MET values corresponding with increasing accelerometer counts was observed, with only one exception for the wrist monitors in this protocol: the interactive computer games (boxing on

Nintendo Wii). This was due to the relatively fast paced movement of the wrists with little torso movement. Similarly, group differences seen with the hip-mounted monitor and the MET output can be explained by the use of the same walking speeds for each age category, and later, the self-selected running speeds; for example higher MET values for older children may have been due to the selection of faster speeds.

Although output from GENEAs at each body location demonstrated excellent criterion validity scores, there was a significant difference between the validity values for wrists and hip mounted monitors, with the hip monitor performing to the same standard as the ActiGraph, and better than the wrist worn monitors. The potential for increased acceptability and compliance that may be associated with wrist-worn monitors may compensate for the slightly lower validity scores observed. Interestingly, these results differ slightly from those found within the adult population; where results showed similar criterion reference validity across all wear locations (Esliger et al., 2011). However the correlations reported for criterion validity ($r = 0.83 - 0.87$) in the adult population were lower than those recorded in the present study. The lower concurrent validity of the wrist-worn GENE compared with the GENE worn at the hip is not surprising as the ActiGraph was also worn at the hip and monitor output differs by body location (Mathie, Coster, Lovell & Celler, 2004).

When validity was assessed by age group all monitors performed consistently across groups, with the output from the hip-mounted monitor showing comparable validity to the ActiGraph GT1M counts in each age category. It is therefore sufficient to say that the GENE demonstrates good criterion referenced and concurrent validity across each age category, whether worn at the hip or wrist. The slightly lower validity of the wrist monitors will be a small compromise if an increase in compliance is evident.

Further analysis via ROC curve showed that the GENE monitors at all three locations were able to successfully distinguish and classify sedentary, moderate and vigorous intensity activity. The hip-mounted GENE provided the most precise discrimination of each intensity level, consistently reporting the highest sensitivity, specificity and AUC. It is important to note that the cut-points are

location specific; this is highlighted by the difference between cut-points for wrist and hip locations. It was noted that the child population cut-points are much higher than those reported with an adult population. Previous research differs on whether age-specific cut-points are required in a child population (Reilly et al., 2008; Trost et al., 2002). This premise was examined in the present study; no significant differences in accelerometer output across age groups for the wrist GENEAs were found, although some differences were observed for walking and running with the hip-mounted GENE. To investigate this further, regression analyses were carried out. These results indicated that neither height nor age added significantly to the variance explained in METs by accelerometer output. Therefore, from the data presented, it appears that the cut-points established for the GENE can be used for the entire age range covered in the present study.

The present study has some limitations that should be addressed. It is possible the skew of the data towards right handed participants may have influenced the validity and cut-points slightly. Although none of the activities used the dominant hand more than the non-dominant hand, some activities may have been influenced by this bias. For example, the Nintendo Wii condition involved boxing, with both hands being used; however observations as to whether the participant favoured the use of one hand more than the other were not made. Secondly, although age and sex-specific resting $\dot{V}O_2$ values were used to calculate METs, individual resting $\dot{V}O_2$ values may have provided a more accurate view of each activity's relative intensity. However, due to the constraints of time and collecting data with this age range, obtaining a true resting metabolic rate would have been difficult.

Future studies should also consider the feasibility of the GENE for use as a measure of habitual physical activity in children, and consider whether handedness impacts upon the validity of the GENE output as a measure of habitual physical activity. The high frequency of data capture possible with the GENE facilitates the use of pattern recognition approaches to classify activity type, potentially overcoming some of the well-documented limitations associated with the use of cut-points to classify activity intensity (as discussed in chapter three and four). Nevertheless, the development of cut-points that can

be used with epoch data to determine activity intensity will allow the use of the GENE wrist-worn accelerometer while more sophisticated pattern recognition methods are being developed. When using the most recent model, the GENEActiv, it is possible to select frequencies of data collection up to 100 Hz. Measurements from different recording frequencies and epoch lengths can be compared with suitable scaling as the number of measures included in the SVMs is equal to the recording frequency multiplied by the epoch length.

7.6. Conclusion

The GENE has high criterion and concurrent validity for the classification of physical activity intensity in children, irrespective of whether it is worn on either wrist or the hip, and has been established as peer to the ActiGraph. Cut-points for use with the GENE when worn at the wrist or hip are presented to enable the classification of sedentary, light, moderate and vigorous intensity activity. The GENE therefore has the potential to be used in future child related studies and the option of wrist-wear may facilitate increased compliance rates therefore reducing the number of participants excluded, which can lead to a bias sample. Additionally, the GENE and the newer GENEActiv have the potential for 24 hours wear, allowing the examination of misclassification through different capture periods. The questions of increased compliance, capture periods and misclassification will be addressed in the following chapter, along with whether increased capture periods has any impact on detected relationships between PA and PWB variables.

CHAPTER EIGHT

24 HOUR PHYSICAL ACTIVITY MONITORING: AN EXAMINATION OF MISCLASSIFICATION AND THE IMPACT UPON RELATIONSHIPS WITH PSYCHOLOGICAL WELL-BEING

The following chapter aims to assess the acceptability of measuring children's PA for 24 hours with a wrist-worn accelerometer, and the implication of increased monitoring time upon the relationships with PWB. A recap of the importance and benefits of longer monitoring periods is provided, followed by the hypotheses of the present study. Subsequently, measurement methods and a description of the protocol used along with data analysis techniques are described. Results are reported for compliance with traditional and 24-hour wear time requirements, followed by an examination of the potential misclassification occurring from reduced sampling periods. Finally, the effect of different sample periods upon relationships between PWB and PA will be reported.

8.1. Introduction

Chapter three of this thesis highlighted the problems that can occur with objective measurement of PA, and the potential impact upon the relationships observed with health variables. Subsequently, chapters four, five and six explored the specific effect of data reduction methods upon the estimation of PA levels, and the relationship between PA and PWB. Specifically, chapter four demonstrated the effect of cut-point classification upon the relationships observed, while chapters five and six demonstrated how the issue of wear time compliance hinders the progression of research concerning associations between objectively measured PA and health outcomes. In addition to this, chapters five and six also demonstrated how methods for overcoming low compliance can in themselves lead to further problems, specifically, selection bias and misclassification of time in activity intensities. Chapter seven described

the validation of a new, wrist-worn monitor, which can potentially reduce non-compliance with wear-time requirements as a result of design features that permit longer periods of uninterrupted wear times.

8.1.1. Bias and misclassification

As demonstrated throughout this thesis, non-compliance with various wear time criteria has the potential to introduce both selection bias and misclassification to a study. Results in chapter five showed that compliers and non-compliers differed on levels of depression, weight and BMI percentiles, while chapter six demonstrated that systematic selection bias can occur between arms of an intervention study potentially altering the true effects of the intervention.

Non-compliance arising from monitor removal occurs for several widely reported reasons, particularly with waist worn devices. These include removal out of necessity, when changing clothes, sleeping and participating in some contact sports (Evenson & Terry, 2009) and for water based activities (e.g. showering or swimming). Other reasons include removal for either physical comfort or social acceptance (Crocker et al., 2001; Audrey et al., 2012). Removal of monitors for any purpose contributes to time in various activity intensities being missed, resulting in a misclassification of time in PA intensities or sedentary behaviour. This was recently examined by Ottervaere et al. (2011) who, through the use of accelerometry combined with non-wear diaries, showed an average underestimation for MVPA of approximately six minutes per day when activities with no-wear were not accounted for. Results from previous chapters of this thesis indicate that under-estimation resulting from non-wear is greatest for sedentary and light activities, although it was unclear whether this was due to selection bias or misclassification.

Misclassification can arise from the exclusion of participants failing to achieve wear time recommendations due to prolonged periods of sedentary behaviour, being falsely classified as non-wear (Rowlands et al., 2010). The possibility for misclassification of PA intensities also arises from capturing only small portions of the day (Catellier et al., 2005); a recent examination of this in adults demonstrated that larger misclassification, particularly of sedentary and light

intensity activity, was apparent with shorter monitoring periods (Herrmann et al., 2012). Similar results were observed in chapters five and six of this thesis. Additionally, misclassification can occur by failing to indicate which portions of the day are captured (e.g. Pulsford et al., in prep) as children's activity patterns vary at different periods of the day (Jago et al., 2005).

The GENEActive, described in chapter seven, and the newer GENEActive (ActivInsights Ltd. Kimbolton, Cambs, UK; www.GENEActive.co.uk), are discrete, wrist-worn, waterproof monitors; qualities which largely negate the need for monitor removal for the reasons listed above. These monitors, therefore, have the potential to increase uninterrupted periods of wear-time, including continuous 24-hour wear for consecutive days. Achieving almost continuous wear for 24 hours reduces the need for imputation methods for missing values (Montori & Guyatt, 2001; Kang et al., 2009) and reduces the risks of selection bias and misclassification of PA. Additionally, continuous wear removes the guesswork involved in distinguishing between non-wear and sedentary time. Although 24-hour continuous wear is theoretically more acceptable with waterproof, wrist-worn monitors, actual compliance with 24-hour wear is unknown in children and adults.

It is hypothesised that the use of a waterproof, wrist-worn device will lead to longer wear – time periods in children compared to commonly reported figures for waist-worn devices. This hypothesis was tested in this chapter. Following this, the study aimed to examine whether longer periods of wear resulted in higher estimated times in PA intensities compared to shorter periods of wear within the same sample. Additionally, the study aimed to explore whether the period of the day captured, matched for wear time, lead to different time accrued in PA intensities. Finally, in a sample of children with 24-hour wear for 7 days; the relationships between PA and PWB were assessed and compared with shorter monitoring periods modelled from the same sample of children. The use of the same sample of children to compare different wear times, rules out misclassification of PA due to selection bias and therefore focuses on misclassification due to variation in wear time only. From the results of chapter five, it was hypothesised that 24-hour wear time would result in higher estimations of time in all activity intensities, and that variations in time spent in

intensities would be observed for different portions of the day. Additionally, it was hypothesised that relationships between PA and PWB would be stronger with 24-hour monitoring than when only 10 hours of measurement are included.

8.2. Method

8.2.1. Participants

Participants consisted of 149 children ($n = 69$ males, $n = 80$ females), aged between 9 and 11 years ($\bar{x} = 10.0 \pm 0.7$), who had a mean (\pm SD) height of 144.3 ± 6.9 cm, and body mass of 39.2 ± 8.6 kg, seated height of 75.6 ± 3.4 cm and BMI Z-score of 0.34 ± 1.01 . The Institutional Ethics Committee granted approval for the study (10th August 2011). Parental informed written consent and the children's assent were obtained prior to data collection. (See appendix 2.5 & 2.6.p. 316 - 321)

8.2.2 Protocol

Participants for the present study were recruited via their schools. Schools within the Devon area were contacted by email to assess their interest in the study, with follow up phone calls made to those schools who did not respond within two weeks. Six of the 32 schools approached consented to be involved. Information packs and consent forms were distributed to all year five and six children within the schools ($n = 232$), of which 64% agreed to participate. Children's height (cm), seated height (cm) and body mass were measured and their BMI Z-score calculated. Participants were asked to fill in the strengths and difficulties questionnaire (SDQ; Goodman et al., 1998) and were asked to wear a GENEActiv accelerometer on their left wrist for a period of eight days, including one familiarisation day. Data collection took place between January and April 2012.

8.2.3. Measures

The SDQ (Goodman et al., 1998) consists of 25 items, measuring five subscales; emotional symptom scale (ES), hyperactivity scale (HA), conduct problem scale (CP), peer problems scale (PP) and pro-social scale (PS). Participants are asked to indicate on a zero to two scale, the extent to which an item is true for them (0 = not true; 1 = somewhat true; 2 = certainly true). Items

are worded to assess children's strengths and difficulties in each particular area, with some items being reverse scored.

Examples of items from each sub-scale are given below.

ES: 'I am often unhappy, down-hearted or tearful'

HA: 'I am restless, I cannot stay still for long'

CP: 'I usually do as I am told'

PP: 'Other people my age generally like me'

PS: 'I am kind to younger children'

Items from the ES, HA, CP and PP subscales are summed in order to obtain an indication of total psychological difficulties. Alternatively, subscales can be assessed individually. Normality values for both individual sub-scales and total difficulties values are presented in table 8.1 below. Previous studies have found the SDQ to be a reliable measure, with individual subscales and total difficulties reporting Cronbach's alpha values between 0.61 and 0.82 (Goodman et al., 1998). Alpha values for the present study ranged from 0.53 to 0.73

Table 8.1. Thresholds for classification of normal / abnormal SDQ values (adapted from Goodman et al., 1998, p.130.)

	Total Difficulties	ES	HA	CP	PP	PS
Normal	0 – 15	0 – 5	0 – 5	0 – 3	0 – 3	6 – 10
Borderline	16 – 19	6	6	4	4	5
Abnormal	20 – 40	7 – 10	7 – 10	5 – 10	5 – 10	0 – 4

The SDQ was selected to overcome the problems faced with the measures used previously. As discussed in chapter 6, many participants' PWB data became spoilt due to a lack of understanding regarding the response process, particularly for the assessment of the self-esteem hierarchy. Additionally, the length of the battery of questionnaires resulted in many questions being unanswered. The inclusion of a wide range of psychological concepts within the SDQ, addressing both maladaptive and positive states, measured with regard to positive and negative feelings (Rothenberger & Woerner, 2004) ensures that

comparability between the SDQ and those measures used in previous chapters should be attainable. The ES subscale can be related to the anxiety & depression indicators previously employed; Stone, Otten, Engles, Vermulst and Janssens (2010) reported correlations for total difficulties ($r = .73$) and emotional symptoms ($r = .63$) with the CDI; though these were for parent reported versions of the questionnaires. Muris, Meester and Van der berg (2003), also reported strong correlations between TD, ES and measures of anxiety, (r 's = 0.72 & 0.73 respectively)

Additionally, the PS scale accounts for the positive aspects of PWB. The measure does however fail to take into account feelings of physical self-perceptions, which may be of particular relevance when attempting to examine the relationships with physical activity (Fox, 1999). However the brevity and simplicity of the scale should ensure a reliable depiction of PWB. Importantly, if missing values are obtained, sub-scales scores can be prorated so long as three items for the subscale have responses, increasing completion rates.

8.2.4. Physical activity measurement

PA was assessed with the new version of the GENEActiv described in chapter six, the GENEActiv. The GENEActiv measuring 43 mm x 40 mm x 13 mm (L x W x H), is attached to a polyurethane strap, and is worn on the wrist, like a watch. It measures between +/- 8 g at a rate of up to 100 Hz and is capable of recording for 60 days with lower frequencies (10 Hz), and up to seven days at 100 Hz. The GENEActiv is comparable to the GENEActiv with the use of SVM-gravity subtracted (J. Langford, Personal communication, 6th November 2012).

8.2.5. Data Analysis

The GENEActiv data was uploaded using specific GENEActiv PC Software (Version 2.2; ActivInsights Ltd, Kimbolton, Cambs, UK). The resultant data was analysed in several ways; primarily, data was converted into 60 second epochs, and a signal vector magnitude created using the equation reported in chapter six. Subsequently, participants were assessed for compliance with required wear time criteria using Microsoft Excel 2010 Non-wear time was established

through the use of a Microsoft Excel macro (developed by ActivInsights, Kimbolton, UK), Primarily, a mean of the x, y & z axis outputs for the current (a) and the previous minute (b) is established; one is then subtracted from the other ($x_b - x_a$, $y_b - y_a$, $z_b - z_a$). Next the algorithm takes the square root of the sum of the absolute of each of the differences. This figure is then multiplied by the sum of the vector magnitudes for this minute. This prevents misclassification of repetitive activities, whereby the mean position remains constant. This is the movement measure for the current minute (a). With the use of a rolling mid-point, the average of 60 minutes of movement measures is created. This rolling mid-point is created by examining the 30 minutes prior to and the 30 minutes after the present minute, hence the 'mid-point', to create the average. The mid-point or 'present minute' then moves to the subsequent minute and again 30 minutes prior and post this minute are examined and an average created, hence the term 'rolling mid-point'. Subsequently, the algorithm takes the minimum value over 60 minutes (determined by the mid-point) of the average movement measure and applies a cut-point (value=1) to the 60 minute minimum, any minutes with a value below the cut-point value are classified as non-wear. Evidence for the validity of this algorithm is currently unavailable. The output provided by the non-wear algorithm is provided in appendix 6 (p. 337).

Despite applying a cut-point to determine wear and non-wear which can be seen as similar to applying a criteria of consecutive '0' counts, the algorithm applied to the GENEActiv data overcomes some of the problems noted previously with the consecutive '0' counts. Primarily, consecutive '0's could be misclassified as periods of non-wear when sedentary behaviour is occurring due to the same output (Rowlands et al., 2010). By creating a movement measure using the algorithm noted above, this problem can be overcome, preventing repetitive activities that record the same mean position being misclassified as non-movement or non-wear. Additionally, the use of all three axis in combination with the sum of the vector magnitudes in determining wear/non-wear creates a more sensitive determination of the presence of movement / non-movement, while the use of counts is less sensitive to small movement and positional change (personal communication, J. Langford, 22/07/2013).

Primarily, compliance with 24-hour wear and the previous recommended wear time (≥ 10 hour for three week and one weekend day; Trost et al., 2000) were obtained. Additionally, for descriptive purposes, compliance with ≥ 22 hours and ≥ 16 hours wear across a number of days is reported.

Once wear time had been established, data files were converted into 1 second epochs, and time in sedentary (not including sleep), light, moderate, vigorous and MVPA for children completing 24 hours for 7 days and ≥ 10 hours for four days were calculated using the cut-points developed in chapter seven through the use of another Excel macro (developed by ActivInsights, Kimbolton, UK). The algorithm used to extract sleep from sedentary time, based on acceleration, arm position was embedded within this macro, while with time of day and ambient light can be used to check sleep time. Validity data for this algorithm is presently unavailable.

Subsequently, only participants with complete observation were included in this study. To assess misclassification resulting from variations in wear time, three periods of 10 hour wear (8am – 6pm; 10am – 8pm & 12pm – 10pm) were extracted from participants who had recorded continuous 24-hour wear for 7 days. Data from the three time periods of each day were extracted from the 1 second epoch csv files and weekly averages of time spent in sedentary, light, moderate, vigorous and MVPA created. Repeated measures ANOVAs were then undertaken to assess whether there were significant differences in estimated time spent in each activity intensity. A Greenhouse-Geisser correction was applied where assumptions of sphericity were violated. For those demonstrating significant differences between groups, post hoc tests were undertaken with the application of a Bonferroni correction.

Pearson's correlation coefficients were used to assess relationships between complete observation, the selected 10 hour periods, and PWB variables. Only the relationships between the 8am – 6pm period are given in the results, however correlation matrices for the other 10 hour wear periods are given in appendix 4.5 (p. 335). Only those with complete observation were used so as to assess the impact of misclassification upon the observed results rather than selection bias, which was assessed in chapters five and six.

8.3. Results

8.3.1. Compliance data

Children's compliance was assessed for both the minimum recommended requirements and the full wear time of seven days at 24 hours. Figure 8.1 shows the number of files that were excluded prior to assessment and the compliance of the children remaining in the analysis. 78 children (56%) were found to have complete observation, while 131 (94%) children met the minimum recommended requirements (≥ 10 hours for three week and one weekend day). Further to this, Table 8.2 outlines the number of children achieving 24 hours, ≥ 22 hours, ≥ 16 hours and ≥ 10 hours on any numbers of days.

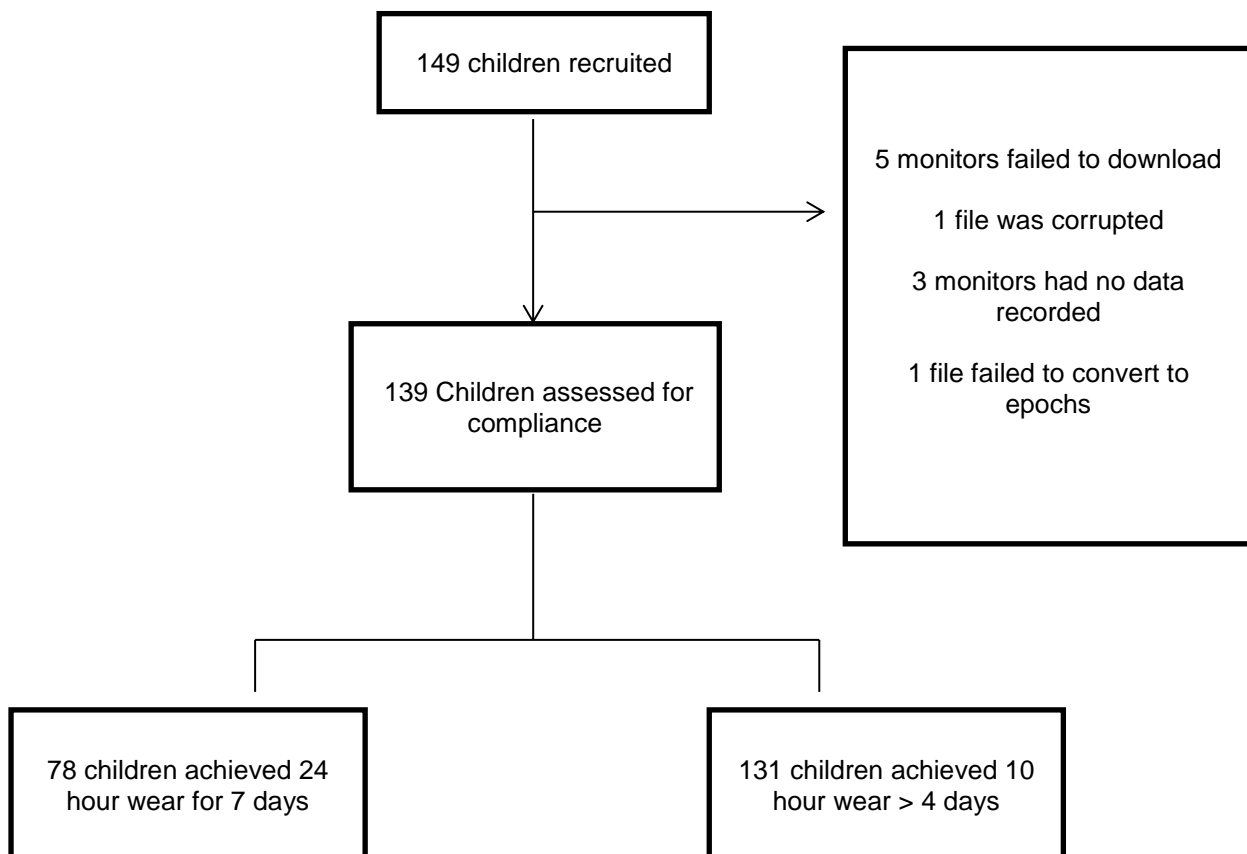


Figure 8.1. Shows the number of children recruited, those eligible to be assessed for compliance and subsequently the number of children achieving 24-hour wear for 7 days and ≥ 10 hours for three week and one weekend day.

As can be seen from table 8.2, 56.1% of children achieved complete observation, while 89.2% achieved at least 10 hours of monitoring for seven days. Interestingly, around 87% of children achieved at least 16 hours of monitoring for seven days suggesting that the majority of waking hours may be captured. Wear time increases as monitoring periods and number of monitoring days becomes more lenient, with 97% having at least one day of 24-hour monitoring.

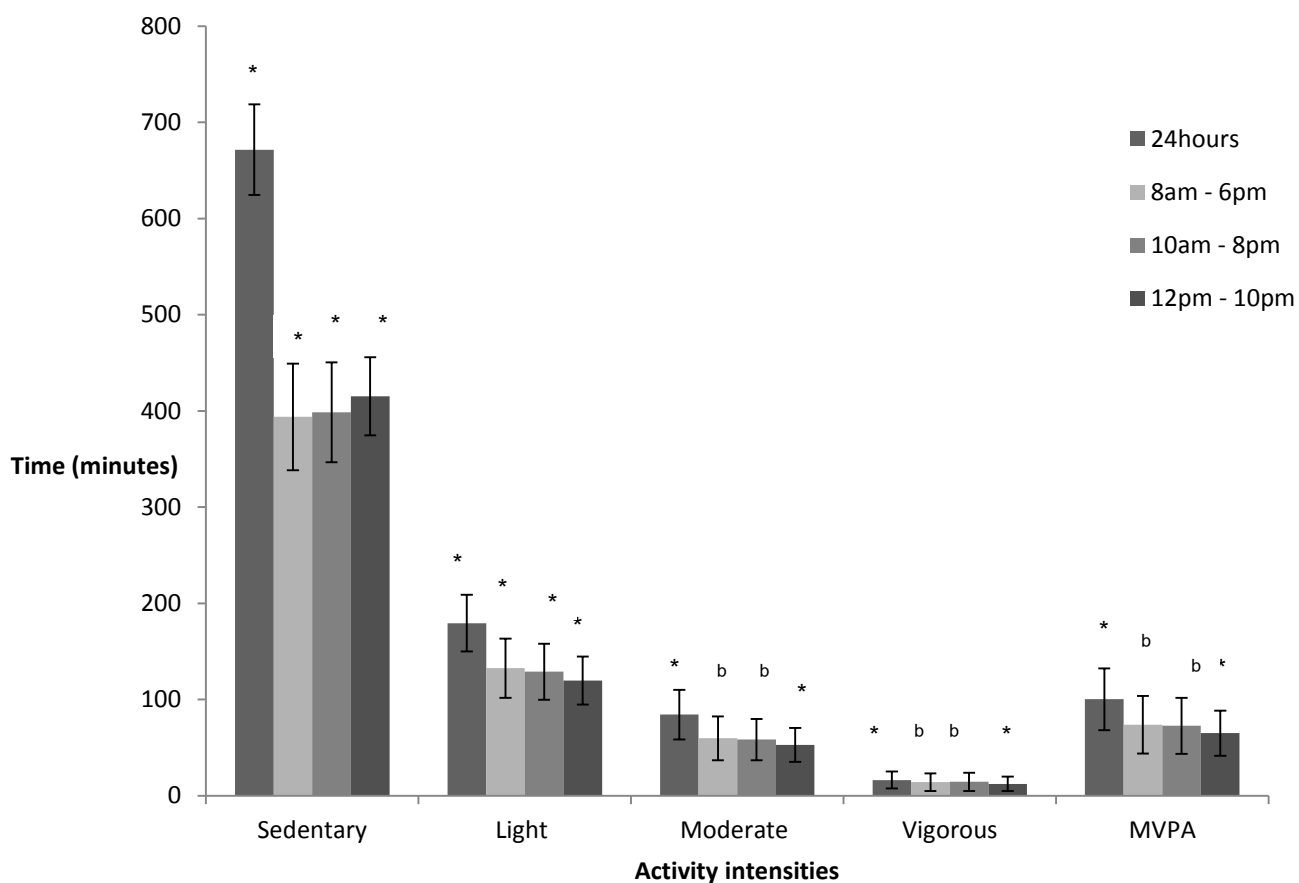
Table 8.2. Number and percentage of children complying with different wear time criteria at each day, along with cumulative percentages for the inclusion of more lenient 'day' criteria

	24-hour wear			>22 hours wear			>16 hour wear			>10 hour wear		
	N ^o . of children	%	Cum. %	No. of children	%	Cum. %	No. of children	%	Cum. %	No. of children	%	Cum. %
7 days	78	56.1	56.1	107	77	77	121	87.1	87.1	124	89.2	89.2
6 days	32	23	79.1	15	10.8	87.8	3	2.1	89.2	3	2.2	91.4
5 days	8	5.8	84.9	6	4.3	92.1	5	3.6	92.8	4	2.9	94.3
4 days	9	6.4	91.3	2	1.4	93.5	3	2.2	95	1	0.7	95
3 days	3	2.2	93.5	1	0.7	94.2	0	0	95	0	0	95
2 days	2	1.4	94.9	1	0.7	94.9	2	1.4	96.4	1	0.7	95.7
1 day	3	2.2	97.1	3	2.2	97.1	1	0.7	97.1	4	2.9	98.6
0 days	4	2.9	100	4	2.9	100	4	2.9	100	2	1.4	100
Total	139	100	100	139	100	100	139	100	100	139	100	100

8.3.2. Time in activity intensities

For those children achieving complete observation (24-hour monitoring for seven days, $n = 78$), mean (\pm SD) time spent in sedentary, light, moderate, vigorous and MVPA intensities were recorded as 671.6 (\pm 47.0) minutes, 179.3 (\pm 29.5) minutes, 84.3 (\pm 25.8) minutes, 16.26 (\pm 8.8) minutes and 100.3 (\pm 32.1) minutes, respectively. Alternatively, time accumulated in activity intensities for participants achieving ≥ 10 hours for four days ($n = 131$) were 665.0 (\pm 62.9) minutes sedentary; 177.6 (\pm 32.9) minutes in light; 81.3 (\pm 22.7) minutes in moderate; 15.8 (\pm 8.0) minutes in vigorous and 97.1 (\pm 28.3) minutes in MVPA.

Time accumulated in different intensities for complete observation and the extracted 10 hour periods for only those participants achieving complete observation are shown in figure 8.2. Time not accounted for from 24-hour monitoring was categorised as sleep time. Repeated measures ANOVA showed significant differences between wear periods for accumulated sedentary time ($F_{GG(1.356,104.402)} = 1911.366$, $p < 0.001$), differences were apparent between each wear period. Similar results were seen for light activity with differences between each of the wear periods ($F_{GG(1.261, 97.113)} = 221.784$ $p < 0.001$). Differences were also seen between time accumulated in moderate ($F_{GG(1.269,97.712)} = 100.969$, $p <.001$), vigorous ($F_{GG(1.846,142.164)} = 28.890$, $p < 0.001$) and MVPA ($F_{GG(1.328,102.293)} = 91.445$, $p < 0.001$), however for these higher intensities, differences were seen between all wear periods except between 8am – 6pm and 10am – 8pm. Table 7.3 shows the differences relative to 24-hour wear for each 10 hour period.



* Significantly different to all other wear time periods; ^b significantly different from 24 hours and 12pm – 10pm wear periods

Figure 8.2. Mean minutes per day spent in activity intensities for different periods of wear time in participants achieving complete observation

Table 8.3. Minutes difference in time accumulated in activity intensities relative to 24-hour monitoring

	Minutes difference relative to 24hour wear			
	24-hours (minutes)	8am - 6pm	10am - 8pm	12pm - 10pm
Sedentary	671.6	- 277.76	- 273.17	- 256.37
Light	179.3	- 46.78	- 50.36	- 59.56
Moderate	84.25	- 24.6	- 25.98	- 31.52
Vigorous	16.26	- 2.22	- 1.88	- 3.95
MVPA	100.31	- 26.63	- 27.66	- 35.29

8.3.3. Relationship between Physical activity and Psychological well-being

Table 8.4 displays the means and standard deviations of PWB variables for the participants achieving full wear time. Participants were assessed as to whether they were classed as 'normal', 'borderline' or 'abnormal' for each subscale of the SDQ. Percentages of children in each boundary are also presented in table 8.4.

Table 8.4 Means and standard deviations and the percentage of children classed as normal, borderline and abnormal for SDQ subscales of participants achieving full wear time (n = 78).

	ES	CP	HA	PP	PS	TD ⁵
Mean	2.70	1.96	4.42	1.90	7.99	10.98
SD	2.09	1.99	2.24	1.90	1.90	5.36
Normal (%)	92.2	83.10	67.50	84.40	88.30	84.40
Borderline (%)	2.60	10.40	11.70	5.20	6.50	9.10
Abnormal (%)	5.20	6.50	20.80	10.40	5.20	6.50

Relationships between PA and PWB were established while using different wear periods within the same sample of children (those achieving complete observation) in order to assess the impact of misclassification due to low wear time, rather than selection bias. Pearson's correlation coefficients for PA (24-hour monitoring and 8am-6pm monitoring) and PWB are shown in table 8.5. No significant correlations were observed between PA intensities and PWB for either of the wear time periods. Though for the 24-hour period, relationships between time in light intensity and peer problems approached significance ($p = .057$), as did time in vigorous and pro social behaviour ($p = .068$). These relationships did not however approach significance with the application of the 10 hour wear time. Table 8.5 shows relationships for just one of the 10 hour periods (8am – 6pm) calculated. Similar values were observed for the other 10

⁵ ES , emotional symptoms; CP, conduct problems; HA, hyperactivity; PP, peer problems; PS, pro-social; TD, total difficulties

hour periods examined; no significant relationships were apparent (see appendix 4.5 for additional correlation tables).

Table 8.5 Correlation matrix for 1 second epochs for 10 hour PA capture periods and 24 hours for 7 days with PWB variables (n =78)

	ES	CP	HA	PP	PS	TD ⁶
24 hour capture period						
Time in Sedentary	.098	.039	.104	.170	-.023	.154
Time in light	.086	.006	-.116	-.217	.124	-.088
Time in moderate	-.109	-.012	-.101	-.133	.018	-.135
Time in vigorous	-.196	.187	.107	-.066	-.208	.014
Time in MVPA	-.141	.042	-.052	-.125	-.042	-.105
10 hour capture period (8am – 6pm)						
Time in Sedentary	-.083	-.084	.013	.028	-.085	-.048
Time in light	.181	.072	-.030	-.045	.094	.068
Time in moderate	.001	.039	-.041	-.015	.149	-.008
Time in vigorous	-.116	.170	.129	.020	-.174	.078
Time in MVPA	-.035	.081	.008	-.005	.061	.017

⁶ ES, emotional symptoms; CP, conduct problems; HA, hyperactivity scale; PP, peer problems; PS, pro social scale; TD, total difficulties

8.4. Discussion

This study aimed to assess children's compliance with 24-hour accelerometer wear, and to investigate the misclassification that occurs with shorter capture periods in comparison to 24-hour wear. The main findings of this study are that 24-hour monitoring for seven days was achieved in over half of the participants, over 90% of participants achieved four days of 24-hour monitoring, while at least one day of 24-hour monitoring was achieved by 97% of participants. Very high rates of compliance (95%) with traditional wear time criteria were also evident. Misclassification occurred between complete observation and 10 hour capture periods for time spent in different PA intensities. There was no significant association between time spent at any PA intensity, based on any wear time period, and PWB measured by SDQ.

8.4.1. Compliance

The results from the present study show that 56% of participants met the full wear time criteria of 24 hours for seven days and 79% for six days. Further, 77% met the slightly more lenient criteria of ≥ 22 hours per day for 7 days. Importantly, 95% of children complied with the previously recommended wear time of ≥ 10 hours for 3 week and 1 weekend day (Troost et al., 2000).

The compliance observed with the maximum wear time criteria of seven days at 24 hours in the present study is higher than compliance previously reported for wear time requirements (e.g. Colley et al., 2010), but similar compliance rates with other required wear times have been reported. For example, Van Coevering et al. (2005) note 50% compliance with wear time required between 8am to 9pm (13 hours) on weekdays and between midday to 9pm on weekends for seven days with waist worn devices. Using the 16 hour criteria, closer to that employed by Van Coevering and colleagues, 87.1% of participants in the present study achieved seven days of monitoring. The large percentage of children achieving near optimal wear times allow for more accurate estimates of children's habitual PA and reduces the chance of misclassification.

Previous chapters within this thesis have reported low compliance rates for waist worn devices with minimum wear time requirements; chapters four and

five reported compliance of 75% for wear time of >8 hours for 3 week and 1 weekend day. Additionally, chapter five showed compliance rates as low as 7.9% with ≥ 10 hours wear for 7 days. The present study shows substantially higher compliance rates with similar groups of children, especially when considering the minimum wear time of > 10 hours; where 89% of children achieved 10 or more hours of valid wear time for 7 days, and only 5% of participants failed to achieve the minimum days of wear (3 week and 1 weekend). The compliance rates with minimum wear time in the present study are also higher than those studies using more lenient minimum wear time criteria. For example previous studies have reported 84% compliance with a minimum of any 3 days at ≥ 10 hours (Riddoch et al., 2007). The substantial increase in compliance reported in this chapter, compared to previous studies within this thesis and from the published literature, support the hypothesis presented by Rowlands and Stiles (2012) and Phillips et al. (2013) that monitor placement on the wrist will increase compliance to existing wear time protocols as well as much stricter ones. As a consequence, wrist-worn monitors can help overcome issues such as reduced statistical power, misclassification and selection bias. This also reduces the need to employ imputation methods which can further underestimate potential effects (Montori & Guyatt, 2001).

Reasons for monitor removal in the present study were mainly attributed to instruction from outside authority during contact sport such as rugby; however some participants were also instructed to remove monitors during activity such as swimming. Non-wear due to instruction by others has previously been noted as a prominent reason for removal of waist worn monitors (Crocker et al., 2001). Therefore, despite the increase in compliance from the participants, some activities, which are likely to count towards the MVPA guidelines (Department of Health, 2011), will have been missed, resulting in a potential underestimation of time in higher intensity activities.

In chapter five, it was noted that low compliance with wear-time criteria at multiple time points needs to be overcome in ways that can reduce bias and misclassification before the effect of PA upon PWB in children can be fully understood; though the present study has only addressed this matter at one

time point, high compliance due to ease of wear may overcome study attrition through non-wear (Audrey et al., 2012; Chapter five of the present thesis).

8.4.2. Time in physical activity intensities

Time in PA intensities was assessed for those achieving complete observation, and those achieving minimal wear time requirements. The results demonstrate that an increased compliance with longer wear times results in less misclassification. Specifically, PA for participants achieving the more lenient wear time was only slightly underestimated in respect to 24-hour wear. This is likely due to the majority of the children far exceeding the minimum requirement of 10 hours (e.g. 16 hour / 22 hours) as noted in table 7.2.

Time in individual PA intensities was also established for specific 10 hour periods for participants achieving complete observation. The results indicate that time in all intensities was significantly underestimated based on any 10 hour capture period. Differences between 24-hour and 10 hour capture periods were largely apparent for sedentary and light activity. Though differences were also seen at higher intensities, they were relatively small compared to sedentary and light intensity. Despite this, these slight differences may lead to important underestimates of MVPA in surveillance studies.

The results of the present study not only highlight the underestimation of PA and sedentary behaviour apparent when capturing only small portions of the day, but also show the timing of the observed period is also important. More specifically, significant differences in time accumulated in activity intensities were apparent between 10 hour capture periods, depending on the timing of the 10 hour period. Differences were apparent between all 10 hour periods for sedentary time, with the latest period (12pm – 10pm) recording the most sedentary time (415 minutes), and the lowest time for the remaining intensities. For light activity the period between 8am and 6pm recorded the most time, while differences between 8am – 6pm and 10am – 8pm were not significant for moderate, vigorous and MVPA.

The underestimation of PA and sedentary time highlighted in the present study expands upon research that has also examined misclassification. Primarily, the results agree with those of Catellier et al. (2005) who reported underestimation of time in MVPA, noting that for 12 h wear time, average time in MVPA was 159 minutes, whereas for 10 h and 8 h, time in MVPA was 147 and 148 minutes, respectively. MVPA in the present study ranged from 100 minutes (24-hour wear) to 65 minutes (12pm – 10pm). Further, the results of this study support the findings of Herrmann et al. (2012), who reported an underestimation of 135 minutes sedentary time, 94 minutes in light intensity activity, 7 minutes in moderate intensity activity and 0.5 minutes in vigorous intensity activity in simulated 10 hour wear relative to 14 hours wear.

The extent of underestimation reported in this study, between 24-hour wear time and shorter wear times, suggest that other studies, using short wear time inclusion criteria, are likely to have underestimated time in all activity intensities. For example, time spent sedentary for ≥ 10 hour wear for > 4 days at baseline in chapter five of this thesis was an average of 575 minutes per day, whereas the 24-hour monitoring in the present study showed sedentary time of 671 minutes per day, even though the children were drawn from similar sampling frames. Furthermore, previously published studies have reported time spent sedentary at only around 430 minutes per day when using wear time criteria of 3 days at ≥ 10 hours (Riddoch et al., 2007).

Time accumulated in MVPA in the present study is higher than that reported in previous studies. For example, studies examining MVPA levels in UK children report median values between 20 – 26 minutes per day, based on wear times as low as ≥ 6 hours for ≥ 3 days (Basterfield et al., 2011; 2012) and ≥ 3 days at ≥ 10 hours (Riddoch et al., 2007; Basterfield et al., 2011; 2012). The MVPA levels presented earlier in the thesis (chapter four and five), using similar waist worn devices and wear time criteria, provide support for the studies noted above. The present study used a different monitor so the absolute values are not necessarily comparable. Irrespective, the differences between the time spent in intensities derived from 24-hour wear time and those derived using 10 hour wear from the same data (100 minutes per day of MVPA compared to 65 – 74 minutes, respectively) suggests that previous findings may have

underestimated MVPA. Thus, the results of this chapter suggest that at least some of the differences in MVPA reported previously and those reported herein are due to misclassification in earlier studies. Despite this, minutes of MVPA in the present study are substantially lower than those reported in European children (Riddoch et al., 2004); 192 and 160 minutes MVPA for boys and girls respectively. However, the high quantity of MVPA recorded in these European children may largely be a result of the use of low cut-points used to classify MVPA (e.g. Trost et al., 2011; Chapter four of this thesis). The results presented in this chapter indicate that sedentary behaviour and MVPA in UK children may have previously been underestimated due to low wear times. However, comparisons with other studies also highlight the well-established effects of the differential use of data reduction techniques on prevalence estimates.

The effect of periods of the day on misclassification has not previously been explored in detail. The results highlight that misclassification does not only occur with reduced monitoring periods, but is also affected by which period of the day is captured. The results of the present study suggest that periods captured later in the day will record higher levels of sedentary time than earlier capture periods, indicating that sedentary behaviours occur more in the evenings. Studies with average wear times of 10-12 hours per day that do not consider later periods of the day may unknowingly misclassify PA and sedentary time. This in turn may lead to errors in studies of the correlates of PA that are likely to be different for children before, during and after school. The fact that all 10 hour periods underestimated time spent in each intensity compared to 24 h monitoring indicates that much longer capture periods are needed in order to estimate children's PA precisely.

As noted previously, although just over half of the participants achieved 24-hour wear for seven days, a great deal more achieved 24-hour monitoring for fewer days (e.g. six days; $n = 79\%$) or had worn the monitor for a substantial portion of the day (e.g. > 22 hours) for 7 days ($n = 76\%$). The high percentage of participants achieving prolonged periods of monitoring suggest that a wrist-worn, waterproof accelerometer can significantly reduce non-wear time and the

accompanying problems of estimating non-wear time, selection bias and misclassification of PA.

8.4.3. Physical activity and psychological well-being.

Similar to patterns seen in chapter five, the results of the present study varied depending upon the wear time requirements. The use of the same sample allows examination of the effect of misclassification alone with no effect of selection bias. Interestingly no significant relationships were apparent between time in PA intensities and PWB variables for the 24-hour wear time or for any of the 10 hour wear periods. These results support some previous findings, yet contradict others. For example, no significant relationships between higher intensities and PWB were apparent in chapter four of this thesis, yet significant relationships were apparent between time in sedentary and depression, and other variables depending upon cut-points employed. Additionally, Parfitt et al. (2009) demonstrated relationships between vigorous PA and the neurotic disorder of anxiety, a construct comparable to the emotional symptoms measured in the present study. Page et al. (2010) reported negative relationships between MVPA and emotional symptoms on the SDQ, while small associations between PA and pro-social behaviour have been noted in previous studies using objective (Sebire et al., 2011), and self-reported PA (Brodersen, Steptoe, Williamson & Wardle, 2005). Allison et al. (2005) found that vigorous activity was negatively related to problems in social functioning (i.e. more pro-social behaviour), a relationship that approached significance with the 24-hour wear. Kantomaa, Tammelin, Ebeling and Taanila (2008) reported that higher levels of MVPA are associated with fewer emotional symptoms (anxiety and depression) in boys, whilst moderate levels of MVPA may appear to be sufficient to reduce emotional symptoms in girls. The fact that no relationships were found for either wear time agrees with the findings of Johnson et al. (2008), and Hume et al. (2011) who all reported no relationships between PA and PWB, with the exception of Johnson, whose negative relationship between sedentary and depression was dismissed as a statistical anomaly by the authors. The apparent lack of significant relationships in the present study and absence of large differences in magnitude of correlation coefficients, where selection bias has been ruled out by using the same sample, suggest that

differences in associations between PA and PWB due to different wear time criteria, seen in other chapters of this thesis and other published studies, is due to selection bias rather than misclassification resulting from variations in wear time.

8.4.4 Limitations and strengths

There are limitations in the present study that should be addressed in future studies. The main limitation is the use of a convenience rather than a random sample. The schools and children agreeing to participate in the study may have been more motivated to comply with the study requirements than the majority of schools who declined. This may also mean that the range of values reported for the measure of PWB may be different from those found in a larger, more representative sample of children. Strengths of the study include the high compliance rates achieved with accelerometry data and the 100% completion rates of questionnaires. For the SDQ, no subscale totals were missing for any children ($n = 149$).

Future studies should take advantage of the 24-hour wear time achieved with a waterproof, wrist-worn device, and attempt to assess differences in time accumulated in specific portions of the day (e.g. school vs. after school), addressing their relationships with health variables such as PWB, as previous studies have attempted to address this with only limited wear and days available (e.g. Sebire et al., 2011).

8.5. Conclusion

The present study has demonstrated that low wear times resulting from non-compliance can be overcome with a wrist-worn, waterproof accelerometer designed to be worn 24-hours a day, seven days a week. The high compliance with wear can overcome issues of selection bias and misclassification that result from low wear time compliance. Even in children not managing to achieve 24-hour wear, high levels of compliance with near 24-hour wear time was achieved and misclassification reduced. The present study suggests that time spent in each activity intensity, particularly sedentary behaviour and light intensity activity, has likely previously been underestimated in UK children due

to limited wear time. As sedentary behaviour is believed to have relationships with health variables (e.g. Dunstan, Thorp & Healy, 2010; Atkin et al., 2012; Webb et al., 2013) independent to those of PA, this finding may be of particular importance. Additionally, the study shows that the equivocal results from studies examining the relationship between PA intensity and PWB may in part be due to the misclassification of PA at varying levels based on the data reduction methods employed. Now that a highly wearable accelerometer is available, more precise estimates of PA prevalence are possible, as is a better understanding of the relationship between PA intensity and health outcomes.

CHAPTER NINE

GENERAL DISCUSSION

The aim of this thesis was to explore how bias and misclassification introduced during the data reduction process with accelerometer data impact upon the observed relationship between PA and PWB in children. In order to do so, five studies were undertaken, using a variety of study designs, including cross-sectional, longitudinal, experimental and validation. The main findings of the thesis show that misclassification and selection bias vary depending on the data reduction decisions; this error significantly alters the observed relationships between PA and PWB and hinders the exploration of mechanistic properties in experimental research. Further, the findings suggest that variations in observed relationships between PA and PWB in children occur more as a result of selection bias rather than misclassification. The magnitude and direction of relationships between PA and PWB are dependent upon data reduction decisions undertaken and the compliance with those decisions, which prevents further understanding of the true relationship. Findings from the present thesis also demonstrate that selection bias and misclassification resulting from non-wear can largely be overcome with the use of a wrist-worn, waterproof device that children are willing to wear for extended periods.

Prior to discussion of the main findings of this thesis, it is important to note that variations in data reduction methods have led to considerable uncertainty about the prevalence of inactivity in children, this is particularly noticeable with cut-point classification. A number of studies have examined this (e.g. McLure et al., 2009; Stone et al., 2009; Guinhouya et al., 2013). Most recently, Guinhouya, and colleagues (2013) literature review demonstrated that between 1 and 100% of participants met the MVPA recommendations depending on the data reduction decisions regarding cut-points. Similarly, Massé et al, (2005) reported that the amount of MVPA accumulated in adults and therefore the proportion of adults meeting the recommended guidelines also varied depending upon data reduction decisions including non-wear classification, number of valid days and minimum wear criteria per day; minutes of MVPA slightly increased as more stringent criteria were employed. The authors noted that between 73% and 94%

for participants did not meet recommendations with varying data reduction decisions. Since the disease burden of physical inactivity is a function of prevalence and relative risk, it is imperative that a precise measure of inactivity is available to gain the best estimate of prevalence. An underestimate of the true prevalence of physical inactivity, likely if self-report is relied upon, may lead to a lack of appropriate resources being allocated to reducing inactivity and therefore a failure to tackle the disease burden. Further, if inaccurate estimates of physical inactivity are used, the true nature of the relationship with disease may be underestimated that in turn will lead to an underestimate of relative risks.

9.1. Effect of misclassification due to accelerometer cut-points

The first objective of this thesis was to examine the effect that choice of accelerometer cut-point for classification of activity intensity had upon the relationship between PA intensities and PWB in children. This aim was addressed in chapter four, by applying multiple published activity intensity cut-points to obtain time accumulated in each activity intensity then calculating correlation coefficients with PWB. Misclassification of time in intensities occurred across all activity intensities. As expected cut-points with the highest lower boundary for any given intensity resulted in the lowest time accrued in that intensity.

Two significant relationships between PA and PWB were observed; time spent sedentary was significantly and negatively related to depression scores for five out of the six cut-points (Treuth et al., 2004; Mattocks et al., 2007; Evenson et al., 2008; Stone et al., 2009; Freedson/Trost; r 's = -.291 to -.304) and time in light activity was significantly related to perceptions of sports competence for three out of the six cut-points (Treuth et al., 2004; Evenson et al., 2008; Stone et al., 2009; r 's = -.283 to -.300). Interestingly, relationships found to be significant with the use of some cut-points, did not significantly differ from one another yet were attenuated, becoming non-significant with the application of other published cut-points.

The inverse relationship observed between time spent sedentary and depression was contrary to previous findings noted using self-reported PA and

some using accelerometers (e.g. Parfitt et al. 2009). However, Johnson et al. (2008) and Page et al. (2010) also reported inverse relationships between sedentary time and depression. These conflicting findings provide further evidence for the notion raised at the end of chapter two - that the relationships observed depend upon the exposure measure employed. One limitation of PA measured with accelerometers is that the actual behaviour undertaken is not known. For example, time sedentary watching your favourite TV show may have a different relationship with PWB compared to time spent sedentary at school for example.

The results suggest that attenuation of observed relationships occurred with the application of cut-points that employed higher thresholds for activity intensities. For example, the use of the Puyau et al. cut-point for sedentary behaviour ($<800 \text{ counts}\cdot\text{min}^{-1}$) resulted in the relationship between sedentary time and depression becoming non-significant. However such a high threshold indicates that light activities may have been included in this relationship. Additionally, the application of Mattocks et al., Puyau et al. and Freedson/Trost equation resulted in attenuation of the relationship between light activity and sports competence; Mattocks et al., (2007) had the highest upper-bound threshold for light activity ($<3581 \text{ counts}\cdot\text{min}^{-1}$), potentially allowing moderate activities to be misclassified as light. The Puyau cut-points not only have a high lower bound threshold ($\geq 800 \text{ counts}\cdot\text{min}^{-1}$) but also a high upper bound threshold ($<3200 \text{ counts}\cdot\text{min}^{-1}$), therefore capturing less light activity and potentially more moderate activity.

Despite non-significant relationships between higher intensities and PWB variables, the effect of misclassification is still apparent, with variations in both the magnitude and direction of relationships observed. For example, a borderline significant relationship was observed between MVPA, defined by the Stone cut-points, and perceptions of physical strength ($r = .239, p = .055$), when using the Freedson/Trost cut-points to define MVPA, the magnitude was considerably reduced ($r = 0.053$).

The impact of cut-point classification of physical activity intensities demonstrated in chapter four, supports the findings of previous studies (e.g. Reilly et al., 2008; McClure et al., 2009; Stone et al., 2009; Fischer et al., 2012),

yet allowed the problem to be addressed in greater detail through the inclusion of a wider range of published cut-points and the examination of a range of activity intensities, as opposed to just MVPA. Additionally, the main findings demonstrate that the magnitude and significance of relationships between PA and PWB vary due to the differences in time recorded in activity intensities resulting from the application of different published cut-points. Chapter four supports previous studies examining the effect of cut-point classification upon PA and health (e.g. Stone et al., 2009; Atkin et al., 2013; Bailey et al., 2013) and further demonstrates these findings are applicable to relationships between PA and psychological health as well as PA and physiological health.

9.2. The effect of selection bias and misclassification as result of wear time requirements.

The second aim of this thesis was to address how rules concerning wear time requirements can result in variations in prospective relationships. Chapters five and six examined the degree of non-compliance with different wear time criteria and the selection bias and misclassification that resulted from this.

9.2.1. Compliance, selection bias and misclassification across multiple time points

Chapter five addressed the second aim of the thesis by assessing children's compliance with two popular wear time criteria (≥ 10 or ≥ 8 hours for three weekdays and one weekend day) over three time points. Selection bias occurring at each time point as a result of non-compliance was also assessed, along with the misclassification of time accumulated in PA intensities when employing more lenient wear time criteria. Compliance was higher for the more lenient criteria (> 8 hours for four days) at all time points, and a decrease in compliance was observed across all time points regardless of wear time criteria. 75% of participants complied with the eight hour criteria at baseline, decreasing to 49% at mid-point and 36.9% at end-point. For the 10 hour criteria, 50% of participants complied at baseline reducing to 35.6% and 27.4% for mid-point and end-point measurement respectively. The results of the present study concur with the previous findings that note higher compliance with more lenient

wear time (Colley et al., 2010), and a reduction in compliance over multiple wear periods (Audrey et al., 2012).

The results of chapter five reported differences between compliers and non-compliers with the 10 hour wear criteria. Compliers generally reported higher PWB at baseline and the mid-point measure than non-compliers. Theoretically, employing more lenient wear time criteria should reduce selection bias due to the inclusion of more participants, however the results of chapter five did not show this. The selection bias occurring for the 10 hour criteria was also apparent with the eight hour criteria, with further selection bias in anthropometric measures at the mid-point measurement. In addition to selection bias, an underestimation of time in activity intensities occurred when the eight hour criterion was employed. This was apparent across all measurement points, with more substantial misclassification occurring at lower intensities (i.e. sedentary and light activity intensities).

The misclassification apparent between the wear time criteria noted in chapter five supports the findings of previous studies that have noted a difference in the time accumulated at activity intensities as a result of wear time criteria (e.g. Catellier et al., 2005; Herrmann et al., 2012). The study builds on the previously published work by using estimates of PA based on data accumulated over a week, rather than simulated data created from a singular reference day, as used by Herrmann et al. (2012). Additionally, chapter five allowed an exploration of the compromise that researchers may feel they achieve by reducing wear time requirements in an attempt to increase participant numbers. In reality, employing more lenient wear time requirements not only increases misclassification, but can result in even greater differences between compliers and non-compliers in outcomes of interest, though some of this is likely to be due to the selection bias in the loss to follow up between baseline and mid-point.

The longitudinal relationship between baseline PA and follow up PWB provide partial support for previous longitudinal studies. Baseline PA (moderate, vigorous and MVPA) predicted physical self-worth at follow up, as previously reported (Raudsepp et al., 2005), yet baseline PA did not significantly predict follow up scores on any other PWB variable. Previous studies using self-report

have noted longitudinal relationships between PA and depression (e.g. Motl et al., 2004; Wiles et al., 2008; Neissaar & Raudsepp et al., 2011), though the results of chapter five support Hume et al. (2011) and Rothan et al. (2010) who reported no significant prospective relationships between PA and depression. These equivocal findings may be a result of different types of exposure measures.

In terms of the prospective relationship that was significant, the magnitude of the relationship varied depending upon the wear time criteria employed. Time accumulated in moderate intensity predicted 5.8% of the variance in PSW with the eight hour wear criteria, but predicted 14.4% of the variance when the 10 hour wear criteria were used. Similar results were apparent with MVPA; the 8 hour wear predicted 6.4% of the variance in PSW, while when the 10 hour wear criteria were used 16.4% of the variance in PSW was predicted. Time spent in vigorous activity also predicted a significant proportion of the variance in PSW (10.9%) but only when the 10 hour wear criteria were employed.

The impact of selection bias and misclassification occurring through choice of wear time rules and the resultant compliance have not previously been examined in relation to the effect upon PA and health variable relationships. Chapter five therefore makes an important contribution to the literature by demonstrating that an under-estimation of time in activity intensities and the introduction of selection bias can result in an underestimation of the longitudinal relationships between PA and PWB, resulting in some significant relationships being observed as non-significant. However, from the data presented in chapter 5 it is unclear as to whether the variation in relationships between PA and PWB arises from selection bias or through the misclassification of PA due to the capturing of shorter periods of the day.

9.2.2. Differential compliance and selection bias across trial arms.

Chapter six aimed to address the second aim of this thesis by examining whether differential compliance and selection bias were apparent across trial arms of an intervention. In order to do so, a PA intervention aimed at increasing PWB through changing PA in children was undertaken. Schools assigned to moderate intensity activity and vigorous intensity activity interventions

demonstrated consistently lower compliance across three measurement periods than was observed in the control condition, though differences were only significant for the 8 hour mid-point measurements, which occurred four weeks into the intervention. Yet compliance decreased for all conditions across subsequent time points. These results are consistent with those reported in chapter 5 and previous studies (Audrey et al., 2012).

Selection bias was seen to occur within groups at each time point, with compliers and non-compliers to wear time criteria differing on PWB and anthropometric measures across different time points. Selection bias varied for condition, time point and wear time criteria. For example, at baseline compliers to the 10 hour wear time criteria in the control condition had significantly lower depression than non-compliers from the same condition. Similar differences were not apparent for the vigorous or moderate condition for the 10 hour criteria. Additionally, this specific selection bias (for control condition) did not occur with the 8 hour wear criteria or for the mid-point measure with either criterion. It should be noted however that despite differential selection bias occurring within groups, no differences were observed between the compliers of each condition. The employment of a more lenient wear time resulted in an increased selection bias, and a misclassification of PA intensities, though the extent of the misclassification observed for baseline and at mid-point may be more reliable than observed at end-point due to the small participant numbers in each group.

The results of chapter six showed no intervention effects upon PWB, however low compliance in each group at follow up resulted in an inability to undertake mediation analysis in order to determine whether the intervention had altered the proposed mechanism (change in PA). Due to this, the differential selection bias apparent within groups and the effect of misclassification of time in activity intensities upon the causal relationship could not be tested statistically. It appears that an exploration of the differential compliance, selection bias and misclassification that occurs between trial arms of an intervention has not previously been undertaken with respect to objective PA measurement. This study therefore provides important information for researchers wishing to undertake trials using objective measurement;

indicating that trials should be powered to account not only for loss to follow up but also loss to non-compliance. Further, researchers should be aware of the potential bias introduced during completers only analysis with physical activity data that cannot be controlled for by randomisation (Gaus & Mauche, 2013).

9.3. Validation and calibration of a wrist-worn accelerometer.

The next aim of this thesis was to validate and calibrate a wrist-worn accelerometer that has the capability to overcome the issues previously reported in earlier chapters of this thesis including low compliance and thus selection bias and misclassification of time in PA intensities. This aim was addressed in chapter seven by establishing concurrent validity against the ActiGraph GT1M, criterion validity against indirect calorimetry and creating cut-points using ROC curve analysis for GENE A accelerometers placed at the wrist and hip.

The results demonstrated good concurrent validity for the new GENE A accelerometer against the ActiGraph GT1M, reporting strong correlation coefficients for monitors at the hip and at each wrist ($r = .795$ to $.990$). Additionally, monitors worn at all locations demonstrated high criterion validity relative to indirect calorimetry ($r = .880$ to $.970$). ROC curve analysis showed high classification accuracy for each activity intensity at all locations (AUC = 0.92 to 0.99). Monitors positioned at the wrist showed slightly lower validity and classification accuracy than the waist-worn monitors. However, the potential for increased compliance may outweigh the slightly lower validity apparent with the wrist-worn monitors.

The classification accuracy of the wrist worn GENE A monitors was higher than observed in other studies using different wrist-worn accelerometers (Wockets) with adult populations (Rosenberger et al., 2013). Rosenberger and colleagues noted AUC values of 0.73 , with sensitivity and specificity values of 89% and 32% for a wrist mounted monitor when discriminating sedentary behaviour from physical activity. The authors noted more precise discrimination of behaviour for the waist worn Wocket; similar to that noted in chapter seven with the GENE A. High classification accuracy has also been reported for the

GENEA with other populations (Esliger et al., 2011; Van Hees et al., 2011; Zhang et al., 2012).

Concerns with wrist worn physical activity monitoring have recently been noted. For example, Rosenberger and colleagues (2013) suggested that misclassification of activities may occur due to a disproportional increase in wrist motion compared to an increase in energy expenditure. The potential for this is larger during sedentary behaviours, where some participants gesticulate while others remain still, possibly resulting in a misclassification of sedentary activities into those of a higher intensity (Rosenberger et al., 2013). This type of disparity between wrist motion and energy expenditure was observed in chapter seven, where participants were asked to undertake a boxing activity on the Nintendo wii (Nintendo, Windsor, UK); this activity involved relatively fast paced movement of the wrists with little torso movement, resulting in a high output for the monitors mounted at the wrist rather than the hip (left wrist = 48.48; right wrist = 50.25; waist = 7.03 g-seconds) and a low MET value (2.34). Additionally, it was noted by Rosenberger and colleagues (2013) that wrists may become constrained from free movement during ambulatory activities (e.g. carrying bags), resulting in lower accelerometer output in comparison to the energy expenditure required. However, these problems may potentially be overcome by examining the effects of monitor placement on the dominant or non-dominant wrist, as the dominant wrist is more likely to become constrained.

This chapter contributes to the wider literature by establishing the validity of an accelerometer that may be used to gain increased compliance and longer periods of wear in children. Despite the concerns noted with the use of wrist-worn accelerometers, good validity and high classification accuracy have been demonstrated by the GENE A placed at the wrist. The benefits of higher compliance which may accompany wrist placement, compared to waist worn devices, more than compensates for the slightly reduced precision by significantly reducing the selection bias and misclassification resulting from non-compliance. Additionally, the creation of cut-points that can be used with epoch data to determine activity intensity will allow the use of the GENE A wrist-worn accelerometer while more sophisticated pattern recognition methods are being developed; the use of pattern recognition techniques may overcome the issues

with cut-points previously noted in chapters three and four. Discussion of pattern recognition is provided later in this chapter.

9.4. Misclassification through shorter capture periods

The final aim of the thesis was to assess the acceptability of 24-hour PA monitoring and address the effect of PA misclassification upon the PA and PWB relationship independently from the effect of selection bias. This was addressed in chapter eight, by extracting three different 10 hour periods of monitoring from children who had achieved 24-hour monitoring for seven days, and establishing relationships between time accumulated at each PA intensity in each of the 10 hour periods and PWB, then repeating this with time accumulated from the 24 hour monitoring. The three periods of 10-hours also allowed examination of misclassification occurring between different periods of the day even when the same minimum wear-time was used.

The main findings show that 24-hour monitoring for seven days is acceptable in children, with 56% of participants achieving complete observation. Over 90% of children achieved four days of 24-hour monitoring, while 77% of children achieved >22hours monitoring for 7 days, demonstrating higher compliance than previously noted with waist worn devices both within this thesis and the wider literature (Crocker et al., 2001; Van Coevering et al., 2005; Page et al., 2010; Colley et al., 2011). Recently, a database designed to encompass large amounts of children's accelerometer data, the International Children's Accelerometry Database (ICAD), was established. The ICAD has a total of 45,190 accelerometer files available for analysis from studies undertaken in Europe, USA and South America. Assessment of wear time showed that 49% of boys and 51% of girls aged 9 to 11 had at least 12 hours of accelerometer data for more than four days (Sherar et al., 2011); 91% of participants wore the GENEActiv for 24 hours a day for four days, demonstrating the increased compliance achieved with the wrist-worn GENEActiv and thus the ability to acquire substantially longer capture periods. These compliance rates also appear higher than has recently been reported with other wrist-based devices, which have shown 70 – 80% compliance with six or more day monitoring, with a median wear time of 21 – 22 hours per day (Freedson & John, 2013). For a similar criteria (> 22 hours for six days), compliance in chapter eight was 87.8%

Importantly, the results highlighted the substantial misclassification that occurred for time spent in PA intensities established from 10 hour wear periods over 7 days compared to time in PA intensities established from 24-hour wear (average over 7 days). In addition, time in PA intensities accumulated in 10 hour periods from different portions of the day resulted in variations in time recorded in each intensity. The greatest underestimation occurred for sedentary time and time in light intensity activity. It appears from the results that more sedentary time is accumulated later in the day, with mean differences of 278 minutes between 24 hours and time accumulated in 8am – 6pm but only 256 minutes between 24 hours and 12pm – 10pm. Conversely, the majority of MVPA is accumulated during earlier periods of the day; capture periods from 8am to 6pm and 10am to 8pm showed no significant differences between time accumulated in MVPA (73 minutes & 72 minutes, respectively), but significantly fewer minutes of MVPA were recorded with a later capture period (12pm – 10pm; 65 minutes). These results support the findings of previous studies that have noted how children's patterns of activity vary across the day with the majority of high intensity activity accumulated during the late afternoon (3pm – 7pm) (Jago et al., 2005) and, further, have highlighted the extent to which misclassification may occur by capturing different portions of the day (Catellier et al., 2005; Pulsford et al., In prep).

Finally, chapter eight showed no significant relationships between PA and PWB variables with either 24-hour capture or 10 hour capture periods, however the magnitude of relationships varied slightly depending upon the capture period employed; stronger relationships were generally apparent between PA captured in the 24-hour period and PWB than were noted for PA captured in the 10 hour period. This study extends the wider literature by specifically examining the misclassification apparent between different capture periods compared to complete observation. Though previous studies have noted misclassification of time in activity intensities due to short wear periods (e.g. Catellier et al., 2005; Herrman et al., 2012), differences were established between shorter wear periods in which misclassification may still occur. For example, 8, 10 and 12 hours were compared to 14 hour reference days in Herrmann et al. (2012) and comparison was made between 12, 10 and 8 hours noted in the Catellier et al. (2005) study. The examination of underestimation between usual recommended

wear times (10 hours) and complete observation (24 hours) allows for a more detailed depiction of the misclassification problem. Further, the findings of the study indicate that PA and PWB are not as susceptible to change by misclassification, as they may be to selection bias indicated through earlier studies in this thesis.

9.5. Overall summary

This thesis has provided a detailed exploration of the error associated with decisions regarding the data reduction process necessary with accelerometry, and the impact that this error has upon estimates of the prevalence of PA and the relationship between PA and PWB. Non-compliance with various wear time criteria introduces selection bias that alters estimates of PA prevalence and measures of PWB. Reducing wear times, to try and improve compliance, means that varying degrees of the day are not observed and therefore misclassification of time spent in different activity intensities occurs. Trying to disentangle the magnitude of the effect of selection bias and misclassification due to non-wear is problematic as any comparisons between different wear times means different samples. Even when the sample is restricted to children with maximum wear (24 hours and 7 consecutive days) and other wear times are modelled from the same sample, the sample is likely to have a reduced variation in exposure and outcome measures compared to all children who provided any data. This restricts the potential to detect associations. High precision of estimates of PA is essential in fully understanding relationships with PWB but higher precision requires extended periods of wear. With waist worn devices, high wear time criteria leads to non-compliance that in turn introduces considerable selection bias. To counter the selection bias researchers are tempted to reduce wear time requirements yet the penalty is greater misclassification. Both selection bias and misclassification have an impact on associations between PA and health.

The results of the present thesis are limited in determining the true nature of the relationship between PA and PWB, as equivocal results are presented throughout. Results showed both support for and evidence against relationships previously reported for studies using both self-report (e.g. Motl et al., 2004; Ussher et al., 2007; Goldfield et al., 2011; Sund et al., 2011) and objective measures

(e.g. Parfitt et al., 2005; Johnson et al., 2008; Parfitt et al., 2011; Page et al., 2010; Hume et al., 2011). As described above, it appears that the observed relationships are heavily influenced by the choice of exposure measure and the data processing decisions that accompany the use of objective measures such as accelerometry. Achieving high compliance with precise, objective measures of PA will allow researchers to get closer to the 'true' relationship between PA and PWB that to date has eluded them. In addition, the ability to determine specific behaviours via pattern recognition, as opposed to intensity and volume, may further add to the understanding of PA and PWB relationships. As mentioned previously, some of the differences in the findings of existing studies may be due to the behaviour specific relationship between PA and well-being.

9.6. Strengths and limitations

This thesis has explored in detail the influence of the data reduction processes necessary with accelerometry on measures of PA. The use of accelerometers in physical activity research is widespread, yet relatively few studies acknowledge that data reduction procedures impact the results they report. Even fewer studies explore, or are able to explore, the exact nature of the effect that these decisions have. The main strength of this thesis is that the exploration of these issues has indicated the extent to which they can impact upon results. The misclassification that occurs with different wear time requirements, capture periods and cut point use has been explicitly examined along with the exploration of selection bias, showing that these decisions will directly affect the accumulation of time in activity intensities and therefore the conclusions given in any study. The implications of this indicate that researchers should no longer merely acknowledge the potential impact, but should undertake sensitivity analysis. This would involve all studies applying different data reduction techniques and reporting the subsequent effect on results. This would make for a more transparent reporting of results and permit better comparisons between studies.

The results of this thesis not only inform researchers looking to establish relationships with health variables, but also those undertaking population surveillance, assessing the effectiveness of interventions and examining temporal trends in PA (Rennie & Wareham, 1998). The examination of selection

bias and misclassification prior to assessing the impact upon relationships indicates that any studies aiming to use accelerometers should take note of the misclassification and bias occurring through a variety of data reduction decisions, rather than continuing to focus purely upon misclassification through cut-point selection. In addition to this, the influence of data reduction processes has been considered in cross-sectional, longitudinal and experimental designs, a strength of this thesis. The results demonstrated that decisions made may not only alter relationships at a single time point, but changes in selection bias occur across time points. This can result in variations in observed relationships depending upon the measurement point used in the subsequent analysis.

The ability to explore selection bias and misclassification both jointly and independently is a further strength of this thesis. Misclassification can often occur either as a result of selection bias, or as a result of shorter monitoring periods, and period of the day captured. The examination of the two concepts independently enabled the conclusion that selection bias probably has a greater impact on results than misclassification due to short observation periods.

Following this, the thesis highlighted the importance of decisions made about capture periods for PA monitoring, a concept that has previously been noted (Catellier et al., 2005) but has received little attention, perhaps due to difficulties in obtaining samples with maximum wear time. That can now largely be overcome as 24 hour monitoring is feasible with waterproof, wrist-worn monitors (Chapter eight). Future studies should carefully consider monitoring periods as well as minimum wear time when undertaking PA measurement.

Despite the ability to determine the effects of data processing decisions upon the observed relationship with PWB, a number of limitations are apparent. Primarily, this thesis has only addressed a few of the decisions necessary when using accelerometry data. The effect of epochs and determination of non-wear have not been assessed.

9.6.1. Effect of Epochs

As noted in chapter three, epochs vary between one second and one minute. Despite short epochs being recommended for research with children (Ward et

al., 2005) due to the highly transitory nature of children's activity (Bailey et al., 1995), some monitors are incapable of high frequency monitoring due to limited battery and memory capacity. Although memory limitations that led to the use of large epochs have largely been solved, some studies still employ larger epochs than necessary (e.g. Huberty et al., 2011; Van Cauwenberge et al., 2011). It may be that increased misclassification is apparent with the use of longer epochs, with time in PA intensities being under-estimated (Troost et al., 2005; Rowlands 2007), especially higher intensities which occur in shorter bouts (Nilsson et al., 2002; Vale et al., 2009). This underestimation may in turn alter the effects of the observed relationships between PA and PWB. For example, if time in MVPA was underestimated due to a longer epochs, the relationship observed in chapter five (MVPA predicted PSW) may have been attenuated, becoming non-significant. However results from this thesis have demonstrated that misclassification alone may have limited impact upon PA and health relationships.

9.6.2. Non-wear determination

Throughout chapters four, five and six, non-wear was determined as 30 minutes of consecutive '0' counts based on the findings of Rowlands et al., (2010). Determination of non-wear can have a direct effect on whether a day is classed as valid or not, therefore alteration of non-wear definitions will influence the rate of compliance with different wear time requirements, in turn altering the sample included and contributing to the selection bias within each sample. Using shorter non-wear periods, such as the 20 minutes recommended by Esliger et al. (2005), could have resulted in lower compliance for both \geq eight hours wear and ≥ 10 hours wear in chapters four, five and six of this thesis. This alteration may have changed the selection bias and in turn the relationships observed. The use of a shorter non-wear period may have also resulted in underestimation of time in sedentary activity; sedentary behaviours occurring for periods of 20 minutes would be classed as non-wear (Rowlands et al., 2010).

Additionally, longer periods of non-wear such as the 180 minutes of '0' counts employed by Van Coevering et al. (2005) are likely to result in an overestimation of sedentary time, greater compliance with wear time requirements, and potentially a reduction in selection bias. This may have

resulted in smaller variations in the relationships observed between PA and PWB in chapter five and would possibly have allowed for examination of the mechanisms in chapter six. Using such long periods to determine non-wear may have also prevented the systematic differences for compliance across trial conditions in chapter six. Despite this, the decision to use the 30 minutes non-wear criteria consistently across most studies allowed comparability between chapters of this thesis, and was based on the most biologically plausible values as determined by Rowlands et al. (2010). The exception to this was the non-wear determination of the GENEActiv, which due to software limitations was based on a 60 minute criteria.

Future studies employing devices such as the ActiGraph or accelerometers requiring similar decisions, may want to consider the effect of non-wear periods upon compliance, selection bias and misclassification and as a result the observed relationship between PA and health variables. The algorithm used to detect non-wear in chapter eight with the GENEActiv accelerometer has yet to be validated. Multiplying the square root of the sum of the absolute of each of the differences between current and previous epochs by the sum of the vector magnitudes prevents misclassification of repetitive activities as periods of non-movement (Personal communication with J.Langford; 12.04.2013). This overcomes the problems with determination of non-wear noted above. Validation of the non-wear algorithm is required prior to further assessment of compliance with the GENEActiv.

9.6.3. Differential effect of sex

In addition to further data reduction decisions, a limitation of this thesis is that the impact of sex upon relationships was not considered. During chapter two, it was noted that many studies examined the relationship between PA and PWB separately for sex; as boys generally report more PA than girls (Sallis et al., 2000; Pate et al., 2002; Riddoch et al., 2007; Van der Hosrt et al., 2007) and some PWB variables show higher prevalence in girls (Green et al., 2005). The apparent sex differences were not examined in the present study due to relatively small sample sizes ($n = 49$, Chapter four; $n = 76$ chapter five). The examination of sex differences within chapter six especially would have been impractical, due to the already small samples in each arm of the trial across

each time point. Although chapter eight contained a larger sample ($n = 149$), the influence of sex was beyond the scope of this chapter, rather the study aimed to assess the acceptability of longer monitoring periods and the effect of misclassification on relationships across all those who had achieved complete observation. Differential effects may have been observed by splitting the sample by sex which may have impacted on the conclusions. It is recommended that future studies examine whether selection bias and misclassification systematically differ between males and females.

9.6.4. Sample size

A further limitation that was mainly noticeable in chapter six of this thesis is the sample size recruited. Though the number of participants was adequate to detect differences according to the power calculation performed, the low numbers of participants complying with accelerometer wear time criteria meant that examination of end-point PA data could not be undertaken. Following this, it was not possible to undertake mediation analysis to determine whether the proposed mechanism to increase PWB had been altered. In addition, only three schools were recruited to the study, the inclusion of more schools and in turn increased participant numbers would likely have enabled the examination of end-point data and mediation analysis to take place. Finally, the study described in chapter six, was not randomised, therefore differences between intervention groups could have been apparent, though controlling for baseline measures when undertaking the analysis should have negated any baseline differences. Future studies should therefore be powered sufficiently to detect differences and account for loss of participants through non-compliance

9.7. Future directions

This thesis has demonstrated the problems of misclassification and selection bias that are inherent in the measurement of PA using accelerometers when less than maximum compliance is achieved. Some recommendations for future research have already been provided, including the use of sensitivity analysis to highlight the effect of data reduction decisions, examination of additional data reduction decisions (e.g. epoch length and non-wear criteria), consideration of monitoring periods, investigation of whether compliance and selection bias

differs between boys and girls, and exploration of monitor placement on dominant / non-dominant wrist. However a number of other directions should be considered in an attempt to obtain precise PA measurement.

9.7.1. Imputation methods

It is clear that longer monitoring periods of accurate PA monitoring are available with the use of wrist-worn monitors such as the GENEActiv, indicating that future studies should endeavour to employ these or similar, wrist-worn, waterproof monitors. The increased compliance with such monitors decreases the possibility of misclassification of activity intensities through longer capture periods, and decreases selection bias through increased compliance and reductions in excluded participants. Despite these benefits, PA monitoring with these monitors is not flawless; further calibration and validity studies are required to refine estimates of the energy costs of different physical activities in a range of populations.

Despite substantial increases in compliance with current, common wear time criteria, it is still the case that only 56% of participants achieved 24 hour monitoring for 7 days. While this number is encouraging compared to previous levels of compliance, there remains a deficit between ideal (complete observation of all participants) and actual compliance. Therefore, there will always be some gaps in the recorded data that may lead to at least some misclassification and selection bias. One way to retain high wear time criteria and reduce missing data is to use imputation techniques for missing data (Kang et al., 2009).

Imputation is a relatively new development in the area of PA measurement, with few studies utilising the available methods. The simplest method of imputation involves using the mean counts from either the existing group or remaining individual data of the same day (e.g. Monday) or the same type of day (e.g. weekend) to fill in the missing data points (Kang et al., 2009). A variety of imputation methods exist, yet as very few studies have employed them, their value remains relatively uncertain. It may be argued that imputation techniques can be used to overcome the low compliance apparent with waist worn devices, however considering that imputation methods utilise the observed activity

values to predict missing values (Catellier et al., 2005), it follows that more accurate estimates of PA can be obtained with fewer missing data points and shorter missing time periods (Lee, 2013), a luxury that is available through wrist worn monitoring. In addition, combining information from 'invalid' day data with imputed data can result in less biased estimates of PA (Lee, 2013). For example, 77% of participants in chapter eight achieved 7 days of >22 hour monitoring; considering that non-wear may occur at different periods of the day, the availability of large amounts of data for each day allows for more accurate imputation, than if participants only had data 10 hours for 4 days.

Future studies should therefore combine data obtained from wrist-worn monitors with imputation methods in order to obtain complete observation for all participants, providing highly accurate estimates of PA due to the longer periods of monitoring available prior to imputation. This combination would reduce the possibility for misclassification and eradicate selection bias occurring through non-compliance, in turn allowing for more accurate assessment of the true nature of relationships between PA and PWB. The accuracy of these methods could be validated using complete observation samples, by removing segments of data to simulate non-wear, similar to previous studies that have explored imputation with more lenient wear time (Catellier et al., 2005; Kang et al., 2009; Lee et al., 2013).

In addition, the use of imputation methods would allow further examination of selection bias upon PA estimates, as those children with incomplete PA observation may undertake less PA (Catellier et al., 2005). This could only be accurately examined with the use of imputation to obtain accurate PA estimates for the non-compliers. However, limitations may be present with imputation techniques, for example when missing data does not occur at random, the estimates at the population level may be biased (Catellier et al., 2005), the possibility to test this in real data collection scenarios however is limited (Lee, 2013).

9.7.2. Raw data and pattern recognition

The use of monitors such as the GENEActiv and the ActiGraph GT3X+ allow some data reduction decisions to be bypassed as an output of pre-filtered raw acceleration data is available, though decisions about wear time are still required. The use of raw acceleration data can overcome some of the well-documented limitations associated with cut-points of counts·minutes⁻¹ to classify activity intensity (as discussed in chapter three and four). Using pre-filtered data removes the chance of misclassification through the use of cut-points, and allows for greater comparability between studies, as the output is given in the 'SI unit' of m·s⁻² or g (1g = 9.81m·s⁻²).

In addition to the availability of raw data, the high frequency of data capture possible with the GENEActiv, and other high sampling rate accelerometers facilitates the use of pattern recognition approaches to classify activity type rather than just movement. It may be possible therefore to establish relationships between particular behaviours and health variables such as PWB. For example, through pattern recognition, it may be possible to note that walking at 4 km·h⁻¹ for 20 minutes per day can reduce feelings of depression. Whereas running at 10 km·h⁻¹ for 30 minutes per day is related to increased perceptions of body attractiveness. The exploration of pattern recognition technology may also overcome some of the potential pitfalls expressed by Rosenberg et al. (2013) associated with wrist-worn accelerometry previously discussed. Pattern recognition technology is in its infancy, but several studies have examined the determination of domains, or specific behaviours from both epoch count data (Poerber, Staudenmayer, Raphael & Freedson, 2006; Staudenmayer, Poerber, Crouter, Bassett & Freedson, 2009) and raw acceleration data (e.g. Bonomi, Goris, Yin & Westerp, 2009; Oshima et al., 2010; Zhang, Rowlands, Murray & Hurst, 2012). Advancement of such methods may improve PA estimates in free-living environments (Freedson & John, 2013) and the ability to determine relationships with health variables.

9.8. Practical implications

Despite the need for future studies to further advance PA monitoring to aid the establishment of true PA and health relationships and explore dose-response relationships, the results presented in this thesis have a number of practical implications for researchers, policy makes and clinicians.

9.8.1. Implications for researchers

Firstly, researchers should select wrist worn, waterproof accelerometers rather than waist worn, non-waterproof accelerometers to maximise wear time compliance. When waist worn accelerometers are used, researchers should undertake some form of sensitivity analysis to explore how their data reduction decisions influence the relationship observed. In particular, researchers should consider how each of the following can result in variations in observed relationships.

- Misclassification as a result of activity intensity cut-points
- Selection bias as a result of valid day criteria,
- Misclassification as a result of valid day criteria,
- Selection bias that systematically differs across trial arms of an intervention,
- Misclassification as a result of non-standardised periods of the day used for data collection.

A further important implication of this thesis is that studies looking to assess the effectiveness of an intervention should be powered to account for loss of data due to non-compliance as well as loss to follow up, while also being aware of systematic differences in compliance and selection bias occurring between arms of a trial that can influence the conclusions drawn, again this may be a more pertinent issue for researchers employing waist based devices.

9.8.2 Implications for policy makers

The findings of this thesis raise important issues and practical implications that those creating policies regarding physical activity should take into consideration. In particular, objective physical activity measurement can provide a more precise estimate of the true population prevalence of physical inactivity. This will give a better estimate of the magnitude of the problem of inactivity that at present, based on self-report, is considerably underestimated. This in turn will inform decisions regarding the appropriate amount of resources required to tackle the problem. This thesis has demonstrated that gaining accurate assessment of PA in children does not come about purely through the use of objective measurement. Rather, the findings presented have highlighted that policy makers should consider the data reduction decisions used within accelerometry prior to creating recommendations regarding PA.

Similarly to the implications for researchers, the following should be taken into account with regard to PA policy.

- Length and specific period of measurement can result in variations of prevalence estimates.
- Data reduction decisions resulting in selection bias can result in misclassification of estimates of the population prevalence of physical activity and sedentary time.
- Variations in intensity boundaries (cut-points) between studies results in limited comparability and reduce the external validity of population surveillance.
- Studies employing wrist-worn, waterproof devices are likely to provide more precise estimates of PA due to less misclassification and selection bias as a result of increased compliance and extended wear time.

9.8.3 Implications for clinicians

Clinicians may wish to measure patients' physical activity to guide advice on disease prevention, aid in the diagnosis of disease, or monitor progress in the management of disease. In all cases, precise measurement is required to avoid misclassification and diagnosis as well as inappropriate advice. This thesis has

shown that very low levels of physical activity, the main risk factor for disease, are most likely to be underestimated as a result of misclassification due to low wear time, it is imperative to achieve high wear time for correct classification of activity. Furthermore, it is likely that self reported physical activity in a clinical setting will be more subject to social desirability bias compared to less personalised assessments such as population surveillance.

Therefore, clinicians should:

- Measure physical activity objectively
- Select wear times that capture all waking hours for at least 7 days.

According to the results of this thesis this requires:

- A waterproof, wrist worn accelerometer
- An accelerometer that has sufficient battery life to accommodate 24 hours a day and a minimum of 7 days continuous wear.

9.9. Conclusion

This thesis has highlighted how data reduction decisions necessary with accelerometry data can impact on estimates of PA prevalence and the observed relationship between PA and PWB, and thus can hinder the progression of research through inaccurate estimates of activity and selection bias as a result of non-compliance. Though this thesis addresses psychological well-being, the principles examined are likely to affect the observed relationship with other health variables. Overall this thesis has shown that misclassification through cut-points, wear time requirements and capture periods can alter the magnitude of observed relationships. Selection bias is of greater importance, and appears to affect the observed temporal relationships between PA and PWB more, than misclassification of time in activity intensities. This thesis has shown it is possible to overcome these issues through the increased compliance with extended wear periods associated with the use of wrist-worn accelerometry, though further research is needed in order to refine minor problems that accompany wrist-based monitoring, such as the potential constraints of the wrist during movement. Until such a time that complete observation of PA is likely for all study participants, there remains a risk of at

least some misclassification and selection bias as a result of missing data, though the magnitude of this risk is significantly reduced with wrist-worn devices. Nevertheless, the true nature of the relationship between PA and PWB remains unknown; conclusions drawn regarding these variables appear to be relevant only to the methodology reported in each study. This thesis has made progress in identifying threats to internal validity and taking steps to overcome them and thus has made a significant contribution to accelerometer-based PA research.

References

- Adams, S.A., Matthews, C.E., Ebbeling, C.B., Moore, C.G., Cunningham, J.E., Fulton, J., & Herbert, J.R. (2005). The effect of social desirability and social approval on self-reports of physical activity. *American Journal of Epidemiology*, 161 (4), 389 – 398. doi: 10.1093/aje/kwi054
- Adeniyi, A.f., Okafor, N.C., & Andeniyi, C.Y. (2011). Depression and physical activity in a sample of Nigerian adolescents: Levels, relationships and predictors. *Child and Adolescent Psychiatry and Mental Health*, 5 (16). doi: 10.1186/1753-2000-5-16
- Ahn, S., & Fedewa, A.L. (2011). A meta-analysis of the relationship between children's physical activity and mental health. *Journal of Pediatric Psychology*, 36(4), 385 – 397. doi: 10.1093/jpepsy/jsq107
- Ainsworth, B.E., Haskell, W.L., Herrmann, S.D., Meckes, N., Bassett, D.R., Tudor Locke, C., Greer, J.L., Vezina, J., Whitt- Glover, M.C. & Leon, A.S., (2011) Compendium of Physical Activities: a second update of codes and MET values. *Medicine and Science in Sports and Exercise*, 43(8),1575-1581.
- Ainsworth, B.E., Haskell, W.L., Whitt, M.C., Irwin, M.L., Swartz, A.M., Strath, S.J., O'Brien, W.L., Bassett, D.R., Schmitz, K.H., Emplainscourt, P.O., Jacobs, D.R., & Leon, A.S. (2000). Compendium of Physical Activities: An update of activity codes and MET intensities. *Medicine and Science in Sports and Exercise*, 32 (Suppl),S498-S516.
- Aldana, S.G., Sutton, L.D., Jacobson, B.H., & Quirk, M.G. (1996). Relationships between Leisure time physical activity and perceived stress. *Perceptual and Motor Skills*, 82, 315 – 321. doi: 10.2466/pms.1996.82.1.315
- Alhassan, S., Lyden, K., Howe, C., Kozey-Keadle, S., Nwaokemeleh, O., & Freedson, P.S. (2012). Accuracy of accelerometer regression models in predicting energy expenditure and METs in children and youth. *Pediatric Exercise Science*, 24, 519 – 536.
- Allison, K.R., Adlaf, E.M., Irving, H.M., Hatch, J.L., Smith, T.F., Dwyer, J.J. & Goodman, J. (2005). Relationship of vigorous physical activity to psychological distress among adolescents. *Journal of Adolescent Health*, 37 (2), 164 - 166. doi:10.1016/j.jadohealth.2004.08.017
- Alonso, A., Seguí – Gómez, M., Irala, J.D., Sánchez-Villegas, A., Beunza, J.J., & Martínez – Gonzalez, M.A. (2006). Predictors of follow-up and assessment of selection bias from dropouts using inverse probability weighting in a cohort of university graduates. *European Journal of Epidemiology*, 21, 351 – 358. doi:10.1007/s10654-006-9008-y
- American Psychiatric Association (1994). *Diagnosis and statistics manual of mental disorders, Fourth edition*. Washington, DC.

- Anderson, L.B., Harro, M., Sardinha, L.B., Froberg, K., Ekelund, U., Brage, S., & Anderssen, A.A. (2006), Physical activity and clustered cardiovascular risk in children: a cross-sectional study (The European Youth Heart Study). *The Lancet*, 368 (9532), 299 – 304.
- Armstrong, H. (1998). Young people's physical activity patterns as assessed by heart rate monitoring. *Journal of Sport Science*, 16, s9 – s16.
- Armstrong, N. & Welsman, J. (1997). *Young People and Physical Activity*. Oxford, UK : Oxford University Press.
- Atkin, A.J., Ekelund, U., MØller, N.C., Froberg, K., Sardinha, L.B., Anderson, L.B., & Brage, S. (2013). Sedentary time in children: influence of accelerometer processing on health relations. *Medicine and Science in Sport and Exercise*, epub ahead of print. doi: 10.1249/MSS.0b013e318282190e
- Atkinson, G., Davison, R.C.R. & Nevill, A.M (2005). Performance characteristics of gas analysis systems: what we know and what we need to know. *International Journal of Sports Medicine*, 26, s2 – s10. doi: 10.1055/s-2004-830505
- Audrey, S., Bell, S., Hughes, R., & Campbell, R. (2012). Adolescent perspectives on wearing accelerometers to measure physical activity in population-based trials. *European Journal of Public Health*, doi:10.1093/eurpub/cks081
- Bailey, D.P., Boddy, I.M., Savory, L.A., Denton, S.J., & Kerr, C.J. (2013). Choice of activity-intensity classification thresholds impacts upon accelerometer-assessed physical activity-health relationships in children. *PLoS One*, 8 (2), e57101. doi:10.1371/journal.pone.0057101
- Bailey, R.C., Olson, J., Pepper, S.L., Porszasz, J., Barstow, T.J., & Cooper, D.M. (1995). The level and tempo of children's physical activities: an observational study. *Medicine and Science in Sports and Exercise*, 27, 1033-1041.
- Bandura, A. (1977). Self- Efficacy: Toward a unifying theory of behaviour change. *Psychological Review*, 84 , 191 – 215.
- Baquet, G., Stratton, G., Van Praagh, W., & Berthoin, S. (2007). Improving physical activity assessment in prepubertal children with high-frequency accelerometry monitoring: A methodological issue. *Preventive Medicine*, 44(2),143 – 147. <http://dx.doi.org/10.1016/j.ypmed.2006.10.004>
- Baranowski, T., Masse, L.C., Ragan, B., & Welk, G. (2008). How many days was that? We're Still Not Sure, But We're Asking the Question Better! *Medicine and Science in Sport and Exercise*, 40 (7suppl), s544 – s 459, doi:10.1249/MSS.0b013e31817c6651
- Basset, D.R., & Strath, S.J. (2002). Use of Pedometers to assess physical activity. In G.Welk, *Physical Activity Assessments for Health – Related Research* (pp. 164 – 175). Champagne, IL: Human kinetics.

- Basterfield, L., Adamson, A.J., Parkinson, K.N., Maute, U., Li, P.X., & Reilly, J.J. (2008). Surveillance of physical activity in the UK is Flawed: Validation of the Health survey for England physical activity questionnaire. *Archives of Disease in Childhood*, 93, 1054 – 1058. doi:10.1136/adc.2007.135905
- Basterfield, L., Pearce, M.S., Adamson, A.J., Fray, J.K., Parkinson, K.N., Wright, C.M., & Reilly, J.J. (2012). Physical Activity, Sedentary Behavior, and Adiposity in English Children. *American Journal of Preventive Medicine*, 42(5), 445 – 451. doi: 10.1016/j.amepre.2012.01.007
- Bauman, A., & Merom, D. (2002). Measuring and surveillance of physical activity in Australia-an introductory guide, *Australian Epidemiologist*, 5.
- Bauman, A., Phongsavan, P., Schoeppe, S., & Owen, N. (2006). Physical activity measurement – a primer for health promotion. *Promotion and Education*, 13 (2), 92 – 103. doi: 10.1177/10253823060130020103
- Baumeister, R.F., Campbell, J.D., Krueger, J.I., & Vohs, K.D. (2003). Does High self- esteem cause better performance, interpersonal success, happiness or healthier lifestyles? *Psychological Science in the Public Interest*, 4(1), 1 – 44. doi: 10.1111/1529-1006.01431
- Baumeister, R.F., Smart, L., & Boden, J.M. (1996). Relation of threatened egotism to violence and aggression: the dark side of high self-esteem. *Psychological Review*, 103 (1), 5 – 33. doi: 10.1037/0033-295X.103.1.5
- Bech P.(1998). *Quality of life in the psychiatric patient*. London: Mosby-Wolfe.
- Bech P. (2001). Male depression: stress and aggression as pathways to major depression. In A. Dawson & A. Tylee (eds.) *Depression: Social and economic timebomb*. London: BMJ Books. p.63-66.
- Bech, P., Olsen, L.R., Kjoller, M., & Rasmussen, N.K. (2003). Measuring well-being rather than the absence of distress symptoms: a comparison of the SF-36 mental health subscale and the WHO-Five Well-being scale. *International Journal of Methods in Psychiatric research*, 12 (2), 85 – 90.
- Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561 – 571.
- Beets, M.W., Bornstein, D., Beighle, A., Cardinal, B.J., & Morgan, C.F. (2010). Pedometer-measured physical activity patterns of youth, A 13-country review. *American Journal of Preventive Medicine*, 38 (2), 208 – 216. doi: 10.1016/j.amepre.2009.09.045
- Begin, L., Budniok, T., Vaksman, G., Boussard-Delbecque, L., Michaud, L., Turck, D., & Gottrand, F. (2000). Simplification of the method of assessing daily and nightly energy expenditure in children, using heart rate monitoring calibrated against open circuit indirect calorimetry. *Clinical Nutrition*, 19 (6), 425 – 435. doi:10.1054/clnu.2000.0147

- Belmaker, R.H., & Agam, G. (2008). Major Depressive disorder. *New England Journal of Medicine*, 358(1), 55 – 68. doi: 10.1056/NEJMra073096
- Berggren, G., & Christensen, E.H. (1950). Heart rate and body temperature as indices of metabolic rate during work. *Arbeitsphysiologie*, 14, 255 – 260.
- Biddle, S. (2000). Emotion, mood and physical activity. In S. Biddle., K.Fox., & S.H. Boutcher. (Ed). *Physical Activity and Psychological Well Being*. (pp. 63 – 87). London, Uk: Routledge.
- Biddle, S., & Asare, M. (2011). Physical activity and mental health in children and adolescents: a review of reviews. *British Journal of Sports Medicine*, 45 (11), 886 – 895. doi: 10.1136/bjsports-2011-090185
- Biddle, S.J., Gorely, T., Pearson, N., & Bull, F.C. (2011). An assessment of self-reported physical activity instruments in young people for population surveillance:Project ALPHA. *International Journal of Behavioural Nutrition and Physical Activity*, 2, 8 – 1. doi: 10.1186/1479-5868-8-1
- Biddle, S., & Mutrie, N. (2008). *Psychology of physical activity: Determinants, Well-being and interventions* (2nd Ed.) London, UK: Routledge.
- Birkeland, M., Melkevik, O., Holsen, I., & Wold, B. (2012). Trajectories of global self- esteem development during adolescence. *Journal of Adolescence*, 35 (1), 43 – 54. doi: 10.1016/j.adolescence.2011.06.006
- Birkeland, M., Torsheim, T., & Wold, B. (2009). A longitudinal study of the relationship between leisure time physical activity and depressed mood among adolescents. *Psychology of Sport and Exercise*, 10, 25 – 34. doi 10.1016/j.psychsport.2008.01.005
- Blair, A., Stewart, P., Lubin, J.H., & Forastiere, F. (2007). Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. *American Journal of Industrial Medicine*, 50(3), 199 – 207. doi: 10.1002/ajim.20281
- Blumenthal, J.A., Michael, A., & Babyak, A., (2007). Exercise and pharmacotherapy in the treatment of major depressive disorder. *Psychosomatic Medicine*, 69, 587 – 596.
- Bonhauser, M., Fernandez, G., Püschel, K., Yanez, F., Montero, J., Thompson, B., & Conronado, G. (2005). Improving physical fitness and emotional well-being in adolescents of low socioeconomic status in Chile: results of a school based control trial. *Health Promotion International*, 20 (2), 113-122. doi: 10.1093/heapro/dah603
- Bonomi, A.G., Goris, A.H., Yin, B., & Westerterp, K.R. (2009). Detection of type, duration and intensity of physical activity using an accelerometer. *Medicine and Science in Sport and Exercise*, 41 (9), 1770 – 1777.

- Bouchard, C., Tremblay, A., Leblanc, C., Lortie, G., Savard, R., & Theriault, G. (1983). A method to assess energy expenditure in children and adults. *The American Journal of Clinical Nutrition*, 37, 461 – 467.
- Bradburn, N. (1969). *The Structure of Psychological well-being*. Chicago, IL: Aldine.
- Brage, S., Wedderkopp, N., Franks, Anderson, L., & Froberg, K. (2003). Reexamination of validity and reliability of the CSA monitor in walking and running. *Medicine and Science in Sport and Exercise*, 35 (8), 1447 – 1454.
- Brenner, H., & Blettner, M. (1993). Misclassification bias arising from random error in exposure measurement: implications for dual measurement strategies. *American Journal of Epidemiology*, 138 (6), 453 – 461.
- Brodersen, N. H., Steptoe, A. Williamson, S., & Wardle, J. (2005). Sociodemographic, developmental, environmental, and psychological correlates of physical activity and sedentary behavior at age 11 to 12. *Annals of Behavioural Medicine*, 29(1), 2 – 11. doi:10.1207/s15324796abm2901_2
- Brosnahan, J., Steffen, L.M., Lytle, L., Patterson, J., & Boostrom, A. (2004). The relation between physical activity and mental health among Hispanic and non-hispanic white adolescents. *Archives of Pediatrics & Adolescent Medicine*, 158 (8), 818 – 823. doi:10.1001/archpedi.158.8.818
- Brown, J., Kreiger, N., Darlington, G.A., & Sloan, M. (2001). Misclassification of exposure: Coffee as a surrogate for caffeine intake. *American Journal of Epidemiology*, 153 (8), 815 – 820. doi: 10.1093/aje/153.8.815
- Brown, T.A., O'Leary, T.A., & Barlow, D.H. (2001). Generalized anxiety disorder. In D.H. Barlow (Ed.), *Clinical handbook of psychological disorders: A step-by-step treatment manual* (3rd ed.). New York: Guilford Press.
- Brown, H.E., Pearson, N., Braithwaite, R.E., Brown, W.J., & Biddle, S.J.H. (2013). Physical activity interventions and depression in children and adolescents. *Sports Medicine*, 43, 195 – 206. doi: 10.1007/s40279-012-0015-8
- Brown, W.H., Pfeiffer, K., Mclver, K.L., Dowda, M., Almeida, J., & Pate, R. (2006). Assessing preschool physical activity: An observational system for recording physical activity in children-Preschool version (OSRAC-P). *Research Quarterly for Exercise and Sport Science*, 77, 167 – 176.
- Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*, 84 (4), 822-848. doi: 10.1037/0022-3514.84.4.822
- Brown, W.J., Trost, S.G., Bauman, A., Mummery, K., & Owen, N. (2004). Test-retest reliability of four physical activity measures used in population surveys. *Journal of Science and Medicine in Sport*, 7 (2), 205 – 215.

- Bubbar, V.K., & Kreder, H.J. (2006). The Intention-to-Treat Principle: A Primer for the Orthopaedic Surgeon. *The Journal of Bone and Joint Surgery*, 88 (9), 2097 – 2099. doi: 10.2106/JBJS.F.00240.top
- Burdette, H.L., Whitake, R.C., & Daniels, S.R. (2004). Parental report of outdoor playtime as a measure of physical activity in preschool-aged children. *Archives of Pediatric and Adolescent Medicine*, 158 (4), 353 – 357.
- Burnet, J., Sabastian, C.M., Chaiton, M., Barnett, T., O'Loughlin, E., Low, N.C.P., & O'Loughlin, J.L. (2013). The associations between past and current physical activity and depressive symptoms in young adults: a 10 year prospective study. *Annals of Epidemiology*, 23 (1), 25 – 30. doi: 10.1016/j.annepidem.2012.10.006
- Butte, N.F., Ekelund, U., & Westerp, K.R. (2012). Assessing physical activity using wearable monitors: measures of physical activity. *Medicine and Science in Sport and Exercise*, 44 (1suppl), s5 – s12., doi: 10.1249/mss.0b013e3182399c0e
- Byrne, N.M., Hills, A.P., Hunter, G.R., Weinsier, R.L., & Schutz, Y. (2005). Metabolic equivalent: one size does not fit all. *Journal of Applied Physiology*, 99, 1112 – 1119. doi: 10.1152/jappphysiol.00023.2004
- Calfas, K.J., & Taylor, W.C. (1994). Effects of physical activity on psychological variables in adolescents. *Pediatric Exercise Science*, 6, 406 – 423.
- Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait- multimethod matrix. *Psychological Bulletin*, 56, 81 – 105.
- Cao, H., Qian, Q., Weng, T., Yuan, C., Sun, Y., Wang, H., & Tao, F. (2011). Screen time, physical activity and mental health among urban adolescents in China. *Preventive Medicine*, 53 (4-5), 316 – 320. doi:10.1016/j.ypmed.2011.09.002
- Catellier, D.J., Hannan, P.J., Murray, D.M., Addy, C.L., Conway, T.L., Yang, S., & Rice, J.C. (2005). Imputation of missing data when measuring physical activity by accelerometry. *Medicine and Science in Sport and Exercise*, 37 (11s), s555 – s562. doi:10.1249/01.mss.0000185651.59486.4e
- Carter, J., & Jeukendrup, A.E. (2002). Validity and reliability of three commercially available breath-by-breath respiratory systems. *European Journal of Applied Physiology*, 86, 435 – 441.
- Carvajal, S.C., Wiatrek, D.E., Evans, R.I., Knee, C.R., & Nash, S.G. (2000). Psychosocial determinants of the onset and escalation of smoking: cross-sectional and prospective findings in the multi-ethnic middle school samples. *Journal of Adolescent Health*, 27 (4), 255 – 265.
- Caspersen, C.J., Powell, K.E., & Christenson, G.M. (1985) Physical activity, exercise and physical fitness: definitions and distinctions for health related research. *Public Health Reports*, 100, 126 -131.

- Celis-Morales, C.A., Perez-Bravo, F., Ibanez, L., Salas, C., Bailey, M.E., & Gill, J.M. (2012). Objective vs. self-reported physical activity and sedentary time: effects of measurement method on relationships with risk biomarkers. *PLoS One*, 7(5), e36345. doi: 10.1371/journal.pone.0036345
- Chamberlain, J.M., & Hagga, D.A.F. (2001). Unconditional self-acceptance and psychological health. *Journal of Rational-Emotive & Cognitive-Behaviour Therapy*, 19 (3), 163 – 176.
- Chapman, C.N., Love, E., Burgess, S., & Lahav, M. (2010). Consumer usage of a digital memory device prototype. *Proceedings of the 43rd Hawaii International Conference on System Sciences*. Accessed via <http://research.microsoft.com/enus/um/cambridge/projects/sensecam/publications.htm>
- Chalder, M., Wiles, N.J., Campbell, J., Hollinghurst, S.P., Haase, A.M., Taylor, A.H., Fox, K.R...Lewis, G. (2012). Facilitated physical activity as a treatment for depressed adults: randomised controlled trial. *British Medical Journal*, 334; e2758. doi: 10.1136/bmj.e2758
- Chaumeton, N., Duncan, S.C., Duncan, T.E., & Strycker, L.A. (2011). A measurement model of youth physical activity using pedometer and self, parent and peer reports. *International Journal of Behavioural Medicine*, 18 (3), 209 – 215. doi: 10.1007/s12529-010-9118-5
- Chen, K.Y. & Bassett, D.R. (2005). The technology of accelerometry-based activity monitors: current and future. *Medicine and Science in Sport and Exercise*, 37(11suppl), s490 – s500.
- Chen, K.Y., Janz, K.F., Zhu, W., & Brychta, R.J. (2012). Re-defining the roles of sensors in objective physical activity monitoring. *Medicine and Science in Sport and Exercise*, 44 (s1), s13 – s23. doi:10.1249/MSS.0b013e3182399bc8
- Chu, E.Y., McManus, A.M., & Yu, C.C. (2007). Calibration of the RT3 accelerometer for ambulation and non-ambulation in children. *Medicine and Science in Sport and Exercise*, 39 (1), 2085 – 2091. doi: 10.1249/mss.0b013e318148436c
- Cohen, J. (1992). A power primer. *Quantitative Methods in Psychology*, 112 (1), 155 – 159.
- Cole, T.J., Faith, M.S., Pietrobelli, A., & Heo, M. (2005). What is the best measure of adiposity change in growing children: BMI, BMI %, BMI z score or BMI centile? *European Journal of Clinical Nutrition*, 59, 419 – 425. doi:10.1038/sj.ejcn.1602090
- Colley, R., Gorber, S.C., & Tremblay, M. (2010). Quality control and data reduction procedures for accelerometry-derived measures of physical activity. *Health Reports*, 21, 63 – 64.

- Collings, P.J., Brage, S., Ridgeway, C.L., Harvey, N.C., Godfrey, K.M., Inskip, H.M., Cooper, C., Wareham, N.J., & Ekelund, U. (2013). Physical activity intensity, sedentary time and body composition in pre-schoolers. *American Journal of Clinical Nutrition*, 97, 1020 – 1028. doi: 10.3945/ajcn.112.045088
- Cooper, C., & Bebbington, P. (2006). Mental health. In M. Bajekal, V. Osborne, M. Yar, H. Meltzer (Eds). *Focus on Health*, Hampshire, Uk: Palgrave Macmillan.
- Conn, V.S. (2010). Anxiety outcomes after physical activity interventions: Meta-analysis findings. *Nursing Research*, 59 (3), 224 – 231. doi: 10.1097/NNR.0b013e3181dbb2f8
- Corder, K., Ekelund, U., Steele, R.M., Wareham, N.J., & Brage, S.(2008). Assessment of physical activity in youth. *Journal of Applied Physiology*, 105 (3), 977 – 987. doi: 10.1152/jappphysiol.00094.2008
- Craft, L.L., & Landers, D.M. (1998). The effect of exercise on clinical depression and depression resulting from mental illness: a meta-analysis. *Journal of Sport and Exercise Psychology*, 20, 339 – 357.
- Craft, L.L., & Perna, F.M. (2004). The benefits of exercise for the clinically depressed. *Primary Care Companion to the Journal of Clinical Psychiatry*, 6, 104 – 111.
- Crocker, P.R., Bailey, D.A., Faulkner, R.A., Kowalski, K.C., & McGrath, R. (1997). Measuring general levels of physical activity: preliminary evidence for the physical activity questionnaire for older children. *Medicine and Science in Sport and Exercise*, 29 (10), 1334 – 1349.
- Crocker, P., Eklund, R., & Kowalski, K. (2000). Children's Physical activity and physical self-perceptions. *Journal of Sport Science*, 18, 383 – 394.
- Crocker, P., Holowachuk, D., & Kowalski, K. (2001). Feasibility of using the tritrac motion sensor over a 7 day trial with older children. *Pedatric Exercise Science*, 13, 70 – 81.
- Crocker, J., & Park, L.E. (2004). The costly pursuit of self-esteem. *Psychological Bulletin*, 130 (3), 392 – 414. doi: 10.1037/0033-2909.130.3.392
- Crocker, J., & Knight, K. M. (2005). Contingencies of self-worth. *Current Directions in Psychological Science*, 14 (4), 200-203. DOI: 10.1111/j.0963-7214.2005.00364.x
- Crocker, P.R.E., Sabiston, C.M., Kowalski, K.C., McDonough, M.H., & Kowalski, N. (2006). Longitudinal assessment of the relationships between Physical self concept and health-related behaviour and emotion in adolescent girls. *Journal of Applied Sport Psychology*, 18, 185 – 200. Doi: 10.1080/10413200600830257
- Crown, D.P., & Marlow, D. (1964). *The Approval Motive*. New York: Wiley.

- Cuddihy, T., Michaud-Tomson, L., Jones, E.K., & Johnston, A.O. (2006). Exploring the relationship between daily steps, Body mass index and Physical self-esteem in female Australian Adolescents. *Journal of Exercise and Fitness*, 4(1), 25 – 35.
- Dale, D., Corbin, C.B., & Dale, K.S. (2000). Restricting opportunities to be active during school time: do children compensate by increasing physical activity levels after school?. *Research Quarterly for Exercise & Sport*, 71 (3), 240 – 248. <http://dx.doi.org/10.1080/02701367.2000.10608904>
- Dale, D., Welk, G., & Matthews, C.E. (2002). Methods for assessing physical activity and challenges for research. In. G.Welk, *Physical Activity Assessments for Health – Related Research* (p. 19 – 33). Champagne,IL: Human kinetics
- Daley, A.J., Copeland, R.J., Wright, N.P., Roalfe, A., & Wales, J.K.H. (2006). Exercise therapy as a treatment for psychopathologic conditions in obese and morbidly obese adolescents: A randomised control trial. *Pediatrics*, 118, 2126. doi: 10.1542/peds.2006-1285
- Deci, E.L. & Ryan, R.M. (2008). Hedonia, Eudaimonia, and well-being: An introduction. *Journal of Happiness Studies*, 9, 1 – 11. doi: 10.1007/s10902-006-9018-1
- Department of Health (2011). Start active, Stay Active, A report on physical activity for health from the four home countries' Chief Medical Officers. <http://www.bhfactive.org.uk/userfiles/Documents/startactivestayactive.pdf>
- Deverill, C., & King, M. (2009). Common mental disorders. In S.McManus, H.Meltzer, T. Brugha, P. Bebbington, R.Jenkins. *Adult psychiatric morbidity in England, 2007: Results of a household survey*. The NHS information centre for Health and social care. ([www.ic.nhs.uk/pubs/psychiatric morbidity07](http://www.ic.nhs.uk/pubs/psychiatric_morbidity07)).
- Diener, E. (2009). Subjective well-being. In E. Diener (Ed.), *The Science of Well-being*. New York: Spring.
- Diener, E., Scollon, C.N., & Lucas, R.E. (2003). The evolving concept of Subjective well-being: the multifaceted nature of happiness. *Advances in cell Aging and Gerontology*, 15, 187 – 219
- Dishman, R.K., Hales,D.P., Pfeiffer, K.A., Felton, G.A., Saunders,R., Ward, D.S., Dowda, M., Pate, R.R. (2006). Physical self-concept and self-esteem mediate cross-sectional relations of physical activity and sport participation with depression symptoms among adolescent girls. *Health Psychology*, 25 (3), 396 – 407. doi: 10.1037/0278-6133.25.3.396
- Dodge, R., Daly, A.P., Huyton, J. & Sanders, L.D. (2012). The challenge of defining well-being. *Journal of Well-Being*, 2 (3), 222-235. doi:10.5502/ijw.v2i3.4

- Dollman, J., Okely, A.D., Hardy, L., Timperio, A., Salmon, J., & Hills, A.P. (2009). A hitchhiker's guide to assessing young people's physical activity: Deciding what method to use. *Journal of Science and Medicine in Sport*, 12, 518 – 525. doi: 10.1016/j.jsams.2008.09.007
- Donaghy, M., & Taylor, A.H. (2010). Should practitioners promote physical activity as a treatment for depression? *Journal for the Royal College of Physicians of Edinburgh*, 40, 132 – 135. doi:10.4997/JRCPE.2010.223
- Dopp, R.R., Mooney, A.J., Armitage, R., & King, C. (2012). Exercise for adolescents with depressive disorders: A feasibility Study. *Depression Research and Treatment*. doi: 10.1155/2012/257472.
- Dorsey, K.B., Herrin, J., & Krumholz, H.M. (2011) Patterns of moderate and vigorous physical activity in obese and overweight compared with non-overweight children. *International Journal of Pediatric Obesity*, 6 (2-2), e547-55. doi: 10.3109/17477166.2010.490586
- Doty, D.H., & Glick, W.H. (1998). Common methods bias: does common methods variance really bias results? *Organizational Research Methods*, 1(4), 374 – 406. doi: 10.1177/109442819814002
- Douglas, C.G. (1911). A method for determining the total respiratory exchange in man. *Journal of Physiology*, 42, 17 – 18.
- Duncan, M.J., Al-Nakeeb, Y., Nevill, A.M., & Jones, M.V. (2006). Body dissatisfaction, body fat and physical activity in British children. *International Journal of Pediatric Obesity*, 1(2), 89 – 95. doi: 10.1080/17477160600569420
- Dunn, A.L., Trivedi, M.H., Kampert, J.B., Clark, C.G., & Chambliss, H.O. (2005). Exercise treatment for depression: efficacy and does response. *American Journal of Preventive Medicine*, 28, 1 – 8. doi:10.1016/j.amepre.2004.09.003
- Dunstan, D.W., Thorp, A.A., & Healy, G.N. (2011). Prolonged sitting: is it a distinct coronary heart disease risk factor? *Current Opinion in Cardiology*, 26 (5), 412 – 9. doi: 10.1097/HCO.0b013e3283496605
- Ebata, A.T., & Moos, R.H. (1991). Coping and adjustment in distressed and healthy adolescents. *Journal of Applied Developmental Psychology*, 12, 33 – 54.
- Edwards, S.D., Ngcobo, H.S.B., Edwards, D.J., & Palavar, K. (2005). Exploring the relationship between physical activity, psychological well-being and physical self-perceptions in different exercise groups. *South African Journal for Research in Sport, Physical Education and Recreation*, 27 (1), 75 – 90.
- Egger, M., Davey-Smith, G., & Phillips, A.N. (1997). Meta-analysis principles and procedures. *British Medical Journal*, 315 (7121), 1533 – 1537.

- Ekeland, E., Heian, F., & Hagen, K.B. (2005). Can exercise improve self-esteem in children and young people? A systematic review of randomised controlled trials. *British Journal of Sports Medicine*, 39, 792 – 798. doi: 10.1136/bjism.2004.017707
- Ekelund, U., Aman, J., & Westerterp, K. (2003). Is the ArteACC index a valid indicator of free-living physical activity in adolescents? *Obesity Research*, 11 (6), 793 – 801.
- Ellis, A. (1994). *Reason and Emotion in psychotherapy, A comprehensive method of treating human Disturbances*. New York, NY: Birch Lane Press.
- Epstein, L.H., Paluch, R.A., Kalakanis, L.E., Goldfield, G.S., Cerny, F.J., & Roemmich, J.N. (2001). How much activity do youth get? a quantitative review of heart-rate measured activity. *Pediatrics*, 108;e44. doi: 10.1542/peds.108.3.e44
- Esliger, D.W., Copeland, J.L., Barnes, J.D., & Tremblay, M.S. (2005). Standardizing and optimizing the use of accelerometer data for free-living physical activity monitoring. *Journal of Physical Activity and Health*, 2 (3), 366 - 383.
- Esliger, D.W., Rowlands, A.V., Hurst, T.L., Catt, M., & Eston, R.G. (2011). Validation of the GENE accelerometer. *Medicine and Science in Sport and Exercise*, 43 (6), 1085 – 1093.
- Esliger, D.W., & Tremblay, M.S. (2007). Physical activity and inactivity profiling: the next generation. *Applied Physiology, Nutrition and Metabolism*, 32 (suppl E2), s195 – s207.
- Eston, R.G., Rowlands, A.V., & Ingledew, D.K. (1998). Validity of heart rate, pedometry and accelerometry for predicting the energy cost of children's activities. *Journal of Applied Physiology*, 84, 362 – 371.
- Evenson, K.R., Catellier, D.J., Gill, K., Ondark, K.S., & McMurray, R.G. (2008). Calibration of two objective measures of physical activity for children. *Journal of Sport Sciences*, 26 (14), 1557 – 1565. doi: 10.1080/02640410802334196
- Fairclough, S.J., Boddy, L.M., Ridgers, N.D., & Stratton, G. (2012). Weight status associations with physical activity intensity and physical self-perceptions in 10- to 11- year – old children. *Pediatric Exercise Science*, 24 (1), 100 – 112.
- Field, A. (2009). *Discovering Statistics using SPSS (3rd ed)*. London; Sage Publications.
- Finch, A.J., Montgomery, L.E., & Deardorff, P.A. (1974). Reliability of the state-trait anxiety with emotionally disturbed children. *Journal of Abnormal Child Psychology*, 2, 67 – 69.

- First, M.B., Spitzer, R.L., Gibbon, M., & Williams, J.B.W. (1995). The structured clinical interview for DSM-III-R Personality disorders (SCID-II).Part 1: Description. *Journal of Personality Disorders*, 9 (2), 83 – 9. doi: 10.1521/pedi.1995.9.2.83
- Fisher, C., Yildirim, M., Salmon, J., & Chinapaw, M.J.M. (2012). Comparing different accelerometer cut-points for sedentary time in children. *Pediatric Exercise Science*, 24, 220 – 228.
- Fitzgibbon, M.L., Stolley, M.R., Schiffer, L.A., Braunschweig, C.L., Gomez, S.L., Van Horn, L., & Dyer, A.R. (2011). Hip-Hop to Health Jr. Obesity Prevention Effectiveness Trial: Post intervention Results. *Obesity*, 19(5), 994 – 1003. doi:10.1038/oby.2010.314
- Flegal, K.M., Keyl, P.M., & Nieto, F.J. (1991). Differential misclassification arising from nondifferential errors in exposure measurement. *American Journal of Epidemiology*, 134 (10), 1233-1244.
- Folkman, S., & Moskowitz, J.T. (2004). Coping: pitfalls and promise. *Annual Review of Psychology*, 55, 745 – 774. doi: 10.1146/annurev.psych.55.090902.141456
- Fosgate, G.T. (2006). Non-differential measurement error does not always bias diagnostic likelihood ratios towards the null. *Emerging Themes in Epidemiology*, 3, doi: 10.1186/1742-7622-3-7
- Fox, K. (1997). *The physical self: from motivation to well-being*. Champagne, IL: Human Kinetics.
- Fox, K. (1999). The influence of physical activity on mental well-being. *Public Health Nutrition*, 2 (3a), 411 – 418.
- Fox, K., (2000). The effects of exercise on self-perceptions and self-esteem. In S. Biddle., K.Fox., & S.H. Boutcher. (Ed). *Physical Activity and Psychological Well-Being*. (pp. 88 – 117). London, Uk: Routledge.
- Fox, K., & Corbin, C., (1989). The physical self-perception profile: Development and preliminary validation. *Journal of Sport and Exercise Psychology*, 11, 408 – 430.
- Freedson, P.S., & John, D. (2013). Comment on “Estimating Activity and Sedentary Behaviour from an Accelerometer on the Hip and Wrist”. *Medicine and Science in Sports and Exercise*, 45 (5), 962 – 963. doi: 10.1249/MSS.0b013e31827f024d
- Freedson, P. S., Melanson, E. L., & Sirard, J. R. (1997). Calibration of the computer science and applications, Inc. accelerometer. *Medicine and Science in Sports and Exercise*, 29, S45.
- Freedson, P., Pober, D., & Janz, K.F. (2005). Calibration of accelerometer output for children. *Medicine and Science in Sports and Exercise*, 37, pp. S523 – S530.

- Gail, M.H. (2005). Bias from Loss to follow up. *Encyclopedia of Biostatistics*. doi: 10.1002/0470011815.b2a03010
- Gaus, W., & Muche, R. (2013). Is a controlled randomised trial the non-plus-ultra design? A contribution to discussion on comparative, controlled, non-randomised trials. *Contemporary Clinical Trials*, 35 (1), 127 – 132. <http://dx.doi.org/10.1016/j.cct.2013.02.011>
- Goldfield, G.S., Henderson, K., Buchholz, A., Obeid, N., Nguyen, H., & Flament, M.F. (2011). Physical activity and psychological adjustment in adolescents. *Journal of Physical Activity and Health*, 8, 157 – 163.
- Goodman, R., Meltzer, H., & Bailey, V. (1998). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *European Child and Adolescent Psychiatry*, 7, 125-130.
- Goodman, R., Ford, T., Richards, H., Gatward, R., & Meltzer, H. (2000). The development and well-being assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 41 (5), 645 – 655.
- Goodman, E., & Whitaker, R.C. (2002). A prospective study of the role of depression in the development and persistence of adolescent obesity. *Pediatrics*, 109 (3), 497 – 504. doi: 10.1542/peds.110.3.497
- Goran, M.I., Poehlman, E.T., & Danforth, E. (1994). Experimental reliability of the doubly labelled water technique. *American Journal of Physiology*, 266 (3), E510-E515.
- Green, H., McGinnity, A., Meltzer, H., Ford, T., & Goodman, R. (2005). *Mental Health of children and young people in Great Britain, 2004*. The Office of National statistics. Basingstoke, Hampshire: Palgrave MacMillan.
- Greiner, M., Pfeiffer, D., & Smith, R.D. (2000). Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventative Veterinary Medicine*, 45 (1-2), 23 – 41.
- Gruber, J.J. (1986). Physical activity and self-esteem development in children: A Meta – analysis. *American Academy of Physical Education Papers*, 19, 330 – 348.
- Guindon, S., & Cappeliez, P. (2010). Contributions of psychological well-being and social support to an integrative model of subjective health in later adulthood. *Ageing International*, 35, 38–60. doi: 10.1007/s12126-009-9050-7
- Guinhouya, B.C., Samouda, H., & de Beaufort, C. (2013). Levels of physical activity among children and adolescents in Europe: a review of physical activity assessed objectively by accelerometry. *Public Health*, 127 (4) , 301 – 311. doi: 10.1016/j.puhe.2013.01.020

- Gustafson, P., & Le Nhu, D. (2002). Comparing the effects of continuous and discrete covariate mismeasurement, with emphasis on the dichotomization of mismeasured predictors. *Biometrics*, 58 (4), 878 – 87.
- Hagger, M.S., Asçi, F.H., & Lindwall, M. (2004). A cross-cultural evaluation of a multidimensional and hierarchical model of physical self-perceptions in three national samples. *Journal of Applied Social Psychology*, 34 (5), 1075 – 1107.
- Hagger, M. S., Ashford, B., & Stambulova, N. (1998). Russian and British children's physical self-perceptions and physical activity participation. *Pediatric Exercise Science*, 10, 137-152.
- Hagger, M.S., Biddle, S.J.H., & Wang, C.K.J. (2005). Physical self-concept in adolescence: Generalizability of a multidimensional, Hierarchical model across gender and grade. *Educational and Psychological Measurement*, 65 (2), 297-322. doi: 10.1177/0013164404272484
- Halfon, N., Larson, K., & Slusser, W. (2013). Associations between Obesity and comorbid mental health, developmental, and physical health conditions in a nationally representative sample of US children aged 10 to 17. *Academic Pediatrics*, 13 (1), 6 – 13. doi: 10.1016/j.acap.2012.10.007
- Hall, A., Hern, R.A., & Fallowfield, L. (1999). Are we using appropriate self-report questionnaires for detecting anxiety and depression in women with early breast cancer? *European Journal of Cancer*, 35 (1), 79 – 85.
- Hamer, M., Stamatakis, E., & Mishra, G. (2009). Psychological distress, television viewing and physical activity in children aged 4 – 12 years. *Pediatrics*, 123 (5), 1263 – 1268, doi: 10.1542/peds.2008-1523
- Hands, B., Parker, H., & Larkin, D. (2006). Physical activity measurement methods for young children: A comparative study. *Measurement in Physical Education and Exercise Science*, 10 (3), 203 – 214. doi: 10.1207/s15327841mpee1003_5
- Hanggi, J., Phillips, L., & Rowlands, A.V. (2013). Validation of the GT3X ActiGraph in children and comparison with the GT1M. *Journal of Science and Medicine in Sport*, 16 (1), 40 – 4. doi: 10.1016/j.jsams.2012.05.012
- Harrell, J., McMurray, R., Bagget, C., Pennell, M., Pearce, P., & Bangdiwala, H. (2005). Energy costs of Physical activity in children and adolescents. *Medicine and Science in Sport and Exercise*, 37, 329 – 336.
- Harrington, R. (2002). Affective Disorders. In M. Rutter & E. Taylor (eds). *Child and Adolescent Psychiatry* (4th Ed). (pp. 463 – 485). Oxford: Blackwell Science.
- Harris, A., Cronkite, R., & Moos, R. (2006). Physical activity, exercise coping and depression in 10 – year cohort study of depressed patients. *Journal of Affective Disorders*, 93, 79 – 85.

- Harter, S. (1982). The perceived competence scale for children. *Child Development*, 53 (1), 87 – 97.
- Harter, S. (1985). *Manual for the Self - Perception profile for children*. Denver: University of Denver
- Harter, S. (1999). *The construction of the self*. New York: Guilford Press
- Harter, S., & Whitesell, N.R. (2003) Beyond the debate: Why some adolescents report stable self-worth overtime and situation, whereas others report changes in self-worth. *Journal of Personality*, 71(6), 1027-1058.
- Heil, D.P., Brage, S., & Rothney, M.P. (2012). Modeling physical activity outcomes from wearable monitors. *Medicine and Science in Sports and Exercise*, 44 (1s), s50 – s60. doi:10.1249/mss.0b013e3182399dcc
- Hendelman, D., Miller, K., Baggett, C., Debold, B.E., & Freedson, P. (2000). Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Medicine and Science in Sports Exercise*, 32 (9 Suppl), S442–S449.
- Henry, C.J.K., Lightowler, H.J., & Al-Hourani, H.M. (2004). Physical activity and levels of inactivity in adolescent females ages 11 – 16 years in the United Arab Emirates. *American Journal of Human Biology*, 16, 346 – 353. doi: 10.1002/ajhb.20022
- Herrmann, S.D., Barreira, T.V., Kang, M., & Ainsworth, B.E. (2012). Impact of accelerometer wear time on physical activity data: a NHANES semisimulation data approach. *British Journal of Sports Medicine*, 0, 1 – 5, doi:10.1136/bjsports-2012-091410.
- Higgins, E.T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, 94 (3), 319 – 340.
- Ho, P.M., Perterson, P.N., & Masoudi, F.A. (2008). Evaluating the evidence: Is there a rigid hierarchy? *Circulation*, 118, 1675 – 1684.
- Hong, X., Li,J., Xu,F., Tse, L. A., Liang, Y., Wang,Z., Yu, I.T., & Griffiths, S. (2009). Physical activity inversely associated with the presence of depression among urban adolescents in regional China. *BMC Public Health*, 9, 148 – 156. doi: 10.1186/1471-2458-9-148.
- Huang,J.S., Norman, G.J., Zabanski, M.F., Calfas, K., & Patrick, K. (2007), Body image and self-esteem among adolescents undergoing an intervention targeting dietary and physical activity behaviours. *Journal of Adolescent Health*, 40 (3), 245 – 251. doi: 10.1016/j.jadohealth.2006.09.026
- Hubble, A.M., & Zumbo, B.D. (1996). A dialect on Validity: Where we have been and where we are going. *The Journal of General Psychology*, 123 (3), 207 – 215.

- Huberty, J.H., Siahpush, M., Beighle, A., Fuhrmeister, E., Silva, P., & Welk, G. (2011). Ready for recess: A pilot study to increase physical activity in elementary school children. *Journal of School Health*, 81 (5), 251 – 257.
- Hume, C., Timperio, A., Veitch, Salmon, J., Crawford, D., & Ball, K. (2011). Physical activity, sedentary behaviour and depressive symptoms among adolescents. *Journal of Physical Activity and Health*, 8 (2), 152 – 156.
- Inchley, J., Kirby, J., & Currie, C. (2011). Longitudinal changes in physical self-perceptions and associations with physical activity during adolescence. *Pediatric Exercise Science*, 23 (2), 237 – 249.
- Intille, S.S., Lester, L., Sallis, J.F., & Duncan, G. (2012). New horizons in sensor development. *Medicine and Science in Sport and Exercise*, 44 (1suppl), s24 – s31. doi: 10.1249/MSS.0b013e3182399c7d
- Jacka, F.N., Pasco, J.A., Williams, L.J., Leslie, E.R., Dodd, S., Nicholson, G.C., Kotowicz, M.A., & Berk, M. (2011). Lower levels of physical activity in childhood associated with adult depression. *Journal of Science and Medicine in Sport*, 14 (3), 222 – 226. doi: 10.1016/j.jsams.2010.10.458
- Jacob, M.L., & Storch, E.A. (2013). Pediatric Obsessive–Compulsive Disorder: A Review for Nursing Professionals. *Journal of Child and Adolescent Psychiatric Nursing*, 26 (2), 138 – 148. doi: 10.1111/jcap.12029
- Jago, R., Anderson, C.B., Baranowski, T., & Watson, K. (2005). Adolescent patterns of physical activity, differences by gender, day and time of day. *American Journal of Preventative Medicine*, 28 (5), 447 – 452, doi: 10.1016/j.amepre.2005.02.007
- Jago, R., Watson, K., Baranowski, T., Zakeri, I., Yoo, S., Baranowski, J., & Conry, K. (2006). Pedometer reliability, validity and daily activity targets among 10- to 15- year old boys. *Journal of Sport Science*, 24 (3), 241 – 251. doi:10.1080/02640410500141661
- Jago, R., Zakeri, I. Barnanowski, T., & Waton, K. (2007). Decision boundaries and receiver operator characteristic curves: New methods for determining accelerometer cut points. *Journal of Sport Science*, 25, 937 – 944. doi:10.1080/02640410600908027
- James, W. (1890). Principles of Psychology. Chicago: Encyclopedia Britannica.
- Janz, K.F. (2002). Use of heart rate monitoring to assess physical activity. In G.Welk (eds). *Physical Activity Assessments for Health – Related Research*. (p.143 – 158).Champaign, IL: Human kinetics
- Janz, K.F., Kwon, S., letuchy, E.M., Eichenberger Gilmore, J.M., Burns, T.L., Torner, J.C., Willing, M.C., & Levy, S.M. (2009). Sustained effect of early physical activity on body fat mass in older children. *American Journal of Preventive Medicine*, 37 (1), 35 – 40. doi: 10.1016/j.amepre.2009.03.012

- Johnson, C.C., Murray, D.M., Elder, J.P., Jobe, J.B., Dunn, A.L., Kubik, M., Voorhee, C., & Schachter, K. (2008). Depressive symptoms and physical activity in adolescents girls. *Medicine and Science in Sport and Exercise*, 40 (5), 818 – 826. doi: 10.1249/MSS.0b013e3181632d49
- Kang, M., Rowe, D.A., Barreira, T.V., Robinson, T.S., & Mahar, M.T. (2009). Individual information-centred approach for handling physical activity missing data. *Research Quarterly for Exercise and Sport*, 80 (2), 131 – 137.
- Kantomaa, M.T., Tammelin, T.H., Ebeling, H.E. & Taanila, A.M. (2008). Emotional and behavioural problems in relation to physical activity in youth. *Medicine and Science in Sport and Exercise*, 40 (10), 1749 – 1756. doi: 10.1249/MSS.0b013e31817b8e82.
- 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 Information Technology in Biomedicine*, 10(1), 156-67.
- Katzmarzyk, P.T., & Tremblay, M.S. (2007). Limitations of Canada's physical activity data: implications for monitoring trends. *Applied Physiology, Nutrition and Metabolism*, 32 (suppl 2E), s185 – s194.
- Kelly, P., Doherty, A., Berry, E., Hodges, S., Batterham, A.M., & Foster, C. (2011). Can we use digital life-log images to investigate active and sedentary travel behaviour? Results from a pilot study. *International Journal of Behavioural Nutrition and physical activity*, 20 (8), doi: 10.1186/1479-5868-8-44
- Kelly, P., Doherty, A.R., Hamilton, A., Matthews, A., Batterham, A.M., Nelson, M, Foster, C., & Cowburn, G. (2012). Evaluating the feasibility of measuring travel to school using a wearable camera. *American Journal of Preventative Medicine*, 43 (5), 546 – 550. doi: 10.1016/j.amepre.2012.07.027
- Kessler, R.C., Chiu, W., Delmer, O., & Walters, E.E. (2005). Prevalence, Severity and Comorbidity of twelve-month DSM-IV Disorders in the national comorbidity survey replication (NCS-R). *Archive of General Psychiatry*, 62 (6), 617 – 627.
- Kim, Y., Beets, M.W., & Welk, G.J. (2012). Everything you wanted to know about selecting the “right” accelerometer cut-points for youth, but ...: A systematic review. *Journal of Science and Medicine in Sport*, 15, 311 – 321. doi:10.1016/j.jsams.2011.12.001
- King, A.C., Parkinson, K.N., Adamson, A.J., Lilian, M., Besson, H., Reilly, J.J., & Basterfield, L. (2011). Correlates of objectively measured physical activity and sedentary behaviour in English children. *European Journal of Public Health*, 21 (4), 424 – 431. doi: 10.1093/eurpub/ckq104

- Klesges, L.M., Baranowski, T., Beech, B., Cullen, K., Murray, D.M., Rochon, J., & Pratt, C. (2004). Social desirability bias in self-reported dietary, physical activity and weight concerns measures in 8- to 10- year old African-American girls: results from the girls health enrichment multisite studies (GEMS). *Preventive Medicine, 38*, s78 – s87.
- Knight, J.F., Bristow, H.W., Anastopoulou, S., Baber, C., Schwitz, A., & Arvanitis, T.N. (2007). Use of accelerometer data collected from a wearable system. *Personal and Ubiquitous Computing, 11*, 117 – 132. doi: 10.1007/s00779-006-0070-y
- Kohl, H.W., Fulton, J.E., & Caspersen, C.J. (2000). Assessment of Physical activity among children and adolescents: A review and synthesis. *Preventive Medicine, 31*, s54 – s76. doi: 10.1006/pmed.1999.0542
- Kovacs, M., & Beck, A. (1977). An empirical clinical approach towards definition of childhood depression. In J.Schulterbrandt & A. Raskin (Ed.), *Depression in children* (pp. 1 - 25). New York: Raven.
- Kowalski, K.C., Crocker, P.R.E., & Donen, R.M. (2004). *The physical activity questionnaire for older children (PAQ-C) and adolescents (PAQ-A) Manual*. Saskatoon, Canada.
- Kowalski, K. C., Crocker, P. R. E., & Faulkner, R. A. (1997). Validation of the Physical Activity Questionnaire for Older Children. *Pediatric Exercise Science, 9*, 174-186.
- Kreuser, F., Kromeyer-Hauschild, K., Gollhofer, A., Korsten-Reck, U., & Rottger, K. (2013). “obese equals lazy?” analysis of the association between weight status and physical activity in children. *Journal of Obesity*, doi: 10.1155/2013/437017
- Larun, L., Nordheim, L.V., Ekelund, E., Hagen, K.B., & Heian, F. (2009). Exercise in prevention and treatment of anxiety and depression among children and young people (review). *The Cochrane Database of systematic reviews, 3*, doi:10.1002/14651858.CD004691.pub2
- Lau, P.W.C., Yu, C.W., Lee, A., & Sung, R.Y.T. (2004). The physiological and psychological effects of resistance training on Chinese obese adolescents. *Journal of Exercise Science, 2* (2), 115 – 120.
- Lazarus, R., & Folkman, S. (1984). *Stress, Appraisal and Coping*. New York, NY: Springer Publishing company.
- Lee, P.H. (2013). Data imputation for accelerometer-measured physical activity: the combined approach. *American Journal of Clinical Nutrition, 97*, 965 – 971. doi: 10.3945/ajcn.112.052738
- Lee, A., & Hankin, B.L. (2009). Insecure attachment, dysfunctional attitudes, and low self-esteem predicting prospective symptoms of depression and anxiety during adolescence. *Journal of Clinical Child and Adolescent Psychology, 38* (2), 219 – 231.

- Lee, I.M., Sesso, H.D., Oguma, Y., & Paffenbarger, R.S. (2003). Relative intensity of physical activity and risk of coronary heart disease. *Circulation*, 107 (8), 1110 – 1116.
- Lindwall, M., Asçi, F.H., Palmeira, A., Fox, K., & Hagger, M.S. (2011). The importance of importance in the physical self: Support for the theoretically appealing but empirically elusive model of James. *Journal of Personality*, 79 (2), 303 – 334.
- Lindwall, M., Renmark, M., Halling, A., Berglund, J., & Hassmen, P. (2007). Depression and exercise in elderly men and women: findings from the Swedish national study on aging and care. *Journal of Aging Physical Activity*, 15, 41 – 55.
- Livingstone, M.B.E., (1997). Heart rate monitoring: the answer for assessing energy expenditure and physical activity in population studies? *British Journal of Nutrition*, 78, 869 – 871.
- Loprinzi, P.D., & Cardinal, B.J. (2011). Measuring children's physical activity and sedentary behaviours. *Journal of Exercise Science and Fitness*, 9 (1), 15 – 23.
- Loprinzi, P.D., Lee, H., Cardinal, B.J., Crespo, C.J., Andersen, R.E., & Smit, E. (2012). The relationship of actigraph accelerometer cut-points for estimating physical activity with selected health outcomes: results from NHANES 2003-06. *Research Quarterly for Exercise and Sport Science*, 83 (3), 422 – 430.
- Long, B.C., & Van Stavel, R. (1995). Effects of exercise training on anxiety: A meta analysis. *Journal of Applied Sport Psychology*, 7, 167 – 189.
- Löwe, B., Spitzer, R.L., Gräfe, K., Kroenke, K., Quenter, A., Stephan, Z., Buchholz, C., Witte, S., & Herzog, W. (2004). Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. *Journal of Affective Disorders*, 78 (2), 131 – 140.
- Lubans, D.R., Morgan, P.J., & Tudor-locke, C. (2009). A systematic review of studies using pedometers to promote physical activity. *Preventive Medicine*, 48 (4) , 307 – 315. doi: 10.1016/j.ypmed.2009.02.014
- Macinnes, D.L. (2006). Self-esteem and self-acceptance: an examination into their relationship and their effect on psychological health. *Journal of Psychiatric and Mental Health Nursing*, 13, 483 – 489.
- Mackintosh, K.A., Fairclough, S.J., Stratton, G., & Ridgers, N.D. (2012). A Calibration protocol for population-specific accelerometer cut-points in children. *PLoS One*, 7 (5), e36919 doi: 10.1371/journal.pone.0036919
- Magnusson, K.T., Sigurgeirsson, I. Sveinsson, T., & Johannsson, E. (2011). Assessment of a two-year school-based physical activity intervention among 7 – 9 year old children. *International Journal of Behavioural Nutrition and Physical Activity*, 8, 138 – 149. doi: 10.1186/1479-5868-8-138

- Malhotra, N. K., Kim, S.S., & Patil, A. (2006), Common Method Variance in IS Research: A Comparison of Alternative Approaches and a Reanalysis of Past Research. *Management Science*, 52 (12),1865-1883. doi: 10.1287/mnsc.1060.0597
- Mann, M., Hosman, C.M., Schaalma, H.P., & de Vries, N.K. (2004). Self – esteem in a broad-spectrum approach for mental health promotion. *Health Education Research*, 19 (4), 357 – 372
- Marsh, H.W., Richards, G.E., Johnson, S., Roche, L., & Tremayne, P. (1994). Physical self-description questionnaire: Psychometric properties and multitrait- multi method analysis of relations to existing instruments. *Sport and Exercise Psychology*, 16, 270 – 305.
- Marsh, H. W. (1997). The measurement of physical self-concept: A construct validation approach. In K. R. Fox (Ed.), *The physical self: From motivation to well-being* (pp. 27–58). Champaign, IL: Human Kinetic
- Marshall, S.J., Levy, S.S., Tudor-locke, C.E., kolkhorst, F.W., Wooten, K.M., Ji,M., Macera, C.A., & Ainsworth, B.E. (2009). Translating physical activity recommendations into a pedometer-based step Goal, 3000 steps in 30minutes. *American Journal of Preventive Medicine*, 36 (5), 410 – 415. doi:10.1016/j.amepre.2009.01.021
- Martinez-Gomez, D., Gomez-Martinez, S., Ruiz, J.R., Diaz, L.E., Ortega, F.B., Widhalm, K., Cuenca-Garcia, M... & Marcos, A. (2012). Objectively-measured and self-reported physical activity and fitness in relation to inflammatory markers in European adolescents: the HELENA Study. *Atherosclerosis*, 221 (1), 260 – 7, doi: 10.1016/j.atherosclerosis.2011.12.032
- Massé, R., Poulin, C., Dassa, C., Lambert, J., Bélair, S., & Battaglini, M.A.(1998) Elaboration and validation of a tool to measure psychological well-being. *Canadian Journal of Public Health*, 89 (5), 352–7.
- Masse, L.C., Fuemmeler, B.F., Andersen, C.B., Matthes, C.E., Trost, S.G., Cattelier, D.J., & Treuth, M. (2005). Accelerometer Data Reduction: A Comparison of Four Reduction Algorithms on Select Outcome Variables. *Medicine and Science in Sport and Exercise*, 37 (11supplement), s544 – s554.
- Matthews, C.E., Chen, K.Y., Freedson, P.S., Buchowski, M.S., Beech, B.M., Pate, R.R., & Troiano, R.P. (2008). Amount of time spent in sedentary behaviours in the United States, 2003-2004. *American journal of Epidemiology*, 167 (7), 875 – 881. doi: 10.1093/aje/kwm390
- Matthews, C.E., Hagströmer, Pober, D.M., & Bowles, H.R. (2012). Best practices for using physical activity monitors in population-based research. *Medicine and Science in Sport and Exercise*, 44 (1s), s68 – s76.

- Mattocks, C., Leary, S., Ness, A., Deere, K., Saunders, J., Tilling, K., Kirkby, J., Blair, S.N., & Riddoch, C. (2007). Calibration of an accelerometer during free-living activities in children. *International Journal of Pediatric Obesity*, 2 (4), 218 – 226. doi: 10.1080/17477160701408809
- Maughan, B., Brock, A., & Ladva, G. (2004). Mental Health. In The Office of National Statistics. *The Health of Children and Young People*.
- McAuley, E., Blissmer, B., Katula, J., Duncan, T.E., & Mihalko, S.L. (2000). Physical activity, self-esteem and self-efficacy relationships in older adults: A randomised controlled trial. *Annals of Behavioural Medicine*, 22 (2), 131 – 139.
- McAuley, E., Mihalko, S.L., & Bane, S.M. (1997). Exercise and self esteem in middle-aged adults: Multidimensional relationships and physical fitness and self-efficacy influences. *Journal of Behavioural Medicine*, 20, 67 – 83.
- McAuley, E., & Rudolph, D.L (1995). Physical activity, aging, and psychological well-being. *Journal of Aging and Physical Activity*, 3, 67-96.
- McCarthy, H.D., Ellis, S.M., & Cole, T.J. (2003) Central overweight and obesity in British youth aged 11–16 years. *British Medical Journal*, 326, 624–628.
- McClain, J.J., Abraham, T.L., Brusseau, T.A., & Tudor-Locke, C. (2008). Epoch length and accelerometer outputs in children: comparison to direct observation. *Medicine and Science in Sport and Exercise*, 40 (12), 2080 – 2087. doi: 10.1249/MSS.0b013e3181824d98
- McClure, S.A., Summerbell, C.D., & Reilly, J. (2009). Objectively measured habitual physical activity in a highly obesogenic environment. *Child: Care, Health & Development*, 35 (3), 369 – 375. doi:10.1111/j.1365-2214.2009.00946.x
- McClure, A.C., Tanski, S.E., Kingsbury, J., Gerrard, M., & Sargent, J.D. (2010). Characteristics associated with low self-esteem among US adolescents. *Academic Pediatrics*, 10 (4), 238 – 244. doi: 10.1016/j.acap.2010.03.007
- McGee, R., & Williams, S. (2000). Does low self-esteem predict health compromising behaviours among adolescents. *Journal of Adolescence*, 23 (5), 569 – 582.
- McIver, K.I., Brown, W.H, Pfeiffer, K.A., Dowda, M., & Pate, R.R. (2009). Assessing children's physical activity in their homes: the observational system for recording physical activity in children-home. *Journal of Applied Behaviour Analysis*, 42 (1), 1 – 16. doi: 10.1901/jaba.2009.42-1
- McKenzie, T.L. (2002). Use of direct observation to assess physical activity. In G.Welk, *Physical Activity Assessments for Health – Related Research*. (pp.179-191) Champagne IL: Human kinetics.

- McKenzie, T.L., Marshall, S.J., Sallis, J.F., & Conway, T.L. (2000). Leisure-time physical activity in school environments: an observational study using SOPLAY. *Preventive Medicine*, 30 (1), 70 – 77. doi:10.1006/pmed.1999.0591
- McKenzie, T.L., Sallis, J.F., Nader, P.R., Patterson, T.L., Elder, J.P., Berry, C.C., Rupp, J.W...&Nelson, J.A. (1991). BEACHERS: An observational system for assessing children's eating and physical activity behaviours and associated events. *Journal of Applied Behaviour Analysis*, 24(1), 141 – 151. doi: 10.1901/jaba.1991.24-141
- McLaughlin, J.E., King, G.A., Howley, E.T., Bassett, D.R., & Ainsworth, B.E. (2001). Validation of the cosmed K4B2 portable metabolic system. *International Journal of Sport Medicine*, 22, 280 – 284.
- McManus, S., Meltzer, H., Brugha, T., Bebbington, P., & Jenkins, R. (2009). *Adult psychiatric morbidity in England, 2007: Results of a household survey*. The NHS information centre for Health and social care. (www.ic.nhs.uk/pubs/psychiatric morbidity07).
- McMurray, R.G., Harrell, J.S., Bradley, C.B., Webb, J.P., & Goodman, E. (1998). Comparison of a computerised physical activity recall with a triaxial motion sensor in middle-school youth. *Medicine and Science in Sport and Exercise*, 30 (8), 1238 – 1245.
- McNamara, E., Hudson, Z., & Taylor, S.J.C. (2010). Measuring activity levels of young people: the validity of pedometers. *British Medical Bulletin*, 95, 121 – 137. doi: 10.1093/bmb/ldq016
- McPhie, M.L., & Rawana, J.S. (2012). Unravelling the relation between physical activity, self esteem and depressive symptoms among early and late adolescents: A mediation analysis. *Mental Health and Physical Activity*, 5, 43 – 49. doi: 10.1016/j.bbr.2011.03.031
- Mead, G.E., Morley, W., Campbell, P., Greig, C.A., McMurdo, M.E.T., & Lawlor, D.A. (2009). Exercise for depression. *Mental Health and Physical Activity*, 2, 95 – 96.
- Meade, A.W., Watson, A.M., & Kroustalis, C.M. Assessing common method bias in organisational research. *Paper presented at the 22nd Annual Meeting of the Society for Industrial and Organisational Psychology, New York*
- Meng, X.-L., Rosenthal, R., & Rubin, D.B., (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, 111, 172–175.
- Molner, F.J., Hutton, B., & Fergusson, D. (2008). Does analysis using "last observation carried forward" introduce bias in dementia research? *Canadian Medical Association Journal*, 179 (8), 751 – 753. doi: 10.1503/cmaj.080820

- Monshouwer, K., Have, M.T., Poppel, M., Kemper, H., & Vollebergh, W. (2012). Possible mechanisms explaining the association between physical activity and mental health: Findings from the 2001 Dutch health behaviour in school – aged children survey. *Clinical Psychological Science*, doi: 10.1177/2167702612450485
- Montori, V.M., & Guyatt, G.H. (2001). Intention-to-treat principle. *Canadian Medical Association Journal*, 165 (10), 1339 – 1341.
- Montoye, H.J., Kemper, H.C.G., Saris, W.H.M., & Washburn, R.A. (1996). *Measuring Physical activity and energy expenditure*. Champaign, IL: Human Kinetics.
- Morgan, C.F., Pangrazi, R.P., & Beighle, A. (2003). Using pedometers to promote physical activity in physical education. *Journal of Physical Education & Dance*, 74 (7), 33 – 38. <http://dx.doi.org/10.1080/07303084.2003.10609235>
- Motl, R.W., Birnbaum, A.S., Kubik, M.Y., & Dishman, R.K. (2004). Naturally occurring changes in physical activity are inversely related to depressive symptoms during early adolescence. *Psychosomatic Medicine*, 66, 336 – 342.
- Muris, P., Meesters, C., & Van den Berg, F. (2003). The Strengths and Difficulties Questionnaire (SDQ)--further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *European Child and Adolescent Psychiatry*, 12 (1), 1 – 8. doi: 10.1007/s00787-003-0298-2
- Murray, C.L.J., & Lopez, A.D. (1997). Alternative projections of mortality and disability by cause 1990–2020: Global Burden of Disease Study. *The Lancet*, 149, 1498 – 1504.
- Naicker, K., Galambos, N.L., Zeng, Y., Senthithilselvan, A., & Colman, I., (2013). Social, demographic, and health outcomes in the 10 years following Adolescent depression. *Journal of Adolescent Health*, 52 (5), 533 – 538. doi: 10.1016/j.jadohealth.2012.12.016
- Neissaar, I., & Raudsepp, L. (2011). Changes in Physical activity, self-efficacy and depressive symptoms in adolescent girls. *Pediatric Exercise Science*, 23, 331 - 343.
- Nichols, J. (2012). Moving from self-esteem to self-acceptance. *Kai Tiaki Nursing New Zealand*, 18 (8), 28 – 29.
- Niemi, I. (1993). Systematic error in behavioural measurement: comparing results from interview and time budget studies. *Social Indicators Research*, 30 (2/3), 229 – 244.
- Nihill, G. F. J., Lubans, D. R., & Plotnikoff, R. C. (2013). Associations between sedentary behaviour and self-esteem in adolescent girls from schools in low income communities. *Mental Health and Physical Activity*, 6 (1), 30 - 35. doi:10.1016/j.mhpa.2012.02.003

- Nilsson, A., Ekelund, U., Yngve, A., & Sjöström, M. (2002). Assessing physical activity among children with accelerometers using different time sampling intervals and placements. *Pediatric Exercise Science*, 14, 87 – 96.
- Norris, R., Carrol, D., & Cochrane, R. (1992). The effects of physical activity and exercise training on psychological stress and well-being in an adolescent population. *Journal of Psychosomatic Research*, 36 (1), 55 – 65.
- North, T.C., McCullagh, P., & Tran, Z.V. (1990). Effect of exercise on depression. *Exercise and Sport Science Reviews*, 18 (1), 379 – 416.
- Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Osei – Tutu, K., & Campagna, P.D. (1998) Psychological benefits of continuous vs intermittent moderate intensity exercise. *Medicine and Science in Sport and Exercise*, 30, s117.
- Oshima, Y., Kawauchi, K., Tanaka, S., Ohkawara, K., Hikiyama, Y., Ishikawa-Takata, K., & Tabata, I. (2010). Classifying household and locomotive activities using a triaxial accelerometer. *Gait and Posture*, 31 (3), 370 – 374. doi: 10.1016/j.gaitpost.2010.01.005
- Ott, A.E., Pate, R.R., Trost, S.G., Ward, D.S. & Saunders, R. (2000). The Use of Uniaxial and Triaxial Accelerometers to Measure Children's "Free-Play" Physical Activity. *Pediatric Exercise Science*, 12 (4), 360 – 370.
- Ottevaere, C., Huybrechts, I., De Meester, F., De Bourdeaudhuij, I., Cuenca-Garcia, M., & De Henauw, S. (2011). The use of accelerometry in adolescents and its implementation with non-wear time activity diaries in free-living conditions. *Journal of Sports Science*, 29 (1), 103 - 113. doi: 10.1080/02640414.2010.521169
- Ozdoba, R., Corbin, C., & Le Masurier, G. (2004) Does reactivity exist in children when measuring activity levels with unsealed pedometers. *Pediatric Exercise Science*, 16, 158-166.
- Page, A.S., Cooper, A.R., Griew, P., & Jago, R. (2010). Children's screen viewing is related to psychological difficulties irrespective of physical activity. *Pediatrics*, 126 (5), e1011-7. doi: 10.1542/peds.2010-1154
- Paluska, S.A., & Schwenk, T.L. (2000). Physical activity and mental health: current concepts. *Sports Medicine*, 29, 167 – 180.
- Parfitt, G., & Eston, R. (2005). The relationship between children's habitual physical activity level and psychological well-being. *Acta Paediatrica*, 94: 1791 - 1797. Doi:10.1080/08035250500268266
- Parfitt, G., Pavey, T., & Rowlands, A. (2009). Children's physical activity and psychological health: The relevance of intensity *Acta Paediatrica*, 98, 1037 – 1043. doi:10.1111/j.1651-2227.2009.01255.x

- Park, N. (2004). The role of subjective well-being in positive youth development. *The ANNALS of the American Academy of Political and Social Science*, 591. doi: 10.1177/0002716203260078
- Parry – Langdon, N. (2008). *Three years on: survey of the development and emotional well – being of children and young people*. The Office of National statistics.
- Pate, R.R., Almeida, M.J., McIver, K.L., Pfeiffer, K.A., & Dowda, M. (2006). Validation and calibration of an accelerometer in preschool children. *Obesity*, 14 (11), 2000 – 2006.
- Pate, R.R., Freedson, P.S., Sallis, J.F., Taylor, W.C., Sirard, J., Trosts, S.G., & Dowda, M. (2002). Compliance with Physical activity guidelines: Prevalence in a population of children and youth. *Annals of Epidemiology*, 12 (5), 303 – 308.
- Pate, R. R., Ross, R., Dowda, M., Trost, S.G., & Sirard, J.R. (2003). Validation of a 3-day physical activity recall instrument in female youth. *Pediatric Exercise Science*, 15, 257-265
- Paul, D.R., Kramer, M., Moshfegh, A.J., Baer, D.J., & Rumpler, W.V. (2007). Comparison of two different physical activity monitors. *BMC Medical Research Methodology*, 7, doi:10.1186/1471-2288-7-26
- Pelcovitz, D., Van der Kolk, B., Roth, S., Mandel, F., Kaplan, S., & Resick, P. (1997). Development of a criteria set and a structured interview for disorders of extreme stress (SIDES). *Journal of Traumatic Stress*, 10 (1), 3 – 16.
- Perry, A.M., Hendrick, P.A., Hale, L., Baxter, G.D., Milosavljevic, S., Dean, S.G., McDonough, S.M., & Hurley, D.A. (2010). Utility of the RT3 triaxial accelerometer in free living: An investigation of adherence and data loss. *Applied Ergonomics*, 41, 469 – 476.
- Petty, K.H., Davis, C.L., Tkacz, J., Young-Hyman, D., & Waller, J.L. (2009). Exercise effects on depressive symptoms and self-worth in over-weight children: A randomised controlled trial. *Journal of Pediatric Psychology*, 34 (9), 929 – 939. Doi: 10.1093/jpepsy/jsp007
- Petruzzello, S.J., Landers, D.M., Hatfield, B.D., Kubitz, K.A., & Salazar, W. (1991). A meta-analysis on the anxiety-reducing effects of acute and chronic exercise. *Outcomes and Measures. Sports Medicine*, 11 (3), 143 – 182.
- Phillips, D., (1973). *Abandoning Method, Sociological studies in methodology*. London: Jossey-Brass Publishers.
- Phillips, L.R.S., Parfitt, G., & Rowlands, A.V. (2013). Calibration of the GENEA accelerometer for assessment of physical activity intensity in children. *Journal of Science and Medicine in Sport*, 16 (2), 124 – 128. doi: 10.1016/j.jsams.2012.05.013

- 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 Sport and Exercise*, 38 (9), 1626 – 1634. doi: 10.1249/01.mss.0000227542.43669.45
- Podsakoff, P.M., Mackenzie, S.B., Lee, J.Y., & Podsakoff, N.P. (2003). Common method Biases in behavioural research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88 (5), 879 – 903. doi: 10.1037/0021-9010.88.5.879
- Pollard, E.L., & Lee, P.D. (2003). Child Well-being: A systematic Review of the literature. *Social Indicators Research*, 61 (1), 59 – 78. doi: 10.1023/A:1021284215801
- Puhl, J., Greaves, K., Hoyt, M., & Baranowski, T. (1990). Children's activity rating scale (CARS): description and calibration. *Research Quarterly for Exercise and Sport*, 61 (1), 26 – 36.
- Pulsford, R.M., Cortina-Borja, M., Rich, C., Kinnafick, F.E., Dezateux, C., & Griffiths, L.J. (2011). Actigraph accelerometer-defined boundaries for sedentary behaviour and physical activity intensities in 7 year old child. *PLoS One*, 6 (8), e21822. doi: 10.1371/journal.pone.0021822
- Pulsford, R.M., Griew, P., Page, A., Cooper, A., & Hillsdon, M. (In prep). Socioeconomic position and childhood sedentary time: Evidence from the PEACH study.
- Puyau, M. R., Adolph, A. L., Firoz, A. V., & Butte, N. F. (2002) Validation and calibration of physical activity monitors in children. *Obesity Research*, 10 (3), 150–157.
- Raudsepp, L., Liblik, R., & Hannus, A. (2002). Children's and adolescents' physical self-perceptions as related to moderate to vigorous physical activity and physical fitness. *Pediatric Exercise Science*, 14, 97 – 106.
- Raudsepp, L., Neissaar, I., & Kull, M. (2013). A longitudinal assessment of the links between physical activity and physical self-worth in adolescent females. *European Journal of Sport Science*, <http://dx.doi.org/10.1080/17461391.2013.775349>
- Raustorp, A., Ståhle, A., Gudasic, H., Kinnunen, A., & Mattsson, E. (2005). Physical activity and self-perception in school children assessed with the children and youth physical self-perception profile. *Scandinavian Journal of Medicine & Science in Sports*, 15 (2), 126 - 134.
- Ratray, J., & Jones, M.C. (2007). Essential elements of questionnaire design and development. *Journal of clinical Nursing*, 16, 234 – 243.
- Reilly, J.J. (2006). Tackling the obesity epidemic: New Approaches. *Archives of Disease in Childhood*, 91 (9), 724 – 726. doi: 10.1136/adc.2006.098855

- Reilly, J.J., Coyle, J., Kelly, L., Burke, G., Grant, S., & Paton, J.Y. (2003). An objective method for measurement of sedentary behaviour in 3- to 4-year olds. *Obesity Research*, 11 (10), 1155 – 1158. doi: 10.1038/oby.2003.158
- Reilly, J.J., Penpraze, V., Hislop, J., Davies, G., Grant, S., & Paton, J.Y. (2008). Objective Measurement of physical activity and sedentary behaviour: review with new data. *Archive of Disease in Childhood*, 93 (7), 614-619. doi: 10.1136/adc.2007.133272
- Rennie, K.L., & Wareham, N.J. (1998). The assessment of physical activity in individuals and populations: why try to be more precise about how physical activity is assessed?. *International Journal of Obesity and Related Metabolic Disorders*, s2, s30 – s38.
- Richardson, W. Stewart-Brown, S. & Wilcock, E., & Oldfield, M., & Thorogood, M. (2011). Utility of accelerometers to measure physical activity in children attending an obesity treatment intervention. *Journal of Obesity*, 398918 doi: 10.1155/2011/398918
- Riddoch, C., Bo Andersen, L., Wedderkopp, N., Harro, M., Klasson-Heggbø, L., Sardinha, L.B., Cooper, A., & Ekelund, U. (2004). Physical activity levels and patterns of 9- and 15- yr old European children. *Medicine and Science in Sport and Exercise*, 36 (1), 86 – 92. doi: 10.1249/01.MSS.0000106174.43932.92
- Riddoch, C.J., Mattocks, C., Deere, K., Saunders, J., Kirkby, J., Tilling, K., Leary, S.D., Blair, S.N., & Ness, A.R. (2007). Objective measurement of levels and patterns of physical activity. *Archives of Disease in Childhood*, 92 (11), 963 – 969. doi: 10.1136/adc.2006.112136
- Ridely, K., & Olds, T. (2008). Assigning energy costs to activities in children: A review and Synthesis. *Medicine and Science in Sport and Exercise*, 40, 1439 – 1446. doi: 10.1249/MSS.0b013e31817279ef
- Rimer, J., Dwan, K., Lawlor, D.A., Greig, C.A., McMurdo, M., Morley, W., & Mead, G.E., Exercise for depression (Review). *Cochrane Database of Systematic Reviews*, 7, doi: 10.1002/14651858.CD004366.pub5
- Rohde, P., Lewinsohn, P.M., & Seeley, J.R. (1997). Comparability of telephone and face to face interviews in assessing axis I and II disorders. *American Journal of Psychiatry*, 154, 1593 – 1598.
- Romanzini, M., Petroski, E.L., & Reichert, F.F. (2011). Accelerometers thresholds to estimate physical activity intensity in children and adolescents: a systematic review. *Brazilian Journal of Kinanthropometry and Human Performance*, 14 (1), 101 – 113.
- Rosdahl, H., Gullstrand, L., Salier-Eriksson, J., Johansson, P., & Schantz, P. (2010). Evaluation of the Oxycon mobile metabolic system against the douglas bag method. *European Journal of Applied Physiology*, 109, 159 – 171. Doi: 10.1007/s00421-009-1326-9.

- Rosenberg, M., Schooler, C., Schoenbach, C., & Rosenberg, F. (1995). Global self-esteem and specific self esteem: Different concepts, different outcomes. *American Sociological Review*, 60 (1), 141 – 156.
- Rosenberger, M.E., Haskell, W.L., Albinali, F., Mota, S. Nawyn, J., & Intille, S. (2013). Estimating Activity and Sedentary Behaviour from an Accelerometer on the Hip or Wrist. *Medicine and Science in Sport and Exercise*, 45 (5), 964 – 975. doi: 10.1249/MSS.0b013e31827f0d9c
- Rothon, C., Edwards, P., Bhul, K., Viner, R.M., Taylor, S., & Stansfield, S.A. (2010). Physical activity and depressive symptoms in adolescent: a prospective study. *BMC Medicine*, 8 (32), doi: 10.1186/1741-7015-8-32.
- Rothonberger, A., & Woerner, W. (2004). Strengths and Difficulties Questionnaire (SDQ)--evaluations and applications. *European Child Adolescent Psychiatry*, 13 (suppl 2), 1 – 2.
- Rowlands, A.V. (2007). Accelerometer assessment of physical activity in children: an update. *Pediatric Exercise Science*, 19, 252-266.
- Rowlands, A.V., Esliger, D.W., Eady, J., &Eston, R.G. (2010). Empirical evidence to inform decisions regarding identification of non-wear periods from accelerometer habitual physical activity data. In: Bacquet, G. and Bethoin, S. *Children and Exercise XXV*. Routledge, London, 219-222.
- Rowlands, A.V., & Eston, R.G. (2007). Measurement and interpretation of physical activity of children. *Journal of Sports Science and Medicine*, 6, 270-276.
- Rowlands, A.V., Ingledew, D.K., & Eston, R.G. (2000). The effect of type of activity measure on the relationship between body fatness and habitual physical activity in children: A meta-analysis. *Annals of Human Biology*, 27, 479-497.
- Rowlands, A.V., Powell, S.M., Humphries, R., & Eston, R.G. (2006). The effect of accelerometer epoch on physical activity output measures. *Journal of Exercise Science and Fitness*, 4, 51-57.
- Rowlands, A.V., & Stiles, V.H. (2012). Accelerometer counts and raw acceleration output in relation to mechanical loading. *Journal of Biomechanics*, 45 (3), 448-454 doi: 10.1016/j.jbiomech.2011.12.006
- Rowlands, A.V., Thomas, P.W.M., Eston, R.G., & Topping R. (2004). Validation of the RT3 triaxial accelerometer for the assessment of physical activity. *Medicine and Science in Sport and Exercise*, 36, 518-524.
- Ryan, R.M., & Deci, E.L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *The Annual Review of Psychology*, 52, 141 – 166.

- Ryff, C.D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, 57 (6), 1069 – 1081.
- Sallis, J.F., Haskell, W.L., Wood, P.D., Fortman, S.P., Rogers, T., Balir, S.N., & Paffenbarger, R.S. (1985). Physical activity assessment methodology in the five-city project. *American Journal of Epidemiology*, 121 (1), 91 – 106.
- Sallis, J.F., Prochaska, J.J., & Taylor, W.C. (2000). A review of correlates of physical activity of children and adolescents. *Medicine and Science in Sport and Exercise*, 32 (5), 963 – 975.
- Sallis, J.F., & Saelens, B.E. (2000). Assessment of physical activity by self-report: status, limitations and future directions. *Research Quarterly for Exercise and Sport*, 71 (s2), s1 – s14.
- Sardinha, L.B., Andersen, L.B., Andersen, S.A., Quiterio, A.L., Ornelas, R., Froberg, K., Riddoch, C.J., & Ekelund, U. (2008). Objectively Measured Time Spent Sedentary Is Associated With Insulin Resistance Independent of Overall and Central Body Fat in 9- to 10-Year-Old Portuguese Children, *Diabetes Care*, 31 (3), 569-575. doi: 10.2337/dc07-1286
- Sasaki, J.E., John, D., & Freedson, P. (2011). Validation and comparison of the ActiGraph activity monitors. *Journal of Science and Medicine in Sport*, 14, 411 – 416. doi: 10.1016/j.jsams.2011.04.003
- Saylor, C.F., Finch, A.J., Spirito, A., & Bennett, B. (1984). The children's depression inventory: A systematic evaluation of psychometric properties. *Journal of Consulting and Clinical psychology*, 52, 955 – 967.
- Scaini, S.S., Ogliari, A., Eley, T.C., Zavos, H.M.S., & Battaglia, M. (2012). Genetic and Environmental contributions to separation anxiety: A meta – analytic approach to twin data. *Depression and Anxiety*, 29 (9), 754 – 761.
- Schmalz, D.L., Deane, G.D., Birch, L.L., & Davison, L.K. (2007). A longitudinal assessment of the links between physical activity and self-esteem in early adolescent non-Hispanic females. *Journal of Adolescent Health*, 41 (6), 559 – 565. doi: 10.1016/j.jadohealth.2007.07.001
- Schneider, M., Dunton, G.F., & Cooper, D.M. (2008). Physical activity and physical self-concept among sedentary adolescent females; An intervention study. *Psychology of Sport and Exercise*, 9 (1), 1 – 14. doi: 10.1016/j.psychsport.2007.01.003
- Schoeller, D.A. (1999). Recent advances from application of doubly labelled water to measurement of human energy expenditure. *The Journal of Nutrition*, 129, 1765 – 1768.
- Scully, D., Kremer, J., Meade, M., Graham, R., & Dudgeon, K. (1998). Physical exercise and psychological well-being: a critical review, *British Journal of Sport Medicine*, 38, 111 -120.

- Sebire, S.J., Jago, R., Fox, K., Page, A.S., Brockman, R., & Thompson, J. (2011). Associations between children's functioning and physical activity participation are not mediated by social acceptance: a cross-sectional study. *International Journal of Behavioural Nutrition and Physical Activity*, 8, doi: 10.1186/1479-5868-8-106
- Seery, M.D., Blascovich, J., Weisbuch, M., & Vick, S.B. (2004). The relationship between self-esteem level, self-esteem stability, and cardiovascular reactions to performance feedback. *Journal of Personality and Social Psychology*, 87(1), 133 – 145. doi: 10.1037/0022-3514.87.1.133
- Segal, D.L., Hersen, M., & Van Hasselt, V.B. (1994). Reliability of the structured clinical interview for DSM-III-R: An evaluative review. *Comprehensive Psychiatry*, 35 (4), 316 – 327.
- Seligman, M.E.P. (2011). *Flourish- A new understanding of happiness and well-being – and how to achieve them*. London: Nicholas Brealey Publishing.
- Sequeira, M.M., Rickenbach, M., Wietlisbach, V., Tullen, B., & Schutz, Y. (1995). Physical activity assessment using a pedometer and its comparison with a questionnaire in a large population survey. *American Journal of Epidemiology*, 142 (9), 989 – 998.
- Shah, H., & Marks, N. (2004). *A well-being manifesto for a flourishing society*. London: The new economics society.
- Shephard, R.J. (2001). Absolute versus relative intensity of physical activity in a dose-response context. *Medicine and Science in Sport and Exercise*, 33 (6), s400 – s418.
- Sherar, L.B., Griew, P., Esliger, D.W., Cooper, A.R., Ekelund, U., Judge, K., & Riddoch, C. (2011). International children's accelerometry database (ICAD): design and methods. *BMC Public Health*, 11, 485. doi: 10.1186/1471-2458-11-485.
- Sica, G.T. (2006). Bias in Research studies. *Radiology*, 238, 780 – 789. doi: 10.1148/radiol.2383041109
- Silverstone, P.H. (1994). Poor efficacy of the hospital anxiety and depression scale in the diagnosis of major depressive disorder in both medical and psychiatric patients. *Journal of Psychosomatic Research*, 38 (5), 441 – 450.
- Simonson, D.C., & DeFronzo, R.A. (1990). Indirect calorimetry: Methodological and interpretative problems. *American Journal of Physiology*, 258, e399 - e412.
- Singleton, N., Bumpstead, R., O'Brien, M., Lee, A., & Meltzer, H. (2001). *Psychiatric morbidity among adults living in private households*. London: TSO
- Sirard, J.R., & Pate, R., (2001). Physical activity assessment in children and Adolescents. *Sports Medicine*, 31, 439 – 454.

- Sirard, J. R., Trost, S. G., Pfeiffer, K. A., Dowda, M., & Pate, R. R. (2005). Calibration and evaluation of an objective measure of physical activity in prechool children. *Journal of Physical Activity and Health*, 2 (3), 345-357.
- Sirard, J.R., & Slater, M.E. (2009). Compliance with wearing physical activity accelerometers in high school students. *Journal of Physical Activity and Health*, 6 (1s), s148 – 155.
- Sjoberg, R.L., Nilsson, K.W., & Leppert, J. (2005). Obesity, shame and depression in school-aged children: a population-based study. *Pediatrics*, 116 (3). E389 - 392 doi: 10.1542/peds.2005-0170
- Slootmaker, S.M., Schuit, A.J., Chinapaw, M.J., Seidell, J.C., & Van Nechelen, W. (2009). Disagreement in physical activity assessed by accelerometer and self-report in subgroups of age, gender, education and weight status. *International journal of Behavioural Nutrition and Physical activity*, 6, doi: 10.1186/1479-5868-6-17
- Smucker, M. R., Caighead, E. Craighead, L.W., & Green, B.J. (1986). Normative and reliability data for the children's depression inventory. *Journal of Abnormal Child Psychology*, 14, 25 – 39.
- Speakman, J.R. (1998). The history and the theory of doubly labeled water technique. *American Journal of Clinical Nutrition*, 68 (Supplement), s923 – s928.
- Spence, J.C., McGannon, K.R., & Poon, P. (2005). The effect of exercise on global self-esteem. *Journal of Sport and Exercise Psychology*, 27, 311 – 334.
- Spielberger, C. (1973). *Manual for the State - Trait Anxiety Inventory for Children*. Palo Alto, CA: Consulting Psychologists.
- Spielberger, C.D., Gorsuch, R.L., & Lushene, R.E. (1970). *STAI manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Spitzer, R.L., Williams, J.B., Gibbon, M., & First, M.B (1992). The Structured Clinical Interview for DSM-III-R (SCID). *Archives of General Psychiatry*, 49, 624–629.
- Sonstroem, R.J. (1997a). Physical activity and self-esteem. In W.P. Morgan (eds), *Physical activity and Mental Health* (p127 – 143). Washington, DC: Taylor & Francis.
- Sonstroem, R.J. (1997b). The physical self-system: A mediator of exercise and self-esteem. In K.R.Fox (eds). *The physical self: From motivation to well-being* (p 3 – 26). Champaign, IL: Human Kinetics.
- Sonstroem, R.J., & Morgan, W.P. (1989) Exercise and self-esteem: rationale and model. *Medicine and Science in Sport and Exercise*, 21 (3), 329 – 337.

- Sowislo, J.F., & Orth, U. (2013). Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. *Psychological Bulletin*, 139 (1), 213 – 240. doi: 10.1037/a0028931
- Stamatakis, E., Zaninotto, P., Falaschetti, E., Mindell, J., & Head, J. (2010). Time trends in childhood and adolescent obesity in England from 1995 to 2007 and projections of prevalence to 2015. *Journal of Epidemiology and Community Health* 64 (2). doi: 10.1136/jech.2009.098723
- Starling, R.D. (2002). Use of Doubly labelled water and indirect calorimetry to assess physical activity. In G. Welk (eds). *Physical Activity Assessments for Health – Related Research* (pp.197 – 208). Champagne, IL: Human kinetics
- Sterne, J.A.C., & Davey-Smith, G. (2001). Sifting the evidence – what’s wrong with significance tests? *British Medical Journal*, 322 (7280), 226 – 231.
- Steptoe, A., & Butler, N. (1996). Sport participant and emotional well-being in adolescents. *The Lancet*, 347, 1789 – 1792.
- Steptoe, A., Wardle, J., Fuller, R., Holte, A., Justo, J., Sanderamn, R., & Wichstrøm, L. (1997). Leisure-time physical exercise: Prevalence, attitudinal correlates, and behavioural correlates among young Europeans from 21 countries. *Preventive Medicine*, 26 (6), 845 – 854.
- Stevens, J. (2002). Repeated measures analysis, Chapter 13. *Applied Multivariate Statistics for the Social Sciences*, 4th ed (p. 520–521). New Jersey, Lawrence Erlbaum Associates
- Stewart-Brown, S., & Janmohamed, K. (2008). Warwick-Edinburgh Mental Well-being scales (WEMWBS) User guide.
- Stone, L. L., Otten, R., Engles, R.C., Vermulst, A.A., & Janssens, J.M. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: a review. *Clinical Child and Family Psychology Review*, 13 (3), 254 – 274. doi: 10.1007/s10567-010-0071-2
- Stone, M.R., Rowlands, A.V., & Eston, R.G. (2009). Relationships between accelerometer-assessed physical activity and health in children: impact of the activity-intensity classification method. *Journal of Sports Science and Medicine*, 8, 136-143.
- Stracker, L., & Campbell, A. (2012). Translation equations to compare ActiGraph GT3X and Actical accelerometers activity counts. *BMC Medical research Methodology*, 12, doi: 10.1186/1471-2288-12-54
- Strawbridge, W.J., Deleger, S., Roberts, R.E., & Kaplan, G.A. (2002). Physical activity reduces the risk of subsequent depression in older adults. *American Journal of Epidemiology*, 156, 328 – 334.

- Streiner, D.L. (2003). Starting at the beginning: an introduction of coefficient Alpha and internal consistency. *Journal of personality Assessment*, 80 (1), 99 – 103.
- Studenmayer, 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. doi: 10.1152/jappphysiol.00465.2009
- Sund, A.M., Larsson, B., & Wichstrøm, L. (2011). Role of Physical and sedentary activities in the development of depressive symptoms in early adolescence. *Social Psychiatry and Psychiatric Epidemiology*, 46 (5), 431 – 441. doi: 10.1007/s00127-010-0208-0
- Sutton-Hamilton, S. & Armando, J. (2008). Oppositional Defiant Disorder. *American Family Physician*, 78 (7), 861 – 866.
- Swain, D.P., Leutholtz, B.C., King, M.E., Haas, L.A., & Branch, J.D. (1998). Relationship between % heart rate reserve and % VO₂ reserve in treadmill exercise. *Medicine and Science in Sports and Exercise*, 30 (2), 318 – 321.
- Taylor, A., (2000). Physical activity, anxiety and stress. In S. Biddle., K.Fox., & S.H. Boutcher. (Ed). *Physical Activity and Psychological Well-Being*. (pp. 10 - 45). London, Uk: Routledge
- Taylor, E., Dopfner, M., Sergeant, J., Asherson, P., Banaschewski, T., Buitelaar, J., & Zuddas, A. (2004). European clinical guidelines for Hyperkinetic disorder – first upgrade. *European Child Adolescent Psychiatry*. 13(s1), 1/7 – 1/30.
- Telford, A., Salmon, J., Jolley, D. & Crawford. (2004). Reliability and validity of physical activity questionnaires for children: The children's Leisure activities study survey (CLASS). *Pediatric Exercise Science*, 16, 64 – 78.
- Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J., & Stewart-Brown, S. (2007). The warwick-edinburgh mental well-being scale (WEMWBS): development and UK validation. *Health and Quality of Life Outcomes*, 5, 63. doi:10.1186/1477-7525-5-63
- Teychenne, M., Ball, K., & Salmon, J. (2008). Physical activity and likelihood of depression in adults: A review. *Preventative Medicine*, 46, 397 – 411.
- The Mental health foundation (2005): *Lifetime Impacts: Childhood and Adolescent Mental Health, Understanding The Lifetime Impacts*.
- Thompson, P.D., Buchner, D., Pina, I.L., Balady, G.J., Williams, M.A., Marcus, B.H., Berra, K...& Wenger, N.K. (2003). Exercise and Physical Activity in the Prevention and Treatment of Atherosclerotic Cardiovascular Disease. *Circulation*, 107, 3109 – 3116. doi: 10.1161/01.CIR.0000075572.40158.77

- Tobias, J.H., Steer, C.D., Mattocks, C.G., Riddoch, C., & Ness, A.R. (2007). Habitual levels of physical activity influence bone mass in 11-year-old children from the United Kingdom: findings from a large population-based cohort. *Journal of Bone Mineral Research*, 22 (1), 101–109. doi:10.1359/jbmr.060913
- Tomson, L.M., Pangrazi, R.P., Friedman, G., & Hutchinson, N. (2003). Childhood depressive symptoms, physical activity and health related fitness. *Journal of Sport & Exercise Psychology*, 25 (4), 419 – 439.
- Torres, E.R., Sampsel, C.M., Ronis, D.L., Neighbors, H.W., & Gretebeck, K.A. (2013). Leisure time physical activity in relation to depressive symptoms in African-americans: Results from the National survey of American life. *Preventive Medicine*, 56 (6), 410 - 412 doi: 10.1016/j.ypmed.2013.02.013
- Tremblay, M.S., & Williams, J.D. (2003). Is the Canadian childhood obesity epidemic related to physical activity? *International Journal of Obesity*, 27 (9), 111 – 1105.
- Treuth, M.S., Schmitz, K., Catellier, D.J., McMurray, R.G., Murray, D.M., Almeida, M.J., Going, S., Norman, J.E., & Pate, R. (2004). Defining accelerometer thresholds for activity intensities in adolescent girls. *Medicine and Science in Sport and Exercise*, 36, 1259 – 1266.
- Troiano, R., Berrigan, D., Dodd, K., Masse, L., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise*; 40, 181-8. doi: 10.1249/mss.0b013e31815a51b3
- Trost, S.G., Loprinzi, P.D., Moore, R., & Pfeiffer, K.A. (2011). Comparison of accelerometer cut-points for predicting activity intensity in youth. *Medicine and Science in Sport and Exercise*, 43 (7), 1360 – 1368. doi: 10.1249/MSS.0b013e318206476e
- Trost, S., McIver, K.L., & Pate, R.R. (2005). Conducting accelerometer-based activity assessments in field-based research. *Medicine and Science in Sport and Exercise*, 37 (11)s0, s531 – s543. doi: 10.1249/01.mss.0000185657.86065.98
- Trost, S., Pate, R., Freedson, P., Sallis, J., & Taylor, W. (2000). Using Objective physical activity measures with youth: How many days of monitoring are needed? *Medicine and Science in Sports and Exercise*, 32, 426 – 431.
- Trost, S.G., Pate, R.R., Sallis, J.F., Freedson, P.S., Taylor, W.C., Dowda, M., & Sirard, J.(2002). Age and gender differences in objectively measured physical activity in youth. *Medicine and Science in Sport and Exercise*, 34, 350 – 355. doi:10.1097/00005768-200202000-00025
- 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 Sport and Exercise*, 30 (4), 629 – 633.

- Trucker, P., & Gilland, J. (2007). The effect of season and weather on physical activity: a systematic review. *Public Health*, 121 (12), 909 – 22.
- Tudor-Locke, C., Craig, C.L., Beets, M.W., Belton, S. Cardon, G.M. Duncan, S., Hatano, Y... & Blair, S.N. (2011). How many steps/days are enough? For children and adolescents. *International Journal of behavioural Nutrition and Physical Activity*, 28 (8), 78 doi: 10.1186/1479-5868-8-78
- Tudor-Locke, C., McClain, J.J., Hart, T.L., Sisson, S.B., & Washington, T.L. (2009). Pedometry methods for assessing free-living youth. *Research Quarterly for Exercise and Sport*, 80 (2), 175 – 184.
- Vale, S., Santos, R., Silva, P., Soares-Miranda, L., & Mota, J. (2009). Preschool children physical activity measurement: importance of Epoch length choice. *Pediatric Exercise Science*, 21, 413 – 420.
- Van Cauwenberghe, E., Labarque, V., Trost, S.G., de Bourdeaudhuij, I., & Cardon, G. (2011). Calibration and comparison of accelerometer cut-points in pre-school children. *International Journal of Pediatric Obesity*, 6, e582-589. doi:10.3109/17477166.2010.526223
- Van Coevering, P., Harnack, L., Schmitz, K., Fulton, J.E., Galuska, D.A., & Gao, S. (2005). Feasibility of using accelerometers to measure physical activity in Young adolescents. *Medicine and Science in Sport and Exercise*, 37, 867 – 871. doi: 10.1249/01.MSS.0000162694.66799.FE
- Van Der Horst, K., Paw, M.J.C.A., Twisk, J.W.R., & Van Mechelen, W. (2007). A brief review on correlates of physical activity and sedentariness in youth. *Medicine and Science in Sport and Exercise*, 39 (8), 1241 – 1250. doi: 10.1249/mss.0b013e318059bf35
- VanHees, L., Lefevre, J., Philippaerts, R., Martens, M., Huygens, W., Troosters, T., & Beunen, G. (2005). How to assess physical activity? How to assess physical fitness. *European Journal of Cardiovascular Prevention and Rehabilitation*, 12, 102 – 114. doi: 10.1097/01.hjr.0000161551.73095.9c
- Van Hees, V.T., Renström, F., Wright, A., Gradmark, A., Catt, M., Chen, K.Y. Löff, M. ...& Franks, P.W. (2011). Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. *PLoS One*, 6 (7), e22922. doi:10. 1371/journal.pone.0022922
- Vanhelst, J., Beghin, L., Rasoamanana, P., Theunynck, D., Meskini, T. , Lliescu, C., Duhamel, A., Turck, D., & Gottrand, F. (2010) Calibration of the RT3 accelerometer for various patterns of physical activity in children and adolescents. *Journal of Sports Sciences*, 28 (4), 381– 387 doi: 10.1080/02640410903508821

- Vanhelst, J., Beghin, L., Duhamel, A., Bergman, P., Sjöström, 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 Medical Research Methodology*, 12. doi: 10.1186/1471-2288-12-26
- Verbestel, V., Van Cauwenberghe, E., De Coen, V., Maes, L., de Bourdeaudhuij, I., & Cardon, G. (2011). Within- and between-day variability of objectively measured physical activity in preschoolers. *Pediatric Exercise Science*, 23 (3), 366 – 378.
- Vincent, W.L. (2005). *Statistics in Kinesiology*, Champaign, IL; Human Kinetics.
- Ward, D.S., Evenson, K.R., Vaughn, A., Rodgers, A.B., & Troiano, R.P (2005). Accelerometer use in physical activity: best practices and research recommendations. *Medicine and Science in Sport and Exercise*, 37 (11s), s582 – 588. doi: 10.1249/01.mss.0000185292.71933.91
- Wardle, J., & Cooke, L., (2005). The impact of obesity on psychological well-being. *Best Practice & Research Clinical Endocrinology & Metabolism*, 19 (3), 421 – 440. doi:10.1016/j.beem.2005.04.006
- Warr, P. (2012). How to think about and measure psychological well-being. In M.Wang, R.R. Sinclair & L.E. Tetrick (eds). *Research Methods in Occupational Health Psychology*. New York: Psychology Press/Routledge.
- Warren, J.M., Ekelund, U., Besson, H., Mezzani, A., Geladas, N. & Vanhees, L. (2010). Assessment of physical activity – a review of methodologies with reference to epidemiological research. *European Journal of Cardiovascular Prevention and Rehabilitation*, 17 (2), 127 – 139. doi:10.1097/HJR.0b013e32832ed875.
- Waterman, A.S. (1993). Two concepts of happiness: Contrasts of personal expressiveness (eudiamonia) and hedonic enjoyment. *Journal of Personality and Social Psychology*, 64 (4), 678 – 691.
- Waters, E., Stewart-Brown, S., & Fitzpatrick, R. (2003). Agreement between adolescent self-report and parent reports of health and well-being: results of an epidemiological study. *Child: Care, Health and Development*, 29 (6), 501 – 509. doi: 10.1046/j.1365-2214.2003.00370.x
- Watkinson, C., van Sluijs, E.M.F., Sutton, S., Hardeman, W., Corder, K., & Griffin, S.J. (2010). Overestimation of physical activity level is associated with lower BMI: a cross-sectional analysis. *International Journal of Behavioural Nutrition and Physical Activity*, 7, doi: 10.1186/1479-5868-7-68.
- Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin*, 96, 465–490.

- Watson, D., Clark, L.A., & Tellegen, A. (1988). Development and validation of Brief measures of Positive and Negative Affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54 (6), 1063 – 1070.
- Watson, D., Suls, J., & Haig, J. (2002). Global self-esteem in relation to structural models of personality and affectivity. *Journal of Personality and Social Psychology*, 83(1), 185 – 197. doi: 10.1037/0022-3514.83.1.185
- Webb, O.J., Benjamin, C.C., Gammon, C., McKee, H.C., & Biddle, S.J.H. (2013). Physical activity, sedentary behaviour and physical self-perceptions in adolescents girls: A mediation analysis. *Mental Health and Physical Activity*, 6, 24 – 29.
- Weissman, M.M., & Klerman, G.L. (1992). DEPRESSION: Current Understanding and Changing Trends. *Annual Review of Public Health*, 13, 319 – 339. doi: 10.1146/annurev.pu.13.050192.001535
- Welk, G., (2002). *Physical Activity Assessments for Health – Related Research*. Champagne, IL: Human kinetics
- Welk, G. (2005). Principles of design and analysis for the calibration of accelerometry-based activity monitors. *Medicine and Science in Sport and Exercise*, 37 (11suppl), s501 – s511. doi:10.1249/01.mss.0000185660.38335.de
- Welk, G., & Corbin, C.B. (1995). The validity of the Tritrac-R3D Activity Monitor for the assessment of physical activity in children. *Research Quarterly for Exercise and Sport*, 66 (3), 202 – 209.
- Welk, G., McClain, J., & Ainsworth, B.E. (2012). Protocols for evaluating equivalency of accelerometry-based activity monitors. *Medicine and Science in Sport and Exercise*, 44 (1s), s39 – s49. doi: 10.1249/MSS.0b013e3182399d8f
- Whaley, M.H., Burbaker, P.H., & Otto, R.M. (Eds). (2006) ACSM's Guidelines for exercise testing and prescription (7th ed). Baltimore, ML; Lippincott, Williams & Wilkin.
- Whitehead, J. R. (1995). A study of children's physical self-perceptions using an adapted physical self-perception profile questionnaire. *Pediatric Exercise Science* 7, 132 – 151.
- Whitehead, J.R., & Corbin, C.B. (1991). Effects of Fitness test type, teacher and gender on exercise intrinsic motivation and physical self-worth. *Journal of School Health*, 61 (1), 11 – 16. doi: 10.1111/j.1746-1561.1991.tb07850.x
- Wild, L.G., Flisher, A.J., Bhana, A., & Lombard, C. (2004). Associations among adolescent rick behaviours and self-esteem in six domains. *Journal of Child Psychology and Psychiatry*, 45 (8), 1454 – 67. Doi: 10.1111/j.1469-7610.2004.00330.x

- Wiles, N.J., Jones, G.T., Haase, A.M., Lawlor, D.A., Macfarlane, G.J. & Lewis, G. (2008). Physical activity and emotional problems amongst adolescents. *Social Psychiatry & Psychiatric epidemiology*, 43, 765 – 772. DOI 10.1007/s00127-008-0362-9
- Willmore, J., & Costill, D. (2005). *Physiology of Sport and Exercise (3rd ed)*. Champagne, IL: Human Kinetics
- Winefield, H.R., Gill, T.K., Taylor, A.W., & Pilkington, R.M. (2012). Psychological well-being and psychological distress: is it necessary to measure both? *Psychology of Well-Being, Theory, Research and Practice*, 2 (3). doi: 10.1186/2211-1522-2-3.
- Wise, L.A., Adams – Campbell, L.L., Palmer, J.R., & Rosenberg, L. (2006). # Leisure Time Physical Activity in Relation to Depressive Symptoms in the Black Women’s Health Study. *Annals of Behavioural Medicine*, 32 (1), 68 – 76. doi: 10.1207/s15324796abm3201_8
- Wisocki, P.A., Handen, B., & Morse, C.K. (1986). The Worry Scale as a measure of anxiety among home bound and community active elderly. *The Behaviour Therapist*, 5, 91–95.
- World Health Organization (2001d). *The World Health Report. Mental Health: New Understanding, New Hope*. Geneva: World Health Organization
- Ussher, M.H., Owen, C.G., Cook, D.G., & Whincup, P.H. (2007). The relationship between physical activity, sedentary behaviour and *Psychiatric Epidemiology*, 42 (10), 851 – 856. doi:10.1007/s00127-007-0232-x
- Yang, C.C., & Hsu, Y.L. (2010). A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors*, 10 (8), 7772 – 1188. doi: 10.3390/s100807772
- Yildirim, M., Veloigne, M., de Bourdeaudhuij, I., Androutsos, O., Manios, Y., Felso, R. Kovács, E... & Chinapaw, M.J.M. (2011). Study Protocol of physical activity and sedentary behaviour measurement among school children by accelerometry – cross-sectional survey as part of the ENERGY – project. *BMC Public Health*, 11, 182 doi: 10.1186/1471-2458-11-182
- Zhang, S., Rowlands, A.V., Murray, P., & Hurst, T.L. (2012). Physical activity classification using the GENEA wrist-worn Accelerometer. *Medicine and Science in Sport and Exercise Science*, 44 (4), 742 – 748. doi: 10.1249/MSS.0b013e31823bf95c
- Zhang, K., Werner, P., Sun, M., Pi-Sunyer, F.X., & Boozer, C.N. (2003). Measurement of human daily physical activity. *Obesity Research*, 11 (1), 33-40.

Appendix

	Page
1. Ethics approval	
1.1. Ethics approval for chapters 4, 5 and 6.....	294
1.2. Ethics approval for chapter 7	285
1.3. Ethics approval for chapter 8.....	286
2. Informed consent	
2.1. Parent information sheets and consent pertaining to chapters 4, 5 and 6.....	297
2.2. Child information sheets and consent pertaining to chapters 4, 5 and 6	302
2.3. Parent information sheets and consent pertaining to chapter 7.....	307
2.4. Child information sheets and consent pertaining to chapter 7.....	312
2.5. Parent information sheets and consent pertaining to chapter 8	316
2.6. Child information sheets and consent pertaining to chapter 8	319
3. Questionnaires	
3.1. Child Depression Inventory (CDI)	322
3.2. State-Trait Anxiety Inventory for Children (STAIC).....	325
3.3. Physical self-perception profile for children (PSPP-CY), PSW & GSW.....	326
3.4. Strengths and Difficulties questionnaire (SDQ).....	330
4. Further analysis	
4.1. Worked example of correlated correlation coefficients (chapter four)	331
4.2. Worked example of correlated correlation coefficients (Chapter seven).....	332
4.3. Regression tables (chapter five).....	333
4.4 Regression tables (chapter seven).....	334
4.5. Correlation tables (chapter eight).....	335
5. Gantt chart for completion	336
6. Screen view for GENEActiv macro	337

Appendix 1.1. Ethics approval chapters four, five and six



SPORT AND HEALTH SCIENCES

College of Life and Environmental Sciences
St Luke's Campus
Heavitree Road
Exeter UK EX1 2LU

t +44 (0) 1392 722807 (General Enquiries)
+44 (0) 1392 722896 (Education Enquiries)
f +44 (0) 1392 724726
e sshs-school-office@exeter.ac.uk
w sshs.exeter.ac.uk

Certificate of Ethical Approval

Title: Proposal 5 (20/10/10) "The effect of activity intensity on children's psychological well being"

Applicants: Dr Ann Rowlands, Dr Gaynor Parfitt, Ms Lisa Phillips (PhD Researcher)

The proposal was reviewed by a Representative on the Committee.

Decision: The proposal was approved from January 2011 to July 2011

Signature:

A handwritten signature in blue ink, appearing to be 'D. Wilkerson'.

Date: 13-05-13

Name/Title of Ethics Committee Reviewer: Dr. D. Wilkerson

Your attention is drawn to the attached paper which reminds the researcher of information that needs to be observed when Ethics Committee approval is given.

Appendix 1.2. Ethics approval for chapter 7



SPORT AND HEALTH SCIENCES

College of Life and Environmental Sciences
St Luke's Campus
Heavitree Road
Exeter UK EX1 2LU

t +44 (0) 1392 722807 (General Enquiries)
+44 (0) 1392 722896 (Education Enquiries)
f +44 (0) 1392 724726
e sshs-school-office@exeter.ac.uk
w sshs.exeter.ac.uk

Certificate of Ethical Approval

Title: Proposal 2 (21/10/09) "Calibration and validation of accelerometers with children"
Applicants: Dr Ann Rowlands, Dr Gaynor Parfitt with Ms Lisa Phillips (PhD Researcher)
and Ms Charlotte Benjamin (PhD Researcher)

The proposal was reviewed by a Representative on the Committee.

Decision: The proposal was approved from October 2009 to July 2011

Signature:

A handwritten signature in blue ink, appearing to read 'D. Wilkerson'.

Date:

13-05-13

Name/Title of Ethics Committee Reviewer: Dr. D. Wilkerson

Your attention is drawn to the attached paper which reminds the researcher of information that needs to be observed when Ethics Committee approval is given.

Appendix 1.3. Ethics approval for chapter eight



SPORT AND HEALTH SCIENCES

College of Life and Environmental Sciences
St Luke's Campus
Heavitree Road
Exeter UK EX1 2LU

t +44 (0) 1392 722807
f +44 (0) 1392 724726
e sshs-school-office@exeter.ac.uk
w sshs.exeter.ac.uk

Certificate of Ethical Approval

Title: "The use of a wrist worn activity monitor to assess the relationship between Physical activity and psychological well being"

Applicants: Lisa Phillips (PGR student), Dr Melvyn Hillsdon, Dr Ann Rowlands, Rebecca Calver (MSc Student)

The proposal was reviewed by a Representative on the Committee.

Decision: The proposal was approved from September 2011 to December 2012

Signature: 

Date: 03-07-12

Name/Title of Ethics Committee Reviewer: Dr. D. Wilkerson

Your attention is drawn to the attached paper which reminds the researcher of information that needs to be observed when Ethics Committee approval is given.

Appendix 2.1. Parental information for intervention schools, control school and consent for chapters four, five and six



School of Sport and Health Sciences
St Luke's Campus
Heavitree Road
EXETER, EX1 2LU
General Office +44 (0)1392 262896
Fax +44(0)1392 264706

January 2011

Information Sheet

Dear Parent/ Guardian,

We are inviting the participation of the children in year 5, to take part in a PhD research study looking at Physical Activity and Health. We welcome all activity levels from sedentary to very active and all fitness levels as well.

What is the purpose of the study?

Following on from previous research that has shown physical activity to be beneficial to children's psychological well-being (such as improving their self-esteem and happiness) we aim to pinpoint which physical activity intensity (medium / high) will have the most beneficial effect.

What will my son/daughter be asked to do?

The study will be split into three sections.

Section A will be used to gain baseline measures of the children's current physical activity. The children will be asked to wear a physical activity monitor for a week. They will also be asked to complete a number of questionnaires, assessing how they feel about themselves. Children will also have their height and weight measured. These measures will be taken individually and will be kept confidential.

Section B will last for 8 weeks, during which your child will undertake 3 physical activity sessions each week. Each activity session will last no more than 30minutes. During this time they may be asked to wear the physical activity monitors again for a week. They will also be asked to fill in the questionnaires again.

Section C will be a repeat of section A, it is used to assess whether there have been any changes in how you child feels and the amount of activity they do.

What will happen with the information?

All information obtained will be stored on computer in coded form and individual results will be confidential to the participant and the research team. Results of this study may be published but any data included will in no way be linked to any specific participant. It is important to note that you may withdraw your child's participation in the study at any time without any disadvantage to you of any kind.

If you, or your child, have any further queries regarding the above study please do not hesitate to contact us

The Ethics Committee of the School of Sport and Health Sciences has reviewed and approved this study.

Yours Sincerely,

Lisa Phillips

School of Sport and Health Sciences
University of Exeter
St. Luke's Campus
Heavitree Road
Exeter, EX1 2LU
Email: lnsp201@exeter.ac.uk
Tel: 07957356209

Additional contacts:

Dr. Ann Rowlands
Tel: 01392 262878
Email: A.V.Rowlands@exeter.ac.uk

Assoc. Prof Gaynor Parfitt

Email: C.G.Parfitt@exeter.ac.uk



School of Sport and Health Sciences
 St Luke's Campus
 Heavitree Road
 EXETER, EX1 2LU
 General Office +44 (0)1392 262896
 Fax +44 (0)1392 264706

January 2011

Information Sheet

Dear Parent/ Guardian,

We are inviting the participation of the children in year 5, to take part in a PhD research study looking at Physical Activity and Health. We welcome all activity levels from sedentary to very active and all fitness levels as well.

What is the purpose of the study?

Following on from previous research that has shown physical activity to be beneficial to children's psychological well-being (such as improving their self-esteem and happiness) we aim to assess whether providing children with information about health living and leading active lifestyles will increase their activity and therefore effect how they feel.

What will my son/daughter be asked to do?

The study will be split into three sections.

Section A will be used to gain baseline measures of the children's current physical activity. The children will be asked to wear a physical activity monitor for a week. They will also be asked to complete a number of questionnaires, assessing how they feel about themselves. Children will also have their height and weight measured. These measures will be taken individually and will be kept confidential.

Section B involves your child attending a talk about the benefits of leading a healthy lifestyle. The talk should last no more than 30minutes, where you child will gain ideas about how to be more active during everyday life and the benefits that this could bring both physically and psychologically.

Section C will be a repeat of section A and will be conducted 4 weeks and 9 weeks after the initial assessment. This section is used to assess whether there have been any changes in how you child feels and the amount of activity they do.

What will happen with the information?

All information obtained will be stored on computer in coded form and individual results will be confidential to the participant and the research team. Results of this study may be published but any data included will in no way be linked to any specific participant. It is important to note that you may withdraw your child's participation in the study at any time without any disadvantage to you of any kind.

If you, or your child, have any further queries regarding the above study please do not hesitate to contact us.

The Ethics Committee of the School of Sport and Health Sciences has reviewed and approved this study.

Yours Sincerely,

Lisa Phillips

School of Sport and Health Sciences
University of Exeter
St. Luke's Campus
Heavitree Road
Exeter, EX1 2LU
Email: lrsp201@exeter.ac.uk
Tel: 07957356209

Additional contacts:

Dr. Ann Rowlands
Tel: 01392 262878
Email: A.V.Rowlands@exeter.ac.uk

Assoc. Prof Gaynor Parfitt

Email: C.G.Parfitt@exeter.ac.uk



Consent Form for parents

Please complete and sign if you **DO** wish your son/daughter to take part in this study.

I have read the information sheet concerning this study and understand what it is about. All my questions have been answered to my satisfaction. I understand my son/daughter will have their physical activity measured at varying points throughout the study, and will be required to complete questionnaires about their feelings. I understand that I am free to request further information at any stage.

I know that:

1. My son/daughters' participation in the study is entirely voluntary.
2. My son/daughter will be required to take part in physical activity sessions that may be of a high intensity.
3. My son/daughter will be required to answer questions about their feelings.
4. My son/daughter will have his/her height and weight measured
5. My son/daughter is free to withdraw from the study at any time without giving reason.
6. The results will be stored on computer in coded form and individual results will be confidential.
7. The results of the study may be published but my son/daughter's anonymity will be preserved.

Please sign and return the slip below if you consent to your son/daughter taking part in the study.

Signed Parent/ Guardian / Care giver.....

Date.....

Name of son/daughter (please print).....

Name parent/ guardian/ care giver

Contact telephone number(s).....

Contact email(s).....

This study has been reviewed and approved by the Ethics Committee of the School of Sport and Health Sciences.

Do we have permission to contact your son/daughter for further testing at a later date?

YES NO

Appendix 2.2. Child information and consent for chapters four, five and six



School of Sport and Health Sciences
 St Luke's Campus
 Heavitree Road
 EXETER, EX1 2LU
 General Office +44 (0)1392 262896
 Fax +44

January 2011
 (0)1392 264706

Information Sheet

Dear Pupil,

Thank you for showing an interest in taking part in this study. This sheet will tell you a bit more about the study and what we would like you to do. If you decide not to take part it will not change your relationship with the research team or your school.

What is the project about?

We are inviting pupils, both male and female, in year 5 to take part in a Physical Activity and Health research study. Everyone is welcome to take part in the study whether you do lots of activity and exercise or not much at all.

What do we want to find out?

We would like to know how the exercise you will be doing makes you feel about your self, and if it affects how you feel generally.

What do I have to do?

The study will be split into 3 sections

- Section A
 - This will last for one week, during which time you will be asked to wear a gadget that monitors how much activity you do.
 - You will also be asked to fill in some questionnaires about how you feel.
 - You will also have your height and weight measured (this will be done individually).
- Section B
 - This part will last for 8 weeks, where you will take part in a physical activity session 3 times a week.
 - Each activity session should last no more than 30minutes.
 - During this section, you may be asked to wear the activity monitor and fill out the questionnaires again.
- Section C
 - This is similar to section A; you will be asked to wear the gadgets again and to fill in the questionnaires about how you feel. And measurements taken again.

Can I change my mind?

Yes! You can stop the study at any time without having to give a reason.

What will we do with the information?

All the information collected will be stored on a computer and the results will be confidential to the University research team.

What if I have any questions?

If you have any questions then please feel free to ask any of the researchers at any time.

What do I have to do next?

If you have read and understood everything that we want you to do and are happy to take part please sign the consent form, which is attached to this sheet. Your parent/guardian must also sign the form we have given to them.

The Ethics Committee of the School of Sport and Health Sciences has reviewed and approved this study.

Thank you,

Lisa Phillips

School of Sport and Health Sciences
University of Exeter
St. Luke's Campus
Heavitree Road
Exeter, EX1 2LU
Email: lrsp201@exeter.ac.uk
Tel: 07957356209

Additional contacts:

Dr. Ann Rowlands
Tel: 01392 262878
Email: A.V.Rowlands@exeter.ac.uk

Dr. Gaynor Parfitt

Email: C.G.Parfitt@exeter.ac.uk



School of Sport and Health Sciences
 St Luke's Campus
 Heavitree Road
 EXETER, EX1 2LU
 General Office +44 (0)1392 262896
 Fax +44

January 2011
 (0)1392 264706

Information Sheet

Dear Pupil,

Thank you for showing an interest in taking part in this study. This sheet will tell you a bit more about the study and what we would like you to do. If you decide not to take part it will not change your relationship with the research team or your school.

What is the project about?

We are inviting pupils, both boys and girls, in years 5 to take part in a Physical Activity and Health research study. Everyone is welcome to take part in the study whether you do lots of activity and exercise or not much at all.

What do we want to find out?

We would like to know how the exercise you will be doing makes you feel about your self, and if it affects how you feel generally.

What do I have to do?

The study will be split into 3 sections

- Section A
 - This will last for one week, during which time you will be asked to wear a small gadget that monitors how much activity you do.
 - You will also be asked to fill in some questionnaires about how you feel
 - You will also have your height and weight measured (this will be done individually).
- Section B
- You will be asked to listen to a talk about the benefits of leading a healthy lifestyle. The talk should last no more than 30minutes, will give you ideas about how to be more active during everyday life and how this can help your body and mind feel better.
- Section C
 - This is similar to section A; you will be asked to wear the gadgets again and to fill in the questionnaires about how you feel and take measurements again.

Can I change my mind?

Yes! You can stop the study at any time without having to give a reason.

What will we do with the information?

All the information collected will be stored on a computer and the results will be confidential to the University research team.

What if I have any questions?

If you have any questions then please feel free to ask any of the researchers at any time.

What do I have to do next?

If you have read and understood everything that we want you to do and are happy then no further action is required, the research will be in touch with your school to arrange a start date for the study. However, if you decide that you do not wish to take part, please sign the opt out form which is attached to this sheet.

The Ethics Committee of the School of Sport and Health Sciences has reviewed and approved this study.

Thank you,

Lisa Phillips

School of Sport and Health Sciences
University of Exeter
St. Luke's Campus
Heavitree Road
Exeter, EX1 2LU
Email: Insp201@exeter.ac.uk
Tel: 07957356209

Additional contacts:

Dr. Ann Rowlands
Tel: 01392 262878
Email: A.V.Rowlands@exeter.ac.uk

Associate Professor. Gaynor Parfitt
Email: C.G.Parfitt@exeter.ac.uk



Consent Form for pupils

Please complete and sign if you **DO** to take part in this study.

I have read the information sheet concerning this study and understand what it is about and all my questions have been answered. I understand that I will have my physical activity measured at some points during the study, and I will be asked to fill in some questionnaires about how I feel. I understand that I am free to request further information at any point.

I know that:

1. My participation in the study is entirely voluntary.
2. I am free to withdraw from the study at any time without giving reason.
3. I will be asked to answer questions about my feelings
4. I may be required to take part in high intensity physical activity sessions
5. The results will be stored on computer in coded form and my name will not be included.
6. The results of the study may be published but my personal identity will not be known.

If you **DO** wish to take part in this study please sign below and return this form to the researcher.

Signed Date.....

Name (please print).....

Name parent/ guardian/ care giver

Contact telephone number(s).....

This study has been reviewed and approved by the Ethics Committee of the School of Sport and Health Sciences.

Do we have permission to contact you for further testing at a later date? YES NO

Appendix 2.3. Parent information and consent forms for chapter seven



School of Sport and Health Sciences
St Luke's Campus
Heavitree Road
EXETER, EX1 2LU
General Office +44 (0)1392 262896
Fax +44

July 2010
(0)1392 264706

Information Sheet

Dear School/ Parent/ Guardian,

We are inviting the participation of 60 children, both male and female, aged 8 - 14 years to participate in a Physical Activity and Health research study. We welcome all activity levels from sedentary to very active and all fitness levels as well.

What is the purpose of the study?

An accurate measure of physical activity is necessary for determining the impact of physical activity on health. The main purpose of the study is to test how well a newly designed activity monitor performs when measuring various forms of physical activity. The new activity monitor will be compared to a variety of existing devices.

What will my son/daughter be asked to do?

When your child arrives at the laboratory a researcher will record the following details:

- 1) Full name
- 2) Date of birth
- 3) Sex
- 4) Handedness (left or right)
- 5) Contact information (address, phone number(s), email (s) if applicable)

Next, he/she will be asked to complete two short questionnaires about:

- a) His/her physical activity level
- b) His/her readiness for physical activity

Next, the researcher will take some key body measurements:

- 1) Height
- 2) Seated height
- 3) Weight
- 4) Waist size

Next, your child will be fitted with a series of activity monitors on the wrists, waist, and ankle as well as a heart rate monitor and face mask to measure the air breathed. Then they will be briefed about the physical activity testing protocol which includes up to 11 of the following semi-structured activities: Next your son/daughter will be asked to do the following activities; lying down, watching a DVD, playing a computer game, walking at a slow pace (on a running machine), walking at a fast pace, and then running at two different speeds. Each activity will last for three minutes, apart from lying down (which is for 10) and there will be chance to rest between each activity. The last thing we would like your son/daughter to do is a running test to establish their level of fitness. This is a maximal

test where your son/daughter will run on a treadmill on a gradient until they cannot run anymore. It is a hard test and they may sweat a lot, their heart rate will increase and they will feel tired at the end.

Some participants may be asked to come back to the university to do a few more outdoor activities. If your son/daughter is asked to come back some of the activities they may do will include walking outside, running on the field and kicking a ball around. As with the first session they will be asked to wear the same equipment again (activity monitors, heart rate monitor and the face mask).

Finally, your child's health information will be summarized and returned to you at a later date.

Who can take part?

Due to the nature of the study your son/daughter must be 'apparently healthy' without muscular-skeletal injury.

Can I stay with my son/daughter?

Yes. You are more than welcome to remain with your child throughout the study. However, trained individuals will be present at all times, all of whom have full CRB checks.

What will happen with the information?

All information obtained will be stored on computer in coded form and individual results will be confidential to the participant and the research team. Results of this study may be published but any data included will in no way be linked to any specific participant. It is important to note that you may withdraw your child's participation in the study at any time without any disadvantage to you of any kind. The data will be stored for up to five years so that we can gain as much information from the results as possible. After this time all data will be destroyed.

If you, or your child, have any further queries regarding the above study please do not hesitate to contact us.

The Ethics Committee of the School of Sport and Health Sciences has reviewed and approved this study.

Yours Sincerely,

Charlotte Benjamin and Lisa Phillips

School of Sport and Health Sciences
University of Exeter
St. Luke's Campus
Heavitree Road
Exeter, EX1 2LU
Email: ccb204@exeter.ac.uk, lrsp201@exeter.ac.uk
Tel: Charlotte : 07748282923, Lisa : 07957356209

Additional contacts:

Dr. Ann Rowlands

Tel: 01392 262878

Email: A.V.Rowlands@exeter.ac.uk

Associate Professor Dr. Gaynor Parfitt

Tel: 01392 262869

Email: C.G.Parfitt@exeter.ac.uk



Consent Form for Parents

Please complete and sign if you **DO** wish your child to take part in this study.

I have read the information sheet concerning this study and understand what it is about. All my questions have been answered to my satisfaction. I understand my child will have their general health assessed and their physical activity measured during specific activities of daily living. I understand that I am free to request further information at any stage.

I know that:

1. My son/daughters' participation in the study is entirely voluntary.
2. My son/daughter will be required to perform various activities of daily living while wearing:
 - a. a heart rate monitor
 - b. a series of activity monitors on their wrist, waist, and ankle
 - c. a face mask to allow the measurement of air breathed
3. My son/daughter will be required to have his/her weight, height, sitting height, waist size, and body composition measured
4. My son/daughter is free to withdraw from the study at any time without giving reason.
5. The results will be stored on computer in coded form and individual results will be confidential.
6. The results of the study may be published but my son/daughter's anonymity will be preserved.

If you **DO** wish to take part in this study please sign below and return this form to the researcher.

Signed Parent/ Guardian / care giver.....

Date.....

Name Child (please print).....

Name parent/ guardian/ care giver

Contact telephone number(s).....

Contact email(s).....

This study has been reviewed and approved by the Ethics Committee of the School of Sport and Health Sciences.

Do we have your permission to contact you in the future for follow-up research?

YES NO

I understand that my son/daughter may be asked to return for further testing at a later date?

YES NO

I give permission for my son/daughter's photograph to be taken for the production of a certificate?

YES NO

Appendix 2.4. Child information and consent forms for chapter seven



School of Sport and Health Sciences
St Luke's Campus
Heavitree Road
EXETER, EX1 2LU
General Office +44 (0)1392 262896
Fax +44

July 2010
(0)1392 264706

Information Sheet

Dear Pupil,

Thank you for showing an interest in taking part in this study. This sheet will tell you a bit more about the study and what we would like you to do. If you decide not to take part it will not change your relationship with the research team or your school.

What is the project about?

We are inviting the participation of 60 volunteers, both male and female, aged 8 - 14 years to take part in a Physical Activity and Health research study. Everyone is welcome to take part in the study whether you do lots of activities and exercise or not much at all.

What do we want to find out?

We have two new measuring devices that can be used to measure how much activity you do and we want your help to see how well they work.

What do I have to do?

When you arrive at the laboratory at the University a researcher will ask you the following details:

- a) Full name
- b) Date of birth
- c) Handedness (whether you write with your left or right hand)
- d) Any contact information you may have (address, phone number(s), email(s) if you have one)

Next, you will be asked to complete two short questionnaires about:

- c) Your physical activity level
- d) Your readiness for physical activity

Next, the researcher will take some key body measurements:

- 1 Height
- 2 Seated height
- 3 Weight
- 4 Waist size

After we've got this information from you, you will be fitted with some activity monitors. These will be put on your waist and wrists. As well as wearing these we will ask you to wear a heart rate monitor and a face mask to measure the air you breathe out.

Next you'll be asked to do the following activities; lying down, watching a DVD, playing a computer game, walking at a slow pace (on a running machine), walking at a fast pace, and then running at two different speeds. Each activity will last for three minutes, apart from lying down (which is for 10) and there will be chance to rest between each activity.

The last thing we will ask you to do is a running test to see how fit you are. This is called a maximal test and we are going to see how long you can run on a treadmill for before you get too tired to carry on. It is a hard test and you may find that you sweat a lot during it, your heart will beat faster and you will feel tired at the end of it. But this will be followed by a rest period.

Some of you may be asked to come back to the university to do a few more outdoor activities. If you are asked to come back, some of the activities you may do will include walking outside, running on the field and kicking a ball around. As with the first session you will be asked to wear the same equipment again (activity monitors, heart rate monitor and the face mask).

As a thank you for taking part we will provide you with a certificate with a photograph of you in all the gear!

Who can take part?

Because part of the study involves being able to run for quite a long time on a treadmill you must be able to run for at least ten minutes otherwise we won't be able to work out your fitness level properly. So if you think you can run for that long and don't mind getting a bit sweaty and tired then you can definitely take part!

Can I change my mind?

Yes! You can stop the study at any time without having to give a reason.

What will we do with the information?

All the information collected will be stored on a computer and the results will be confidential to the University research team. We will keep the results we get from you for up to five years so that we can use as many of your results as possible to find out new and interesting things.

What if I have any questions?

If you have any questions then please feel free to ask any of the researchers listed below at any time.

What do I have to do next?

If you have read and understood everything that we want you to do and are happy to take part please sign the consent form, which is attached to this sheet. Your parent/guardian must also sign the form we have given to them.

The Ethics Committee of the School of Sport and Health Sciences has reviewed and approved this study.

Thank you,

Charlotte Benjamin and Lisa Phillips

School of Sport and Health Sciences
University of Exeter
St. Luke's Campus
Heavitree Road
Exeter, EX1 2LU
Email: ccb204@exeter.ac.uk, lrsp201@exeter.ac.uk
Tel: Charlotte : 07748282923, Lisa : 07957356209

Additional contacts:

Dr. Ann Rowlands
Parfitt

Tel: 01392 262878

Email: A.V.Rowlands@exeter.ac.uk

Associate Professor Dr. Gaynor

Tel: 01392 262869

Email: C.G.Parfitt@exeter.ac.uk

Consent Form for Child

Please complete and sign if you **DO** wish to take part in this study.

I have read the information sheet concerning this study and understand what it is about. All my questions have been answered to my satisfaction. I understand that I will have my general health assessed and my physical activity measured during specific activities of daily living. I understand that I am free to request further information at any stage.

I know that:

1. My participation in the study is entirely voluntary.
2. I will be required to perform various activities of daily living while wearing:
 - a. a heart rate monitor
 - b. a series of activity monitors on my wrist, waist, and ankles
 - c. a face mask to allow the measurement of expired air
3. I will be required to have my weight, height, sitting height, waist size, and body composition measured
4. I am free to withdraw from the study at any time without giving reason.
5. The results will be stored on computer in coded form and individual results will be confidential
6. The results of the study may be published but my anonymity will be preserved.

If you **DO** wish to take part in this study please sign below and return this form to the researcher.

Signed

Date.....

Name (please print).....

Contact telephone number(s).....

This study has been reviewed and approved by the Ethics Committee of the School of Sport and Health Sciences.

Do we have your permission to contact you in the future for follow-up research?
YES NO

I understand that I may be asked to complete a further session at a later date.
YES NO

Please tick the box if you would like to receive a certificate with your photograph on

Appendix 2.5. Parent information and consent forms for chapter eight



Sport and Health Sciences
College of Life and Environmental Sciences
St Luke's Campus
Heavitree Road
EXETER, EX1 2LU
General Office +44 (0)1392 262896

January 2012
(0)1392 264706

Fax +44

Information Sheet

Dear Parent/ Guardian,

We are inviting the participation of 128 children, both male and female, aged 8 - 12 years to participate in a Physical Activity and Health research study entitled 'The use of a wrist worn activity monitor to assess the relationship between Physical activity and psychological well being'. We welcome all activity levels from sedentary to very active and all fitness levels as well.

What is the purpose of the study?

An accurate measure of physical activity is necessary for determining the impact of physical activity on health. The main purpose of the study is to test how well a newly designed activity monitor performs when worn in everyday life, and how easy it is for children to wear. A secondary purpose for the study is to see if there is a relationship between the amount of physical activity people do and how they feel about themselves.

What will my son/daughter be asked to do?

At the start of the study, your son/daughter will be asked to provide some basic details:

- 1) Full name
- 2) Date of birth
- 3) Sex
- 4) Contact information (address, phone number(s), email(s) if applicable)

Next, he / she will have their height and weight measured and will be asked to complete a short questionnaire about how they feel about themselves. After that, the researcher will provide your son/daughter with a physical activity monitor (that looks like a watch), which they will be asked to wear for seven consecutive days. The monitor is waterproof and can be worn at all times.

After the seven days, a researcher will return to collect the activity monitors and ask your son/daughter to fill in another short questionnaire asking how comfortable the monitor was to wear and how easy or difficult wearing the monitor was, so as to gain some feedback on possible things to change in the future.

What will happen with the information?

All information obtained will be stored on computer in coded form and individual results will be confidential to the participant and the research team. Results of this study may be published but any data included will in no way be linked to any specific participant. It is important to note that you may withdraw your child's participation in the study at any time without any disadvantage to you of any kind.

If you, or your child, have any further queries regarding the above study please do not hesitate to contact us.

The Sport and Health Sciences Ethics Committee has reviewed and approved this study.

Yours Sincerely,



Lisa Phillips

Sport and Health Sciences
College of Life and Environmental science
University of Exeter
St. Luke's Campus
Heavitree Road
Exeter, EX1 2LU
Email: lrsp201@exeter.ac.uk
Tel: 07957356209

Additional contacts:

Dr. Melvyn Hillsdon
Tel: 01392 722868
Email: M.Hillsdon@exeter.ac.uk



Sport and Health Sciences
 College of Life and Environmental Sciences
 St Luke's Campus
 Heavitree Road
 Exeter, UK, EX1 2LU

Consent Form for parents

Please complete and sign if you **DO** wish your son/daughter to take part in this study.

I have read the information sheet concerning this study and understand what it is about. All my questions have been answered to my satisfaction. I understand my son/daughter will have their physical activity measured for a seven day period. I understand that I am free to request further information at any stage.

I know that:

1. My son/daughters' participation in the study is entirely voluntary.
2. My son/daughter will be required to wear an activity monitor for 7 days
3. My son/daughter will have their height and weight measured
4. My son/daughter will be asked to complete questionnaires about:
 - a. Their feelings
 - b. Their thoughts on the activity monitor
5. My son/daughter is free to withdraw from the study at any time without giving reason.
6. The results will be stored on computer in coded form and individual results will be confidential.
7. The results of the study may be published but my son/daughter's anonymity will be preserved.

If you **DO** wish to take part in this study please sign below and return this form to the researcher.

Signed Parent/ Guardian / care giver.....

Date.....

Name of son/daughter (please print).....

Name parent/ guardian/ care giver

Contact telephone number(s).....

Contact email(s).....

This study has been reviewed and approved by the Sport and Health Sciences Ethics Committee.

Do we have permission to contact your son/daughter for further testing at a later date?

YES NO

Appendix 2.6. Child information and consent for chapter eight



Sport and Health Sciences
College of Life and Environmental Sciences
St Luke's Campus
Heavitree Road
EXETER, EX1 2LU
General Office +44 (0)1392 262896
Fax +44

November 2011
(0)1392 264706

Information Sheet

Dear Pupil,

Thank you for showing an interest in taking part in this study. This sheet will tell you a bit more about the study and what we would like you to do. If you decide not to take part it will not change your relationship with the research team or your school.

Project title:

The use of a wrist worn activity monitor to assess the relationship between Physical activity and psychological well being

What is the project about?

We are inviting the participation of 128 volunteers, both boys and girls, aged 8 - 12 years to take part in a Physical Activity and Health research study. Everyone is welcome to take part in the study whether you do lots of activities and exercise or not much at all.

What do we want to find out?

We have a new gadget (that looks a bit like a watch) that can be used to measure how much activity you do and we want your help to see how well they work and how easy they are to wear.

What do I have to do?

All you will be asked to do is to fill in some simple information about yourself (name, age, gender) and have your height and weight measured. Then you will be asked to wear the activity monitor on your wrist for seven days and fill out two questionnaires. The first questionnaire asks about your feelings and how you view yourself. This will be completed when you receive your activity monitor. The second questionnaire will ask you how you felt about wearing the monitor and will be completed after the seven day wear period, when we come to collect the monitors.

Can I change my mind?

Yes! You can stop the study at any time without having to give a reason.

What will we do with the information?

All the information collected will be stored on a computer and the results will be confidential to the University research team.

What if I have any questions?

If you have any questions then please feel free to ask any of the researchers listed below at any time.

What do I have to do next?

If you have read and understood everything that we want you to do and are happy to take part please sign the consent form, which is attached to this sheet. Your parent/guardian must also sign the form we have given to them.

The Sport and Health Sciences Ethics Committee has reviewed and approved this study.

Thank you,

Lisa Phillips

Sport and Health Sciences
College of Life and Environmental Sciences
University of Exeter
St. Luke's Campus
Heavitree Road
Exeter, EX1 2LU
Email: lrsp201@exeter.ac.uk
Tel: 07957356209

Additional contacts:

Dr. Melvyn Hillsdon
Tel: 01392 722868
Email: M.Hillsdon@exeter.ac.uk



Sport and Health sciences
 College of Life and Environmental Science
 St Luke's campus
 University of Exeter
 Exeter, Devon, UK
 EX1 2LU

Assent Form for child

Please complete and sign if you **DO** wish to take part in this study.

I have read the information sheet concerning this study and understand what it is about. All my questions have been answered. I understand that I will have my physical activity measured for seven days. I understand that I am free to ask for further information at any stage.

I know that:

1. My participation in the study is completely my choice, and I can stop taking part at any time.
2. I will be required to wear an activity monitor for 7 days
3. I will have my height and weight measured.
4. I will be asked to complete two questionnaires about :
 - a. My feelings
 - b. My thoughts on the activity monitor
5. I am free to withdraw from the study at any time without giving reason.
6. The results will be stored on computer in coded form and my name will not be included.
7. The results of the study may be published but my personal identity will not be known.

If you **DO** wish to take part in this study please sign below and return this form to the researcher.

Signed

Date.....

Name (please print).....

Contact telephone number(s).....

This study has been reviewed and approved by the Sport and Health Sciences Ethics Committee.

Do we have your permission to contact you in the future for follow-up research?

YES No

Appendix 3.1. Child Depression Inventory (CDI)

<p><u>Item 1</u></p> <p><input type="checkbox"/> I am sad once in a while</p> <p><input type="checkbox"/> I am sad many times</p> <p><input type="checkbox"/> I am sad all the time</p> <p><u>Item 2</u></p> <p><input type="checkbox"/> Nothing will ever work out for me</p> <p><input type="checkbox"/> I am not sure if things will work out for me</p> <p><input type="checkbox"/> Things will work out for me</p> <p><u>Item 3</u></p> <p><input type="checkbox"/> I do most things ok</p> <p><input type="checkbox"/> I do many things wrong</p> <p><input type="checkbox"/> I do everything wrong</p> <p><u>Item 4</u></p> <p><input type="checkbox"/> I have fun in many things</p> <p><input type="checkbox"/> I have fun in some things</p> <p><input type="checkbox"/> Nothing is fun at all</p> <p><u>Item 5</u></p> <p><input type="checkbox"/> I am bad all the time</p> <p><input type="checkbox"/> I am bad many times</p> <p><input type="checkbox"/> I am bad once in a while</p> <p><u>Item 6</u></p> <p><input type="checkbox"/> I think about bad things happening to me once in a while</p> <p><input type="checkbox"/> I worry that bad things will happen to me</p> <p><input type="checkbox"/> I am sure that terrible things will happen to me</p> <p><u>Item 7</u></p> <p><input type="checkbox"/> I hate myself</p> <p><input type="checkbox"/> I do not like myself</p> <p><input type="checkbox"/> I like myself</p> <p><u>Item 8</u></p> <p><input type="checkbox"/> All bad things are my fault</p> <p><input type="checkbox"/> Many bad things are my fault</p> <p><input type="checkbox"/> Bad things are not usually my fault</p>	Please leave blank
---	--------------------------

Item 9

- I feel like crying every day
- I feel like crying many days
- I feel like crying once in a while

Item 10

- Things bother me all the time
- Things bother me many times
- Things bother me once in a while

Item 11

- I like being with people
- I do not like being with people many times
- I do not want to be with people at all

Item 12

- I cannot make up my mind about things
- It is hard to make up my mind about things
- I make up my mind about things easily

Item 13

- I look ok
- There are some bad things about my looks
- I look ugly

Item 14

- I have to push myself all the time to do school work
- I have to push myself many times to do school work
- Doing school work is not a big problem

Item 15

- I have trouble sleeping every night
- I have trouble sleeping many nights
- I sleep pretty well

Item 16

- I am tired once in a while
- I am tired many days
- I am tired all the time

Item 17

- Most days I do not feel like eating
- Many days I do not feel like eating
- I eat pretty well

Item 18

- I do not worry about aches and pains
- I worry about aches and pains many times
- I worry about aches and pains all the time

Item 19

- I do not feel alone
- I feel alone many times
- I feel alone all the time

Item 20

- I never have fun at school
- I have fun at school only once in a while
- I have fun at school all the time

Item 21

- I have plenty of friends
- I have some friends but I wish I had more
- I do not have any friends

Item 22

- My school work is alright
- My school work is not as good as before
- I do badly in subjects I used to be good in

Item 23

- I can never be as good as other kids
- I can be as good as other kids if I want to
- I am just as good as other kids

Item 24

- Nobody really loves me
- I am not sure if anybody loves me
- I am sure that somebody loves me

Item 25

- I usually do what I am told
- I do not do what I am told most times
- I never do what I am told

Item 26

- I get along with people
- I get into fights many times
- I get into fights all the time

Appendix 3.2. State – trait anxiety inventory for children (STAIC)

1. I worry about making mistakes	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often
2. I feel like crying	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often
3. I feel unhappy	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often
4. I have trouble making up my mind	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often
5. It is difficult for me to face my problems	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often
6. I worry too much	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often
7. I get upset at home	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often
8. I am shy	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often
10. Unimportant thoughts run through my mind and bother me	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often
11. I worry about school	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often
12. I have trouble deciding what to do	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often
13. I notice my heart beats fast	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often
14. I am secretly afraid	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often
15. I worry about my parents	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often
16. My hands get sweaty	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often
17. I worry about things that may happen	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often
18. It is hard for me to fall asleep at night	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often
19. I get funny feelings in my stomach	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often
20. I worry about what others think of me	<input type="checkbox"/> hardly ever	<input type="checkbox"/> sometimes	<input type="checkbox"/> often

**Appendix 3.3. Self-perception profile for children (PSPP-CY),
including PSW and GSW subscales**

	Really true for me	Sort of true for me		BUT		Sort of true for me	Really true for me
1.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids do very well at all kinds of sports	BUT	Other kids don't feel that they are very good when it comes to sport	<input type="checkbox"/>	<input type="checkbox"/>
2.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids are often unhappy with themselves	BUT	Other kids are pretty pleased with themselves	<input type="checkbox"/>	<input type="checkbox"/>
3.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids wish they could be a lot better at sports	BUT	Other kids feel that they are good enough at sports	<input type="checkbox"/>	<input type="checkbox"/>
4.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids don't like the way they are leading their life	BUT	Other kids do like the way they are leading their life	<input type="checkbox"/>	<input type="checkbox"/>
5.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids think they could do well at just about any new sports activity they haven't tried before	BUT	Other kids are afraid they might not do well at sports they haven't ever tried	<input type="checkbox"/>	<input type="checkbox"/>
6.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids are happy with themselves as a person	BUT	Other kids are often not happy with themselves	<input type="checkbox"/>	<input type="checkbox"/>
7.	<input type="checkbox"/>	<input type="checkbox"/>	In games and sports some kids usually watch instead of play	BUT	Other kids usually play rather than watch.	<input type="checkbox"/>	<input type="checkbox"/>
8.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids like the kind of person they are	BUT	Other kids often wish they were someone else	<input type="checkbox"/>	<input type="checkbox"/>
9.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids feel that they are better than others their age at sports	BUT	Other kids don't feel they can play as well	<input type="checkbox"/>	<input type="checkbox"/>
10.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids are very happy being the way they are	BUT	Other kids wish they were different	<input type="checkbox"/>	<input type="checkbox"/>
11.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids don't do well at new outdoor games	BUT	Other kids are good at new games right away	<input type="checkbox"/>	<input type="checkbox"/>
12.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids are not very	BUT	Other kids think the way	<input type="checkbox"/>	<input type="checkbox"/>

		happy with the way they do a lot of things		they do things is fine			
13.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids feel uneasy when it comes to doing vigorous physical exercise	BUT	Other kids feel confident when it comes to doing vigorous physical exercise	<input type="checkbox"/>	<input type="checkbox"/>
14.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids are confident about how their bodies look physically	BUT	Other kids feel uneasy about how their bodies look physically	<input type="checkbox"/>	<input type="checkbox"/>
15.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids feel that they lack strength compared to other kids their age	BUT	Other kids feel that they are stronger than other kids their age	<input type="checkbox"/>	<input type="checkbox"/>
16.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids are proud of themselves physically	BUT	Other kids don't have much to be proud of physically	<input type="checkbox"/>	<input type="checkbox"/>
17.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids have a lot of stamina for vigorous physical exercise	BUT	Other kids soon get out of breath and have to slow down or quit	<input type="checkbox"/>	<input type="checkbox"/>
18.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids feel that they have a good looking (fit - looking) body compared to other kids	BUT	Other kids feel that compared to most, their body doesn't look so good	<input type="checkbox"/>	<input type="checkbox"/>
19.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids think that they have stronger muscles than kids their age	BUT	Other kids feel that they have weaker muscles than other kids their age	<input type="checkbox"/>	<input type="checkbox"/>
20.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids are happy with how they are and what they can do physically	BUT	Other kids are unhappy with how they are and what they can do physically	<input type="checkbox"/>	<input type="checkbox"/>
21.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids don't have much stamina and fitness	BUT	Other kids have lots of stamina and fitness	<input type="checkbox"/>	<input type="checkbox"/>
22.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids find it difficult to keep their bodies looking good physically	BUT	Other kids find it easy to keep their bodies looking good physically	<input type="checkbox"/>	<input type="checkbox"/>
23.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids lack confidence when it comes to strength activities	BUT	Other kids are very confident when it comes to strength activities	<input type="checkbox"/>	<input type="checkbox"/>
24.	<input type="checkbox"/>	<input type="checkbox"/>	Some kids don't feel very confident about themselves physically	BUT	Other Kids feel really good about themselves physically	<input type="checkbox"/>	<input type="checkbox"/>

- | | | | | | | | |
|-----|--------------------------|--------------------------|--|-----|---|--------------------------|--------------------------|
| 25. | <input type="checkbox"/> | <input type="checkbox"/> | Some kids try to take part in energetic physical exercise whenever they can | BUT | Other kids try to avoid to energetic exercise if they can | <input type="checkbox"/> | <input type="checkbox"/> |
| 26. | <input type="checkbox"/> | <input type="checkbox"/> | Some kids are pleased with the appearance of their body | BUT | Other kids wish that their bodies looked in better shape physically | <input type="checkbox"/> | <input type="checkbox"/> |
| 27. | <input type="checkbox"/> | <input type="checkbox"/> | When strong muscles are needed, some kids are the first to step forwards | BUT | Other kids are the last to step forward when strong muscles are needed | <input type="checkbox"/> | <input type="checkbox"/> |
| 28. | <input type="checkbox"/> | <input type="checkbox"/> | Some kids have a positive feeling about themselves physically | BUT | Other kids feel somewhat negative about themselves | <input type="checkbox"/> | <input type="checkbox"/> |
| 29. | <input type="checkbox"/> | <input type="checkbox"/> | Some kids soon have to quit running and exercising because they get tired | BUT | Other kids can run and do exercises for a long time without getting tired | <input type="checkbox"/> | <input type="checkbox"/> |
| 30. | <input type="checkbox"/> | <input type="checkbox"/> | Some kids feel that they are often admired for their good looking bodies | BUT | Other kids feel that they are seldom admired for the way their bodies look | <input type="checkbox"/> | <input type="checkbox"/> |
| 31. | <input type="checkbox"/> | <input type="checkbox"/> | Some kids feel that they are not as good as others when physical strength is needed | BUT | Other kids feel that they are among the best when it is physical strength that is needed | <input type="checkbox"/> | <input type="checkbox"/> |
| 32. | <input type="checkbox"/> | <input type="checkbox"/> | Some kids wish that they could feel better about themselves physically | BUT | Other kids always seem to feel good about themselves physically | <input type="checkbox"/> | <input type="checkbox"/> |
| 33. | <input type="checkbox"/> | <input type="checkbox"/> | When it comes to activities like running, some kids are able to keep going | BUT | Other kids soon have to quit to take a rest | <input type="checkbox"/> | <input type="checkbox"/> |
| 34. | <input type="checkbox"/> | <input type="checkbox"/> | Some kids don't like how their bodies look physically | BUT | Other kids are pleased with how their bodies look physically | <input type="checkbox"/> | <input type="checkbox"/> |
| 35. | <input type="checkbox"/> | <input type="checkbox"/> | Some kids think that they are strong, and have good muscles compared to other kids their age | BUT | Other kids think that they are weaker, and don't have such good muscles as kids their age | <input type="checkbox"/> | <input type="checkbox"/> |

36. Some kids are very satisfied with themselves physically **BUT** Other kids are often dissatisfied with themselves physically

Appendix 3.4. Strengths and difficulties questionnaire (SDQ)

	Not True	Somewhat True	Certainly True
I try to be nice to other people. I care about their feelings	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am restless, I cannot stay still for long	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I get a lot of headaches, stomach-aches or sickness	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I usually share with others (food, games, pens etc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I get very angry and often lose my temper	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am usually on my own. I generally play alone or keep to myself	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I usually do as I am told	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I worry a lot	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am helpful if someone is hurt, upset or feeling ill	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am constantly fidgeting or squirming	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I have one good friend or more	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I fight a lot. I can make other people do what I want	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am often unhappy, down-hearted or tearful	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other people my age generally like me	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am easily distracted, I find it difficult to concentrate	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am nervous in new situations. I easily lose confidence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am kind to younger children	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am often accused of lying or cheating	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other children or young people pick on me or bully me	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I often volunteer to help others (parents, teachers, children)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I think before I do things	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I take things that are not mine from home, school or elsewhere	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I get on better with adults than with people my own age	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I have many fears, I am easily scared	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I finish the work I'm doing. My attention is good	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Appendix 4.1. Worked example for correlated correlation coefficients
(chapter four)**

$$Z = (Zr1 - Zr2) \left\{ \frac{N-3}{2(1-rx)h} \right\}^{0.5}$$

Where:

N = number of subjects

Zr1 = Fisher's Zr for 1st predictor variable

Zr2 = Fisher's Zr for 2nd predictor variable

rx = Correlation between two predictor variables

h = $(1 - f\bar{r}^2) / (1 - \bar{r}^2)$

Where:

f = $(1 - rx) / [2(1 - \bar{r}^2)]$

\bar{r}^2 = $(r_1^2 + r_2^2) / 2$

Example demonstrated for the correlations present between light activity intensity and perceptions of sports competence when using Stone et al. (2009) cut-points and Evenson et al. (2008) cut-points

N = 49

Zr1 (Stone) = -.293

Zr2 (Evenson) = -.310

rx = .983

r_1^2 = 0.08

r_2^2 = 0.09

\bar{r}^2 = $(0.08 + 0.09) / 2$
= 0.086

f = $(1 - .983) / [2(1 - 0.086)]$
= $0.017 / 1.828$
= 0.0093

h = $(1 - 0.0007998) / (1 - 0.086)$
= $0.999 / 0.914$
= 1.093

Z = $(-.293 + .310) \{(49 - 3) / [2(1 - .983)1.093]\}^{0.5}$
= $0.017 \{46 / 0.037\}^{0.5}$
= 0.017×35.26
= 0.599

p = 0.55

Appendix 4.2. Worked example for correlated correlation coefficients (chapter seven)

The example given below is for assessment of differences between criterion validity of the GENEAt the hip location compared to the GENEAt the left wrist for all participants.

$$N = 44$$

$$Zr1(\text{left wrist}) = 1.528$$

$$Zr2(\text{Waist}) = 2.014$$

$$rx = 0.880$$

$$r_1^2 = 0.828$$

$$r_2^2 = 0.931$$

$$\begin{aligned} \bar{r}^2 &= (r_1^2 + r_2^2) / 2 \\ &= (0.828 + 0.931) / 2 \\ &= 0.879 \end{aligned}$$

$$\begin{aligned} f &= (1 - rx) / [2(1 - \bar{r}^2)] \\ &= (1 - .880) / [2(1 - .879)] \\ &= 0.12 / 0.241 \\ &= 0.498 \end{aligned}$$

$$\begin{aligned} h &= (1 - f\bar{r}^2) / (1 - \bar{r}^2) \\ &= (1 - 0.438) / (1 - 0.879) \\ &= 0.562 / 0.129 \\ &= 4.357 \end{aligned}$$

$$\begin{aligned} Z &= (zr1 - zr2) \{ (N-3) / [2 (1-rx) h] \}^{0.5} \\ &= (1.528 - 2.014) \{ (44-3) / [2 (1 - .880) 4.257] \}^{0.5} \\ &= -0.486 \{ 41 / [2 \times 0.12 \times 4.257] \}^{0.5} \\ &= -0.486 \{ 41 / 1.022 \}^{0.5} \\ &= -0.486 \times 6.334 \\ &= -3.07 \end{aligned}$$

$$p = 0.0021$$

Appendix 4.3. Non-significant regression analysis from chapter Five

Below is the regression output for non-significant predictions between baseline Pa and follow up PWB in chapter five over different wear times. Baseline PWB was accounted for in all models prior to inputting baseline PA.

	$r^2\Delta$	b	t	p
Physical strength				
<i>10 hour criteria</i>				
Time in moderate	.033	.018	1.738	.092
Time in MVPA	.033	.008	1.730	.094
<i>8 hour criteria</i>				
Time in moderate	.012	.012	1.202	.236
Time in MVPA	.015	.007	1.329	.190
Physical conditioning				
<i>10 hour criteria</i>				
Time in moderate	.034	.019	1.240	.224
<i>8 hour criteria</i>				
Time in moderate	.037	.023	1.703	.095
Time in vigorous	.031	.015	1.546	.129
Time in MVPA	.047	.013	1.935	.059

Appendix 4.4 Regression tables from chapter seven

Hierarchical regression analyses presented below showed that no additional variance in METs was accounted for by height or age ($p > 0.05$) at any wear location over and above that accounted for by GENE output.

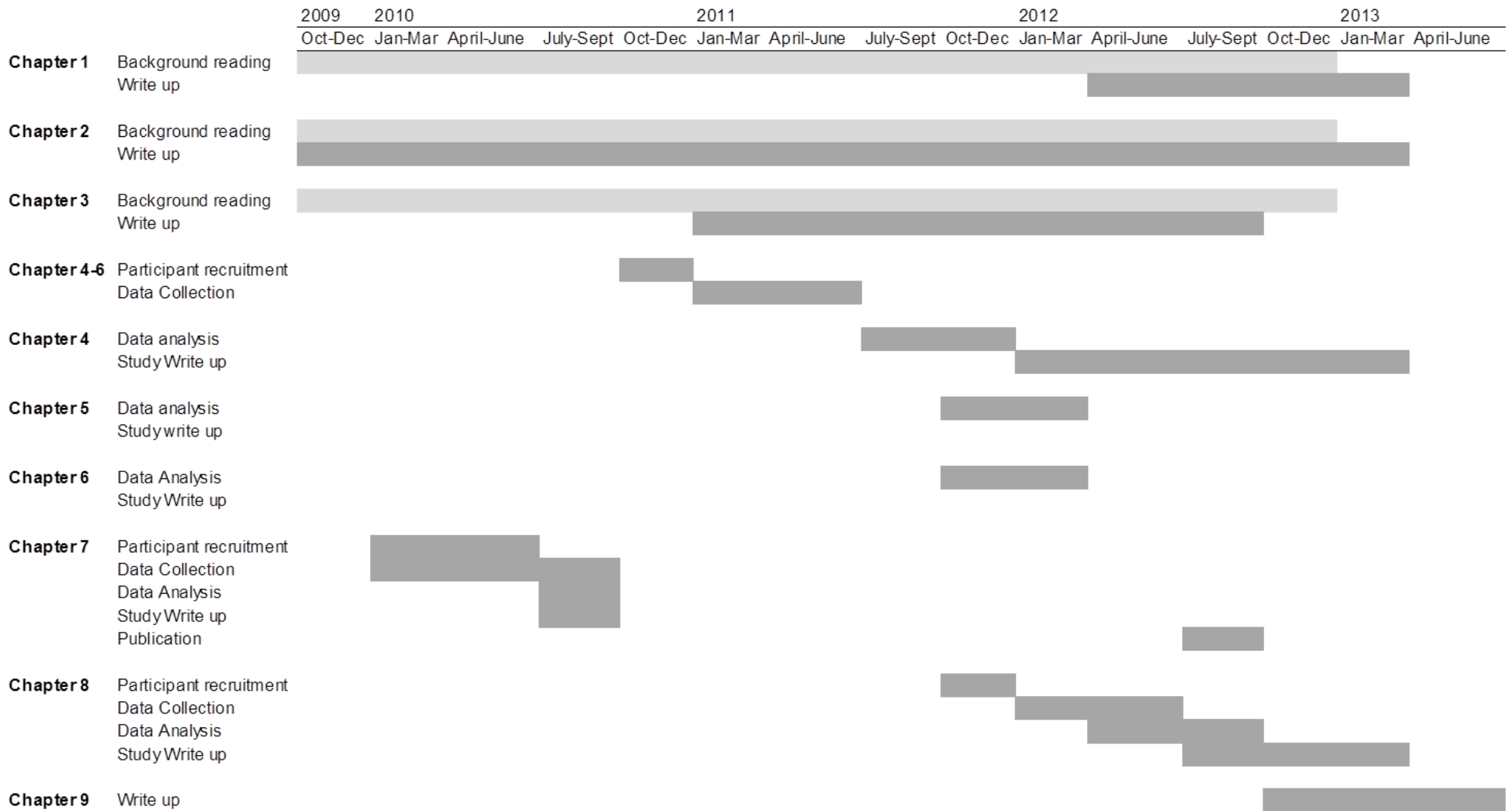
	<i>r</i>	<i>r</i> ²	<i>r</i> ² Δ	<i>b</i>	<i>t</i>	<i>p</i>
DV: METS						
<i>Model</i>						
GENEA right wrist	.418	.175	.175	.720	8.480	<.001
Height	.421	.177	.003	-.420	-1.047	.296
<i>Model</i>						
GENEA right wrist	.418	.175	.175	.727	8.561	<.001
Age	.418	.175	.000	-.510	-.181	.857
<i>Model</i>						
GENEA left wrist	.151	.023	.023	.093	2.538	.012
Height	.160	.026	.003	-.446	-.995	.320
<i>Model</i>						
GENEA left wrist	.151	.023	.023	.101	2.780	.006
Age	.151	.023	.000	-.571	-.184	.854
<i>Model</i>						
GENEA waist	.456	.208	.208	.818	9.428	<.001
Height	.457	.209	.001	-.285	-.722	.471
<i>Model</i>						
GENEA waist	.456	.208	.208	.826	9.542	<.001
Age	.456	.208	.000	.388	.140	.889

Appendix 4.5. Correlation tables for different capture periods with PWB for chapter eight

	ES	CP	HA	PP	PS	TD⁷
10 hour capture period (10am – 8pm)						
Time in Sedentary	-.086	-.068	.024	.028	-.098	-.038
Time in light	.173	.057	-.036	-.047	.095	.056
Time in moderate	.018	.029	-.056	-.017	.160	-.012
Time in vigorous	-.105	.134	.105	.032	-.121	.063
Time in MVPA	-.021	.065	-.008	-.002	.079	.012
10 hour capture period (12pm – 10pm)						
Time in Sedentary	-.085	-.085	.049	.042	-.124	-.029
Time in light	.155	.058	-.056	-.065	.103	.035
Time in moderate	.022	.042	-.078	-.019	.185	-.016
Time in vigorous	-.106	.170	.105	.034	-.110	.077
Time in MVPA	-.017	.086	.034	-.004	.105	.013

⁷ ES, emotional symptoms; CP, conduct problems; HA, hyperactivity scale; PP, peer problems; PS, pro social scale; TD, total difficulties

Appendix 5.1. Gantt chart



Appendix 6. Screen view of non-wear algorithm for GENEActiv

Microsoft Excel interface showing a spreadsheet with a formula bar and a data table. The formula bar contains: `=COUNTIF(OFFSET(N4,(A5-2)*1440+midnight_row-3,0,1440,1),AV$3)/60`

	Date	Time	x axis mean	y axis mean	z axis mean	Lux	Button	Temp	SVMgs	X axis stdev	Y axis stdev	Z axis stdev	Light	Activity Level	Current Movement	calc	calc	BED	Current Moveme nt * SVM	1hr Average	Min. of 1hr Average	NO WEAR			
1		9.00																							
2																									
3																									
5726	2012-03-2	08:46	-0.314	0.414	0.776	0	0	26.1	292	0	0	0	0	sedentary	0.88	1	301	1	FALSE	255.89	249.76	106	FALSE	0	0
5727	2012-03-2	08:47	-0.229	0.342	0.786	0	0	26.0	370	0	0	0	5	sedentary	0.15	1	66.5	1	FALSE	56.55	247.92	105	FALSE	0	0
5728	2012-03-2	08:48	-0.485	0.279	0.177	0	0	25.8	232	0	0	1	0	sedentary	0.96	1	263	1	FALSE	223.63	243.53	101	FALSE	0	0
5729	2012-03-2	08:49	-0.258	0.865	-0.366	0	0	25.7	82	0	0	0	0	sedentary	0.52	0	49.9	1	FALSE	42.40	238.66	101	FALSE	0	0
5730	2012-03-2	08:50	-0.120	0.794	-0.202	0	0	25.8	88	0	0	0	0	sedentary	0.48	1	49.8	1	FALSE	42.35	235.57	100	FALSE	0	0
5731	2012-03-2	08:51	-0.738	0.047	0.345	0	0	25.9	526	0	1	0	18	light	0.90	1	560	1	FALSE	475.94	228.21	92	FALSE	0	0
5732	2012-03-2	08:52	-0.650	0.639	0.222	0	0	25.8	152	0	0	0	0	sedentary	0.75	1	134	0	FALSE	113.54	223.80	90	FALSE	0	0
5733	2012-03-2	08:53	-0.791	0.491	-0.104	0	0	26.0	128	0	0	0	0	sedentary	0.78	1	118	0	FALSE	100.51	221.71	89	FALSE	0	0
5734	2012-03-2	08:54	-0.830	0.066	0.385	0	0	26.0	143	0	0	0	0	sedentary	0.16	1	26.5	0	FALSE	22.55	218.89	88	FALSE	0	0
5735	2012-03-2	08:55	-0.817	0.019	0.504	0	0	25.8	99	0	0	0	0	sedentary	0.29	0	33.9	0	FALSE	28.79	216.28	88	FALSE	0	0
5736	2012-03-2	08:56	-0.377	0.308	0.131	0	0	25.7	211	1	0	1	0	sedentary	0.60	2	148	0	FALSE	125.93	213.96	86	FALSE	0	0
5737	2012-03-2	08:57	0.122	0.763	-0.468	0	0	25.7	65	0	0	0	0	sedentary	0.60	1	45.8	0	FALSE	38.97	211.09	85	FALSE	0	0
5738	2012-03-2	08:58	0.490	0.698	-0.515	0	0	25.7	53	0	0	0	0	sedentary	0.51	0	31.7	0	FALSE	26.96	207.03	85	FALSE	0	0
5739	2012-03-2	08:59	-0.028	0.129	-0.177	0	0	25.9	2014	1	1	1	16	moderate	0.87	3	2050	0	FALSE	1742.74	202.99	55	FALSE	0	0
5740	2012-03-2	09:00	-0.465	-0.582	0.243	0	0	25.8	978	0	1	0	10	light	0.85	1	982	0	FALSE	834.69	201.32	42	FALSE	0	0
5741	2012-03-2	09:01	-0.302	-0.819	0.192	0	0	25.9	862	0	0	0	13	light	0.35	1	359	0	FALSE	304.98	192.61	36	FALSE	0	0
5742	2012-03-2	09:02	-0.413	-0.404	0.134	0	0	26.1	568	0	1	1	29	light	0.50	1	332	0	FALSE	282.05	189.03	32	FALSE	0	0
5743	2012-03-2	09:03	-0.548	-0.091	0.170	0	0	26.1	771	0	1	0	2	light	0.46	2	420	0	FALSE	356.98	183.02	26	FALSE	0	0
5744	2012-03-2	09:04	-0.556	-0.177	0.049	9	0	26.2	1698	1	1	1	61	bed	0.46	2	927	0	TRUE	788.36	172.85	13	FALSE	0	0
5745	2012-03-2	09:05	-0.645	-0.101	0.438	17	0	26.2	1073	0	0	1	120	bed	0.61	1	774	0	TRUE	657.72	163.28	2	FALSE	0	0
5746	2012-03-2	09:06	0.169	0.990	-0.049	104	0	26.0	38	0	0	0	106	bed	1.19	0	53.5	0	TRUE	45.48	159.00	1	FALSE	0	0
5747	2012-03-2	09:07	0.171	0.990	-0.047	104	0	25.6	38	0	0	0	106	no wear	0.07	0	2.95	0	TRUE	2.51	152.58	1	TRUE	1	1
5748	2012-03-2	09:08	0.171	0.990	-0.044	104	0	25.2	37	0	0	0	106	no wear	0.06	0	2.48	0	TRUE	2.10	145.92	1	TRUE	1	2
5749	2012-03-2	09:09	0.172	0.991	-0.042	105	0	24.8	38	0	0	0	106	no wear	0.05	0	2.35	0	TRUE	2.00	138.12	1	TRUE	1	3
5750	2012-03-2	09:10	0.172	0.991	-0.040	105	0	24.5	39	0	0	0	106	no wear	0.05	0	2.18	0	TRUE	1.86	133.78	1	TRUE	1	4
5751	2012-03-2	09:11	0.172	0.991	-0.038	105	0	24.2	39	0	0	0	106	no wear	0.05	0	2.36	0	TRUE	2.01	126.69	1	TRUE	1	5
5752	2012-03-2	09:12	0.172	0.992	-0.036	104	0	23.8	41	0	0	0	106	no wear	0.05	0	2.29	0	TRUE	1.95	122.63	1	TRUE	1	6
5753	2012-03-2	09:13	0.172	0.992	-0.035	103	0	23.6	42	0	0	0	104	no wear	0.04	0	2.14	0	TRUE	1.82	115.29	1	TRUE	1	7
5754	2012-03-2	09:14	0.172	0.992	-0.033	101	0	23.3	42	0	0	0	104	no wear	0.04	0	1.98	0	TRUE	1.69	113.99	1	TRUE	1	8
5755	2012-03-2	09:15	0.172	0.993	-0.032	101	0	23.0	43	0	0	0	104	no wear	0.04	0	2.28	0	TRUE	1.94	110.24	1	TRUE	1	9
5756	2012-03-2	09:16	0.172	0.993	-0.031	101	0	22.9	44	0	0	0	104	no wear	0.03	0	1.7	0	TRUE	1.44	105.99	1	TRUE	1	10
5757	2012-03-2	09:17	0.172	0.993	-0.030	101	0	22.7	44	0	0	0	104	no wear	0.04	0	1.93	0	TRUE	1.64	105.05	1	TRUE	1	11
5758	2012-03-2	09:18	0.172	0.993	-0.029	101	0	22.5	44	0	0	0	104	no wear	0.03	0	1.8	0	TRUE	1.53	101.33	1	TRUE	1	12
5759	2012-03-2	09:19	0.172	0.993	-0.028	101	0	22.2	45	0	0	0	104	no wear	0.03	0	1.74	0	TRUE	1.48	100.64	1	TRUE	1	12
5760	2012-03-2	09:20	0.173	0.993	-0.027	101	0	22.2	45	0	0	0	104	no wear	0.03	0	1.74	0	TRUE	1.48	99.95	1	TRUE	1	12
5761	2012-03-2	09:21	0.173	0.993	-0.026	100	0	22.1	46	0	0	0	101	no wear	0.03	0	1.52	0	TRUE	1.29	92.03	1	TRUE	1	12
5762	2012-03-2	09:22	0.173	0.993	-0.026	98	0	22.0	46	0	0	0	101	no wear	0.03	0	1.51	0	TRUE	1.29	90.16	1	TRUE	1	12
5763	2012-03-2	09:23	0.173	0.993	-0.025	98	0	21.7	46	0	0	0	101	no wear	0.03	0	1.72	0	TRUE	1.46	88.51	1	TRUE	1	12
5764	2012-03-2	09:24	0.173	0.993	-0.024	98	0	21.7	47	0	0	0	101	no wear	0.03	0	1.56	0	TRUE	1.33	88.14	1	TRUE	1	12

Excel status bar: Average: 3.6 hrs Count: 9 Sum: 21.4 hrs 100%