

Statistical methods for analyzing physical activity data

by

Nicholas Koenen Beyler

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
Sarah M. Nusser, Co-major Professor
Wayne A. Fuller, Co-major Professor
Daniel J. Nordman
Gregory J. Welk
Cindy L. Yu

Iowa State University

Ames, Iowa

2010

Copyright © Nicholas Koenen Beyler, 2010. All rights reserved.

DEDICATION

To my family and friends, especially my parents, for their constant moral and financial support of my education.

ACKNOWLEDGEMENTS

This research was supported by Grant Number HL091024 from the National Institutes of Health. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

I am grateful to my major professors, Sarah Nusser and Wayne Fuller, for taking the time to guide me through all the challenges and obstacles that have presented themselves during my tenure at Iowa State. I would like to thank Alicia Carriquiry, Dan Nordman, Greg Welk, and Cindy Yu for serving on my committee and providing me with useful feedback on my research. I owe a special thanks to my fellow graduate students and the Survey Research Services staff at CSSM for their moral support.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vii
LIST OF FIGURES	x
ABSTRACT	xi
CHAPTER 1 A REVIEW OF PHYSICAL ACTIVITY MEASUREMENT	
METHODS AND MODELS	1
1.1 Introduction	1
1.2 Physical Activity Metrics	2
1.2.1 Metabolic Equivalents	2
1.2.2 Energy Expenditure	3
1.2.3 Physical Activity Groups	4
1.2.4 Other Metrics	5
1.3 Physical Activity Instruments	5
1.3.1 Laboratory Instruments	5
1.3.2 Monitor-based Instruments	6
1.3.3 Self-report Instruments	9
1.4 Modeling Physical Activity	11
1.4.1 Variation and Bias in Physical Activity Measurement	11
1.4.2 A Measurement Error Model	14
1.4.3 Physical Activity Models	16
1.4.4 Dietary Intake Models	19

1.5	Summary	28
CHAPTER 2 A REGRESSION MODEL FOR MODERATE TO VIGOR-		
	OUS PHYSICAL ACTIVITY (MVPA)	31
2.1	Introduction	31
2.2	NHANES Physical Activity Data	32
2.3	Regression Model for Female Physical Activity	36
2.3.1	Model Development	37
2.3.2	EGLS Estimator	40
2.3.3	Test for Full vs. Reduced Models	42
2.3.4	Test for Survey Weights	43
2.3.5	Test for EGLS Variances vs. Stratified Cluster Variances	44
2.3.6	Model Comparisons Across Samples	44
2.3.7	Model for Full Sample	47
2.4	Estimation and Prediction of Daily MVPA	50
2.5	Discussion	53
CHAPTER 3 A METHOD FOR ESTIMATING USUAL DAILY ENERGY		
	EXPENDITURE PARAMETERS	55
3.1	Introduction	55
3.2	Methodology	60
3.2.1	Transformation to Normality and Test for Nuisance Effects	61
3.2.2	Group-Level Measurement Error Model	65
3.2.3	Population-Level Model	71
3.2.4	Estimating Parameters of Usual Daily EE in the Original Scale	74
3.3	Application to PAMS Data	77
3.3.1	Survey Design	77
3.3.2	Female Daily EE	79
3.3.3	Methodology for the PAMS Sample	80
3.4	Discussion	91

APPENDIX A	CONSISTENCY OF AN EGLS REGRESSION ESTIMA-	
	TOR FOR A STRATIFIED CLUSTER DESIGN	94
APPENDIX B	REGRESSION ESTIMATION FOR STRATIFIED CLUS-	
	TER DESIGN	102
APPENDIX C	TEST FOR WEIGHTED AND UNWEIGHTED ESTIMA-	
	TORS FOR THE POPULATION-LEVEL MODEL	105
BIBLIOGRAPHY	109

LIST OF TABLES

Table 2.1	Demographic decomposition of the NHANES samples	34
Table 2.2	Percentiles for the distribution of final survey weights for the NHANES samples	34
Table 2.3	Weighted means (standard errors) for average daily MVPA (min/day) measured by accelerometer for the NHANES samples	35
Table 2.4	Count (percent) of individuals with zero (0) and positive (> 0) esti- mated average daily MVPA for the NHANES samples	37
Table 2.5	Variables for model	38
Table 2.6	Estimated regression coefficients for model (2.4) fit to three female groups in the 2003-2004 NHANES sample using OLS	40
Table 2.7	Estimated regression coefficients for model (2.6) fit to the 2003-2004 NHANES sample using OLS	42
Table 2.8	Percentiles for the distribution of estimated error variances for the 2003- 2004 NHANES sample	42
Table 2.9	EGLS and stratified cluster (Str. Clus.) standard errors (SE) for model (2.6) fit to the 2003-2004 NHANES sample using EGLS	45
Table 2.10	ANOVA for the EGLS standardized residuals	45
Table 2.11	Estimated regression coefficients for model (2.6) fit to the 2003-2004 and 2005-2006 NHANES samples using EGLS	46
Table 2.12	Estimated regression coefficients for models (2.6) and (2.12) fit to the full NHANES sample using EGLS	48

Table 2.13	Estimates and standard errors for models (2.17) and (2.18) fit using EGLS	50
Table 2.14	Example estimated average daily accelerometer MVPA in groups of the female population for 33 minutes of reported MVPA	52
Table 2.15	Example predicted average daily accelerometer MVPA for an individual in the female population for 33 minutes of reported MVPA (standard errors are in parentheses and 95% prediction intervals are in brackets)	53
Table 3.1	Method of moments estimators for group g	69
Table 3.2	Elements D_{glk} in the derivative matrix \hat{D}_g , where $f_1 = m_{12} - 0.25m_{34}$ and $f_2 = m_{11} - 0.25m_{33}$	70
Table 3.3	Percentiles for PAMS survey weights for the preliminary female sample	78
Table 3.4	Demographic characteristics of the female PAMS sample	80
Table 3.5	P-values for the linear regression models fit to the log-transformed data	82
Table 3.6	Age groups	83
Table 3.7	Estimated measurement error model parameters (and standard errors) for the mean of daily EE (μ_g), the mean of reported daily EE (μ_{yg}), and the slope for population-level reporting bias (β_{1g})	84
Table 3.8	Estimated variance components (and standard errors) from the measurement error model for mean daily EE (σ_{tg}^2), within-individual EE variation in daily EE (σ_{dg}^2), measurement error variation from the monitor (σ_{ug}^2) and the 24PAR (σ_{eg}^2), and reporting-bias variation from the recall (σ_{rg}^2)	85
Table 3.9	Parameter estimates (standard errors) for the population-level model .	87
Table 3.10	Parameter estimates (standard errors) for the daily EE group means (μ_g) and slope parameters (β_{1g}) based on the fitted population-level model	88

Table 3.11	Parameter estimates (standard errors) for the group mean of usual daily EE in the original scale (\bar{T}_g), the group standard deviation of usual daily EE in the original scale (S_{Tg}), the group proportion of individuals with less than 1750 kcal/d of usual daily EE ($p_{low,g}$), and the group proportion of individuals with more than 3250 kcal/d of usual daily EE ($p_{high,g}$)	92
Table C.1	Parameter estimates for $\hat{\lambda}_1$ and $\hat{\lambda}_2$	107
Table C.2	Sample sizes and population control totals for females in the 8 PAMS strata (control totals are from the 2000 U.S. Census)	108

LIST OF FIGURES

Figure 2.1	Plot of average daily MVPA for the 2003-2004 NHANES sample (dashed line is the identity line)	36
Figure 2.2	Plot of standardized group means and standard deviations	48
Figure 2.3	Plot of standardized group means and standard deviations for the restricted model	51
Figure 3.1	Boxplots of daily EE from the monitors and 24PARs and scatter plot of individual means of daily EE from the monitors and 24PARs (dashed line is the identity line)	80
Figure 3.2	Estimated lines relating mean daily EE and reported daily EE for age groups 1 and 4 (points are the individual means of measured EE in the log scale; dashed lines are the estimated lines and dotted lines are the identity lines)	88
Figure 3.3	Estimated densities of usual daily EE for age groups 1 - 4	90
Figure 3.4	Estimated densities of usual daily EE, individual means of daily monitor EE, and individual means of daily 24PAR EE for age group 1 (age < 43)	91

ABSTRACT

Physical activity is any bodily movement that results in caloric expenditure. One important aspect of physical activity research is the assessment of usual (i.e., long-term average) physical activity in the population, in order to better understand the links between physical activity and health outcomes. Daily or weekly measurements of physical activity taken from a sample of individuals are prone to measurement errors and nuisance effects, which can lead to biased estimates of usual physical activity parameters. Fortunately, statistical models can be used to account and adjust for these errors in order to give more accurate estimates of usual physical activity parameters.

In this dissertation we develop statistical methods for estimating parameters of usual physical activity. In Chapter 1 we outline metrics and instruments used for physical activity assessment, and review current approaches for modeling usual physical activity and usual dietary intake for regularly consumed food components. In Chapter 2 we develop a model for physical activity data from the National Health and Nutrition Examination Survey (NHANES). A linear regression is defined to model objective monitor-based physical activity as a function of self-reported physical activity variables and demographic variables. The fitted model is used to estimate mean daily physical activity levels for demographic groups in the population. In Chapter 3 we develop a method for estimating usual daily energy expenditure parameters from data collected using a self-report instrument and an objective monitoring device. Our method is an extension of existing methods that utilize measurement error models. We illustrate our method with preliminary data from the Physical Activity Measurement Survey (PAMS) collected using a SenseWear Pro armband monitor and a 24-hour physical activity recall.

CHAPTER 1 A REVIEW OF PHYSICAL ACTIVITY MEASUREMENT METHODS AND MODELS

1.1 Introduction

Physical activity is any bodily movement produced by skeletal muscles that results in caloric expenditure (Caspersen et al. 1985). Participation in regular physical activity has been linked to health benefits such as reduced risk of obesity (Grundy et al. 1999), improved mental health (Kritz-Silverstein et al. 2001), and improved cognitive function (Yaffe et al. 2001). Moreover, lack of participation in regular physical activity has been linked to a number of health concerns such as heart disease (Berlin and Colditz 1990), diabetes (Manson et al. 1991), osteoporotic fractures (Kannus 1999), and cancers (Lee 2003; Thune et al. 1997; Friedenreich and Orenstein 2002). Due to the recognized relationships between physical activity and various health outcomes, researchers have established recommendations for physical activity engagement. A recent report from the Surgeon General recommends that people engage in at least 150 minutes of moderate-intensity physical activity per week, or alternatively, at least 75 minutes of vigorous-intensity physical activity per week (U.S. Department of Health and Human Services 2008).

With these recommendations in place, researchers must assess whether or not individuals in the population adhere to the recommendations and whether or not the recommendations are reasonable in terms of improving health. Measuring and analyzing physical activity data from individuals in the population can be challenging. First, researchers must decide how to measure physical activity from individuals in the population. A variety of physical activity metrics are used to measure and assess physical activity. Second, researchers must decide how to analyze physical activity data collected from a sample of individuals. One important

goal in physical activity research is the assessment of usual or habitual physical activity in the population (Shephard 2003). Usual physical activity can be thought of as an individual's long-term average physical activity, such as his or her average daily physical activity level over the course of a year, and cannot be measured directly because of day-to-day variation in physical activity and errors in physical activity measurements. Statistical models can be used for estimating usual physical activity parameters in a population by adjusting for measurement errors and other forms of variation that exist in physical activity measurements. The estimated parameters can then be used to evaluate adherence to physical activity guidelines and to study the relationships between usual physical activity and health outcomes.

In this chapter, we review current approaches for measuring and modeling physical activity data. In Section 1.2 we define common metrics used for physical activity assessment. In Section 1.3 we present some of the instruments that are used to measure physical activity. In Section 1.4 we examine statistical methods used for modeling physical activity and related methods used for modeling dietary intake. In Section 1.5 we give a summary of the remaining chapters.

1.2 Physical Activity Metrics

Physical activity can be measured as the energy cost required to engage in the activity, as the time spent in a specific activity or behavior, or via indirect outcomes that are correlated with physical activity metrics. In this section, we review the methods used for measuring physical activity.

1.2.1 Metabolic Equivalents

An important concept in physical activity measurement is the metabolic equivalent (MET). One MET is the energy expenditure at an individual's resting state, which is defined to be approximately 3.5 ml/kg/min of oxygen consumption (Welk 2002a). Intensities of activities are measured in MET units, which are defined relative to the baseline level of 1 MET. The Compendium of Physical Activities (Ainsworth et al. 2000; Ainsworth et al. 1993) lists MET values associated with various activities and can be used to quantify the intensities of activities

reported by individuals in free-living situations. A limitation of using MET values from the compendium to standardize physical activity levels across individuals is that the process fails to account for adaptability of the body to physical activity (Welk 2002a). That is, the same level of activity can be perceived differently for a physically fit person and an unfit person. Running at 5 mph may barely quantify as “physical activity” for a physically fit person, but may be considered very challenging for an unfit person.

1.2.2 Energy Expenditure

One of the most common physical activity metrics is energy expenditure. Energy expenditure is a measure of the energy cost of physical activity (Schutz et al. 2001) and can be expressed in kilocalories (kcal) per unit of time (e.g., kcal/d or kcal/wk). One MET is equivalent to approximately 1 kcal/kg/hr, so that an individual with body weight W (in kilograms) who engages in T hours of an activity with MET value V expends $K = VTW$ kcals of energy for the activity during those T hours (Ainsworth 2009). The accumulation of kcals for all non-resting activities (i.e., activities with associated MET values greater than one) during the course of a day represents physical activity energy expenditure (PAEE) and is often measured in kcal/d (Schutz et al. 2001). The sum of PAEE and daily resting energy expenditure (REE) (kcals expended during rest) is known as total energy expenditure (TEE), which is also measured in kcal/d (Schutz et al. 2001). Physical activity level (PAL) is the ratio of TEE to REE and is a useful alternative to PAEE as an index of energy expenditure related to physical activity over a 24-hour period (Schutz et al. 2001).

Energy expenditure can also be measured in MET-hours (or MET-minutes) per unit of time. If an individual engages in an activity with MET value V for T hours, he or she engages in VT MET-hours of activity. MET-hours can be accumulated for all activities (resting and non-resting activities) during the course of a day or can just be accumulated for non-resting activities. If both resting and non-resting activities are considered, a measure of energy expenditure in MET-hours is similar to TEE, and an individual who is at a MET level of 1 for the entire day engages in 24 MET-hours of activity. If only non-resting activity is considered,

a measure of energy expenditure in MET-hours is similar to PAEE, and the same individual at a MET level of 1 for the entire day engages in 0 MET-hours of activity. Whether to express energy expenditure in kilocalories or in MET-hours is a decision left to the researcher. If there is interest in comparing energy intake to energy expenditure, kilocalories is the preferred unit of measure. If there is interest in comparing the intensity levels for various types of activity, or in comparing physical activity levels across individuals with varying weights, MET-hours may be the preferred unit of measure.

1.2.3 Physical Activity Groups

Oftentimes researchers are interested in the amount of time people spend engaging in activity that is classified into activity groups or behaviors, where the activity groups are usually defined by intensity level. Most researchers classify an activity as light intensity if it has a MET value in the 1-3 range, as moderate intensity if it has a MET value in the 3-6 range, and as vigorous intensity if it has a MET value greater than 6 (Troiano et al. 2008; Crouter et al. 2006; Ainsworth et al. 2000). A common metric used to classify activity by intensity is time spent in moderate to vigorous physical activity (MVPA), which measures the amount of time individuals engage in activity at or above 3 METs in a day or week or month (Troiano et al. 2008). This metric is important for evaluating the adherence of physical activity guidelines, which are defined by intensity level of activity.

Physical activity groups may also be defined by factors other than intensity of activity. For example, researchers may be interested in measuring the amount of time people spend in activity in specific contexts, such as for an individual's occupation, transportation, leisure, household chores, and exercise. This type of physical activity assessment is gaining in popularity as interest in studying sedentary behaviors from occupational and household settings grows (Ainsworth 2009).

1.2.4 Other Metrics

When physical activity metrics are unavailable, other metrics related to physical activity can be used to indirectly assess physical activity. Heart rate (HR) measured in average beats per minute is one metric that is related to physical activity (Schutz et al. 2001) and is usually measured with heart rate monitors (Janz 2006). Accelerometers measure activity intensity via the average number of counts per minute (Troiano et al. 2008; Welk 2002b), and pedometers measure the number of steps taken in a day (Ainsworth 2009; Janz 2006). These types of metrics can be analyzed directly or can be converted into estimates of physical activity levels using calibration functions, which are discussed in Section 3.2 (Schutz et al. 2001; Welk 2005; Crouter et al. 2006; Moy et al. Submitted).

1.3 Physical Activity Instruments

In this section, we describe three types of instruments used for measuring physical activity: laboratory instruments, monitor-based instruments, and self-report instruments.

1.3.1 Laboratory Instruments

Three common laboratory-based methods used for measuring physical activity are doubly labeled water (DLW), direct calorimetry, and indirect calorimetry. For the DLW method, individuals drink water containing isotopically labeled hydrogen and oxygen atoms on multiple occasions and provide urine samples before and after drinking the water (Starling et al. 1999). Usually the final urine sample is collected 14 days after the water is first administered (Bratteby et al. 1998). An estimate of TEE is obtained by comparing carbon dioxide production in pre-dose and post-dose urine samples (Bratteby et al. 1998). Estimates of PAEE and PAL can be indirectly obtained by using an estimate of REE from some external source such as indirect calorimetry (Bratteby et al. 1998; Starling et al. 1999; Bouten et al. 1996). The DLW method is often referred to as the “gold standard” for measuring energy expenditure (Lagerros and Lagiou 2007; Bouten et al. 1996; Bratteby et al. 1998; Starling et al. 1999; Moy et al. Submitted) and can be used to measure energy expenditure in free-living subjects without

influencing daily routines (Bratteby et al. 1998). But the DLW method is also very costly to implement (Johnson et al. 1998; Starling et al. 1999) and only provides estimates of TEE for a one or two week period (Bouten et al. 1996). Thus, the DLW method cannot be used for measuring MVPA or other physical activity variables related to the behavior or context of physical activity.

With direct calorimetry, energy expenditure is measured through production of heat from individuals who are contained in special chambers (LaPorte et al. 1985). Direct calorimetry is accurate for measuring energy expenditure, but is also expensive, and limits measurement to the laboratory environment (LaPorte et al. 1985; Lagerros and Lagiou 2007). With indirect calorimetry, energy expenditure is measured by the consumption of oxygen (LaPorte et al. 1985), where individuals are required to wear a face mask or a mouthpiece with a nose clip and a container that collects expired air (LaPorte et al. 1985). Like direct calorimetry, indirect calorimetry is accurate for measuring energy expenditure, but is also expensive and unrealistic for measurement under free-living conditions (Lagerros and Lagiou 2007). Because of the limitations associated with direct and indirect calorimetry, DLW remains the only stand alone “gold standard” for measuring energy expenditure in free-living subjects.

1.3.2 Monitor-based Instruments

Monitor-based instruments are instruments that individuals wear on their bodies as they go about their day. The monitors record information related to an individual’s activity by keeping track of bodily movements and other bodily functions, such as heart rate and body temperature. Accelerometers, pedometers, heart rate monitors, and multi-sensor devices are all instruments used to measure physical activity from individuals in free-living conditions. In this section, we review research on accelerometers and multi-sensor devices, which are used most often in contemporary physical activity studies (Welk 2002b; Moy et al. Submitted). See Schutz et al. (2001) for information on heart rate monitors and Ainsworth (2009) for information on pedometers.

To date, the most commonly used monitor-based instrument for measuring physical activity

is the accelerometer (Welk 2002b; Janz 2006; Ward et al. 2005; Welk 2005; Trost et al. 2005; Strath et al. 2005). Accelerometers are usually worn on the waist or hips (Welk 2002b; Ward et al. 2005; Ainsworth 2009), but can also be worn on the wrist or ankle (Ward et al. 2005). Accelerometers measure acceleration, which is the change in velocity over time (Welk 2002b). The data produced by the accelerometer are intensity counts, where an increasing number of counts reflects more intense activity (Ainsworth 2009). There are many commercially available accelerometers on the market (Ward et al. 2005; Welk 2005). Some monitors measure acceleration in only one direction, while other monitors measure acceleration in multiple dimensions (Welk 2002b; Welk et al. 2004). The most widely used accelerometer is the Actigraph, which is a one-dimensional accelerometer that measures vertical acceleration (Leenders et al. 2006; Welk 2002b; Troiano et al. 2008).

Accelerometer research is extensive and a number of studies have investigated the reliability and validity of various accelerometers used for field-based research (Welk 2002b; Trost et al. 2005; Ward et al. 2005; Welk et al. 2004). Some important points made in the literature are:

- no one accelerometer is vastly superior to another (Trost et al. 2005; Ward et al. 2005)
- selecting a type of accelerometer is primarily an issue of practicality (Trost et al. 2005)
- using multiple accelerometers on any one individual as opposed to a single accelerometer may be beneficial (Strath et al. 2005), but one monitor will suffice in most cases (Troiano 2005; Trost et al. 2005)
- the trunk (i.e., hip or lower back) is the best place to wear an accelerometer (Trost et al. 2005; Ward et al. 2005)
- 3 to 5 days of monitoring is required to reliably estimate usual or habitual activity in adults (Trost et al. 2005)
- 4 to 9 days of monitoring is required to reliably estimate usual or habitual activity in children and adolescents (Trost et al. 2005)

- accelerometers do a better job of measuring general locomotor tasks as opposed to upper-body movements (Welk 2002b)
- different monitors produce output that is measured in different units making it difficult to compare results across studies (Welk 2002b).

One area of accelerometer research that has gained popularity is calibration research (Welk 2005). Calibration, as defined by physical activity researchers, is the conversion of accelerometer intensity counts into useful physical activity metrics such as energy expenditure or MVPA (Welk 2002b; Welk 2005). Calibrating intensity counts into energy expenditure usually involves the development of a regression equation that defines a linear relationship between intensity counts and energy expenditure, where estimates of energy expenditure are obtained by plugging intensity counts into the fitted regression equation (Welk 2002b). Unfortunately, energy expenditure estimates obtained from fitted linear regression equations have been found to be fairly inaccurate when applied to individuals who wear accelerometers in free-living situations (Welk 2002b). As a consequence, Crouter et al. (2006) consider using two regression equations to estimate separately energy expenditure for walking and running activity and for leisure time activity. Calibrating intensity counts into MVPA involves determining intensity count cutpoints to represent moderate and vigorous intensity (Welk 2002b). For example, Freedson et al. (1998) consider counts per minute from the Actigraph of 1951 or lower, 1952-5724, and 5725 or higher to represent activity of light, moderate, and vigorous intensity, respectively. Other methods for converting intensity counts into time-based physical activity metrics have been proposed by Nichols et al. (1999) and Hendelman et al. (2000), among others. Unfortunately, having multiple methods for measuring MVPA from accelerometers makes it difficult to compare results across studies (Welk 2002b).

Advances in technology have led to the development of multi-channel or multi-sensor devices that utilize pattern recognition algorithms to estimate physical activity (Moy et al. Submitted). Three such devices are the Actiheart, the SenseWear Pro armband monitor, and the Intelligent Device for Energy Expenditure and Physical Activity (IDEEA) monitor (Moy et al. Submitted). The Actiheart uses integrated information on heart rate and acceleration to

estimate PAEE (Moy et al., Submitted). Studies have shown that integrating heart rate and motion sensor information improves accuracy of PAEE estimates (Strath et al. 2001), but this technique is still difficult to implement under free-living conditions (Moy et al. Submitted). The SenseWear Pro armband monitor is a wireless armband worn on the upper arm that integrates information from two accelerometers and a variety of heart and pulse sensors (Moy et al. Submitted). The SenseWear monitor is of minimal burden to researchers and survey participants, registers upper body movements typically missed by hip-worn accelerometers, and is highly accurate for estimating PAEE (Jakicic et al. 2004; Fruin and Rankin 2004). However, this monitor is also inadequate for detecting certain types of activities, such as bicycling, and must be taken off during showering and swimming (Moy et al. Submitted). The IDEEA monitor is composed of 5 integrated sensors connected by wires that are placed on different parts of the body. This device can measure physical activity fairly accurately (Zhang et al. 2003) and can store a large amount of data (Moy et al. Submitted), but is also fairly expensive (at least 3-5 times more expensive than the Actiheart and SenseWear monitor) and is a significant burden to survey participants since multiple sensors must be placed all over the body (Moy et al. Submitted).

1.3.3 Self-report Instruments

With self-report instruments, individuals are asked to report on their activities. Individuals may be asked to recall activities from a previous day, week, or month, or may be asked to keep a log or record of their activities as they go about their day. Information on activity type (e.g., aerobic, anaerobic, occupational, household), frequency (e.g., number of times per week), intensity (e.g., energy cost), and duration (e.g., how many minutes per occasion) can be gathered using self-report instruments (Matthews 2002). The four general classes of self-report instruments are records/logs, global self-reports, recall questionnaires, and quantitative history questionnaires (Matthews 2002; Ainsworth 2009).

With physical activity records or logs, individuals provide detailed information on physical activities as they occur during the day. Logs can provide fairly accurate information on physical

activity because the activities are reported on as they occur, which reduces the likelihood of misreporting on activity. However, activity logs can be a significant burden to individuals (Ainsworth 2009) and may influence individuals to engage in more intense activity than normal, which is a phenomenon known as reactivity (Matthews 2002).

With global self-reports, individuals are asked to provide a generic classification of their usual activity patterns over a long period of time period (e.g., a year) via a small number of questions (Matthews 2002). These types of instruments rely more on generic memories (i.e., recollections of general events or patterns of events from the past) instead of episodic memories (i.e., specific recollections of individual and innumerable autobiographical events), and are therefore reliant on individuals' abilities to accurately assess their own usual physical activity (Matthews 2002). Global self-reports are used primarily as screening tools in clinical settings and are not very useful for understanding type, frequency, intensity, and duration of activity (Matthews 2002).

Recall questionnaires ask individuals to recall their activity from the recent past (e.g., the previous day or week). These questionnaires are usually short (5 to 15 minutes) and are designed to classify individuals into broad physical activity categories (Matthews 2002; Ainsworth 2009). Recall questionnaires are useful for classifying activity into groups (e.g., exercise, leisure, occupation, transportation activity) and for assessing type, frequency, intensity, and duration of activities (Matthews 2002; Ainsworth 2009). Recalling activity from a previous day or week reduces the effects of reactivity compared to records or logs, but can still be difficult for respondents. The entire process of answering a question from a questionnaire requires question comprehension (i.e., an understanding of the question), a decision about the question (i.e., if the question is clear and answerable), retrieval from memory (i.e., the gathering of information to answer the question), and response generation (i.e., organizing the memories into a verbal or written response) (Matthews 2002). Such a complex process can often lead to misreporting on activity (Ainsworth 2009; Matthews 2002).

Quantitative history questionnaires are more detailed than recall questionnaires and require individuals to respond to anywhere from 15 to 60 questions about physical activity from their

past (Matthews 2002; Ainsworth 2009). These instruments are useful for estimating energy expenditure and MVPA from the previous day (Matthews et al. 2000), week (Sallis et al. 1985) and month (Dipietro et al. 1993), and are also useful for gathering information on where activities are occurring (e.g., at home, at work, or in transit). These types of questionnaires usually take considerable time to administer and may be inappropriate for some large-scale surveys settings (Ainsworth 2009).

1.4 Modeling Physical Activity

In many physical activity studies, researchers are interested in studying usual or habitual physical activity in a population (Shephard 2003), where usual physical activity broadly refers to long-term average physical activity. More specifically, an individual’s usual daily energy expenditure is his or her average daily energy expenditure over a long period of time, such as one year. An individual’s measurement of physical activity from a day or week will be different from his or her usual physical activity level because of daily changes in physical activity and because of measurement errors. Consequently, using unadjusted physical activity measurements to estimate usual daily physical activity in the population may lead to biased estimates of usual physical activity parameters. Statistical models can be utilized to account and adjust for the errors and biases in physical activity data, which allows for more accurate estimation of usual physical activity parameters. In this section, we highlight sources of variation and bias in physical activity data (Section 1.4.1), provide a brief introduction to measurement error models using a simple example (Section 1.4.2), and discuss statistical modeling approaches that are relevant to physical activity research (Section 1.4.3) and dietary intake research (Section 1.4.4).

1.4.1 Variation and Bias in Physical Activity Measurement

The difference in observed physical activity and usual physical activity for an individual is generally attributed to measurement errors and nuisance effects, while the between-individual variation in usual physical activity may be influenced by other demographic indicators such as age, gender, and race/ethnicity. We define measurement error to be the difference between a

measurement of physical activity and the actual value of physical activity for a given day. For example, if Y_{ij} is a measurement of energy expenditure in MET-minutes for individual i on day j and T_{ij} is the actual energy expenditure in MET-minutes for individuals i on day j , then the measurement error in the measurement is $E_{ij} = Y_{ij} - T_{ij}$. Measurement error can exist in physical activity data collected from any type of instrument. Measurement error in DLW measurements is minimal and is usually due to bodily changes that occur naturally (Schoeller and van Santen 1982). Measurement error in monitor-based measurements is mainly due to the inability of monitors to capture the full range of activities in which an individual engages (Welk et al. 2004; Moy et al. Submitted). For example, if an individual wears an accelerometer around his waist and engages in activity with lots of upper arm movement, the accelerometer may not register all of the upper arm movement, and the measurement of activity may be less than the actual amount of activity the individual engaged in, resulting in measurement error. Errors in measurements from monitors may also come from calibration, where monitor data are converted into physical activity metrics (Welk 2002b). If the calibration function used to convert monitor data into measurements of energy expenditure is not properly specified, measurements may be inaccurate. Measurement error in self-report data exists because individuals do not always accurately report on their activity. Social desirability effects may influence individuals to report more activity than they actually do (Adams et al. 2005; Warnecke et al. 1997). Cognitive limitations associated with recalling activity may cause individuals to underreport on their activity if they forget certain activities they engaged in during the previous day or week (Bassett et al. 2000; Matthews 2002). The terminology used in physical activity questionnaires may be confusing to individuals and lead to misreporting on activity (Sallis and Saelens 2000). For example, if a questionnaire asks an individual to report on his or her moderate intensity activity, the individual may interpret moderate activity as only activity related to exercise and not activity related to household chores such as mowing the lawn or cleaning the house, which may be performed at a moderate intensity level. Reactivity is another factor that may lead to measurement error in either accelerometer or self-reported physical activity data, because individuals may engage in, or report engaging in, more activity than they would

do normally (Matthews 2002).

Individuals deviate from their usual physical activity levels on a short-term basis. That is, an individual may be more or less active than he or she usually is on any given day. For example, if T_{ij} is the actual energy expenditure in MET-minutes for individual i on day j and T_i is the usual daily energy expenditure of individual i , then the deviation in actual energy expenditure relative to usual daily energy expenditure is $D_{ij} = T_{ij} - T_i$. This difference between actual physical activity and usual physical activity can be attributed to nuisance effects, which cause individuals to change their physical activity habits on a short-term basis. One nuisance effect considered by researchers is seasonality (Matthews et al. 2001; Levin et al. 1999). An individual may be more active than he or she usually is in the summer because the warmer summer weather allows the individual to engage in more outdoor activity. On the other hand, an individual may be less active than he or she usually is in the winter because the colder winter weather keeps the individual indoors more often. Another nuisance effect is day-of-week effect (Matthew et al. 2002). An individual may be more active than he or she usually is on the weekend when there is more time to exercise and may be less active on weekdays when there is less time to exercise because of work. Time-in-sample effect is a third factor that may lead to differences between actual physical activity and usual physical activity. An individual may change his or her physical activity in response to participating in a survey.

Individuals in the population have varying levels of usual physical activity because some individuals are, on average, more active than other individuals. Demographic factors such as gender (Troiano et al. 2008; Ainsworth 2009; Ferrari et al. 2007), age (Troiano et al. 2008; Irwin et al. 2001; Ainsworth 2009), race and/or ethnicity (Marshall et al. 2007), and educational status (Hebert et al. 2002; Lagerros et al. 2006) are all factors that may be associated with variation in usual physical activity. For example, men may be more active than women because of social pressures, which encourage men to engage in more intense activity on a regular basis. Younger adults in the population may be more active than older adults because younger adults are generally more capable of engaging in more intense activity for longer periods of time than older adults. Individuals from a certain race or ethnicity group

may be, on average, more active than individuals in the general population because of cultural or ethnic traditions that encourage engagement in intense physical activity. Individuals with less education may be more active than individuals with higher education because many of the jobs performed by individuals with less education may be more physically strenuous than jobs performed by individuals with more education.

1.4.2 A Measurement Error Model

Measurement error models (Fuller 1987; Carroll et al. 2006) can be used to model the variation and bias in physical activity data and to estimate parameters of usual daily physical activity for the population. In this section, we present a simple measurement error model to motivate the use of measurement error models for assessment of physical activity data. More complex measurement error models are presented in Sections 1.4.3 and 1.4.4.

Suppose that a simple random sample of n individuals is selected and measured for physical activity using a monitoring device that is known to provide fairly accurate measurements of physical activity. Let T_i be the true usual daily physical activity for individual i and let Y_{ij} be a monitor measurement of physical activity for individual i on day j , where $j = 1, 2$. Assume that T_i and Y_{ij} are given in the same units (e.g., kcal/d or MET-hours/d). Consider the measurement error model,

$$Y_{ij} = T_i + e_{ij}. \quad (1.1)$$

Under this model, the term e_{ij} accounts for the difference between Y_{ij} and T_i due to measurement error and day-to-day variability in physical activity. Model (1.1) is often referred to as the classical measurement error model (Carroll et al. 2006).

The parameters of model (1.1) can be estimated given model assumptions. For example, assume that $T_i \stackrel{ind}{\sim} (\mu_T, \sigma_T^2)$, $e_{ij} \stackrel{ind}{\sim} (0, \sigma_e^2)$, and that $Cov(T_i, e_{ij}) = 0$ for all i and j . Let

$$\mathbf{Z}_i = \begin{pmatrix} \frac{Y_{i1} + Y_{i2}}{2} \\ Y_{i1} - Y_{i2} \end{pmatrix},$$

and let

$$\mathbf{m}_1 = n^{-1} \sum_{i=1}^n \mathbf{Z}_i$$

and

$$\mathbf{m}_2 = (n-1)^{-1} \sum_{i=1}^n (\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})'$$

Then

$$E\{\mathbf{m}_1\} = \begin{pmatrix} \mu_T \\ 0 \end{pmatrix}$$

and

$$E\{\mathbf{m}_2\} = \begin{pmatrix} \sigma_T^2 + 0.5\sigma_e^2 & 0 \\ 0 & 2\sigma_e^2 \end{pmatrix}.$$

By equating the sample moments to their expectations, we obtain the method of moments estimators

$$\begin{aligned} \hat{\mu}_T &= m_1, \\ \hat{\sigma}_T^2 &= m_{11} - 0.25m_{22}, \\ \hat{\sigma}_e^2 &= 0.5m_{22}, \end{aligned}$$

where m_1 is the first element in \mathbf{m}_1 , and m_{11} and m_{22} are the first and second diagonal elements in \mathbf{m}_2 , respectively.

Under model (1.1), a measurement of physical activity is assumed to be unbiased for true usual physical activity for individual i in that

$$E\{Y_{ij}|i\} = T_i.$$

This assumption may be violated when self-report instruments are used to measure physical activity, because individuals are known to misreport on their activity (see Section 1.4.1). An alternative measurement error model for self-report measurements is

$$Y_{ij} = \beta_0 + \beta_1 T_i + e_{ij}, \tag{1.2}$$

where the parameters β_0 and β_1 account for a systematic linear bias in the self-reported measurements. Given the same model assumptions as before,

$$E\{Y_{ij}|i\} = \beta_0 + \beta_1 T_i,$$

and Y_{ij} is a biased measure of usual physical activity when $(\beta_0, \beta_1) \neq (0, 1)$. Measurements from an unbiased reference instrument are needed to estimate the bias parameters β_0 and β_1 from model (1.2). Models of this form are discussed in more detail in the following sections.

1.4.3 Physical Activity Models

To date, use of measurement error models in assessment of physical activity data is limited to two papers, Ferrari et al. (2007) and Spiegelman et al. (1997). In both papers, the authors develop models as a means to validate self-report instruments for measuring physical activity. These models provide context for measurement error model development in Chapter 3.

1.4.3.1 Ferrari et al. Model

Ferrari et al. (2007) developed a model for sources of measurement error in physical activity data obtained from a study conducted at the Alberta Cancer Board from 2002-2003 (Friedenreich et al. 2006). One of the goals of the study was to validate a self-administered physical activity questionnaire which measured physical activity over the course of one year. One hundred and fifty four individuals were recruited to complete the study. During the course of one year, each study participant wore an accelerometer for four 1-week periods approximately 12 weeks apart. After wearing the accelerometer, each participant completed a physical activity log during a second 1-week period. At the end of the year, each participant completed the physical activity questionnaire.

The authors define T_i to be true usual weekly physical activity in MET-hours/week for individual i , and define a three-equation measurement error model relating true usual activity

to measured activity for the three instruments as

$$\begin{aligned} Q_i &= \alpha_Q + \beta_Q T_i + \epsilon_{Qi} \\ R_{ij} &= T_i + \epsilon_{Rij} \\ A_{ij} &= \alpha_A + \beta_A T_i + \epsilon_{Aij}, \end{aligned} \tag{1.3}$$

where Q_i is measured activity in MET-hours/week from the questionnaire for individual i , R_{ij} is measured activity in MET-hours/week from the log for individual i during week j , A_{ij} is measured activity in MET-hours/week from the accelerometer for individual i during week j , and $\epsilon_{Qi} \sim (0, \sigma_{\epsilon_Q}^2)$, $\epsilon_{Rij} \sim (0, \sigma_{\epsilon_R}^2)$, and $\epsilon_{Aij} \sim (0, \sigma_{\epsilon_A}^2)$ for all i and j . The fixed α and β parameters in the model capture the systematic component of measurement error, while the ϵ terms in the model capture the random component of measurement error. The authors assume that true usual activity is uncorrelated with the measurement error terms, that $Cov\{\epsilon_{Qi}, \epsilon_{Aij}\} = Cov\{\epsilon_{Rij}, \epsilon_{Aij}\} = 0$ for all i and j , and that $Cov\{\epsilon_{Qi}, \epsilon_{Rij}\} \neq 0$. To identify parameters from the model, one of the instruments must serve as a reference instrument that is assumed to provide unbiased measurements of usual activity. The authors chose the physical activity log as the reference instrument over the accelerometer because of concerns about the ability of accelerometers to accurately measure certain types of activity (Matthews 2005).

The authors are primarily interested in estimation of a slope attenuation factor for the physical activity questionnaire, which is defined as the slope in the linear calibration model

$$T_i = \lambda_0 + \lambda_{QT} Q_i + \xi_i,$$

where ξ_i is a random error term with 0 mean. Under model (1.3), the attenuation factor is

$$\lambda_{QT} = \frac{\beta_Q \sigma_T^2}{\beta_Q^2 \sigma_T^2 + \sigma_{\epsilon_Q}^2}.$$

A value of λ_{QT} close to one would indicate that there is little effect from measurement error when studying the relationship between true activity and measured activity using the questionnaire. On the other hand, a value of λ_{QT} close to zero would suggest a considerable effect from measurement error that may limit the ability to estimate usual physical activity from the questionnaire measurements without a bias adjustment.

The physical activity measurements from the Alberta study were log transformed for model fitting. To account for gender effects, the log-transformed measurements were regressed on gender and the residuals from the fitted regressions were used for model fitting. The measurement error model (1.3) was fit using maximum likelihood under the assumption that the random model terms were normally distributed. The estimated attenuation factor for the overall sample was 0.13 with a confidence interval of (0.05, 0.23). For men, the estimate was 0.23 with a confidence interval of (0.09, 0.41), and for women, the estimate was 0.07 with a confidence interval of (-0.03, 0.18). Given these results, the authors conclude that there is evidence of bias in both the female and male measurements of physical activity using the questionnaire.

1.4.3.2 Spiegelman et al. Model

Spiegelman et al. (1997) also consider a measurement error model for validating a physical activity questionnaire. The model is developed for physical activity data that come from the Health Professionals Follow-up Study (Grobbee et al. 1990). Study participants were measured for physical activity in MET-hours/week using physical activity logs and physical activity questionnaires. Each participant completed a physical activity log during four 1-week periods over the course of 1 year and completed a questionnaire at the end of the year, which asked about frequency and duration of activities from the past year. A measurement of physical fitness was also taken from each study participant based on change in pulse rate before and after a step test. The measurement error model is

$$\begin{aligned} X_{ij} &= T_i + e_{Xij} \\ Z_i &= a + bT_i + e_{Zi} \\ W_i &= c + dT_i + e_{Wi}, \end{aligned} \tag{1.4}$$

where $T_i \sim (\mu_T, \sigma_T^2)$ is true usual physical activity in MET-hours/week for individual i , X_{ij} is the j th unbiased measurement of physical activity in MET-hours/week for individual i from a physical activity log, Z_i is a measure of physical activity in MET-hours/week for individual i from a physical activity questionnaire, W_i is a measure of physical fitness for individual i (units not given), and $e_{Xij} \sim (0, \sigma_{eX}^2)$, $e_{Zi} \sim (0, \sigma_{eZ}^2)$, and $e_{Wi} \sim (0, \sigma_{eW}^2)$ are random

measurement error terms. The parameters a and b account for a systematic measurement error in the questionnaire measurements and the parameters c and d account for a linear relationship between physical fitness and usual physical activity. The authors assume zero correlation between true usual activity (T_i) and each of the random measurement error terms (e_{Xij} , e_{Zi} , and e_{Wi}) for all i and j . The authors also assume that $Cov\{e_{Xij}, e_{Zi}\} \neq 0$, but that $Cov\{e_{Xij}, e_{Wi}\} = Cov\{e_{Zi}, e_{Wi}\} = 0$ for all i and j .

The model (1.4) was fit to the physical activity data using method of moments. The slope attenuation factor of the physical activity questionnaire is

$$\lambda_{ZT} = \frac{b\sigma_T^2}{b^2\sigma_T^2 + \sigma_{eZ}^2},$$

and was estimated to be 0.30 with a 95% confidence interval of (0.21, 0.39). This estimated attenuation factor is similar to the estimated attenuation factor for men given in Ferrari et al. (2007), which was 0.23 with a 95% confidence interval of (0.09, 0.41). Given these results, there is evidence of bias in the physical activity questionnaire.

There are a number of similarities in the Spiegelman et al. and Ferrari et al. models. Both models are developed to investigate the validity of a physical activity questionnaire. Both models are three equation models and account for replicate measures of physical activity from a reference instrument, which is assumed to provide unbiased measurements of usual physical activity. Similar model assumptions are also considered in both papers to allow for model identifiability. The Spiegelman et al. model includes an additional model equation for an instrumental variable (physical fitness), while the Ferrari et al. model includes an additional model equation for an alternative measure of physical activity from an accelerometer.

1.4.4 Dietary Intake Models

In many dietary intake studies, researchers are interested in assessment of usual (long-term average) intake of nutrients and foods in a population (Nusser et al. 1996; Carriquiry 2003; Dodd et al. 1996). Because usual intakes are unobservable, daily measurements of intake are taken from individuals in the population. Like physical activity measurements, food and

nutrient intake measurements are subject to measurement error and other nuisance effects. Hence, measurement error models are developed to account and adjust for the errors.

Compared to the physical activity literature, the dietary intake literature offers more extensive research on measurement error model development. Models have been considered for intake variables that are consumed on a nearly daily basis, such as nutrients and energy intake, and for intake variables that are episodically consumed, such as foods. We limit our review to models for dietary intake variables that are consumed on a nearly daily basis because these models are more appropriate for the activity metrics we consider in our research. We review methods used for estimating the distributions of usual intake of nutrients (Nusser et al. 1996; Dodd et al. 2006; Carriquiry 2003) and methods used for estimating the error structure in observed nutrient intake data obtained from multiple instruments (Kipnis et al. 2003; Spiegelman et al. 2005; Rosner et al. 2008).

1.4.4.1 Estimating Usual Intake Distributions of Nutrients

One objective in dietary intake research is to estimate the distribution of usual intake of nutrients for a population (Carriquiry 2003). There is considerable within-individual variation in daily intake of nutrients, which when unaccounted for, induces excess variation and bias in estimated distributions of usual daily intake for a population. To address this issue, the National Research Council (1986) proposed a method for estimating usual daily intake distributions, which involved shrinking the individual mean intakes towards the group mean intake. This method is often referred to as the NRC method (Carriquiry 2003; Dodd et al. 2006). A more extensive method for estimating usual intake distributions of nutrients, known as the ISU method, was developed by Nusser et al. (1996) as an alternative to the NRC method. Dodd et al. (2006) outlines an abbreviated version of the ISU method, known as the best power (BP) method, which was used by Nusser et al. (1996) in a simulation study to evaluate the use of a semi-parametric transformation to normality. In this section, we review each of these methods.

First, we describe the NRC method. Let Y_{ij} be a measure of nutrient intake from a 24-

hour recall for individual i on day j , where $j = 1, \dots, d$. If the data are not nearly normal, a log or power transformation is applied to the data to better approximate normality. Let $y_{ij} = h(Y_{ij})$ be the value of Y_{ij} in the transformed scale and let $h(\cdot)$ be the log or power transformation that produces data that are more nearly normally distributed. To estimate usual daily intake values, the individual means are shrunk towards the group mean. The adjusted usual daily intake value for individual i is

$$\hat{t}_i = \bar{y}_{..} + \frac{\hat{\sigma}_t^2}{\hat{\sigma}_t^2 + \hat{\sigma}_e^2/d}(\bar{y}_{i.} - \bar{y}_{..}),$$

where $\bar{y}_{..}$ is the overall mean of the intake measurements, $\bar{y}_{i.}$ is the mean of the intake measurements of individual i , $\hat{\sigma}_t^2$ is the estimated inter-individual variance of the intake measurements, and $\hat{\sigma}_e^2$ is the estimated within-individual variance of the intake measurements. An estimate of usual daily intake in the original scale, \hat{T}_i , is obtained by applying the inverse of the transformation for normality, so that $\hat{T}_i = h^{-1}(\hat{t}_i)$, where $h(\cdot)$ is the transformation used to approximate normality. The set of back-transformed values, $\{\hat{T}_i\}$, can be used to obtain an empirical estimate of the usual daily intake distribution. Quantiles, means, and standard deviations of usual daily nutrient intake can be estimated from this empirical distribution.

A number of concerns with the NRC method have been discussed in the literature. The simple power or log transformation used in the NRC method may not approximate normality well for nutrient intake data, and in many cases more complex transformations are necessary to achieve normality (Nusser et al. 1996; Carriquiry 2003). Also, using the simple inverse of the power or log transformation to estimated usual daily nutrient intake values in the original scale can introduce bias in the original-scale intake distribution. Because the mean of a log or power transformed variable is not equal to the transformed mean of the original-scale variable, the NRC method will generate biased estimates of usual intake parameters in the original scale (Carriquiry 2003). The NRC method also assumes that the within-individual variances of daily intake are homogeneous across individuals, which is not necessarily guaranteed.

The ISU method (Nusser et al. 1996) was developed to account for some of the concerns raised about the NRC method. In the ISU method, the original nutrient intake data (Y_{ij}) are transformed into the normal scale in a series of steps. First, the daily intake data are

transformed using a power transformation. The “best” power for this transformation is selected by minimizing the error sum of squares

$$\sum_{i=1}^n \sum_{j=1}^d (U_{ij} - \alpha_0 - \alpha_1 Y_{ij}^\gamma)^2$$

over a grid of γ values, where U_{ij} is the normal score for the ij th observation in the dataset, and α_0 and α_1 are estimated for each value of γ . Let γ^* be the value that minimizes the error sum of squares based on the grid search and let $y_{ij}^* = Y_{ij}^{\gamma^*}$. If γ^* is zero, then $y_{ij}^* = \log(Y_{ij})$. Next, the transformed data are adjusted for nuisance effects, such as day-of-week, interview mode, and interview sequence effects. A model is fit to the y_{ij}^* data containing variables for these nuisance effects. The adjusted value for daily intake of individual i on day j is $y_{ij}^{**} = (1/\hat{y}_{ij}^*) \bar{y}_{.1} y_{ij}^*$, where \hat{y}_{ij}^* is the predicted value of y_{ij}^* from the fitted model with nuisance effects and $\bar{y}_{.1}$ is the mean of the y_{ij}^* values for the first interview day. The data are adjusted to the mean of the first interview day because the data are believed to be more accurate on the first interview day (Nusser et al. 1996). Next, a grafted cubic polynomial is fit to the (U_{ij}, y_{ij}^{**}) pairs. The number of join points used to construct the polynomial is chosen to be the minimum number of join points required to make the value of the Anderson-Darling test statistic less than or equal to a cutoff value of 0.58 (p-value of 0.15) when applied to the data from the polynomial fit (Nusser et al. 1996). The grafted polynomial is used instead of a log or power transformation because the polynomial adjustment gives a better approximation to normality, especially in the tails of the distribution. Let y_{ij} be the estimated value of y_{ij}^{**} from the polynomial fit.

After the transformation to normality, the next step in the ISU method is to fit the measurement error model

$$y_{ij} = t_i + u_{ij},$$

where $t_i \sim N(\mu_t, \sigma_t^2)$, $u_{ij} \sim N(0, \sigma_{ui}^2)$, and $\sigma_{ui}^2 \sim (\mu_A, \sigma_A^2)$. The u_{ij} are assumed to be independent given i , and t_i and u_{lj} are assumed to be independent for all i , l , and j . The distribution of σ_{ui}^2 accounts for heterogeneity in the within-individual error variances. The parameters in the model are estimated using method of moments. Let $\hat{\mu}_t$ and $\hat{\sigma}_t^2$ be estimates

of μ_t and σ_t^2 , respectively, from the model fit. Then, the estimated distribution of usual daily nutrient intake in the normal scale is $N(\hat{\mu}_t, \hat{\sigma}_t^2)$.

The final step of the ISU method is to transform usual intake values in the normal scale back to the original scale. Let $g(\cdot)$ denote the transformation taking the adjusted observed intakes to normality, let \ddot{T}_i denote the true usual daily intake for individual i in the original scale, and let \ddot{t}_i denote the true usual daily intake for individual i in the normal scale. Then

$$\ddot{T}_i = E\{y|t = \ddot{t}_i\} = E\{g^{-1}(t + u)|t = \ddot{t}_i\} = h(\ddot{t}_i),$$

where $h(\cdot)$ is the implicit transformation taking the normal-scale usual daily intake values into the original scale that must be estimated by approximating the conditional expectation of y given t for a set of values \ddot{t}_i and then fitting a grafted polynomial to the (\ddot{T}_i, \ddot{t}_i) pairs (Nusser et al. 1996). The set of \ddot{t}_i values used for this procedure is a set of 400 values, where the first five moments of the set of values match the first five moments of a $N(0, \hat{\sigma}_t^2)$ distribution. At each value of \ddot{t}_i , the usual intake value in the original scale is approximated by

$$\ddot{T}_i = \sum_{l=-4}^4 w_l g^{-1}(\ddot{t}_i + c_l),$$

where the nine points c_l and the nine weights w_l , with $\sum w_l = 1$, are constructed such that the first five moments of the discrete nine-point distribution match the first five estimated moments of the conditional distribution of $\ddot{t} + u$ conditional on \ddot{t} (Nusser et al. 1996). The 400 \ddot{T}_i values provide an estimated usual daily intake distribution and a grafted cubic polynomial fit to the pairs (\ddot{T}_i, \ddot{t}_i) , denoted by \hat{h} , is an estimator of the transformation taking the normal-scale usual daily intake values (\ddot{t}_i) into the original scale. The estimated function \hat{h} can be used to compute usual intake values in the original scale using normal-scale usual intake values.

Finally, we give a brief review of the BP method as given in Dodd et al. (2006). This method is similar to the NRC method, but uses a bias correction for transforming estimated usual daily intake values from the normal scale back into the original scale. In the method, the nutrient intake values in the original scale, Y_{ij} , are transformed using a power transformation, where the transformed values approximate normality. Let $y_{ij} = g(Y_{ij})$ be the transformed value of Y_{ij} , where $g(\cdot)$ is the transformation function. The measurement error model of y_{ij} is

given by

$$y_{ij} = t_i + w_{ij},$$

where $t_i \sim N(\mu_t, \sigma_t^2)$, $w_{ij} \sim N(0, \sigma_w^2)$, and $Cov(t_i, w_{ij}) = 0$ for all i and j . A set of t_i^* values are generated from the $N(\hat{\mu}_t, \hat{\sigma}_t^2)$ distribution, where $\hat{\mu}_t$ and $\hat{\sigma}_t^2$ are estimates of μ_t and σ_t^2 from the fitted model. A value for t_i^* in the original scale is

$$T_i^* = h(t_i^*) + (1/2)h''(t_i^*)\hat{\sigma}_w^2,$$

where $h(\cdot) = g^{-1}(\cdot)$, $h''(\cdot)$ is the second derivative of $h(\cdot)$, and $\hat{\sigma}_w^2$ is the estimate of σ_w^2 from the measurement error model. This derivation is based on the second-order Taylor expansion

$$\begin{aligned} T_i &= E\{h(t+w)|t=t_i\} \\ &\approx h(E\{t+w|t=t_i\}) + h'(E\{t+w|t=t_i\})E\{(t+w) - E\{t+w|t=t_i\}|t=t_i\} \\ &\quad + (1/2)h''(E\{t+w|t=t_i\})E\{(t+w) - E\{t+w|t=t_i\}|t=t_i\}^2 \\ &= h(E\{t+w|t=t_i\}) + (1/2)h''(E\{t+w|t=t_i\})Var\{t+w|t=t_i\} \\ &= h(t_i) + (1/2)h''(t_i)\sigma_w^2. \end{aligned}$$

The BP method offers a simple alternative to the ISU method because a single power transformation is used to approximate normality and to back transform values into the original scale instead of a two-stage transformation involving a cubic polynomial. But, the BP method may not be appropriate when nutrient intake data cannot be made approximately normal using a simple power or log transformation (Carriquiry 2003). Also, the BP method does not account for heterogeneous error variances across individuals, which may exist in nutrient intake data.

1.4.4.2 Measurement Error Structure in Nutrient Intake Models

A second objective in dietary intake research is to evaluate the validity of self-report instruments for measuring usual daily intake of nutrients. In many dietary intake studies, food frequency questionnaires are used to measure nutrient intake from a large sample of individuals and 24-hour dietary recalls are used as an unbiased reference instruments to calibrate

or adjust for biases in the the food frequency questionnaires. This approach has been shown to be problematic, since 24-hour dietary recalls, like food frequency questionnaires, may give biased estimates of nutrient intake. Researchers have developed measurement error models to investigate the error structure in nutrient intake data as a means for evaluating the validity of both food frequency questionnaires and 24-hour dietary recalls. We review one such model presented in Kipnis et al. (2003).

The Kipnis et al. (2003) model was developed for data from the Observing Protein and Energy Nutrition (OPEN) study, where approximately 500 adults aged 40-69 years completed multiple food frequency questionnaires and 24-hour dietary recalls during September 1999 to March 2000. Every participant was also measured for energy intake using doubly labeled water (DLW) and protein intake using urinary nitrogen measurements. A subsample of the participants provided multiple DLW and urinary nitrogen measurements. In our review, we consider the Kipnis et al. model for measurements of energy intake.

Let T_i denote true, usual daily energy intake for individual i and let Q_{ij} , F_{ij} , and M_{ij} be estimated energy intake for individual i on day j using a food frequency questionnaire (FFQ), a 24-hour dietary recall, and a reference biomarker (doubly labeled water), respectively. The model equation for the FFQ-derived intake is

$$Q_{ij} = \beta_{Q0} + \beta_{Q1}T_i + \mu_{Qj} + r_i + \epsilon_{ij},$$

where μ_{Qj} is a time-specific effect for the j th measurement, $\beta_{Q0} + \beta_{Q1}T_i + r_i$ represents the within-person bias in the measurement with a systematic component ($\beta_{Q0} + \beta_{Q1}T_i$) and random component (r_i), and ϵ_{ij} represents within-person variation. The random terms in the model are $T_i \sim (\mu_T, \sigma_T^2)$, $r_i \sim (0, \sigma_r^2)$, and $\epsilon_{ij} \sim (0, \sigma_\epsilon^2)$ and are assumed to be uncorrelated for all i and j . The remaining terms in the model, μ_{Qj} , β_{Q0} , and β_{Q1} , are fixed. The model equation for energy intake measured from the the dietary recall is

$$F_{ij} = \beta_{F0} + \beta_{F1}T_i + \mu_{Fj} + s_i + u_{ij}$$

and is similar to the model equation for the FFQ in that it contains a time-specific group effect term, μ_{Fj} , an individual-level bias model with systematic and random components,

$\beta_{F0} + \beta_{F1}T_i + s_i$, and a within-person error term, u_{ij} . $T_i \sim (\mu_T, \sigma_T^2)$, $s_i \sim (0, \sigma_s^2)$, and $u_{ij} \sim (0, \sigma_u^2)$ are assumed to be uncorrelated for all i and j and the remaining terms in the equation, μ_{Fj} , β_{F0} , and β_{F1} , are fixed. The third and final equation for the biomarker is

$$M_{ij} = T_i + \mu_{Mj} + v_{ij},$$

where μ_{Mj} is a time-specific group effect term, $v_{ij} \sim (0, \sigma_v^2)$ is a within-person error term, and T_i and v_{ij} are assumed to be uncorrelated. The authors assume that T_i is uncorrelated with the individual-level bias terms, r_i and s_i , for all i . The terms r_i and s_i are assumed to be correlated with each other, and are assumed to be uncorrelated with the model error terms ϵ_{ij} , u_{ij} , and v_{ij} for all i and j . The ϵ_{ij} , u_{ij} , and v_{ij} terms are assumed to be uncorrelated with each other except when measurements are taken contemporaneously, in which case the pairs (ϵ_{ij}, u_{ij}) , (ϵ_{ij}, v_{ij}) , and (u_{ij}, v_{ij}) are assumed to be correlated.

The Kipnis et al. model was fit to the OPEN energy intake data using maximum likelihood under the assumption of normality. Before fitting, the energy intake measurements were log transformed to better approximate normality. Extreme outlying values were excluded from the analysis. To evaluate the validity of the food frequency questionnaire against the DLW measurements, the authors estimated the slope attenuation factor

$$\lambda_Q = \frac{\beta_{Q1}\sigma_T^2}{\beta_{Q1}^2\sigma_T^2 + \sigma_r^2 + \sigma_\epsilon^2}$$

for males and females separately. The estimate was 0.080 for males with a standard error of 0.025 and was 0.039 for females with a standard error of 0.028. To evaluate the validity of the food frequency questionnaire against the 24-hour dietary recall, the authors assume the 24-hour dietary recall is the unbiased reference instrument and fit the reduced model

$$\begin{aligned} Q_{ij} &= \beta_{Q0} + \beta_{Q1}T_i + \mu_{Qj} + r_i + \epsilon_{ij}, \\ F_{ij} &= T_i + \mu_{Fj} + u_{ij}, \end{aligned}$$

to the OPEN data and estimate the same attenuation factor λ_Q . The estimated attenuation factors based on the reduced model are higher for males (0.230 with a standard error of 0.037) and females (0.128 with a standard error of 0.044), relative to the estimates based on the

full model with the biomarker. The authors conclude that the attenuation factor λ_Q may be overestimated using only the 24-hour recall as a reference instrument and not the DLW biomarker, because of the potential bias in the 24-hour recall.

The structure of the Ferrari et al. model (1.3) is similar to the structure of the Kipnis et al. model. Both models include model equations for potentially biased measurements. The Kipnis et al. model includes equations for measurements from a food frequency questionnaire and 24-hour dietary recall and the Ferrari et al. model includes equations for measurements from a physical activity questionnaire and accelerometer. Both models also include a model equation for an unbiased reference instrument, which is DLW in the Kipnis et al. model and is a physical activity log in the Ferrari et al. model. Using DLW as the reference instrument seems more appropriate, since the physical activity log may give biased measurements of physical activity due to the nature of self-reporting.

The Ferrari et al. model assumes that the measurement errors from the questionnaire and physical activity log have a nonzero correlation, but that the measurement errors from the accelerometer are uncorrelated with the measurement errors from the self-reports. The Kipnis et al. model assumes that the individual-level bias terms from the FFQ and 24-hour recall are correlated and assumes that the measurement error terms are only correlated when the measurements are taken contemporaneously. Assumptions of this nature are necessary for model identifiability, which allows for estimation of model parameters given sample data. Other measurement error models have been given in the literature that go beyond the three-equation model structure presented in Kipnis et al. (2003) and Ferrari et al. (2007) in order to identify a larger set of model parameters. For example, Spiegelman et al. (2005) consider models where a fourth model equation is included for measurements from an instrumental variable to allow for identifiability of additional model parameters. Rosner et al. (2008) consider models where covariate information, such as BMI and smoking status, is included in the model equations to estimate relationships between the covariates and nutrient intake.

1.5 Summary

In this chapter, we have reviewed some of the methods used to measure physical activity and presented some of the models used to estimate usual (long-term average) physical activity and dietary intake variables. The general consensus in the literature is that self-report instruments are the most practical type of instrument for measuring physical activity from individuals in the population because self-report instruments are inexpensive to implement in large-scale surveys and are often of little burden to survey participants. But, self-report instruments are also subject to significant measurement errors and biases because individuals tend to misreport on their activity due to a variety of factors. As a result, assessment of physical activity using unadjusted self-report data may lead to significant biases, particularly in usual physical activity parameters.

In more recent research, monitoring devices have been considered as an alternative or companion to self-report instruments because monitors measure physical activity objectively (e.g., without self-reporting biases) and most contemporary monitors are small enough to be worn without much of a burden to the survey participants. When multiple concurrent measurements of physical activity are taken from individuals in the sample using a monitor instrument and self-report instrument, measurement error models can be used to estimate the various sources of variation and bias in the data and to estimate usual physical activity parameters after adjusting for excess variation and bias due to measurement error and nuisance effects. Measurement error model research has been well established for assessing dietary intake variables (Section 1.4.4), but to date, has only been considered in two papers for assessing physical activity variables (Section 1.4.3). The Ferrari et al. and Spiegelman et al. models presented in Section 1.4.3 assume that physical activity logs provide unbiased measurements of physical activity, which may be violated due to the nature of self-reporting on activity. The models are also considered for convenience samples, and not probability samples from the population.

The goal of our research is to build upon the measurement methods and models in the literature in order to develop models for physical activity data. We consider methods that account

for the bias associated with self-report instruments and allow researchers to make inferences about physical activity in subpopulations or target groups of the population. In Chapter 2 we develop a model for physical activity data from the National Health and Nutrition Examination Survey (NHANES). The NHANES data are collected from a representative sample of the United States using a questionnaire, which asks survey participants to report on their moderate to vigorous physical activity (MVPA) from the previous 30 days. The questionnaire data are subject to significant reporting errors because survey participants tend to have a difficult time remembering and accurately reporting on their physical activity over the course of a 30-day period. A convenience subsample of the survey participants wore accelerometers for a week after completing the questionnaire to provide a monitor-based measure of daily MVPA to go with their self-report measure of daily MVPA. Using these data, we develop a linear regression model that models accelerometer-based daily MVPA as a function of self-reported activity variables and other demographic variables. The fitted model can be used to estimate mean daily MVPA levels of demographic groups in the population. Using the model to estimate mean daily MVPA in groups of the population is a reasonable alternative to using the unadjusted self-report measurements of MVPA because of the significant reporting errors observed in the unadjusted self-report data.

In Chapter 3 we develop a method for estimating usual daily energy expenditure parameters from physical activity data collected using a self-report instrument and an unbiased objective monitoring device for at least a subsample of study respondents. Our method extends the methods considered in Sections 1.4.3 and 1.4.4, which utilize measurement error models for estimating usual physical activity and dietary intake variables. In our approach, a measurement error model is fit to daily measurements of total energy expenditure. Parameters of usual daily energy expenditure are estimated for subpopulations that may be determined by gender, age, or race/ethnicity. Researchers can then use the parameter estimates to compare EE behaviors across these subpopulations. We illustrate our method with preliminary data from a sample of females in the Physical Activity Measurement Survey (PAMS). The PAMS data are collected from a 24-hour physical activity recall and the SenseWear Pro armband monitor. Parameters

of usual daily EE are estimated for 4 age groups from the female sample.

CHAPTER 2 A REGRESSION MODEL FOR MODERATE TO VIGOROUS PHYSICAL ACTIVITY (MVPA)

2.1 Introduction

Accurate assessment of physical activity is a well-established public health priority (U.S. Department of Health and Human Services 1996). Estimates of physical activity for individuals in a group or population are often calculated using self-report instruments (Matthews 2002; Ainsworth 2009), which are relatively inexpensive to administer to large samples of the population. However, estimates from self-report instruments are also prone to significant measurement errors and biases due to the subjective nature of reporting on physical activity and instrument limitations (Matthews 2002; Sallis and Saelens 2000; Adams et al. 2005). Monitoring devices such as accelerometers (Welk 2002; Ward et al. 2005) offer more objective measurements of physical activity than self-report instruments, but are more expensive to implement in large-scale surveys and are often more of a burden to survey participants (Ward et al. 2005; Matthews 2005). In most large-scale surveys that include physical activity measurement (e.g., BRFSS, NHIS) the full sample is measured for physical activity using a self-report instrument. In some surveys (e.g., NHANES) a subsample is also measured for physical activity using monitor instruments. To obtain more objective estimates of physical activity for the full sample, the self-report measurements from the full sample can be adjusted or calibrated using statistical models that are estimated from the subsample containing both self-report and monitor-based measurements.

In the 2003-2004 and 2005-2006 cycles of NHANES, individuals from a subsample are measured for physical activity via accelerometers and physical activity questionnaires. We use the subsample of female adults (age 20 and older) from the 2003-2004 cycle of NHANES

to develop a linear regression model relating accelerometer physical activity to self-reported physical activity. We then fit the same model to the 2005-2006 female sample of adults to see if the model based on analyses of 2003-2004 data is generalizable to the 2005-2006 data, and develop a final model for female physical activity using both samples. In Section 2.2 we describe the NHANES physical activity data. In Section 2.3 we develop the regression model for female physical activity. In Section 2.4 we present examples for estimating and predicting physical activity. We conclude with a discussion in Section 2.5.

2.2 NHANES Physical Activity Data

NHANES is an ongoing survey of the United States civilian non-institutionalized population sponsored by the National Center for Health Statistics (NCHS), a branch of the Centers for Disease Control and Prevention (CDC). Survey participants provide health and nutrition data during interviews and medical examinations. In the 2003-2004 and 2005-2006 cycles of NHANES, physical activity data were collected for a subsample of survey participants using a physical activity questionnaire and Actigraph accelerometers.

The NHANES sample design is a stratified cluster design, where clusters or primary sampling units (PSUs) are selected from geographic strata that are subdivisions of the United States and individuals are selected from within the PSUs. To protect the confidentiality of survey participants, pseudo-strata and pseudo-PSUs are created for the NHANES samples and are used in place of the actual strata and PSUs for variance estimation. In the remainder of the presentation we will refer to the pseudo-PSUs and pseudo-strata as PSUs and strata, respectively.

The physical activity questionnaire asks participants to recall their physical activity from the past 30 days. Participants report on frequency and duration of activities related to transportation to and from work or school, or to do errands, activities related to household maintenance (e.g., raking leaves, mowing the lawn), and activities related to leisure (e.g., exercise, sports, and hobbies) (U.S. Department of Health and Human Services 2009). For reports on transportation and household activities, participants are asked to specify frequency and

duration of activities they do for at least 10 minutes that are at a moderate intensity level or higher. For reports on leisure activities, participants are asked to specify separately type, frequency, and duration of activities they do for at least 10 minutes that are at moderate and vigorous intensity levels. In the questionnaire documentation, moderate intensity activities are defined as activities causing “light sweating or a slight to moderate increase in heart rate or breathing” and vigorous intensity activities are defined as activities causing “heavy sweating or large increases in breathing or heart rate” (U.S. Department of Health and Human Services 2009). For each participant, an estimate of average daily time spent in moderate to vigorous physical activity (MVPA) is computed by adding up the minutes of reported moderate and vigorous transportation, household, and leisure activity from the 30 days of recall and dividing the total number of minutes by 30.

After completing the physical activity questionnaire, participants are asked to wear accelerometers for a week to further monitor their physical activity. The accelerometers are worn around the waist during all waking hours of the day and are taken off during water activities such as swimming and showering. The accelerometers measure duration and intensity of movement in activity counts and the activity counts are translated into periods of little or no intensity activity, moderate intensity activity, and vigorous intensity activity. The threshold for moderate intensity activity is 2020 counts and the threshold for vigorous intensity activity is 5999 counts (Troiano et al. 2008). Only periods of moderate and vigorous intensity activity lasting at least 10 minutes are considered. For each participant, an estimate of average daily MVPA is computed by adding up the minutes of measured moderate and vigorous activity and dividing by the total number of days worn.

For our analysis, we consider female participants age 20 years and older who completed the questionnaire and wore an accelerometer for at least 10 or more hours on 4 or more days. In the 2003-2004 NHANES there are 1569 such females (after removing an outlier), which we will denote as the 2003-2004 NHANES sample. In the 2005-2006 NHANES there are 1522 such females, which we will denote as the 2005-2006 NHANES sample. The outlier in the 2003-2004 NHANES sample was identified as having unrealistic physical activity reports.

The demographic decompositions of each NHANES sample are provided in Table 2.1. Both samples are distributed fairly uniformly across age groups. Over 90% of each sample contains participants who classify themselves as either non-Hispanic black, Mexican American, or non-Hispanic white.

Table 2.1 Demographic decomposition of the NHANES samples

2003-2004 Sample		2005-2006 Sample
Age Group	Count (%)	Count (%)
20-29	219 (14)	268 (18)
30-39	240 (15)	255 (17)
40-49	257 (16)	270 (18)
50-59	219 (14)	222 (14)
60-69	285 (18)	247 (16)
70-79	201 (13)	146 (10)
80+	148 (10)	114 (7)
Total	1569 (100)	1522 (100)
Race/Ethnicity	Count (%)	Count (%)
Black	260 (17)	331 (22)
Mexican	319 (20)	310 (20)
White	879 (56)	766 (50)
Other Hispanic	48 (3)	46 (3)
Other	63 (4)	69 (5)
Total	1569 (100)	1522 (100)

A survey weight is computed for each individual in the NHANES samples. The initial survey weight is the inverse of the individual's probability of being included in the sample. The final weight is the individual's initial weight adjusted for nonresponse and post-stratified to match 2000 U.S. Census population control totals for gender, age group, and race/ethnicity group. Percentiles for the distribution of the final survey weights are given in Table 2.2 for each of the NHANES samples.

Table 2.2 Percentiles for the distribution of final survey weights for the NHANES samples

Percentile	0	10	25	50	75	90	100
2003-2004 Survey Weight	1,569	5,214	16,888	28,254	50,485	70,621	105,962
2005-2006 Survey Weight	1,261	7,524	15,584	28,108	55,388	75,147	117,833

Survey weighted means of average daily accelerometer MVPA are given in Table 2.3 for the NHANES samples. Means are computed separately for age groups and race/ethnicity groups. Stratified cluster standard errors are computed for the means to take into account the NHANES complex sample design. See the SAS documentation on PROC SURVEYMEANS (SAS Institute 2009). In both samples, there is a noticeable drop in mean estimated average daily MVPA for the oldest age group. In the 2005-2006 sample, there is a noticeable difference in mean estimated average daily MVPA for the “White and Other” group compared to the Black and Mexican groups.

Table 2.3 Weighted means (standard errors) for average daily MVPA (min/day) measured by accelerometer for the NHANES samples

Age Group	2003-2004 Sample Mean (SE)	2005-2006 Sample Mean (SE)
20-40	6.93 (0.80)	5.81 (0.75)
41-60	5.81 (0.56)	5.86 (0.59)
61-75	5.04 (0.66)	4.74 (0.78)
76+	1.15 (0.44)	1.08 (0.40)
Race/Ethnicity	2003-2004 Sample Mean (SE)	2005-2006 Sample Mean (SE)
Black	5.07 (1.08)	3.32 (0.45)
Mexican	5.70 (0.52)	3.82 (0.48)
White and Other	5.90 (0.51)	5.71 (0.44)

A plot comparing average daily MVPA estimated from the accelerometers and average daily MVPA estimated from the questionnaires is provided in Figure 2.1 for the 2003-2004 NHANES sample. The plot omits 3 individuals with extreme questionnaire-based estimates above 700 minutes/day. There is a modest positive linear association between the accelerometer and questionnaire estimates based on the Pearson correlation coefficient ($r = 0.23$). This linear association is not noticeable in the plot. A majority of the points in the plot lie above the dashed identity line (about 76% of the points), suggesting that most individuals report more average daily MVPA than the accelerometers record. The plot comparing accelerometer and questionnaire average daily MVPA for the 2005-2006 NHANES sample shows a similar relationship. The Pearson correlation coefficient is also similar ($r = 0.27$).

There is a significant number of individuals in the NHANES samples with zero estimated

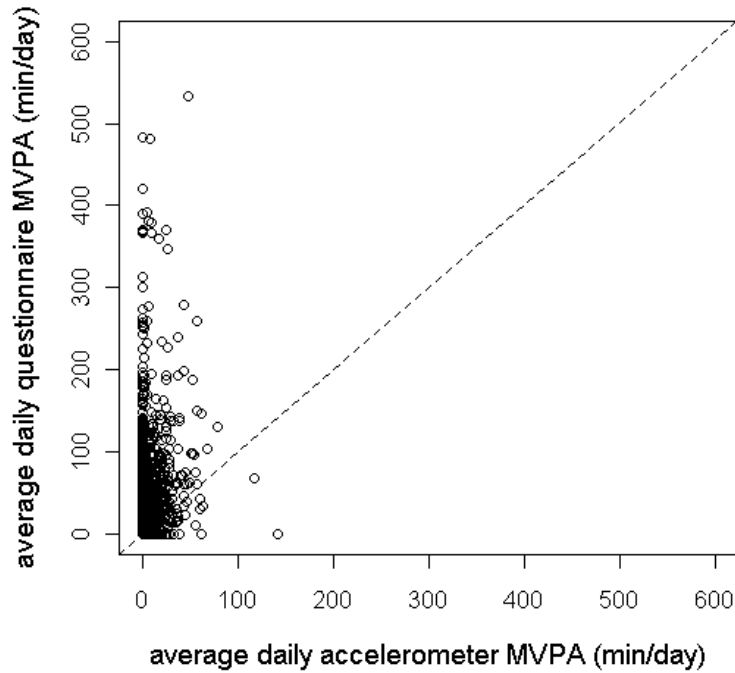


Figure 2.1 Plot of average daily MVPA for the 2003-2004 NHANES sample (dashed line is the identity line)

average daily MVPA (Table 2.4). Over 60% of the accelerometer measurements of average daily MVPA are zero in both NHANES samples. Around 20% of the questionnaire measurements are also zero in both samples. About half of the participants in both samples have contradicting estimates in that the estimate is zero based on one instrument and positive based on the other.

2.3 Regression Model for Female Physical Activity

In this section we develop a linear regression model for female physical activity. In our analyses, we use ordinary least squares (OLS) estimators to develop a preliminary model (Section 2.3.1) and use estimated generalized least squares (EGLS) estimators (Section 2.3.2) to estimate model parameters and compute standard errors for the final models. In our final analyses, we use EGLS procedures rather than design-based procedures, because EGLS may be more familiar to physical activity researchers than design-based estimation and because the

Table 2.4 Count (percent) of individuals with zero (0) and positive (> 0) estimated average daily MVPA for the NHANES samples

2003-2004 NHANES sample			
	0 Quest. MVPA	> 0 Quest. MVPA	Total
0 Accel. MVPA	270 (17)	702 (45)	972 (62)
> 0 Accel. MVPA	55 (3)	542 (35)	597 (38)
Total	325 (20)	1244 (80)	1569 (100)
2005-2006 NHANES sample			
	0 Quest. MVPA	> 0 Quest. MVPA	Total
0 Accel. MVPA	242 (16)	728 (48)	970 (64)
> 0 Accel. MVPA	41 (3)	511 (33)	552 (36)
Total	283 (19)	1239 (81)	1522 (100)

results for EGLS are similar to those for design-based estimation. We justify the use of EGLS procedures in Section 2.3.5.

2.3.1 Model Development

The physical activity and demographic variables we use for model development are given in Table 2.5. The model response variable is average daily MVPA measured by accelerometer. This variable is given in the original scale so that the model can be used for estimation of average daily accelerometer MVPA. The self-report physical activity variables *trans*, *mod*, and *vig* were truncated at their respective 99th percentiles in the original scales to account for cases of extreme over-reporting on physical activity. We did not include a variable for reported household activity because the estimated regression coefficient on the variable was non-significant in all of the models considered in our preliminary analyses. The self-report variables are in the cube-root scale. The cube-root transformation provided a better model fit than the square root and fourth root transformations. Using the *age* variable, we define the variables

$$age1 = \begin{cases} 30 & \text{if } age < 30 \\ 60 - age & \text{if } 30 \leq age \leq 60 \\ 0 & \text{if } age > 60 \end{cases} \quad (2.1)$$

and

$$age2 = \begin{cases} 15 & \text{if } age < 60 \\ 75 - age & \text{if } 60 \leq age \leq 75 \\ 0 & \text{if } age > 75. \end{cases} \quad (2.2)$$

These variables are defined so that the estimated coefficients are positive in the estimated model. The variable

$$mexblack = mex + black \quad (2.3)$$

is an indicator variable for being either non-Hispanic black or Mexican American. Whenever it is reasonable based on significance tests, we use the *mexblack* indicator variable in our analyses instead of separate indicator variables for *mex* and *black*, because preliminary tests suggest that the full models with the *mex* and *black* indicator variables are not significantly different than the reduced models with the *mexblack* indicator variable.

Table 2.5 Variables for model

Variable	Description
<i>y</i>	average daily accelerometer MVPA
<i>trans*</i>	self-reported average daily transportation activity
<i>mod*</i>	self-reported average daily moderate leisure activity
<i>vig*</i>	self-reported average daily vigorous leisure activity
<i>age</i>	age at time of screening for NHANES
<i>age1</i>	age group variable defined by (2.1)
<i>age2</i>	age group variable defined by (2.2)
<i>mex</i>	1 if Mexican American, 0 otherwise
<i>black</i>	1 if non-Hispanic black, 0 otherwise
<i>mexblack</i>	indicator variable defined by (2.3)

*Truncated at 99th percentiles and transformed to the cube-root scale

In developing a regression model, we first fit linear regression models for each of three race/ethnicity groups for females: non-Hispanic black, Mexican American, and other, where other includes non-Hispanic whites. Then we fit a final model for the full female sample using the information from the three initial model fits. The model for the three race/ethnicity groups

is

$$\begin{aligned} y_{hij} &= \beta_0 + \beta_1 trans_{hij} + \beta_2 mod_{hij} + \beta_3 vig_{hij} + \beta_4 age1_{hij} + \beta_5 age2_{hij} + e_{hij} \\ &= \mathbf{x}'_{hij} \boldsymbol{\beta} + e_{hij}, \end{aligned} \quad (2.4)$$

where the model variables are defined in Table 2.5 and the hij indexing on the variables refers to individual j , $j = 1, \dots, m_{hi}$, in PSU i , $i = 1, \dots, n_h$, in stratum h , $h = 1, \dots, H$. See Appendix B. For preliminary analyses, model (2.4) is fit to each of the three female race/ethnicity groups in the 2003-2004 NHANES sample using ordinary least squares (OLS). The OLS estimator is

$$\hat{\boldsymbol{\beta}}_{OLS} = \left(\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \mathbf{x}_{hij} \mathbf{x}'_{hij} \right)^{-1} \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \mathbf{x}_{hij} y_{hij}. \quad (2.5)$$

Standard errors of the estimated regression coefficients are computed using the Taylor linearization variance of $\hat{\boldsymbol{\beta}}_{OLS}$ given in Appendix B. The estimates and standard errors from the model fits are given in Table 2.6. The intercept coefficients in each of the model fits are non-significant and removed from consideration for the full female model. Similarly, the estimated coefficients on *age1* are all non-significant and removed from consideration for the full female model. The estimated coefficient on *mod* is larger for the “other” sample compared to the black and Mexican samples. To account for the difference, we define the model variable

$$mexblackmod = (mexblack)(mod)$$

for the full female model. The estimated coefficient on *vig* is smaller for the Mexican sample compared to the black and other samples. To account for the difference, we define the model variable

$$mexvig = (mex)(vig)$$

for the full female model. To account for a potential interaction between race/ethnicity and age, we also define the model variable

$$mexblackage = (mexblack)(age2)$$

for the full female model. The version of model (2.4) for the combined sample of all females is

$$\begin{aligned}
y_{hij} &= \beta_1 trans_{hij} + \beta_2 mod_{hij} + \beta_3 vig_{hij} + \beta_4 age2_{hij} + \beta_5 mexblackmod_{hij} \\
&\quad + \beta_6 mexvig_{hij} + \beta_7 mexblackage_{hij} + e_{hij} \\
&= \mathbf{x}'_{hij} \boldsymbol{\beta} + e_{hij}.
\end{aligned} \tag{2.6}$$

Estimates for the regression coefficients from this model are presented in Section 2.3.6.

Table 2.6 Estimated regression coefficients for model (2.4) fit to three female groups in the 2003-2004 NHANES sample using OLS

Variable	Black Est (SE)	Mexican Est (SE)	Other Est (SE)
Intercept	-1.788 (0.950)	0.667 (0.912)	-0.483 (0.433)
<i>trans</i>	1.231 (0.606)	2.436 (0.750)	2.250 (0.496)
<i>mod</i>	0.549 (0.431)	0.193 (0.389)	1.120 (0.245)
<i>vig</i>	3.328 (1.236)	0.565 (0.663)	2.051 (0.410)
<i>age1</i>	0.016 (0.092)	0.054 (0.055)	-0.057 (0.037)
<i>age2</i>	0.215 (0.129)	0.138 (0.112)	0.193 (0.058)

2.3.2 EGLS Estimator

Because of evidence of heterogeneity in the estimated error variances in preliminary analyses, we consider an estimated generalized least squares (EGLS) estimator for estimating regression coefficients from the full female model. The EGLS estimator of $\boldsymbol{\beta}$ for (2.6) is

$$\hat{\boldsymbol{\beta}}_{EGLS} = \left(\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \mathbf{x}_{hij} \hat{v}_{hij}^{-1} \mathbf{x}'_{hij} \right)^{-1} \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \mathbf{x}_{hij} \hat{v}_{hij}^{-1} y_{hij}, \tag{2.7}$$

where \hat{v}_{hij} is an estimator of v_{hij} and v_{hij} is the variance of e_{hij} in (2.6). Given regularity conditions, $\hat{\boldsymbol{\beta}}_{EGLS}$ is consistent for $\boldsymbol{\beta}$. See Theorem 1 in Appendix A. For calculation, we often use an alternative form of the EGLS estimator in (2.7). The estimator

$$\hat{\boldsymbol{\beta}}_{EGLS}^* = \left(\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \mathbf{x}_{*hij} \mathbf{x}'_{*hij} \right)^{-1} \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \mathbf{x}_{*hij} y_{*hij}, \tag{2.8}$$

where $y_{*hij} = \hat{v}_{hij}^{-1/2} y_{hij}$ and $\mathbf{x}_{*hij} = \hat{v}_{hij}^{-1/2} \mathbf{x}_{hij}$, is equivalent to the estimator in (2.7) and has the appealing form of an OLS estimator. An estimator of the variance of $\hat{\beta}_{EGLS}^*$ is

$$\hat{V}(\hat{\beta}_{EGLS}^*) = \left(\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \mathbf{x}_{*hij} \mathbf{x}_{*hij}' \right)^{-1} (n-p)^{-1} \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \hat{e}_{*hij}^2, \quad (2.9)$$

where $\hat{e}_{*hij} = y_{*hij} - \mathbf{x}_{*hij} \hat{\beta}_{EGLS}^*$.

We obtain estimates of \hat{v}_{hij} by fitting a variance model. Let

$$\hat{y}_{hij} = \mathbf{x}_{hij}' \hat{\beta}_{OLS} \quad (2.10)$$

be the estimate of y_{hij} and let

$$\hat{e}_{hij} = y_{hij} - \hat{y}_{hij} \quad (2.11)$$

be the residual value of e_{hij} for individual j in PSU hi when model (2.6) is fit to the 2003-2004 NHANES sample, where $\hat{\beta}_{OLS}$ is given in (2.5). The estimates of elements in $\hat{\beta}_{OLS}$ are given in Table 2.7 with standard errors computed using the Taylor linearization variance of $\hat{\beta}_{OLS}$ given in Appendix B. The nonlinear model

$$\hat{e}_{hij}^2 = \alpha_0 + \alpha_1 (\hat{y}_{hij})^{\alpha_2} \quad (2.12)$$

is fit using OLS, where \hat{y}_{hij} is defined in (2.10) and \hat{e}_{hij} is defined in (2.11). The estimated coefficients of α_0 , α_1 , and α_2 are 14.34, 6.67, and 1.41, respectively. The nonlinear model (2.12) is refit using weighted least squares, where the weights are the inverses of the estimated values from the initial nonlinear model fit. The second fitting of model (2.12) with variance weights accounts for heterogeneity in the errors from the first fitting. The estimated coefficients of α_0 , α_1 , and α_2 are 5.18 (10.59), 12.15 (6.76), and 1.18 (0.23), respectively, for the second fitting, where the standard errors of the estimates are given in parentheses. An estimate of v_{hij} is then

$$\hat{v}_{hij} = 5.18 + 12.15 (\hat{y}_{hij})^{1.18}, \quad (2.13)$$

where \hat{y}_{hij} is defined in (2.10). Percentiles of the distribution of estimated variances are given in Table 2.8. As a check to see if there is additional variability in the estimated variances

which can be accounted for by the model variables in (2.6), we fit the model

$$\frac{(\hat{e}_{hij}^2 - \hat{v}_{hij})}{\hat{v}_{hij}} = \mathbf{x}_{hij}'\boldsymbol{\eta},$$

to the 2003-2004 NHANES sample using OLS, where \mathbf{x}_{hij} is given in (2.6). None of the estimated regression coefficients were significant in the model fit, giving evidence that the model variables do not account for any additional variability in the estimated variances. The EGLS estimator (2.8) is fit to the 2003-2004 NHANES sample using the estimated variances given by (2.13). Standard errors of the estimated regression coefficients are computed using the EGLS variance in (2.9). The estimated regression coefficients and standard errors are given in the first column of Table 2.11.

Table 2.7 Estimated regression coefficients for model (2.6) fit to the 2003-2004 NHANES sample using OLS

Variable	2003-2004 Est (SE)
<i>trans</i>	2.017 (0.228)
<i>mod</i>	1.081 (0.199)
<i>vig</i>	2.202 (0.239)
<i>age2</i>	0.100 (0.035)
<i>mexblackmod</i>	-0.747 (0.338)
<i>mexvig</i>	-1.316 (0.652)
<i>mexblackage</i>	0.092 (0.052)

Table 2.8 Percentiles for the distribution of estimated error variances for the 2003-2004 NHANES sample

Percentile	0	10	25	50	75	90	100
Estimated Error Variance	5	12	28	59	130	206	516

2.3.3 Test for Full vs. Reduced Models

We conduct a test to see if model (2.6) for the full female sample is significantly different than model (2.4) for the three female race/ethnicity groups. In the test, we consider model (2.6) as a reduced model of model (2.4) when model (2.4) is fit to three separate race/ethnicity

samples and use F-test procedures for comparing full and reduced models. The test statistic is

$$F = \frac{(SSE_{fem} - SSE_{sep})/(df_{fem} - df_{sep})}{(SSE_{sep}/df_{sep})},$$

where SSE_{fem} is the error sum of squares (SSE) when model (2.6) is fit to the full 2003-2004 sample, SSE_{sep} is the sum of the SSEs when model (2.4) is fit to each of the three race/ethnicity groups, df_{fem} is the degrees of freedom of SSE_{fem} , and df_{sep} is the degrees of freedom of SSE_{sep} . The sum of squares are computed from fitting model (2.6) to the full sample and model (2.4) to each of the three race/ethnicity groups using the EGLS estimator (2.8). The F statistic is 1.15 with (11, 1551) degrees of freedom and has a corresponding p-value of 0.32. Hence, there is little evidence to suggest that model (2.6) for the full female sample is different than model (2.4) for each of the three female race/ethnicity groups.

2.3.4 Test for Survey Weights

The EGLS estimator (2.8) used to estimate the regression coefficients may be biased if the error terms in model (2.6) are correlated with the survey weights. To test to see if the survey weights are significant in the regression estimation, we consider the test procedure from Appendix B for the EGLS estimator and variance. Let

$$\begin{aligned} y_{*hij} &= \mathbf{x}'_{*hij}\boldsymbol{\beta} + w_{hij}^*\mathbf{x}'_{*hij}\boldsymbol{\gamma} + a_{*hij} \\ &= (\mathbf{x}'_{*hij}, w_{hij}^*\mathbf{x}'_{*hij})(\boldsymbol{\beta}', \boldsymbol{\gamma}')' + a_{*hij} \\ &= \mathbf{z}'_{*hij}\boldsymbol{\lambda} + a_{*hij}, \end{aligned} \tag{2.14}$$

where \mathbf{x}_{*hij} and y_{*hij} are defined in (2.8),

$$w_{hij}^* = (w_{hij} - \bar{w})/\bar{w},$$

and a_{*hij} is a model error term. A test for the hypothesis that $\boldsymbol{\gamma} = \mathbf{0}$ is

$$F = p^{-1}\hat{\boldsymbol{\lambda}}_2'\hat{\mathbf{V}}(\hat{\boldsymbol{\lambda}})^{-1}_2\hat{\boldsymbol{\lambda}}_2, \tag{2.15}$$

where $\hat{\boldsymbol{\lambda}}_2$ is the lower p elements of

$$\hat{\boldsymbol{\lambda}} = \left(\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \mathbf{z}_{*hij} \mathbf{z}'_{*hij} \right)^{-1} \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \mathbf{z}_{*hij} y_{*hij},$$

and $\hat{V}(\hat{\lambda})_{22}$ is the lower right $p \times p$ submatrix of $\hat{V}(\hat{\lambda})$, where $\hat{V}(\hat{\lambda})$ is given by (2.9) with \mathbf{z}_{*hij} replacing \mathbf{x}_{*hij} and $\hat{a}_{*hij} = y_{*hij} - \mathbf{z}_{*hij}'\hat{\lambda}$ replacing \hat{e}_{*hij} . Under the null hypothesis that $\gamma = \mathbf{0}$ and given regularity conditions, F in (2.15) is approximately an F with p and $n - 2p$ degrees of freedom, where p is the dimension of γ and n is the number of elements in the sample.

When the extended model (2.14) is fit to the 2003-2004 NHANES sample, the F statistic from (2.15) is 1.72 with (7, 1555) degrees of freedom and a p-value of 0.10. Based on the test results there is little evidence to suggest that the EGLS estimator is biased for β in model (2.6).

2.3.5 Test for EGLS Variances vs. Stratified Cluster Variances

In our analyses we have used EGLS variances instead of stratified cluster variances, where stratified cluster variances account for the complex sample design. In Table 2.9 we give the EGLS variances and stratified cluster variances for the EGLS estimates given in the first column of Table 2.11. The estimated variances are similar for most of the model variables. We ran an analysis of variance (ANOVA) on the EGLS residuals from the 2003-2004 sample to check if the residuals were significantly different by strata and PSU (Table 2.10). The residuals used in the ANOVA are of the form

$$\hat{e}_{*hij} = y_{*hij} - \mathbf{x}_{*hij}'\hat{\beta}_{EGLS}^*,$$

where y_{*hij} , \mathbf{x}_{*hij} , and $\hat{\beta}_{EGLS}^*$ are defined in (2.8). Neither the strata effect nor the PSU effect are significant in the ANOVA based on F-tests.

2.3.6 Model Comparisons Across Samples

As part of our analyses, we want to determine if the models developed using the 2003-2004 female NHANES sample (models (2.6) and (2.12)) give similar results when they are fit to the 2005-2006 female NHANES sample. Following the procedures from Section 2.3.1 and 2.3.2, we fit model (2.6) to the 2005-2006 NHANES sample using the OLS estimator given by (2.5). Model estimates and residuals are computed using (2.10) and (2.11) based on the estimated regression coefficients from the OLS model fit. Model (2.12) is then fit first using OLS and

Table 2.9 EGLS and stratified cluster (Str. Clus.) standard errors (SE) for model (2.6) fit to the 2003-2004 NHANES sample using EGLS

Variable	Est	EGLS SE	Str. Clus. SE
<i>trans</i>	1.754	0.297	0.292
<i>mod</i>	1.130	0.172	0.189
<i>vig</i>	2.065	0.318	0.343
<i>age2</i>	0.108	0.022	0.027
<i>mexblackmod</i>	-0.882	0.286	0.411
<i>mexvig</i>	-0.925	0.728	0.436
<i>mexblackage</i>	0.093	0.036	0.037

Table 2.10 ANOVA for the EGLS standardized residuals

Source	DF	Sum of Squares	Mean Square	F
Strata	14	16.73	1.19	1.26
PSU(Strata)	15	15.25	1.02	1.07
Error	1539	1460	0.95	
Total	1568	1492		

second using weighted least squares, where the weights are the inverses of the predicted values from the first model fit (to account for heterogeneity in the error variances). The estimates of α_0 , α_1 , and α_2 are 5.79 (5.87), 8.96 (4.38), and 1.28 (0.21), respectively, for the second model fit, where the standard errors are given in the parentheses. An estimate of the error variance for element hij in the 2005-2006 NHANES sample is

$$\hat{v}_{hij} = 5.79 + 8.96(\hat{y}_{hij})^{1.28}.$$

Finally, model (2.6) is fit to the 2005-2006 female sample using the EGLS estimator (2.8). Standard errors for the estimates are computed using the EGLS variance (2.9). The estimated regression coefficients and standard errors are given in the second column of Table 2.11.

From Table 2.11, we see that the results are relatively similar for the two samples. The regression coefficients on *trans*, *mod*, *vig*, *age2*, and *mexblackmod* have similar estimates and standard errors. The regression coefficients on *mexvig* and *mexblackage* are more dissimilar across samples than the other regression coefficients, but the estimated coefficients are both negative for *mexvig* and both positive for *mexblackage*. To test if β in (2.6) is the same for

the two NHANES samples, we define the test statistic

$$F = p^{-1}(\hat{\beta}_{1,EGLS}^* - \hat{\beta}_{2,EGLS}^*)'[\hat{V}(\hat{\beta}_{1,EGLS}^*) + \hat{V}(\hat{\beta}_{2,EGLS}^*)]^{-1}(\hat{\beta}_{1,EGLS}^* - \hat{\beta}_{2,EGLS}^*), \quad (2.16)$$

where $\hat{\beta}_{1,EGLS}^*$ is the EGLS estimator for model (2.6) fit to the 2003-2004 NHANES sample, $\hat{\beta}_{2,EGLS}^*$ is the EGLS estimator for model (2.6) fit to the 2005-2006 NHANES sample, and $\hat{V}(\hat{\beta}_{1,EGLS}^*)$ and $\hat{V}(\hat{\beta}_{2,EGLS}^*)$ are the corresponding estimated variance matrices of $\hat{\beta}_{1,EGLS}^*$ and $\hat{\beta}_{2,EGLS}^*$, respectively, computed using (2.9). If we assume that the two NHANES samples are selected independently, under the null hypothesis that β is the same for both samples and given regularity conditions, F in (2.16) is approximately distributed as an F with p and $n_1 + n_2 - 2p$ degrees of freedom, where p is the dimension of β and n_1 and n_2 are the sample sizes. The F statistic computed for the NHANES samples is 1.81 with (7, 3077) degrees of freedom and a p-value of about 0.08. Hence, there is modest evidence suggesting that the β vector in (2.6) is different for the 2003-2004 and 2005-2006 samples. We may expect some difference in the estimates, because the model was developed based only the 2003-2004 data and not the 2005-2006 data. But, given the test results, the bias from variable selection appears to be modest.

Table 2.11 Estimated regression coefficients for model (2.6) fit to the 2003-2004 and 2005-2006 NHANES samples using EGLS

Variable	2003-2004 Est (SE)	2005-2006 Est (SE)
<i>trans</i>	1.754 (0.297)	1.449 (0.221)
<i>mod</i>	1.130 (0.172)	1.260 (0.182)
<i>vig</i>	2.065 (0.318)	1.415 (0.265)
<i>age2</i>	0.108 (0.022)	0.094 (0.016)
<i>mexblackmod</i>	-0.882 (0.286)	-1.080 (0.231)
<i>mexvig</i>	-0.925 (0.728)	-0.014 (0.602)
<i>mexblackage</i>	0.093 (0.036)	0.013 (0.025)

We can use the same testing procedure given above to test for the difference in $\alpha = (\alpha_0, \alpha_1, \alpha_2)'$ from the variance model (2.12) across sample years. The test statistic is

$$F = p^{-1}(\hat{\alpha}_{0304} - \hat{\alpha}_{0506})'[\hat{V}_{0304} + \hat{V}_{0506}]^{-1}(\hat{\alpha}_{0304} - \hat{\alpha}_{0506}),$$

where $\hat{\alpha}_{0304}$ and $\hat{\alpha}_{0506}$ are the estimates of α for the 2003-2004 and 2005-2006 samples, respec-

tively, $\hat{\mathbf{V}}_{0304}$ and $\hat{\mathbf{V}}_{0506}$ are the estimated covariance matrices of $\hat{\boldsymbol{\alpha}}_{0304}$ and $\hat{\boldsymbol{\alpha}}_{0506}$, respectively, and p is the dimension of $\boldsymbol{\alpha}$. The F test statistic is 0.14 on (3, 3085) degrees of freedom with a p-value of 0.94, giving little evidence to suggest that the variance model is different across samples.

2.3.7 Model for Full Sample

Given the test results from Section 2.3.6, we combine the samples and fit the models to the full data set. We use similar procedures to the procedures of Sections 2.3.1 and 2.3.2. First, we fit model (2.6) using EGLS, where the weights are the inverses of the estimated variances given by (2.13) and (2.16) for elements in the 2003-2004 sample and 2005-2006 sample, respectively. Second, we fit the variance model (2.12) using EGLS and the estimates and residuals from the initial model fit of (2.6), where the weights are the inverses of the estimated variances given by (2.13) and (2.16). Third, we refit model (2.12) using EGLS, where the weights are the inverses of the estimated values from the initial model fit. Fourth, we fit model (2.6) using EGLS with the estimated variances from the second fitting of model (2.12). The estimated regression coefficients and standard errors are given in Table 2.12. The estimates are similar to the estimates from Table 2.11, but have smaller standard errors given that the model is fit with a larger sample. Using the test described in Section 2.3.4, we test to see if the survey weights are significant in the regression estimation for the full sample. The F test statistic is 1.51 on (7, 3077) degrees of freedom with a p-value of 0.16. Hence, the evidence suggests that the survey weights have little influence on the expected value of the estimates.

To evaluate the EGLS model fit of (2.6) to the full sample, we look at group means and standard deviations of the standardized residuals. Let $\hat{y}_{*hij} = \mathbf{x}'_{*hij} \hat{\boldsymbol{\beta}}_{EGLS}^*$ and $\hat{e}_{*hij} = y_{*hij} - \hat{y}_{*hij}$ for individual hij in the full sample, where \mathbf{x}_{*hij} , y_{*hij} , and $\hat{\boldsymbol{\beta}}_{EGLS}^*$ are defined in (2.8). We sorted $\{\hat{y}_{*hij}, \hat{e}_{*hij}\}$ by \hat{y}_{*hij} , divided the data into ten groups of approximately equal size, and computed group means of \hat{y}_{*hij} and \hat{e}_{*hij} and group standard deviations of \hat{e}_{*hij} . The first group in each of the samples is restricted to be the set of $\{\hat{y}_{*hij}, \hat{e}_{*hij}\}$ values, where \hat{y}_{*hij} is equal to zero so that the group mean of \hat{y}_{*hij} is zero by default. Based on our model, individuals

Table 2.12 Estimated regression coefficients for models (2.6) and (2.12) fit to the full NHANES sample using EGLS

Model (2.6)		Model (2.12)	
Variable	Est (SE)	Coefficient	Est (SE)
<i>trans</i>	1.552 (0.191)	α_0	1.579 (3.484)
<i>mod</i>	1.184 (0.129)	α_1	12.417 (3.548)
<i>vig</i>	1.692 (0.218)	α_2	1.203 (0.135)
<i>age2</i>	0.104 (0.012)		
<i>mexblackmod</i>	-1.033 (0.170)		
<i>mexvig</i>	-0.335 (0.497)		
<i>mexblackage</i>	0.035 (0.020)		

older than 75 who report zero activity will have \hat{y}_{*hij} values of zero. Plots of the standardized group means and standard deviations are given in Figure 2.2. The plot on the left compares group means of the estimates to group means of the residuals. The plot on the right compares group means of the estimates to group standard deviations of the residuals. In the plot on the right, the standard deviation of the residuals for the first group is much lower than the standard deviations of residuals for the other nine groups. This occurs because a large majority of the residuals are zero when \hat{y}_{*hij} is zero. That is, the vast majority of individuals are estimated to have zero MVPA from the fitted model also measure zero accelerometer MVPA.

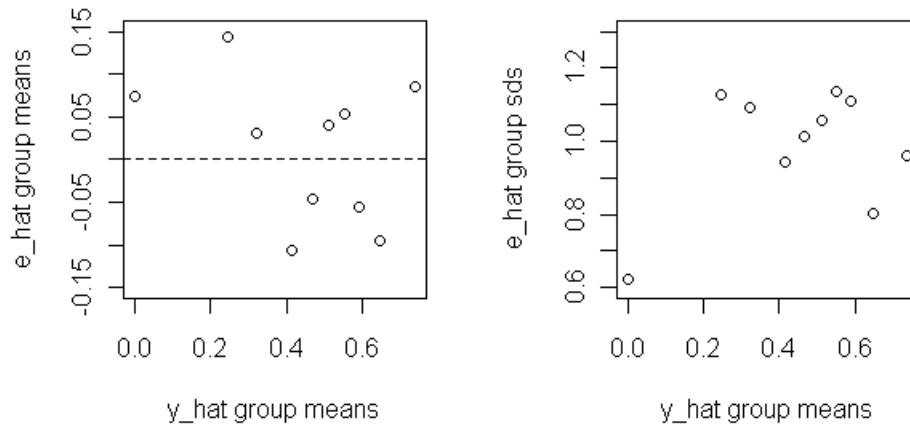


Figure 2.2 Plot of standardized group means and standard deviations

We consider an additional model adjustment to account for estimation of zero accelerometer MVPA. In the full sample there are 219 individuals with $\hat{y}_{*hij} = 0$ from the model fit. The sample mean and variance of measured accelerometer-based activity is 0.093 (min/day) and 0.612 (min/day)², respectively, for the same 219 individuals. We use these estimates as the intercepts in our mean and variance models. To implement the restrictions, we refit models (2.6) and (2.12) and fix the intercept in model (2.6) at 0.093 and the intercept in the variance model (2.12) at 0.612. The restricted models are

$$\begin{aligned} y_{hij} &= 0.093 + \beta_1 trans_{hij} + \beta_2 mod_{hij} + \beta_3 vig_{hij} + \beta_4 age2_{hij} + \beta_5 mexblackmod_{hij} \\ &\quad + \beta_6 mexvig_{hij} + \beta_7 mexblackage_{hij} + e_{hij} \\ &= 0.093 + \mathbf{x}'_{hij} \boldsymbol{\beta} + e_{hij} \end{aligned} \quad (2.17)$$

and

$$v_{hij} = 0.612 + \alpha_1 (\hat{y}_{hij})^{\alpha_2}, \quad (2.18)$$

where \hat{y}_{hij} is the estimated value from model (2.17) for y_{hij} .

To get estimates for the new means model we first regress $y_{hij} - 0.093$ on the model variables (\mathbf{x}_{hij}) in (2.17) using EGLS, where we use the estimated variances from the fitted variance model given in Table 2.12. The final set of estimates are computed using estimated variances from the restricted variance model (see below). We fit model (2.17) using only the data with positive \hat{y}_{*hij} values from the model fit of (2.6) given in Table 2.12. Let $\hat{\boldsymbol{\beta}}_{res}$ denote the vector of fitted regression coefficients from model (2.17). The estimated variance matrix of $\hat{\boldsymbol{\beta}}_{res}$ is

$$(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' [\mathbf{I} \hat{\sigma}_e^2 + \mathbf{J} \mathbf{J}' \hat{\sigma}_{e0}^2] \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}, \quad (2.19)$$

where \mathbf{X} is the matrix of \mathbf{x}_{*hij} variables from the model, \mathbf{I} is an identity matrix, $\mathbf{J} \mathbf{J}'$ is a matrix of 1's, $\hat{\sigma}_e^2$ is the estimated error variance from the model fit for the residuals with positive estimated MVPA, and $\hat{\sigma}_{e0}^2$ is the estimated variance of the mean of individuals with zero estimated MVPA.

Next, we fit model (2.18) using the squared residuals and \hat{y}_{hij} values from the model fit of (2.6), where the response variable is $\hat{e}_{hij}^2 - 0.612$. We fit this model with EGLS, using the

estimated variances from the variance model given in Table 2.12. We do the fitting using only the data with positive \hat{y}_{*hij} values from the model fit of (2.6) given in Table 2.12. The variance model is refit using the inverse of the estimated values from the first fit as the weights in the second fit to account for heterogeneity in the errors. Let $\hat{\alpha}_{res}$ denote the vector of fitted regression coefficients from model (2.18). The estimated variance matrix of $\hat{\alpha}_{res}$ is approximated using the form given by (2.19), where the rows in \mathbf{X} are the partial derivatives of $\alpha_1(\hat{y}_{hij})^{\alpha_2}$ with respect to α_1 and α_2 evaluated at $\hat{\alpha}_{res}$. The estimated regression coefficients and standard errors for models (2.17) and (2.18) are given in Table 2.13. Plots of the group means and standard deviations of the standardized residuals for the restricted model are given in Figure 2.3. The groups used to construct the plots in Figure 2.2 were also used to construct the plots in Figure 2.3.

Table 2.13 Estimates and standard errors for models (2.17) and (2.18) fit using EGLS

Model (2.17)		Model (2.18)	
Variable	Est (SE)	Coefficient	Est (SE)
Intercept (restricted)	0.093 (0.002)	Intercept (restricted)	0.612 (0.285)
<i>trans</i>	1.546 (0.221)	α_1	13.681 (3.510)
<i>mod</i>	1.181 (0.171)	α_2	1.167 (0.090)
<i>vig</i>	1.791 (0.245)		
<i>age2</i>	0.095 (0.017)		
<i>mexblackmod</i>	-0.984 (0.194)		
<i>mexvig</i>	-0.587 (0.500)		
<i>mexblackage</i>	0.050 (0.021)		

2.4 Estimation and Prediction of Daily MVPA

The fitted models (2.17) and (2.18) with estimated regression coefficients given in Table 2.13 can be considered for estimating overall group means of MVPA in groups of the female population and for predicting average daily MVPA for individuals in the population. First we consider estimation of a group mean of MVPA. Let

$$\mathbf{c}_g = (trans_g, mod_g, vig_g, age2_g, mexblackmod_g, mexvig_g, mexblackage_g)'$$

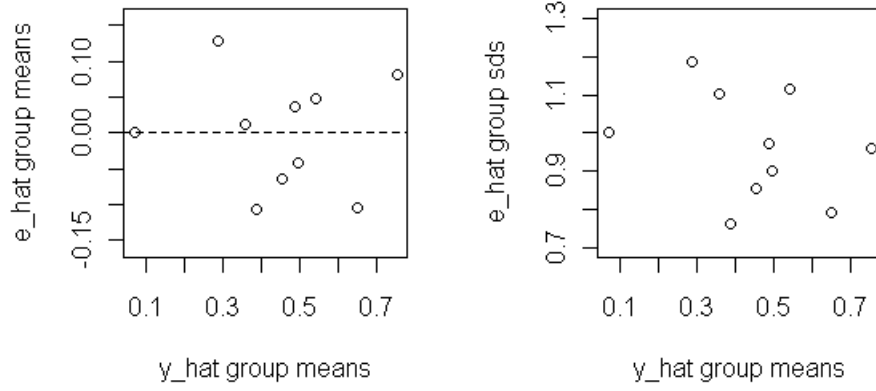


Figure 2.3 Plot of standardized group means and standard deviations for the restricted model

be a vector of model variables that define a group g in the female population. For example, the group may be defined by Mexican females younger than 60 who report zero activity. The group mean of MVPA is $(1, \mathbf{c}'_g)(0.93, \boldsymbol{\beta}')'$, where $\boldsymbol{\beta}$ is given in model (2.17). An estimate of the group mean is

$$\hat{y}_g = (1, \mathbf{c}'_g)(0.093, \hat{\boldsymbol{\beta}}'_{res})',$$

where $\hat{\boldsymbol{\beta}}_{res}$ is the estimated vector of regression coefficients given in Table 2.13. The standard error of the estimate is

$$SE(\hat{y}_g) = \sqrt{0.612 + \mathbf{c}'_g \hat{V}\{\hat{\boldsymbol{\beta}}_{res}\} \mathbf{c}_g},$$

where $\hat{V}\{\hat{\boldsymbol{\beta}}_{res}\}$ is the EGLS variance of $\hat{\boldsymbol{\beta}}_{res}$. For illustration, consider the group of females in the population who report 7 minutes of average daily transportation activity, 18 minutes of average daily moderate leisure activity, and 8 minutes of average daily vigorous leisure activity, for a total of 33 minutes of average daily MVPA. Using our model, we can estimate average daily MVPA for subgroups of this female group based on the reports, race/ethnicity, and age (Table 2.14). For example, the estimated mean of average daily MVPA is 9.3 minutes for the group of black females younger than 60 who report 33 minutes of activity. The estimated

means in the table are much smaller than the reported estimate of 33 minutes of MVPA that is based on the questionnaire. The means are different based on age and race/ethnicity.

Table 2.14 Example estimated average daily accelerometer MVPA in groups of the female population for 33 minutes of reported MVPA

	Race/Ethnicity		
Age	Black	Mexican	Other
< 60	9.3 (1.2)	8.1 (1.4)	11.2 (1.3)
70	7.9 (1.1)	6.7 (1.4)	10.2 (1.2)
> 75	7.1 (1.1)	6.0 (1.3)	9.7 (1.2)

Next, we consider prediction of average daily MVPA for an individual in the female population. Let \mathbf{x}_k denote the vector of model variables for individual k in the female population. The predicted average daily MVPA for the individual is

$$\hat{y}_k = (1, \mathbf{x}'_k)(0.093, \hat{\boldsymbol{\beta}}'_{res})'$$

with a standard error of

$$SE(\hat{y}_k) = \sqrt{\hat{v}_k + \mathbf{x}'_k \hat{V}\{\hat{\boldsymbol{\beta}}_{res}\} \mathbf{x}_k},$$

where

$$\hat{v}_k = 0.612 + 13.681\hat{y}_k^{1.167}$$

is the estimated error variance of individual k given by the fitted variance model. Suppose that we want to predict average daily MVPA for a hypothetical individual who reports 7 minutes of average daily transportation activity, 18 minutes of average daily moderate leisure activity, and 8 minutes of average daily vigorous leisure activity, for a total of 33 minutes of average daily MVPA. For illustration, we consider different ages and race/ethnicity groups for the individual. The predicted values and standard errors for these groups are given in Table 2.15. The predicted values are the same as the estimated values in Table 2.14 because we are using the same values for the report variables *trans*, *mod*, and *vig* in the computations. The standard errors in Table 2.15, however, are larger than the standard errors in Table 2.14

because in Table 2.15 we give predictions of average daily MVPA for an individual, which are less precise than the estimated means of average daily MVPA for a group in the population. The standard errors are large because of large \hat{v}_k terms that are estimated from the variance model.

In Table 2.15 we also give 95% prediction intervals for the predicted values. In preliminary analyses, a cube root transformation was shown to give approximately normal data for nonzero accelerometer MVPA. Thus, we construct prediction intervals in the cube root scale and transform the interval limits to the original scale. For a predicted value \hat{y}_k in the original scale, the 95% prediction interval in the cube root scale is $\hat{y}_k^{1/3} \pm 1.96\sqrt{\hat{V}(\hat{y}_k^{1/3})}$, where

$$\hat{V}(\hat{y}_k^{1/3}) \approx \left[\frac{1}{3} \hat{y}_k^{-2/3} \right] \hat{V}(\hat{y}_k)$$

by the delta method and $\hat{V}(\hat{y}_k)$ is the variance of \hat{y}_k . The 95% prediction interval in the original scale is obtained by taking the cube of the lower and upper bounds of the cube root scale interval. The lower bounds of all the prediction intervals were close to zero. Lower bounds that were less than zero were set to zero in Table 2.15.

Table 2.15 Example predicted average daily accelerometer MVPA for an individual in the female population for 33 minutes of reported MVPA (standard errors are in parentheses and 95% prediction intervals are in brackets)

	Race/Ethnicity		
Age	Black	Mexican	Other
< 60	9.3 (13.6) [0.0, 70.5]	8.1 (12.7) [0.0, 67.4]	11.2 (15.2) [0.0, 75.7]
70	7.9 (12.4) [0.0, 66.2]	6.7 (11.3) [0.0, 62.9]	10.2 (14.4) [0.0, 73.0]
> 75	7.1 (11.7) [0.0, 64.0]	6.0 (10.6) [0.0, 60.6]	9.7 (14.0) [0.0, 71.6]

2.5 Discussion

The results given in Table 2.14 suggest that the female physical activity model developed in Section 2.3 is reasonable for estimating means of average daily MVPA for groups in the female

population. We recommend that researchers use the model when estimating means of average daily MVPA in the population instead of using the questionnaire data because the activity reports from the questionnaire data overestimate means in the population. This is illustrated in our example, where the mean of average daily MVPA based on the reports was given as 33 minutes and the estimated means from the model were between 6 and 11 minutes depending on the age and race/ethnicity group being considered. The results given in Table 2.15 suggest that the female physical activity model has large relative variance for predicting the average daily MVPA for individuals in the female population. In our example, the standard errors are larger than the predicted average daily MVPA values for each age and race/ethnicity group considered and the 95% prediction intervals cover a large range of values.

Based on our analyses, the self-report variables from the NHANES questionnaire are not very good indicators of individual average daily MVPA. In general, the questionnaire estimates of average daily MVPA were larger and more variable than the accelerometer estimates of average daily MVPA. This is illustrated in Figure 2.1, which shows some extreme questionnaire-based estimates of MVPA. The results suggest that a redesign of the questionnaire be considered. For example, the questionnaire could be redesigned by asking survey participants to report on their activity from the previous day instead of the previous 30 days. Research has shown that individuals have more difficulty reporting on activity over a long period of time, such as 30 days, than they do for shorter periods of time, such as a day or week (Matthew 2002).

Given the current NHANES design, each survey participant in a subsample provides one set of physical activity measurements via the questionnaire and accelerometer. If multiple accelerometer measurements were available for some of the survey participants, one could use measurement error models to model the between-person variation and within-person variation in the physical activity data and develop methods for estimating physical activity parameters in the population based on the estimated between-person variation in physical activity. This line of research is developed in Chapter 3.

CHAPTER 3 A METHOD FOR ESTIMATING USUAL DAILY ENERGY EXPENDITURE PARAMETERS

3.1 Introduction

Assessment of usual or habitual physical activity is important for studying relationships between physical activity and health and for determining appropriate physical activity guidelines to maintain good health (Shephard 2003). One component of this assessment involves estimation of usual daily energy expenditure (EE) parameters. EE is a measure of the energy cost associated with physical activity (Schutz et al. 2001). An individual's usual daily EE is his or her average daily EE over a long period of time, such as one year. From a statistical perspective, usual daily EE of individual i is

$$T_i = E\{T_{ij}|i\},$$

where T_{ij} is the actual daily EE of individual i on day j .

The instruments most commonly used to measure daily EE from individuals in the population are self-report instruments (Ainsworth 2009; Matthews 2002) and monitoring devices (Welk 2002; Moy et al. Submitted), both of which provide imperfect measurements of usual daily EE. An observed measurement of daily EE for individual i on day j , defined as Y_{ij} , will differ from the usual daily EE for individual i , T_i , because of nuisance effects (Matthews et al. 2001; Matthews et al. 2002) and measurements errors (Ainsworth 2009; Welk 2002). Nuisance effects, such as seasonality and day-of-week effect, exist because individuals vary their physical activity habits on a daily basis. Measurement errors from monitoring devices are due to the inability of monitors to accurately capture the full range of activities (Welk et al. 2004) and the imperfect conversion process of monitor data into EE estimates (Welk 2002). Measurement

errors from self-report instruments are due to such factors as social desirability effects (Adams et al. 2005), difficulty in understanding concepts of survey questions (Sallis and Saelens 2000), and cognitive limitations for recalling activity from the past (Matthews 2002). The difference between actual daily EE and usual daily EE may be defined as

$$D_{ij} = T_{ij} - T_i$$

for individual i on day j , and can be attributed to nuisance factors. For example, if individual i was more active than he or she usually is on day j , then $D_{ij} > 0$. The difference between measured and actual daily EE may be defined as

$$E_{ij} = Y_{ij} - T_{ij},$$

and can be attributed to measurement errors. For example, if individual i reports more activity than he or she actually did on day j using a self-report instrument, then $E_{ij} > 0$. The total difference between observed EE (Y_{ij}) and usual daily EE (T_i) is then

$$\begin{aligned} Y_{ij} - T_i &= T_{ij} - T_i + Y_{ij} - T_{ij} \\ &= D_{ij} + E_{ij}, \end{aligned}$$

for individual i on day j , which is the sum of the nuisance effect (D_{ij}) and the measurement error effect (E_{ij}).

Failure to account for the measurement error and nuisance effects in daily EE measurements may lead to biased estimates of usual daily EE parameters. Troiano et al. (2008) demonstrate the potential for bias in self-reported physical activity measurements using physical activity data from the 2003-2004 NHANES sample. The percent of individuals in the U.S. population who adhere to physical activity guidelines set by the U.S. Department of Health and Human Services was estimated separately using accelerometer measurements and questionnaire-based measurements of physical activity from the NHANES sample. Less than 10% of individuals age 12 and older were estimated to adhere to the physical activity guidelines based on the accelerometer measurements, while over 50% of individuals were estimated to adhere to the same guidelines according to the questionnaire measurements of physical activity (Troiano et

al. 2008). These results suggest that individuals may overreport on their activity, which can lead to over-estimation of physical activity levels in the population, and that the accelerometers may underreport on individuals' activity, since accelerometers do not capture the full range of activity. Ferrari et al. (2007) show evidence of bias in measurements of EE from a physical activity questionnaire. The authors fit a measurement error model to data from a sample of 154 adults in a study conducted at the Alberta Cancer Board (Friedenreich et al. 2006), where each adult provided four weekly measurements of EE from an accelerometer, four weekly measurements of EE from a physical activity log, and one measurement of EE from a questionnaire. All EE measurements were in MET-hours/week. The estimated attenuation factor for the questionnaire, which assesses the ability to measure usual EE from the questionnaire measurements, was 0.13 with a 95% confidence interval of (0.05, 0.23). Given that the estimate is close to 0, the authors conclude that there is evidence of bias in the physical activity questionnaire measurements.

The potential for bias is also a concern in dietary intake studies because, as in measurements of EE or physical activity, measurements of nutrient and food intakes are prone to measurement error and nuisance effects. Nusser et al. (1996) show that using the unadjusted individual means of daily intake measurements from a 24-hour recall to estimate a distribution function of usual daily intake can lead to biased inferences regarding dietary status, and suggest an alternative approach for estimating usual intake distributions which accounts for the measurement error and nuisances effects in daily intake measurements using statistical models. Kipnis et al. (2003) provide evidence of bias in both food frequency questionnaires and 24-hour food recalls for measuring usual intake of energy and protein, and suggest the use of reference instruments such as doubly labeled water or urinary nitrogen for calibrating these self-report instruments.

In this paper, we develop a method for estimating usual daily EE parameters that accounts for the measurement error and nuisance effects in observed EE data. In our method, parameters of usual daily EE are estimated from a sample of individuals in the population, where each individual provides replicate concurrent measurements of daily EE using a reference instru-

ment, such as a multi-sensor monitor, and a self-report instrument, such as a 24-hour recall. Like some of the other methods in the physical activity and dietary intake literature (Ferrari et al. 2007; Nusser et al. 1996; Kipnis et al. 2003), our method adjusts for the measurement error and nuisance effects associated with observed values of EE using measurement error models. Like the models presented in Ferrari et al. (2007) and Kipnis et al. (2003), our models also account for systematic reporting biases from a self-report instrument. Unlike these other methods, our method includes a procedure for estimating usual daily EE parameters simultaneously for distinct groups in the population, which may be defined by gender, age, and race/ethnicity. This extension allows researchers to compare EE across groups that are of interest in physical activity assessment.

Our method consists of several steps, which we briefly outline in this section. The steps are used to estimate and remove measurement error and bias in the EE data before estimating usual daily EE parameters. In the first step of our method, we transform the EE data to approximate normality and test for the presence of a variety of nuisance factors. In our analyses, a log transformation gives approximately normal data, but in other cases, a power transformation or a more complex semiparametric transformation may be necessary to approximate normality. The transformation is important because the normality assumption is required to model the distribution of usual daily EE.

We test for nuisance effects in the transformed data by fitting separate linear regression models to the EE measurements from the reference instrument and self-report instrument, which include nuisance effects parameters. Common nuisance effects to consider are day-of-week effect (e.g., weekday vs. weekend), time-in-sample effect (e.g., first vs. second replicate), and seasonality (e.g., summer vs. winter). In our analyses, we consider only variables for day-of-week effect and time-in-sample effect in our models and not variables for seasonality because individuals in the preliminary sample we use are measured for EE in the same season. If a nuisance effect is significant in the fitted linear regression models, the estimated effect is removed from the EE data and the remainder of the analyses are conducted with the adjusted EE data. If a nuisance effect is non-significant in the fitted models, the EE data are not

adjusted for that effect. The procedure for transforming the data to normality and testing for nuisance effects is described in Section 3.2.1.

In the next step, models are fit to the adjusted normal-scale EE data to account for sources of variation and bias in the data and to estimate parameters of the usual daily EE distribution. Assessment of usual daily EE in subpopulations (hereafter referred to as groups) is often of interest to public health researchers. In our method, groups can be defined by gender, age, race/ethnicity, or other factors with the goal of comparing model parameters for EE behaviors across these groups. After the groups are identified, a group-level measurement error model is fit to each group using method of moments. The same measurement error model is fit to each group so that parameter estimates can be compared across groups. A population-level model is then developed based on the group-level estimates so that the total number of model parameters may be reduced. If there is evidence that a group-level model parameter is similar across groups, the parameter may be pooled across the groups. If there is evidence of a systematic trend in a group-level parameter across groups, the trend can be accounted for with fewer parameters in the population-level model. Once the population-level model is specified, the model is fit to group-level moment estimators using estimated generalized least squares and estimated daily EE parameters are obtained. The group-level and population-level models are developed in Sections 3.2.2 and 3.2.3, respectively.

As a final step of our method, we give a procedure for estimating a distribution of usual daily EE for each group in the original scale. For the procedure, daily EE values are generated from an estimated normal-scale distribution of mean daily EE for the group, and the generated values are transformed to the original scale to create an estimated distribution of usual daily EE in the original scale. In our presentation we give the procedure for a log transformation, but other procedures will be needed if a power or semiparametric transformation was used in the transformation to normality. The procedure is given in Section 3.2.4.

Our method is developed to account for complex sample designs by incorporating weights into the analyses. Each individual i in the sample is assigned a weight of w_i which reflects the individual's probability of selection based on the sample design and the model parameters

are estimated using weighted-estimation approaches. When it is of interest to the researcher to conduct an unweighted (or equal-weight) analyses, where the weights are set to 1 for all individuals in the sample, the researcher should first compare the unweighted and survey-weighted analyses to see if the results are different (i.e., to see if the sample design is informative to the analyses). In Appendix C we give a test for comparing the unweighted and survey-weighted estimators for parameters of the population-level model given in Section 3.2.3. In our analyses of the preliminary PAMS data, the test is non-significant at the 0.05 level, giving evidence that the unweighted and survey-weighted analyses provide similar results. Thus, we give results for an unweighted analysis of the PAMS data in Section 3. Results from the fitted population-level model based on the survey-weighted analyses are given in Appendix C.

This chapter is outlined as follows. First we develop our method in Section 3.2. Then, we illustrate our method by estimating usual daily EE parameters from a preliminary sample of females in the Physical Activity Measurement Survey (PAMS) in Section 3.3. We give a discussion of the results in Section 3.4.

3.2 Methodology

In this section, we develop a method for estimating usual daily EE parameters from a sample of n individuals who provide daily EE measurements. We assume that each individual i has a survey weight w_i , which reflects the individual's probability of selection, and that each individual is measured for daily EE on two days. On each measurement day, the individuals are measured for EE using an unbiased reference instrument, such as a multi-sensor monitoring device, and a self-report instrument, such as a 24-hour recall. Let X_{ij} be the measurement of EE for individual i on day j from the reference instrument and let Y_{ij} be the measurement of EE for individual i on day j from the self-report instrument, where $j = 1, 2$. The complete set of measurements for individual i is $(X_{i1}, X_{i2}, Y_{i1}, Y_{i2})$. Define T_i to be the true usual daily EE for individual i in the original scale. Assume that T_i , X_{ij} , and Y_{ij} are all given in the same units (e.g., kilocalories per day or MET-minutes per day).

We divide the population into groups, which are chosen so that usual daily EE parameters

can be compared across the groups. The groups may be defined by any variable, but EE levels have been shown to differ by gender, age, and race/ethnicity (Matthews 2002; Ferrari et al. 2007; Marshall et al. 2007; Ainsworth 2009).

3.2.1 Transformation to Normality and Test for Nuisance Effects

In our method, we assume normality when fitting measurement error models (Section 3.2.2). Thus, the first step in our method is to transform the original-scale daily EE data (X_{ij} and Y_{ij}) to approximate normality. Let

$$x_{ij}^* = h(X_{ij})$$

and

$$y_{ij}^* = h(Y_{ij}),$$

where $h(\cdot)$ is a continuous function and the set of x_{ij}^* values and the set of y_{ij}^* values are both approximately normal. The same transformation, $h(\cdot)$, for both the reference EE data and self-report EE data is assumed. In practice, the choice of $h(\cdot)$ will depend on the EE data from the sample. For EE data, a log transformation may be sufficient. Using EE data in the log scale for analyses is appealing from a subject matter perspective because log-scale data are often considered in physical activity research to approximate normality. Ferrari et al. (2007) consider weekly measurements of EE in the log scale to approximate normality. If normality cannot be achieved using log-scale EE data, a power transformation or a more complex transformation such as the semiparametric transformation proposed by Nusser et al. (1996) may be required.

A Shapiro-Wilk test (Shapiro and Wilk 1965) can be used to test the normality of the transformed x_{ij}^* values and y_{ij}^* values in unweighted data. In our procedure, we consider the transformed data to be approximately normal if the p-values of the two test statistics are greater than 0.10. If at least one of the p-values is less than or equal to 0.10, other transformations will be considered. When considering survey-weighted data, an approximate equal-weight sample can be created from the survey-weighted sample before testing for normality. A procedure

for creating an equal-weight sample from a survey-weighted sample is given in Section 2.4 of Nusser et al. (1996).

Daily EE data include nuisance effects that are not of interest in estimating usual daily EE parameters. Some nuisance effects are day-of-week effect, time-in-sample effect, and seasonality (Matthews et al. 2001; Matthews et al. 2002). To test for nuisance effects in the EE data, we define linear regression models for the reference monitor and self-report EE data. The regression models include variables for the nuisance effects and variables for other demographic factors that are potentially related to EE, such as age, gender, and race/ethnicity, so that the estimated nuisance effects do not include the effects from these other factors. Let x_{ij}^* and y_{ij}^* be the daily EE values from the reference instrument and self-report instrument in the normal scale, respectively, for individual i on day j and let

$$\begin{aligned} x_{ij}^* &= \mathbf{z}_{1,ij}' \boldsymbol{\gamma}_{x,1} + \mathbf{z}_{2,ij}' \boldsymbol{\gamma}_{x,2} + \epsilon_{x,ij} \\ &= (\mathbf{z}_{1,ij}', \mathbf{z}_{2,ij}') (\boldsymbol{\gamma}_{x,1}', \boldsymbol{\gamma}_{x,2}')' + \epsilon_{x,ij} \\ &= \mathbf{z}_{ij}' \boldsymbol{\gamma}_x + \epsilon_{x,ij} \end{aligned} \tag{3.1}$$

and

$$\begin{aligned} y_{ij}^* &= \mathbf{z}_{1,ij}' \boldsymbol{\gamma}_{y,1} + \mathbf{z}_{2,ij}' \boldsymbol{\gamma}_{y,2} + \epsilon_{y,ij} \\ &= (\mathbf{z}_{1,ij}', \mathbf{z}_{2,ij}') (\boldsymbol{\gamma}_{y,1}', \boldsymbol{\gamma}_{y,2}')' + \epsilon_{y,ij} \\ &= \mathbf{z}_{ij}' \boldsymbol{\gamma}_y + \epsilon_{y,ij} \end{aligned} \tag{3.2}$$

define the regression models for the nuisance effects, where $\mathbf{z}_{1,ij}$ is a vector of nuisance variables, $\mathbf{z}_{2,ij}$ is a vector of other variables of interest, $\boldsymbol{\gamma}_x$ is the vector of model parameters for model (3.1), $\boldsymbol{\gamma}_y$ is the vector of model parameters for model (3.2), and $\epsilon_{x,ij} \sim (0, \sigma_{x,i}^2)$ and $\epsilon_{y,ij} \sim (0, \sigma_{y,i}^2)$ are model error terms for individual i on day j . We assume that $\epsilon_{x,ij}$ and $\epsilon_{x,i'j'}$ are independent for $i \neq i'$ and that $\epsilon_{y,ij}$ and $\epsilon_{y,i'j'}$ are independent for $i \neq i'$. The models are fit by weighted least squares, where the weights are the survey weights (w_i). The weighted estimators for $\boldsymbol{\gamma}_x$ and $\boldsymbol{\gamma}_y$ are

$$\hat{\boldsymbol{\gamma}}_x = \left(\sum_{i=1}^n \sum_{j=1}^2 \mathbf{z}_{ij} w_i \mathbf{z}_{ij}' \right)^{-1} \sum_{i=1}^n \sum_{j=1}^2 \mathbf{z}_{ij} w_i x_{ij}^* \tag{3.3}$$

and

$$\hat{\gamma}_y = \left(\sum_{i=1}^n \sum_{j=1}^2 z_{ij} w_i z'_{ij} \right)^{-1} \sum_{i=1}^n \sum_{j=1}^2 z_{ij} w_i y_{ij}^*, \quad (3.4)$$

respectively. If the sample is selected using a complex design, the design should be accounted for when estimating the variances of $\hat{\gamma}_x$ and $\hat{\gamma}_y$. For a stratified design, where individuals are selected from each of H strata and there are n_h individuals selected from stratum h , $h = 1, \dots, H$, an estimated Taylor linearization variance can be constructed for the estimated vector of regression coefficients $\hat{\gamma}_x$. Given the design with replicate measurements from each individual, the individuals are treated as clusters and the replicate measurements are treated as elements within clusters. The estimated variance of $\hat{\gamma}_x$ is

$$\hat{V}(\hat{\gamma}_x) = \left(\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^2 z_{hij} w_{hi} z'_{hij} \right)^{-1} \hat{G}_{x,WLS} \left(\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^2 z_{hij} w_{hi} z'_{hij} \right)^{-1}, \quad (3.5)$$

where

$$\begin{aligned} \hat{G}_{x,WLS} &= \frac{n-1}{n-p} \sum_{h=1}^H \frac{n_h}{n_h-1} \sum_{i=1}^{n_h} (\mathbf{s}_{hi.} - \bar{\mathbf{s}}_{h..})(\mathbf{s}_{hi.} - \bar{\mathbf{s}}_{h..})', \\ \mathbf{s}_{hij} &= \mathbf{z}_{hij} w_{hi} \hat{\epsilon}_{x,hij}, \\ \hat{\epsilon}_{x,hij} &= x_{hij}^* - \mathbf{z}'_{hij} \hat{\gamma}_x, \\ \mathbf{s}_{hi.} &= \sum_{j=1}^2 \mathbf{s}_{hij}, \\ \bar{\mathbf{s}}_{h..} &= n_h^{-1} \sum_{i=1}^{n_h} \mathbf{s}_{hi.}, \end{aligned}$$

p is the dimension of γ_x , and x_{hij} , w_{hi} , and \mathbf{z}_{hij} are the values of x_{ij} , w_i , and \mathbf{z}_{ij} for individual i in stratum h , respectively. The estimated variance of $\hat{\gamma}_y$ may be defined in a similar manner as

$$\hat{V}(\hat{\gamma}_y) = \left(\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^2 z_{hij} w_{hi} z'_{hij} \right)^{-1} \hat{G}_{y,WLS} \left(\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^2 z_{hij} w_{hi} z'_{hij} \right)^{-1}, \quad (3.6)$$

where

$$\begin{aligned}
\hat{G}_{y,WLS} &= \frac{n-1}{n-p} \sum_{h=1}^H \frac{n_h}{n_h-1} \sum_{i=1}^{n_h} (\mathbf{t}_{hi.} - \bar{\mathbf{t}}_{h..})(\mathbf{t}_{hi.} - \bar{\mathbf{t}}_{h..})', \\
\mathbf{t}_{hij} &= \mathbf{z}_{hij} w_{hi} \hat{\epsilon}_{y,hij}, \\
\hat{\epsilon}_{y,hij} &= y_{hij}^* - \mathbf{z}_{hij}' \hat{\gamma}_y, \\
\mathbf{t}_{hi.} &= \sum_{j=1}^2 \mathbf{t}_{hij}, \\
\bar{\mathbf{t}}_{h..} &= n_h^{-1} \sum_{i=1}^{n_h} \mathbf{t}_{hi.},
\end{aligned}$$

and y_{hij} is the value of y_{ij} for individual i in stratum h .

We consider t-tests to test for the significance of the individual nuisance effects in the fitted models. Without loss of generality, we consider the procedure for removing a single nuisance effect from the EE data. Let $z_{k,ij}$ be the model variable for nuisance effect k , $\gamma_{x,k}$ be the regression coefficient on $z_{k,ij}$ in model (3.1), $\hat{\gamma}_{x,k}$ be the weighted least squares estimator of $\gamma_{x,k}$, and let $se(\hat{\gamma}_{x,k})$ be the standard error of $\hat{\gamma}_{x,k}$ computed using the Taylor variance (3.5). A test statistic for $H_0 : \gamma_{x,k} = 0$ is

$$t = \hat{\gamma}_{x,k} / se(\hat{\gamma}_{x,k}). \quad (3.7)$$

We treat the nuisance effect as significant at the 0.05 level if the absolute value of t is greater than the upper .025 quantile of a t distribution with $n - p$ degrees of freedom. A significant nuisance effect is removed by computing an adjusted value of x_{ij}^* as

$$x_{ij}^{**} = x_{ij}^* - (z_{k,ij} - \bar{z}_{k,..}) \hat{\gamma}_{x,k},$$

where $\bar{z}_{k,..}$ is the weighted mean of the $z_{k,ij}$ in the sample. In general, if a nuisance effect is significant at the 0.05 level in model (3.1) and/or in model (3.2), it is removed from the x_{ij}^* data and the y_{ij}^* data. Once any significant nuisance effects are removed from the data, the researcher should check to make sure the adjusted data are still approximately normal. If the normality assumption no longer holds, alternative transformations should be considered. In what follows, let x_{ij} and y_{ij} be the EE values for X_{ij} and Y_{ij} in the normal scale, respectively, after being adjusted for significant nuisance effects.

3.2.2 Group-Level Measurement Error Model

The next step in our method is parameter estimation for a group-level measurement error model. The group-level model is used to estimate daily EE parameters for each group. Groups may be defined by gender, age, race/ethnicity or any other factors of interest to the researcher. In this section, we present a group-level measurement error model and develop estimators for the model.

Assume that G groups are considered for the analyses, and let g denote the g th group. Further assume that the EE measurements from group g and group g' are uncorrelated for $g \neq g'$. Let μ_g be the mean of daily EE in the normal scale for group g and let $\mu_g + t_{gi}$ be the mean daily EE for individual i in the normal scale, where $t_{gi} \sim N(0, \sigma_{tg}^2)$. The distribution of mean daily EE in the normal scale is then given by $N(\mu_g, \sigma_{tg}^2)$ for group g . On any given day j , individual i in group g will have an actual daily EE value of t_{gij} in the normal scale. We assume that the daily deviations from the individual's mean daily EE are additive. Thus, our model for t_{gij} is

$$t_{gij} = \mu_g + t_{gi} + d_{gij},$$

where $d_{gij} \sim N(0, \sigma_{dg}^2)$ is individual i 's deviation from his or her mean daily EE on day j in the normal scale. On days where individual i is more active than usual, d_{gij} will be positive, and on days where individual i is less active than usual, d_{gij} will be negative. We assume that t_{gi} and d_{gij} are uncorrelated for all g , i , and j . That is, we assume that an individual's mean activity is unrelated to his or her within-individual variation in activity on a day-to-day basis. Given this assumption, the variance of t_{gij} ,

$$\begin{aligned} V\{t_{gij}\} &= V\{\mu_g + t_{gi} + d_{gij}\} \\ &= \sigma_{tg}^2 + \sigma_{dg}^2, \end{aligned}$$

is the sum of the mean daily EE variance (σ_{tg}^2) and the within-individual variance (σ_{dg}^2).

Let x_{gij} be a measure of daily EE in the normal scale for individual i on day j in group g from an unbiased reference instrument, such as a multi-sensor monitoring device. We assume

that the reference instrument gives an unbiased measurement of daily EE in the normal scale,

$$x_{gij} = \mu_g + t_{gi} + d_{gij} + u_{gij}, \quad (3.8)$$

where $u_{gij} \sim N(0, \sigma_{ug}^2)$ is random measurement error for individual i on day j in group g . We assume that u_{gij} is uncorrelated with t_{gi} and d_{gij} for all g , i , and j , and hence, the variance of x_{gij} is

$$\begin{aligned} V\{x_{gij}\} &= V\{\mu_g + t_{gi} + d_{gij} + u_{gij}\} \\ &= \sigma_{tg}^2 + \sigma_{dg}^2 + \sigma_{ug}^2. \end{aligned}$$

Let y_{gij} be a measurement of daily EE in the normal scale for individual i on day j in group g from a self-report instrument such as a 24-hour recall. We assume that the self-report measure y_{gij} is potentially biased for actual daily EE in the normal scale and represent y_{gij} as

$$y_{gij} = \mu_{yg} + \beta_{1g}(t_{gi} + d_{gij}) + r_{gi} + e_{gij}, \quad (3.9)$$

where μ_{yg} is the group mean of daily EE in the normal scale from the self-report instrument, β_{1g} is the slope that accounts for the systematic error in the relationship between self-report and actual daily EE in group g , $r_{gi} \sim N(0, \sigma_{rg}^2)$ is a term that represents individual i 's deviation from the group-level mean, and $e_{gij} \sim N(0, \sigma_{eg}^2)$ is the remaining measurement error in the self-report for individual i on day j in group g . We assume that the model terms r_{gi} and e_{gij} are uncorrelated with each other, with t_{gi} and d_{gij} , and with u_{gij} from model (3.8) for all g , i , and j . Like model (3.8), model (3.9) assumes an additive linear relationship between measured EE and mean daily EE in the normal scale. Unlike model (3.8), model (3.9) includes a different overall mean, μ_{yg} , and a slope term, β_{1g} , to account for systematic error that may arise from self-reporting EE.

To identify the parameters of the measurement error model given by equations (3.8) and (3.9), we assume that the reference measure gives an unbiased measurement of mean daily EE. This assumption may not be reasonable if the measurement is from a monitor that is known to have bias. For example, it is recognized that accelerometers are unable to capture some types of physical activity (Welk et al. 2004) and may give biased measurements of daily EE. The

assumption of an unbiased reference measure is more reasonable if measurements come from a multi-sensor device such as the SenseWear armband monitor. SenseWear monitors have been shown to provide accurate measurements of daily EE in free-living conditions when compared to doubly labeled water, which is considered a gold standard for measuring EE (Moy et al. Submitted; Calabro et al. 2009).

We use method of moments to derive estimators of the parameters for the group-level measurement error model. The estimators are given as weighted estimators, where w_{gi} is the weight for individual i in group g . Let the 8-dimensional parameter vector for group g be defined by

$$\boldsymbol{\theta}_g = (\mu_g, \mu_{yg}, \beta_{1g}, \sigma_{tg}^2, \sigma_{dg}^2, \sigma_{ug}^2, \sigma_{eg}^2, \sigma_{rg}^2)'. \quad (3.10)$$

To compute estimators for $\boldsymbol{\theta}_g$, we consider summary statistics based on

$$\mathbf{Z}_{gi} = \begin{pmatrix} \bar{x}_{gi.} \\ \bar{y}_{gi.} \\ x_{gi1} - x_{gi2} \\ y_{gi1} - y_{gi2} \end{pmatrix}, \quad (3.11)$$

where

$$\bar{x}_{gi.} = \frac{x_{gi1} + x_{gi2}}{2}$$

and

$$\bar{y}_{gi.} = \frac{y_{gi1} + y_{gi2}}{2}.$$

We define \mathbf{Z}_{gi} in this manner because \mathbf{Z}_{gi} provides an algebraically simpler covariance matrix than the observed data vector $(x_{gi1}, x_{gi2}, y_{gi1}, y_{gi2})'$. Given the model assumptions, the expected value of \mathbf{Z}_{gi} is

$$E\{\mathbf{Z}_{gi}\} = \begin{pmatrix} \mu_g \\ \mu_{yg} \\ 0 \\ 0 \end{pmatrix} \quad (3.12)$$

and the variance of \mathbf{Z}_{gi} is

$$V\{\mathbf{Z}_{gi}\} = \begin{pmatrix} \sigma_{gt}^2 + \frac{1}{2}\sigma_{gd}^2 + \frac{1}{2}\sigma_{gu}^2 & \beta_{1g}\sigma_{tg}^2 + \frac{1}{2}\beta_{1g}\sigma_{dg}^2 & 0 & 0 \\ & \beta_{1g}^2(\sigma_{tg}^2 + \frac{1}{2}\sigma_{dg}^2) + \sigma_{rg}^2 + \frac{1}{2}\sigma_{eg}^2 & 0 & 0 \\ & & 2(\sigma_{dg}^2 + \sigma_{ug}^2) & 2\beta_{1g}\sigma_{dg}^2 \\ \text{symmetric} & & & 2(\beta_{1g}^2\sigma_{dg}^2 + \sigma_{eg}^2) \end{pmatrix}. \quad (3.13)$$

The sample mean of \mathbf{Z}_{gi} is

$$\mathbf{m}_{1g} = \begin{pmatrix} m_{1g} \\ m_{2g} \\ 0 \\ 0 \end{pmatrix}, \quad (3.14)$$

where

$$\begin{aligned} m_{1g} &= \frac{\sum_{i=1}^{n_g} w_{gi} x_{gi}}{\sum_{i=1}^{n_g} w_{gi}}, \\ m_{2g} &= \frac{\sum_{i=1}^{n_g} w_{gi} y_{gi}}{\sum_{i=1}^{n_g} w_{gi}}, \end{aligned}$$

and n_g is the number of individuals in group g . The sample variance of \mathbf{Z}_{gi} is

$$\mathbf{m}_{2g} = \frac{\sum_{i=1}^{n_g} w_{gi} (\mathbf{Z}_{gi} - \bar{\mathbf{Z}}_g)(\mathbf{Z}_{gi} - \bar{\mathbf{Z}}_g)'}{\sum_{i=1}^{n_g} w_{gi}},$$

where

$$\bar{\mathbf{Z}}_g = \frac{\sum_{i=1}^{n_g} w_{gi} \mathbf{Z}_{gi}}{\sum_{i=1}^{n_g} w_{gi}}$$

is the group sample mean of the \mathbf{Z}_{gi} . For deriving the method of moments estimating equations,

we write

$$\mathbf{m}_{2g} = \begin{pmatrix} m_{11g} & m_{12g} & 0 & 0 \\ & m_{22g} & 0 & 0 \\ & & m_{33g} & m_{34g} \\ \text{sym.} & & & m_{44g} \end{pmatrix}, \quad (3.15)$$

where the sample moments m_{13g} , m_{14g} , m_{23g} , and m_{24g} are set to zero since their corresponding population moments in (3.13) are all zero.

The estimating equations are

$$\mathbf{m}_{1g} = E\{\mathbf{Z}_{gi}\}$$

and

$$\mathbf{m}_{2g} = V\{\mathbf{Z}_{gi}\},$$

where \mathbf{m}_{1g} and \mathbf{m}_{2g} are defined by (3.14) and (3.15), respectively, and $E\{\mathbf{Z}_{gi}\}$ and $V\{\mathbf{Z}_{gi}\}$ are defined by (3.12) and (3.13), respectively. There are eight model parameters and eight unique first and second moments in these equations, which allows for identification of each model parameter as a function of the sample moments. The method of moments estimators are given in Table 3.1. In what follows, we let

$$\hat{\boldsymbol{\theta}}_g = (\hat{\mu}_g, \hat{\mu}_{yg}, \hat{\beta}_{1g}, \hat{\sigma}_{tg}^2, \hat{\sigma}_{dg}^2, \hat{\sigma}_{ug}^2, \hat{\sigma}_{eg}^2, \hat{\sigma}_{rg}^2)' \quad (3.16)$$

denote the method of moments estimator for the parameter vector $\boldsymbol{\theta}_g$ in (3.10).

Table 3.1 Method of moments estimators for group g

Parameter	Estimator
μ_g	$\hat{\mu}_g = m_{1g}$
μ_{yg}	$\hat{\mu}_{yg} = m_{2g}$
β_{1g}	$\hat{\beta}_{1g} = (m_{12g} - 0.25m_{34g}) / (m_{11g} - 0.25m_{33g})$
σ_{tg}^2	$\hat{\sigma}_{tg}^2 = m_{11g} - 0.25m_{33g}$
σ_{dg}^2	$\hat{\sigma}_{dg}^2 = [m_{34g}(m_{11g} - 0.25m_{33g})] / [2(m_{12g} - 0.25m_{34g})]$
σ_{ug}^2	$\hat{\sigma}_{ug}^2 = 0.5m_{33g} - [m_{34g}(m_{11g} - 0.25m_{33g})] / [2(m_{12g} - 0.25m_{34g})]$
σ_{eg}^2	$\hat{\sigma}_{eg}^2 = 0.5m_{44g} - [m_{34g}(m_{12g} - 0.25m_{34g})] / [2(m_{11g} - 0.25m_{33g})]$
σ_{rg}^2	$\hat{\sigma}_{rg}^2 = m_{22g} - 0.25m_{44g} - [(m_{12g} - 0.25m_{34g})^2] / [m_{11g} - 0.25m_{33g}]$

A Taylor series approximation is used to derive an estimated variance matrix for $\hat{\boldsymbol{\theta}}_g$. The approximation is given by

$$\hat{V}\{\hat{\boldsymbol{\theta}}_g\} = \hat{\mathbf{D}}_g \hat{V}\{\mathbf{m}_g\} \hat{\mathbf{D}}_g', \quad (3.17)$$

where $\hat{\mathbf{D}}_g$ is a matrix of derivatives for the method of moments estimators evaluated at the method of moments estimates and $\hat{V}\{\mathbf{m}_g\}$ is an estimated variance of the sample moments

$$\mathbf{m}_g = (m_{1g}, m_{2g}, m_{11g}, m_{12g}, m_{22g}, m_{33g}, m_{34g}, m_{44g})'. \quad (3.18)$$

To derive the matrix of derivatives, let m_{gk} denote the k th element in \mathbf{m}_g for $k = 1, \dots, 8$ and let $b_l(\mathbf{m}_g)$ be a function of \mathbf{m}_g that represents the l th method of moments estimator in Table 3.1 for $l = 1, \dots, 8$. Then, define $\hat{\mathbf{D}}_g$ to be an 8×8 matrix of derivatives for the sample moments, where element lk in $\hat{\mathbf{D}}_g$ is

$$D_{g lk} = \frac{\partial b_l(\mathbf{m}_g)}{\partial m_{gk}}$$

for $l = 1, \dots, 8$ and $k = 1, \dots, 8$. The values for $D_{g lk}$ are given in Table 3.2.

Table 3.2 Elements $D_{g lk}$ in the derivative matrix $\hat{\mathbf{D}}_g$, where $f_1 = m_{12} - 0.25m_{34}$ and $f_2 = m_{11} - 0.25m_{33}$

	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$
$l = 1$	1	0	0	0	0	0	0	0
$l = 2$	0	1	0	0	0	0	0	0
$l = 3$	0	0	$\frac{-f_1}{f_2^2}$	$\frac{1}{f_2}$	0	$\frac{f_1}{4f_2^2}$	$\frac{-1}{4f_2}$	0
$l = 4$	0	0	1	0	0	-0.25	0	0
$l = 5$	0	0	$\frac{m_{34}}{2f_1}$	$\frac{-m_{34}f_2}{2f_1^2}$	0	$\frac{-m_{34}}{8f_1}$	$\frac{2f_1f_2+0.5m_{34}f_2}{4f_1^2}$	0
$l = 6$	0	0	$\frac{-m_{34}}{2f_1}$	$\frac{m_{34}f_2}{2f_1^2}$	0	$0.5 + \frac{m_{34}}{8f_1}$	$-\frac{2f_1f_2+0.5m_{34}f_2}{4f_1^2}$	0
$l = 7$	0	0	$\frac{m_{34}f_1}{2f_2^2}$	$\frac{-m_{34}}{2f_2}$	0	$\frac{-m_{34}f_1}{8f_2^2}$	$\frac{-m_{12}+0.5m_{34}}{2f_2}$	0.5
$l = 8$	0	0	$\frac{f_1^2}{f_2^2}$	$\frac{-2f_1}{f_2}$	1	$\frac{-f_1^2}{4f_2^2}$	$\frac{f_1}{2f_2}$	-0.25

The variance of \mathbf{m}_g can be estimated using a Horvitz-Thompson variance to account for the sample design. The Horvitz-Thompson variance estimator is

$$\hat{V}\{\mathbf{m}_g\} = \sum_{i=1}^{n_g} \sum_{k=1}^{n_g} \pi_{ik}^{-1} (\pi_{ik} - \pi_i \pi_k) w_{gi} \mathbf{s}_{gi} w_{gk} \mathbf{s}_{gk}',$$

where π_i is the first order inclusion probability of individual i into the sample, π_{ik} is the second order inclusion probability of individuals i and k into the sample, w_{gi} is the survey weight for

individual i in group g , and

$$\mathbf{s}_{gi} = \begin{pmatrix} \bar{x}_{gi.} \\ \bar{y}_{gi.} \\ (\bar{x}_{gi.} - m_{1g})^2 \\ (\bar{x}_{gi.} - m_{1g})(\bar{y}_{gi.} - m_{2g}) \\ (\bar{y}_{gi.} - m_{2g})^2 \\ (x_{gi1} - x_{gi2})^2 \\ (x_{gi1} - x_{gi2})(y_{gi1} - y_{gi2}) \\ (y_{gi1} - y_{gi2})^2 \end{pmatrix} \quad (3.19)$$

is the vector of summary statistics for individual i .

3.2.3 Population-Level Model

In the previous section we developed estimators which can be used to estimate the group-level model parameters, including the group mean (μ_g) and variance (σ_{tg}^2) of daily EE in the normal scale. Although it is of interest to estimate separate parameters for each of the G groups, it is possible that the group-level parameters can be modeled across the groups to form a population-level model with a reduced number of parameters. In this section, we outline a procedure for developing a population-level model from the group-level model parameters. First we give the general form of the model and an estimator for the model parameter vector. Then we illustrate how the model can be formulated.

The population-level model is defined by a set of functions that model the group-level parameters in $\boldsymbol{\theta}_g$ given by (3.10) as functions of a new set of parameters defined for the population. The set of functions and population-level model parameters are formulated based on an analysis of the group-level parameter estimates. We illustrate how one can formulate the model later in this section. The general form of the population-level model is

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\lambda} + \mathbf{e}. \quad (3.20)$$

In the model,

$$\mathbf{y} = (\hat{\boldsymbol{\theta}}'_1, \dots, \hat{\boldsymbol{\theta}}'_G)'$$

is the $8G$ -dimensional vector of the estimated group-level model parameters, where $\hat{\boldsymbol{\theta}}_g$ is given by (3.16) for group g , $g = 1, \dots, G$. $\boldsymbol{\lambda}$ is the q -dimensional vector of parameters for model (3.20), where $q < 8G$ so that the total number of parameters from the group-level models is smaller for the population-level model. \mathbf{Z} is a $(8G \times q)$ design matrix for the model representing coefficients that define the set of functions that relate the $8G$ group-level estimated parameters to linear functions of the q population-level parameters. The variance of the vector of error terms, $\mathbf{e} \sim (\mathbf{0}, \mathbf{V})$, is estimated by

$$\hat{\mathbf{V}} = \text{blockdiag}(\hat{V}\{\hat{\boldsymbol{\theta}}_1\}, \dots, \hat{V}\{\hat{\boldsymbol{\theta}}_G\}), \quad (3.21)$$

where $\hat{V}\{\hat{\boldsymbol{\theta}}_g\}$ is given by (3.17) for $g = 1, \dots, G$. The estimated variance (3.21) is appropriate under the assumption that the EE measurements are uncorrelated across groups. With an estimated variance $\hat{\mathbf{V}}$, the population-level model can be estimated using estimated generalized least squares (EGLS). The EGLS estimator of $\boldsymbol{\lambda}$ is

$$\hat{\boldsymbol{\lambda}} = (\mathbf{Z}'\hat{\mathbf{V}}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{V}}^{-1}\mathbf{y}, \quad (3.22)$$

and an estimated variance of the estimator is

$$\hat{V}\{\hat{\boldsymbol{\lambda}}\} = (\mathbf{Z}'\hat{\mathbf{V}}^{-1}\mathbf{Z})^{-1}. \quad (3.23)$$

We illustrate how model (3.20) can be formulated by considering a simple reduced version of (3.20) for the group-level model parameters μ_g and σ_{tg}^2 . Suppose that the estimates of σ_{tg}^2 are similar across all groups defined by age. One may decide to use a common σ_t^2 in the population-level model by defining the function

$$\sigma_{tg}^2 = \sigma_t^2, \quad (3.24)$$

where σ_t^2 represents the mean daily EE variance in the normal scale for the population. Also suppose that the μ_g are linearly related to the mean age of age group. One can express the relationship by

$$\mu_g = \mu_0 + \theta A_g, \quad (3.25)$$

where μ_0 is a baseline parameter for the mean daily EE in the population, A_g is the mean age in age group g , and θ represents the linear relationship in μ_g in relation to the mean age in a group. Focusing only on the daily EE mean and variance, the population-level model representing (3.24) and (3.25) is given as

$$\begin{pmatrix} \hat{\mu}_1 \\ \hat{\sigma}_{t1}^2 \\ \hat{\mu}_2 \\ \hat{\sigma}_{t2}^2 \\ \vdots \\ \hat{\mu}_G \\ \hat{\sigma}_{tG}^2 \end{pmatrix} = \begin{pmatrix} 1 & A_1 & 0 \\ 0 & 0 & 1 \\ 1 & A_2 & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & A_G & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_0 \\ \theta \\ \sigma_t^2 \end{pmatrix} + \mathbf{e}, \quad (3.26)$$

where $\hat{\mu}_g$ is the estimated value of μ_g and $\hat{\sigma}_{tg}^2$ is the estimated value of σ_{tg}^2 from the group-level model for $g = 1, \dots, G$. In practice, one would incorporate the other group-level model parameters in the population-level model by defining functions for the parameters, similar to the way in which the functions (3.24) and (3.25) were defined for μ_g and σ_{tg}^2 . In Section 3.3 we define a more complete population-level model for the group-level model parameters using preliminary data from PAMS.

After an initial fitting, the population-level model may be re-formulated. For example, if the parameter θ in (3.26) is non-significant in the fitted population-level model, the θA_g term may be dropped from the model. Alternatively, a new function relating μ_g and group mean age (A_g) could be considered in the population-level model.

Using the final population-level model, we estimate model parameters for each group. In the simple population-level model given by (3.24) and (3.25), the estimated mean of daily EE in the normal scale for group g is $\hat{\mu} + \hat{\theta}A_g$ and the estimated variance of mean daily EE in the normal scale for group g is $\hat{\sigma}_t^2$, where $\hat{\mu}$, $\hat{\theta}$, and $\hat{\sigma}_t^2$ are estimated from model (3.26) and A_g is the group mean of age for group g . Let $\hat{\mu}_g$ and $\hat{\sigma}_{tg}^2$ denote the estimated mean daily EE and estimated variance of daily EE in the normal scale for group g , $g = 1, \dots, G$, from the population-level model. The estimated distribution of mean daily EE in the normal scale for

group g is then $N(\hat{\mu}_g, \hat{\sigma}_{tg}^2)$. Estimates of other model parameters can also be obtained using the population-level model, including estimates of the slope parameters relating actual daily EE to self-reported EE ($\hat{\beta}_{1g}$), estimates of the group means of self-reported EE ($\hat{\mu}_{yg}$), and estimates of the variance components that account for day-to-day variation in daily EE ($\hat{\sigma}_{dg}^2$), measurement error variation in the reference instrument ($\hat{\sigma}_{ug}^2$) and self-report instrument ($\hat{\sigma}_{eg}^2$), and random variation due to self-reporting ($\hat{\sigma}_{rg}^2$).

3.2.4 Estimating Parameters of Usual Daily EE in the Original Scale

Researchers are often interested in estimating parameters of usual daily EE in the original scale for subpopulations. In the final step of our method, we develop a procedure for generating estimated distributions of usual daily EE in the original scale for each group and describe how to estimate some parameters of usual daily EE using the estimated distributions. To estimate distributions of usual daily EE, a set of values for each group are generated from the estimated normal-scale distribution of mean daily EE. The conditional expectation of the original-scale daily EE values conditional on the normal-scale daily EE values is estimated. This procedure accounts for the potential bias in transforming an individual's daily EE from the normal scale to the original scale. A simple back-transformation of the mean daily EE values using the inverse of the transformation used to approximate normality would give potentially biased values of usual daily EE in the original scale because the mean of a nonlinearly transformed variable is not equal to the transformed mean of the original-scale variable. The set of original scale values given by the estimated conditional expectation are then used to estimate a distribution of usual daily EE in the original scale for each group and to estimate usual daily EE parameters in the original scale.

For group g , we first randomly generate a set of $m = 100,000$ daily EE values, $\ddot{t}_{g1}, \dots, \ddot{t}_{gm}$ from the estimated $N(\hat{\mu}_g, \hat{\sigma}_{tg}^2)$ distribution. Next, we transform the generated values to usual daily EE values in the original scale. Given the measurement error model for daily EE, the usual daily EE value for individual i in group g is

$$\ddot{T}_{gi} = E\{h^{-1}(t_{gi} + d_{gij} + u_{gij}) | t_{gi} = \ddot{t}_{gi}\},$$

where $h(\cdot)$ is the transformation from Section 3.2.1 taking the daily EE values into the normal scale and $h^{-1}(t_{gi} + d_{gij} + u_{gij})$ represents the monitor daily EE value in the original scale. When $h(\cdot) = \log(\cdot)$, the usual daily EE value is

$$\begin{aligned}\ddot{T}_{gi} &= E\{\exp(t_{gi} + d_{gij} + u_{gij})|t_{gi} = \ddot{t}_{gi}\} \\ &= \exp(\ddot{t}_{gi})E\{\exp(d_{gij} + u_{gij})|t_{gi} = \ddot{t}_{gi}\} \\ &= \exp(\ddot{t}_{gi})\exp[(1/2)(\sigma_{dg}^2 + \sigma_{ug}^2)]\end{aligned}\tag{3.27}$$

since t_{gi} , d_{gij} , and u_{gij} are assumed to be uncorrelated, and $\exp(d_{gij} + u_{gij})$ has a lognormal distribution with mean $\exp[(1/2)(\sigma_{dg}^2 + \sigma_{ug}^2)]$ under the assumption that $d_{gij} \sim N(0, \sigma_{dg}^2)$ and $u_{gij} \sim N(0, \sigma_{ug}^2)$. An estimate of \ddot{T}_{gi} is given by substituting $\hat{\sigma}_{dg}^2$ and $\hat{\sigma}_{ug}^2$ for σ_{dg}^2 and σ_{ug}^2 , respectively.

If $h(\cdot)$ is a function other than the log function used to achieve normality, the transformation taking the normal-scale EE values into the original scale may be approximated using other methods. For example, Dodd et al. (2006) consider the Taylor expansion

$$\ddot{T}_i = g(\ddot{t}_i) + (1/2)g''(\ddot{t}_i)(\sigma_w^2)$$

for transforming normal-scale nutrient intake values into the original scale, where $g(\cdot)$ is the inverse of a power transformation or a Box-Cox transformation with second derivative $g''(\cdot)$, \ddot{t}_i is a normal-scale nutrient intake value, and σ_w^2 is the within-individual variance of the nutrient intake values in the normal scale. Nusser et al. (1996) give a procedure for taking usual intake values from the normal scale into the original scale when a semiparametric transformation is initially used to achieve normality. These procedures are reviewed in more detail in Section 1.4.4 of Chapter 1.

The set of usual daily EE values in the original scale, $\ddot{T}_{g1}, \dots, \ddot{T}_{gm}$, can be used to estimate usual daily EE parameters in the original scale for each group. For example, the estimated mean and variance of usual daily EE in the original scale for group g are

$$\bar{T}_g = m^{-1} \sum_{i=1}^m \ddot{T}_{gi}$$

and

$$S_{T_g}^2 = (m-1)^{-1} \sum_{i=1}^m (\ddot{T}_{gi} - \bar{T}_{g.})^2,$$

respectively. The original scale values can also be used to estimate the proportion of individuals in the group above or below some EE threshold value. For example, the estimated proportion of individuals below an EE value of T_{val} is

$$\hat{p}_{T_{val}} = m^{-1} \sum_{i=1}^m I(\ddot{T}_{gi} < T_{val}), \quad (3.28)$$

where $I(\ddot{T}_{gi} < T_{val})$ is 1 if $\ddot{T}_{gi} < T_{val}$ and is 0 otherwise. When the transformation to normality is a log transformation, it is not necessary to use the generated values, $\ddot{T}_{g1}, \dots, \ddot{T}_{gm}$, to estimate these parameters in the original scale because the distribution generating these values is a lognormal distribution multiplied by a constant, as given by (3.27). The distribution can be used to directly estimate the mean, variance, and quantiles.

Estimated variances of the estimated usual daily EE parameters can be obtained using delete-1 jackknife variance estimation (Section 4.2 of Fuller 2009). A jackknife variance estimator is given in Appendix C for the population-level model parameter vector $\boldsymbol{\lambda}$ for a stratified design. The jackknife procedure in Appendix C can be extended to estimate the variance of original-scale usual daily EE parameters. Using the replicate estimates of $\boldsymbol{\lambda}$ given in Appendix C, we compute replicate sets of $m = 100,000$ original-scale usual daily EE values using the procedure defined above and compute replicate estimates of the usual daily EE parameter of interest. For illustration, let $\bar{T}_{g.}$ be the estimated mean of usual daily EE in group g and let $\bar{T}_{g.}^{(hi)}$ be the h th replicate estimate of $\bar{T}_{g.}$ for individual i in stratum h . The estimated jackknife variance of $\bar{T}_{g.}$ is

$$\hat{V}\{\bar{T}_{g.}\} = \sum_{h=1}^H N_h^{-1} (N_h - n_h) n_h^{-1} (n_h - 1) \sum_{i=1}^{n_h} (\bar{T}_{g.}^{(hi)} - \bar{T}_{g.})^2,$$

where n_h is the number of sampled individuals in stratum h and N_h is the total number of individuals in stratum h . The jackknife variance estimator is not always appropriate for nonsmooth functions such as sample quantiles (Section 4.2 of Fuller 2009). In our analyses, sample quantiles such as $\hat{p}_{T_{val}}$ given in (3.28) are computed using the \ddot{T}_{gi} values to approximate

a distribution where quantiles are smooth functions of the estimated parameters of model (3.20).

3.3 Application to PAMS Data

In this section, we use the method described in Section 3.2 to estimate usual daily EE parameters using preliminary EE data from a sample of 171 females from the Physical Activity Measurement Survey (PAMS). In Section 3.3.1, we describe the PAMS survey design. In Section 3.3.2, we present the daily EE data for the sample of 171 females. In Section 3.3.3, we use the methodology from Section 3.2 to estimate usual daily EE parameters for four age groups from the female sample.

3.3.1 Survey Design

The Physical Activity Measurement Survey (PAMS) is a survey conducted in four Iowa counties (Black Hawk, Dallas, Marshall, and Polk) starting in Fall of 2009. A multi-stage stratified probability design is used to select individuals from the counties. There are 2 strata per county, for a total of 8 strata. In each county, one stratum is a “high minority” defined by Census tracts that have relatively high percentages of minorities and the other stratum is a “low minority” stratum defined by Census tracts that have relatively low percentages of minorities. The “high minority” strata are oversampled to achieve a higher percentage of minorities in the sample. Households in each stratum are systematically selected from a white pages listing of telephone numbers. Every three months (quarter of a year), a new household sample is selected. The preliminary data we use in our analyses are from the first quarter of an eight quarter sample. We refer to this sample as the preliminary sample in the remainder of the presentation.

A screening interview is used to randomly select 1 eligible adult in each household to participate in the survey. To be eligible, the adult has to be between the ages of 21 and 70, capable of physical activity engagement, and competent to be interviewed. After agreeing to participate in the study, each respondent in the sample provides EE data from a SenseWear

armband monitor and a 24-hour physical activity recall (24PAR) on two measurement days. To be considered independent, the measurement days are randomly assigned approximately two to three weeks apart. On the assigned measurement days, the respondent wears the armband monitor for the full 24 hours of the day, except for water activities such as swimming and showering. The day following a measurement day, the respondent is contacted to complete a 24PAR by telephone. During the 24PAR, the respondent reports on the activities he or she engaged in during the measurement day.

A survey weight is computed for each individual in the sample. The base weight for individual i from household k in stratum h is

$$w_{hki} = \frac{N_h}{n_h} n_{hk}, \quad (3.29)$$

where N_h is the total number of households listed for stratum h , n_h is the number of households selected from stratum h , and n_{hk} is the number of eligible adults in household k in stratum h . For the preliminary sample of females, the base weights are adjusted for nonresponse by stratum, and then post-stratified to match the 2000 U.S. Census totals for 20 - 69 year-olds by county and gender. The final weight for female i in the sample from household k in stratum h is

$$w_{hki}^{ps} = \left(\frac{N_h n_{hk}}{n_{R,h}} \right) \frac{N_h^f}{\sum_{k=1}^{n_h} \frac{N_h n_{hk}}{n_{R,h}}},$$

where N_h^f is the 2000 U.S. Census total for adult females age 20 - 69 in stratum h and $n_{R,h}$ is the number of respondent households in stratum h , with $n_{R,h} \leq n_h$. Percentiles for the final survey weights for females in the preliminary PAMS sample are given in Table 3.3. The sample sizes and population control totals for each of the strata are given in Table C.2 in Appendix C.

Table 3.3 Percentiles for PAMS survey weights for the preliminary female sample

Percentile	0	10	25	50	75	90	100
Survey Weight	113	210	271	527	1053	2107	3838

Measurements of daily EE in kilocalories per day (kcal/d) are computed from the monitor and the 24PAR for each respondent on each measurement day. The monitors contain internal algorithms that estimate daily EE based on the activity data that are recorded and the respondent's height and weight. The activities reported using the 24PAR are assigned metabolic equivalent (MET) intensity levels using a modified version of the Compendium of Physical Activities (Ainsworth et al. 2000; Ainsworth et al. 1993). The activities are converted into an estimate of daily EE using the conversion equation

$$1 \text{ MET} = 0.0175 \text{ kcal/kg/min.}$$

To illustrate the use of the equation, suppose that a respondent weighing 70 kg reports engaging in an activity at a MET level of 4 for 10 minutes. The EE associated with this activity is estimated to be $4 \times 0.0175 \times 70 \times 10 = 49$ kcals.

3.3.2 Female Daily EE

For our analyses, we consider data from females in the preliminary PAMS sample, who were measured for daily EE during October 2009 to December 2009. One hundred and seventy one females were measured for daily EE from the monitor and 24PAR on two measurement days. Each female in this sample wore the monitor for at least 85% of the day and reported activity for at least 85% of the day for each measurement day. Over 90% of the females (154 out of 171) in this sample wore the monitor for more than 95% of the day and reported on activity for 95% of the day. Activity that occurred during time unaccounted for by the monitor or recall was estimated to be at the individual's resting rate of 1 MET or 0.0175 kcal/kg/min.

Unweighted demographic characteristics of the sample are given in Table 3.4. The median age in the sample is 53 and about half of the females in the sample are between age 40 and 60. Only a small portion of the sample is composed of Hispanics and blacks. Just under 40% of the females in the sample have college degrees and just under 20% of the females in the sample are self-identified smokers.

In Figure 3.1, we give side-by-side boxplots that compare the distributions of the monitor and 24PAR EE data in the sample. Both distributions are skewed to the right, but the

Table 3.4 Demographic characteristics of the female PAMS sample

Median (IQR) Age	53 (18)
Age Range	23-70
Count (%) of Hispanics	3 (1.8)
Count (%) of blacks	16 (9.4)
Count (%) of College Graduates	66 (38.6)
Count (%) of Smokers	31 (18.1)

skewness is more extreme for the 24PAR data. In Figure 3.1 we also present a scatter plot of the individual means of daily EE from the 24PAR versus monitor EE. The plot suggests that 24PAR EE is over-estimated in relation to monitor EE.

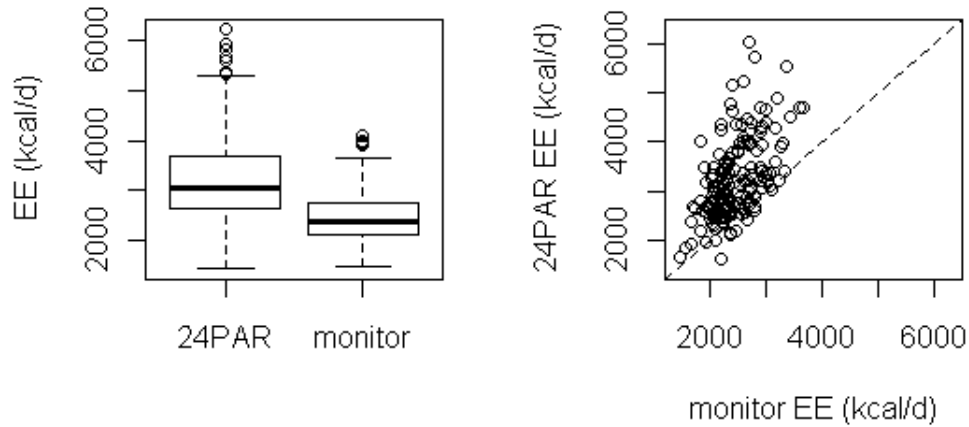


Figure 3.1 Boxplots of daily EE from the monitors and 24PARs and scatter plot of individual means of daily EE from the monitors and 24PARs (dashed line is the identity line)

3.3.3 Methodology for the PAMS Sample

In this section, we use the methodology from Section 3.2 to estimate usual daily EE parameters from the preliminary female PAMS sample. For simplicity, we consider an unweighted analyses in our presentation because the unweighted and survey-weighted results were shown to be similar based on the test in Appendix C. The results from fitting the population-level

model using the survey weights are given in Appendix C. As in Section 3.2, let X_{ij} denote the original scale EE from the monitor and let Y_{ij} denote the original scale EE from the 24PAR for individual i on day j .

3.3.3.1 Transformation to Normality and Check for Nuisance Effects

First, we transform the daily EE data to approximate normality. Because log transformations are often used for analyses in physical activity research (Ferrari et al. 2007), we consider the log transformation to approximate normality for the PAMS EE data. Let $x_{ij} = \log(X_{ij})$ be daily EE from the monitor and let $y_{ij} = \log(Y_{ij})$ be daily EE from the 24PAR in the log scale for individual i on day j . Shapiro-Wilk test statistics are computed for the set of x_{ij} values and set of y_{ij} values from the sample using SAS statistical software (SAS Institute 2009). The p-values for the test statistics are 0.25 and 0.21 for the set of x_{ij} and y_{ij} values, respectively. Therefore, the log transformed values are used for model fitting.

Next, we check for nuisance effects in the log-transformed daily EE data by fitting linear regression models containing covariates for day-of-week effect, time-in-sample effect, and demographic variables. We include variables for day-of-week effect and time-in-sample effect in the models because we suspect that an individual may have different EE values depending on the day of the week (e.g., weekday vs. weekend) and depending on whether the value is the first or second observation for a respondent (e.g., replicate 1 vs. 2). We include demographic variables for age, race/ethnicity, education, and smoking status in the models because we suspect that EE levels may vary by these factors. In preliminary fits we also included variables for town size and number of adults in the household, but these variables were non-significant and are not considered in this presentation. The model variable for day-of-week effect is an indicator variable, which takes a value of 1 if day j is Saturday or Sunday and takes a value of 0 otherwise. The model variable for time-in-sample effect is an indicator variable, which takes a value of 1 if day j is the first measurement day of the individual and a value of 0 otherwise. The model variable for age is the actual age of the individual. The variable age squared is also included in the model to account for a quadratic relationship between daily EE and age

observed in exploratory analyses. The model variables for the other demographic variables are indicator variables for Hispanic, black, college graduate, and smoker.

The models are fit using the weighted least squares estimators given by (3.3) and (3.4) in Section 3.2.1, where the weights w_i are all set to 1. Estimated variances for the regression coefficients are computed using the Taylor variances given by (3.5) and (3.6) in Section 3.2.1. Test statistics are computed using (3.7) and p-values are computed for each of the test statistics. The p-values are given in Table 3.5. The nuisance effects (day of week and time in sample) are not significant in either of the model fits (p-values greater than 0.05), and, given these results, we do not adjust the PAMS EE data for nuisance effects. The nuisance effects are also non-significant when the models are fit using the survey weights (results not shown).

The results in Table 3.5 suggest that age squared is a significant indicator of daily EE from the monitor and there is suggestive evidence of a difference for blacks vs. non-blacks in the monitor data. Being Hispanic is a significant indicator of daily EE from the 24PAR at the 0.05 level. In this presentation, we consider age groups when conducting our group-level analyses, but not race/ethnicity groups. There are only 3 Hispanics and 16 blacks in the preliminary female sample. As more data become available, we will consider groups for race/ethnicity in our analyses.

Table 3.5 P-values for the linear regression models fit to the log-transformed data

	Monitor Model (x_{ij})	24PAR Model (y_{ij})
Variable	P-value	P-value
Intercept	< 0.01	< 0.01
Day of Week	0.87	0.69
Time in Sample	0.83	0.54
Age	0.12	0.54
Age Squared	0.03	0.46
Hispanic	0.18	0.01
Black	0.07	0.12
College Graduate	0.61	0.23
Smoker	0.62	0.88

3.3.3.2 Age Groups

In preliminary analyses, daily EE measurements from the monitor were shown to vary according to age. Based on these results, we define age groups for the group-level measurement error models. To form groups, we divide the sample into four age groups of approximately equal size (Table 3.6). For the remainder of the presentation, we will denote the age groups as groups 1 - 4, where 1 is the youngest age group and 4 is the oldest age group.

Table 3.6 Age groups

Age Group g	Age Range	Average Age	Sample Size
1	23 - 42	34.3	44
2	43 - 52	48.6	40
3	53 - 59	55.6	43
4	60 - 70	64.6	44

3.3.3.3 Group-Level Model

Once the groups have been determined, the next step in our method is to estimate the group-level model parameters. The measurement error model given by equations (3.8) and (3.9) is fit to each of the four age groups using method of moments as described in Section 3.2.3. The method of moments estimators are given in Table 3.1 in Section 3.2.2. Standard errors for the parameter estimates are computed using the Taylor series variance estimator given by (3.17), $\hat{V}\{\hat{\boldsymbol{\theta}}_g\} = \mathbf{D}_g \hat{V}\{\mathbf{m}_g\} \mathbf{D}_g'$. Due to the small number of individuals in each of the 4 age groups, we ignore the stratified design in computing the estimated variance of the sample moments, $\hat{V}\{\mathbf{m}_g\}$, and instead use the estimated variance for a simple random sample (ignoring the finite population correction) defined by

$$\hat{V}\{\mathbf{m}_g\} = n_g^{-1}(n_g - 1)^{-1} \sum_{i=1}^{n_g} (\mathbf{s}_{gi} - \bar{\mathbf{s}}_g)(\mathbf{s}_{gi} - \bar{\mathbf{s}}_g)',$$

where \mathbf{s}_{gi} is given by (3.19) and $\bar{\mathbf{s}}_g$ is the mean of the \mathbf{s}_{gi} in group g . The parameter estimates and standard errors from the measurement error models for each group are given in Tables 3.7 and 3.8.

Table 3.7 contains the estimated group-level measurement error model parameters for the mean of daily EE (μ_g), the mean of reported daily EE (μ_{yg}) and the slope relating mean daily EE to reported daily EE (β_{1g}). The estimated means of daily EE decrease by age group, suggesting that older females tend to have lower levels of mean daily EE compared to younger females. The estimated slope parameters also decrease by age group, suggesting that the relationship between average levels of mean daily EE and reported daily EE may be a function of age. The estimated means of reported daily EE are larger than the estimated means of daily EE, suggesting over-reporting in daily EE for all age groups. Unlike the daily EE means, the reported daily EE means do not show much of a trend across age groups. Given these results, we model the decreasing trends in the estimated daily EE means and the estimated slope parameters in the population-level model and estimate a common mean for reported daily EE in the next section.

Table 3.7 Estimated measurement error model parameters (and standard errors) for the mean of daily EE (μ_g), the mean of reported daily EE (μ_{yg}), and the slope for population-level reporting bias (β_{1g})

Parameter	Group 1	Group 2	Group 3	Group 4
μ_g	7.8421 (.0240)	7.8104 (.0230)	7.7595 (.0283)	7.7182 (.0241)
μ_{yg}	8.0616 (.0435)	8.0300 (.0326)	8.0656 (.0398)	8.0318 (.0315)
β_{1g}	1.2970 (.2103)	0.9226 (.1984)	0.8433 (.2081)	0.6982 (.1145)

Table 3.8 contains the estimated group-level measurement error model parameters for the variance components from models (3.8) and (3.9). No systematic trends in the components are discernible. In preliminary analysis of possible models for the variance components, the largest differences between age groups were non-significant (results not shown). As more data become available, evidence of relationships or differences in the variance components across age groups may surface. For this analyses, we assume constant variance components across age groups in the population-level model (next section).

Table 3.8 Estimated variance components (and standard errors) from the measurement error model for mean daily EE (σ_{tg}^2), within-individual EE variation in daily EE (σ_{dg}^2), measurement error variation from the monitor (σ_{ug}^2) and the 24PAR (σ_{eg}^2), and reporting-bias variation from the recall (σ_{rg}^2)

Parameter	Group 1	Group 2	Group 3	Group 4
σ_{tg}^2	0.0211 (.0059)	0.0170 (.0048)	0.0296 (.0077)	0.0235 (.0062)
σ_{dg}^2	0.0089 (.0027)	0.0044 (.0022)	0.0025 (.0015)	0.0065 (.0028)
σ_{ug}^2	0.0056 (.0028)	0.0047 (.0027)	0.0067 (.0028)	0.0044 (.0034)
σ_{eg}^2	0.0042 (.0041)	0.0066 (.0022)	0.0064 (.0024)	0.0079 (.0028)
σ_{rg}^2	0.0363 (.0088)	0.0221 (.0049)	0.0416 (.0095)	0.0259 (.0085)

3.3.3.4 Population-Level Model

Given the results from the fitted group-level models (Table 3.7 and 3.8), we develop a population-level model for daily EE. We model the daily EE mean for age group g as

$$\mu_g = \mu_0 + \theta A_g, \quad (3.30)$$

where μ_0 is a baseline parameter for the daily EE mean in the population, A_g is the mean age of age group g minus the overall mean age for the sample (see Table 3.6), and θ is a parameter to estimate the linear trend in the daily EE mean. We model the bias slope parameters as a function of mean age,

$$\beta_{1g} = \beta_1 + \beta_3 A_g, \quad (3.31)$$

where β_1 is the baseline slope for the population and β_3 accounts for the linear trend in the slopes across age groups. We model the group means of reported EE as

$$\mu_{yg} = \mu_y + \beta_{1g}(\mu_g - \mu_0),$$

where μ_y is the overall mean of reported EE and $\beta_{1g}(\mu_g - \mu_0)$ accounts for the deviation in the group-level reported EE mean from the overall mean. Given models (3.30) and (3.31), the model for the mean of reported EE can be written as

$$\mu_{yg} = \mu_y + (\beta_1 + \beta_3 A_g)\theta A_g. \quad (3.32)$$

The group-level variance components are related to population-level variance components through the system of equations

$$\begin{aligned}
\sigma_{tg}^2 &= \sigma_t^2 \\
\sigma_{dg}^2 &= \sigma_d^2 \\
\sigma_{ug}^2 &= \sigma_u^2 \\
\sigma_{eg}^2 &= \sigma_e^2 \\
\sigma_{rg}^2 &= \sigma_r^2,
\end{aligned} \tag{3.33}$$

for $g = 1, \dots, 4$.

The population-level model is given by (3.30) - (3.33), where the parameters on the left side of the equations are replaced by their respective estimates from the estimated parameter vector

$$\hat{\boldsymbol{\theta}}_g = (\hat{\mu}_g, \hat{\mu}_{yg}, \hat{\beta}_{1g}, \hat{\sigma}_{tg}^2, \hat{\sigma}_{dg}^2, \hat{\sigma}_{ug}^2, \hat{\sigma}_{eg}^2, \hat{\sigma}_{rg}^2)',$$

The vector of population-level model parameters from (3.30) - (3.33) is

$$\boldsymbol{\lambda} = (\mu_0, \mu_y, \theta, \beta_1, \beta_3, \sigma_t^2, \sigma_d^2, \sigma_u^2, \sigma_e^2, \sigma_r^2)'$$

The model is nonlinear since (3.32) is a nonlinear function of population-level model parameters. We fit the model using nonlinear EGLS. When the population-level model is linear, the model parameters are estimated using the approach from Section 3.2.3.

The parameter estimates and standard errors from the fitted model are given in Table 3.9. Each of the model parameters is significant at the 0.05 level. There is evidence of a linear trend across age groups in the daily EE mean (represented by θ), and evidence of a linear trend across age groups in the slope parameter (represented by β_3). The estimated mean of daily EE (μ_0) appears to be smaller than the estimated mean of reported daily EE (μ_y), indicating over-reporting bias in daily EE from the 24PAR. The estimated variance for individual reporting effects (σ_r^2) is large relative to the other estimated variances components. The estimated inter-individual variance in usual daily EE (σ_t^2) is about 4 times larger than the estimated within-individual variance in daily EE (σ_d^2).

Table 3.9 Parameter estimates (standard errors) for the population-level model

Parameter	Est	(SE)
μ_0	7.8004	(.0096)
μ_y	8.0544	(.0141)
100θ	-0.3767	(.0977)
β_1	0.9531	(.0705)
$100\beta_3$	-1.6351	(.4515)
$100\sigma_t^2$	2.1556	(.2492)
$100\sigma_d^2$	0.4969	(.0898)
$100\sigma_u^2$	0.5011	(.1126)
$100\sigma_e^2$	0.6558	(.1172)
$100\sigma_r^2$	2.1072	(.2954)

The parameter estimates from the population-model can be used to estimate normal-scale mean daily EE values in each of the 4 age groups. The estimated means are computed from equation (3.30) as $\hat{\mu}_g = \hat{\mu}_0 + \hat{\theta}A_g$, where $\hat{\mu}_0$ and $\hat{\theta}$ are given in Table 3.9. Standard errors for the estimated means are given by

$$se(\hat{\mu}_g) = \sqrt{\mathbf{c}'_g \hat{V}\{\hat{\boldsymbol{\lambda}}\} \mathbf{c}_g},$$

where

$$\mathbf{c}'_g = (1, 0, A_g, 0, 0, 0, 0, 0, 0, 0)$$

and $\hat{V}\{\hat{\boldsymbol{\lambda}}\}$ is the estimated variance matrix for the population-level model. The estimates and standard errors are given in Table 3.10. The parameter estimates from the population-level model can also be used to estimate the slope parameters for each of the age groups based on equation (3.31). The estimates (and standard errors) for the slope parameters are given in Table 3.10. Note that the estimated means and slope parameters in Table 3.10 are similar to the estimated means and slope parameters in Table 3.7 for the fitted group-level models.

Figure 3.2 illustrates the relationships between mean daily EE and reported daily EE in the youngest and oldest age groups. In the youngest age group (group 1), females with higher levels of usual daily EE tend to have a greater discrepancy between their reported and mean

Table 3.10 Parameter estimates (standard errors) for the daily EE group means (μ_g) and slope parameters (β_{1g}) based on the fitted population-level model

Parameter	Est	(SE)
μ_1	7.8625	(.0187)
μ_2	7.8087	(.0098)
μ_3	7.7824	(.0108)
μ_4	7.7484	(.0167)
β_{11}	1.2228	(.1167)
β_{12}	0.9889	(.0740)
β_{13}	0.8748	(.0673)
β_{14}	0.7275	(.0791)

daily EE, while in the oldest age group (group 4), females with higher levels of mean daily EE tend to have a smaller discrepancy between their reported and mean daily EE.

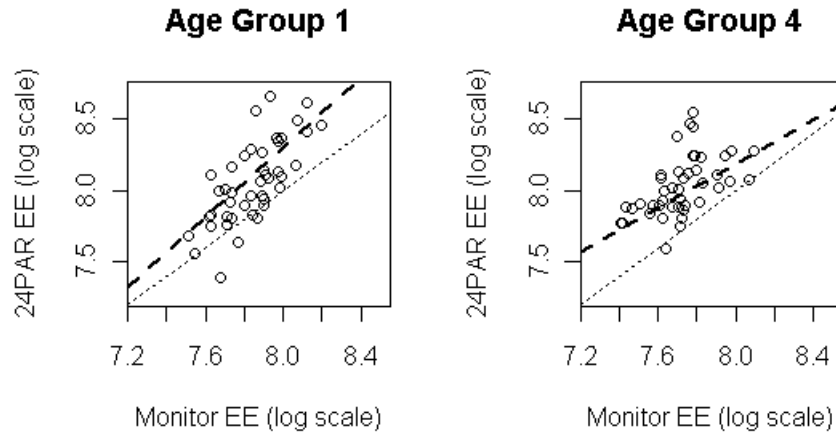


Figure 3.2 Estimated lines relating mean daily EE and reported daily EE for age groups 1 and 4 (points are the individual means of measured EE in the log scale; dashed lines are the estimated lines and dotted lines are the identity lines)

3.3.3.5 Estimated Usual Daily EE Parameters in the Original Scale

As a final step in our analyses, we present plots with estimated distributions of usual daily EE in the original scale and give estimated parameters of usual daily EE in the original scale.

Given the estimated daily EE means, $\hat{\mu}_1, \dots, \hat{\mu}_4$, in Table 3.10 and the estimated mean daily EE variance, $\hat{\sigma}_t^2$, in Table 3.9, the estimated distributions of mean daily EE in the normal scale are $N(7.8625, 0.0216)$, $N(7.8087, 0.0216)$, $N(7.7824, 0.0216)$, and $N(7.7484, 0.0216)$ for age groups 1 - 4, respectively. The estimated distributions of usual daily EE are computed using the procedure given in Section 3.2.4. For each group, $m = 100,000$ values, $\ddot{t}_{g1}, \dots, \ddot{t}_{gm}$, are generated from the estimated normal distribution and each value is transformed into the original scale using equation (3.27), where estimates of σ_{dg}^2 and σ_{ug}^2 , $\hat{\sigma}_d^2 = 0.0050$ and $\hat{\sigma}_u^2 = 0.0050$, respectively, are obtained from the fitted population-level model. Figure 3.3 gives the estimated density functions for each of the estimated age group distributions of usual daily EE. Density values for the plots are computed using the R function *density()*, which computes an empirical distribution function over a grid of points and uses a linear approximation to evaluate the densities at the specified points. There is a slight right skew in each of the density functions. The distributions shift to the left as age group goes from 1 to 4, which is a consequence of the estimated trend in the usual daily EE means across the age groups.

In Figure 3.4 we give the estimated density of usual daily EE in kcal/d for age group 1 (age < 43), along with the estimated densities for the individual means of daily EE from the monitor and the 24PAR. The estimated density function based on the monitor means has slightly more dispersion than the estimated density of usual daily EE due to the measurement error variance in the monitor model. The monitor density has the same mean as the usual daily EE density, which is a result of the assumption that the monitor gives unbiased measurements of daily EE. The estimated density from the 24PAR has a larger spread than either the usual daily EE density or the monitor means density due to the excess variability in the 24PAR EE values. The 24PAR density is also shifted to right relative to the two other density functions due to the over-reporting bias in the 24PAR for age group 1.

Using the estimated distributions of usual daily EE in the original scale, we can estimate usual daily EE parameters. The group-level parameters we consider are the mean of usual daily EE in the original scale (\bar{T}_g), the standard deviation of usual daily EE in the original scale (S_{Tg}), the proportion of individuals with less than 1750 kcal/d of usual daily EE ($p_{low,g}$),

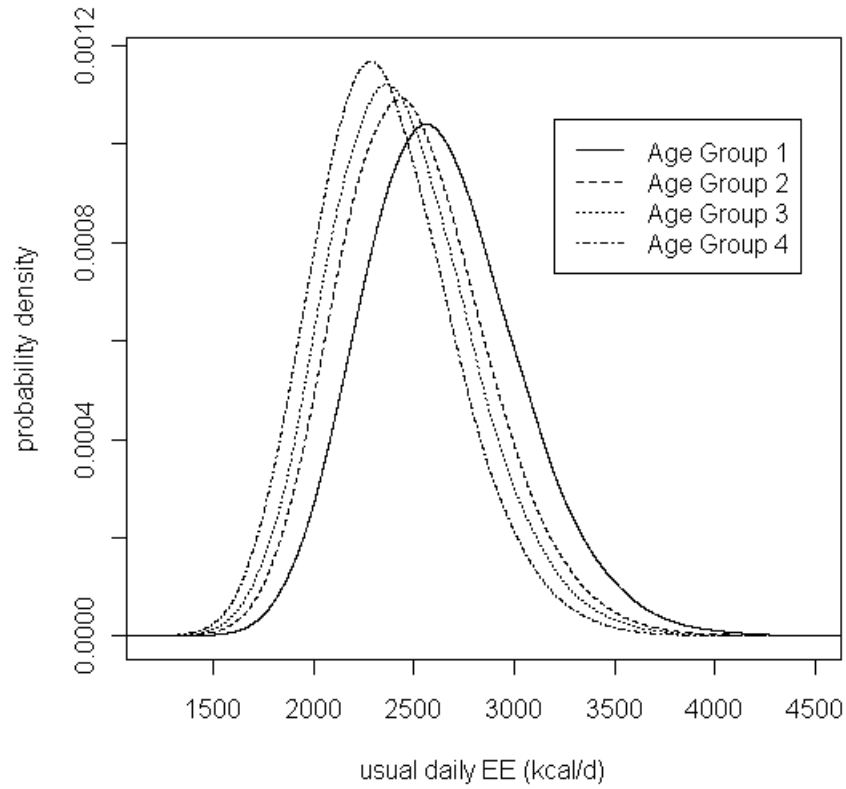


Figure 3.3 Estimated densities of usual daily EE for age groups 1 - 4

and the proportion of individuals with more than 3250 kcal/d of usual daily EE ($p_{high,g}$). The cutoff values of 1750 kcal/d and 3250 kcal/d were chosen for illustrative purposes and do not necessarily have any significance from a public health perspective. The estimates are given in Table 3.11 for each of the age groups. Standard errors of the estimates are computed using a delete-1 jackknife, which is described in Section 3.2.4. For each individual i , a replicate set of $m = 100,000$ original-scale usual daily EE values is generated and estimates of $\bar{T}_{g.}$, S_{Tg} , $p_{low,g}$, and $p_{high,g}$ are computed. Then standard errors are computed using the replicate jackknife estimates. The means of usual daily EE in the original scale decline by age group. Similarly, there is a decrease in standard deviations of usual daily EE in the original scale across age group. The tail estimates of $p_{low,g}$ and $p_{high,g}$ also reflect the decrease in usual daily EE across age group.

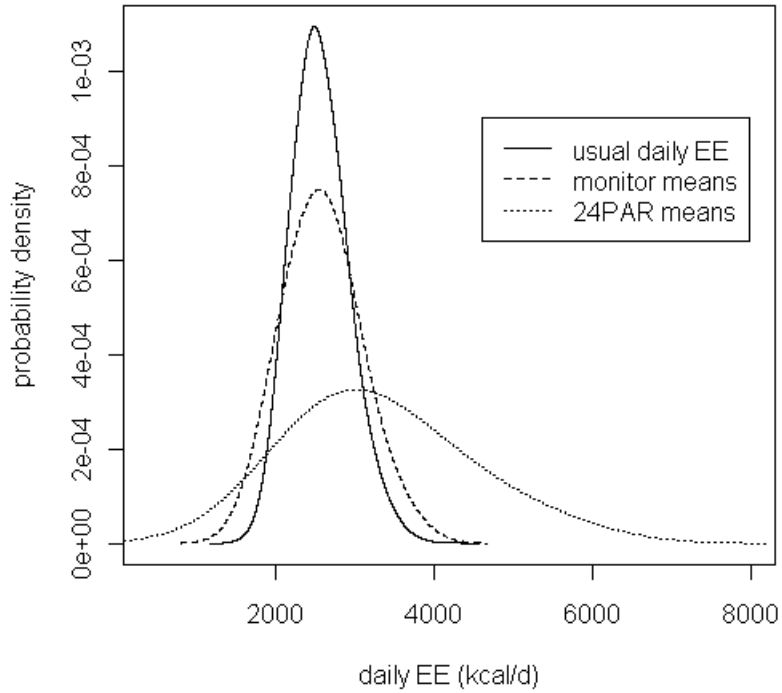


Figure 3.4 Estimated densities of usual daily EE, individual means of daily monitor EE, and individual means of daily 24PAR EE for age group 1 (age < 43)

3.4 Discussion

In this chapter, we have presented a method for estimating usual daily EE parameters, where daily EE measurements are adjusted for measurement error and nuisance effects using measurement error models. Our method is an extension of existing methods proposed in the literature for estimating usual physical activity parameters (Ferrari et al. 2007) and usual intake parameters (Nusser et al. 1996; Kipnis et al. 2003). A useful feature of our analysis is estimation of usual daily EE parameters for groups of the population. To implement our method, multiple concurrent measurements of daily EE must be available from an unbiased reference instrument, such as a multi-sensor monitoring device, and a self-report instrument, such as a 24-hour recall. The reference instrument is assumed to give unbiased measurements

Table 3.11 Parameter estimates (standard errors) for the group mean of usual daily EE in the original scale (\bar{T}_g), the group standard deviation of usual daily EE in the original scale (S_{Tg}), the group proportion of individuals with less than 1750 kcal/d of usual daily EE ($p_{low,g}$), and the group proportion of individuals with more than 3250 kcal/d of usual daily EE ($p_{high,g}$)

	\bar{T}_g	S_{Tg}	$100p_{low,g}$	$100p_{high,g}$
Age Group	Est (SE)	Est (SE)	Est (SE)	Est (SE)
1	2640 (68)	391 (46)	0.31 (0.32)	6.85 (3.51)
2	2503 (41)	368 (41)	0.87 (0.66)	3.16 (1.75)
3	2435 (45)	360 (41)	1.46 (0.99)	2.11 (1.34)
4	2355 (62)	348 (37)	2.51 (1.76)	1.20 (0.88)

of usual daily EE for model identification purposes.

A number of interesting points were identified by the analysis of the PAMS data in Section 3.3. An important amount of the variation in daily EE measured from the 24PAR is due to individual-level reporting biases. Almost half of the variation in the 24PAR EE data is due to individual-level reporting bias given the results from Table 3.9. Hence, individuals tend to misreport on their daily EE from the previous day, which could be due to cognitive limitations associated with recalling activity from the past (Matthews 2002). Researchers should use caution when making inferences on self-reported EE data because of the potential for bias and excess variation in the data.

The results from the female PAMS sample also suggest that the within-individual variation in daily EE is small relative to the inter-individual variation in usual daily EE. In Table 3.9, the estimated usual daily EE variance in the normal scale ($100\sigma_t^2$) is about 2 and the estimated within-individual variance of daily EE in the normal scale ($100\sigma_d^2$) is about 0.5. Hence, the inter-individual variation in usual daily EE is about four times larger than the within-individual variation in daily EE. This result is contrary to results from the dietary intake literature, which indicate that there is much more within-individual variation in dietary intake than there is inter-individual variation (Nusser et al. 1996; Carriquiry 2003).

In our analyses, there was evidence of a decrease in mean usual daily EE as age increases. The youngest age group (age 21 - 42) had the largest estimated mean of usual daily EE, while

the oldest age group (age 60 - 70) had the smallest estimated mean of usual daily EE. Similar results are given in Ferrari et al. (2007), which show lower levels of estimated EE in older age groups relative to younger age groups. The estimated slope parameters, which compare usual daily EE to reported daily EE in the groups, also decreased with age. The more active females in the youngest age group tend to have larger discrepancies between their usual daily EE and reported daily EE, while the more active females in the oldest age group tend to have smaller discrepancies between their usual daily EE and reported daily EE (Figure 3.2).

A long-term goal of our research is to use estimated usual daily EE distributions to estimate usual daily EE parameters in the original scale and to infer about EE behaviors of individuals in the population. The analyses we have presented here offer an example of what might be done to estimate usual daily EE parameters in the original scale (Table 3.11). Future work should involve a more thorough development of the methodology we have considered in this paper using EE data from a larger sample of the population.

APPENDIX A CONSISTENCY OF AN EGLS REGRESSION ESTIMATOR FOR A STRATIFIED CLUSTER DESIGN

Theorem 1 below contains conditions for the consistency of an estimated generalized least squares (EGLS) regression estimator for a stratified two-stage cluster sampling design. The conditions are developed to account for estimation of a regression estimator using data from NHANES, which is considered in Chapter 2.

For the NHANES sample design, clusters or primary sampling units (hereafter referred to as clusters) are selected from geographic strata that are subdivisions of the United States, and individuals are then selected from within the clusters through a multi-stage selection process. To account for this design in the theorem, we assume that stratified finite populations are realizations from a stratified infinite superpopulation with a fixed number of strata and that clusters in the finite population strata are independent realizations from the infinite superpopulation strata. In multi-stage sample designs like the NHANES design, regression error terms are often correlated within clusters. In the theorem, we allow for the error terms from model (A.2) to be correlated within clusters.

In the proof of Theorem 1 we rely on Corollary 5.1.1.2 in Fuller (1976), which states that

$$X_n = O_p(a_n)$$

for a sequence of random variables, $\{X_n\}$, and a sequence of numbers, $\{a_n\}$, that satisfy

$$E\{(X_n - E\{X_n\})^2\} = O(a_n^2).$$

This corollary is a consequence of Chebyshev's inequality.

Theorem 1

Let $\{\mathcal{F}_r\}$ be a sequence of stratified populations each with H strata. Let the finite population in stratum h of the r th stratified population be a realization of N_{rh} clusters from the infinite superpopulation, where $N_{rh} \geq N_{r-1,h}$ and

$$N_r = \sum_{h=1}^H N_{rh}$$

is the total number of clusters in the r th finite population. Let $N_r \rightarrow \infty$ as $r \rightarrow \infty$ and

$$\lim_{r \rightarrow \infty} N_{rh}/N_r = c_h, \quad (\text{A.1})$$

where $0 < c_h \leq 1$ for all h . Let $\mathbf{z}_{rhij} = (y_{rhij}, \mathbf{x}'_{rhij})'$ be a $(p+1)$ -dimensional random vector associated with element j , $j = 1, \dots, M_{rhi}$, in cluster i , $i = 1, \dots, N_{rh}$, in stratum h , $h = 1, \dots, H$, of the r th population, where M_{rhi} is the total number of elements in cluster rhi and $M_{rhi} \geq 2$ for all r , h , and i . Let the vector of cluster totals

$$\mathbf{z}_{rhi.} = \sum_{j=1}^{M_{rhi}} \mathbf{z}_{rhij},$$

have absolute $4 + \delta$ moments, for $\delta > 0$, and be independent with mean $\boldsymbol{\mu}_{rh}$ and covariance matrix $\boldsymbol{\Sigma}_{rh}$ for all r , h and i . Let y_{rhij} and \mathbf{x}_{rhij} in \mathbf{z}_{rhij} be related through the model

$$y_{rhij} = \mathbf{x}'_{rhij} \boldsymbol{\beta} + e_{rhij}, \quad (\text{A.2})$$

where $\boldsymbol{\beta}$ is a p -dimensional vector of regression coefficients, $e_{rhij} \sim (0, v_{rhij})$ is independent of $\mathbf{x}_{rh'i'j'}$ for all h, h', i, i', j , and j' , e_{rhij} is independent of $e_{rh'i'j'}$ when $hi \neq h'i'$, and

$$0 < M_{v1} < v_{rhij} < M_{v2} < \infty, \quad (\text{A.3})$$

for positive constants M_{v1} and M_{v2} and all r , h , i , and j .

Let a stratified simple random sample of clusters be selected from the r th finite population, where n_{rh} clusters are selected from stratum rh , $n_{rh} \geq 2$, $n_{rh} \geq n_{r-1,h}$, and

$$n_r = \sum_{h=1}^H n_{rh}$$

is the total number of clusters in the sample. Let $n_r \rightarrow \infty$ as $r \rightarrow \infty$,

$$\lim_{r \rightarrow \infty} n_{rh}/n_r = c_h \quad (\text{A.4})$$

for $h = 1, \dots, H$, and

$$\lim_{r \rightarrow \infty} n_r / N_r = f, \quad (\text{A.5})$$

where $0 < f \leq 1$. Also, let m_{rhi} elements be selected from cluster rhi in the sample, where m_{rhi} is the smallest integer greater than or equal to $g_{rhi}M_{rhi}$ such that $0 < g_{rhi} \leq 1$ for all r , h , and i .

Let $\mathbf{q}_{rhij} = \mathbf{x}_{rhij} v_{rhij}^{-1} \mathbf{x}_{rhij}'$ and let

$$\begin{aligned} \hat{\mathbf{M}}_{rq} &= n_r^{-1} \sum_{h=1}^H \sum_{i=1}^{n_{rh}} \sum_{j=1}^{m_{rhi}} \mathbf{q}_{rhij}, \\ \mathbf{M}_{rq,N} &= N_r^{-1} \sum_{h=1}^H \sum_{i=1}^{N_{rh}} \sum_{j=1}^{M_{rhi}} \mathbf{q}_{rhij}, \end{aligned}$$

and

$$\mathbf{M}_q = E\{\mathbf{M}_{rq,N}\}.$$

Similarly, let $\mathbf{u}_{rhij} = \mathbf{x}_{rhij} v_{rhij}^{-1} e_{rhij}$ and let

$$\begin{aligned} \hat{\mathbf{M}}_{ru} &= n_r^{-1} \sum_{h=1}^H \sum_{i=1}^{n_{rh}} \sum_{j=1}^{m_{rhi}} \mathbf{u}_{rhij}, \\ \mathbf{M}_{ru,N} &= N_r^{-1} \sum_{h=1}^H \sum_{i=1}^{N_{rh}} \sum_{j=1}^{M_{rhi}} \mathbf{u}_{rhij}, \end{aligned}$$

and

$$\mathbf{M}_u = E\{\mathbf{M}_{ru,N}\}.$$

Let \hat{v}_{rhij} be an estimator of v_{rhij} for element rhi , which satisfies

$$\hat{\mathbf{M}}_{r\hat{q}} - \hat{\mathbf{M}}_{rq} = O_p(n_r^{-1/2}) \quad (\text{A.6})$$

and

$$\hat{\mathbf{M}}_{r\hat{u}} - \hat{\mathbf{M}}_{ru} = O_p(n_r^{-1/2}), \quad (\text{A.7})$$

where

$$\begin{aligned}\hat{\mathbf{M}}_{r\hat{q}} &= n_r^{-1} \sum_{h=1}^H \sum_{i=1}^{n_{rh}} \sum_{j=1}^{m_{rhi}} \hat{\mathbf{q}}_{rhij}, \\ \hat{\mathbf{M}}_{r\hat{u}} &= n_r^{-1} \sum_{h=1}^H \sum_{i=1}^{n_{rh}} \sum_{j=1}^{m_{rhi}} \hat{\mathbf{u}}_{rhij},\end{aligned}$$

$\hat{\mathbf{q}}_{rhij} = \mathbf{x}_{rhij} \hat{v}_{rhij}^{-1} \mathbf{x}'_{rhij}$, and $\hat{\mathbf{u}}_{rhij} = \mathbf{x}_{rhij} \hat{v}_{rhij}^{-1} e_{rhij}$. Let

$$\hat{\beta}_{EGLS} = \left(\sum_{h=1}^H \sum_{i=1}^{n_{rh}} \sum_{j=1}^{m_{rhi}} \hat{\mathbf{q}}_{rhij} \right)^{-1} \sum_{h=1}^H \sum_{i=1}^{n_{rh}} \sum_{j=1}^{m_{rhi}} \mathbf{x}_{rhij} \hat{v}_{rhij}^{-1} y_{rhij}. \quad (\text{A.8})$$

Then $\hat{\beta}_{EGLS} - \beta = O_p(n_r^{-1/2})$.

Proof

Since $\mathbf{q}_{rhij} = \mathbf{x}_{rhij} v_{rhij}^{-1} \mathbf{x}'_{rhij}$, by the assumption that \mathbf{x}_{rhij} in \mathbf{z}_{rhij} has finite $4 + \delta$ moments and assumption (A.3), it follows that

$$\mathbf{q}_{rhi.} = \sum_{j=1}^{m_{rhi}} \mathbf{q}_{rhij},$$

has finite $2 + \delta$ moments and that the $\mathbf{q}_{rhi.}$ are independent and share a common covariance matrix $\Sigma_{q,rh}$ for all i in stratum rh . Let $\sigma_{q,rh,kl}$ be the kl th element of $\Sigma_{q,rh}$ for $k, l = 1, \dots, p$. Then, the kl th element of $\text{Var}\{\mathbf{M}_{rq,N} - \mathbf{M}_q\}$ is

$$\begin{aligned}\text{Var}\{\mathbf{M}_{rq,N} - \mathbf{M}_q\}_{kl} &= N_r^{-2} \sum_{h=1}^H \sum_{i=1}^{N_{rh}} \sigma_{q,rh,kl} \\ &= N_r^{-2} \sum_{h=1}^H N_{rh} \sigma_{q,rh,kl} \\ &= N_r^{-1} \sum_{h=1}^H (N_{rh}/N_r) \sigma_{q,rh,kl} \\ &= O(N_r^{-1}),\end{aligned}$$

for all k and l by assumption (A.1). By Corollary 5.1.1.2 in Fuller (1976), it follows that

$$\mathbf{M}_{rq,N} - \mathbf{M}_q = O_p(N_r^{-1/2}).$$

Since $\mathbf{u}_{rhij} = \mathbf{x}_{rhij} v_{rhij}^{-1} e_{rhij}$, by the model assumptions of (A.2) and by assumption (A.3), similar arguments can be used to show that

$$\mathbf{M}_{ru,N} - \mathbf{M}_u = O_p(N_r^{-1/2}).$$

Since $\mathbf{q}_{rhi.}$ has finite $2 + \delta$ moments and the $\mathbf{q}_{rhi.}$ are independent with common covariance matrix $\Sigma_{q,rh}$, it follows that the partial sum of $\mathbf{q}_{rhi.}$,

$$\mathbf{q}_{m,rhi.} = \sum_{j=1}^{m_{rhi}} \mathbf{q}_{rhij},$$

has finite $2 + \delta$ moments and the $\mathbf{q}_{m,rhi.}$ are independent. Let

$$M_{qm,rhi,kl} = \max\{Var\{\mathbf{q}_{m,rhi.}\}_{kl}, Var\{\mathbf{q}_{rhi.}\}_{kl}, |Cov\{\mathbf{q}_{m,rhi.}, \mathbf{q}_{rhi.}\}_{kl}|\},$$

where $Var\{\mathbf{q}_{m,rhi.}\}_{kl}$ is the kl th element of the variance of $\mathbf{q}_{m,rhi.}$, $Var\{\mathbf{q}_{rhi.}\}_{kl}$ is the kl th element of the variance of $\mathbf{q}_{rhi.}$, and $Cov\{\mathbf{q}_{m,rhi.}, \mathbf{q}_{rhi.}\}_{kl}$ is the kl th element of the covariance of $\mathbf{q}_{m,rhi.}$ and $\mathbf{q}_{rhi.}$. Let

$$M_{qm,rh,kl} = \max_i \{M_{qm,rhi,kl}\}.$$

Then, the kl th element of $Var\{\hat{\mathbf{M}}_{rq} - \mathbf{M}_{rq,N}\}$ is

$$\begin{aligned} Var\{\hat{\mathbf{M}}_{rq} - \mathbf{M}_{rq,N}\}_{kl} &= n_r^{-2} \sum_{h=1}^H \sum_{i=1}^{n_{rh}} Var\{\mathbf{q}_{m,rhi.}\}_{kl} + N_r^{-2} \sum_{h=1}^H \sum_{i=1}^{N_{rh}} Var\{\mathbf{q}_{rhi.}\}_{kl} \\ &\quad - 2n_r^{-1} N_r^{-1} \sum_{h=1}^H \sum_{i=1}^{n_{rh}} Cov\{\mathbf{q}_{m,rhi.}, \mathbf{q}_{rhi.}\}_{kl} \\ &\leq n_r^{-2} \sum_{h=1}^H \sum_{i=1}^{n_{rh}} M_{qm,rh,kl} + N_r^{-2} \sum_{h=1}^H \sum_{i=1}^{N_{rh}} M_{qm,rh,kl} \\ &\quad + 2n_r^{-1} N_r^{-1} \sum_{h=1}^H \sum_{i=1}^{n_{rh}} M_{qm,rh,kl} \\ &= n_r^{-1} \sum_{h=1}^H (n_{rh}/n_r) M_{qm,rh,kl} + N_r \sum_{h=1}^H (N_{rh}/N_r) M_{qm,rh,kl} \\ &\quad + 2N_r^{-1} \sum_{h=1}^H (n_{rh}/n_r) M_{qm,rh,kl} \\ &= O(n_r^{-1}) + O(N_r^{-1}) + O(N_r^{-1}) \\ &= O(n_r^{-1}) \end{aligned}$$

by assumptions (A.1), (A.4), and (A.5). By Corollary 5.1.1.2 in Fuller (1976), it follows that

$$\hat{\mathbf{M}}_{rq} - \mathbf{M}_{rq,N} = O_p(n_r^{-1/2}).$$

By similar arguments,

$$\hat{\mathbf{M}}_{ru} - \mathbf{M}_{ru,N} = O_p(n_r^{-1/2}).$$

It follows that,

$$\begin{aligned} \hat{\mathbf{M}}_{rq} - \mathbf{M}_q &= \hat{\mathbf{M}}_{rq} - \mathbf{M}_{rq,N} + \mathbf{M}_{rq,N} - \mathbf{M}_q \\ &= O_p(n_r^{-1/2}) + O_p(N_r^{-1/2}) \\ &= O_p(n_r^{-1/2}) \end{aligned}$$

and

$$\begin{aligned} \hat{\mathbf{M}}_{ru} - \mathbf{M}_u &= \hat{\mathbf{M}}_{ru} - \mathbf{M}_{ru,N} + \mathbf{M}_{ru,N} - \mathbf{M}_u \\ &= O_p(n_r^{-1/2}) + O_p(N_r^{-1/2}) \\ &= O_p(n_r^{-1/2}), \end{aligned}$$

by assumption (A.5). Also,

$$\begin{aligned} \hat{\mathbf{M}}_{r\hat{q}} - \mathbf{M}_q &= \hat{\mathbf{M}}_{r\hat{q}} - \hat{\mathbf{M}}_{rq} + \hat{\mathbf{M}}_{rq} - \mathbf{M}_q \\ &= O_p(n_r^{-1/2}) + O_p(n_r^{-1/2}) \\ &= O_p(n_r^{-1/2}), \end{aligned}$$

by assumption (A.6). By a Taylor expansion,

$$\begin{aligned} \hat{\mathbf{M}}_{r\hat{q}}^{-1} &= \mathbf{M}_q^{-1} + (\hat{\mathbf{M}}_{r\hat{q}} - \mathbf{M}_q) \mathbf{h}'(\mathbf{M}_{rq}^*) \\ &= \mathbf{M}_q^{-1} + O_p(n_r^{-1/2}), \end{aligned}$$

where \mathbf{M}_{rq}^* is on the line segment joining $\hat{\mathbf{M}}_{r\hat{q}}$ and \mathbf{M}_q and $\mathbf{h}(\mathbf{M}_{rq}^*)$ is the vector of derivatives of $\hat{\mathbf{M}}_{r\hat{q}}^{-1}$ with respect to the elements in $\hat{\mathbf{M}}_{r\hat{q}}$ evaluated at \mathbf{M}_{rq}^* . It follows that

$$\hat{\mathbf{M}}_{r\hat{q}}^{-1} - \mathbf{M}_q^{-1} = O_p(n_r^{-1/2}). \quad (\text{A.9})$$

By the assumptions of model (A.2),

$$\begin{aligned} E\{\mathbf{u}_{rhij}\} &= E\{\mathbf{x}_{rhij}v_{rhij}^{-1}e_{rhij}\} \\ &= 0 \end{aligned}$$

for all r, h, i , and j , and it follows that

$$\mathbf{M}_u = \mathbf{0},$$

since the \mathbf{u}_{rhi} are independent for all r, h , and i . Then,

$$\begin{aligned} \hat{\mathbf{M}}_{ru} &= \mathbf{M}_u + O_p(n_r^{-1/2}) \\ &= O_p(n_r^{-1/2}), \end{aligned}$$

and by assumption (A.7),

$$\begin{aligned} \hat{\mathbf{M}}_{r\hat{u}} &= \hat{\mathbf{M}}_{ru} + O_p(n_r^{-1/2}) \\ &= O_p(n_r^{-1/2}). \end{aligned} \tag{A.10}$$

Thus, for the EGLS estimator in (A.8),

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{EGLS} - \boldsymbol{\beta} &= \left(\sum_{h=1}^H \sum_{i=1}^{n_{rh}} \sum_{j=1}^{m_{rhi}} \hat{\mathbf{q}}_{rhij} \right)^{-1} \sum_{h=1}^H \sum_{i=1}^{n_{rh}} \sum_{j=1}^{m_{rhi}} \mathbf{x}_{rhij} \hat{v}_{rhij}^{-1} y_{rhij} - \boldsymbol{\beta} \\ &= \left(\sum_{h=1}^H \sum_{i=1}^{n_{rh}} \sum_{j=1}^{m_{rhi}} \hat{\mathbf{q}}_{rhij} \right)^{-1} \sum_{h=1}^H \sum_{i=1}^{n_{rh}} \sum_{j=1}^{m_{rhi}} \mathbf{x}_{rhij} \hat{v}_{rhij}^{-1} (y_{rhij} - \mathbf{x}'_{rhij} \boldsymbol{\beta}) \\ &= \left(\sum_{h=1}^H \sum_{i=1}^{n_{rh}} \sum_{j=1}^{m_{rhi}} \hat{\mathbf{q}}_{rhij} \right)^{-1} \sum_{h=1}^H \sum_{i=1}^{n_{rh}} \sum_{j=1}^{m_{rhi}} \mathbf{x}_{rhij} \hat{v}_{rhij}^{-1} e_{rhij} \\ &= \left(n_r^{-1} \sum_{h=1}^H \sum_{i=1}^{n_{rh}} \sum_{j=1}^{m_{rhi}} \hat{\mathbf{q}}_{rhij} \right)^{-1} n_r^{-1} \sum_{h=1}^H \sum_{i=1}^{n_{rh}} \sum_{j=1}^{m_{rhi}} \mathbf{x}_{rhij} \hat{v}_{rhij}^{-1} e_{rhij} \\ &= \hat{\mathbf{M}}_{r\hat{q}}^{-1} \hat{\mathbf{M}}_{r\hat{u}} \\ &= O_p(n_r^{-1/2}), \end{aligned}$$

by (A.9) and (A.10), and the proof is complete.

Comment

For assumptions (A.6) and (A.7) to hold, a sufficient condition is that

$$\hat{v}_{rhij} = v_{rhij} + O_p(n_r^{-1/2}). \quad (\text{A.11})$$

In practice, \hat{v}_{rhij} is derived by fitting model (A.2) using ordinary least squares and using the squared residuals from the model fit to estimate the parameter vector $\boldsymbol{\eta}$ in the variance model

$$v_{rhij} = v(\boldsymbol{x}_{rhij}, \boldsymbol{\eta}),$$

where v is a known, continuous function. For this procedure to satisfy (A.11), the estimator of $\boldsymbol{\eta}$ must be consistent for $\boldsymbol{\eta}$ at most of order $n_r^{-1/2}$. See Lemma 5.7.1 of Fuller (1976).

APPENDIX B REGRESSION ESTIMATION FOR STRATIFIED CLUSTER DESIGN

In this appendix we present regression estimators, variances of the estimators, and a test statistic for comparing the weighted and unweighted estimators for a regression model. The regression model is presented in Chapter 2 for the NHANES stratified cluster design. We consider the model set up from Theorem 1 in Appendix A. To reduce notational complexity we ignore the index r representing the r th finite population generated from the infinite superpopulation.

Consider the linear regression model (A.2),

$$y_{hij} = \mathbf{x}'_{hij}\boldsymbol{\beta} + e_{hij}, \quad (\text{B.1})$$

where $\boldsymbol{\beta}$ is a p -dimensional vector of unknown regression coefficients, y_{hij} is average daily accelerometer MVPA, \mathbf{x}_{hij} is a p -dimensional function of covariates, and $e_{hij} \sim (0, v_{hij})$ with a positive finite variance v_{hij} for individual j , $j = 1, \dots, m_{hi}$, in PSU i , $i = 1, \dots, n_h$, in stratum h , $h = 1, \dots, H$. Assume that e_{hij} is independent of $\mathbf{x}_{h'i'j'}$ for all h, h', i, i', j , and j' , and that e_{hij} is independent of $e_{h'i'j'}$ when $hi \neq h'i'$. The ordinary least squares (OLS) estimator of $\boldsymbol{\beta}$ for (B.1) is

$$\hat{\boldsymbol{\beta}}_{OLS} = \left(\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \mathbf{x}_{hij} \mathbf{x}'_{hij} \right)^{-1} \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \mathbf{x}_{hij} y_{hij}, \quad (\text{B.2})$$

and the weighted least squares (WLS) estimator of $\boldsymbol{\beta}$ for (B.1) is

$$\hat{\boldsymbol{\beta}}_{WLS} = \left(\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \mathbf{x}_{hij} w_{hij} \mathbf{x}'_{hij} \right)^{-1} \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \mathbf{x}_{hij} w_{hij} y_{hij}, \quad (\text{B.3})$$

where w_{hij} is the survey weight for individual j in PSU hi . To take into account the sample design, the variances of (B.2) and (B.3) can be estimated using the Taylor linearization form

available in SAS and STATA and given in Fuller (1984). The estimated Taylor linearization variance of $\hat{\beta}_{OLS}$ is

$$\hat{V}(\hat{\beta}_{OLS}) = \left(\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \mathbf{x}_{hij} \mathbf{x}'_{hij} \right)^{-1} \hat{\mathbf{G}}_{OLS} \left(\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \mathbf{x}_{hij} \mathbf{x}'_{hij} \right)^{-1}, \quad (\text{B.4})$$

where

$$\begin{aligned} \hat{\mathbf{G}}_{OLS} &= \frac{n-1}{n-p} \sum_{h=1}^H \frac{n_h}{n_h-1} \sum_{i=1}^{n_h} (\mathbf{r}_{hi.} - \bar{\mathbf{r}}_{h..})(\mathbf{r}_{hi.} - \bar{\mathbf{r}}_{h..})', \\ \mathbf{r}_{hij} &= \mathbf{x}_{hij} \hat{e}_{OLS,hij}, \\ \hat{e}_{OLS,hij} &= y_{hij} - \mathbf{x}'_{hij} \hat{\beta}_{OLS}, \\ \mathbf{r}_{hi.} &= \sum_{j=1}^{m_{hi}} \mathbf{r}_{hij}, \\ \bar{\mathbf{r}}_{h..} &= n_h^{-1} \sum_{i=1}^{n_h} \mathbf{r}_{hi.}, \end{aligned} \quad (\text{B.5})$$

n is the total number of elements in the sample, and p is the dimension of β . The factor $(n-1)/(n-p)$ is a variance adjustment term used to reduce small sample bias (Hidiroglou, Fuller, and Hickman 1980). The form of the estimated variance (B.4) is appropriate given the model assumptions that the error terms are uncorrelated across PSUs and that the PSUs are simple random samples from the strata. The estimated Taylor linearization variance of $\hat{\beta}_{WLS}$ is

$$\hat{V}(\hat{\beta}_{WLS}) = \left(\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \mathbf{x}_{hij} w_{hij} \mathbf{x}'_{hij} \right)^{-1} \hat{\mathbf{G}}_{WLS} \left(\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \mathbf{x}_{hij} w_{hij} \mathbf{x}'_{hij} \right)^{-1},$$

where

$$\begin{aligned} \hat{\mathbf{G}}_{WLS} &= \frac{n-1}{n-p} \sum_{h=1}^H \frac{n_h}{n_h-1} \sum_{i=1}^{n_h} (\mathbf{s}_{hi.} - \bar{\mathbf{s}}_{h..})(\mathbf{s}_{hi.} - \bar{\mathbf{s}}_{h..})', \\ \mathbf{s}_{hij} &= \mathbf{x}_{hij} w_{hij} \hat{e}_{WLS,hij}, \\ \hat{e}_{WLS,hij} &= y_{hij} - \mathbf{x}'_{hij} \hat{\beta}_{WLS}, \\ \mathbf{s}_{hi.} &= \sum_{j=1}^{m_{hi}} \mathbf{s}_{hij}, \end{aligned}$$

and

$$\bar{\mathbf{s}}_{h..} = n_h^{-1} \sum_{i=1}^{n_h} \mathbf{s}_{hi.}.$$

Given regularity conditions, the WLS estimator is asymptotically unbiased for β , while the OLS estimator can be biased for β if the model error terms e_{hij} are correlated with the survey weights w_{hij} (Fuller 2009, page 350-1). To test for the bias in the OLS estimator, one can test for $\gamma = \mathbf{0}$ in the extended model

$$y_{hij} = \mathbf{x}'_{hij}\beta + w_{hij}^*\mathbf{x}'_{hij}\gamma + a_{hij},$$

where a_{hij} is equal to e_{hij} if $\gamma = \mathbf{0}$,

$$w_{hij}^* = \frac{w_{hij} - \bar{w}}{\bar{w}},$$

and \bar{w} is the mean of the survey weights (Fuller 2009, page 352). Define $\mathbf{z}_{hij} = (\mathbf{x}'_{hij}, w_{hij}^*\mathbf{x}'_{hij})'$ and $\boldsymbol{\theta} = (\beta', \gamma')'$. The OLS estimator of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_{OLS}$, is given by (B.2), with \mathbf{z}_{hij} replacing \mathbf{x}_{hij} . The estimated Taylor linearization variance of $\hat{\boldsymbol{\theta}}_{OLS}$, $\hat{V}(\hat{\boldsymbol{\theta}}_{OLS})$, is given by (B.4), with \mathbf{z}_{hij} replacing \mathbf{x}_{hij} and $\mathbf{z}_{hij}\hat{a}_{OLS,hij}$ replacing \mathbf{r}_{hij} in (B.5), where $\hat{a}_{OLS,hij} = y_{hij} - \mathbf{z}'_{hij}\hat{\boldsymbol{\theta}}_{OLS}$.

The test statistic for $\gamma = \mathbf{0}$ is

$$F(p, m) = p^{-1}\hat{\boldsymbol{\theta}}_2'\hat{V}(\hat{\boldsymbol{\theta}}_{OLS})_{22}^{-1}\hat{\boldsymbol{\theta}}_2,$$

where $\hat{\boldsymbol{\theta}}_2$ is the lower p elements of $\hat{\boldsymbol{\theta}}_{OLS}$, $\hat{V}(\hat{\boldsymbol{\theta}}_{OLS})_{22}$ is the lower right $p \times p$ submatrix of $\hat{V}(\hat{\boldsymbol{\theta}}_{OLS})$, p is the dimension of γ , and m is the number of PSUs minus the number of strata for the sample. Under the null hypothesis that $\gamma = \mathbf{0}$, $F(p, m)$ is approximately distributed as an F with p and m degrees of freedom. This result follows from result (17) in Fuller (1984).

APPENDIX C TEST FOR WEIGHTED AND UNWEIGHTED ESTIMATORS FOR THE POPULATION-LEVEL MODEL

In this appendix, we give a test to compare equal-weight estimators and survey-weighted estimators for parameters of the population-level model given in Section 3.2.3 of Chapter 3. In the first part of the appendix, we develop the test procedure. In the second part of the appendix, we give test results for the population-level model given in Section 3 of Chapter 3, which was estimated using the preliminary first quarter sample of females from PAMS.

Test Procedure

Consider the method developed in Section 3.2 of Chapter 3 and assume that a complex sample design is used to select n individuals into the sample. For the population-level model given by (3.20) in Chapter 3, let $\hat{\boldsymbol{\lambda}}_1$ be the EGLS estimator of the population-level model parameter vector $\boldsymbol{\lambda}$ given by (3.22) in Chapter 3 and computed using equal weights (i.e., $w_i = 1$ for all i), and let $\hat{\boldsymbol{\lambda}}_2$ be the EGLS estimator of $\boldsymbol{\lambda}$ computed using survey weights. We consider a test for $E\{\hat{\boldsymbol{\lambda}}_1 - \hat{\boldsymbol{\lambda}}_2\} = \mathbf{0}$ to determine if the results are similar for equal weights and survey weights. A test statistic to test for $E\{\hat{\boldsymbol{\lambda}}_1 - \hat{\boldsymbol{\lambda}}_2\} = \mathbf{0}$ is

$$F = (p)^{-1}(\hat{\boldsymbol{\lambda}}_1 - \hat{\boldsymbol{\lambda}}_2)'[\hat{\mathbf{V}}_{11} + \hat{\mathbf{V}}_{22} - \hat{\mathbf{V}}_{12} - \hat{\mathbf{V}}_{21}]^{-1}(\hat{\boldsymbol{\lambda}}_1 - \hat{\boldsymbol{\lambda}}_2), \quad (\text{C.1})$$

where

$$\hat{\mathbf{V}}\{\hat{\boldsymbol{\lambda}}_{full}\} = \begin{pmatrix} \hat{\mathbf{V}}_{11} & \hat{\mathbf{V}}_{12} \\ \hat{\mathbf{V}}_{21} & \hat{\mathbf{V}}_{22} \end{pmatrix}$$

is an estimated variance of

$$\hat{\boldsymbol{\lambda}}_{full} = (\hat{\boldsymbol{\lambda}}_1', \hat{\boldsymbol{\lambda}}_2')'$$

and p is the dimension of $\boldsymbol{\lambda}$. Under the null hypothesis that $E\{\hat{\boldsymbol{\lambda}}_1 - \hat{\boldsymbol{\lambda}}_2\} = \mathbf{0}$ and given regularity conditions, the F statistic in (C.1) is approximately distributed as an F distribution with p and $n - 2p$ degrees of freedom.

The variance required for the test can be estimated using jackknife variance estimation. For a stratified sample design, let H be the number of strata, let n_h be the number of individuals sampled from stratum h , and let N_h be the total number of individuals in stratum h , for $h = 1, \dots, H$. The jackknife variance is estimated by computing n replicate estimators of $\hat{\boldsymbol{\lambda}}_{full}$. Let

$$\hat{\boldsymbol{\lambda}}_{full}^{(hi)} = \begin{pmatrix} \hat{\boldsymbol{\lambda}}_1^{(hi)} \\ \hat{\boldsymbol{\lambda}}_2^{(hi)} \end{pmatrix}$$

be the h th replicate estimator of $\hat{\boldsymbol{\lambda}}_{full}$, where $\hat{\boldsymbol{\lambda}}_1^{(hi)}$ is the h th replicate of $\hat{\boldsymbol{\lambda}}_1$ and $\hat{\boldsymbol{\lambda}}_2^{(hi)}$ is the h th replicate of $\hat{\boldsymbol{\lambda}}_2$. The estimator $\hat{\boldsymbol{\lambda}}_2^{(hi)}$ is computed using the h th set of replicate survey weights defined by deleting individual i in stratum h . The replicate survey weight of individual i' in stratum h' from the h th set of replicate weights is

$$w_{h'i'}^* = \begin{cases} 0 & \text{if } h' = h \text{ and } i' = i \\ (c_1/c_2)w_{h'i'} & \text{if } h' = h \text{ and } i' \neq i \\ w_{h'i'} & \text{if } h' \neq h \end{cases} \quad (\text{C.2})$$

where $w_{h'i'}$ is the original survey weight of individual i' in stratum h' ,

$$c_1 = \sum_{i'=1}^{n_{h'}} w_{h'i'},$$

and $c_2 = c_1 - w_{hi}$. The h th replicate equal weight of individual i' in stratum h' is given by setting $w_{h'i'} = 1$ in (C.2) for all $h'i' \neq hi$. The replicate estimators are computed just as the original estimators, but are computed with the replicate weights instead of the original weights. The estimated jackknife variance for the stratified design is

$$\hat{V}\{\hat{\boldsymbol{\lambda}}_{full}\} = \sum_{h=1}^H N_h^{-1}(N_h - n_h)n_h^{-1}(n_h - 1) \sum_{i=1}^{n_h} (\hat{\boldsymbol{\lambda}}_{full}^{(hi)} - \hat{\boldsymbol{\lambda}}_{full})(\hat{\boldsymbol{\lambda}}_{full}^{(hi)} - \hat{\boldsymbol{\lambda}}_{full})', \quad (\text{C.3})$$

where the multipliers $N_h^{-1}(N_h - n_h)n_h^{-1}(n_h - 1)$ are included to account for the sample selection within strata. See Section 4.2 of Fuller (2009).

Application to PAMS Data

The test described above is applied to the population-level model in Section 3 of Chapter 3 using the preliminary sample of females from PAMS. See Section 3.3 for a description of the sample and a description of the population-level model. Let $\hat{\lambda}_1$ be the estimated vector of model parameters for the population-level model computed with equal weights and let $\hat{\lambda}_2$ be the estimated vector of model parameters for the model computed with the survey weights for the PAMS sample. The estimated parameters are given in Table C.1. A jackknife variance for $\hat{\lambda}_{full} = (\hat{\lambda}'_1, \hat{\lambda}'_2)'$ is computed using equation (C.3), where the replicate jackknife estimators of $\hat{\lambda}'_1$ and $\hat{\lambda}'_2$ are computed for the stratified PAMS design. The standard errors from the jackknife variance are given in Table C.1.

Table C.1 Parameter estimates for $\hat{\lambda}_1$ and $\hat{\lambda}_2$

Parameter	Estimates (SEs) for $\hat{\lambda}_1$	Estimates (SEs) for $\hat{\lambda}_2$
μ_0	7.7940 (.0103)	7.7849 (.0145)
μ_y	8.0564 (.0144)	8.0477 (.0222)
100θ	-0.2409 (.1060)	-0.2467 (.1147)
β_1	0.9950 (.0914)	1.1310 (.1497)
$100\beta_3$	-1.7035 (.6221)	-2.2653 (.8841)
$100\sigma_t^2$	1.9868 (.2553)	1.5915 (.3291)
$100\sigma_d^2$	0.4949 (.1105)	0.6523 (.2672)
$100\sigma_u^2$	0.5186 (.1231)	0.5667 (.1989)
$100\sigma_e^2$	0.6175 (.1277)	0.5264 (.2870)
$100\sigma_r^2$	2.0934 (.3062)	2.0306 (.3643)

The values of n_h and N_h used to compute the jackknife variance are given in Table C.2. The F statistic given by (C.1) is computed to be 1.203 for the PAMS sample on 10 and 151 degrees of freedom with a p-value of 0.293. Given these results, there is little evidence suggesting that the equal-weight and survey-weighted estimators of the population-level model are different for the preliminary female PAMS sample.

Table C.2 Sample sizes and population control totals for females in the 8 PAMS strata (control totals are from the 2000 U.S. Census)

Stratum	Sample Size	Population Size
Black Hawk (low minority)	32	30543
Black Hawk (high minority)	24	7979
Dallas (low minority)	6	6914
Dallas (high minority)	26	5424
Marshall (low minority)	5	8178
Marshall (high minority)	10	3609
Polk (low minority)	27	95948
Polk (high minority)	41	20082

BIBLIOGRAPHY

- Adams, S.A., Matthews, C.E., Ebbeling, C.B., Moore, C.G., Cunningham, J.E., Fulton, J., Hebert, J.R. (2005). The effect of social desirability and social approval on self-reports of physical activity. *American Journal of Epidemiology*, **161**, 389-398.
- Ainsworth, B.E. (2009). How do I measure physical activity in my patients? Questionnaires and objective methods. *British Journal of Sports Medicine*, **43**, 6-9.
- Ainsworth, B.E., Haskell, W.L., Leon, A.S., Jacobs, D.R., Montoye, H.J., Sallis, J.F., Paffenbarger, R.S. (1993). Compendium of physical activities: classification of energy costs of human physical activities. *Medicine and Science in Sports and Exercise*, **25**, 71-80.
- Ainsworth, B.E., Haskell, W.L., Whitt, M.C., Irwin, M.L., Swartz, A.M., Strath, S.J., O'Brien, W.L., Bassett, D.R., Schmitz, K.H., Emplaincourt, P.O., Jacobs, D.R., Leon, A.S. (2000). Compendium of physical activities: an update of activity codes and MET intensities. *Medicine and Science in Sports and Exercise*, **32**, S498-S516.
- Bassett, D.R., Cureton, A.L., Ainsworth, B.E. (2000). Measurement of daily walking distance – questionnaire versus pedometer. *Medicine and Science in Sports and Exercise*, **32**, 1018-1023.
- Berlin, J.A. and Colditz, G.A. (1990). A meta-analysis of physical activity in the prevention of coronary heart disease. *American Journal of Epidemiology*, **132**, 612-628.
- Bouten, C., Verboeket-van de Venne, W., Westerterp, K.R., Verduin, M., Janssen, J.D. (1996). Daily physical activity assessment: comparison between movement registration and doubly labeled water. *Journal of Applied Physiology*, **81**, 1019-1026.

- Bratteby, L.-E., Sandhagen, B., Fan, H., Enghardt, H., Samuelson, G. (1998). Total energy expenditure and physical activity as assessed by the doubly labeled water method in Swedish adolescents in whom energy intake was underestimated by 7-d diet records. *American Journal of Clinical Nutrition*, **67**, 905-911.
- Calabro, M.A., Welk, G.J., Carriquiry, A.L., Nusser, S.M., Beyler, N.K., Matthews, C.E. (2009). Validation of a computerized 24-hour physical activity recall (24PAR) instrument with pattern-recognition activity monitors. *Journal of Physical Activity and Health*, **6**, 211-220.
- Carriquiry, A.L. (2003). Estimation of usual intake distributions of nutrients and foods. *Journal of Nutrition*, **133**, 601S-608S.
- Carroll, R.J., Ruppert, D., Stefanski, L.A., Crainiceanu, C.M. (2006), *Measurement Error in Nonlinear Models: A Modern Perspective*. Boca Raton, FL: Chapman and Hall.
- Casperson, 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**, 125-131.
- Crouter, S.E., Clowers, K.G., Bassett, D.R. (2006). A novel method for using accelerometer data to predict energy expenditure. *Journal of Applied Physiology*, **100**, 1324-1331.
- Dipietro, L., Caspersen, C.J., Ostfeld, A.M., Nadel, E.R. (1993). A survey for assessing physical activity among older adults. *Medicine and Science in Sports and Exercise*, **25**, 628-642.
- Dodd, K.W., Guenther, P.M., Freedman, L.S., Subar, A.F., Kipnis, V., Midthune, D., Tooze, J.A., Krebs-Smith, S.M. (2006) Statistical methods for estimating usual intake of nutrients and foods: a review of the theory. *Journal of the American Dietetic Association*, **106**, 1640-1650.
- Ferrari, P., Friedenreich, C., Matthews, C.E. (2007). The role of measurement error in estimating levels of physical activity. *American Journal of Epidemiology*, **166**, 832-840.

- Freedson, P.S., Melanson, E., Sirard, J. (1998). Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine and Science in Sports and Exercise*, **30**, 777-781.
- Friedenreich, C.E. and Orenstein, M.R. (2002). Physical activity and cancer prevention: etiology evidence and biological mechanisms. *Journal of Nutrition*, **132**, 3456S-3464S.
- Friedenreich, C.M., Courneya, K.S., Neilson, H.K., et al. (2006) Reliability and validity of the Past Year Total Physical Activity Questionnaire. *American Journal of Epidemiology*, **163**, 959-970.
- Fruin, M.L., Rankin, J.W. (2004). Validity of a multi-sensor armband in estimating rest and exercise energy expenditure. *Medicine and Science in Sports and Exercise*, **36**, 1063-1069.
- Fuller, W.A. (1976). *Introduction to Statistical Time Series*. New York: Wiley.
- Fuller, W.A. (1984). Least squares and related analyses for complex survey designs. *Survey Methodology*, **10**, 97-118.
- Fuller, W.A. (1987). *Measurement Error Models*. New York: Wiley.
- Fuller, W.A. (2009). *Sampling Statistics*. New York: Wiley.
- Grobbee, D.E., Rimm, E.B., Giovannucci, E., et al. (1990). Coffee, caffeine, and cardiovascular disease in men. *New England Journal of Medicine*, **323**, 1026-32.
- Grundy, S.M., Blackburn, G., Higgins, M., Lauer, R., Perri, M.G., Ryan, D. (1999). Physical activity in the prevention and treatment of obesity and its comorbidities. *Medicine and Science in Sports and Exercise*, **31**, S502.
- Hebert, J.R., Ebbeling, C.B., Matthews, C.E., Hurley, T.G., Ma, Y., Druker, S., Clemow, L. (2002). Systematic errors in middle-aged women's estimates of energy intake: comparing three self-report measures to total energy expenditure from doubly labeled water. *Annals of Epidemiology*, **12**, 577-586.

- Hendelman, D., Miller, K., Baggett, C., Debold, E., Freedson, P. (2000). Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Medicine and Science in Sports and Exercise*, **32**, S442-S449.
- Hidioglou, M.A., Fuller, W.A., and Hickman, R.D. (1980). *SUPERCARP-Sixth Edition*. Statistical Laboratory, Survey Section, Iowa State University, Ames, IA.
- Irwin, M.L., Ainsworth, B.E., Conway, J.M. (2001). Estimation of energy expenditure from physical activity measures: determinants of accuracy. *Obesity Research*, **9**, 517-525.
- Jakicic, J.M., Marcus, M., Gallagher, K.I., Randall, C., Thomas, E., Goss, F.L., Robertson, R.J. (2004). Evaluation of the SenseWear Pro Armband to assess energy expenditure during exercise. *Medicine and Science in Sports and Exercise*, **36**, 897-904.
- Janz, K.F. (2006). Physical activity in epidemiology: moving from questionnaire to objective measurement. *British Journal of Sports Medicine*, **40**, 191-192.
- Johnson, R.K., Russ, J., Goran, M.I. (1998). Physical activity related energy expenditure in children by doubly labeled water as compared with the Caltrac accelerometer. *International Journal of Obesity*, **22**, 1046-1052.
- Kannus, P. (1999). Preventing osteoporosis, falls, and fractures among elderly people. *British Medical Journal*, **318**, 205-206.
- Kipnis, V., Subar, A.F., Midthune, D., Freedman, L.S., Ballard-Barbash, R., Troiano, R.P., Bingham, S., Schoeller, D.A., Schatzkin, A., Carroll, R.J. (2003). Structure of dietary measurement error: results of the OPEN Biomarker Study. *American Journal of Epidemiology*, **158**, 14-21.
- Kritz-Silverstein, D., Barrett-Connor, E., Corbeau, C. (2001). Cross-sectional and prospective study of exercise and depressed mood in the elderly. *American Journal of Epidemiology*, **153**, 596-603.

- Lagerros, Y.T., Lagiou, P. (2007). Assessment of physical activity and energy expenditure in epidemiological research of chronic diseases. *European Journal of Epidemiology*, **22**, 353-362.
- Lagerros, Y.T., Mucci, L.A., Bellocco, R., Nyren, O., Balter, O., Balter, K.A. (2006). Validity and reliability of self-reported total energy expenditure using a novel instrument. *European Journal of Epidemiology*, **21**, 227-236.
- LaPorte, R.E., Montoye, H.J., Caspersen, C.J. (1985). Assessment of physical activity in epidemiologic research: problems and prospects. *Public Health Reports*, **100**, 131-146.
- Lee, I.-M. (2003). Physical activity and cancer prevention-data from epidemiologic studies. *Medicine and Science in Sports and Exercise*, **35**, 1823-1827.
- Leenders, N.Y., Sherman, W.M., Nagaraja, H.N. (2006). Energy expenditure estimated by accelerometry and doubly labeled water: do they agree? *Medicine and Science in Sports and Exercise*, **38**, 2165-2172.
- Levin, S., Jacobs, D.R., Ainsworth, B.E., Richardson, M.T., Leon, A.S. (1999). Intra-individual variation and estimates of usual physical activity. *Annals of Epidemiology*, **9**, 481-488.
- Manson, J.E., Rimm, E.B., Stampfer, M.J., Colditz, G.A., Willet, W.C., Krolewski, A.S., Rosner, B., Hennekens, C.H., Speizer, F.E. (1991). Physical activity and incidence of non-insulin-dependent diabetes mellitus in women. *Lancet*, **338**, 774-778.
- Marshall, S.J., Jones, D.A., Ainsworth, B.E., Reis, J.P., Levy, S.S., Macera, C.A. (2007). Race/ethnicity, social class, and leisure-time physical inactivity. *Medicine and Science in Sports and Exercise*, **39**, 44-51.
- Matthews, C.E. (2005). Calibration of accelerometer output for adults. *Medicine and Science in Sports and Exercise*, **37**, S512-S522.
- Matthews, C.E. (2002). Use of self-report instruments to assess physical activity. In: Welk, G.J. ed. *Physical Activity Assessments for Health Related Research*. Champaign: Human Kinetics.

- Matthews, C.E., Ainsworth, B.E., Thompson, R.W., Bassett, D.R. (2002). Sources of variance in daily physical activity levels as measured by an accelerometer. *Medicine and Science in Sports and Exercise*, **34**, 1376-1381.
- Matthews, C.E., Hebert, J.R., Freedson, P.S., Stanek, E.J., Merriam, P.A., Ebbeling, C.B., Ockene, I.S. (2001). Sources of variance in daily physical activity levels in the Seasonal Variation of Blood Cholesterol Study. *American Journal of Epidemiology*, **153**, 987-995.
- Matthews, C.E., Hebert, J.R., Freedson, P.S., Stanek, E.J., Ockene, I.S., Merriam, P.A. (2000). Comparing physical activity assessment methods in the Seasonal Variation of Blood Cholesterol Levels Study. *Medicine and Science in Sports and Exercise*, **32**, 976-984.
- Moy, K.L., Calabro, M.A., McClain, J., Welk, G.J. Comparison of multi-channel pattern recognition monitors for estimating energy expenditure – a brief review. Submitted.
- National Research Council. (1986). *Nutrient Adequacy: Assessment Using Food Consumption Surveys*. Washington D.C: National Academy Press.
- Nichols, J.F., Morgan, C.G., Sarkin, J.A., Sallis, J.F., Calfas, K.J. (1999). Validity, reliability, and calibration of the Tritrac accelerometer as a measure of physical activity. *Medicine and Science in Sports and Exercise*, **31**, 908-912.
- Nusser, S.M., Carriquiry, A.L., Dodd, K.W., and Fuller, W.A. (1996). A semiparametric transformation approach to estimating usual daily intake distributions. *Journal of the American Statistical Association*, **91**, 1440-9.
- Rosner, B., Michels, K.B., Chen, Y.-H., Day, N.E. (2008). Measurement error for nutritional exposures with correlated measurement error: use of the method of triads in a longitudinal setting. *Statistics in Medicine*, **27**, 3466-3489.
- Sallis, J., Haskell, W., Wood, P., Fortmann, S., Rodgers, T., Blair, S., Paffenbarger, R. (1985). Physical activity assessment methodology in the Five-City Project. *American Journal of Epidemiology*, **121**, 91-106.

- 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**, 1-14.
- SAS Institute, Inc. (2009). SAS Online Doc, version 9.2. Cary, NC: SAS Institute, Inc.
- Schoeller, D.A., van Santen, E. (1982). Measurement of energy expenditure in humans by doubly labeled water method. *Journal of Applied Physiology*, **53**, 955-959.
- Schutz, Y., Weinsier, R.L., Hunter, G.R. (2001). Assessment of free-living physical activity in humans: an overview of currently available and proposed new measures. *Obesity Research*, **9**, 368-379.
- Shapiro, S.S. and Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, **52**, 591-611.
- Shephard, R.J. (2003). Limit to the measurement of habitual physical activity by questionnaires. *British Journal of Sports Medicine*, **37**, 197-206.
- Spiegelman, D., Schneeweiss, S., McDermott, A. (1997). Measurement error correction for logistic regression models with an “alloyed gold standard.” *American Journal of Epidemiology*, **145**, 184-196.
- Spiegelman, D., Zhao, B., Kim, J. (2005). Correlated errors in biased surrogates: study designs and methods for measurement error correction. *Statistics in Medicine*, **24**, 1657-1682.
- Starling, R.D., Matthews, D.E., Ades, P.A., Poehlman, E.T. (1999). Assessment of physical activity in older individuals: a doubly labeled water study. *Journal of Applied Physiology*, **86**, 2090-2096.
- Strath, S.J., Bassett, D.R., Swartz, A.M., Thompson D.L. (2001). Simultaneous heart rate-motion sensor technique to estimate energy expenditure. *Medicine and Science in Sports and Exercise*, **33**, 2118-2123.

- Strath, S.J., Brage, S., Ekelund, U. (2005). Integration of physiological and accelerometer data to improve physical activity assessment. *Medicine and Science in Sports and Exercise*, **37**, S563-S571.
- Thune, I., Brenn, T., Lund, E., Gaard, M. (1997). Physical activity and the risk of breast cancer. *New England Journal of Medicine*, **336**, 1269-1275.
- Troiano, R.P. (2005). A timely meeting: objective measurement of physical activity. *Medicine and Science in Sports and Exercise*, **37**, S487-S489.
- Troiano, R.P., Berrigan, D., Dodd, K.W., Masse, L.C., Tilert, T., McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise*, **40**, 181-188.
- Trost, S.G., McIver, K.L., Pate, R.R. (2005). Conducting accelerometer-based activity assessments in field-based research. *Medicine and Science in Sports and Exercise*, **37**, S531-S543.
- U.S. Department of Health and Human Services (2008). *2008 Physical Activity Guidelines for Americans*. Atlanta, Georgia: U.S. Department of Health and Human Services, Public Health Service, CDC, National Center for Chronic Disease and Prevention and Health Promotion.
- U.S. Department of Health and Human Services (2009). *NHANES - National Health and Nutrition Examination Survey Homepage*. Available: <http://www.cdc.gov/nchs/nhanes.htm>. Last accessed 3 September 2009.
- U.S. Department of Health and Human Services (1996). *Physical Activity and Health: A Report of the Surgeon General*. Atlanta, Georgia: U.S. Department of Health and Human Services, Public Health Service, CDC, National Center for Chronic Disease Prevention and Health Promotion.
- 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 Sports and Exercise*, **37**, S582-S588.

- Warnecke, R.B., Johnson, T.P., Chavez, N., Sudman, S., O'Rourke, D.P., Lacey, L., Horm, J. (1997). Improving questionnaire wording in surveys of culturally diverse populations. *Annals of Epidemiology*, **7**, 334-342.
- Welk, G.J. (2002a). Introduction to physical activity research. In: Welk, G.J. ed. *Physical Activity Assessments for Health Related Research*. Champaign: Human Kinetics.
- Welk, G.J. (2005). Principles of design and analyses for the calibration of accelerometry-based activity monitors. *Medicine and Science in Sports and Exercise*, **37** S501-S511.
- Welk, G.J. (2002b). Use of accelerometry-based activity monitors to assess physical activity. In: Welk, G.J. ed. *Physical Activity Assessments for Health Related Research*. Champaign: Human Kinetics.
- Welk G.J., Schaben, J.A., Morrow, J.R. (2004). Reliability of accelerometry-based activity monitors: a generalizability study. *Medicine and Science in Sports and Exercise*, **36**, 1637-1645.
- Yaffe, K., Barnes, D., Nevitt, M., Lui, L., Covinsky, K. (2001). A prospective study of physical activity and cognitive decline in elderly women. *Archives of Internal Medicine*, **161**, 1703-1708.
- Zhang, K., Werner, P., Sun, M., Pi-Sunyer, F.X., Boozer, C.N. (2003). Measurement of human daily physical activity. *Obesity Research*, **11**, 33-40.