ADVANCING ACCELEROMETRY-BASED PHYSICAL ACTIVITY

MONITORS: QUANTIFYING MEASUREMENT ERROR AND IMPROVING

ENERGY EXPENDITURE PREDICTION

By

Megan Pearl Rothney

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

In partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Engineering

May, 2007

Nashville, Tennessee

Approved by:

Professor Kong Chen

Professor Robert Roselli

Professor Frank Harrell Jr.

Professor Richard Shiavi

Professor Paul Harris

Professor Charles Matthews

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

Appendix

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AC – Actical accelerometer

ACSM – American college of sports medicine

AG – ActiGraph accelerometer

ANN – Artificial neural network

BMI – Body mass index

BMR – Basal metabolic rate

CV – Coefficient of variability

DEXA – Dual x-ray absorptometry

ECG – Electrocardiogram

EE – Energy expenditure

EE_{ACT} – Energy expenditure associated with physical activity

FW – Filter weighting function

HR – Heart Rate

MAE – Mean absolute error

MET – Metabolic Equivalent

MSE – Mean squared error

PA – Physical activity

PAL – Physical activity level

PAR – Physical activity ratio

RMR – Resting metabolic rate

RQ – Respiratory quotient

TEE – Total energy expenditure

TEF – Thermic effect of food

VCO_2 – Carbon dioxide production rate

VO_2—Oxygen consumption rate

CHAPTER I

INTRODUCTION

Motivation

Obesity is an increasing epidemic in the Western world with nearly two-thirds of Americans classified as either overweight (Body mass index, BMI, between 25.0 and 29.9 kg/m^2) or obese (BMI $\geq$ 30.0 kg/m^2) (2). Obesity carries with it the risk of developing several chronic conditions such as Type II Diabetes (11), coronary artery disease (10), and osteoarthritis (8). Analytically, the weight changes that lead to obesity represent an imbalance between energy intake and energy expenditure (EE). Though many factors, both behavioral and genetic, are capable of altering energy balance, one of the most readily modified components is the energy expenditure associated with physical activity (EE$_{ACT}$). Accurate assessment of physical activity is therefore a critical step in assessing the role of physical activity (PA) in body weight maintenance both on an individual basis and as a tool in developing and validating public health recommendations for daily PA levels.

The Vanderbilt University Energy Balance Laboratory is a unique environment comprised of physicians and research scientists devoted to the accurate assessment of EE and the development of hardware and software tools to carry these measurements from the laboratory into free-living. Accurate measurements of EE obtained in the fast response indirect room calorimeter, housed in the General Clinical Research Center, allow for subject-specific metabolic profiles to be developed for a wide variety of

subjects (12). These results represent an accurate assessment tool for daily energy balance in the laboratory. While inpatient measurements represent the gold standard in metabolic assessment, there is an acute need to develop tools that can be translated into a free-living environment that are capable of assessing the duration and intensity of daily PA patterns, and are accurate in their prediction of the metabolic costs of these activities.

Accelerometers have been considered a candidate for assessment of PA due to the physical relationship between acceleration and work (9) as well as the relatively low cost and simplicity of use. By relating accelerometer output to measured EE using indirect calorimeters, relationship can be established between the intensity of PA and the resulting change in EE (4-7). When subjects are then asked to wear accelerometers while they complete their normal daily routine, and the resulting data is analyzed using prediction equations developed in a laboratory setting, a prediction of their PA patterns are developed. However, when the first approaches were subjected to laboratory and short (several hours duration) field validation, significant errors emerged between EE measurements and predictions (1). These errors are likely a result of the limited duration of studies used for model development (7), and use of small, homogenous subject populations (5). Additionally, data from commercially-available accelerometers is coarse, typically one sample per minute, which limits the complexity of the relationship that can be established between their output and EE.

Though accelerometers have been considered a promising tool for objectively assessing PA in a free-living environment, both the measurement devices and analysis paradigms to date have not been able to realize this promise. Two main areas can be identified as shortcomings in existing technologies: accelerometer hardware variability

and analytic tools relating acceleration to EE. To this end, we will explore the capabilities of existing analysis routines to predict EE on a daily basis by comparing the monitor output to data acquired in a diverse population of subjects in a room calorimeter. Additionally, we will attempt to characterize the hardware filtering in three commercially available accelerometer devices by subjecting them to a range of mechanically generated accelerations. These experiments will help us to better understand the unit of activity counts, which is the reporting unit used in most commercially available accelerometry-based PA monitors. Further, we will explore a new modeling approach for the prediction of EE using raw (32 Hz) acceleration signals. Current data acquisition and analysis techniques using accelerometry-based PA monitors discard a large amount of potentially useful information by utilizing only an integrated signal, generally reported over a one minute interval, though device sampling is usually performed at a much greater temporal resolution (3). This model will be developed using a data set that is large both with respect to the number and heterogeneity of subjects included as well as the length of each study visit.

Overview

The research discussed in this thesis was performed to achieve three main objectives: (1) To understand the strengths and weaknesses of published relationships between acceleration and EE in twenty-four hour clinical data using market-available accelerometer technologies, (2) To characterize the device-specific hardware filtering in three commercial accelerometers subjected to mechanically generated accelerations, and

(3) To develop novel methods for prediction of EE using raw acceleration signals and artificial neural network (ANN) modeling.

Chapter II provides a brief introduction to and clinical relevance of obesity, a description of energy balance as well as introducing the measurement principles involved in metabolic monitoring and PA assessment. Instrumentation used throughout the thesis will be introduced and brief qualitative descriptions of currently used and proposed analytic techniques will be discussed. The goal of this chapter is to enable readers to understand both the direction and the outcomes of this body of research. More detailed mathematical descriptions directly applicable to the modeling approaches used can be found in Appendix A.

Chapter III explores the predictive accuracy of three commonly used accelerometers and seven of their associated EE prediction algorithms when applied to twenty-four hour studies where EE was constantly measured using a room calorimeter. This work represents a bridge between short structured activity protocols that have typically been used for model development and free-living studies where no reference standard is available on a minute-by-minute basis. Most of the regression equations were able to accurately assess the percent of the study visit spent in four PA intensity categories. However, most regressions seemed to under-predict the intensity of some data which resulted in an under prediction of the daily physical activity level (PAL). Results from this experiment were further explored in Appendix B where the EE is predicted using each model, and the impact of the baseline EE on the daily prediction errors is explored.

Chapter IV describes the response of three generations of a commonly used accelerometer, the ActiGraph, subjected to a wide range of accelerations generated by altering both the frequency and radius of oscillation on a modified bench-top orbital shaker. Results from this study are used to understand both the response profile for each monitor generation as well as a study of the inter-monitor variability between the generations. The ActiGraph monitor demonstrated a nonlinear response when activity counts are related to acceleration. We were able to detect statistically significant point differences between the three generations of monitors at a number of acceleration values. These differences in counts corresponded to altered EE prediction of, on average, 0.30 kcal/min (~ 130 kcal/day) for activity count values less than 4000, which account for greater than 90% of our study visit. Inter-monitor variability was consistent for monitors of each generation once a low frequency threshold had been passed (generally around 40 RPM). Inter-monitor variability was reduced in the newest version of the ActiGraph, the GT1M, recommending its use in studies where measurements must be made with multiple monitor units. Similar data for two additional activity monitors (Actical, and RT3) as well as preliminary data relating the three activity monitors can be found in Appendix C.

To fundamentally change the classical paradigm of using summed acceleration signals to model EE, Chapter V proposes a new approach where raw acceleration data (32 Hz) collected from a bi-axial hip-mounted accelerometer is used to drive an ANN modeling approach. To accomplish this goal, raw signal data is reduced into a small number of summary statistics (features) for each minute of data. These statistics, combined with subject characteristics, are used as inputs to a three-layer (one hidden

5

layer) network. EE prediction accuracy is compared with results obtained from a

traditional (one minute epoch) analysis approach as well as compared with a multi-

accelerometer array. Results of this study indicate that by using raw data at the hip,

significant improvements can be realized relative to a hip only model developed using

integrated data. Improvement was also found relative to a proprietary EE prediction

algorithm developed using five sensor sites. This work suggests that ANN modeling has

the potential to reduce error in prediction of EE as well as encouraging the notion that

signal characteristics other than the integral over one minute may be useful in predicting

EE. Additional results from the generalized ANN model can be found in Appendix D.

Individual ANN models have been developed for each subject to study the improvements

in model performance that can be achieved by using the data on a per subject basis. These

experiments showed that individual models for EE using an ANN are able to improve

upon the error values achieved using a generalized model. These results can be found in

Appendix E.

Chapter VI provides a summary of this body of work, global conclusions, and

future directions.

References

1.   **Bassett DR, Jr., Ainsworth BE, Swartz AM, Strath SJ, O'Brien WL, and King GA.** Validity of four motion sensors in measuring moderate intensity physical activity. *Medicine and science in sports and exercise* 32: S471-480, 2000.

2.   **Brooks G, Fahey T, and Baldwin K.** Exercise Physiology: Human Bioenergetics and Its Applications. New York: MacGraw Hill, 2005, p. 617-645.

3.   **Chen KY and Bassett DR, Jr.** The technology of accelerometry-based activity monitors: current and future. *Medicine and science in sports and exercise* 37: S490-500, 2005.

4.   **Chen KY and Sun M.** Improving energy expenditure estimation by using a triaxial accelerometer. *J Appl Physiol* 83: 2112-2122, 1997.

5.   **Freedson PS, Melanson E, and Sirard J.** Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine and science in sports and exercise* 30: 777-781, 1998.

6.   **Heil DP.** Predicting activity energy expenditure using the Actical activity monitor. *Research quarterly for exercise and sport* 77: 64-80, 2006.

7.   **Hendelman D, Miller K, Baggett C, Debold E, and Freedson P.** Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Medicine and science in sports and exercise* 32: S442-449, 2000.

8.   **Hinton R, Moody RL, Davis AW, and Thomas SF.** Osteoarthritis: diagnosis and therapeutic considerations. *American family physician* 65: 841-848, 2002.

9.   **Ozkaya N and Nordin M.** *Fundementals of Biomechanics: Equilibrium, Motion, and Deformation.* New York, NY: Springer Science, 1999.

10.   **Poirier P, Giles TD, Bray GA, Hong Y, Stern JS, Pi-Sunyer FX, and Eckel RH.** Obesity and cardiovascular disease: pathophysiology, evaluation, and effect of weight loss. *Arteriosclerosis, thrombosis, and vascular biology* 26: 968-976, 2006.

11.   **Sharma AM.** The obese patient with diabetes mellitus: from research targets to treatment options. *The American journal of medicine* 119: S17-23, 2006.

12.   **Sun M, Reed GW, and Hill JO.** Modification of a whole room indirect calorimeter for measurement of rapid changes in energy expenditure. *J Appl Physiol* 76: 2686-2691, 1994.

CHAPTER II

BACKGROUND

Obesity and its clinical significance

Obesity is characterized by an excess of body fat. It can be diagnosed using either percent body fat or body mass index (BMI) depending on the availability of appropriate measurement equipment. Using body fat definitions, women with greater than 33% and men with greater than 25% body fat are considered obese (16). More commonly used in the clinic are determinations made using BMI, which is computed as weight in kilograms divided by the square of height in meters. BMI is divided into four broad classifications: underweight (BMI < 18.5 kg/m^2), healthy (18.5 < BMI < 25 kg / m^2), overweight (25 < BMI < 30 kg/ m^2), and obese (BMI > 30 kg/ m^2) (1). Approximately 66.3% of American adults are classified as either overweight or obese (2). The prevalence of obesity has increased from 12.8% to 32% since 1960 (32, 51).

Epidemiological research has shown correlations between obesity and several chronic diseases including coronary artery disease (54), hypertension (17), type II diabetes (63), and osteoarthritis (40). Obese individuals also have a shorter life expectancy (31) and run a higher risk of depression (71). Health care costs for the obese are greater than for individuals with healthy BMI, with reported associations between obesity and days of hospitalization, number and cost of outpatient doctor visits, and increased spending on prescription drugs (57). The estimated health care cost of obesity in 2003 was $130 billion (16).

Energy Balance

Body weight regulation occurs by balancing energy intake with total energy expenditure (TEE) in a process referred to as energy balance. Energy intake can be defined as any food, beverage, or calorie-containing nutritional supplement that is ingested. TEE is made up of three distinct components. The basal metabolic rate (BMR) is the energy cost associated with maintaining life in a resting condition. It can be approximated using physical attributes with any number of predictive equations (33), or it can be measured in the laboratory using calorimetry (18). BMR accounts for approximately 60% of TEE in an average adult (44). The thermic effect of food (TEF) is the component of EE resulting from the increase in metabolism due to digestion and absorption of nutrition, which accounts for approximately 10% of TEE. TEF is dependant on both the size and macronutrient content of a meal (69). The remaining energy expenditure is associated with physical activity (EE_{ACT}). Physical activity (PA) can be defined as active movements or motions performed by the body and therefore encompass a broad spectrum of activities from those with a low intensity such as fidgeting to high intensity intervals such as running (45).

When, over time, energy intake and energy expenditure are not matched, fluctuations in body weight occur. Positive energy balance is defined as the condition where energy intake exceeds TEE and will result, in time, in weight gain. Conversely, negative energy balance occurs when there is not adequate energy intake to compensate for the TEE (16). Accurate characterization of each component of energy balance is an important step in understanding body weight regulation.

Measuring Energy Expenditure

Human energy expenditure is most commonly measured using calorimeters (47). Calorimeters rely on the thermodynamic conservation principle that energy cannot be created or destroyed (43). Thus, changes in the energy content of a closed system must be due to conversion of energy from one form to another. Two types of calorimeters, direct and indirect, exist and both can be used for quantifying human energy expenditure. Direct calorimeters measure the temperature change of a known volume of water due to the heat production of the human subject (61, 62). This method is rarely used for clinical data collection due to the technical challenges and costs involved. Indirect calorimeters are more frequently used. In indirect calorimeters, the differences between the subjects' inspired and expired air are characterized as oxygen (O_2) and carbon dioxide (CO_2) fractions in the air they breathe. EE is then be computed from these fractional gas concentrations using standard equations (42, 67).

Several types of instrumentation have been developed to compute energy expenditure using indirect calorimetry. One commonly used device is the metabolic cart (24), which includes devices such as the Vmax Encore (Viasys; Yorba Linda CA), and the Quark (Cosmed, Rome IT) systems. Subjects are asked to wear either a mouthpiece and nose clip or a mask over their nose and mouth. The mouthpiece or mask is connected to an external gas analyzer via tubing. Subjects engage in exercise and the gas analyzer returns several parameters such as oxygen consumption (VO_2), carbon dioxide production (VCO_2), the respiratory quotient (RQ; ratio of VCO_2 to VO_2) and EE. Metabolic carts are in widespread clinical use. They are relatively easy to use, have been well validated (using test-retest validations), and are relatively inexpensive. Because they are

commercially manufactured, an onsite technical expert is not required to keep them operational. Metabolic carts do have some limitations. The subject is tethered to the gas analyzer and therefore has a limited ability to perform locomotion or sport specific activities. There is also discomfort associated with the mouthpiece or mask which limits the test duration and subject compliance.

An alternative approach is the portable indirect calorimeter. These systems use the same measurement principles as the metabolic cart but they have been designed to be lighter weight and battery powered (28, 48). Examples of portable calorimeters are the Oxycon Mobile (Viasys, Yorba Linda, CA) and K4B^2 (Cosmed; Rome IT). The design of the portable calorimeters allows subjects to wear the gas analyzer unit in a harness either on the back or the chest. While these systems suffer from the same types of restrictions in experimental duration (two hours maximum) and subject discomfort as the metabolic carts, they do allow subjects to engage in most types of PA (water sports are excluded). While the increase in the variety of PA modes is desirable for investigators, subjects using these devices are required to carry an added load associated with the gas analyzer, battery, and harness. While this load is generally small (< 5 lbs), there may be effects on exercise performance particularly in subjects with a low body weight.

Longer measurements of energy expenditure can be made in a room calorimeter (Figure 1). In these custom devices, subjects are sealed into a room where gas concentrations are constantly measured. All hardware is housed outside the calorimeter where it can be monitored by a trained operator. Compared to metabolic carts and portable calorimeter units, room calorimeters allow a wide variety of activities to be performed without the discomfort of a mask or mouthpiece. Some activities cannot be

performed due to the size of the room (e.g. baseball, golf). Activities that would change

the gas content of the air are also excluded. Room calorimeters are expensive to build and

maintain and require onsite technical expertise. They are also limited in their temporal

resolution since the volume of the gas in the calorimeter is large relative to the changes in

concentration the subject can generate. As a result, there are only about ten room

calorimeters currently in use in the United States.

Figure 1: Images of a subject engaged in various activities in the Vanderbilt room
calorimeter.

The Vanderbilt whole-room indirect calorimeter is a small, airtight environmental room (2.6 x 3.4 x 2.4 m, 19,500 liters in net air volume) with an entrance door (1 x 2 m) and an air lock (0.6 x 0.3 m) for passing food and other items. The room is equipped with a desk, a chair, an outside window, a toilet, a sink, a telephone, a TV with VCR/DVD player, an audio system/alarm clock, and a fold-down mattress. Oxygen consumption (VO_2) and carbon dioxide production (VCO_2) are calculated by measuring the changes of oxygen and carbon dioxide content of the air inside the calorimeter and by the flow rate of the purged air times multiplied by the concentration of gases. A special multi-channel air sampling system was designed to ensure an even sampling of the gas expired by the subject. Temperature is precisely controlled, while barometric pressure and humidity of the room are monitored (Figure 2). For this investigation, the sensitivity of the system (response time) is very important because EE changes in minutes in response to body movement during PA. Our system can consistently achieve the highest accuracy and fastest response time (>90% recovery in 1 min) compared to other room calorimeter units that are similar in size (67). To our knowledge, these are the most accurate data ever reported using whole room indirect calorimeters with measurement intervals as short as one minute. The accuracy and fast response of this whole-room indirect calorimeter make it possible to study EE_{ACT} in details that were not previously feasible using room calorimeters.
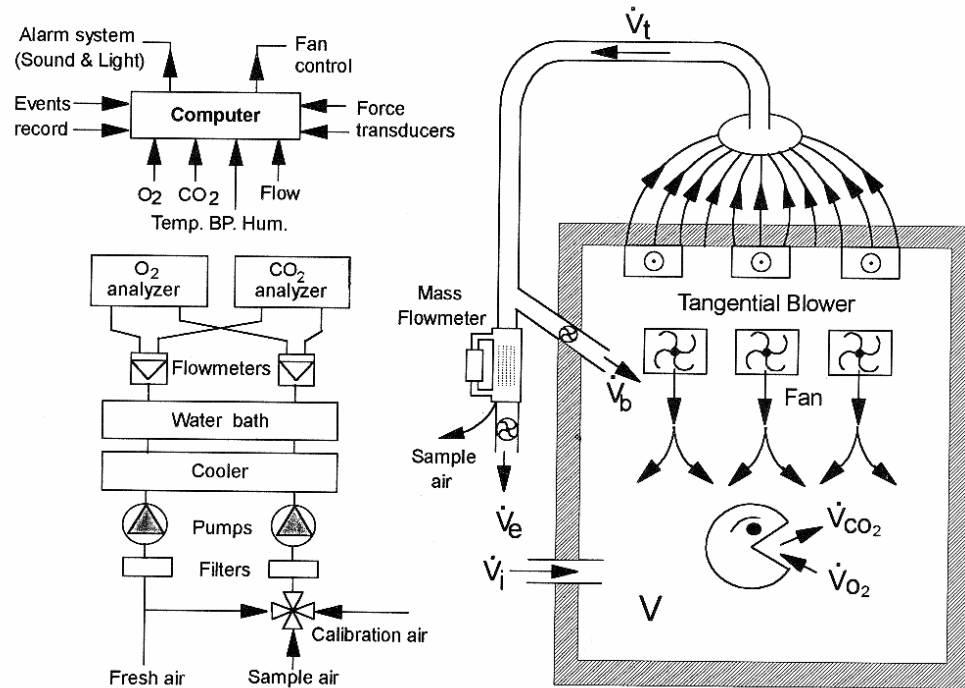
FIG. 1. Schematic drawing of whole room indirect calorimeter at Vanderbilt University is shown after described modifications. System was originally constructed in 1988. BP, barometric pressure; $\dot{V}_T$, total flow; $\dot{V}_E$, expiratory flow; $\dot{V}_I$, inlet flow; $\dot{V}_B$, airflow back to chamber; V, volume; $\dot{V}_{O_2}$, O_2 consumption; $\dot{V}_{CO_2}$, CO_2 production.

Figure 2: Schematic detailing the operation of the Vanderbilt University room calorimeter (67).

An alternative approach which can be used to measure EE is the doubly-labeled water technique (DLW). Subjects are asked to consume water with a known concentration of two stable isotopes: deuterium and oxygen-18. Urine or saliva concentration of the isotopes is measured both before and several days after consumption of the labeled water and the differential clearance rate of the isotopes is used to assess the CO_2 production. Oxygen consumption is then computed using either an estimated or measured RQ (60). DLW only gives information about total energy expenditure over the measurement period, usually 7-10 days (all others can give information on the order of

14

minutes). It is also prohibitively expensive for most researchers with high costs of both the labeled water and the specialized spectroscopy required for data analysis.

Studying Physical Activity

While all components of energy balance must be studied and understood, PA is an ideal component of TEE for investigation since it can be directly controlled by individuals on a daily basis and participation can be systematically altered through exercise prescriptions. Alterations in PA have the capabilities of changing all components of EE (64, 74). Though most obviously participation in exercise leads to increases in EE_{ACT}, in the long term these changes will also result in increases in BMR. BMR is highly correlation with lean body mass (64). After consistent exercise performance, it is common that lean body mass would increase while fat mass decreases (10). TEE has also been shown to increase in individuals who are active relative to inactive control matched counterparts (65). It is therefore desirable to quantify the intensity and duration of PA when attempting to understand the etiology of obesity.

As discussed previously, PA can be measured in the laboratory using calorimetry or DLW. However, in order to develop a tool that is useful for understanding the patterns of activity and its dynamic range in a real world or free-living environment, it is desirable to develop instrumentation that can capture information about the duration and intensity of PA bouts over several days at a time, while still maintaining a well-characterized relationship with minute-by-minute EE. Such devices would ideally be inexpensive, non-invasive (worn externally with a minimum of added weight), and capable of monitoring

activity over a period of days or even weeks. To this end, a number of types of instruments, both qualitative and quantitative, have been developed.

One approach to estimate EE_{ACT} is through the subjective reporting of PA. Broadly, two types of survey methods exist: activity recalls and activity diaries. In activity recalls a trained interviewer asks each subject questions about their occupation, leisure activities, and sport participation (46, 66). This information, coupled with the subject's age, ethnicity, and a subjective assessment of the individual's physical fitness is used to develop an estimate of the subject's TEE. In activity diaries, each day is divided into segments of at least 15 minutes (76, 77). Subjects are asked to account for the primary activity they engaged in during each time period. Researchers collect the survey and assign energy costs to each interval based on the activity reported and a table of PA types and their associated energy costs (5). Surveys are ideal for large studies since the cost per subject is low and many subjects can participate simultaneously. They do however suffer from large inter-subject variability as not all subjects will accurately report their activities (76).

Portable sensor systems offer more quantitative information about PA. These monitors rely on either physical or physiological information to develop an estimate of EE_{ACT}. This class of device includes heart rate (HR) monitors, pedometers, and accelerometers. While more expensive than surveys, these devices are all small, non-invasive, simple to use, and able to monitor PA for several days at a time.

HR can be detected using either a threshold voltage detector (3) or through a feature detection routine (e.g. measuring the RR interval) as part of the collection of an ECG signal (12). EE_{ACT} can be predicted by capitalizing on the linear or close to linear

relationship between HR and EE_{ACT} over a wide variety of aerobic conditions. But, HR measurements have the disadvantage of being affected by a number of external factors such as gender, fatigue, stress, and state of hydration which are not always associated with a proportional change in EE_{ACT}. Individual curves relating HR and EE_{ACT} could therefore be created with high accuracy but generalized modeling is difficult. In addition, HR is also not a good predictor of EE_{ACT} during low intensity (sedentary) activities (14, 23), which account for a large portion of most subjects' daily PA.

Pedometers are devices designed to count steps and estimate distance traveled during locomotion. Vertical acceleration is used to displace a lever arm inside the device. When the lever arm displacement reaches a threshold, it completes an electric circuit, which triggers the device to record a step (27). Pedometers are easy to use, with simple placement instructions and few if any input parameters for the user to specify. They also provide the user with immediate visual feedback indicating the number of steps taken, which can serve as a motivational tool.

The same technical simplicity that makes pedometers attractive to consumers hampers their use as predictors of EE_{ACT}. The primary limitation of pedometers for PA analysis is they only attempt to account for information pertaining to locomotion. They also typically underestimate distance during slow walking and over-estimate it during jogging mostly due to a single stride length assumption for each subject (modeled by sex, weight and height) (8). Another limitation of pedometers is that many devices do not record a time history and therefore cannot be used to evaluate the pattern of PA over time (7).

Over the last twenty-five years accelerometers have become a common sensor used in portable PA monitors. Accelerometers have been chosen because of the proportionality between force and acceleration expressed in Newton's Second Law (F_{NET} = mass*acceleration) (52). Since force can be correlated with mechanical work (Work = force*displacement), and mechanical work requires an input of energy (Work = Change in Energy), acceleration in the vertical plane has been used as an estimate of EE_{ACT}. Acceleration measurements in PA monitors are most commonly made using beam bending piezo-electric accelerometers. Deflection of the beam due to activity causes a transient change in output voltage proportional to the magnitude of the deflection. Accelerometers are usually worn at the subject's hip, a site close to the center of mass, and measure in at least the vertical plane (21). In most commercially available systems, voltage signals are rectified, filtered, and integrated over a user-defined measurement epoch and data are reported as activity counts, an arbitrary unit whose meaning varies between device manufacturers based on device sensitivity and filtering routines (36).

Commercially-available accelerometers can be coarsely divided into first generation and second generation devices. First generation devices are defined as single site, single sensor type (acceleration only) units. These devices are able to record a time history of PA intensity information and have been widely used on adults (35, 58, 72) and children (34, 56) for monitoring PA over the course of days or weeks. Measurement accuracy is limited because accelerometers are only sensitive to movements on the part of the body to which they are attached. Temporal resolution of first-generation accelerometer output is also limited by the memory capacity of the monitors, battery life,

and the data epochs prescribed by the manufacturers. One minute is the standard field measurement interval, which was largely determined by the memory storage requirements of early accelerometer devices, though most modern accelerometers can record in several time epochs between one second and 15 minutes. Processed (filtered) signals are the only ones that end-users are able to access in most first-generation devices. They also have sensitivity primarily in a single plane, potentially limiting detection of sedentary activities, which may not have vertical acceleration components. The three most common first generation accelerometers currently in use are the ActiGraph (formerly MTI/CSA, Fort Walton Beach, FL), the Actical (MiniMitter/Respironics, Bend OR), and the RT3 (Stayhealthy, Monrovia CA).

The ActiGraph is a uni-axial accelerometer device. The 7164 model has 64KB internal random access memory (RAM) for data storage of 22 days at 1-minute epoch. The 71256 model increases the memory storage capacity by 4 times (256KB RAM). The piezo-electrical accelerometer has a dynamic range of 0.05-2.0 G (multiples of the gravitational unit; 9.81 m/s^2) with a frequency response between 0.25-2.5 Hz (70). Recently, a new model, the GT1M with expanded memory (1 MB) was released. A new sensor has been used in the GT1M monitors, however according to the manufacturer the internal filtering has been designed to minimize differences between the GT1M and previous ActiGraph devices. All ActiGraph monitors are configured to allow attachments to the wrist, ankle, or waist.

The Actical is a small omni-directional accelerometer. It is worn at the hip, wrist, or ankle of the subject. The AC64 model has 64KB RAM for data storage that can record PA counts for up to 45 days using the 1-minute epoch. The sampling frequency is 32 Hz,

and sensitivity is 0.01 G. The Actical is a sister technology to the ActiWatch, which has been used extensively in sleep research (9). Though the sensor, memory size, and sampling rate are common to both the Actical and Actiwatch, the ActiWatch uses a peak detection routine to quantify the activity counts per epoch, where the Actical returns the integrated signal over the measurement epoch, which allows for a more balanced representation of the activity in the measurement epoch relative to the peak detection routine (38).

The RT3 Research Tracker is a triaxial accelerometer device, which is built on the technology of the Tritrac-R3D (58), which was the first commercial tri-axial accelerometer for PA assessment. The RT3 is the size of a pager and is worn clipped onto the waist. It uses three orthogonally mounted piezo-electric accelerometers to measure motion in three planes and provides tri-axial vector data in activity count units. The sensor range, data sampling frequency, and A/D converting resolution are proprietary. The manufacture's software calculates a subject's EE_{ACT} using the vector magnitude of the activity counts and a linear regression algorithm, which is also proprietary. The RT3 is capable of collecting and storing data up to 8.5 days (triaxial mode) with a battery life of 30 days.

While a number of first generation devices have been used, the ActiGraph, Actical, and RT3 represent a cross section of device sizes, measurement axes, sampling rates, and frequency bandwidths. A table comparing some of the most relevant specifications of the first generation devices used in these experiments is presented in Table 1. Some of the specifications from the RT3 have appeared in peer-reviewed publications but have not been verified with the manufacturer.

Table 1: Descriptions of first generation physical activity monitors

	ActiGraph 7164	ActiGraph 71256	ActiGraph GT1M	Actical	RT3
Size	51x41x 15mm	51x41x 15mm	53x51x 22mm	28x27x 10mm	71x56x 28mm
Memory	64 Kb	256 Kb	1 Mb	64 Kb	128 Kb*
Sensitive Axes	Uni-axial	Uni-axial	Uni-axial	Omni-directional	Tri-axial
Sampling Rate	10 Hz	10 Hz	10 Hz	32 Hz	(10–12 Hz)*
Bandwidth	0.5-2.5 Hz	0.5-2.5 Hz	0.5-2.5 Hz	0.5-3.2 Hz	0.2-10Hz*

* Unpublished

In order to attempt to address some of the limitations of first generation accelerometers, a second generation of accelerometry-based PA monitors has been developed. Second generation accelerometers are either equipped with an array of accelerometers worn at multiple sites (58, 79, 80), or they incorporate accelerometers alongside other physiologic measurements such as HR (12), near body temperature, and skin impedance (41). Most second generation monitors also allow raw signals to be collected and analyzed, at least over a short period of time. Increased temporal resolution coupled with a larger number of sensors has increased the amount of information available for EE_{ACT} prediction in second generation devices. Three commercially-

available second generation devices are the Actiheart (MiniMitter/Respironics, Bend OR), the Sensewear (Bodymedia, Pittsburgh PA), and the IDEEA monitor (MiniSun, Fresno CA). The Actiheart and Sensewear monitors combine physical and physiological measures, while the IDEEA monitor uses an array of accelerometers attached to the body at multiple sites.

The Actiheart is a small (7mm thick, 33 mm diameter, primary sensor) device worn on the chest that combines simultaneous HR detection and a uni-axial accelerometer. Data storage and acceleration detection occurs at the primary sensor and a wire runs to a smaller secondary sensor that is positioned for ECG acquisition. Heart rate data are collected at 128 Hz while acceleration is collected at 32 Hz. The accelerometer dynamic range is ± 2 G with a frequency response range of 1-7 Hz. Data are reported in either 15 second or one minute epochs, though EE predictions are made only on a minute-by-minute basis. The internal memory size is 128 KB which allows continuous recording for 11 days using a one-minute epoch (13).

The Sensewear Armband is a PA monitor (85x54x20 mm, 85g with an internal lithium-ion battery) which is uniquely contoured to be worn at the upper arm. The internal sensors include an accelerometer as well as sensors for heat flux, galvanic skin response, skin temperature, and near-body ambient temperature. The accelerometer in the armband is a 2-axis accelerometer that utilizes a micro-electro-mechanical sensor device that measures motion. A poly-silicon spring supports a small mass that moves when subjected to external acceleration, i.e., body movements. The scale for the sensor is ±2 G with an 8-bit A/D converter (256 counts at 3.66 mg/count). The sampling rate is 32 Hz and has 512KB RAM of data storage. The manufacture's software calculates a subject's

EE$_{ACT}$ using a proprietary algorithm that combines acceleration, heat flux and other parameters. The Sensewear is capable of collecting and storing data up to 5.5 days at 1-minute epoch (41).

The Intelligent Device for Energy Expenditure and Activity (IDEEA), is an array of five accelerometers (upper sternum, mid thigh of both legs, and under both feet) designed to identify PA types and predict EE (79, 80). For our study, we utilized custom-designed IDEEA monitors with added sensor arrays for upper limbs. Our current study will use this novel device based on accelerometry sensors (20x15x4 mm, 2g) attached at sites on both of the upper arms, back of the both hands (middle of the 3rd metatarsal), and hip (internal accelerometer within the minicomputer), in addition to the current IDEEA sites. Data are transmitted through thin, flexible cables (diameter = 1.7mm) to a minicomputer (70x56x16 mm, weighing 59g, and powered by one AAA battery) clipped at the waist. The sensors on the hands, feet, hip, and sternum are bi-axial accelerometers and the sensors on the thighs and upper arms are uni-axial. The scale of the sensors are ±2 G. Currently, up-to 88 mega bytes (about 2 complete days) of continuous raw data can be stored in the minicomputer for download (via USB or serial port). Although the number of sensors and the wires limit the practical applications of this monitor in the field, the IDEEA2 is the only device that could measure and continuously record the raw signals necessary for our model development. An image of the IDEEA monitor and the sensor sites both for the commercial device and for the IDDEA2 custom device are shown in Chapter 5.

Table 2 shows three second generation activity monitors and provides some relevant specifications. The original IDEEA monitor is shown as this is the model that is commercially available.

Table 2: Description of three second generation activity monitors

	Actiheart	Sensewear	IDEEA
Sensor Types	Uni-axial accelerometer, Heart rate	Bi-axial accelerometer, Heat Flux, Near Body temperature, Galvanic Skin Response	5 accelerometers, mixture of uni-axial and bi-axial
Device Location	Chest (2 suggested locations)	Bicep of right arm	Hip (microprocessor), sensors on chest, mid-thighs, bottom of feet
Sampling Rate	32 Hz (acceleration) 128 Hz (heart rate)	32 Hz	32 Hz

Though a significant amount of research has been dedicated to understanding the output of accelerometers and how this output relates to EE, gaps still exist in our understanding of how to use the information from accelerometers most effectively. One of the major limitations in our current knowledge involves characterization of the relationship between the output of each PA monitor and measured EE. This involves both

validating existing equations in environments other than those that were used in development, but also the development of novel analytic routines to improve EE prediction accuracy. The second research area involves understanding the data filtering processes built in to accelerometers. Explorations in these areas will further reveal the capabilities of each accelerometer device to provide accurate and reproducible EE predictions. Developments in these areas will serve to guide the design of future instrumentation by identifying the sensor types and sites as well as the type of data acquisition and filtering that provide the best prediction potential.

Relating Acceleration to Energy Expenditure

Transforming the output data from accelerometers is essential for EE prediction, since the output units are arbitrary acceleration counts or raw voltage signals. Because activity counts are not standardized across types of PA monitors, this poses an additional challenge to generalized modeling using accelerometers because prediction equations are specific to the type of instrument used. Data analysis to date has existed in three main areas: intensity classification using ranges of metabolic equivalent units (METs), linear regression analysis relating counts and EE_{ACT} or METs, and nonlinear models where data are fitted to a pre-specified functional form.

METs are a unit often used to classify PA intensity because they are represent TEE normalized by resting metabolic rate (RMR). This computation reduces the inter-subject variability for a given type of PA and improves the ability of researchers to compare PA intensity across subjects of different body sizes and compositions. Traditionally, one MET is defined as an oxygen consumption of 3.5 mL/kg/min, which is

assumed to be the energy cost associated with sitting at rest (75). Therefore, it is assumed that all subjects performing a specific type of PA exert approximately the same number of METs even though the EE measurement may be very different. Thus METs provide a method to normalize data across subjects, which should reduce the error introduced into prediction equations when subjects whose physical characteristics, e.g., weight, height, age, differ from the model development set are introduced into a validation experiment.

A compendium of PA types and their associated MET costs is published periodically by the American College of Sports Medicine (ACSM) (4). While each activity type has been assigned a MET value or range of values, many analyses using METs further simplify data by examining ranges of METs. Standard analysis ranges have been designated by ACSM and correspond to light (1-3 METs), moderate (3 – 6 METs), vigorous (6 - 9 METs), and very vigorous (> 9 METs) PA, although additional ranges are sometimes specified (75). Recently, there has been increased interest in characterizing sedentary activities, which can be assigned a range of 1-1.5 METs.

Analysis using METs is the most logical approach to estimate EE in qualitative PA research where the activity type is known but no reference standard has been measured (PA recalls and surveys). From the activity type and the MET values for each activity (5), an estimation of the PA intensity can be made. Categorical analysis techniques are also appealing to researchers who are analyzing quantitative data over a long period of time or from many subjects since they reduce the large number of minute-by-minute data points into a smaller number of summary statistics that can be used to compare data prospectively or cross-sectionally. Early accelerometer research focused on defining cut-points, values of acceleration counts that discriminated between the MET

ranges (35, 39, 50, 66, 68). The cut-points are then used to determine the amount of time subjects spend engaged in PA of a specific intensity, which is useful in determining compliance to public health recommendations or personal PA prescriptions. To develop cut-points, subjects performed specific PA types while researchers made measurements simultaneously with indirect calorimeters and accelerometers.

To develop cut-points, several research groups have used linear regression to determine the relationship between VO_2 and activity counts for the measured modes of PA (35, 39, 50, 68). The count value corresponding to 3, 6, and 9 METs were extracted from the regression fit and these values were used to classify data into three or four intensity categories (> 6 METs is sometimes used). Many sets of cut-points were independently developed for the ActiGraph PA monitor, the most commonly used commercial device to date, however only three will be presented here. The first set of cut-points, developed by Freedson et al (35) was developed using treadmill walking and jogging at three specific speeds. In an attempt to build a model from a more diverse set of data, Hendelman et al (39) developed a second set of cut-points by supplementing the walking and jogging intervals with golfing, and basic household tasks. Swartz et al (68) developed a third alternative where only tasks of daily living such as yard work, walking while carrying a load, housework, and family care activities were measured. Each set of cut-points presents a different picture of how to analyze data acquired from the same piece of equipment. Since the studies used to develop cut-points involve fitting regression lines to all of the measured PA data, the regressions themselves can also be used to predict EE or METs. These early models tend to underestimate the specific MET costs of modes of PA relative to the compendium values in validation experiments (25).

Recognizing that the ability to predict EE from a linear regression model was not ideal for all activities, some researchers have begun to explore more sophisticated analysis approaches. One simple method to slightly increase model complexity, a bi-linear regression approach has been suggested, where sedentary activities were used to build one linear regression and locomotion activities were used to develop the second (38). While this modeling approach addresses the apparent differences in the relationship between counts and VO_2 in the low intensity domain relative to the moderate and vigorous ranges, the current two regression model is discontinuous at the point where the activities were divided. This increases the potential for large error in assigning energy costs to count values that lie close to the discontinuity, which could result from measurement differences across monitor units rather than actual intensity differences between the observations (29).

In a shift away from using a linear function, a power model has been proposed to predict EE_{ACT} from a tri-axial accelerometer (22). This model takes advantage of the additional acceleration information provided by the three acceleration measurements to improve prediction accuracy in the low-intensity region, where uni-axial accelerometers and linear models tend to perform poorly. Validation of this model on a small sample of women wearing the Cosmed K4B2 portable calorimeter showed that the model overestimated the energy costs of walking and jogging, though the total EE estimate was similar to the calorimeter (19). A model based on a quadratic function has been developed in a small population of adolescent girls performing 10 PA types of various intensities in a free-living environment (59). This model has improved performance relative to cut-point approaches for the sample on which it was developed, however, the

quadratic form is not generally considered to be physiologically reasonable as there is the potential for decay in EE assigned to high intensity intervals. To our knowledge, this model has not been validated by other researchers.

In our first study (Chapter III), we validated seven regression equations that have been developed for the ActiGraph, Actical, and RT3 PA monitors. All equations were developed using one-minute epochs and using populations of healthy adults, making them appropriate for comparison with our data. This work is, to our knowledge, the first time that these equations have been independently validated in 24 hour data sets in a large population of subjects asked to perform PA at self-selected paces. These experiments will provide insight into the relative strengths and weaknesses of each regression model.

Accelerometer Variability

In addition to challenges associated with relating accelerometer output to EE, a potential source of error in data acquisition and analysis using accelerometers are the technical specifications and performance of the accelerometer units themselves. Some accelerometer studies have revealed significant inter-monitor variability (15, 73) in clinical applications. This should raise concerns in research where multiple monitors are used for PA assessment and data are compared cross-sectionally or prospectively. The inherent variability introduced by subjects themselves is often confounded by variance introduced by the monitors in such trials. One approach to quantifying these differences is for a single subject to wear multiple monitors (50, 68, 78). But, the number of units

that can be worn and locations are always prohibitive factors in consistently representing monitor response between units of the same monitor type (ActiGraph, Actical, etc).

One approach to developing an understanding of the magnitude of inter-monitor variability is to perform mechanical oscillation experiments. Compared to human trials, mechanical oscillators have the potential to more fully characterize the function of accelerometry-based PA monitors because of the large number of accelerations that can be generated, the ability to record data from multiple monitors simultaneously, and the reproducibility of oscillations between trials. Several groups have used mechanical shakers to determine the inter- and intra- monitor variability in ActiGraph 7164 monitors, the most commonly used portable PA monitor. Fairweather et al developed a mechanical shaker to determine the variability of four ActiGraph monitors at a single acceleration value (30). This study showed high inter-instrument correlations in pair-wise comparisons between the monitors but the results are limited both by a small sample size, and a single acceleration test point. Metcalf et al developed a testing apparatus that captured the response of 23 ActiGraph 7164 monitors to sinusoidal oscillations at two speed profiles (49) These experiments showed low intra- and inter- instrument coefficients of variability (CV) at both the moderate and fast speed. While this study has a larger sample size, it still does not provide a complete characterization of the sensor response to different accelerations. Using a rotational wheel apparatus, Brage et al expanded on this work by exploring the inter- and intra-monitor variability of six ActiGraph 7164 monitors subjected to 51 accelerations generated by modulating both the speed and radius of oscillation (14). This study showed low intra-monitor variability but indicated that the inter-monitor variability was large enough to justify individual monitor

calibration, or specific EE prediction equations for each PA monitor unit. Because of the diversity of tested accelerations, this study provides a paradigm for experiments demonstrating the complete range of response of accelerometry-based PA monitors to mechanical oscillations.

The RT3 (StayHealthy; Monrovia, CA) PA monitor has also been subjected to mechanical oscillation experiments (55). In this study, 23 RT3 monitors were oscillated at three frequencies (2.1, 5.1, and 10 Hz). Difference in response between the three measurement axes as well as inter-monitor variability as a function of frequency of oscillation was explored. Inter-monitor variability decreased as frequency increases, but was larger in all cases than that of the ActiGraph 7164. The y-axis measurements were significantly different than the x and z axis responses with respect to total counts recorded at the two highest frequencies. This study was limited by the small number of frequencies tested, and the radius of oscillation of the shaker used (8.8 mm), which is smaller than the physiological displacement expected for many PA types.

Recently, a study was performed that compared the response of the ActiGraph 7164, the RT3, and a third activity monitor, the Actical (MiniMitter/Respironics; Bend, OR) (29). While these investigators looked at inter-monitor variability in large sample sizes of monitors, only six accelerations were tested (three by modifying radius, and three by modifying frequency), and therefore, no conclusions could be drawn about the overall response profiles of each device, or the relationship between the three monitor types.

While the response pattern and variability of the ActiGraph 7164 have been explored and some data have been collected using the RT3 and Actical, to our knowledge, no researchers to date have performed comprehensive analyses on newer

31

models of the ActiGraph, such as the 71256 and the GT1M, or developed complete response profiles for the RT3 and Actical monitors. It would also be desirable to determine if relationships exist between the monitor types, which might allow results to be compared across monitor types. Explorations in this area could lead to significant improvement in model development by reducing the error in EE_{ACT} models caused by inter-monitor variability, as well as by suggesting populations of subjects or modes of PA where a monitor's hardware filters are likely to limit the accuracy of the response.

In Chapter IV we undertook an extensive series of oscillation experiments to characterize the filtering of three generations of ActiGraph accelerometers. This work helps us to understand sources of error in our measurements that may be directly attributable to the instrumentation, particularly those errors that may result from introducing new equipment types during the course of a long term clinical investigation. Since clinical studies can easily span years, and equipment may be lost or damaged during this time, introducing new equipment is almost unavoidable. Further, through experiments in Appendix B where we compared responses between several PA monitor types we gained further insight into the differences in the meaning of activity counts across monitor types.

Alternate Modeling Strategies

In spite of a large body of work dedicated to the development of relationships between activity monitor output and EE, the major drawback with using accelerometers in the field to date has been in developing predictive models for METs or EE that generalize well to new subject populations and activity types. While linear regression

analysis is simple, the functional form of the model is not ideal, and the outcome of the modeling effort is highly dependant on the subject population and activity set used for model development, with large standard deviations in the generalization performance of group models. More complex models have not yet been widely accepted because of the required equipment, data dimensions, size and diversity of subject population needed for model development and testing, and technical hurdles associated with implementing the modeling techniques. However, recent work suggests that investigators are increasingly willing to explore more complex modeling approaches if higher predictive accuracy can be demonstrated.

Recently, a model has been developed using data collected in 1-second epochs. Data are grouped into two classes based on the CV in 10-second data intervals (26). An exponential regression was performed where CV ≤ 10, and a cubic regression was performed in all other data. MET values associated with individual PA types were improved using this approach. This approach capitalizes on technical improvements in the accelerometers which allow data to be collected in shorter epochs for longer periods of time. Results of this study indicated improvement relative to linear regression models developed using the ActiGraph; however, large errors were still detected for some individuals.

Additionally, a classification scheme using quadratic discriminant analysis and hidden Markov models has been proposed as a proof-of-concept to identify specific PA types (53). Though the discrimination accuracy in the model was high (< 90%), the model was developed on four activity types (rest, computer work, walking at a set pace, and vacuuming) and data were only collected from a small number of subjects (n=6). In

spite of these limitations, models developed using clustering techniques and other nonlinear models suggest that researchers are aware of a need for more robust data acquisition and modeling approaches to improve prediction of type, duration, and intensity of PA.

Building on existing modeling approaches, which have shown benefits from using multi-dimensional (53), nonlinear (22), and high speed (26) accelerometer data, we have established design requirements for development of a new model relating acceleration and EE. Because there is no single "normal" metabolic response, a candidate modeling approach which allows for multi-dimensional data input and allows full interactions between acceleration and subject characteristics may improve prediction of both PA intensity and minute-by-minute EE. Predictions should be flexible, allowing the same acceleration inputs, when matched with different personal characteristics to elicit different predictions, and should be robust in the face of noise introduced both by the acceleration measurement of the device and inter-monitor variability. Machine learning approaches fit these requirements well.

Artificial neural network (ANN) is an information-processing paradigm inspired by the way the densely interconnected, parallel structure of the mammalian brain processes information (37). ANNs are collections of mathematical nodes that emulate some of the observed properties of biological nervous systems and draw on the analogies of adaptive biological learning.
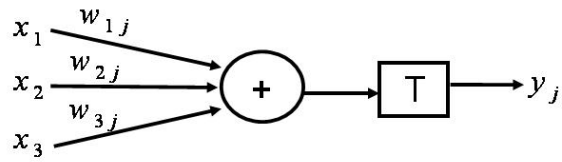
Variable	Definition	Meaning
x	Input	Characteristics of acceleration signal
w	Weight	Relative importance of each input
T	Transfer Function	Scales input
Y	Output	Predicted output (EE, MET, PA Type, etc)

Figure 3: Schematic of a single "artificial neuron" with parameters defined for the EE prediction problem.

The key element of the ANN paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements that are analogous to neurons and are tied together with weighted connections that are analogous to synapses. Each of these artificial neurons requires information inputs, weighting coefficients, transfer functions. Using these pieces of information the neuron can produce an output reflective of the current information supplied to it (Figure 3).

In a complex ANN model, multiple processing units are organized into layers allowing for parallel processing and improving computational efficiency (Figure 4). These connections also allow for interactions between neurons and nonlinearities to be introduced. In a fully connected network, the weight values between neurons of a layer

are all allowed to influence the inputs to the next layer. This allows all layer full

interactions in the model.
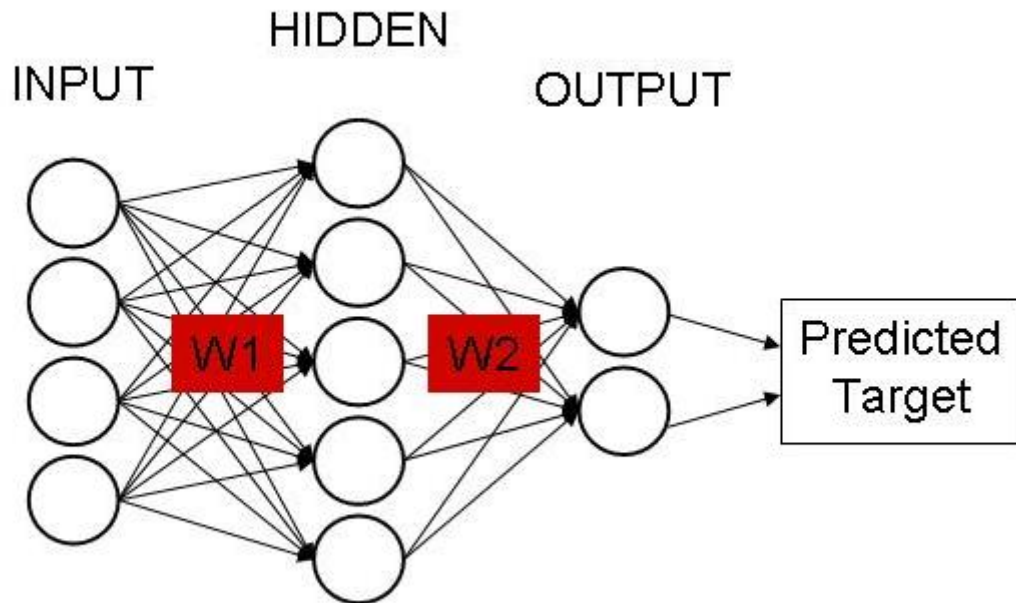
Figure 4: Schematic of multi-layer ANN. In this diagram, the circles labeled "HIDDEN" and "OUTPUT" reflects the summation of layer weights and transformation by the basis functions.

Learning in biological systems involves adjustments to the synaptic connections

that exist between neurons. ANNs also operate in an adaptive fashion. Supervised

learning occurs by example through training, or exposure to a set of input/output data

where the training algorithm iteratively adjusts the connection weights until a pre-

determined acceptable agreement is reached between model-predictive output (e.g.,

predicted EE) and the true output (e.g., measured EE). Thus, these connection weights store the knowledge necessary to solve specific problems.

Although ANNs have been around since the late 1950's, it wasn't until the mid-1980's that algorithms became widely adopted modeling tools (37). Today ANNs are being applied to an increasing number of real-world problems of considerable complexity. They are good pattern recognition engines and robust classifiers, with the ability to generalize in making decisions about noisy input data. They offer solutions to a variety of classification problems such as speech, character and signal recognition (11), as well as function approximation and system modeling where the physical processes are not understood or are highly complex, as is the case in metabolic modeling. The advantage of ANNs lies in their resilience against distortions in the input data and their capability of learning (37). They are often good at solving problems that are too complex for conventional technologies (e.g., problems that do not have a closed form solution or for which a closed form solution is too complex to be found) and are often well suited to problems that people are good at solving, but for which traditional analytic methods are not.

In the case of modeling the metabolic response to exercise, ANN models offer several advantages. First of all, a large number of subject characteristics may need to be combined with movement parameters in order to appropriately assess the magnitude of the expected metabolic response. ANN will allow the appropriate interactions to be emphasized, while reducing the impact of those that do not make a large contribution. Additionally, relative to standard least squares linear regressions, ANN models should be able to achieve better generalization since the higher number of degrees of freedom allow

the model to be more flexible in developing an appropriate model fit for data similar to that used for model training. ANN models can also avoid some problems with over-fitting by terminating model training when error from a validation set begins to increase even if the training error is still decreasing. Finally, by not specifying a functional form *a priori*, bias introduced by any distributional assumptions is minimized and an appropriate model form for the data is realized.

In the field of PA research, ANN models have been applied to gait analysis problems identifying the phase of gait as well as predicting joint angle parameters (20) . Additionally, speed and incline of unconstrained walking were predicted using accelerometer data and an ANN trained using data acquired from treadmill walking (6). To our knowledge, however, ANN approaches have not been used to directly predict EE from acceleration signals.

To test the capabilities of the ANN approach to improve the prediction of EE, we designed an ANN (Chapter V) using raw data from two acceleration channels at the hip and compared the EE predictions and the resulting errors to those from traditional accelerometer models. Additionally, we compared these predictions to those made using the IDEEA monitor, which has access to data from five sites. In Appendix D we demonstrate an ANN using only acceleration signal components, which is used to highlight the importance of subject characteristics in model development, while showing that minute-by-minute validation errors are still lower than those observed with an integrated acceleration response. In Appendix E we demonstrate the capability to develop ANNs on a subject by subject basis, which could serve to further reduce observed errors in EE prediction.

Statement of Purpose

The purpose of this work is to demonstrate the utility of accelerometers for prediction of PA intensity, in particular, EE. This will be accomplished by first demonstrating the performance of first generation accelerometers coupled with existing regression models, by comparing EE predictions to measurements made in a room calorimeter. Resulting predictions will be explored to determine the types of prediction variables for which the monitors and regressions are successful as well as highlighting any systematic discrepancies in the predictions. We will then undertake an exploration of the hardware filtering built in to the same accelerometers to attempt to discern if prediction accuracy and measurement validity is impacted by the hardware filtering in each device. Finally we will pursue a novel analytic approach where raw (32 Hz) acceleration signals are used to drive an ANN model using data from hip sensors. Results from this model will be compared with existing regression approaches to determine the benefit that can be derived from adding additional information into the EE prediction process.

This work addresses several crucial gaps in our understanding of accelerometers and their ability to predict EE. By developing improved tools for quantitative measurement of PA, we will take a step towards being able to accurately predict EE in a free-living environment. This information helps us to understand the role of PA in energy balance and to tackle the epidemic clinical problem of obesity.

References

1.   Executive Summary of the Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults. *Arch Intern Med* 158: 1855-1867, 1998.

2.   http://www.cdc.gov/nchs/fastats/overwt.htm: Center for Disease Control and Prevention, 2006.

3.   **Achten J and Jeukendrup AE.** Heart rate monitoring: applications and limitations. *Sports medicine (Auckland, NZ* 33: 517-538, 2003.

4.   **Ainsworth BE, Bassett DR, Jr., Strath SJ, Swartz AM, O'Brien WL, Thompson RW, Jones DA, Macera CA, and Kimsey CD.** Comparison of three methods for measuring the time spent in physical activity. *Medicine and science in sports and exercise* 32: S457-464, 2000.

5.   **Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, O'Brien WL, Bassett DR, Jr., Schmitz KH, Emplaincourt PO, Jacobs DR, Jr., and Leon AS.** Compendium of physical activities: an update of activity codes and MET intensities. *Medicine and science in sports and exercise* 32: S498-504, 2000.

6.   **Aminian K, Robert P, Jequier E, and Schutz Y.** Incline, speed, and distance assessment during unconstrained walking. *Medicine and science in sports and exercise* 27: 226-234, 1995.

7.   **Bassett DR, Jr., Ainsworth BE, Leggett SR, Mathien CA, Main JA, Hunter DC, and Duncan GE.** Accuracy of five electronic pedometers for measuring distance walked. *Medicine and science in sports and exercise* 28: 1071-1077, 1996.

8.   **Bassey EJ, Dallosso HM, Fentem PH, Irving JM, and Patrick JM.** Validation of a simple mechanical accelerometer (pedometer) for the estimation of walking activity. *European journal of applied physiology and occupational physiology* 56: 323-330, 1987.

9.   **Benson K, Friedman L, Noda A, Wicks D, Wakabayashi E, and Yesavage J.** The measurement of sleep by actigraphy: direct comparison of 2 commercially available actigraphs in a nonclinical population. *Sleep* 27: 986-989, 2004.

10.   **Bingham SA, Goldberg GR, Coward WA, Prentice AM, and Cummings JH.** The effect of exercise and improved physical fitness on basal metabolic rate. *The British journal of nutrition* 61: 155-173, 1989.

11.   **Bishop C.** *Neural Networks for Pattern Recognition*. New York, NY: Oxford Publishing, 1995.

12.   **Brage S, Brage N, Ekelund U, Luan J, Franks PW, Froberg K, and Wareham NJ.** Effect of combined movement and heart rate monitor placement on physical activity estimates during treadmill locomotion and free-living. *European journal of applied physiology* 96: 517-524, 2006.

13.   **Brage S, Brage N, Franks PW, Ekelund U, and Wareham NJ.** Reliability and validity of the combined heart rate and movement sensor Actiheart. *European journal of clinical nutrition* 59: 561-570, 2005.

14.   **Brage S, Brage N, Wedderkopp N, and Froberg K.** Reliability and Validity of the Computer Science and Applications accelerometer in a mechanical setting. *Measurement in Physical Education and Exercise Science* 7: 101-119, 2003.

15.   **Brage S, Wedderkopp N, Franks PW, Andersen LB, and Froberg K.** Reexamination of validity and reliability of the CSA monitor in walking and running. *MedSciSports Exerc* 35: 1447-1454, 2003.

16.   **Brooks G, Fahey T, and Baldwin K.** Exercise Physiology: Human Bioenergetics and Its Applications. New York: MacGraw Hill, 2005, p. 617-645.

17.   **Brown CD, Higgins M, Donato KA, Rohde FC, Garrison R, Obarzanek E, Ernst ND, and Horan M.** Body mass index and the prevalence of hypertension and dyslipidemia. *Obesity research* 8: 605-619, 2000.

18.   **Brown SP MW, Eason JM.** *Exercise Physiology: Basis of Human Movemen in Health and Disease*: Lippincott, Williams, and Wilkins, 2006.

19.   **Campbell KL, Crocker PR, and McKenzie DC.** Field evaluation of energy expenditure in women using Tritrac accelerometers. *Medicine and science in sports and exercise* 34: 1667-1674, 2002.

20.   **Chau T.** A review of analytical techniques for gait data. Part 2: neural network and wavelet methods. *Gait & posture* 13: 102-120, 2001.

21.   **Chen KY and Bassett DR, Jr.** The technology of accelerometry-based activity monitors: current and future. *Medicine and science in sports and exercise* 37: S490-500, 2005.

22.   **Chen KY and Sun M.** Improving energy expenditure estimation by using a triaxial accelerometer. *J Appl Physiol* 83: 2112-2122, 1997.

23.   **Crouter SE, Albright C, and Bassett DR, Jr.** Accuracy of polar S410 heart rate monitor to estimate energy cost of exercise. *Medicine and science in sports and exercise* 36: 1433-1439, 2004.

24.   **Crouter SE, Antczak A, Hudak JR, DellaValle DM, and Haas JD.** Accuracy and reliability of the ParvoMedics TrueOne 2400 and MedGraphics VO2000 metabolic systems. *European journal of applied physiology* 98: 139-151, 2006.

25.   **Crouter SE, Churilla JR, and Bassett DR, Jr.** Estimating energy expenditure using accelerometers. *European journal of applied physiology* 98: 601-612, 2006.

26.   **Crouter SE, Clowers KG, and Bassett DR, Jr.** A novel method for using accelerometer data to predict energy expenditure. *J Appl Physiol* 100: 1324-1331, 2006.

27.   **Crouter SE, Schneider PL, and Bassett DR, Jr.** Spring-levered versus piezo-electric pedometer accuracy in overweight and obese adults. *Medicine and science in sports and exercise* 37: 1673-1679, 2005.

28.   **Duffield R, Dawson B, Pinnington HC, and Wong P.** Accuracy and reliability of a Cosmed K4b2 portable gas analysis system. *Journal of science and medicine in sport / Sports Medicine Australia* 7: 11-22, 2004.

29.   **Esliger DW and Tremblay MS.** Technical reliability assessment of three accelerometer models in a mechanical setup. *Medicine and science in sports and exercise* 38: 2173-2181, 2006.

30.   **Fairweather SC, Reilly JJ, Grant S, Whittaker A, and Paton JY.** Using the computer science and applications (CSA) activity monitor in preschool children. *Pediatric Exercise Science* 11: 413-420, 1999.

31.   **Flegal KM.** Estimating the impact of obesity. *Sozial- und Praventivmedizin* 50: 73-74, 2005.

32.   **Flegal KM, Carroll MD, Ogden CL, and Johnson CL.** Prevalence and trends in obesity among US adults, 1999-2000. *Jama* 288: 1723-1727, 2002.

33.   **Frankenfield D, Roth-Yousey L, and Compher C.** Comparison of predictive equations for resting metabolic rate in healthy nonobese and obese adults: a systematic review. *Journal of the American Dietetic Association* 105: 775-789, 2005.

34.   **Freedson P, Pober D, and Janz KF.** Calibration of accelerometer output for children. *Medicine and science in sports and exercise* 37: S523-530, 2005.

35.   **Freedson PS, Melanson E, and Sirard J.** Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine and science in sports and exercise* 30: 777-781, 1998.

36.   **Gorny S and Allen R.** What is an activity count? A comparison of different methods used in wrist ActiGraphy. *Annual Meeting of the associated professional sleep societies*, Orlando, FL, 1999.

37.   **Haykin S.** *Neural Networks: A comprehensive foundation*. Upper Saddle River, NJ: Prentice Hall, 1999.

38.   **Heil DP.** Predicting activity energy expenditure using the Actical activity monitor. *Research quarterly for exercise and sport* 77: 64-80, 2006.

39.   **Hendelman D, Miller K, Baggett C, Debold E, and Freedson P.** Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Medicine and science in sports and exercise* 32: S442-449, 2000.

40.   **Hinton R, Moody RL, Davis AW, and Thomas SF.** Osteoarthritis: diagnosis and therapeutic considerations. *American family physician* 65: 841-848, 2002.

41.   **Jakicic JM, Marcus M, Gallagher KI, Randall C, Thomas E, Goss FL, and Robertson RJ.** Evaluation of the SenseWear Pro Armband to assess energy expenditure during exercise. *Medicine and science in sports and exercise* 36: 897-904, 2004.

42.   **Jequier E and Schutz Y.** Long-term measurements of energy expenditure in humans using a respiration chamber. *The American journal of clinical nutrition* 38: 989-998, 1983.

43.   **Kondepudi D and Prigogine I.** *Modern Thermodynamics*. New York, NY: Wiley, 1998.

44.   **Levine JA.** Non-exercise activity thermogenesis (NEAT). *Nutrition reviews* 62: S82-97, 2004.

45.   **Levine JA, Lanningham-Foster LM, McCrady SK, Krizan AC, Olson LR, Kane PH, Jensen MD, and Clark MM.** Interindividual variation in posture allocation: possible role in human obesity. *Science* 307: 584-586, 2005.

46.   **Matthews CE, Ainsworth BE, Hanby C, Pate RR, Addy C, Freedson PS, Jones DA, and Macera CA.** Development and testing of a short physical activity recall questionnaire. *Medicine and science in sports and exercise* 37: 986-994, 2005.

47.   **McClean JA TG.** *Animal and Human Calorimetry*. New York: Cambridge University Press, 1990.

48.   **McLaughlin JE, King GA, Howley ET, Bassett DR, Jr., and Ainsworth BE.** Validation of the COSMED K4 b2 portable metabolic system. *International journal of sports medicine* 22: 280-284, 2001.

49.   **Metcalf BS, Curnow JS, Evans C, Voss LD, and Wilkin TJ.** Technical reliability of the CSA activity monitor: The EarlyBird Study. *MedSciSports Exerc* 34: 1533-1537, 2002.

50.   **Nichols JF, Morgan CG, Chabot LE, Sallis JF, and Calfas KJ.** Assessment of physical activity with the Computer Science and Applications, Inc., accelerometer: laboratory versus field validation. *ResQExercSport* 71: 36-43, 2000.

51.   **Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, and Flegal KM.** Prevalence of overweight and obesity in the United States, 1999-2004. *Jama* 295: 1549-1555, 2006.

52.   **Ozkaya N and Nordin M.** *Fundementals of Biomechanics: Equilibrium, Motion, and Deformation.* New York, NY: Springer Science, 1999.

53.   **Pober DM, Staudenmayer J, Raphael C, and Freedson PS.** Development of novel techniques to classify physical activity mode using accelerometers. *Medicine and science in sports and exercise* 38: 1626-1634, 2006.

54.   **Poirier P, Giles TD, Bray GA, Hong Y, Stern JS, Pi-Sunyer FX, and Eckel RH.** Obesity and cardiovascular disease: pathophysiology, evaluation, and effect of weight loss. *Arteriosclerosis, thrombosis, and vascular biology* 26: 968-976, 2006.

55.   **Powell SM and Rowlands AV.** Intermonitor variability of the RT3 accelerometer during typical physical activities. *Medicine and science in sports and exercise* 36: 324-330, 2004.

56.   **Puyau MR, Adolph AL, Vohra FA, Zakeri I, and Butte NF.** Prediction of activity energy expenditure using accelerometers in children. *Medicine and science in sports and exercise* 36: 1625-1631, 2004.

57.   **Quesenberry CP, Jr., Caan B, and Jacobson A.** Obesity, health services use, and health care costs among members of a health maintenance organization. *Arch Intern Med* 158: 466-472, 1998.

58.   **Rowlands AV, Thomas PW, Eston RG, and Topping R.** Validation of the RT3 triaxial accelerometer for the assessment of physical activity. *Medicine and science in sports and exercise* 36: 518-524, 2004.

59.   **Schmitz KH, Treuth M, Hannan P, McMurray R, Ring KB, Catellier D, and Pate R.** Predicting energy expenditure from accelerometry counts in adolescent girls. *Medicine and science in sports and exercise* 37: 155-161, 2005.

60.   **Schoeller DA, Ravussin E, Schutz Y, Acheson KJ, Baertschi P, and Jequier E.** Energy expenditure by doubly labeled water: validation in humans and proposed calculation. *The American journal of physiology* 250: R823-830, 1986.

61.   **Seale JL and Rumpler WV.** Synchronous direct gradient layer and indirect room calorimetry. *J Appl Physiol* 83: 1775-1781, 1997.

62.	**Seale JL, Rumpler WV, and Moe PW.** Description of a direct-indirect room-sized calorimeter. *The American journal of physiology* 260: E306-320, 1991.

63.	**Sharma AM.** The obese patient with diabetes mellitus: from research targets to treatment options. *The American journal of medicine* 119: S17-23, 2006.

64.	**Speakman JR and Selman C.** Physical activity and resting metabolic rate. *The Proceedings of the Nutrition Society* 62: 621-634, 2003.

65.	**Starling RD.** Energy expenditure and aging: effects of physical activity. *International journal of sport nutrition and exercise metabolism* 11 Suppl: S208-217, 2001.

66.	**Strath SJ, Bassett DR, Jr., Ham SA, and Swartz AM.** Assessment of physical activity by telephone interview versus objective monitoring. *Medicine and science in sports and exercise* 35: 2112-2118, 2003.

67.	**Sun M, Reed GW, and Hill JO.** Modification of a whole room indirect calorimeter for measurement of rapid changes in energy expenditure. *J Appl Physiol* 76: 2686-2691, 1994.

68.	**Swartz AM, Strath SJ, Bassett DR, Jr., O'Brien WL, King GA, and Ainsworth BE.** Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Medicine and science in sports and exercise* 32: S450-456, 2000.

69.	**Tai MM, Castillo P, and Pi-Sunyer FX.** Meal size and frequency: effect on the thermic effect of food. *The American journal of clinical nutrition* 54: 783-787, 1991.

70.	**Tryon W and R W.** Fully proportional actigraphy: A new instrument. *Behavior Research Methods Instruments & Computers* 28: 392-403, 1996.

71.	**Wardle J and Cooke L.** The impact of obesity on psychological well-being. *Best practice & research* 19: 421-440, 2005.

72.	**Welk GJ, Blair SN, Wood K, Jones S, and Thompson RW.** A comparative evaluation of three accelerometry-based physical activity monitors. *Medicine and science in sports and exercise* 32: S489-497, 2000.

73.	**Welk GJ, Schaben JA, and Morrow JR, Jr.** Reliability of accelerometry-based activity monitors: a generalizability study. *MedSciSports Exerc* 36: 1637-1645, 2004.

74.	**Westerterp KR, Meijer GA, Janssen EM, Saris WH, and Ten Hoor F.** Long-term effect of physical activity on energy balance and body composition. *The British journal of nutrition* 68: 21-30, 1992.

75.   **Whaley M.** *ACSM's Guidelines for Exercise Presciption and Testing*. Baltimore, MD: American College of Sports Medicine, 2006.

76.   **Wickel EE and Eisenmann JC.** Within- and between-individual variability in estimated energy expenditure and habitual physical activity among young adults. *European journal of clinical nutrition* 60: 538-544, 2006.

77.   **Wickel EE, Welk GJ, and Eisenmann JC.** Concurrent validation of the Bouchard Diary with an accelerometry-based monitor. *Medicine and science in sports and exercise* 38: 373-379, 2006.

78.   **Yngve A, Nilsson A, Sjostrom M, and Ekelund U.** Effect of monitor placement and of activity setting on the MTI accelerometer output. *MedSciSports Exerc* 35: 320-326, 2003.

79.   **Zhang K, Pi-Sunyer FX, and Boozer CN.** Improving energy expenditure estimation for physical activity. *Medicine and science in sports and exercise* 36: 883-889, 2004.

80.   **Zhang K, Werner P, Sun M, Pi-Sunyer FX, and Boozer CN.** Measurement of human daily physical activity. *Obesity research* 11: 33-40, 2003.

CHAPTER III

VALIDITY OF PHYSICAL ACTIVITY INTENSITY PREDICTIONS BY
ACTIGRAPH, ACTICAL, AND RT3 ACCELEROMETERS

Abstract

Accelerometers are promising tools for characterizing physical activity (PA)

patterns in free-living persons. To date, validation of energy expenditure (EE) predictions

from accelerometers has been restricted to short laboratory or simulated free-living

protocols. This study seeks to determine the ability of seven regression equations for

three commercially available accelerometers, ActiGraph (three equations), Actical (two

equations), and RT3 (two equations), to predict summary measures of EE during 24-hour

stays in a room calorimeter in a diverse subject population (n=85). All accelerometers

and regression equations significantly underestimated ($p<0.05$) the daily physical activity

level. When data from the entire visit was divided into four intensity categories, light,

moderate, intense, and very intense, significant ($p<0.05$) over and under predictions were

detected in numerous regression equations and intensity categories, though in most cases

errors were less than 1%. Over almost all monitors and regressions, estimates of time

spent in each intensity category were reliable (differences less than 2% of measurement

interval); however, the regressions with physiologically reasonable intercepts always

underestimated the daily physical activity level. New regression equations should be

developed if accurate prediction of the daily PAL is desired.

Introduction

Physical activity (PA) is widely recognized as an important factor in maintaining healthy body weight. As the prevalence of obesity increases (15), increasing the daily PA level among adults has become an important public health priority. Several specific PA guidelines have been issued in an attempt to help individuals develop appropriate exercise habits. Both the Centers for Disease Control and Prevention and the American College of Sports Medicine (ACSM) have issued recommendations that adults perform moderate intensity PA for at least 30 minutes per day, five days a week (16, 23). Healthy People 2010 encourages adults to engage in at least three 20 minute bouts of vigorous PA each week (14). Some studies suggest that even these guidelines are insufficient to combat weight gain (3). Regardless of the specific PA goal adopted, it is important for researchers to objectively and accurately measure the actual daily participation in PA in order to understand the patterns of PA as well as to characterize the impact of achieving specific PA goals on overall health.

One common method for objective assessment of PA is accelerometry. Accelerometer output can be used to predict gross energy expenditure (EE) (6, 10, 22) or metabolic equivalents (METs) (9, 11, 20), which can be computed by normalizing EE by resting energy expenditure (REE). To simplify interpretation of accelerometer data, cut-points that distinguish intensity categories have been developed with descriptive names that correspond to those used in making public health predictions. Typical MET categories include light (1-3 METs), moderate (3-6 METs), intense/vigorous (6-9 METs), and very intense (> 9 METs) PA (23).

A number of different EE prediction equations exist in the literature to predict both METS and gross EE using minute-by-minute accelerometer output. Some of these equations have yielded cut-points, which serve to discriminate PA intensities without making specific EE predictions. All of these regressions are specific to a particular accelerometer device, such as the ActiGraph (9, 11, 20), the Actical (10), and the RT3 (6). Early equations were developed using only moderate-to-vigorous PA (9), while more recent approaches have incorporated lower intensity lifestyle activities, such as sweeping, house-cleaning, and gardening (10, 11, 20). A number of analytic approaches have been explored to develop robust prediction capabilities including linear regression (9-11, 20), bi-linear regression (10), and a nonlinear, power model (6). An extensive review of experiments designed to develop EE prediction equations and cut-points has recently been published by Matthews (13).

With numerous accelerometer devices on the market and multiple regression equations developed for each device, it is often difficult to select the device and regression equation that will be most appropriate for a specific study (5). Recently, a validation of three accelerometers, ActiGraph, Actical, and AMP-331, and fifteen prediction equations was performed on data acquired from short, structured protocols using portable indirect calorimetry (7), which showed overestimation of the metabolic cost of walking and sedentary activities, while underestimating the cost of most other activities tested. In the current study, we compared the predictive performance of three accelerometry-based PA monitors, ActiGraph, Actical, and RT3, and six EE prediction equations from the literature and one from the device manufacturer. We used EE measured by a room calorimeter during 24-hr periods in a heterogeneous group of

healthy adult volunteers as the reference criteria. Understanding the prediction accuracy of each monitor with respect to room calorimeter data provides an important intermediate step between EE estimates based on fully-structured laboratory protocols and free-living analyses.

Methods

Participants: Eighty-five adults (37 men, 48 women) between the ages of 18 and 70 years completed this study. Subjects were weight stable (< 2kg change in the last year), free of both diseases and medications known to alter EE, were non-smokers, and were free of major orthopedic problems that would limit their ability to perform PA. The characteristics of these subjects are shown in Table 1.

Table 1: Characteristics of study participants represented as median ~ inter-quartile range and (total range)

	All Subjects (n = 85)	**Men (n = 37)**	**Women (n = 48)**
Age (years)	40 ~ 21.0 (20 ~ 69)	40.0 ~ 21.0 (20 ~ 69)	39.0 ~ 22.0 (20 ~ 67)
Height (m)	1.70 ~ 0.13 (1.52 ~ 1.91)	1.77 ~ 0.10 (1.67 ~ 1.91)	1.64 ~ 0.08 (1.52 ~ 1.78)
Weight (kg)	72.0 ~ 22.0 (47.2 ~ 118)	81.5 ~ 19.0 (64 ~ 118)	63.1 ~ 15.2 (47.2 ~ 114)
BMI (kg / m^2)	24.6 ~ 5.15 (16.9 ~ 42.1)	25.4 ~ 3.8 (21.3 ~ 38.5)	22.9 ~ 5.7 (16.92 ~ 42.1)
% Body Fat	28.0 ~ 16.0 (6.7 ~ 57)	23.3 ~ 10.0 (6.7 ~ 45.1)	35.2 ~ 13.6 (11.7 ~ 57)

Experimental Procedures: Volunteers were recruited from the Nashville, TN area using flyers, email distribution lists, and personal contact. Before participation, all subjects

signed an informed consent document approved by the Vanderbilt University Committee for the Protection of Human Subjects. Each subject was asked to complete one overnight stay in the room calorimeter while minute-by-minute activity data was acquired with three hip-mounted accelerometers. Each subject was asked to engage in two structured activity intervals. The morning activity period was comprised of self-paced ambulatory activities, while the afternoon activity period contained sedentary activities, such as deskwork, along with stationary biking. Because hip-worn accelerometers were used, stationary biking was ultimately eliminated from the analysis. Each prescribed activity was performed for ten minutes followed by a ten minute rest period to allow the metabolic rate to return to baseline between activities. During times when no specific activity was prescribed (~ 18 hours), subjects were encouraged to engage in their normal daily physical activity patterns. Resting energy expenditure was computed using the mean of measured sleeping EE. Subjects' body composition was assessed in the week before their study visit using DEXA (GE Lunar Prodigy). Height, measured using a stadiometer, and weight, measured using a calibrated physician's scale, were measured the morning the subject entered the room calorimeter.

Instrumentation

Whole-room indirect calorimetry chamber: Energy-expenditure was computed on a minute-by-minute basis by the Vanderbilt University room calorimeter, which is located within the General Clinical Research Center. This system measures oxygen consumption and carbon dioxide production with high accuracy (system error < 1%). The room calorimeter is an airtight environmental room measuring 2.5 x 3.4 x 2.4 m. The

calorimeter is equipped with a toilet and sink, desk, chair, telephone, television, DVD

player, stereo system, bed, treadmill, and exercise bike. Technical details of the

calorimeter have been previously reported (19).

Accelerometers: During the study visit, subjects were simultaneously outfitted

with three commercially available accelerometers, the ActiGraph (formerly MTI/CSA;

Fort Walton Beach, FL), the Actical (MiniMitter/Respironics; Bend, OR), and the RT3

(StayHealthy; Monrovia, CA). Both the ActiGraph and Actical are primarily sensitive to

motion in one plane (vertical). The RT3 is a tri-axial accelerometer, which reports

activity in each of three orthogonal directions as well as the vector magnitude (VM) of

the three measurements.

Each of these monitors reports activity counts, a device-specific arbitrary unit,

which represents the frequency and amplitude of acceleration events occurring over a

user-defined measurement epoch. Technical specifications for each type of monitor have

been previously reported (10, 17, 21). For this study, all monitors were attached to a belt

secured at the waist with monitors positioned on the right hip, and all data was acquired

in one minute epochs.

Regression Equations: Activity count data for each monitor can be converted to

measures of PA intensity (EE or METS) using a variety of both published and proprietary

equations. Three regression equations were studied for the ActiGraph, while two

equations each were explored for the Actical and RT3 (Table 2). Data were analyzed

using both MET-based categorical predictions and using daily physical activity level

(PAL), which was computed as the average of the minute-by-minute MET prediction

during the study interval.

Table 2: Regression Equations for each PA monitor (ac = activity counts; AC = Actical PA monitor, and AG = ActiGraph PA monitor).

Equation Number	Developer	Activity Monitor	Regression Equation(s)
AC 1	*Heil[10]	Actical	AEE=(0.02779+1.143e-5*ac)*weight(kg)
AC 2	*Heil[10]	Actical	AEE=(0.01217+5.268e-5*ac)*weight(kg) 350≤ac≤1200 AEE=(0.02663+1.107e-5*ac)*weight(kg) ac ≥ 1200
AG 1	Freedson\ Work-Energy[1, 9]	ActiGraph	AEE=9.4e-4*ac+0.1346*weight(kg)-7.37418 ac≤ 1952 EE=1.91e-5*ac*weight(kg) ac>1952
AG 2	Hendelman[11]	ActiGraph	METS=2.922 + 4.09e-4*ac
AG 3	Swartz[20]	ActiGraph	METS=2.606 + 6.863e-4*ac
RT3 1	Stayhealthy	RT3	Proprietary
RT3 2	Chen[6]	RT3	AEE = (ac/76.2)$^{0.533}$ * (0.203 + 6.8e-3*weight(kg))

* ac ≤ 50: AEE = 0; 50 < ac ≤ 350: AEE = 0.007565*weight(kg)

For two equations, the Freedson equation for the ActiGraph (AG 1) and the Chen equation for the RT3 (RT3 2), adaptations were made to the originally published form to make them appropriate for our study data. The Freedson equation was developed specifically for moderate to vigorous intensity PA. The original model, assumed to be valid only for moderate to vigorous activity, was augmented using a formulation of the work-energy theorem presented in the ActiGraph instruction manual (Version 3.2) (1). In order to minimize the impact of this change on our categorical variable analysis, ActiGraph data was analyzed using published cut-points (Table 3) to categorize the data rather than continuous EE predictions when possible, though continuous EE predictions were required for the PAL analysis. For the Chen RT3 equation (RT3 2), which was

originally developed for the TriTrac-R3D accelerometer, a constant correction factor (17) was used to account for data collected with RT3 monitors.

Table 3: Activity Cut-points based on the EE prediction equations for each ActiGraph regression.

Regression	Light\Moderate Cut-point (3 METS)	Moderate/Vigorous Cut-point (6 METS)	Vigorous /Very Vigorous Cut-Point (9 METS)
AG 1	1952	5724	9498
AG 2	190.7	7525.7	14860.6
AG 3	574	4945	9317

Statistical Analysis: Both activity monitor types and specific regression equations were compared as to their ability to accurately predict daily PAL, and time spent in four standard PA intensity ranges, specified in METs. As the standard criteria, the calorimetry-measured PAL values were calculated on a minute-by-minute basis as the ratio between absolute EE and sleeping EE, which was the averaged EE over the entire sleeping duration. Data are presented as median and inter-quartile range for continuous variables, and as proportions for categorical variables. To test the null hypothesis that there is no difference in PAL and the percent of time spent in each intensity category between each regression equation and the calorimeter data, longitudinal analyses were conducted using GEE (12) to take into account the correlation among repeated measurements obtained from individual subjects. Analyses were performed using STATA 9.1 (StataCorp, College Station, TX), and R (www.r-project.org).

Results

PAL over the entire measurement period (21.7 ± 0.41 hours) was computed for

each subject using the room calorimeter EE as well as predicted EE using each activity

monitor and regression equation (Figure1). The median of the measured PAL values was

1.63, indicating that, on average subjects had an active day (PAL > 1.5). All PAL

predictions were significantly different (p < 0.001) from the calorimeter-measured values.

By examining the differences between the calorimeter and each prediction, the outcomes

can be ranked. The smallest magnitude difference occurs with the proprietary regression

equation for the RT3 monitor (RT3 1), which had an absolute difference of 0.07 from the

median daily PAL obtained using the calorimeter. The largest errors are observed when

using the Hendelman (AG 2) and Swartz (AG 3) regressions, which showed significantly

higher predictive PAL than other regressions considered. These large prediction errors

are attributable to their high y-intercept values (Table 2).
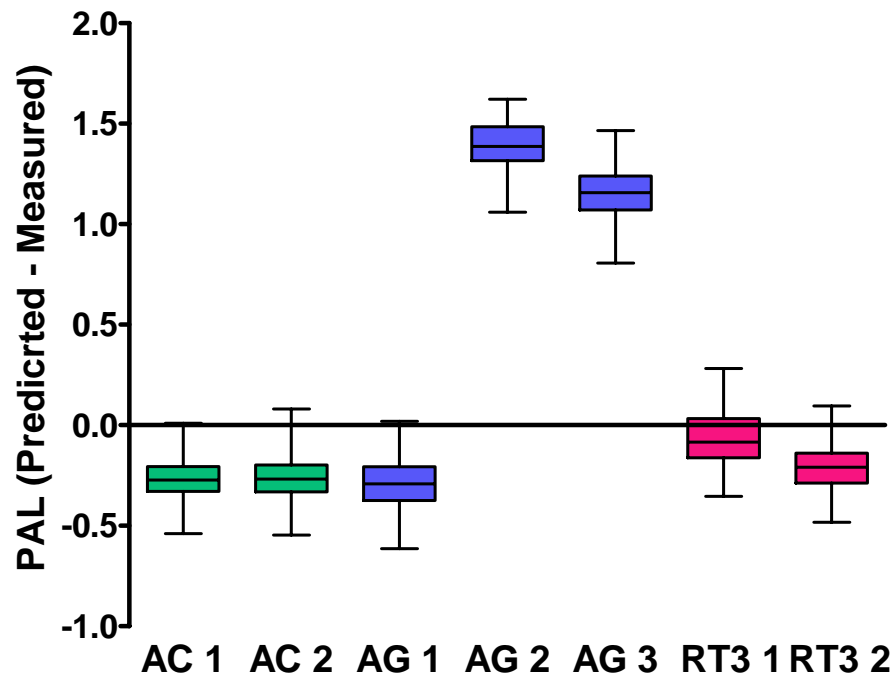
Figure 1: PAL relative to the calorimeter. Under-predictions fall below the x-axis, while over-predictions are above. The mean PAL measured by the calorimeter is 1.63. All PAL values are significantly different from the calorimeter ($p < 0.001$).

Data from each subject was also analyzed to determine the percent of the measurement period associated with each of four PA intensity categories, or MET ranges (Table 5), light (1-3 METs), moderate (3-6 METs), vigorous (6-9 METs), and very vigorous (> 9 METs). On average 95% of the study visit fell between 1-3 METs, 3.5% between 3-6 METs, and 0.75% spent in each of the 6-9 and > 9 MET classes.. The 1-3 MET category was generally well represented by the Actical, which showed no significant differences in either the percent of time measured with the calorimeter and predicted by the single (AC 1) or bi-linear (AC 2) regressions. Both RT3 equations showed ~ 1.5% over-prediction, while the Swartz equation (AG 3) showed ~ 1% under-

prediction. The largest prediction error occurred with the Hendelman cut-points (AG 2), where the mean under-prediction of time spent in light PA was close to 7%. All the models that under-predicted time spent in light activity subsequently over-predicted the time spent in moderate activity and vice versa.
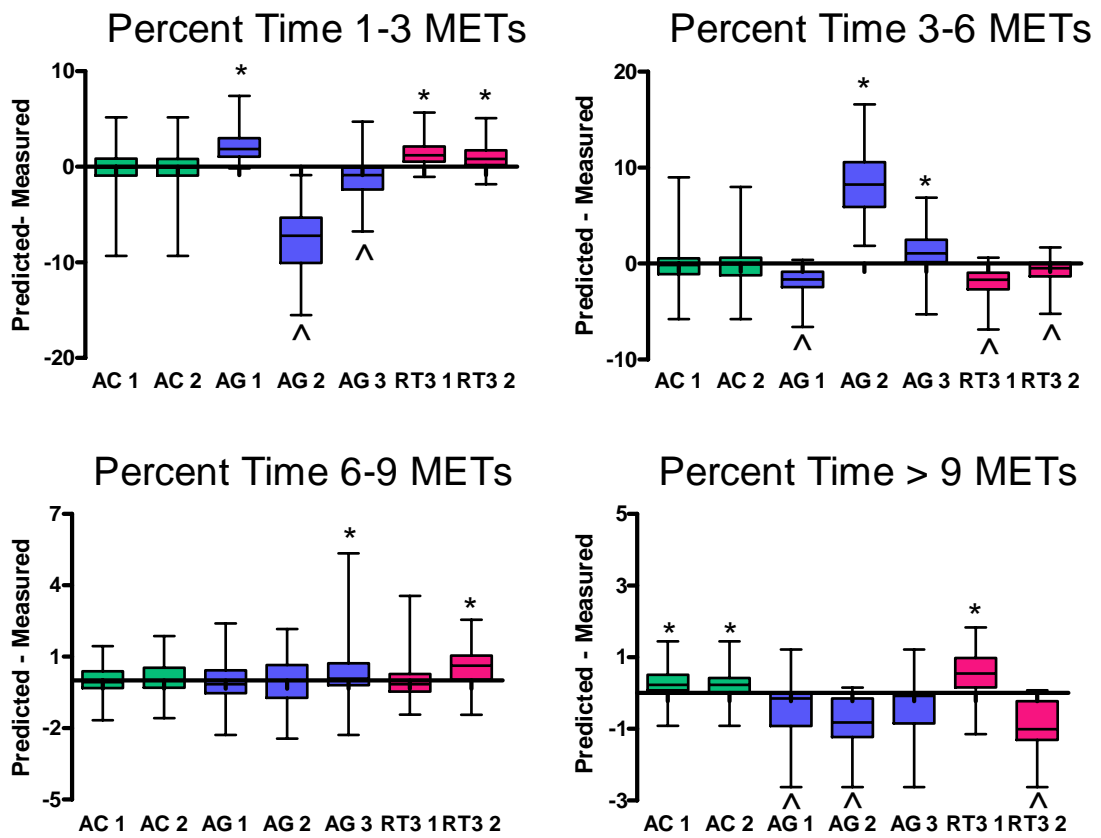
*Significant over-prediction (p<0.01); ^ Significant under-prediction (p<0.01)

Figure 2: Percent of study spent in four intensity classes represented as the difference from the calorimeter.

Discussion

Accelerometry-based portable physical activity monitors are a feasible and objective means of measuring physical activity patterns. Many studies have developed and validated models with various accelerometers to predict activity energy expenditures; however, to our knowledge, the scopes of such studies were mostly limited to short and structured intermittent protocols. In this study using a whole room indirect calorimeter, we validated the ability of the ActiGraph, Actical, and RT3 activity monitors to accurately report group summary statistics relating to time spent in specific PA intensity categories in a heterogeneous group of healthy men and women. Previously published regression equations for each device were explored to discover their relative strengths and weaknesses. The long study duration (~ 22 continuous hours) presents a bridge between short laboratory PA protocols, where all exercise intervals are explicitly specified, and free-living studies by allowing subjects to engage in both prescribed and spontaneous bouts of PA while still providing minute-by-minute correspondence with EE information from the room calorimeter. Analyses were designed to attempt to highlight features that would be of interest to researchers examining long-durations (weeks) of free-living data, where minute-by-minute prediction accuracy is less important than reliable summary measures of each day.

PAL is a measure of the mean EE above resting, which makes it an attractive measurement quantity since it is directly comparable between subjects since it is normalized to REE. Mathematically, accurate predictions of PAL require that intervals in which activity counts are close to zero be assigned an EE close to or equivalent to the REE. Thus, the Hendelman (AG 2) and Swartz (AG 3) regressions are poor choices for

predicting PAL because the large y-intercepts (2.922 and 2.606 METs respectively) cause PAL to be significantly over-predicted when subjects are at rest or engaged in low intensity activities (7). Using all other regressions, PAL was, on average, under-predicted which highlights potential limitations in the regression forms and also reflects that there are some increases in EE where motion was not detected by the accelerometers (thermic effect of food, limb movements, and isometric muscle contractions). The RT3 regressions performed most comparably to the calorimeter, with the best performance associated with the proprietary regression (RT3 1). The higher predicted PAL in both RT3 regressions could be due to the lower proportion of measured zeros by the RT3 (0.50) relative to the Actical (0.59) and ActiGraph (0.61), or could be due to characteristics of this regression, such as a slightly over-predicted baseline value. In the case of the proprietary RT3 regression, this is difficult to isolate the source, since the form of the regression is unknown.

Time spent in MET categories is a summary metric which characterizes the intensity distribution of daily PA, and is a useful tool for assessing whether a daily PA goal has been met in the field. While differences between predicted and measured distributions were generally small, some notable exceptions do occur. In our study, the regression with the largest prediction error was the Hendelman (AG 2) regression, displaying a large difference from the measured light intensity PA (Figure 2). This phenomenon has previously been observed when specific light and moderate PA modes were studied (2, 7) . The Hendelman regression (AG 2), along with the Chen RT3 regression (RT3 2) rarely predicts that any data belongs to the very intense category. If the number of classes is reduced to three by combining all measurements > 6 METS,

which is a frequent analysis construct, the Chen RT3 regression (RT3 2) has good agreement with the calorimeter data with mean time spent in the > 6 MET category predicted at 1.38% relative to the measured 1.66%. This type of correction does not provide the same benefit for the Hendelman regression (AG 2) for the 6-9 MET category, suggesting that there is a systematic bias in the intensity prediction towards moderate intensity PA using this regression. This follows logically from the cut-points (Table 3), where the range of counts associated with moderate activity, 7335 counts, is far wider than the 3772 counts assigned to moderate PA using the Freedson cut-points (AG 1), or the 4371 using the Swartz cut-points (AG 3). The pattern of prediction errors observed in this study agree with previously reported validation results using the ActiGraph regressions in shorter, structured protocols (2, 7).

There are some limitations in this study. First, we did not evaluate all predictive equations available for all the monitors we tested. We restricted our search to commonly used regressions, developed using one-minute epoch data, and chose to only examine one equation, the Freedson equation (AG 1) that only used walking and jogging in model development. Further, when possible we used equations that are built-in to activity monitor software, thereby attempting to isolate the equations that would be most accessible to researchers in the field. A new, nonlinear regression for the ActiGraph has recently been published (8), however, the data collection epoch was one second, and we were unable to validate its performance using this data set. Also, while we frequently referred to our predictions in terms of METs, they are more truly physical activity ratios (PAR) since each subjects' activity data was normalized by a measured REE (18). This difference could explain some discrepancies with regression equations developed using a

constant 3.5 ml O_2/kg/min as the normalization factor. However, there is some recent evidence that the constant normalization factor is not valid for all subjects (4) and PAR may be a more meaningful summary metric. We explored the impact of the value used for normalization by analyzing all of our data using an REE computed with the Harris-Benedict equation (approximately 10% higher than our measured baseline EE, which we took the average EE during the entire sleeping period). Resulting statistical trends for PAL and percent of time spent in each intensity category were unchanged.

It should also be noted that due to our large sample size, both with respect to number of subjects and duration of data collection, many statistically significant differences were detected that may result from absolute differences that are too small to be clinically relevant. Thus, each researcher must examine the magnitude of the difference between predicted and measured values to determine if observed differences are meaningful in the context of their work. Under prediction in PAL may be important, even if absolute differences are small, since a threshold value for an active day may not be predicted even if it is not achieved using these approaches. This was seen frequently in our data. For intensity categorizations, 1% difference between predicted and measured time in any PA category corresponds to approximately 15 minutes. This may not be significant in reporting total time spent in rest or light PA intensities, but may be very important in evaluating moderate to vigorous PA where the total time spent is expected to be relatively small and exercise recommendations frequently involve around 30 minutes of moderate to vigorous PA.

In this study we compared three commercially available accelerometery-based activity monitors and seven EE prediction equations with measured values using a room

indirect calorimeter. Mean PAL was under-predicted in five regressions with a zero or near zero intercept, while it was overestimated using two equations that did not force activity counts of zero to correspond to REE. Despite many performance similarities across monitor types and regressions, specific strengths and weaknesses were found for each, suggesting that no one equation or monitor is superior in all circumstances. Consequently, researchers should consider their outcome goal in determining not only the instrument they use to collect data but also their post-collection processing method. Since data can be safely analyzed using multiple regression approaches, researchers who are interested in more than one type of outcome may determine that more than one regression approach should be employed within a study in order to produce the highest accuracy results for each measurement variable of interest. Additional results from this data set can be found in Appendix B.

References

1.   **ActiGraph.** *Actisoft Analysis Software 3.2 User's Manual.* Fort Walton Beach, FL: MTI Health Services, 2005.

2.   **Bassett DR, Jr., Ainsworth BE, Swartz AM, Strath SJ, O'Brien WL, and King GA.** Validity of four motion sensors in measuring moderate intensity physical activity. *Medicine and science in sports and exercise* 32: S471-480, 2000.

3.   **Blair SN, LaMonte MJ, and Nichaman MZ.** The evolution of physical activity recommendations: how much is enough? *The American journal of clinical nutrition* 79: 913S-920S, 2004.

4.   **Byrne NM, Hills AP, Hunter GR, Weinsier RL, and Schutz Y.** Metabolic equivalent: one size does not fit all. *J Appl Physiol* 99: 1112-1119, 2005.

5.   **Chen KY and Bassett DR, Jr.** The technology of accelerometry-based activity monitors: current and future. *Medicine and science in sports and exercise* 37: S490-500, 2005.

6.   **Chen KY and Sun M.** Improving energy expenditure estimation by using a triaxial accelerometer. *J Appl Physiol* 83: 2112-2122, 1997.

7.   **Crouter SE, Churilla JR, and Bassett DR, Jr.** Estimating energy expenditure using accelerometers. *European journal of applied physiology* 98: 601-612, 2006.

8.   **Crouter SE, Clowers KG, and Bassett DR, Jr.** A novel method for using accelerometer data to predict energy expenditure. *J Appl Physiol* 100: 1324-1331, 2006.

9.   **Freedson PS, Melanson E, and Sirard J.** Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine and science in sports and exercise* 30: 777-781, 1998.

10.   **Heil DP.** Predicting activity energy expenditure using the Actical activity monitor. *Research quarterly for exercise and sport* 77: 64-80, 2006.

11.   **Hendelman D, Miller K, Baggett C, Debold E, and Freedson P.** Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Medicine and science in sports and exercise* 32: S442-449, 2000.

12.   **Liang K and Zeger S.** Longitudinal data analysis using generalized linear models. *Biometrika* 73: 13-22, 1986.

13.   **Matthew CE.** Calibration of accelerometer output for adults. *Medicine and science in sports and exercise* 37: S512-522, 2005.

14.   **McClean JA TG.** *Animal and Human Calorimetry*. New York: Cambridge University Press, 1990.

15.   **Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, and Flegal KM.** Prevalence of overweight and obesity in the United States, 1999-2004. *Jama* 295: 1549-1555, 2006.

16.   **Pate RR, Pratt M, Blair SN, Haskell WL, Macera CA, Bouchard C, Buchner D, Ettinger W, Heath GW, King AC, and et al.** Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *Jama* 273: 402-407, 1995.

17.   **Rowlands AV, Thomas PW, Eston RG, and Topping R.** Validation of the RT3 triaxial accelerometer for the assessment of physical activity. *Medicine and science in sports and exercise* 36: 518-524, 2004.

18.   **Schutz Y, Weinsier RL, and Hunter GR.** Assessment of free-living physical activity in humans: an overview of currently available and proposed new measures. *Obesity research* 9: 368-379, 2001.

19.   **Sun M, Reed GW, and Hill JO.** Modification of a whole room indirect calorimeter for measurement of rapid changes in energy expenditure. *J Appl Physiol* 76: 2686-2691, 1994.

20.   **Swartz AM, Strath SJ, Bassett DR, Jr., O'Brien WL, King GA, and Ainsworth BE.** Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Medicine and science in sports and exercise* 32: S450-456, 2000.

21.   **Tryon W and R W.** Fully proportional actigraphy: A new instrument. *Behavior Research Methods Instruments & Computers* 28: 392-403, 1996.

22.   **Welk GJ, Blair SN, Wood K, Jones S, and Thompson RW.** A comparative evaluation of three accelerometry-based physical activity monitors. *Medicine and science in sports and exercise* 32: S489-497, 2000.

23.   **Whaley Me and al. E.** *ACSM's Guidelines for Exercise Testing and Prescription*. Baltimore, MD: Lippincott, Williams, & Wilkins, 2006.

CHAPTER IV

COMPARING THE PERFORMANCE OF THREE GENERATIONS OF ACTIGRAPH

ACCELEROMETER

Abstract

ActiGraph accelerometers are a useful tool for objective assessment of physical

activity in clinical and epidemiological studies. Several generations of ActiGraph are

being used; however, little work has been done to ensure that measurements are

consistent across generations. This study employed mechanical oscillations to

characterize the dynamic response and inter-monitor variability of three generations of

ActiGraph monitors, from the oldest 7164 (n = 13), 71256 (n = 12), to the newest GT1M

(n = 12). The response due to independent radius (22.1–60.4 mm) and frequency (25–250

RPM) changes were measured, as well as inter-monitor variability within each

generation. The 7164 and 71256 have similar relationships between counts and radius (p

= 0.229) but were significantly different from the GT1M (p < 0.001). The counts vs.

frequency responses were nonlinear in all three generations. Although the patterns were

similar, the differences between generations at various frequencies were significant (p <

0.017), especially in the lower and higher measurement ranges. Inter-monitor variability

was markedly reduced in the GT1M compared to the 7164 and 71256. Other

measurement differences between generations include decreased peak counts and

sensitivity in low frequency detection in the GT1M. The results of this study revealed an

improvement of the inter-monitor variability by the newest GT1M. However, the reduced

sensitivity in low count ranges in the GT1M may not be well suited for monitoring sedentary or light intensity movements. Furthermore, the algorithms for energy expenditure predictions developed using older 7164 monitors may need to modified for the GT1M.

Introduction

Accelerometers are commonly used by researchers to objectively characterize the intensity and duration of physical activity in animals and humans. Due to their portability and data storage capacity, they are particularly useful in clinical and epidemiological studies in the fields of exercise, behavior, nutrition, and obesity. There are several accelerometry-based physical activity monitors commercially available, such as the Actical (Mini Mitter, Bend OR), ActiGraph (Fort Walton Beach, FL), and RT3 (Stayhealthy, Monrovia, CA). Of these, the most studied monitor is the ActiGraph accelerometer, formerly marketed under the Computer Science and Application (CSA) and Manufacture Technology Incorporated (MTI) names. Collectively, ActiGraph monitors have been in use for more than a decade and have undergone several hardware and software revisions.

Extensive work has been dedicated to the development of correlations between ActiGraph data and physiological criteria, such as oxygen consumption during walking and jogging in a laboratory setting (7, 9) Researchers have expanded these studies to include field data using portable indirect calorimeters as the validation criteria. (2, 8, 19) An extensive review of calibration experiments has recently been performed (18). Some studies revealed significant inter-monitor variability in CSA/MTI devices (4, 20). This

raises concerns in field applications where multiple monitors are used for physical activity assessment and data is compared cross-sectionally or prospectively. However, the inherent variability in monitors is often confounded by the variance introduced by the subjects themselves in such validation trials. Even in the case of a single subject wearing multiple monitors (11, 16, 21), the number of units that can be worn and suitable locations are always prohibitive factors.

Compared to human trials, mechanical oscillators have several advantages, such as the large number of accelerations that can be generated, the ability to record data from multiple monitors simultaneously, and the reproducibility of oscillations between trials. Several groups have used mechanical shakers to determine the inter- and intra- monitor variability in ActiGraph 7164 (CSA) monitors. Fairweather et al developed a mechanical shaker to determine the variability between four ActiGraph monitors at a single acceleration value (6), and showed high inter-instrument correlations in pair-wise comparisons between the monitors. However, this study was limited both by a small sample size, and a single acceleration point. Metcalf et al (10) developed a testing apparatus that captured the response of 23 ActiGraph 7164 monitors undergoing sinusoidal oscillations at two speed profiles. These experiments showed low intra- and inter- instrument coefficients of variability (CV = [standard deviation / mean] · 100) at both the moderate and fast speed. While this study has a larger sample size, it does not provide a characterization of the sensor response over a wide range of accelerations. Using a rotational wheel apparatus, Brage et al (3) explored the inter- and intra-monitor variability of six ActiGraph 7164 monitors subjected to 51 accelerations generated by modulating both the frequency and radius of oscillation. This study showed low intra-

monitor variability but indicated that the inter-monitor variability was large enough to justify individual monitor calibration. This study provides a comprehensive approach to characterizing the response patterns of ActiGraph 7164 monitors. Recently, a new investigation was undertaken to compare the response of ActiGraph 7164 monitors with response profiles with response profiles of the Actical and RT3 monitors (5).

To our knowledge, no researchers to date have performed sensor characterization on newer models of the ActiGraph, such as the 71256 (marketed 1999 - 2005) and the GT1M (marketed in 2005), though it has recently been suggested as a potential source of monitor variability that should be explored (5). These analyses could provide insight into variations in the device hardware across monitor generations, which would impact the predictions of physical activity intensity or energy expenditure using devices other than the 7164 monitors. The purpose of this study is to investigate the inter- and intra-generation differences in count response and unit variability in three generations of ActiGraph monitors by applying a wide range of mechanically generated accelerations.

Methods

Instrumentation: The CSA/MTI ActiGraph monitors are small devices (5.1 x 4.1 x 1.5 cm for 7164 and 71256, 5.3 x 5.1 x 2.2 cm for GT1M) that can be worn at the hip, wrist, or ankle (12) which measure acceleration in the vertical plane using a uni-axial piezoelectric accelerometer. Detailed specifications of the hardware and a full description of how the monitor acquires and filters data have been described previously (17). Data output from the ActiGraph is captured in counts from the manufacturer's software (1), which are a measurement unit that accounts for the amplitude and frequency of

acceleration events over each sampling period. Resulting activity counts are then summed over a user-defined epoch, or measurement period. Data collected at different epoch lengths can be compared by multiplying the count output by the ratio of the two epochs (17).

We studied three generations of ActiGraph physical activity monitors: the 7164 (n = 13), the 71256 (n = 12), and the GT1M (n = 12). The 7164 monitors were borrowed from a field investigator, and were at least three years old at the start of the study. These devices were tested using both mechanical acceleration pulses and using the ActiGraph calibrator prior to use and determined to be functional. The 71256 monitors had been in limited use for approximately one year (no more than 10 study visits with each device), while the GT1M monitors were new when these experiments began. All 7164 and 71256 monitors were calibrated using an external calibration apparatus (Model CAL71) according to the manufacturer's guidelines before the experiments. At the conclusion of the experiments, all monitors were again evaluated with the calibrator and none was found to be outside the suggested calibration range. The GT1M monitors are not accommodated by the external calibrator, however, according to the manufacturer's guidelines external calibration is not required for this model.

To generate all accelerations in this study, we selected an orbital shaker (VWR, Catalog Number 57018-755) with a frequency range of $25 - 500$ (± 2) RPM ($0.42 - 8.3$ Hz). An orbital shaker produces a sinusoidal pattern of accelerations and is therefore a reasonable simulator of repetitive human movements such as gait. However, the original 19.05 mm oscillating radius is smaller than the radius of movement that would be expected during normal gait, and higher intensity activities such as jogging

(13). Thus, we modified the orbital shaker to include an adjustable radius ranging from

10.2 – 66.8 mm by adding an arm between the loading plate and the rotor of the shaker.

The rotational frequency capabilities, as measured by an electronic odometer, remain

unchanged after the modification. The loading plate was equipped with three elevated

wooden columns attached by right angle brackets to the loading plate of the mechanical

oscillator. The elevation of the columns allows monitors to be secured on all sides of the

monitor, ensuring fidelity of position within each experimental condition.
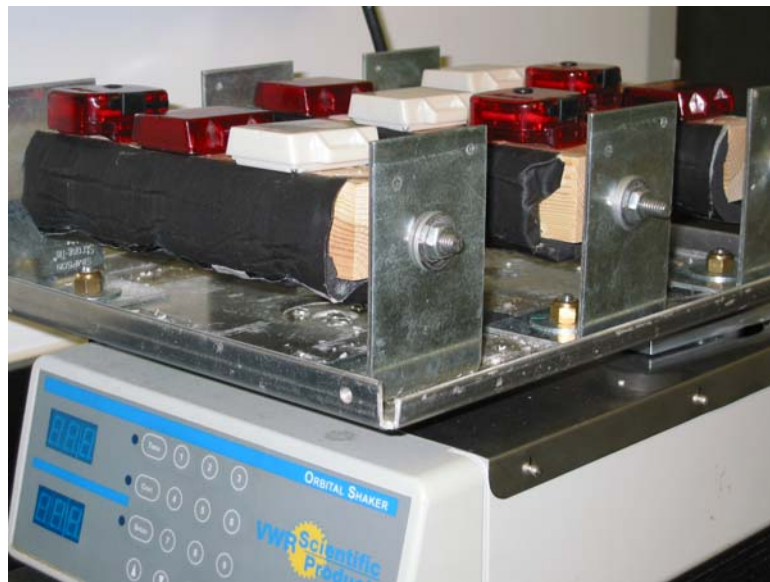
Figure 1: Orbital shaker with three generations of ActiGraph (3 each) placed in the nine
locations used for mechanical oscillation experiments. For experiments, the monitors
were secured to the surface with cable ties.

Experimental Procedure

Location experiment: The purpose of this first experiment was to determine the

uniformity of motion between locations on the loading plate. The elevated columns were

divided into nine positions large enough to accommodate an ActiGraph. Each position was verified for consistency of measurement using a batch of ActiGraph 71256 monitors (n = 9). Each monitor was positioned in all locations and rotated at 60 RPM with an epoch length of 10 seconds for two minutes.

Radius experiment: The purpose of this experiment was to measure the sensor output for five radius values, 22.1, 35.1, 41.6, 46.6, and 60.4 mm in different ActiGraph generations at a constant frequency (150 RPM). Monitors were divided into batches of nine units (Figure 1) from a mixture of generations, and were secured to the table throughout the radius changes. Epoch length was set to one minute, an epoch commonly used in field data collection. Activity counts were measured for six minutes with at least three minutes of rest between successive radii.

Frequency experiment: The purpose of this experiment was to determine the dynamic range of each generation of ActiGraph accelerometer. Monitors were again divided into batches of nine (mixed generations) and subjected to 21 different oscillation frequencies ranging from 25 – 250 RPM at a fixed radius of 46.6 mm. The low end of the frequency range was dictated by the limitations of our mechanical shaker. Although the published high frequency limit for the ActiGraph was 2.5 Hz (150 RPM), we selected the high end of 250 RPM (about 4 Hz) because this is within the range of human movements (14, 15). A two epoch length convention was developed where frequencies less than or equal to 100 RPM were measured using 10 second epochs and frequencies greater than 100 RPM were captured using one minute epochs, with measurements at 60, 80, and 100 RPM collected using both epoch lengths. Short epochs were used to distinguish between noise, isolated spikes of low count values, and continuous signal acquisition. For the

short epoch trials, two minutes of data was collected followed by a one minute rest period, while for the long epoch trials, data was collected for six minutes followed by at least three minutes of rest. Data collected using 10-second epochs were then converted (multiplied by six) into analogous one-minute count values. *Post hoc* analysis showed that the percent difference in counts that were converted from short epochs to long epochs at the three repeated measurement points was less than the intra-monitor CV for each monitor, which suggests that the use of two measurement epochs did not have a significant effect on the reported results.

For each experimental condition, the first and last data points were excluded from analysis in order to eliminate the possibility of transient recording errors associated with oscillator ramp-up/down phases. For the short epoch trials conducted as part of the frequency experiments, the first and last two data points were excluded from analysis.

In additional to present data in terms of frequency (RPM) in our experiments, average acceleration was computed using the following equation (Eq 1) which was presented by Brage et al (2) who used a similar sinusoidal oscillator in a previous study, where acceleration (A) is a function of both the frequency (f) and radius (r) of oscillation.

$$A = 8 \cdot \pi \cdot r \cdot f^2 \qquad \text{(Eq 1)}$$

Moreover, in an attempt to develop a linearly increasing relationship between counts and acceleration for all measured values, results of the frequency experiments were adjusted using the filter weighting (FW) function proposed by Brage et al (4) as the following.

$$FW = 0.06118 \cdot f^2 - 0.5573 \cdot f + 1.453 \qquad\qquad (Eq\ 2)$$

Statistical Analysis: For the location experiment, a non-parametric Kruskal-Wallis test was used to determine if location effects exist. To assess the relationship between radius and counts, linear regression analysis was performed on the output from each generation of monitors. The equality of slopes across generations was assessed by testing if the radius by type effect is equal to zero using an F-test. Contrasts were established to test whether the slope of the 7164 and 71256 are different and if the slope of the GT1M is different than the 7164 and 71256 combined. For the frequency experiments, pair-wise comparisons between any two monitor types were performed using exact Wilcoxon rank sum tests at each measurement frequency with a Bonferroni adjustment to control for multiple comparisons in the overall Type I error rate. The inter- and intra-monitor CV was computed at each frequency. Statistical analyses were performed using SAS for Windows 8.02 (SAS Institute; Cary NC) and R 2.01 (The R foundation; www.r-project.org).

Results

One ActiGraph GT1M monitor exhibited a non-zero response at 25 RPM. This monitor also showed a markedly greater count output at 30 RPM than the other GT1M monitors tested. This result was verified in multiple tests. Because of these differences, this monitor was excluded from all analyses leaving a sample of 11 GT1M monitor units.

Location Experiment

The location validation experiment showed no significant location effect on count values (Table 1) over all monitors ($p = 0.999$). This allowed monitors throughout the rest of the experiments to be tested regardless of their position on the loading plate.

Table 1: Location verification experiment - Mean and Standard deviation (counts/minute) from nine ActiGraph 71256 monitors at each of nine locations on the loading plate are shown.

593.4 ±16.5	594.3 ± 16.7	594.8 ± 17.1
593.9 ± 17.0	594.1 ± 17.6	593.9 ± 16.3
592.6 ±16.2	592.6 ± 16.5	592.6 ± 17.5

Radius Experiment

The mean counts were found to be correlated with radius for each generation of ActiGraph monitors at 150 RPM (Figure 2). Comparisons between groups of monitors reveal that the 7164 and 71256 monitors had similar slopes ($p = 0.229$), but are different from the GT1M ($p < 0.001$).
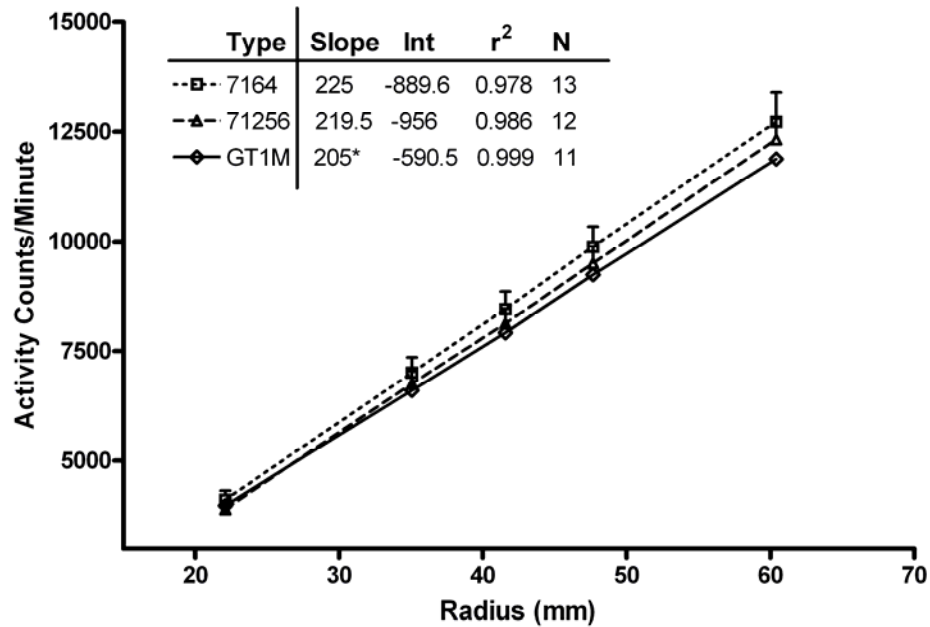
Figure 2: ActiGraph counts as a function of radius for each generation of monitor. This plot contains the mean of monitor means for each generation of monitor at each radius. The error bars represent one standard deviation.

Frequency Experiments

Figure 3 shows the accelerometer output (mean counts) as a function of frequency of oscillation and calculated average acceleration (Eq 1) for the three generations of ActiGraph monitors at a fixed radius. All three generations showed a similar curve shape – a linear region in the middle section, flanked by two nonlinear regions at lower and higher frequencies. At frequencies greater than 160 RPM, all three generations of ActiGraph monitors were significantly different from each other at each frequency tested ($p < 0.017$). In addition, there were several data points at which pair-wise comparisons showed significant differences. The GT1M monitors were significantly different from the 7164 monitors at all frequencies except 120 RPM, and from the 71256 monitors at all

frequencies except 120, 140, and 150 RPM. The 7164 and the 71256 monitors were

significantly different at 30, 35, 50, 60, 90 and 160 RPM.
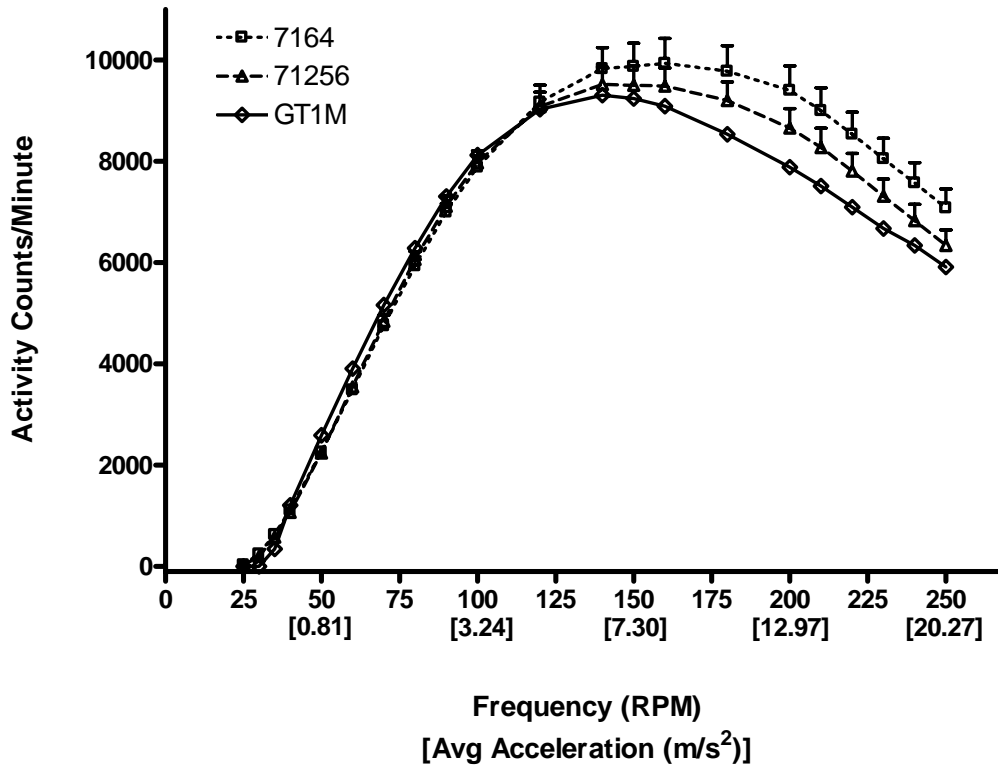
Figure 3: Plot of counts as a function of frequency and acceleration for each monitor
generation over all measured accelerations. Results are mean of all monitors plus one
standard deviation.

One important difference in the behavior of the three generations of ActiGraph

was observed at the lowest test frequencies (Figure 4). While the 7164 and 71256

monitors have a nearly identical response to these oscillations, the GT1M monitor

requires a larger acceleration to record a nonzero activity count response.
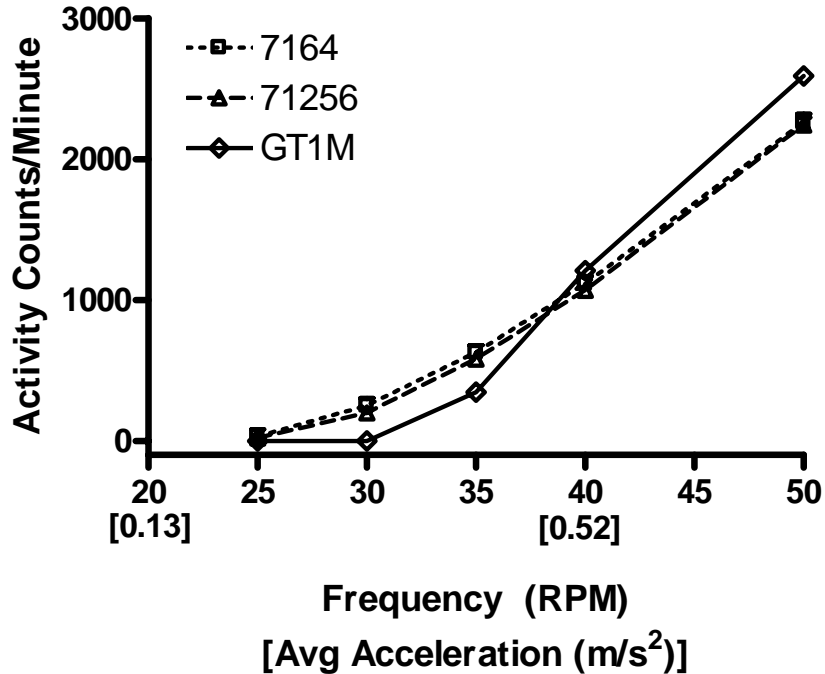
Figure 4: Activity counts as a function of frequency and acceleration for low acceleration values. The GT1M monitor requires a larger acceleration to record a non-zero count response.

Inter-monitor CV values were computed at all measured frequencies. The inter-monitor CV for the GT1M monitors is consistently lower than either the 7164 or 71256 monitors for frequencies greater than 40 RPM (Figure 5). In general, the CV is high for all generations at low frequencies (> 20%) because not all monitors of each generation have a non-zero response in this range (large standard deviation). The GT1M at 25 RPM was an exception because no signal was captured by any of the monitors (hence CV=0).

Intra-monitor CV was small in monitors of all generations, with values on the order of

0.55% for all frequencies greater than 40 RPM.
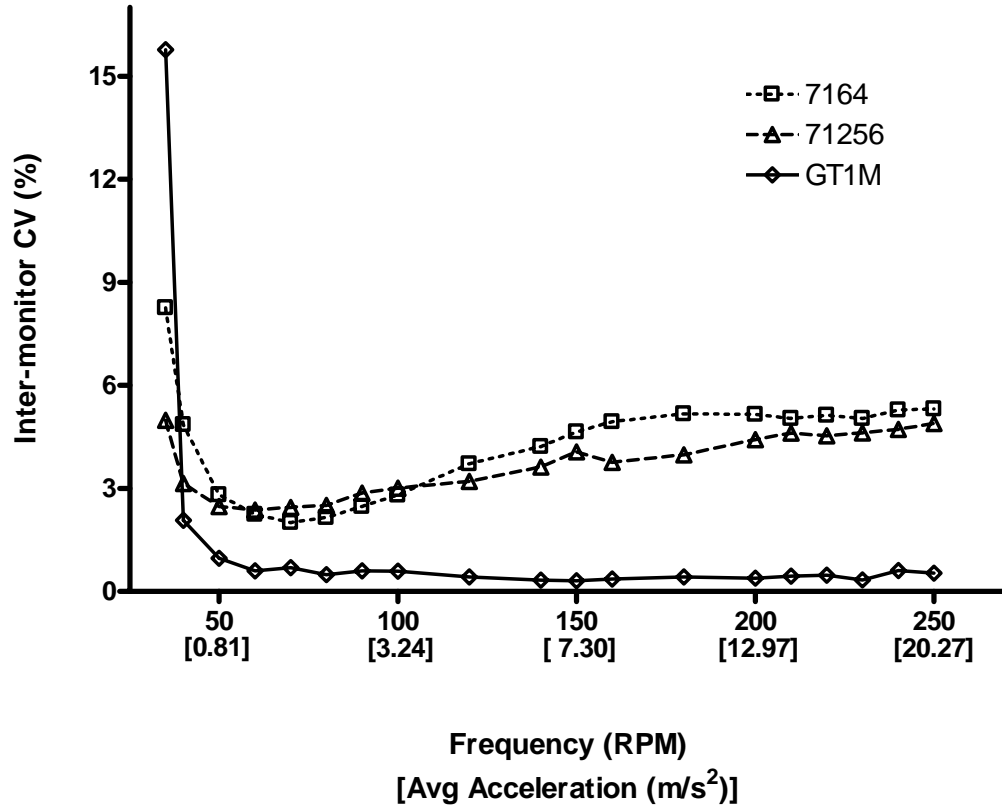
Figure 5: Inter-monitor CV (%) for all frequencies greater than 35 RPM for all three ActiGraph generations. Data from 25 and 30 RPM is not shown because the large CV (> 100%) observed at these frequencies does not well represent the trends over the measurement range.

The exact FW function proposed by Brage (Eq 2) was not adequate to remove all

the signal decay in the high frequency domain (Figure 6a) observed in these experiments.

Therefore, we developed a new filter function using voltage data previously published

(17) over a slightly expanded frequency range, 0.95 – 5 Hz. A least squares quadratic fit

on these data (Eq 3) yielded a regression that was able to correct for nearly all of our

observed count decay at high accelerations (Figure 6b).

$$FW = 0.0592 \cdot f^2 - 0.5493 \cdot f + 1.446 \qquad \text{(Eq 3)}$$

The same filter weighting function was applied to all generations of ActiGraph since, to

our knowledge, no voltage data have been published on the newer models.
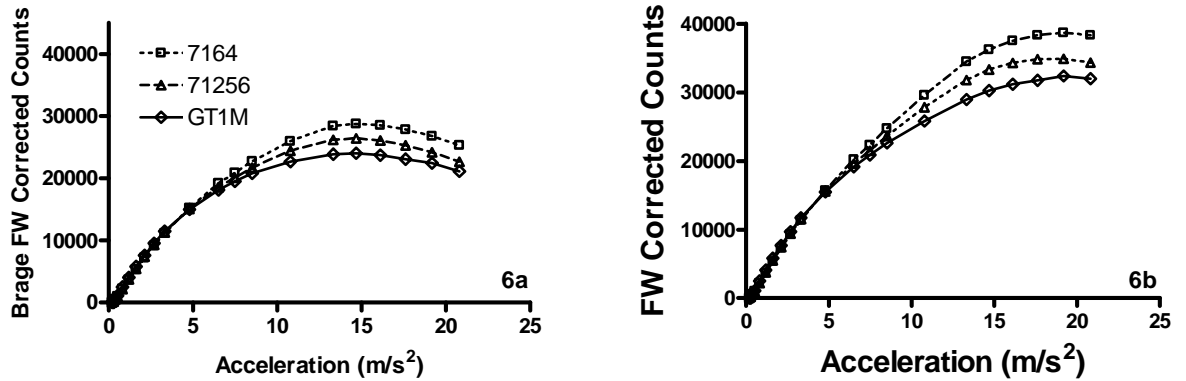
Figure 6: Filter-weight corrected data using quadratic filter weighting function for each generation of ActiGraph monitor. The left panel (6a) shows the results using the Brage weighting function, while the right panel (6b) shows the results using a slightly modified quadratic weighting function designed to attenuate count decay at high accelerations.

Discussion

The ActiGraph line of accelerometers are the most frequently used activity

monitors in the physical activity research field. In this study, we developed an orbital

shaker and measured multiple units of the ActiGraph of three different generations under various acceleration conditions produced by altering both the frequency and radius of oscillation. Although the ActiGraph responses to acceleration changes have similar shapes, statistically significant point differences were found between the three generations. Particularly, the newest monitors, the GT1M, had smaller inter-monitor CV, but lower sensitivity in the low frequency range and reduced amplitude at higher accelerations.

The radius analysis showed that there is a linear relationship between counts and radius for all measured values. This is an expected response based on the theory of rotational motion (Eq 1). However, the slope of the regression is significantly different in the GT1M monitors when compared to older monitors suggesting that there are some differences in either the device sensitivity, or the filtering scheme. To further explore this idea, experiments where acceleration was modulated by frequency changes at a constant radius were used to characterize the dynamic range and filtering of the count output.

Results from the frequency analysis showed that even though the general shape of the curves is similar, the GT1M count output is less than that of the other two monitor generations at low frequencies (<40 RPM), becomes greater for moderate frequencies (around 3-5% between 50 -120 RPM), and decays much more quickly after achieving a lower max count value in the high frequency domain (Figure 3). Although we originally wanted to derive correction factors between different ActiGraph generations, this non-linear relationship between the generational response profiles prohibits us from achieving this goal with a single continuous function. In addition to observations made over the

entire range of tested frequencies, statistically significant point differences in amplitude existed in many of the test conditions.

Using the Freedson equation (7) for predicting EE from activity counts, the observed difference in EE prediction for counts less than 4000 (assumed to be those most frequently attained by subjects) changes in predictions between the 7164 and each of the newer monitor generations were computed. Differences between the 7164 and 71256 showed a consistent reduction in EE prediction of approximately 0.05 kcal/min. Comparisons between the 7164 and GT1M EE predictions ranged from an under-prediction of -0.30 kcals/min at 30 RPM (counts ~ 300) to over-estimation of 0.40 kcal/min at 60 RPM (counts ~ 3500). Differences between 7164 and 71256 monitors are therefore unlikely to be clinically significant; however differences between the 7164 and GT1M predictions could lead to large errors in daily or weekly EE predictions.

Our results from the frequency experiments clearly revealed that a significant improvement in the new GT1M monitor is a reduction in inter-monitor variability. While inter-monitor CV for the 7164 and 71256 was on the order of 3–5 % over all frequencies tested, the GT1M had values of less than 1 % at all frequencies greater than 40 RPM. We conducted additional experiments to reduce the inter-monitor CV in the 7164 and 71256 monitors by using the external calibrator to specify a specific gain value (data not shown). However, even when a specific gain value is specified, the inter-monitor CV of the 7164 and 71256 monitors tested (1.56-5.96%) did not achieve CV as low as that observed in the GT1M, and the process of obtaining this level of control over the calibration factor is time intensive. As such, the ActiGraph GT1M appears to be a more consistent measurement tool than previous ActiGraphs, which could reduce measurement

variability in studies where a large number of physical activity monitor units will be used and data are compared prospectively or cross-sectionally.

The shape of all three of our observed acceleration response curves was similar to the curve shown by Brage et al (4) for the 7164 monitor. However, in order to generate a corrected curve for our data using the previously proposed filter weight correction scheme, a slightly altered weighting function was developed to accommodate our observed decay in high frequency data, which appears to have a slightly larger magnitude than previously observed. This could be due to differences in our mechanical testing devices, differences between the batches of monitors tested, and/or the slightly increased dynamic range used for data collection. Regardless of the exact filter weighting coefficients, this analysis supports earlier findings that the shape of the data response can be modified analytically using a quadratic regression to linearize the response profile and generate a monotonically increasing function (Figure 6). The same weighting function appears to produce similar results across ActiGraph generations. This approach, however, has limited relevance in the field where changes in acceleration can be caused by changes in either the amplitude (radius) or the frequency of oscillation. Because the underlying physical cause of the acceleration change would be unknown, application of a frequency-based correction factor would not be well motivated. Even if changes in acceleration in the field could be linked to specific physical changes, this filter weight correction amplifies the magnitude of the count response, which limits our ability to use existing cut-points or energy expenditure prediction equations without either scaling the raw data or the coefficients in existing analysis routines.

Even though we detected many statistically significant differences in the frequency data, it is important for field researchers to consider how much these differences will impact their research findings. When working with largely sedentary populations, the magnitude of count differences between device generations are expected to be small and the amount of error introduced may be acceptable. However, the higher threshold for non-zero count reporting exhibited by the GT1M may increase the challenge of detecting and discriminating sedentary or light intensity activities; however this threshold should reduce the possibility of that noise would be reported as signal using these devices. Researchers who are using cut-points to divide their data into intensity categories may not see a large change in their outcome using GT1M monitors because only a small number of data points, those with count values similar to the cut-points, are expected to be re-classified if data from GT1M monitors is partitioned with equations developed for the 7164. Differences may be more important if energy expenditure is the desired outcome. In this case, any difference in count value would alter the minute-by-minute energy expenditure prediction. Over the course of a day, small minute-by-minute differences could translate into large differences in total energy expenditure predicted. In such cases, it would be desirable to develop new equations using count data from GT1M monitors. Additionally, researchers who are interested in studying children, who tend to exhibit bursts of high intensity activity, or investigators working with other highly active populations, may be concerned about the lower peak count value for the GT1M monitors.

There are several limitations in our study. While the 71256 and GT1M monitors were purchased in batches and had similar prior use patterns among units of a generation, the 7164 monitors were borrowed from a field investigator. We have no knowledge of

83

their age, prior use, or service history. Since this study did not incorporate a time history component, it is unclear if these factors introduce significant error. The calibration process should minimize sensor drift over time and should identify bad sensors (since their count response would be outside the range of expected output for the calibrator), making it an important check point in the process of evaluating equipment. However, this process is not comprehensive (all possible accelerations are not tested) and some age and wear related issues could remain which may change the response of the monitors. These could explain some of the small changes between 7164 and 71256 monitors we observed. The sample size in our study is also a consideration. While it would be desirable to have larger batches of monitors, our sample sizes were constrained by the practical limitation of the number of monitors we had available on-site and through collaborators.

In conclusion, this study demonstrates the response of three generations of ActiGraph physical activity monitors under a wide range of mechanically produced acceleration conditions. We demonstrated a reduction in inter-monitor variability in the new ActiGraph GT1M monitors, which reduces the error introduced into analyses due to multiple monitor units and represents a significant improvement over previous ActiGraph monitors. Our results also indicated that significant differences in activity count outputs exist between the generations of monitors at numerous acceleration values. Differences between 7164 and GT1M monitors can lead to substantial differences in the prediction of daily EE. However, analyses relying on cut-points will likely only feel a minimal impact of these differences between monitor generations. Differences between 7164 and 71256 monitors are unlikely to be clinically relevant. Ultimately, the end goal of each researcher must be considered when interpreting whether these count differences will introduce an

unacceptable level of error into data analysis. Differences observed in a mechanical

setting should be verified in controlled human experiments to better understand the

impact inter-generational differences have on analysis using ActiGraph monitors.

Additional oscillation results can be found in Appendix C.

References

1.   **ActiGraph.** *Actisoft Analysis Software 3.2 User's Manual*. Fort Walton Beach, FL: MTI Health Services, 2005.

2.   **Bassett DR, Jr., Ainsworth BE, Swartz AM, Strath SJ, O'Brien WL, and King GA.** Validity of four motion sensors in measuring moderate intensity physical activity. *Medicine and science in sports and exercise* 32: S471-480, 2000.

3.   **Brage S, Brage N, Franks PW, Ekelund U, and Wareham NJ.** Reliability and validity of the combined heart rate and movement sensor Actiheart. *European journal of clinical nutrition* 59: 561-570, 2005.

4.   **Brage S, Brage N, Wedderkopp N, and Froberg K.** Reliability and Validity of the Computer Science and Applications accelerometer in a mechanical setting. *Measurement in Physical Education and Exercise Science* 7: 101-119, 2003.

5.   **Esliger DW and Tremblay MS.** Technical reliability assessment of three accelerometer models in a mechanical setup. *Medicine and science in sports and exercise* 38: 2173-2181, 2006.

6.   **Fairweather SC, Reilly JJ, Grant S, Whittaker A, and Paton JY.** Using the computer science and applications (CSA) activity monitor in preschool children. *Pediatric Exercise Science* 11: 413-420, 1999.

7.   **Freedson PS, Melanson E, and Sirard J.** Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine and science in sports and exercise* 30: 777-781, 1998.

8.   **Hendelman D, Miller K, Baggett C, Debold E, and Freedson P.** Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Medicine and science in sports and exercise* 32: S442-449, 2000.

9.   **Melanson EL, Jr. and Freedson PS.** Validity of the Computer Science and Applications, Inc. (CSA) activity monitor. *MedSciSports Exerc* 27: 934-940, 1995.

10.   **Metcalf BS, Curnow JS, Evans C, Voss LD, and Wilkin TJ.** Technical reliability of the CSA activity monitor: The EarlyBird Study. *MedSciSports Exerc* 34: 1533-1537, 2002.

11.   **Nichols JF, Morgan CG, Chabot LE, Sallis JF, and Calfas KJ.** Assessment of physical activity with the Computer Science and Applications, Inc., accelerometer: laboratory versus field validation. *Research quarterly for exercise and sport* 71: 36-43, 2000.

12.   **Nichols JF, Morgan CG, Chabot LE, Sallis JF, and Calfas KJ.** Assessment of physical activity with the Computer Science and Applications, Inc., accelerometer: laboratory versus field validation. *ResQExercSport* 71: 36-43, 2000.

13.   **Orendurff MS, Segal AD, Klute GK, Berge JS, Rohr ES, and Kadel NJ.** The effect of walking speed on center of mass displacement. *JRehabilResDev* 41: 829-834, 2004.

14.   **Schepens B, Willems PA, and Cavagna GA.** The mechanics of running in children. *JPhysiol* 509 (Pt 3): 927-940, 1998.

15.   **Sun M and Hill JO.** A method for measuring mechanical work and work efficiency during human activities. *JBiomech* 26: 229-241, 1993.

16.   **Swartz AM, Strath SJ, Bassett DR, Jr., O'Brien WL, King GA, and Ainsworth BE.** Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Medicine and science in sports and exercise* 32: S450-456, 2000.

17.   **Tryon W and R W.** Fully proportional actigraphy: A new instrument. *Behavior Research Methods Instruments & Computers* 28: 392-403, 1996.

18.   **Welk GJ.** Principles of design and analyses for the calibration of accelerometry-based activity monitors. *Medicine and science in sports and exercise* 37: S501-511, 2005.

19.   **Welk GJ, Schaben JA, and Morrow JR, Jr.** Reliability of accelerometry-based activity monitors: a generalizability study. *MedSciSports Exerc* 36: 1637-1645, 2004.

20. **Welk GJ, Schaben JA, and Morrow JR, Jr.** Reliability of accelerometry-based activity monitors: a generalizability study. *Medicine and science in sports and exercise* 36: 1637-1645, 2004.

21. **Yngve A, Nilsson A, Sjostrom M, and Ekelund U.** Effect of monitor placement and of activity setting on the MTI accelerometer output. *MedSciSports Exerc* 35: 320-326, 2003.

CHAPTER V

AN ARTIFICIAL NEURAL NETWORK MODEL OF ENERGY EXPENDITURE

USING NON-INTEGRATED ACCELERATION SIGNALS

Abstract

Accelerometers are a promising tool for characterizing physical activity (PA) patterns in free-living. The major limitation in their use to date has been a lack of precision in predicting energy expenditure (EE), which may be attributed to the over-simplified time-integrated acceleration signals and subsequent use of linear regression models for EE prediction. In this study, we used bi-axial raw (32 Hz) acceleration signals at the hip to develop a relationship between acceleration and minute-to-minute EE in 102 healthy adults while measurements were simultaneously made in a room calorimeter for nearly 24 hours. Using the pooled data, we extracted 10 features which we determined had the potential to characterize EE intensity, and developed a feed-forward/back-propagation artificial neural network (ANN) model with one hidden layer (12x20x1 nodes). Results of the ANN were compared to predictions using the ActiGraph monitor, a uni-axial accelerometer, and the predictions made by the IDEEA monitor, an array of five accelerometers. All accelerometer predictions were highly correlated with EE ($r^2 > 0.79$). After training and validation (leave –one subject out), the ANN showed significantly reduced mean absolute errors (0.29~0.10 kcal/min), mean squared errors (0.23 ± 0.14 kcal^2/min^2), and difference in total EE (21 ± 115 kcal/day), when compared to both the IDEEA ($p < 0.01$) and the ActiGraph ($p < 0.001$). Thus, ANN combined with raw acceleration signals is a promising approach to link body accelerations to EE. Further

validation is needed to understand the performance of the model for different PA types under free-living conditions.

Introduction

In the last fifteen years, portable accelerometers have been used to characterize the intensity and duration of physical activity (PA), and their output has further been used for the prediction of energy expenditure (EE) (6, 10, 12). Accelerometer devices are typically worn at the hip with the aim of capturing displacement of the subject's center of mass, which is generally associated with moderate to high intensity activities. In order for data to be collected for more than one day, long enough to assess the patterns of PA in free-living, most accelerometery-based PA monitors record one data point per minute, which represents the summation of acceleration events during the minute. The output of PA monitors is reported to the investigator in units of activity counts (22), which are an arbitrary unit specified by each device manufacturer. To our knowledge, this integration (or summation) process was not designed *a priori* for predicting EE rather it was dictated by memory capacity and battery life in early accelerometers. As such, other characteristics of raw acceleration signals may yield better predictive outcomes.

Early modeling approaches to relate activity counts and EE typically assumed a linear relationship between the activity count values and EE measured using indirect calorimeters (10, 13, 15, 20). Linear regression fits were used because of their computational simplicity and ability to well characterize the energy costs of moderate intensity, ambulatory activities. While models based on this strategy provided an excellent first approximation of the relationship between hip-measured acceleration

signals and EE, they have suffered in their generalization to different PA types and subject populations (24). This is because the models were predominantly developed using short protocols containing primarily dynamic modes of PA (walking and jogging) and were developed on homogeneous subject populations. Prediction accuracy of general models also varies greatly between subjects with different personal characteristics (age, height, weight, etc) because identical accelerations may not result in the same metabolic cost for these individuals.

Investigators have sought to improve model accuracy by increasing the amount of information gathered during each measurement epoch. This effort has included adding additional acceleration dimensions at the hip (6, 16), adding sensors to the limbs (wrist and ankle) for more complete movement detection (5, 13), and by coupling physical and physiological information, such as heart rate, near body temperature and skin impedance (4, 14, 18). Using the additional data collected by these devices, more mathematically sophisticated and in some cases more accurate models relating acceleration and energy expenditure have been developed such as multiple linear regressions (12) and nonlinear models (6, 17). Recently, a new model for EE prediction was developed that called for recording data in finer time intervals (one second) using a uni-axial accelerometer rather than collecting more channels or types of sensor data (8). In this model, the second-by-second data variability was used as an initial discrimination tool to decide the specific model to use on each epoch of data. This modeling approach was made possible because of improvements in the data storage capacity and battery life of modern accelerometers. Increases in temporal resolution open the field to new analytic solution techniques that rely upon multiple measurements acquired from each minute of measured activity.

The purpose of this study was to expand on existing EE modeling techniques by capturing raw (32 Hz) acceleration signals from a bi-axial accelerometer worn at the hip. We propose a feature extraction scheme where the dense acceleration signals are reduced to a small number of simple to compute statistical parameters (features) that are well correlated with the minute-by-minute EE measured by a whole room indirect calorimeter. The reduced signal information combined with subject demographics (sex, age, height, weight, BMI, and racial/ethnic background) were used to develop an artificial neural network (ANN) model to predict minute-by-minute EE. Results of the ANN model were compared with both a traditional accelerometer regression equation, and the proprietary output of a commercially available accelerometer array.

Methods

Participants: One hundred and two healthy adults (46 men, 55 women) between the ages of 18 and 70 years completed this study. Subjects were free of both diseases and medications known to alter metabolic rate, major orthopedic limitations, and were non-smokers. The characteristics of these subjects are shown in Table 1.

Table 1: Characteristics of study participants represented as mean ~ inter-quartile range and (total range).

	All Subjects (n = 102)	Men (n = 46)	Women (n = 56)
Age (years)	35 ~ 22.0 (19 – 69)	35.5 ~ 20.5 (20 ~ 69)	35.0 ~ 22.3 (19 ~ 67)
Height (m)	1.70 ~ 0.15 (1.52 ~ 1.91)	1.78 ~ 0.09 (1.67 ~ 1.91)	1.63 ~ 0.08 (1.52 ~ 1.78)
Weight (kg)	72.1 ~ 22.4 (48 ~ 120)	81.4 ~ 18.6 (64 ~ 120)	64.6 ~ 19.5 (48 ~ 114)
BMI (kg / m^2)	25.0 ~ 5.9 (16.9 ~ 42.1)	25.3 ~ 4.4 (19.8 ~ 38.5)	24.4 ~ 7.1 (16.9 ~ 42.1)
% Body Fat	27.6 ~ 17.3 (6.2 ~ 57)	22.4 ~ 9.1 (6.2 ~ 45.1)	36.2 ~ 16.0 (11.7 ~ 57)

Experimental Procedures: Volunteers were recruited from the middle Tennessee area using flyers, email distribution lists, and personal contact. Before participation, all subjects signed an informed consent document approved by the Vanderbilt University Committee for the Protection of Human Subjects. Each subject was asked to stay in the room calorimeter for approximately twenty-four hours while minute-by-minute activity data was acquired with multiple commercially available accelerometery-based PA monitors. Each subject was asked to engage in two structured activity intervals. The morning activity period was comprised of self-paced walking and jogging (both in the room and on the treadmill), while the afternoon activity period contained sedentary activities, such as deskwork, along with stationary biking (Table 2). Each prescribed activity was performed for ten minutes followed by a ten minute rest period to allow the metabolic rate to return to baseline between intervals and to allow *post hoc* discrimination between activity types. During times when no activity was prescribed,

subjects were encouraged to engage in their normal daily PA routine as much as possible.

Subject's height and weight were measured on the morning of the study visit.

Time	
0700	Admission to GCRC: Health & Physical, vital signs, consent Fitted for PA monitors, introduction to room calorimeter
0800	Enter room and start test Breakfast
0930	*Morning PA protocol*: 10-min continuous PA and 10-min rest

Low Intensity	*Moderate Intensity*	*Vigorous Intensity*
Stretch 　Pacing in room 　　Walking 2-3mph Cool down	 Walking 2.5-4.5mph	 Jogging 4-6.5mph* Jogging 4.5-8mph*

1200 — Lunch

1430 — *Afternoon PA protocol*: 10-min continuous PA and 10-min rest

Low Intensity	*Moderate Intensity*	*Vigorous Intensity*
Typing (laptop) Handwriting Moving object (sitting) 　Cleaning (standing) Cool down	 Biking (30-50W)*	 Biking (45-150W)*

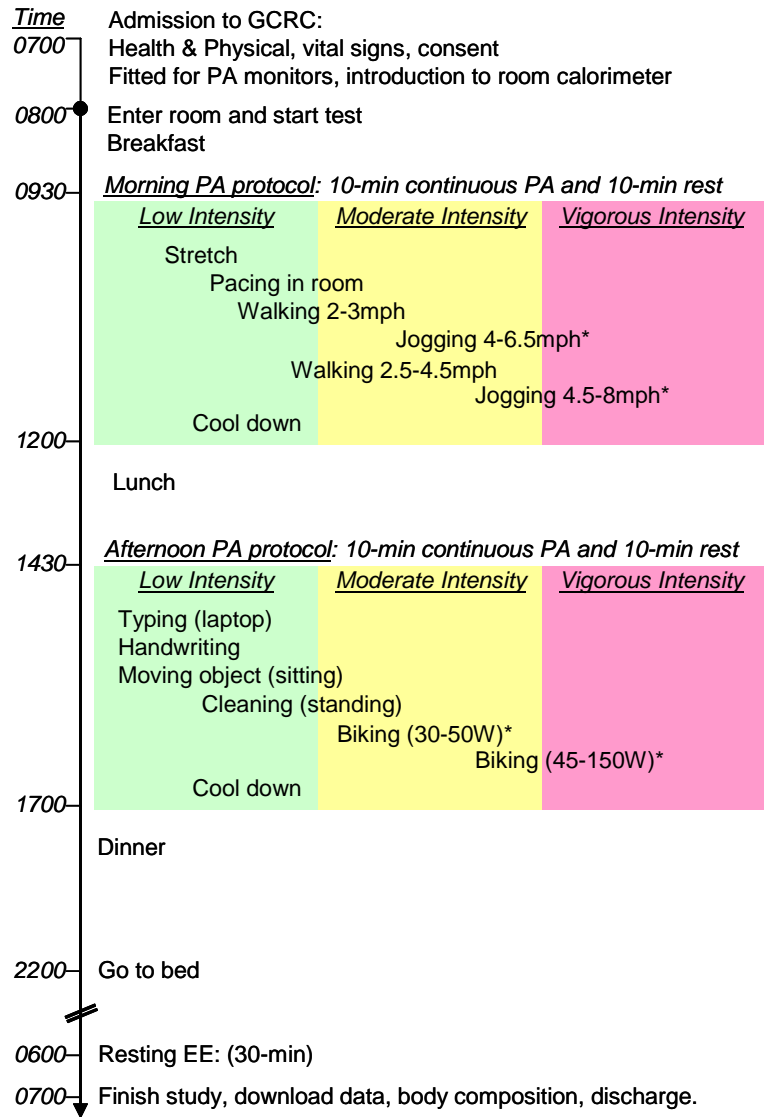| Time	
1700	Dinner
2200	Go to bed
0600	Resting EE: (30-min)
0700	Finish study, download data, body composition, discharge.

Figure 1: Protocol for the metabolic chamber stay. Intervals denoted with (*) indicate recommendations for the range. Subjects are asked to self-pace for these intervals.

Instrumentation

Activity-energy measurement system: Energy-expenditure was computed on a minute-by-minute basis by the Vanderbilt University room calorimeter, which is located within the Vanderbilt General Clinical Research Center. This system measures oxygen consumption and carbon dioxide production with high accuracy (system error < 1%). The room calorimeter is an airtight environmental room measuring 2.5 x 3.4 x 2.4 m. The calorimeter is equipped with a toilet and sink, desk, chair, telephone, television, DVD player, stereo system, bed, treadmill, and exercise bike. While the calorimeter floor contains a force plate and the room has several event markers, information from these systems was not utilized for these experiments. Technical details of the calorimeter have been previously reported (19).

Accelerometers: Subjects were outfitted with both the ActiGraph (Fort Walton Beach, FL) uni-axial accelerometer, and a custom designed activity monitor, which is a derivative of the commercially available IDEEA (MiniSun, Fresno CA) monitor. The commercial IDEEA monitor consists of an array of five accelerometers (20x15x4 mm, 2g) attached to the skin via hypoallergenic tape at the sternum, mid-thigh, and bottom of each foot. Each sensor is wired to a hip pack that serves to synchronize the signals from each channel and store the data. While high accuracy for the IDEEA physical activity type identification routine has been published (26), the study designed to validate the EE prediction routine contained only walking, jogging, and lying down (25), and has therefore not been subjected to a rigorous validation in sedentary PA types.

Figure 2: Illustration of IDEEA sensor locations. Sites shown in black represent the original accelerometer sites, while sites in white are the custom sensor sites.

The custom IDEEA monitor includes all of the sensors from the original configuration but adds recording capability at the hip pack (bi-axial), on each upper arm (uni-axial), and on the top of each hand (bi-axial). Raw data (32 Hz) is collected at each of the custom sites, while integrated signals are recorded by the original IDEEA sensors (Figure 2). In this configuration, data can be acquired continuously throughout our study visits (~ 21 hours). For this study, we used only the raw data from the hip sensors (bi-

axial) for analysis. Since most investigators are only collecting data at the hip, and it would be ideal to collect field data from only one site in order to minimize the inconvenience to the subject, we felt it was important to explore model developments that could be applied to traditional hip-mounted accelerometers before expanding our study goals to include multi-site high frequency data analysis.

The ActiGraph is a uni-axial accelerometer that has been in widespread use for more than a decade. Device specifications have been published elsewhere (23). Data was collected using model 71256 activity monitors using one minute epochs. All monitors were calibrated prior to the start of the study and at 6 subsequent time points to ensure proper calibration. Resulting activity counts were analyzed using a combination of the Freedson equation for moderate to high intensity data and the work-energy theorem for low intensity values. This analysis construct is presented in the ActiGraph instruction manual as a way to use the Freedson equation on all intensities of PA data (1, 10). This equation was chosen since it was developed for one minute epoch data and the ActiGraph software is equipped to compute this regression.

Modeling Approach: Artificial Neural Network (ANN) modeling was selected to relate the features of the raw acceleration signal to measured energy expenditure on a minute-by-minute basis. ANN modeling is an information-processing paradigm inspired by the way the densely interconnected, parallel structure of the mammalian brain processes information (11). Models are developed using a learning process where a series of connection weights, analogous to synapses, are tied to a series of processing elements, analogous to neurons. When presented with data, the model allows weighting coefficients to be updated such that the strength increases if the weight can be used to reduce

prediction error and the connection weight is reduced if it seems to increase error. ANN is a good candidate model when there are a large number of inputs for a small number of outputs or when the ideal functional form of the solution is not known (2).

To implement ANN, we begin by specifying the number of inputs (acceleration or subject characteristic terms), the number of weight values (interactions between the terms), the architecture of the model (how interaction terms are arranged computationally), and the number and type of output parameters (energy expenditure, a single continuous variable) (Figure 3). A single neuron receives multiple inputs, which represent characteristics of the acceleration signal. The relative importance of each input is specified by a weight value. Initially, these values are small random numbers, since the relative importance of each characteristic is unknown. Through the model optimization process, weight values are optimized to predict minute-by-minute EE. The summation of inputs and their associated weight vectors are added and this value is modified by a transfer function, which limits the output range of each neuron is appropriate for the next downstream neuron.

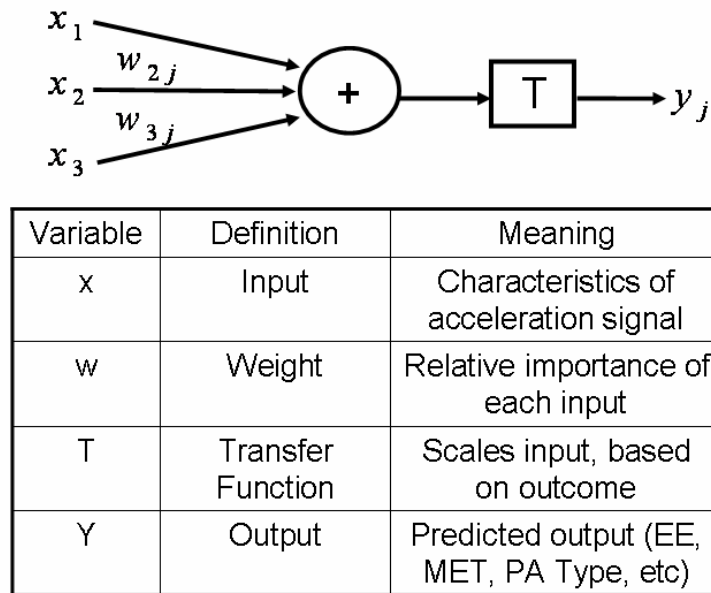| Variable	Definition	Meaning
X	Input	Characteristics of acceleration signal
w	Weight	Relative importance of each input
T	Transfer Function	Scales input, based on outcome
Y	Output	Predicted output (EE, MET, PA Type, etc)

Figure 3: A single artificial neuron. The inputs, which represent characteristics of the acceleration signals, are multiplied by weight vectors, which represent their strengths. The combination of these quantities is used to predict an outcome, in our case, EE.

A single neuron (Figure 3) is not capable of solving difficult problem so multiple neurons are arranged into computational layers. The transfer functions between layers allow for nonlinearities and in the predicted output, while the weight values allow for interaction between all neurons of a layer. Data moves through each of the layers of neurons until the output layer is reached (Figure 4). In a fully connected network each input value is connected by every weight value to the first hidden layer of the network. Data flowing into the nodes (neurons) of the hidden layer is summed and subjected to a transfer function, which transforms the data into an appropriate order of magnitude for

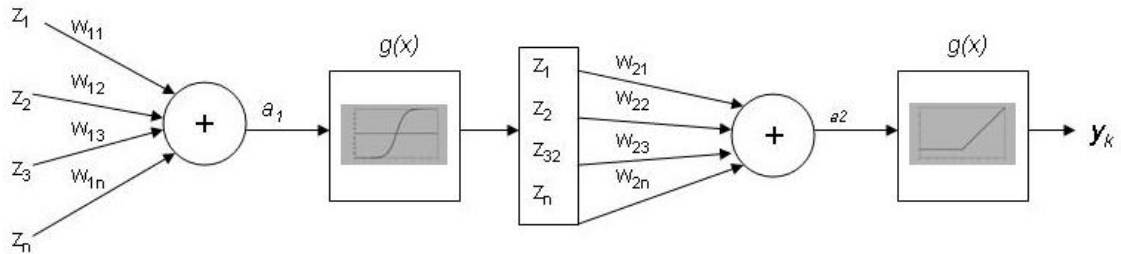the next layer. The transfer function can be any differentiable function, and must be specified by the user.

Figure 4: Schematic for the feed forward network, which includes multiple neuron units formed into computational layers. Each input (z) is multiplied by the layer weights (w) and sent through a summation (a) and transformation (g(x)) to the next layer. The process is repeated until the output (y) is reached. The first basis function is a hyperbolic tangent function, while the second is a positive linear basis function.

The output layer transfer function determines the bounds on the predicted solution. For example, we specified that EE prediction should never be less than zero, so a transfer function is selected that will not allow the model to predict negative values. Network parameters such as the number of neurons and the type of basis function are empirically determined.

The process by which the model derives a solution is referred to as training because the model is presented with examples for which a known measurement value exists. This allows the performance of the model relative to a gold standard to be established. To start, each of the weight values is assigned a small, random value. In order to improve on these random weighting coefficients ANN models "learn" using a feed forward/ back propagation approach. The feed-forward process consists of

computing the transformations between the inputs and outputs as specified by the current weights and transfer functions (Figure 4), with each minute of data from each subject representing a training example since each has an EE measurement. The prediction from the forward pass through the network is compared to the measured value and an error signal is computed. This gradient (rate of change) of the error is used to adjust the weights starting with those in layers closest to the solution, working backwards to the inputs with the size of the weight change dependant on the contribution of the weight to the solution. This constitutes the back-propagation. Back-propagation is an optimization procedure and can be governed by most standard optimization approaches, such as gradient descent, or conjugate gradient methods (7). As the inputs are shown to the network many times and the weights are allowed to adapt, the error signal decreases, and an optimal set of weights can be realized. Training is allowed to continue until a specified error tolerance has been achieved, the validation error, computed on data not used for model development, begins to increase, or a maximum number of computational iterations has been reached.

Once optimized weight values have been achieved, prediction of novel examples consists of using the forward pass through the network (multiplying inputs by layer weights, passing through transfer functions, and predicting a single EE value per minute). This spares the user the bulk of the computational costs and allows prediction to be produced nearly instantaneously for small to moderate sized networks.

For this problem, a 12 node input layer, followed by a 20 node hidden layer, and a single node output layer architecture was used. Hyperbolic tangent basis functions were chosen for the input and hidden layer because of their capability to account for positive

100

and negative correlations in data. Training was performed using a gradient descent training function with a learning rate of 0.01. Validation was performed using leave-one-subject-out cross validation. In this approach, the model was trained on 101 subjects with one subject used to validate the model results (21). Training was ended when the error on the validation set failed to decrease by more that 1e-6 per iteration (after an initial drop), the error gradient fell below 1e-6, or 5000 iterations was reached.

Feature extraction: Feature extraction is the key step in preparing raw data for ANN modeling. The purpose for this step is data reduction. In this study, 1920 (32 samples/sec x 60 seconds per minute) data points are collected by each IDEEA sensor channel for each minute of study data collected. These values all correspond to a single measurement made by the indirect calorimeter. This amount of information quickly becomes cumbersome to analyze, but more importantly, redundant information is contained in the acceleration signals. It is therefore vital that the raw data is reduced into a small number of parameters that carry the most relevant information. We chose to reduce the data into a series of parameters that we felt were both statistically relevant and physically meaningful. Eleven parameters were extracted from each channel of raw data (median, integral, peak intensity, inter-quartile interval, skew, kurtosis, peak CV over any 10 seconds of data, lowest 10 second CV, mean absolute error (MAE), and the summation of signal power above and below 0.7 Hz). The signal power cutoff of 0.7 Hz was determine based on optimizing the division in the power spectral density (PSD) between walking and sedentary tasks in a sample of ten subjects not used for model development. The eleven computed acceleration parameters were then analyzed based on their correlations with one another in order to eliminate redundant information. This step

reduced the inputs to five for each sensor channel. These consist of the peak value, the inter-quartile interval, the lowest coefficient of variability when each minute of data was analyzed in ten second increments, the sum of the signal power below 0.7 Hz (assumed to be associated with sedentary activities), and the sum of the signal power above 0.7 Hz (assumed to be associated with locomotion and other higher intensity activities). These data were joined in the input set by the subjects' sex, age, height, weight, and ethnic background since these features have been shown to impact resting metabolic rate (RMR) and can be easily measured or self-reported (9).

Feature extraction was designed such that the final inputs to the model are quantities researchers are generally familiar with (at least conceptually), and that have meaning outside of this modeling effort, i.e. they correlate with characteristics of PA such as the intensity or variability of movements. Additionally, by using a small number of easily computed data features, the storage requirements for any future activity monitors would be minimized because raw data would not need to be stored, only the relevant computed parameters. This effectively minimizes the amount of internal storage capacity required of the accelerometer while maintaining the quality of information derived from the raw signal. Model development and feature extraction was performed using Matlab 7.01 (Mathworks, Natick MA).

Model validation was performed using a leave-one-subject-out cross validation (Figure 5). In this approach, the data was divided into a training set (n = 101) and a testing set (n = 1). The ANN model was optimized on the training set and the resulting optimized solution was used to predict the EE on a minute-by-minute basis for the single test subject, whose data was not used in model development. This process was repeated

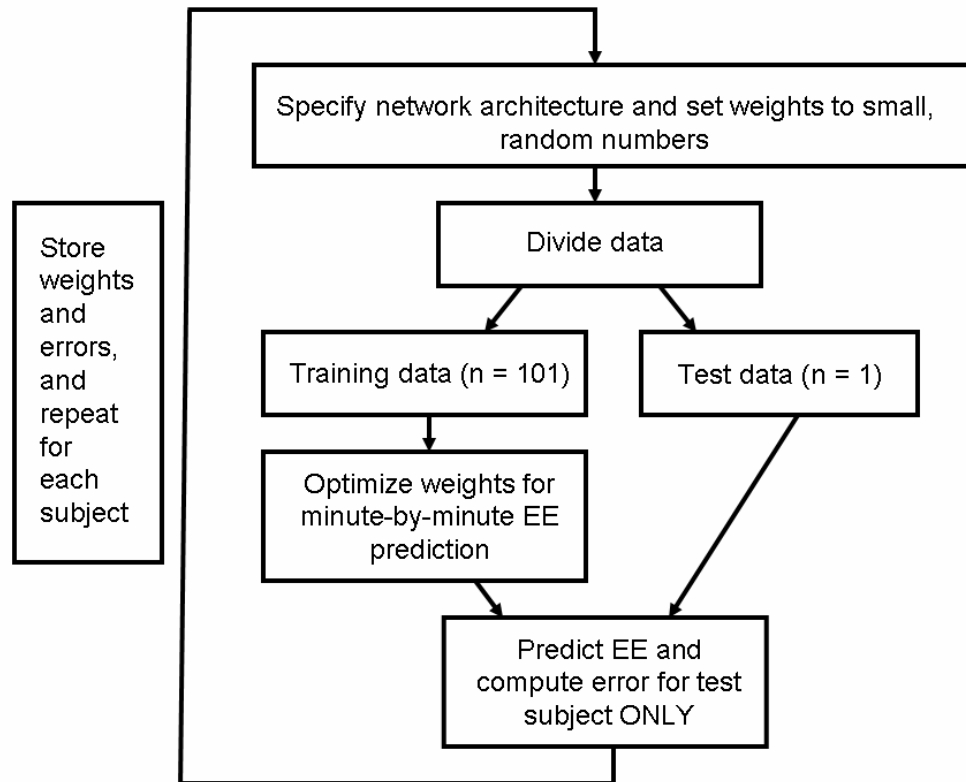102 times and the results reported were derived using the testing data only.

Figure 5: Schematic of the leave-one-subject-out cross validation used for testing the performance of the ANN model.

Statistical Analysis: Data are presented as median, inter-quartile range, and total range. Models (ActiGraph, proprietary IDEEA, ANN) were compared according to the mean absolute error (MAE) (Eq 1), the mean squared error (MSE) (Eq 2), the percent difference between each model and the measured TEE, and the squared Pearson's correlation coefficient (r^2) for each subject over the entire study duration, using ANOVA with post-hoc Tukey tests. Bland-Altman plots (3) were used to examine trends in

prediction accuracy as a function of the TEE. *Post hoc* analysis of the relationships between the difference in TEE measurement and prediction with the subject demographics were performed in order to understand the types of subjects for whom model development is especially challenging.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|(x_i - \mu)\right| \qquad\qquad \text{(Eq 1)}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \qquad\qquad \text{(Eq 2)}$$

Results

Approximately 112,000 data points (minutes of data) were used to train the ANN model. Convergence occurred after an average of 3328 iterations with a range of 475 – 5000 iterations (5000 was the maximum number of iterations allowed for this experiment). The training error as a function of the number iterations showed an exponential decay profile. An error profile from a randomly selected subject is shown in Figure 6. Though the rates of decay may change between subjects, and some subjects may have long plateaus early in the training if the initial weight values place the model in an area of space where the gradient is small, all eventually exhibit an exponential profile where the bulk of the error decay occurs over approximately 50 iterations (Figure 6 inset).   The number of iterations to solution convergence can be altered through the model learning rate, which was set at 0.01 for this model. When larger learning rates

104

were used in pilot experiments for this project and solutions were achieved more rapidly, ultimately, the solutions showed oscillations in the decay profiles indicating that the solution was not stable.
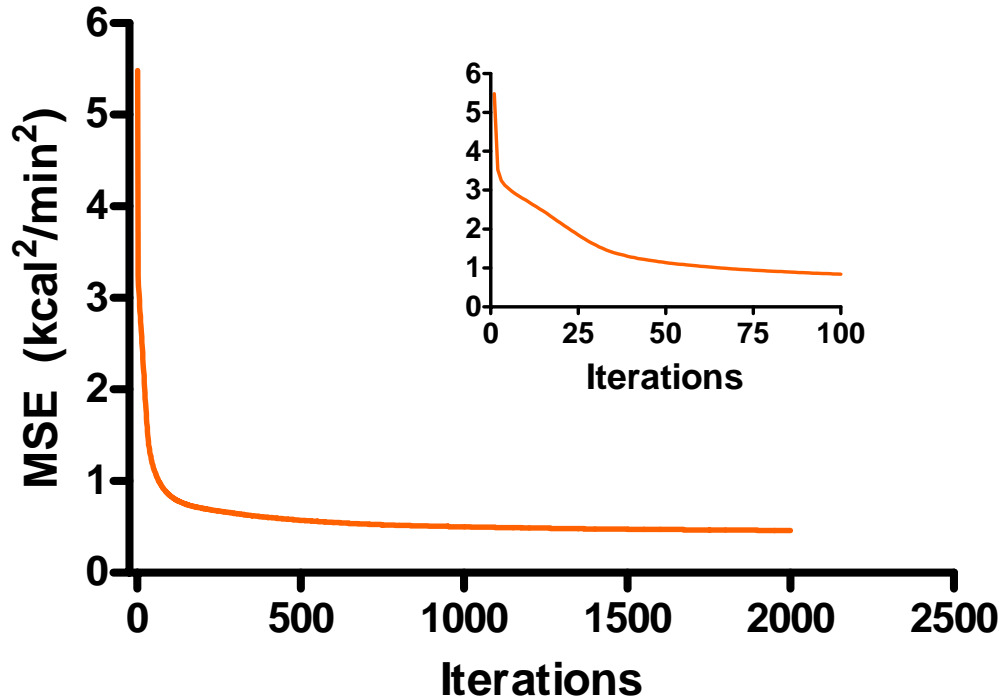

Figure 6: Error decay for a representative model training for the first 2000 iterations. The inset shows the error over the first 100 iterations.

An F-test revealed no statistical correlation ($p = 0.9214$) between the number of iterations to convergence and the resulting validation errors. While the ANN was developed on 102 subjects, only 81 subjects had both IDEEA data and ActiGraph data (data was considered to be complete if at least nine hours of continuous daytime data was

collected). All results shown are based on paired comparisons of the subjects who had both IDEEA and ActiGraph data.

A representative subject's minute-by-minute EE prediction showed that the ANN was able to both characterize the baseline EE as well as prediction of programmed activity intervals and spontaneous PA bouts, though some small errors are observed (Figure 7).
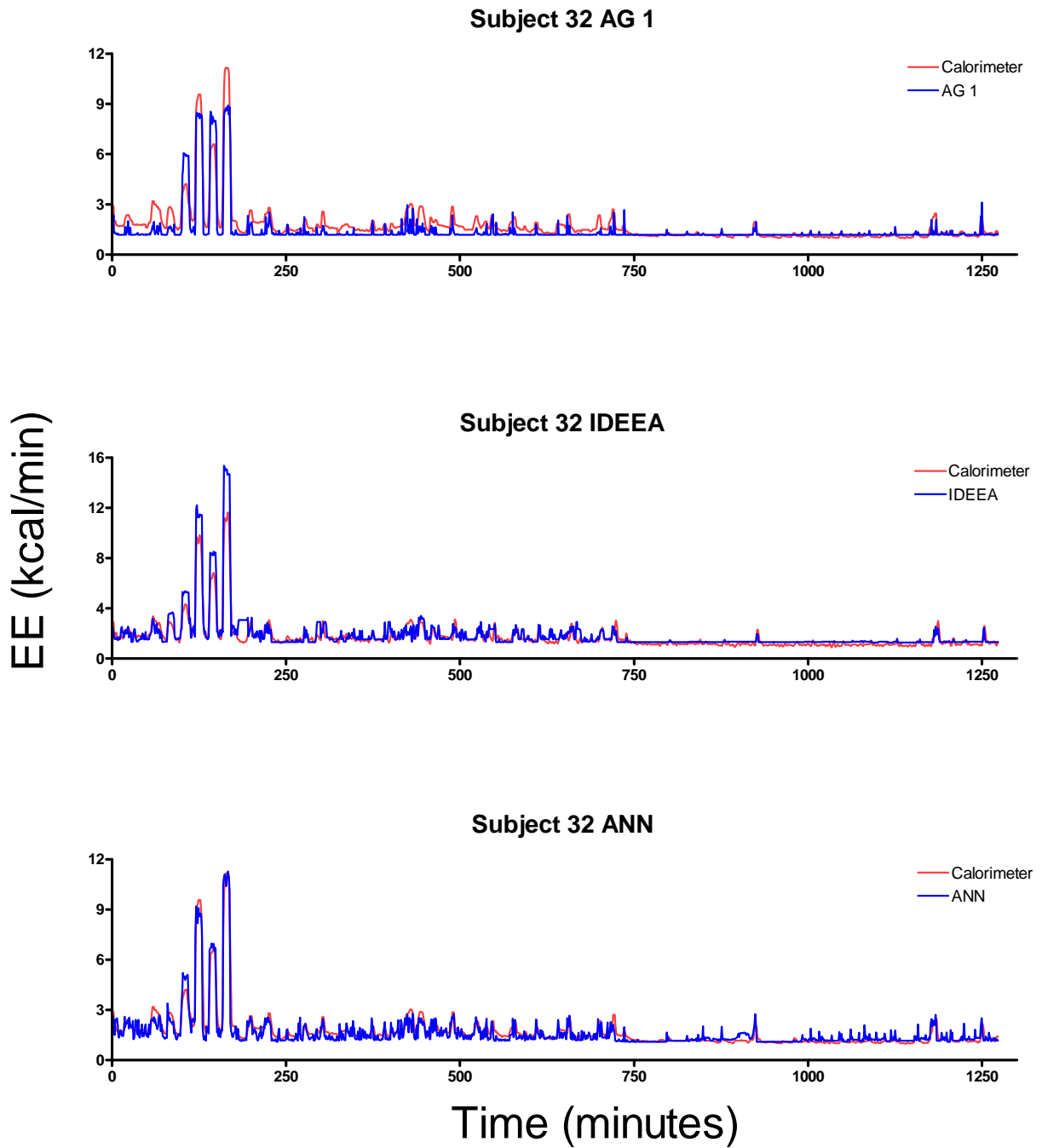
Figure 7: Minute-by-minute EE prediction using ActiGraph (AG 1), IDEEA, and ANN. The top row of graphs shows the entire study visit, while the bottom shows the first 600 minutes of collected data. This subject was a 43 year old male (height – 1.69 m; weight – 76.9 kg; BMI – 26.7 kg/m^2). Using BMI, this individual is characterized as healthy.

Summary statistical measures, the r^2, MAE, MSE (using minute-by-minute EE predictions), and percent difference between the chamber measured and model predicted TEE were computed for the ActiGraph, IDEEA, and ANN on each of the 81 subjects (Figure 8).
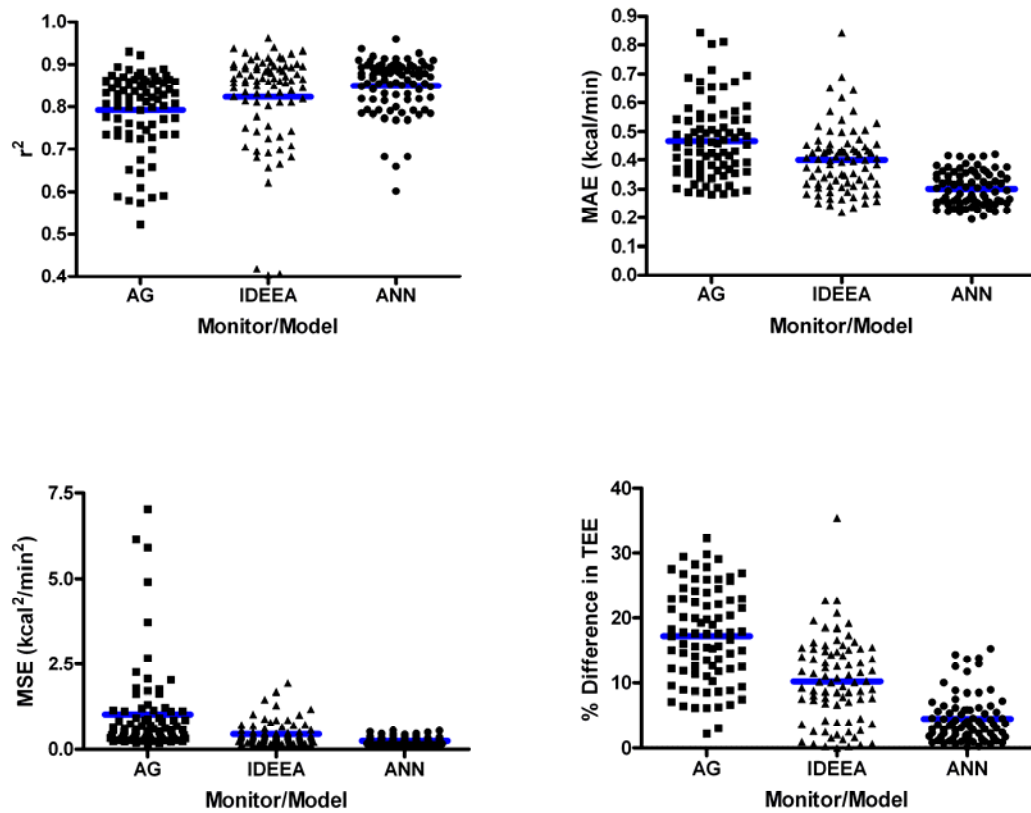

Figure 8: Results of summary analysis of AG, IDEEA, and ANN models. The correlation coefficient, r^2, was higher in the ANN than in the IDEEA or AG model, while MAE, MSE, and % difference in TEE was reduced.

All models showed, on average, high correlation with measured minute-by-minute EE. ANOVA revealed that the r^2 was higher ($p < 0.001$) in the ANN model relative to the other models and that both the mean of the testing set MAE and MSE were

significantly reduced in the IDEEA model relative to the AG model (p < 0.001), with

further reductions (p < 0.01) also seen when the ANN was compared with the IDEEA

monitor proprietary model. The percent difference between TEE measured by the room

calorimeter and that predicted by each model was computed to account for differences in

the number of measurement minutes and PA intensities represented in each subject's

data. Analysis of the percent difference revealed a drop in percent error in the IDEEA

relative to the AG, and a further decrease in error when the ANN was compared to the

IDEEA model (Table 2).

Table 2: Comparison of ActiGraph, IDEEA, and ANN model performance
assessed using summary statistics for 81 common subjects. Data are presented as mean ~
inter-quartile range, along with the total range. All pair-wise comparisons are significant
(p < 0.05 except MSE between IDEEA and ANN model.

	r^2	MAE (kcal/min)	MSE ($kcal^2/min^2$)	% Difference
ActiGraph	$0.82 \sim 0.11$ $(0.55 \sim 0.93)$	$0.46 \sim 0.18$ $(0.28 \sim 0.84)$	$0.56 \sim 0.78$ $(0.18 \sim 7.04)$	$17.59 \sim 11.02$ $(2.20 \sim 32.32)$
IDEEA	$0.86 \sim 0.09$ $(0.40 \sim 0.96)$	$0.40 \sim 0.13$ $(0.22 \sim 0.84)$	$0.35 \sim 0.27$ $(0.11 \sim 1.93)$	$10.09 \sim 9.78$ $(0.02 \sim 35.53)$
ANN	$0.87 \sim 0.08$ $(0.60 \sim 0.96)$	$0.29 \sim 0.10$ $(0.19 \sim 0.42)$	$0.23 \sim 0.14$ $(0.09 \sim 0.55)$	$3.58 \sim 4.04$ $(0.13 \sim 15.35)$

Sex, age, height, weight, and BMI were explored to determine if significant

effects of these characteristics exist in each model. All models had significant bias in the

MAE with respect to sex, height, weight, and BMI (p < 0.001). While the MAE had

significant correlations with height, weight, and BMI for all models, the difference

between predicted and measured total EE was not significantly correlated with these

characteristics using the IDEEA or the ANN (p > 0.05). Age was not significantly

correlated with either MAE or difference in total EE for any of the models.

Bland-Altman plots were used to characterize the ability of each model to predict

TEE (Figure 9). The AG model showed an average difference of 355 ± 240 kcal/day.

There was also a significant trend (p = 0.0351) towards under-prediction of TEE as the

absolute value of TEE increased. The IDEEA model demonstrated a mean difference of

230 ± 209 kcal/ day. There was a trend (p < 0.001) towards over-prediction as TEE

increased. The ANN shows a mean difference of 21 ± 115 kcal/day. No significant trend

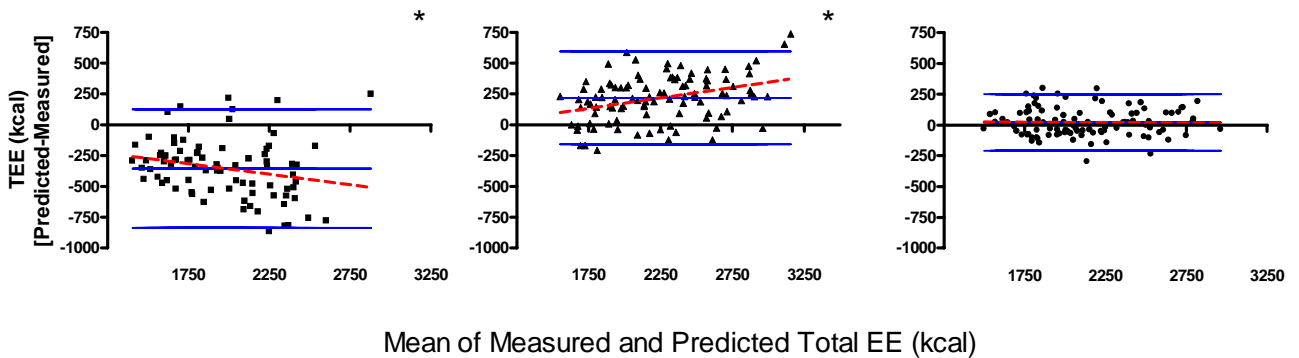(p = 0.86) was observed as a function of magnitude of daily EE.

Figure 9: Bland-Altman plots of TEE using the ActiGraph, IDEEA, and ANN.
Both the mean and the standard deviation of the ANN show improvements relative to the
IDEEA and ActiGraph model.

Discussion

In this study, we investigated a procedure of extracting features from raw

acceleration data collected by hip-worn accelerometers and development of a minute-by-

minute EE prediction model based on ANN techniques. This experiment was developed

as a proof of concept that using features of the acceleration signal other than the integral

and combining them with a flexible modeling approach, models can be developed that

provide reduction in prediction error for EE, while also reducing the inter-subject

variability, which tends to be high in generalized predictive models. The resulting

network model showed a reduction in MAE, MSE, and percent difference in TEE when

compared with both a first generation accelerometer device (ActiGraph), and the

proprietary model used in the IDEEA monitor, a second generation multi-sensor array

system. In addition to reductions in mean errors for the group, the variance was also

reduced, which further suggests the robustness of the ANN modeling approach. Error

reduction observed relative to an accelerometer array suggests that high accuracy

predictions can be achieved using a single sensor, thus minimizing the inconvenience to

the subject while maximizing the utility of the monitoring tool.

While the ActiGraph, proprietary IDEEA model and the ANN model all had high

correlation with the TEE on a minute-by-minute basis, this is more reflective of the

capability of the accelerometer to detect some motion rather than the model accurately

reflecting minute-by-minute EE intensities. As such, it is high for all three monitors.

When measures that reflect the magnitude of the EE differences observed on a minute-

by-minute basis (MAE or MSE) are considered, significant reductions in error were

observed in the IDEEA model relative to the ActiGraph and the ANN relative to both of

the other models. The cumulative effect of these errors can be observed in the percent

difference measurement which shows a reduction of nearly 13% between the ActiGraph

and the ANN, and 5% between the IDEEA monitor and the ANN. The mean of the

difference in TEE was also greatly reduced by the ANN relative to the IDEEA and

ActiGraph with the mean difference in the ANN model being only 21 kcals, suggesting that the model has corrected for some of the baseline offset (resting EE) problems that have been previously observed for the IDEEA model (25).

One of the biggest challenges involved in generalized modeling with accelerometers has been the large standard deviation of predictions across subjects. In this study, we found high individual prediction errors by AG and IDEEA comparing to the measured EE (95% CI [-835 ~ 125] and [-647 ~ 187] kcal/day, respectively). These values are beyond the treatment effect we typically target in sustainable weight loss interventions, which is 100-250 kcal/day. Thus, reducing this measurement error has crucial clinical implications. Large variability in performance across subjects may be related to the fact that standard regression approaches may not have sufficient flexibility to alter predictions when the same acceleration count value is achieved by a subject whose personal characteristics are different than those used for model development. The standard deviation in the TEE observed with the ANN, which allows interactions between characteristics and acceleration terms, showed a reduction of approximately 50% relative to the ActiGraph and nearly 45% relative to the proprietary IDEEA model. However, the characteristics used in the ANN may not be sufficient to account for all modeling challenges in generalized modeling. This is seen as some subject characteristics introduce significant bias in the MAE though the model was given this information in training. Additional information such as body composition and measures of physical fitness such as VO_2 Max would be candidates for the model.

ANN has a number of attractive features for the energy expenditure prediction problem such as the flexibility of predictions across subjects, allowing interactions

between all input terms, and its ability to map multiple inputs (acceleration terms) to a single output (EE) without pre-specifying a functional form (linear, logistic, etc). The major disadvantages of ANN approaches are the computational complexity of the models which may require long training times and a relatively large number of free parameters. Additionally, model training requires a large number of labeled examples, acceleration data from a diverse sample of physical activity types, for which the EE is known, which requires a long data collection period before models can be developed.

While a classical ANN may seem like a black box solution technique, we attempted to minimize this appearance by carefully selecting model inputs that make sense in the context of the EE prediction problem. We chose terms that generally represent the magnitude of movements, frequency of movements, and the variability in motion patterns. These terms were assumed to be characteristic of certain modes and intensities of PA. However, this feature extraction process requires the model developer to make decisions about what data features may be of interest. Alternately, feature extraction can be performed by a standard data reduction technique such as principal component analysis, which achieves data reduction by combining parameters that are linearly related. This process maximizes the amount of the information from the data while eliminating repetitious measurements. The advantage of this technique is its capability to succinctly and consistently reduce data to the desired proportion of the total data variance. The disadvantage is that the reduced data is not comprised of characteristics that would be familiar to researchers; rather it contains features representing agglomerations of measurements.

Perhaps the most challenging aspects of model development are collecting

appropriate model training data and validation. Since there are literally hundreds of modes of physical activity that individuals may engage in and at least that many profiles for subjects' metabolic response to exercise, models will tend to generalize best to data sets comprised of activities similar to those that were used for the original model development. We have attempted to mitigate this factor by (1) asking subjects to self-pace activities and (2) capturing spontaneous bouts of PA. These two steps allow for the collected data to be both diverse in intensity composition, and representative of the activity patterns our subjects would normally engage in. In order to attempt to attempt to minimize the potential error increases associated with applying out model to new subjects (generalization errors), a leave one subject out cross validation procedure was selected. This technique allows the bulk of the collected data to be used in model development relative to a split sample validation where some percentage of the data is withheld for model validation. The data from the validation sample may include unique features that would have impacted the model development had they been available. The model presented here, however meant to prove that in principle, a high dimensional modeling approach such as ANN coupled with feature extraction from raw (32 Hz) acceleration signals can be used for EE prediction on a minute-by-minute basis, and the specific weighting coefficients should not be viewed as final.

Accelerometers have long been considered a promising tool for predicting EE due to their relatively low price, ease of use, and ability to record for many days at a time. This potential has not been fully met to date due to limitations in our ability to relate the output variables from the monitors to EE. Collecting raw acceleration data has the capability of improving the precision of EE prediction by allowing researchers the

flexibility to identify relevant parameters during the feature extraction phase as well as opening the field to high dimensional modeling techniques such as ANN which have the capability of generating more flexible predictions than more traditional modeling techniques. This study has shown a proof of concept that by applying feature extraction and ANN models to bi-axial acceleration data acquired at the hip, minute-by-minute and total EE predictions can be improved. Additional subjects and modes of PA should be acquired to both validate the current model as well as for use in developing a more robust algorithm. Additional results from the hip-only model can be found in Appendix D. Results from individual ANN models can be found in Appendix E.

References

1. **ActiGraph.** *Actisoft Analysis Software 3.2 User's Manual*. Fort Walton Beach, FL: MTI Health Services, 2005.

2. **Bishop C.** *Neural Networks for Pattern Recognition*. New York, NY: Oxford Publishing, 1995.

3. **Bland JM and Altman DJ.** Regression analysis. *Lancet* 1: 908-909, 1986.

4. **Brage S, Brage N, Franks PW, Ekelund U, and Wareham NJ.** Reliability and validity of the combined heart rate and movement sensor Actiheart. *European journal of clinical nutrition* 59: 561-570, 2005.

5.   **Chen KY, Acra SA, Majchrzak K, Donahue CL, Baker L, Clemens L, Sun M, and Buchowski MS.** Predicting energy expenditure of physical activity using hip- and wrist-worn accelerometers. *Diabetes technology & therapeutics* 5: 1023-1033, 2003.

6.   **Chen KY and Sun M.** Improving energy expenditure estimation by using a triaxial accelerometer. *J Appl Physiol* 83: 2112-2122, 1997.

7.   **Chong E and Zak S.** *An Introduction to Optimization*. New York: Wiley, 2001.

8.   **Crouter SE, Clowers KG, and Bassett DR, Jr.** A novel method for using accelerometer data to predict energy expenditure. *J Appl Physiol* 100: 1324-1331, 2006.

9.   **Donahoo WT, Levine JA, and Melanson EL.** Variability in energy expenditure and its components. *Current opinion in clinical nutrition and metabolic care* 7: 599-605, 2004.

10.   **Freedson PS, Melanson E, and Sirard J.** Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine and science in sports and exercise* 30: 777-781, 1998.

11.   **Haykin S.** *Neural Networks: A comprehensive foundation*. Upper Saddle River, NJ: Prentice Hall, 1999.

12.   **Heil DP.** Predicting activity energy expenditure using the Actical activity monitor. *Research quarterly for exercise and sport* 77: 64-80, 2006.

13.   **Hendelman D, Miller K, Baggett C, Debold E, and Freedson P.** Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Medicine and science in sports and exercise* 32: S442-449, 2000.

14.   **Jakicic JM, Marcus M, Gallagher KI, Randall C, Thomas E, Goss FL, and Robertson RJ.** Evaluation of the SenseWear Pro Armband to assess energy expenditure during exercise. *Medicine and science in sports and exercise* 36: 897-904, 2004.

15.   **Nichols JF, Morgan CG, Chabot LE, Sallis JF, and Calfas KJ.** Assessment of physical activity with the Computer Science and Applications, Inc., accelerometer: laboratory versus field validation. *ResQExercSport* 71: 36-43, 2000.

16.   **Rowlands AV, Thomas PW, Eston RG, and Topping R.** Validation of the RT3 triaxial accelerometer for the assessment of physical activity. *Medicine and science in sports and exercise* 36: 518-524, 2004.

17.   **Schmitz KH, Treuth M, Hannan P, McMurray R, Ring KB, Catellier D, and Pate R.** Predicting energy expenditure from accelerometry counts in adolescent girls. *Medicine and science in sports and exercise* 37: 155-161, 2005.

18.   **Strath SJ, Brage S, and Ekelund U.** Integration of physiological and accelerometer data to improve physical activity assessment. *Medicine and science in sports and exercise* 37: S563-571, 2005.

19.   **Sun M, Reed GW, and Hill JO.** Modification of a whole room indirect calorimeter for measurement of rapid changes in energy expenditure. *J Appl Physiol* 76: 2686-2691, 1994.

20.   **Swartz AM, Strath SJ, Bassett DR, Jr., O'Brien WL, King GA, and Ainsworth BE.** Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Medicine and science in sports and exercise* 32: S450-456, 2000.

21.   **Theodordis S and Koutroumbas K.** *Pattern Recognition*. New York, NY: Elsevier Academic Press, 2003.

22.   **Tryon W and R W.** Fully proportional actigraphy: A new instrument. *Behavior Research Methods Instruments & Computers* 28: 392-403, 1996.

23.   **Tryon WW and Williams R.** Fully proportional actigraphy: A new instrument. *Behavior Research Methods Instruments & Computers* 28: 392-403, 1996.

24.   **Welk GJ, Blair SN, Wood K, Jones S, and Thompson RW.** A comparative evaluation of three accelerometry-based physical activity monitors. *Medicine and science in sports and exercise* 32: S489-497, 2000.

25.   **Zhang K, Pi-Sunyer FX, and Boozer CN.** Improving energy expenditure estimation for physical activity. *Medicine and science in sports and exercise* 36: 883-889, 2004.

26.   **Zhang K, Werner P, Sun M, Pi-Sunyer FX, and Boozer CN.** Measurement of human daily physical activity. *Obesity research* 11: 33-40, 2003.

CHAPTER VI

CONCLUSIONS AND FUTURE WORK

Summary

Prediction of EE using accelerometers has been a widely researched topic for the last quarter of a century. In this thesis we have presented a systematic approach to understanding this technology in a diverse subject population using a room calorimeter as a reference measurement. As our first step, we explored model performance based on several market-available portable accelerometers' ability to predict overall physical activity levels over the nearly 24 hour measurement period. The results from this validity study showed significant mismatches between EE predictions using first generation accelerometers and measurements made by the room calorimeter.

We then decided to take the same activity monitors to the bench and comprehensively characterize the contribution of the data filtering process and inter-unit variability when known acceleration loads are applied. These results revealed a limitation in the current sensors and how they process the acceleration signals since the output is not consistent across monitor generations. Further studies using additional activity monitors revealed that the type of hardware filtering is different in different brands of activity monitors and that not all monitors exhibit monotonic behavior, which introduces a challenge in modeling the full scale EE response to activity measured with these accelerometers.

Finally, we developed a new EE prediction approach using accelerometers. We began by deviating from the standard practice of using the time-integrated PA counts and

tested a new paradigm of modeling based on alternative features of the raw accelerometer data. In our final study, we have demonstrated a novel theoretical framework for this approach, using carefully-conducted clinical studies in a large and heterogeneous adult population, with a room calorimeter as gold-standard, and advanced modeling techniques (ANN) in both individual and generalized models.

Through these experiments, we discovered that existing accelerometer models, both linear and nonlinear, are only able to achieve accuracy in predicting some types of summary statistics relating to daily PA, such as the total time spent in each of four MET ranges; however they tended to underestimate the total daily PAL (Chapter III). While some of these errors are do to the functional form relating acceleration to EE, some of these errors could be attributable to the monitor sensitivity or device variability, where between-device CV was sometimes greater than 5% (Chapter IV). This variability is an important contributor to measurement error when multiple monitors are in use, which is common practice in field investigations. We were also able to determine that the ActiGraph monitors demonstrated a nonlinear relationship between activity counts and acceleration with decays in response to high accelerations (Chapter IV), while Actical and RT3 monitors showed increasing activity counts over all accelerations tested (Appendix C). When activity counts were normalized and the monitor response profiles could be compared while neglecting the scaling differences specific to each device, the RT3 and Actical showed similar response profiles. Divergence in the responses to standardized movement inputs delivered by mechanical oscillations between the different types of accelerometers suggests that unique modeling approaches may be needed to relate the output from each monitor to EE.

Finally, we developed a new model using raw data acquired using a hip-mounted bi-axial accelerometer and analyzed using a three layer ANN developed using pooled data from 102 subjects. We demonstrated significant improvements by this approach in predicting EE as seen in reductions in MAE, MSE, and percent difference in TEE (Chapter V). This error reduction was realized relative to traditional linear regression based approaches but was also a significant improvement over a proprietary model using the IDEEA model, which incorporates measurements from an array of five accelerometers. This model demonstrated improvements relative to the linear regression approach even when no subject characteristics were used in model development (Appendix D). Further reductions in error were observed when small individual ANN models were developed (Appendix E); however, development of these models required subject specific training data which require individual calibrations that could hinder field applications. By developing a new generalized model with hip-only data, we have shown an improvement in EE prediction without increasing the burden on the subject. Additional validation and model testing in different subject populations are needed to confirm these promising early results.

Conclusions

In conclusion, we made several advancements in the use of accelerometers through this thesis. The following generalizations can be drawn from these results:

1. Uni-axial accelerometers coupled with linear regressions are generally capable of predicting the time spent in four intensity categories in 24 hour data; however, the

intensity of most types of activity is under predicted leading to a systematic under

prediction in daily PAL if the regression has a physiologically reasonable intercept.

2. Physical activity monitor response should not be assumed to be consistent across

hardware revisions without robust mechanical testing. Changes in the filtering across

monitor revisions have the potential to increase EE prediction error as the same

accelerations are not represented by equivalent device output.

3. Correlations between acceleration and EE are not limited to the time-integral of the

signal. Quantities that characterize both the intensity, variability, and frequency

components of the acceleration signals may be useful.

4. EE is a complex physiological process that cannot be effectively modeled using a

single regression for all subjects. For generalized modeling, a flexible approach such as

ANN may improve the precision of EE predictions.

Future Work

Further work is needed to increase our understanding of the ANN and other

prediction models explored in these studies. First, the existing ANN should be

deconstructed to determine the relative contribution of each of the current features to the

EE prediction. Once this has been accomplished, the feature extraction routine should be

revisiting to attempt to discover additional signal characteristics that could further

improve the prediction of EE. Work should also be expanded to predict EE_{ACT} or METs.

Also, the exact architecture of the network should be validated using more robust statistical validation procedures. It is possible that observed errors could be further reduced in a hip-only model with the addition of new features or a change in the number of free parameters, or the shape of the basis functions.

Model generalization should also be further explored. Though a leave-one-out cross validation was pursued in the general hip-only model presented, all subjects were asked to perform similar PA types, even though variability was introduced by allowing subjects to self-select the intensity levels. Data should be acquired from new modes of PA spanning a wide variety of intensities. Additional subjects, particularly obese subjects and subjects of minority ethnic heritage would also be a welcome addition to the data set. Once the model has been well validated against calorimeter data, it should be validated against total EE in free living over a significant period of time (7-10 days) using DLW to determine the practical use of these models.

While the IDEEA monitor was an ideal test instrument to develop models because of the capability to collect raw data for more than a day, in the future it may be beneficial to develop models that can be used on ActiGraph monitors or other commercially available single sensor units. These monitors are more familiar to researchers, are produced and marketed at a much lower price, and do not have wires connecting sensors to a central microprocessor. Capitalizing on these advantages should decrease the burden on both investigators and subjects while still capturing adequate information for high-precision EE predictions. Currently, the ActiGraph is capable of recording raw data (10 Hz) for only several hours. However, if a feature set based on high frequency voltage signals can be established, and this set of features is comprised of signal components that

can be calculated by hardware, there would be no need to store the raw signals, merely a small subset of signal characteristics per minute. This would allow the ActiGraph, or several other commercially available monitors, to incorporate modeling approaches that capitalize on the benefits of the raw signals in their analysis software without increasing the on-board memory or decreasing the battery life.

As the rate of obesity increases in the western world, the interest in understanding the process of maintaining healthy body weight has become increasingly important. Because PA is a component of energy balance that can be readily modulated it has become a key factor in both individual weight loss prescriptions and public health recommendations. In spite of its widely recognized importance, the ability to quantify patterns of PA has been limited by measurement technology that is often unable to render accurate predictions of EE over the course of days or weeks. This thesis advances the goal of characterizing PA and its associated EE by characterizing the strengths and weaknesses of existing technologies and regression equations but also providing proof of concept of an improved EE prediction approach. This work can ultimately be used to give researchers and physicians insight into the exercise choices (duration, frequency, and intensity of PA) made by their patients and help them to better prescribe an appropriate course of PA to achieve optimal health benefits.

APPENDIX A

ARTIFICIAL NEURAL NETWORK ALGORITHM

For the EE prediction modeling presented in this thesis, artificial neural networks (ANN) were the chosen solution technique. In this appendix, the analytic framework of the feed-forward/back propagation network, as described in Chapter II and implemented in Chapter V of this thesis, will be presented. For all models in this thesis a three layer network was chosen (one hidden layer). Generalized derivations will be presented with additional information added about our specific modeling choices. Some schematic figures are provided to be conceptual representations of the actual networks used in this work because visual representation of our actual multi-layer network is cumbersome; however the written descriptions are accurate to the models used for the experiments.

Implementation of any ANN model begins by specifying the number of layers, the number of computational nodes in each layer (neurons), the basis function (transfer function) used between the layers, the learning rule, and the error function. Additional parameters such as a learning rate or momentum term may be needed based on the learning rule specified (1). The number of layers and nodes is determined empirically, though it is best to choose the smallest network possible to solve the problem. Computational layers can either be classified as input (first), output (final), or hidden layers. Hidden layers are intermediate steps between inputs and outputs and thus the output of the layer is "hidden" from the end user (3).

Figure A1 is a schematic of a two layer network. The feed-forward operation of the network uses each of the inputs (x) and the current state of the weight vectors (w) to predict the network output. The back-propagation step allows the network to iteratively update the weight values until a solution is found that optimally matches all the inputs patterns to the network outputs.
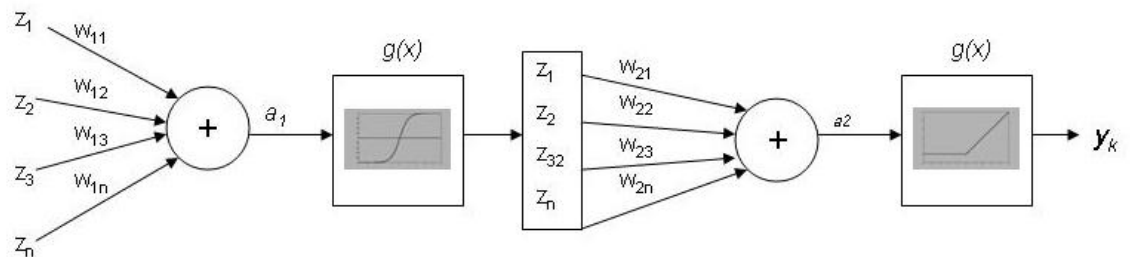

Figure A1: Schematic for the feed forward network. Each input (z) is multiplied by the layer weights (w) and sent through a summation (a) and transformation (g(x)) to the next layer. The process is repeated until the output (y) is reached. The first basis function is a hyperbolic tangent function, while the second is a positive linear basis function.

Notation and derivations are adapted from standard texts of neural networks (1, 3, 4).The net input into each basis function represents the summation of the layer inputs multiplied by the connection weight values between the inputs and the neurons (Equation A.1). Figure A1 shows a single summation site, however, any number of computational nodes can be selected and each neuron would have its own connection weights, which would be related the inputs to the neuron. Weights can either be connected to all computational nodes (fully connected), as we explored in this problem, or the can be connected only to certain neurons as dictated by the user *a priori* (1).

$$a_j = \sum_i w_{ji} z_i \qquad \text{(A.1)}$$

The output from each summation node is transformed by a basis (transfer) function (Equation A.2), which serves to both restrict the magnitude of the layer outputs but also can be used to introduce nonlinearities into the model. The outputs from each basis function form the input vector for the next layer of the network.

$$z_j = g(a_j) \qquad \text{(A.2)}$$

A single basis function or multiple basis functions can be chosen for each layer depending on whether any prior knowledge may suggest that a specific transformation should mediate certain interactions. In our model, a hyperbolic tangent basis function (A.3) was chosen for the input and hidden layer and a positive linear basis function was chosen for the output layer (A.4). The hyperbolic tangent function was chosen because it allows both positive and negative correlations in the data to be realized. This is important because our model includes factors known to be both positively (e.g., increases in limb movements) and negatively (e.g., age) correlated with EE. The positive linear basis function was chosen because it sets a physiological lower limit to predictions from the network (EE cannot be negative). While it does not impose a physiological upper limit, the training data should constrain the upper limit to reasonable values.

$$g(x) = \frac{\left(e^x - e^{-x}\right)}{\left(e^x + e^{-x}\right)} \qquad \text{(A.3)}$$

126

$$g(x) = x \qquad x \geq 0$$
$$g(x) = 0 \qquad x < 0$$

$$(A.4)$$

The process by which network weights are optimized is known as training. Training can either be supervised, meaning that there is a known output for each input pattern, or unsupervised, meaning that the iterative weight updates are an attempt to establish partitions in a data set where relationships within the group of data are not explicitly known (4). Because EE determined by the whole-room indirect calorimeter is recognized as a gold standard value, supervised learning was implemented in our modeling efforts. Two weight update structures exist: batch and online updating. In batch mode, all patterns are presented to the network before any updates are made, where in the online mode weight changes are made as each input is show to the network. Because weight updates are very computationally demanding and we have a large number of input patterns, we chose batch updating for this study.

To begin training, each weight value is assigned a small random value. The entire feed-forward network computation is made and the output for each pattern given the current state of the weight values is obtained. This value is compared to the known solution, or target, and an error term is computed. We chose to optimize on mean squared error (MSE), where the squared error can be computed for each input pattern (A.5) and averaged over all patterns (A.6). MSE causes large prediction errors (greater than 1 kcal/min) to be amplified, which should help in driving the optimization process to fit high intensity data, which may not be represented by as many training examples as other intensity categories. The functional form is also differentiable, which is desirable for

back-propagation, and because the metric is based on a mean, the magnitude of the predicted error is independent of the number of training points.

$$E^n = \frac{1}{2}\left(y\left(x^n; w\right) - t^n\right)^2 \qquad \text{(A.5)}$$

$$E = \frac{1}{N}\sum_n E^n \qquad \text{(A.6)}$$

Error back-propagation reduces the global error by updating the weights from the output end of the network to the input end. In order to determine the appropriate weight update for each individual weight, we must determine the amount of error associated with it. This operation can be expressed using the chain rule of differentiation as the product of the partial derivative of the error for each pattern as a function of the layer inputs, and the partial derivative of the layer inputs as a function of a specific weight value in that layer (A.7).

$$\frac{\partial E^n}{\partial w_{ji}} = \frac{\partial E^n}{\partial a_j}\frac{\partial a_j}{\partial w_{ji}} \qquad \text{(A.7)}$$

We will define the term delta (δ) to express the error for each layer as being the partial derivative of the error for a particular input pattern as a function of the layer inputs (A.8).

$$\delta_j \equiv \frac{\partial E^n}{\partial a_j} \qquad (A.8)$$

By rearranging the terms of the summation of inputs into a computational node (A.1), we can characterize the partial derivative for the layer as a function of each weight value (A.9).

$$\frac{\partial a_j}{\partial w_{ji}} = z_i \qquad (A.9)$$

Combining the terms in A.8 and A.9 we see that the error for each weight is function of δ, which is the derivative of the output of each layer, and z_i, which is the input value to a specific weight (A.10).

$$\frac{\partial E^n}{\partial w_{ji}} = \delta_j z_i \qquad (A.10)$$

In practice, for the output layer, the term δ is a function of the derivative of the output layer transfer function, and the derivative of the error function with respect to each output value (pattern). Thus, if the error function and transfer function are both differentiable, the δ for the output layer can be calculated simply by multiplying the two values (A.11).

$$\delta_k \equiv \frac{\partial E^n}{\partial a_j} = g'(a_k)\frac{\partial E^n}{\partial y_k} \qquad \text{(A.11)}$$

The output layer for our model uses a positive linear basis function, whose derivative is specified by Equation A.12.

$$\begin{aligned} g'(x) &= 1 \qquad & x \geq 0 \\ g'(x) &= 0 \qquad & x < 0 \end{aligned} \qquad \text{(A.12)}$$

The derivative of the MSE function (A.5) is specified by Equation A.13.

$$\frac{\partial E^n}{\partial y_k} = (y_k - t_k) \qquad \text{(A.13)}$$

For the hidden layers computation is slightly more complicated because the error is dependant on both inputs into the layer and errors from the output layer, which are a function of the computational nodes and basis functions in that layer rather than just a single output value as is the case in the output layer (A.14). Thus, the δ for the hidden layer contains the output layer δ (first term). Computation of the output layer delta is therefore essential for computation of the hidden layer error terms, which motivates the back-propagation strategy. The hidden layer delta is more complicated to compute than the output layer since it is modulated by the summation values on both the input and output side of the layer.

$$\delta_j \equiv \frac{\partial E^n}{\partial a_j} = \sum_k \frac{\partial E^n}{\partial a_k} \frac{\partial a_k}{\partial a_j} \qquad \text{(A.14)}$$

By using the general definition of δ (A.8) and substituting Equations A.1 and A.2 into Equation A.14, the delta for a hidden layer can be realized (A.15).

$$\delta_j = g'(a_j) \sum_k w_{kj} \delta_k \qquad \text{(A.15)}$$

Our input and hidden layers were specified by a hyperbolic tangent function, whose derivative is specified by Equation A.16.

$$g'(x) = 1 - \frac{\left(e^x - e^{-x}\right)^2}{\left(e^x + e^{-x}\right)^2} = 1 - g(x)^2 \qquad \text{(A.16)}$$

The derivative of the total error with respect to any weight value can be realized as the summation of the errors with respect that weight over all the training patterns (A.17).

$$\frac{\partial E}{\partial w_{ji}} = \sum_n \frac{\partial E^n}{\partial w_{ji}} \qquad \text{(A.17)}$$

Once the error derivatives have been computed, the training iteration is completed

by updating all the weight values (A.18). The form of the weight update (Δw) depends

on the learning rule.

$$w^{(\tau+1)} = w^{(\tau)} + \Delta w^{(\tau)} \qquad \text{(A.18)}$$

For this problem, a simple gradient descent technique was chosen (2). In this

technique, the direction of the weight change is specified to be the negative of the error

gradient at the current weight state (A.19). The step size is determined by a user specified

learning rate. This is a small value ($0 < \eta \leq 1$), which slows solution progress so that

there is steady progress down the error surface.

$$\Delta w^{(\tau)} = -\eta \nabla E \big|_{w^{(\tau)}} \qquad \text{(A.19)}$$

Gradient descent is the simplest optimization scheme for updating network

weights. It is also a slow search routine as it, by design, takes small incremental steps

down the error gradient rather than searching down the entire line of the gradient for an

ideal step size, or other more sophisticated options. However, it is simple to implement

and does not require the storage of large variables, which is a limitation of some solution

techniques that are able to speed convergence (e.g. conjugate gradient).

Termination of model training, or model convergence, was based on several

factors. The ideal outcome would be for error to reach a value of zero in the training, and

that a zero value would also be maintained in any future testing data. However, in

practice, this is unlikely to ever be the case so a set of three criteria were established. The first was the size of the error gradient. If the error gradient fell below an empirically determined limit (in our case, 1e-6), training was terminated. Small error gradients indicate that each weight update is small (because the gradient size is further reduced by the learning rate), and it is unlikely that meaningful improvement in the solution is being obtained. Small error gradients can be caused by a local error minimum or can be reached at the global minima so the results of terminations based on gradient size must be carefully analyzed. Training was also stopped if the training error increased for five or more successive iterations as this indicates the direction of the weight updates is not improving the quality of the solution, or if the change in the training error decreased less than 1e-6 for a training iteration. Finally, if the testing error, derived from subject data not used to derive the weight updates, increased for five successive iterations, or if the change in the testing error decreased less than 1e-6 for a training iteration, training was terminated, as this was designated as over fitting the model to the training data. Figure A2 is a flow chart summarizing the process of network training.
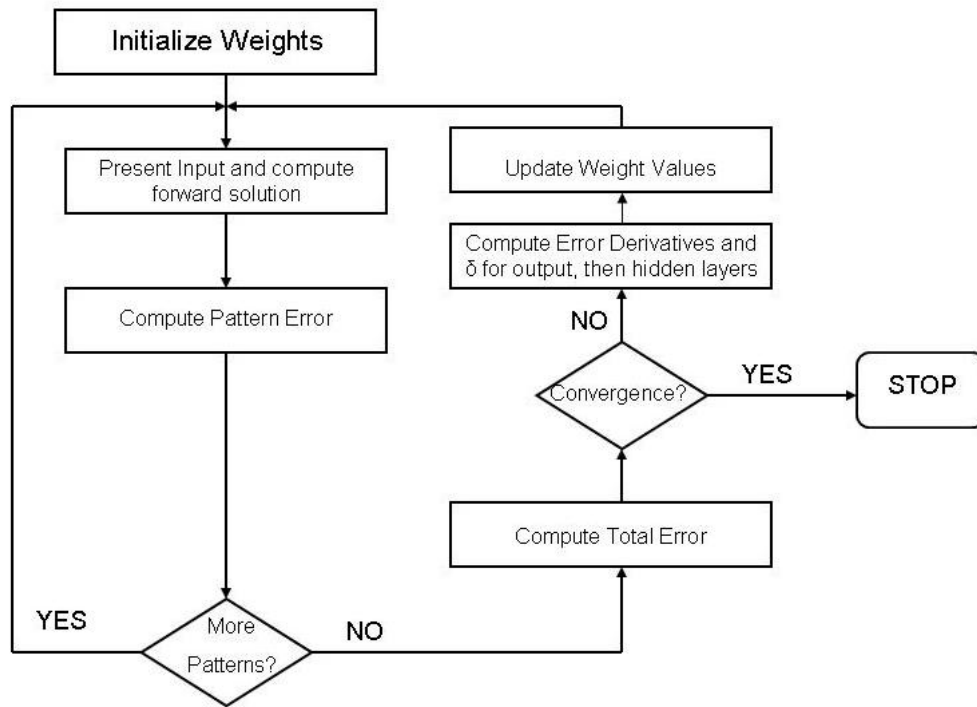
Figure A2: Flow diagram for steps in network convergence. Weights begin with an initial random assignment. Each input is used to determine a pattern error until all patterns have been presented. This total error is used to compute error derivatives and weight updates. This process is repeated until the mode converges.

Model performance was tested using leave-one-out cross validation. In this approach, the network weights are allowed to converge on data from all but one subject (N-1), while data from the single subject that was withheld from the training is used to estimate the generalization performance. This approach is repeated for each subject and generalization error is reported as the mean, standard deviation, and range of the resulting error values. Results of the ANN experiments are presented in Chapter 5 as well as in Appendices D and E.

References

1.   **Bishop C.** *Neural Networks for Pattern Recognition.* New York, NY: Oxford Publishing, 1995.

2.   **Chong E and Zak S.** *An Introduction to Optimization.* New York: Wiley, 2001.

3.   **Haykin S.** *Neural Networks: A comprehensive foundation.* Upper Saddle River, NJ: Prentice Hall, 1999.

4.   **Zurada J.** *Introduction to Artificial Nerual Systems.* St Paul, MN: West Publishing, 1992.

APPENDIX B

ENERGY EXPENDITURE PREDICTION USING ACTICAL AND RT3

In order to completely understand the prediction capabilities of the regression

equations in Chapter III, we expanded the scope of analysis to include direct prediction of

gross EE using the Actical (4) and RT3 equations (1). The Freedson ActiGraph equation

(AG 1) (2) predictions will not be contained here as this information can be found in

Chapter V. The other two ActiGraph equations considered in Chapter III, the Hendelman

equation (AG 2) (5) and the Swartz equation (AG 3) (6) are designed for making

categorical MET predictions. Because they were designed with this goal in mind, they

were allowed to have non-zero y-intercepts, which introduce significant bias into the

predictions of gross EE. The overestimated average physical activity levels (PAL) that

we presented and discussed in Chapter IV were a direct result from this. Thus, only the

results from the two Actical and two RT3 equations will be explored here.

For each regression equation, the same summary statistics computed for the ANN

model were computed (Figure B1). These include the correlation coefficient (r^2), MAE,

MSE, and the percent difference in TEE between each prediction and the measured EE.

All regressions showed high correlation with the measured EE. All MAE and MSE

values were smaller than those computed with the Freedson equation. Both the Actical

equations and the Chen RT3 equation (RT3 2) showed MAE and MSE less than the

IDEEA model but greater than the ANN. The percent difference measurements were

comparable to the ANN for all models except the proprietary RT3 model (RT3 1).

136

Table B1: Summary statistics relating to EE prediction in Actical (AC) and RT3 monitors.

	r^2	MAE (kcal/min)	MSE (kcal^2/min^2)	% Difference
AC 1	0.84 ~ 0.11 (0.42 ~ 0.93)	0.37 ~ 0.11 (0.25 ~ 0.63)	0.35 ~ 0.26 (0.15 ~ 1.50)	3.58 ~ 4.06 (0.03 ~ 17.05)
AC 2	0.83 ~ 0.15 (0.46 ~ 0.94)	0.35 ~ 0.08 (0.24 ~ 0.65)	0.29 ~ 0.20 (0.13 ~ 1.20)	6.04 ~ 4.66 (0.23 ~ 16.55)
RT3 1	0.83 ~ 0.10 (0.52 ~ 0.93)	0.39 ± 0.12 (0.27 ~ 0.66)	0.49 ~ 0.36 (0.15 ~ 1.82)	6.88 ~ 7.00 (0.47 ~ 20.24)
RT3 2	0.83 ~ 0.07 (0.53 ~ 0.89)	0.32 ± 0.10 (0.23 ~ 0.57)	0.27 ~ 0.28 (0.11 ~ 2.25)	6.34 ~ 6.92 (0.25 ~ 17.02)

Because AC 1, AC 2, and RT3 2 are designed to EE_{ACT}, a baseline value was added in order to predict the gross EE for each minute. Initially, a measured RMR was used for these computations but large differences in TEE were observed. We assumed that this could be due to the fact that our RMR values were taken from sleep and as such are about 10% lower than the Harris-Benedict (3) predicted RMR for our subjects. In order to understand the impact of the baseline on EE predictions, the measured RMR was incrementally increased from the measured value to 40% greater than the measured value to understand how each error metric changes as a function of the baseline (Figure B1). The correlation coefficient is not impacted by baseline shifts, so no data for this metric is shown. Optimal results are achieved when the Actical baseline is utilized at 25% above the measured RMR, and the RT3 when RMR used in the model was 10% above the measured RMR, approximately equal to Harris-Benedict predictions.
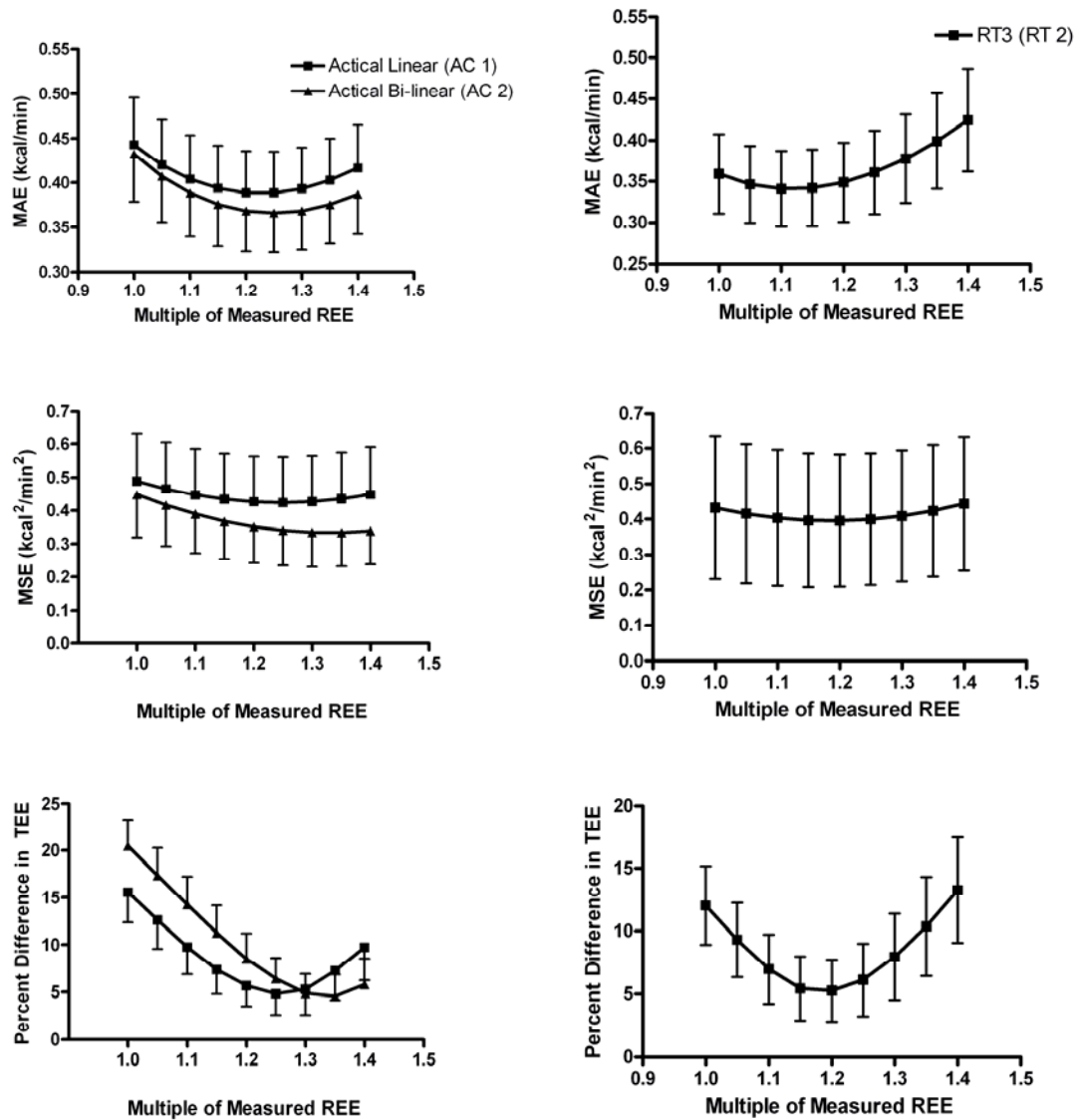
Figure B1: Prediction errors in two Actical equations and one RT3 equation are explored by altering the REE used for predicting minute-by-minute EE.

The results from varying the baseline suggest that transforming regression equations from EE_{ACT} prediction to total EE prediction has the potential to introduce large amounts on unintended errors. The data in total may be best represented by a high baseline, which may serve to balance out other problems in the model, such as under prediction of high intensity activities, and the inability of the accelerometers to detect

some sedentary PA types. The bi-linear Actical regression (AC 2) was particularly susceptible to alterations in the baseline. By optimizing the baseline, we were able to show substantial improvements in model performance, though this baseline shift is not well motivated if RMR is measured. This experiment serves to highlight the difficulties of correctly establishing a baseline and possibly explains the unrealistically high baseline values (y-intercepts) found in some regression equations, such as the Hendelman (AG 2) and Swartz (AG 3) ActiGraph regressions.

In our previous work with the regression equations presented in the literature for several accelerometers (Chapter III), we had considered only their ability to resolve the time spent in PA intensity categories, and their ability to correctly classify data into these categories on a minute-by-minute basis. By considering the EE prediction capabilities of these same regressions we were able to develop a greater understanding of the capabilities of different activity monitors and regression approaches to relate acceleration and EE. From this work we were able to determine that the Actical and RT3 models are more robust than the Freedson equation, and in same cases also more accurate than the proprietary IDEEA model. This work also was able to highlight the pitfalls of converting from EE_{ACT} predictions into gross EE predictions using RMR. In these cases the balance between selecting a physiologically reasonable baseline and improving measures of error must be carefully considered.

References

1.   **Chen KY and Sun M.** Improving energy expenditure estimation by using a triaxial accelerometer. *J Appl Physiol* 83: 2112-2122, 1997.

2.   **Freedson PS, Melanson E, and Sirard J.** Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine and science in sports and exercise* 30: 777-781, 1998.

3.   **Harris J and Benedict F.** A Biometric Study of Basal Metabolism in Man. Washington DC: Carnegie Institute of Washington, 1919, p. 1-266.

4.   **Heil DP.** Predicting activity energy expenditure using the Actical activity monitor. *Research quarterly for exercise and sport* 77: 64-80, 2006.

5.   **Hendelman D, Miller K, Baggett C, Debold E, and Freedson P.** Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Medicine and science in sports and exercise* 32: S442-449, 2000.

6.   **Swartz AM, Strath SJ, Bassett DR, Jr., O'Brien WL, King GA, and Ainsworth BE.** Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Medicine and science in sports and exercise* 32: S450-456, 2000.

APPENDIX C

OSCILLATION RESULTS FROM ACTIGRAPH, ACTICAL, AND RT3 PA

MONITORS

In addition to the extensive oscillation experiments performed on the ActiGraph

PA monitors, detailed in Chapter IV, similar experiments were performed on the Actical

(n = 5) and RT3 (n = 15) PA monitors to establish cross-device references. Frequency

experiment and inter-monitor variability results from these devices, as well as

comparisons with the ActiGraph, are presented here. The same orbital shaker from

Chapter IV was used for these experiments All results for the ActiGraph are for the

GT1M model (n = 11), which is currently the only marketed device, though many old

units (7164 and 71256) are still in clinical use.

When the mean frequency response is compared between PA monitor types,

differences in both the amplitude of the observed count values, and the functional forms

are detected (Figure C1). The nonlinear response of the ActiGraph has been well

documented in this thesis as well as in the literature. However, the observed response

profiles for the Actical and RT3, which to our knowledge have not been reported in the

literature, both show increasing values over the tested frequency range. The response of

the Actical is linear (r^2 = 0.98), though the low frequency response may have some

nonlinearity. The response of the RT3 appears to be nonlinear, with the bulk of the

curvature in the mid-frequency section.
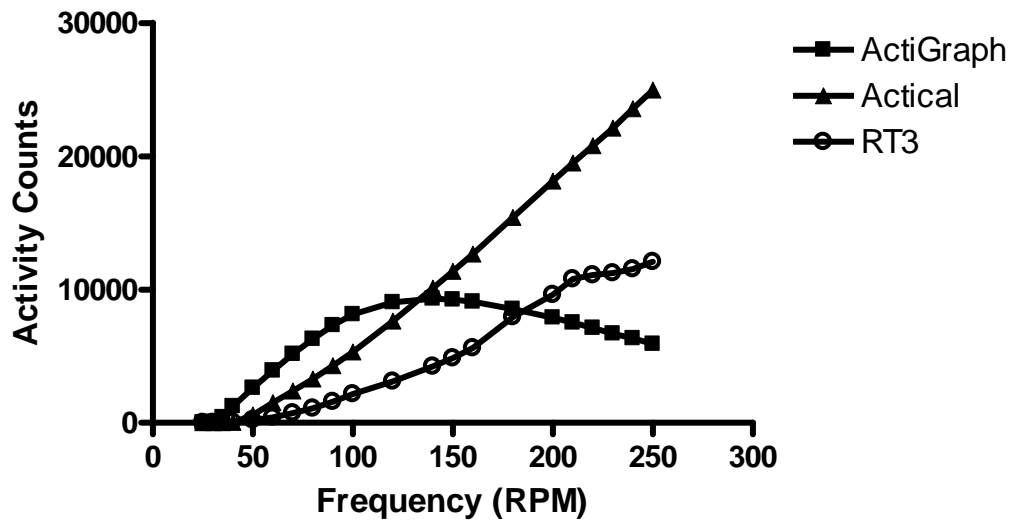
Figure C1: Plot comparing the frequency responses of ActiGraph, Actical, and RT3 PA monitor responses. Frequencies are mechanically generated to be between 25-250 RPM.

Unlike the ActiGraph, both the Actical and RT3 monitors have peak count values that correspond with the highest frequency tested, while the maximum count value for the ActiGraph occurs at 140 RPM with large magnitude of decay at higher frequency values. Because higher frequency human movements should not have a reduced energy cost if all other variables are fixed, we believe that a constantly increasing response or one that increases to an asymptote at high frequencies is the most physiologically reasonable response profile.

Each test point was also analyzed using the acceleration analogous to the test frequencies (Figure C2). This transformation is linear with respect to the radius of oscillation, but quadratic with respect to frequency and is specified in Chapter IV.
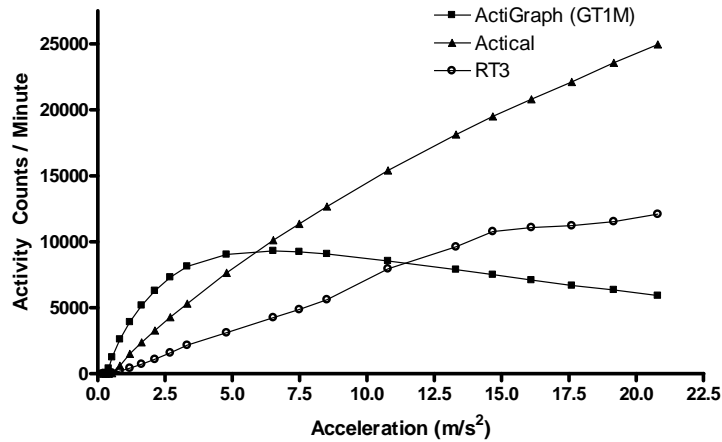
Figure C2: Activity count response of ActiGraph, Actical, and RT3 PA monitors as a function of applied acceleration.

The process of converting from frequency to acceleration does not seem to introduce significant changes in the shape of the response curves, with the exception of the Actical response, which now appears to be nonlinear.

It should be considered that differences observed in the count values themselves are not enough to indicate measurement differences. Counts are an arbitrary unit specified by each device manufacturer and therefore differences in scale are to be expected. It would be desirable to compare the response profiles directly between monitor types in order to attempt to determine if true measurement differences exist between the monitors. To begin to explore these comparisons, we normalized the activity counts of each monitor type by their respective maximum count values (Figure C3).
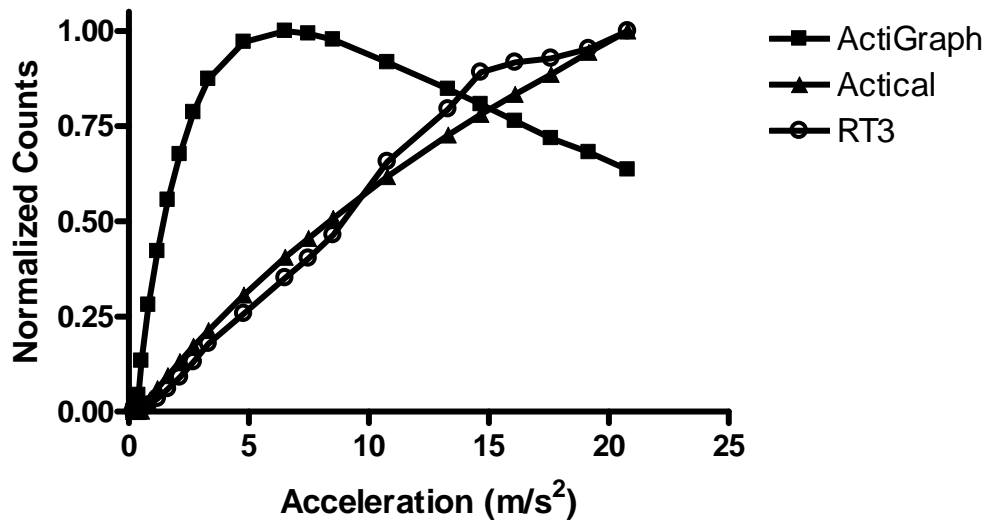
Figure C3: Comparison of the frequency response of ActiGraph, Actical, and RT3 PA monitors when activity counts have been normalized by the maximum count value achieved by monitor type.

The normalized count data seems to suggest that the response profiles of the Actical and RT3 are similar, though there is additional curvature in the RT3 data that does not appear in the Actical data. These profiles are notably different than that exhibited by the ActiGraph, which shows a more rapid increase to the peak value, and significant decay into the high acceleration region where both the Actical and RT3 are still increasing.

Another important characteristic of the response to mechanical oscillation which should be compared between PA monitors is their inter-monitor CV (Figure C4). This calculation expresses how much error is introduced into data analysis by using multiple units of the same type of PA monitor (ActiGraph, e.g.), which is common in field experiments where hundreds of subjects may be recruited for a study. When the full range of frequencies are considered, the data becomes difficult to interpret visually

because in the low frequency range some monitors of each type detect signal and some do not leading to large (CV > 100%) values. A reduced range (frequency > 50 RPM) was considered for visualization purposes.
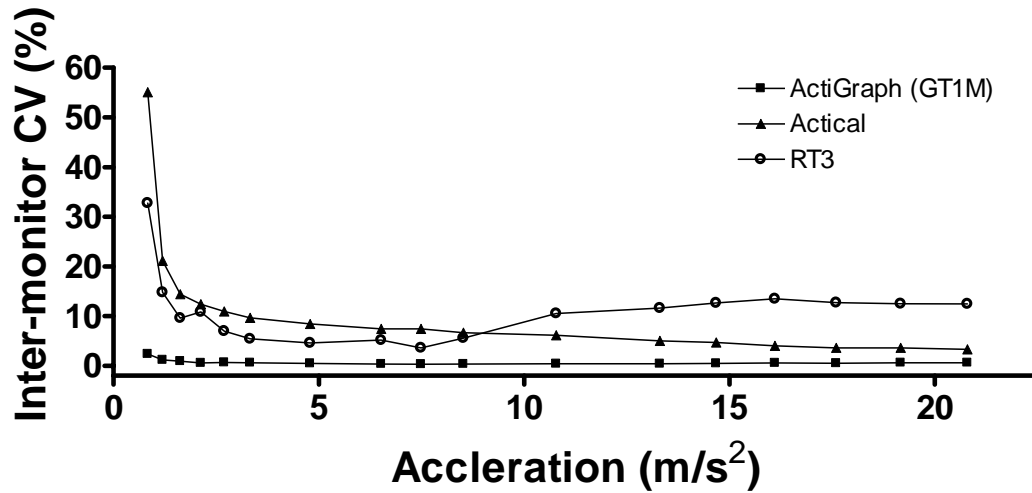
Figure C4: Inter-monitor CV (%) for ActiGraph, Actical, and RT3 PA monitors. Low frequencies (< 50 RPM) were not included because CV for at least one monitor type was greater than 100%.

The CV for the ActiGraph monitor is the lowest for all tested frequencies with values on the order of 0.5%, suggesting a significant advantage for population studies. The Actical shows a pattern of decaying CV as a function of frequency, with the mean value around 10% for frequencies between 60-150 RPM, and around 3% for higher frequencies. The RT3 showed minimum CV for mid frequency values (150RPM) where the CV was around 3.5%. However, the higher frequency measurements yielded CV on the order of 13%, which questions the reliability of measurements for this range of values. The larger CV for the RT3 monitors in the high frequency experiments suggests that some of the nonlinearities in the observed response curve could be attributable to

measurement variability in the sample of monitors and may not be directly related to the device filtering.

It should be considered that only a very small sample was used in the reported Actical results (n=5). This outcome stems from a hardware revision that was made on Actical devices midway through data collection when devices were returned to the manufacturer for calibration. We presented results only from monitors that had the newest device hardware revisions in order to increase the practical applicability of the results. This revision called for a wider range of frequency detection, which should allow higher intensity activities to be accurately characterized. However, the larger frequency band, in general makes the variability harder to control.

In these experiments, the ActiGraph, Actical, and RT3 monitors were subjected to mechanical oscillations in order to compare and contrast their response profiles. While the magnitude of the count values varied greatly between monitors, the Actical and RT3 monitors had similar response profiles, though both are different from the ActiGraph. However, the new ActiGraph had a much lower inter-monitor CV than either the Actical or RT3 monitors.

ADDITIONAL RESULTS: HIP-ONLY ANN

In this Appendix, additional details and results that pertain to ANN training, representations of the final weight values, and additional results representing the prediction accuracy of the ActiGraph, IDEEA, and ANN models within specific activity intensity categories was considered.

During 24-hr stay in the room calorimeter, subjects tend to spend >90% of their time in low intensity activities and the error in prediction models is heavily influenced by small minute-by-minute mismatches in EE predictions during these low intensity intervals. Differences between predictions and measurements for high intensity data can cause large increases in both the mean of the minute-by-minute error measures and large mismatches in total EE even though they contribute only a small percentage of the patterns in the data set. Thus, each of the ANN models was trained on data that has been biased towards the high intensity data by replicating all the data points that corresponded to minutes where the measured EE was greater then 6 kcal/min. These data were presented to the model three times per training iteration.

Feature Extraction

To extract the data features, raw acceleration data was synchronized with output data from the calorimeter by maximizing the correlation coefficient between a down sampled acceleration signal and the EE measured by the room calorimeter.

Table D1: Brief description of candidate features for the ANN model.

Parameter Name	Description	Relationship with Physical Activity
Median	Middle value in an ordered list	Central tendency of the measured activity
Integral	Left hand Riemann sum	Accumulation of measured activity
Peak	Maximum absolute value	Point estimate of the peak activity intensity
Inter-quartile Range	Upper quartile – lower quartile	Variability of movements
Skew	$$g = \frac{\sqrt{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{3/2}}$$	Is data distributed evenly among all PA intensities
Kurtosis	$$k = \frac{n\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2} - 3$$	Does the minute have a single large peak or is the activity more evenly distributed over all measured values
Mean Absolute Error	$$MAE = \frac{1}{n}\sum_{i=1}^{n} abs(x_i - \bar{x})$$	Measurement of the activity variability
CV (1)*	$$CV(1) = \min\left(\frac{\sigma}{\mu}\right)$$	Lowest variability during each minute
CV (2)*	$$CV(2) = \max\left(\frac{\sigma}{\mu}\right)$$	Highest variability during each minute
PSD (1)^	$$psd(1) = sum\left(\left|fft(x)\right|^2\right)\big|_0^{0.7}$$	The amount of total signal power associated with low intensity PA
PSD (2)^	$$psd(2) = sum\left(\left|fft(x)\right|^2\right)\big|_{0.71}^{10}$$	The amount of total signal power associated with moderate to vigorous PA

* Each minute of data was divided into six segments. The CV was computed for each 10 second interval and the smallest and largest values were retained. Analysis was intended to isolate transitions between activity types that may occur within one minute of data

^ 0.7 Hz was used to divide the data between signal power associated with sedentary activities and that associated with dynamic PA. Values between 0.1 and 1.0 Hz were considered and discrimination was tested using hip data only on a sample of five subjects whose data was not used for model development. The value of 0.7 proved to be the best value for these subjects.

Each minute of acceleration data was isolated and used to extract 11 features

(Table D1). These features were intended to represent the PA intensity and variability

over the minute as well as to detect of transitions between PA types happened during each minute, which might impact the overall EE. By computing the correlations between features, the 11 features were reduced to 5 by eliminating one feature of each pair whose correlation was higher than an arbitrarily specified 0.6.

Sample weight distributions

Model validation was performed by leave one subject out cross validation. This process resulted in 102 sets of optimized weighting coefficients. In Figure D1, Hinton plots of the final weight values for a representative subject were generated for qualitative evaluation of the various weighting coefficients between each set of layers of the model. Positive weight values are indicated by green boxes, while negative values are shown in red. The relative size of the box indicates the magnitude of the weight values. Each specific interaction between a layer input weight (x-axis) and the neurons in the next layer (y-axis) can be characterized examined by examining the magnitude and sign of the box specified for the x-y pair. The input characteristics in the order they appear on the x-axis of the input layer Hinton plot (Figure D1) is shown in Table D2.

Table D2: Order of characteristics for input layer Hinton plot.

Variable Number	Variable Name
1	Age
2	Height
3	Weight
4	BMI
5	Sex (Binary – Woman = 0)
6	Caucasian (Binary – Yes = 1)
7	African American (Binary – Yes = 1)
8	Peak (A\P)
9	Inter-Quartile Range (A/P)
10	Lowest 10 second CV (A/P)
11	Signal Power < 0.7 Hz (A/P)
12	Signal Power > 0.7 Hz (A/P)
13	Peak (M\L)
14	Inter-Quartile Range (M/L)
15	Lowest 10 second CV (M/L)
16	Signal Power < 0.7 Hz (M/L)
17	Signal Power > 0.7 Hz (M/L)

A/P – Anterior/Posterior

M/L – Medial/Lateral

For the input layer there are a number of weights that are close to zero, along with several features that seem to have no large weight values associated with them, e.g., features 8 and 10 (peak and lowest 10 second CV of the anterior/posterior data channel). The small magnitude of the weights associated with a single feature could indicate that that feature does not have a large contribution to the total solution. These input features may be good candidates for systematic elimination, which could be performed to decrease the computational complexity of the model. If, however, some of the weight values from an input to a neuron are small, it only indicates that the interactions with this input term are small for a particular neuron.
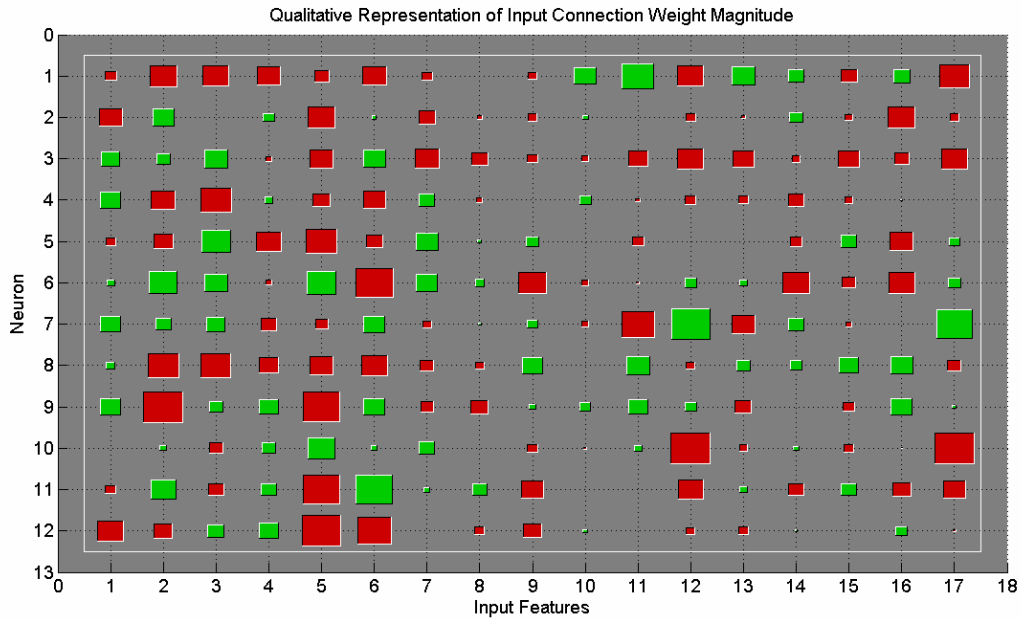
Figure D1: Plot of the signs and relative magnitudes of the weights relative the input features to the first layer of neurons.

In the second layer (Figure D2), fewer weight values have a small magnitude, even though a slightly larger number of total weight values are present. Additionally, there are no inputs to this layer that appear to have only small interactions with all of the neurons of this layer. This does not immediately suggest reducing the size of this layer. Relative to the input layer, the hidden layer shows a greater number of positive weights of high magnitude. This balance is likely due to the basis function limiting the response range from the input layer to values between -1 and 1, while the features used as model inputs were standardized to a mean of zero and standard deviation of one, which would allow some input features to be outside of the range of the basis function.
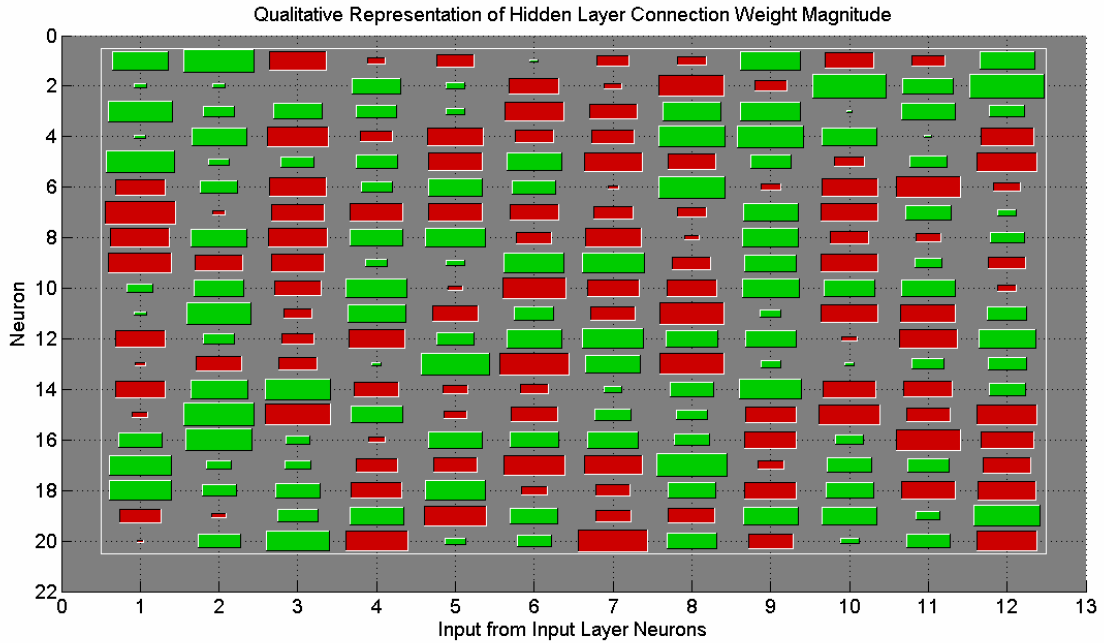
Figure D2: Schematic of the signs and relative magnitudes of the weights from the output end of the input layer to the hidden layer.

The output layer weight schematic shows the 20 weights connecting the hidden layer with the single EE value predicted for each minute (Figure D3). As with the hidden layer, the weights appear to be well balanced between positive and negative values with a variety of weight magnitudes.
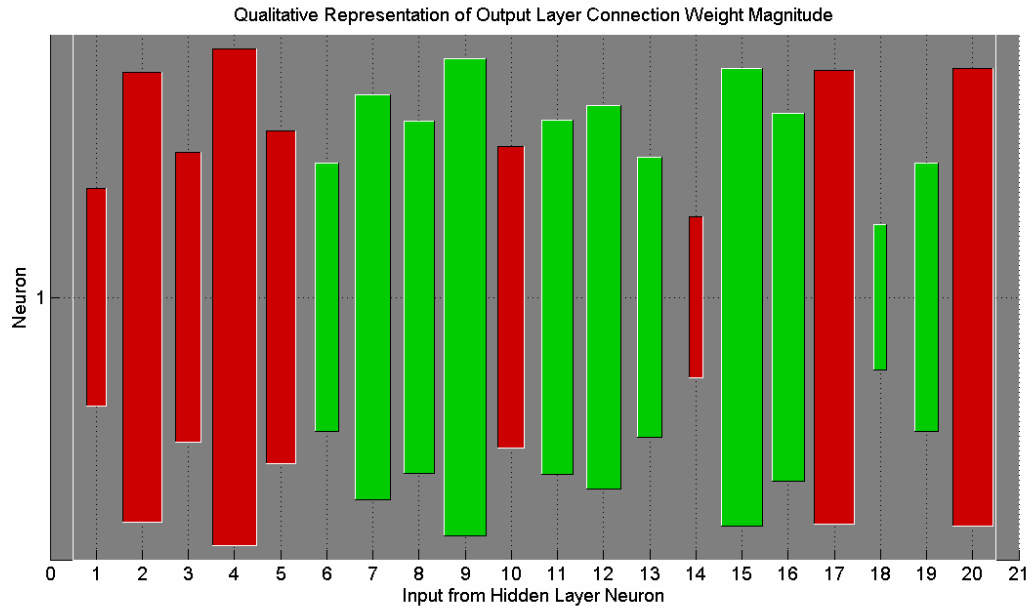
Figure D3: Schematic of sign and relative magnitude of the output layer weights.

Prediction by activity intensity levels

In a further attempt to understand the output data from the ANN model and how it relates to the outputs of the IDEEA model and the ActiGraph mode, the measured data from the chamber was divided into four intensity strata based on the measured resting metabolic rate (RMR). The four categories explored were 1-1.5 time RMR, designated as sedentary, 1.5-3 times RMR, light, 3-6 times RMR, moderate, and greater than 6 times RMR, intense. The mean of the measured and predicted EE for data in these intensity categories was computed and Bland-Altman plots were produced for each model and intensity class (Figure D4).
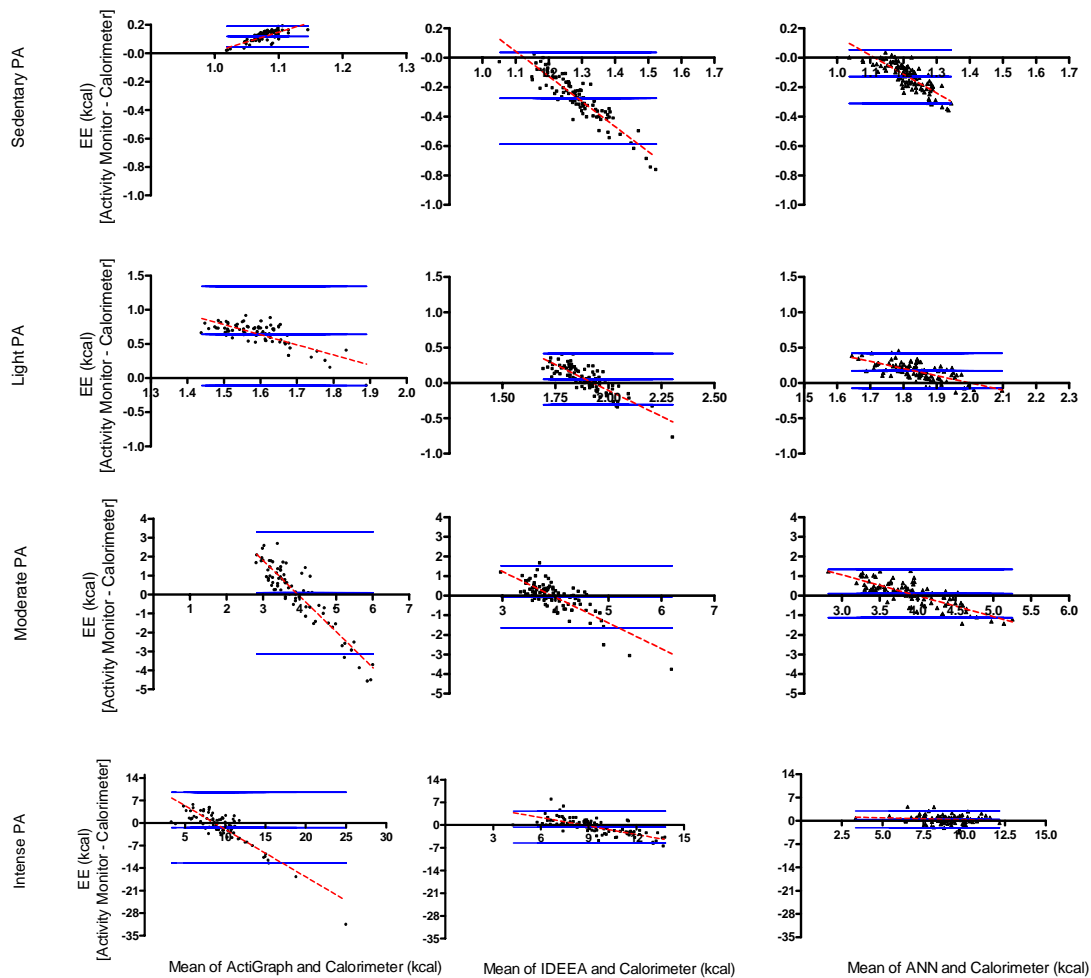
Figure D4: Bland-Altman plots for the ActiGraph, IDEEA, and ANN. Data from the calorimeter was discretized into four intensity categories and the mean of the corresponding data points were used to evaluate predictions in each intensity range.

The intensity stratified Bland-Altman plots show that the Freedson equation for the ActiGraph tends to under-predict sedentary and light PA, while on average the moderate and vigorous PA intervals are well characterized. However, there is a relatively high standard deviation (~1.5 kcals/min for moderate; ~5 kcals/min for vigorous) for these measurements as well as a trend to over predict as the mean intensity of activities contained in these intervals increases. The IDEEA monitor tends to over-predict the EE

associated with sedentary activities. This is likely due to the prediction equation used to establish the RMR. The mean values for the light, moderate, and vigorous EE categories are well represented and standard deviations are improved relative to the ActiGraph for all three classes. Like the IDEEA monitor, our ANN model over-predicted sedentary EE; however, the magnitude of this over-prediction is smaller (~ 0.10 kcals/min) than the IDEEA. The ANN showed the lowest standard deviation of the three models considered in all but the sedentary category (ActiGraph was lowest for this category). The small mean offsets combined with the balance of over-prediction of sedentary EE and under-prediction of light EE likely leads to the small observed differences in total EE.

Acceleration only ANN

In order to understand the potential benefits in EE prediction accuracy that can be realized by allowing interactions between subject characteristics and acceleration signal features within the ANN framework, the acceleration values were isolated and the model was retrained on a reduced (acceleration only) data set. The network architecture, learning rule, and convergence criteria were unchanged from previous experiments. The elimination of the characteristics caused the ANN to have less subject information than AG1 or the proprietary model. AG1 was supplied with a subject specific physiologically reasonable baseline value (measured sleeping EE), and body weight was used as a term in the model. The IDEEA model has access to subjects' sex, age, height, and weight along with signals from the five acceleration channels.

AG1, the proprietary IDEEA model, the acceleration only ANN (ANN 1) and the acceleration plus characteristics model (ANN 2) were compared using the correlation

coefficient (r^2), mean absolute error (MAE), mean square error (MSE), and percent difference in total measured EE. Statistical testing was performed using ANOVA for paired measurements with post hoc testing to identify pair-wise differences (Figure D5).

ANN 1 had a significantly higher r2 that AG 1 ($p < 0.001$) but was not significantly different from the IDEEA or ANN 2. The MAE for ANN 1 was significantly less than AG 1 ($p < 0.001$), not different from the IDEEA model, and was significantly higher than ANN 2 ($p < 0.001$). The same trend was observed in the MSE. The percent difference between the predicted and measured TEE was not significantly different between ANN 1 and either the IDEEA model or AG 1, but was significantly higher than ANN 2.
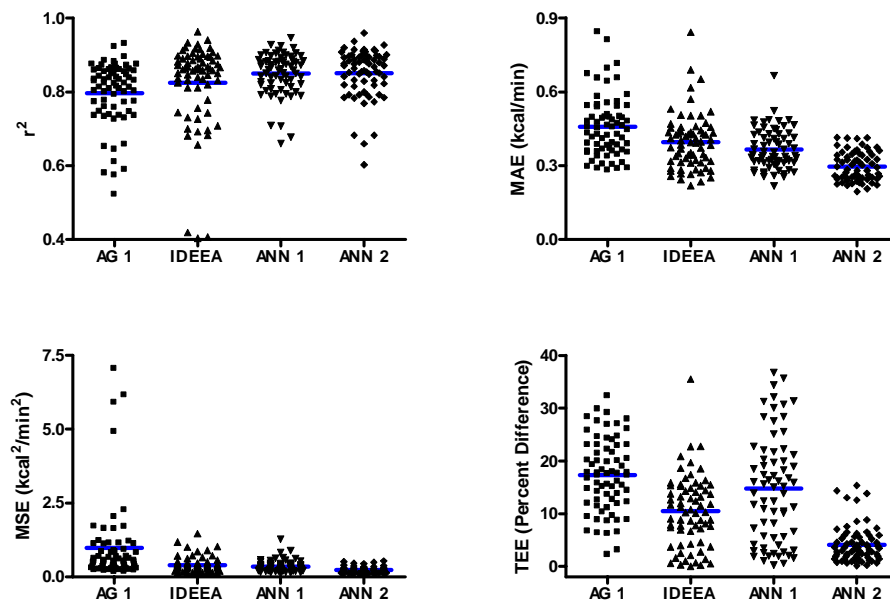
Figure D5: Comparison of error summary measures between the Freedson ActiGraph equation (AG1), the IDEEA proprietary model, an acceleration only ANN model (ANN1), and an acceleration plus characteristics ANN (ANN2).

A Bland-Altman plot of the TEE predicted using ANN 1 relative to the measured EE from the metabolic chamber (Figure D6) revealed that while the mean difference was small (~ 28 kcal), the variability in the response was large ($\sigma = 355$ kcal). This prediction variability was larger than that observed for AG 1, IDEEA, or ANN 2 (See Chapter 5).
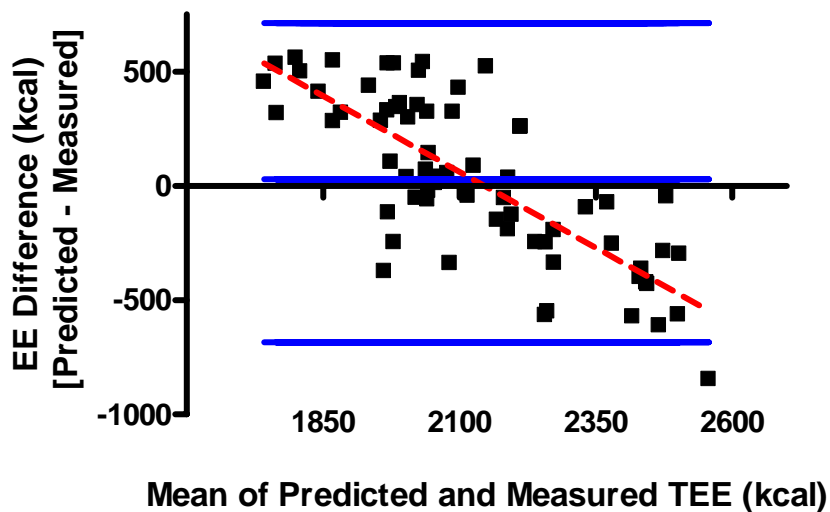
Figure D6: Bland-Altman plot for TEE measured using the metabolic chamber and predictions made using ANN 1

Raw data analysis (Figure D9-D12) showed errors in both the baseline detection and the amplitude of EE prediction are observed using ANN 1. The baseline value tends to be over-predicted in women and lean individuals (Figures D9-D10). While sleeping rest is well characterized in larger individuals (Figures D11-D12), the EE associated with the highest intensity activities is under-predicted. This likely results from the absence of body weight in

the training data, since for a given acceleration load, the energy cost is higher if more body

mass must be moved.
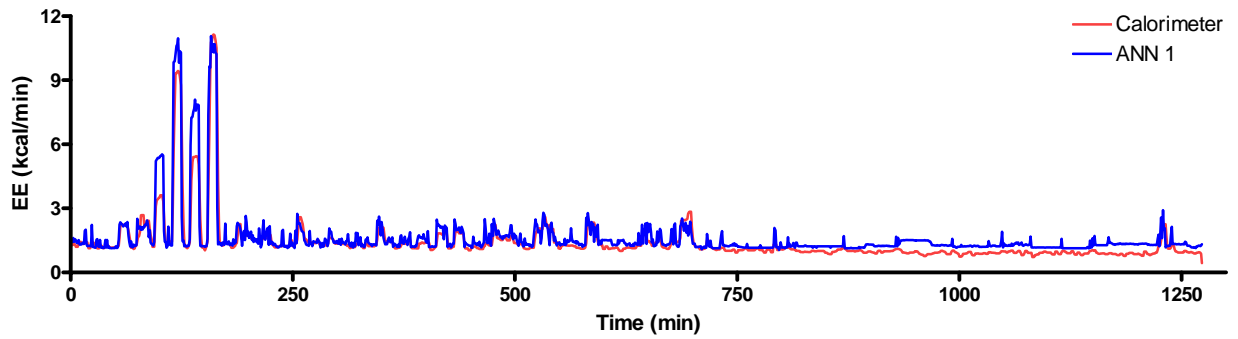
Figure D7: Minute-by-minute EE measurements and predictions for a 33year old female (height = 1.62 m, weight - 58.5 kg, BMI - 22.5 kg/m^2)
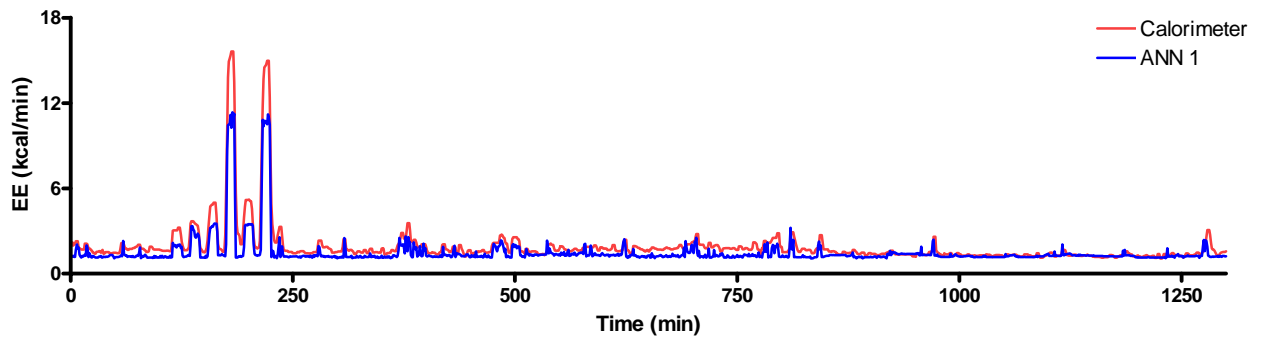
Figure D8: Minute-by-minute EE measurements and predictions for a 35 year old female (height = 1.70 m, weight = 63.2 kg, BMI = 21.9 kg/m^2)
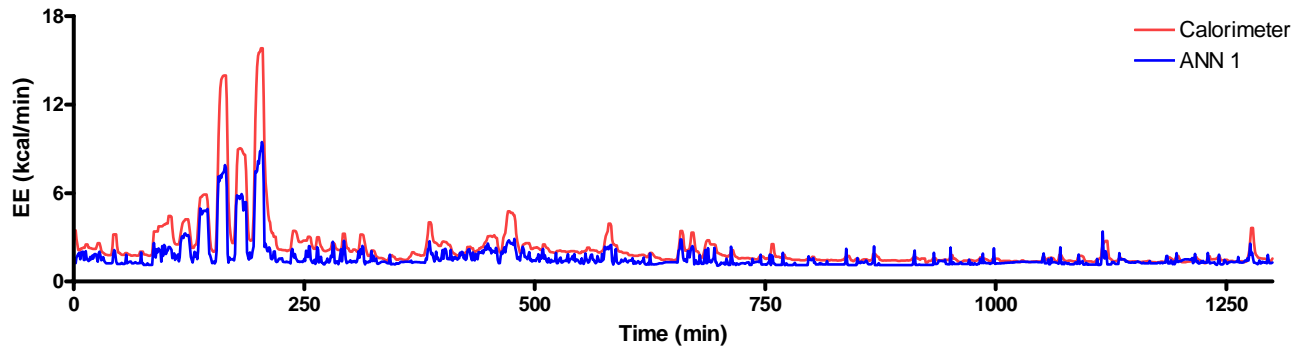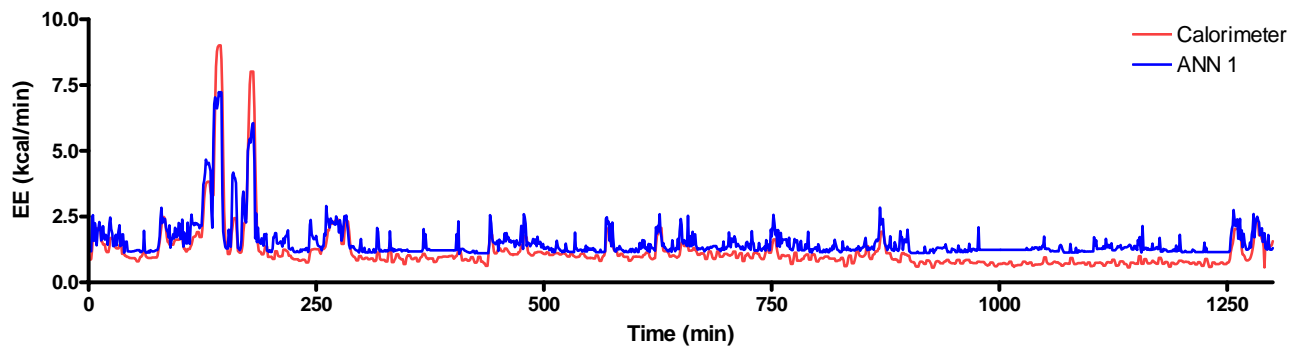
Figure D9: Minute-by-minute EE measurements and predictions for a 40 year old male (height = 1.75 m, weight = 101.5 kg, BMI = 33.14 kg/m^2)

Figure D10: Minute-by-minute EE measurements and predictions for a 48 year old male (height = 1.86 m, weight = 101 kg, BMI = 29.14 kg/m^2)

The acceleration only ANN model was developed to attempt to show that minute-by-minute EE prediction accuracy can be improved using features of the raw acceleration signal relative to models using integrated signals even if the feature based model does not contain any subject information. This was shown through the significant reductions in MAE and MSE observed between AG 1 and ANN 1. In most cases, the ability of the ANN model to predict the EE associated with low intensity EE was improve relative to AG 1. These changes are likely associated with spontaneous PA and characterizing them

may be useful. ANN 1 did not show improved minute-by-minute prediction accuracy relative to the proprietary IDEEA model, which may use subject characteristics in addition to the five channels of acceleration data. The major limitation in ANN 1 was its inability to resolve the RMR. This is not surprising since this value is highly dependant on subject characteristics. Improper RMR predictions caused the percent difference between the TEE measured with the room calorimeter and that predicted with the ANN model to be quite large in some cases. This experiment showed that an acceleration only ANN model is able to achieve higher EE prediction accuracy than an acceleration model that was developed using integrated data and incorporating a measured baseline and body weight and achieve comparable predictive accuracy to a five sensor array which incorporates sex, age, height, and weight. This model, however, did not achieve the minute-by-minute accuracy observed when subject characteristics were combined with acceleration parameters.

Minute-by-Minute ANN Results

Minute-by-minute EE predictions (Figure D11-D20) are presented for 10 subjects (5 male; 5 female). Four subjects were considered lean (BMI ≤ 25 kg/m^2), two subjects were borderline between healthy and overweight (25 < BMI ≤ 30 kg/m^2), and four are characterized as obese (BMI > 30 kg/m^2). Results are shown from the ActiGraph (AG1), the proprietary IDEEA algorithm, and the hip only, acceleration plus characteristics ANN model. Both the full time series (~ 1300 minutes) and the first 600 minutes, where most of the moderate to vigorous intensity PA occurs, are shown.

In viewing the full time series, it should be noted that AG1 was supplied a subject specific sleeping rest value (mean of the measured EE during sleep), which explains the good agreement between measurement and prediction during this time period. The IDEEA model and the ANN both have information (sex, age, height, weight, and BMI), which has been found to be correlated with resting EE, but the models were not directly supplied a value. As such, there is more variability in the error associated with the baseline in these two models.

The major improvements in both the IDEEA and the ANN model, relative to AG 1, is increased detection and characterization of spontaneous low intensity PA, which is frequently assigned an activity count of zero using the ActiGraph. For this data set, the ANN more consistently fits the moderate to vigorous intensity peaks in the walking and jogging data, though some prediction errors are still detected. This is true for individuals of a wide range of BMI.
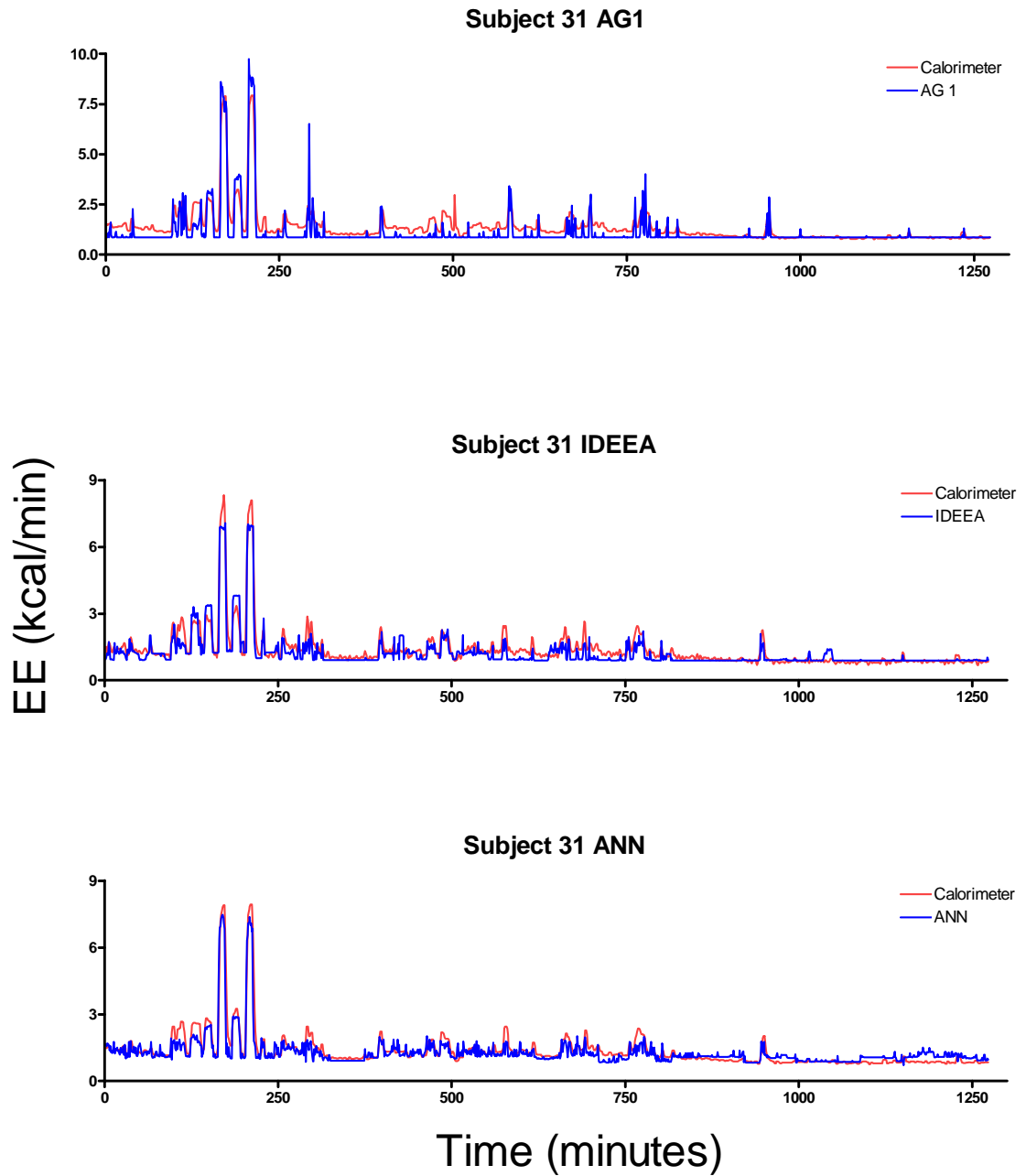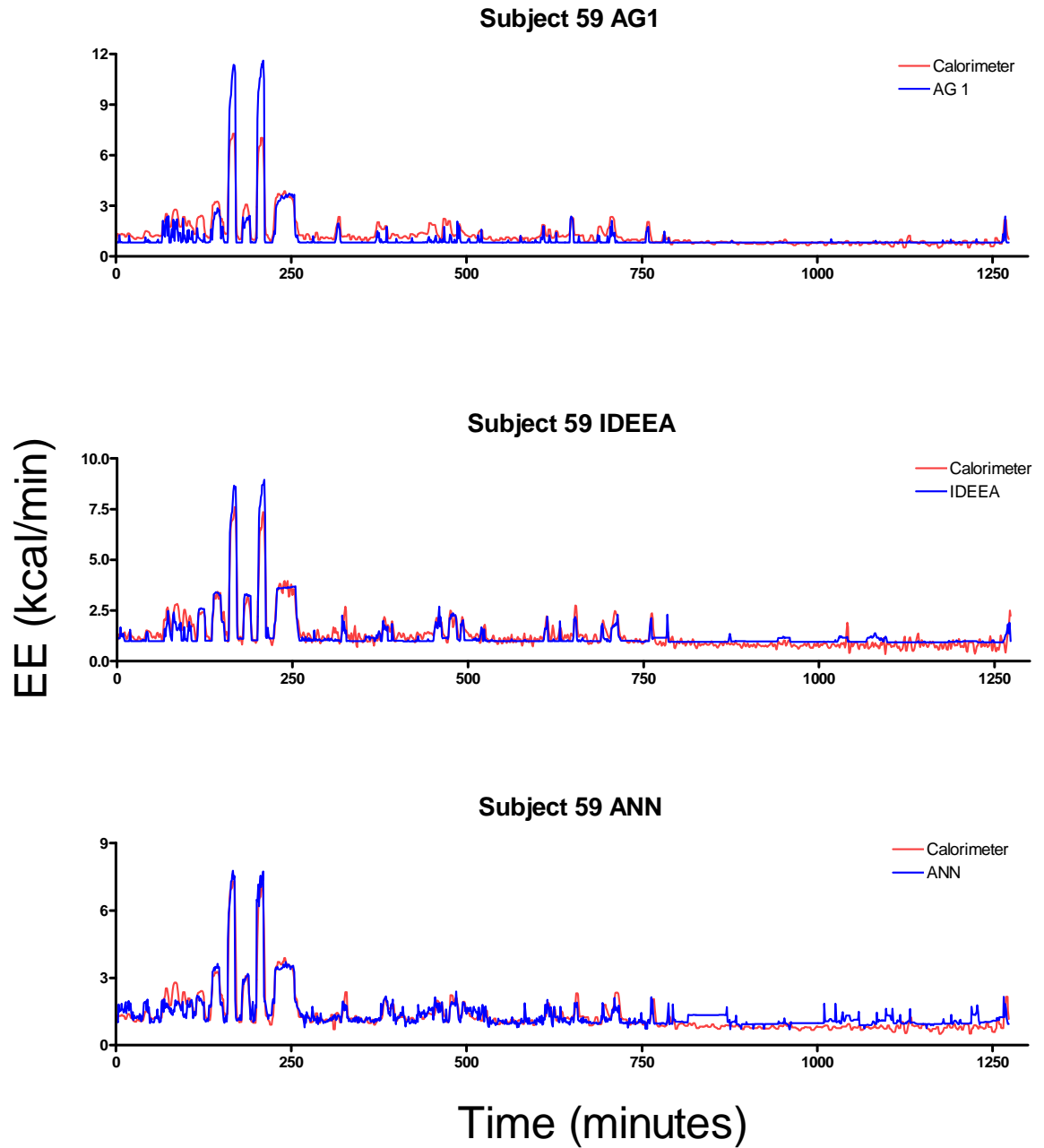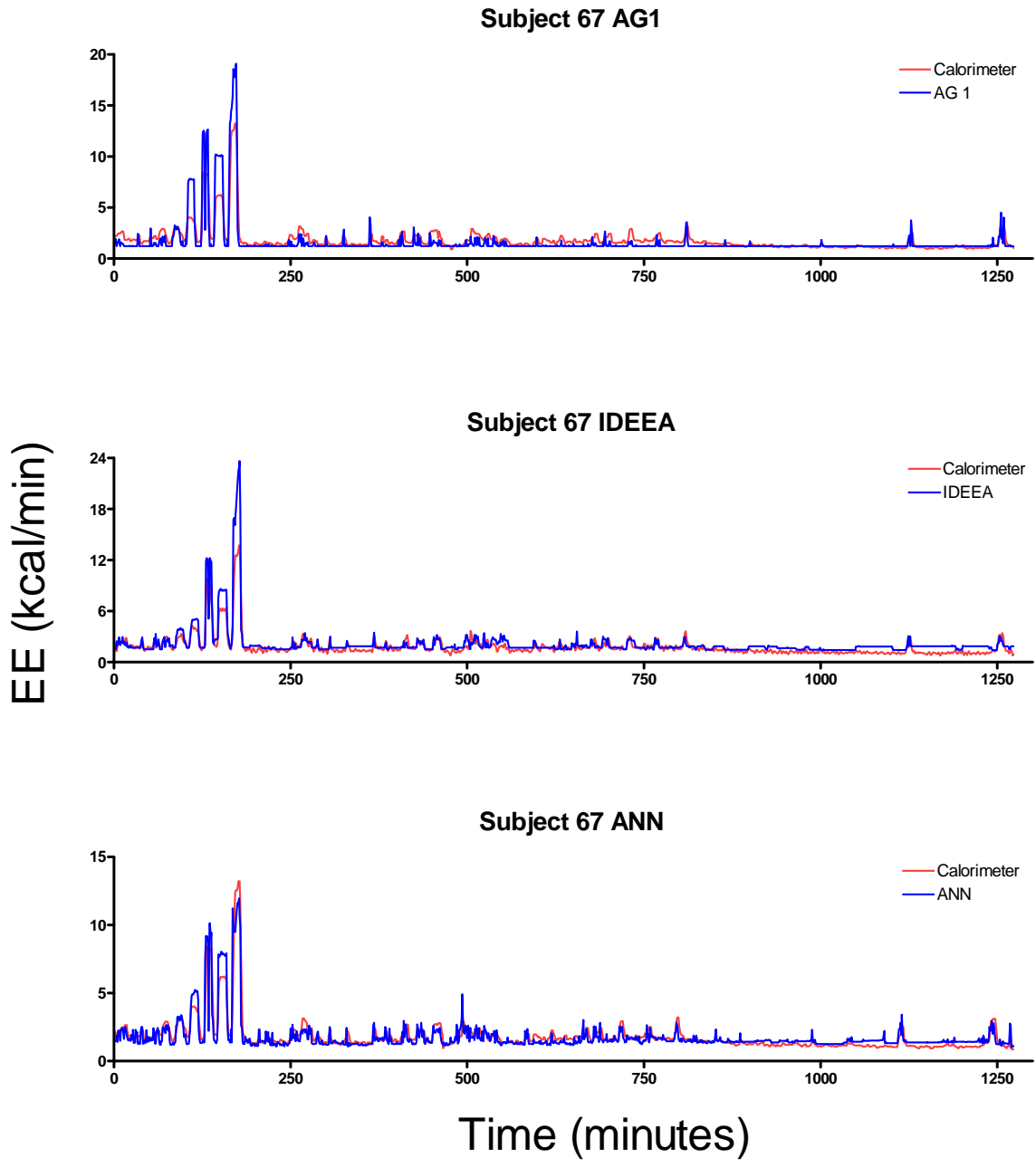
Figure D11: Minute-by-minute EE prediction using ActiGraph (AG 1), IDEEA, and ANN. The top row of graphs shows the entire study visit, while the bottom shows the first 600 minutes of collected data. This subject was a 30 year old female (height – 1.62m; weight – 48.2 kg; BMI – 18.4 kg/m^2). Using BMI, this individual is characterized as borderline underweight/healthy.
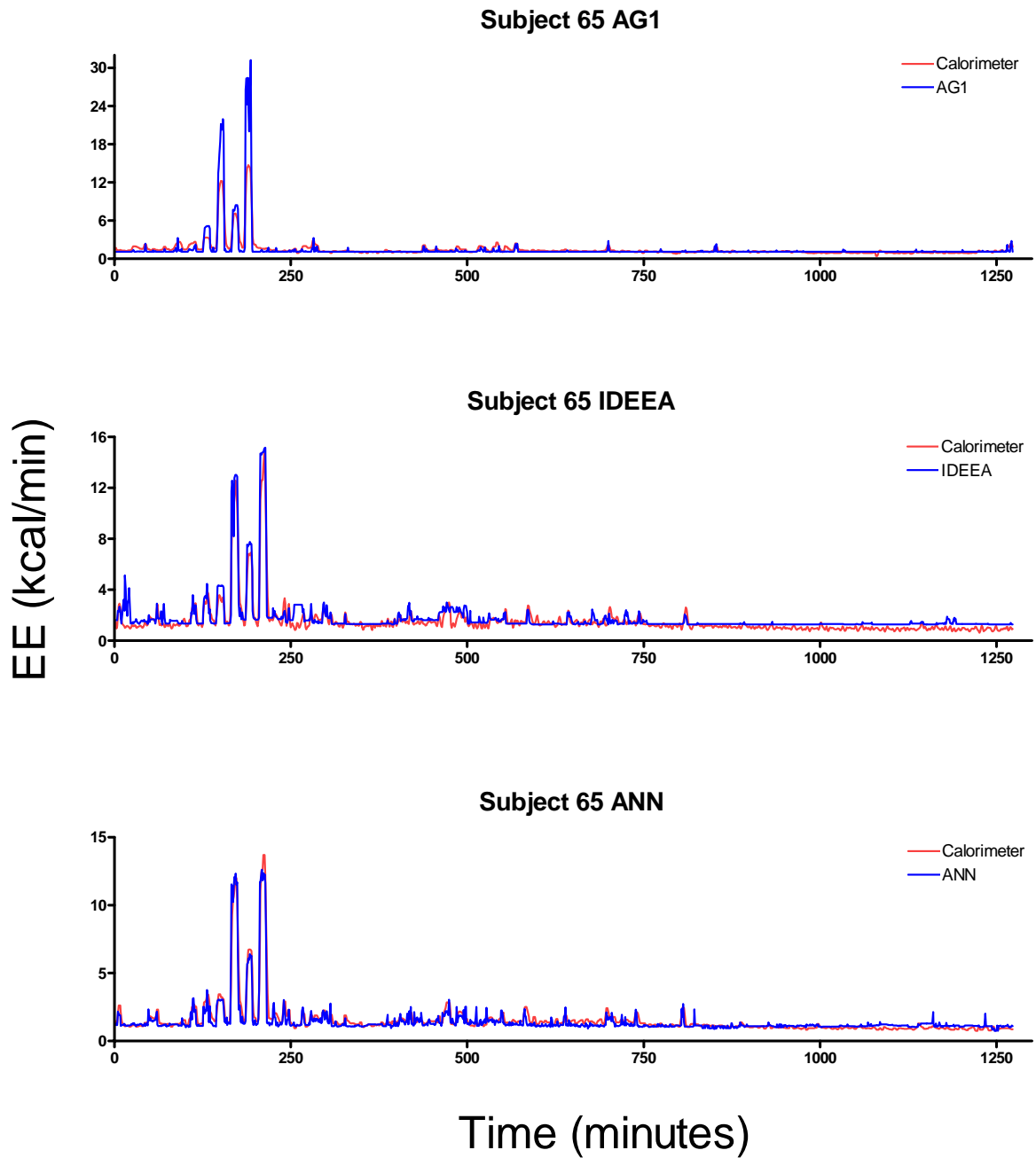
Figure D12: Minute-by-minute EE prediction using ActiGraph (AG 1), IDEEA, and ANN. The top row of graphs shows the entire study visit, while the bottom shows the first 600 minutes of collected data. This subject was a 49 year old female (height – 1.62m; weight – 55.6 kg; BMI – 21.2 kg/m^2). Using BMI, this individual is characterized as healthy.
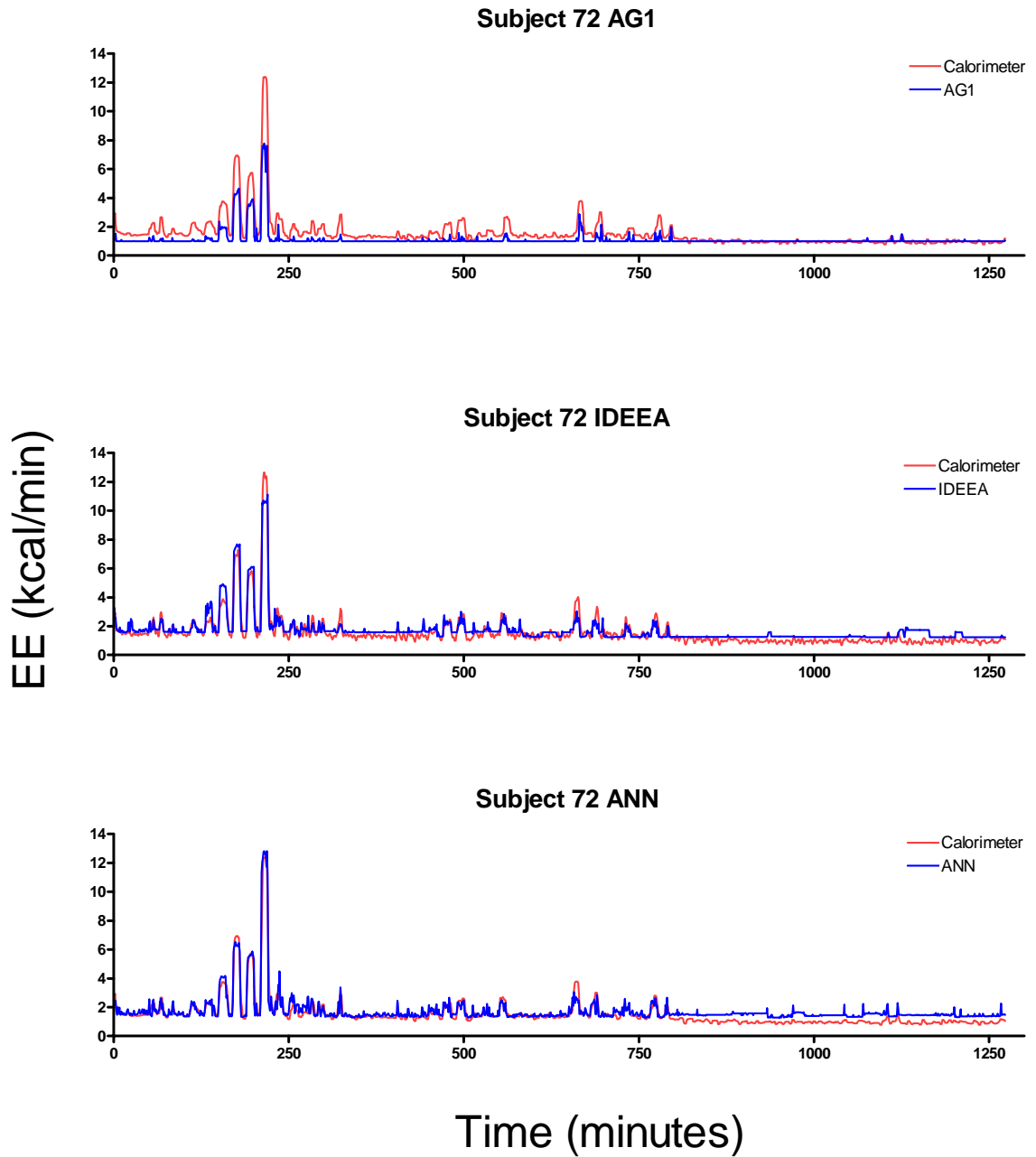
Figure D13: Minute-by-minute EE prediction using ActiGraph (AG 1), IDEEA, and ANN. The top row of graphs shows the entire study visit, while the bottom shows the first 600 minutes of collected data. This subject was a 47 year old male (height – 1.85m; weight – 73.5 kg; BMI – 21.5 kg/m^2). Using BMI, this individual is characterized as healthy.

Figure D14: Minute-by-minute EE prediction using ActiGraph (AG 1), IDEEA, and ANN. The top row of graphs shows the entire study visit, while the bottom shows the first 600 minutes of collected data. This subject was a 23 year old male (height – 1.72m; weight – 67.0 kg; BMI – 22.6 kg/m^2). Using BMI, this individual is characterized as healthy.
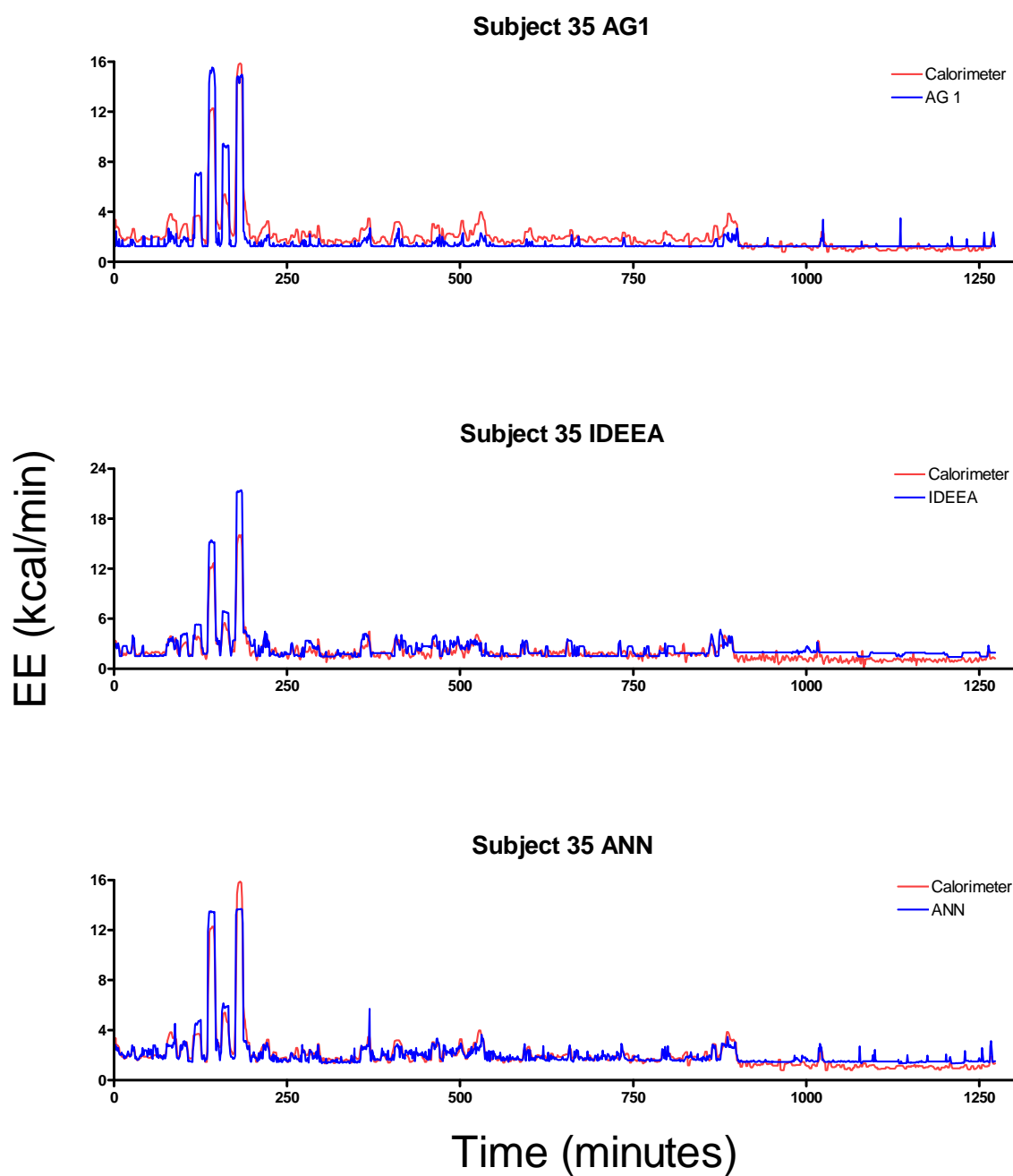
Figure D15: Minute-by-minute EE prediction using ActiGraph (AG 1), IDEEA, and ANN. The top row of graphs shows the entire study visit, while the bottom shows the first 600 minutes of collected data. This subject was a 27 year old female (height – 1.78m; weight – 76.8 kg; BMI – 24.3 kg/m^2). Using BMI, this individual is characterized as healthy to borderline overweight.
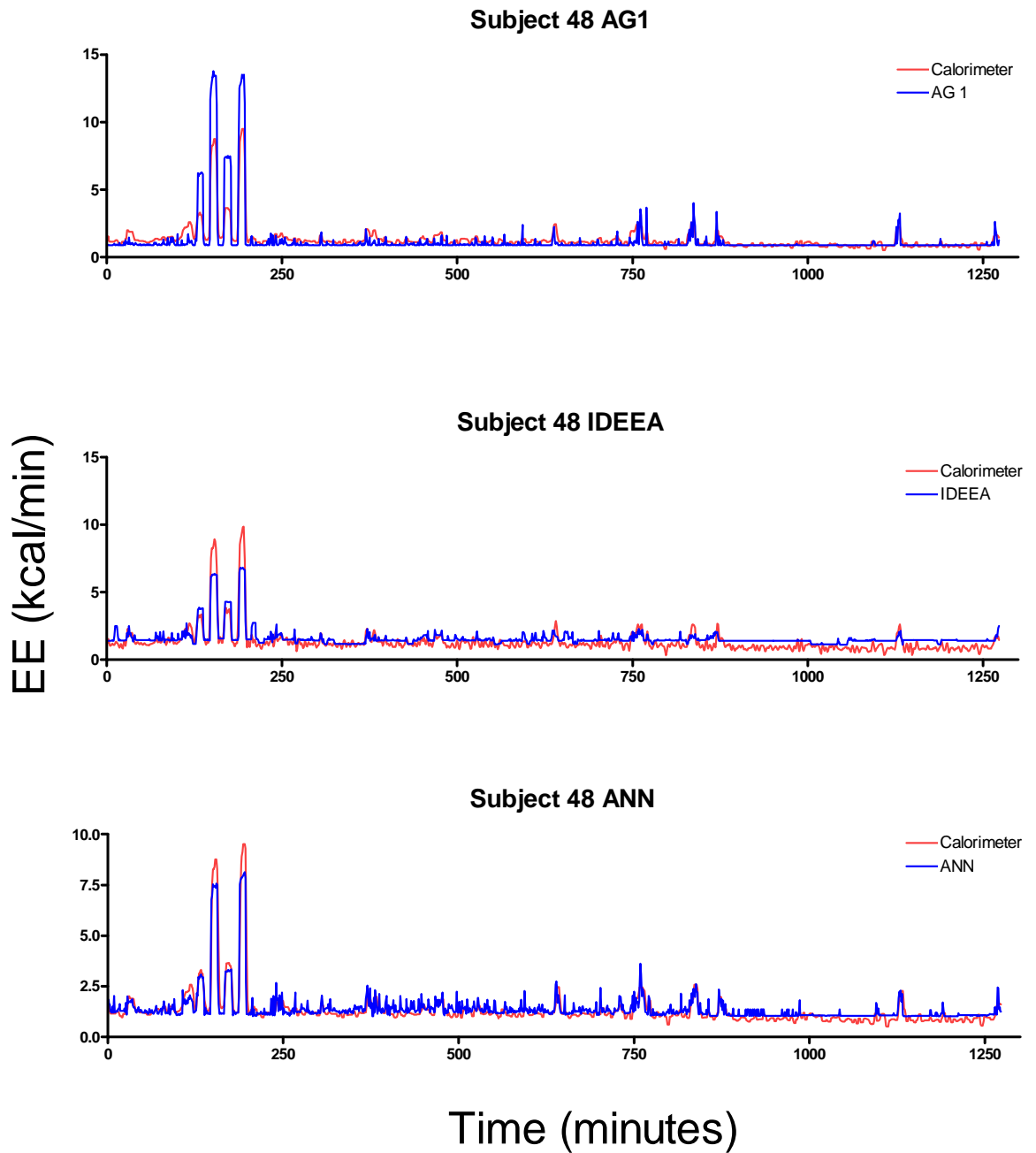
Figure D16: Minute-by-minute EE prediction using ActiGraph (AG 1), IDEEA, and ANN. The top row of graphs shows the entire study visit, while the bottom shows the first 600 minutes of collected data. This subject was a 29 year old male (height – 1.84 m; weight – 83.5 kg; BMI – 25.3 kg/m^2). Using BMI, this individual is characterized as healthy to borderline overweight.

Figure D17: Minute-by-minute EE prediction using ActiGraph (AG 1), IDEEA, and ANN. The top row of graphs shows the entire study visit, while the bottom shows the first 600 minutes of collected data. This subject was a 24 year old female (height – 1.53m; weight – 76.0 kg; BMI – 32.46 kg/m^2). Using BMI, this individual is characterized as obese.
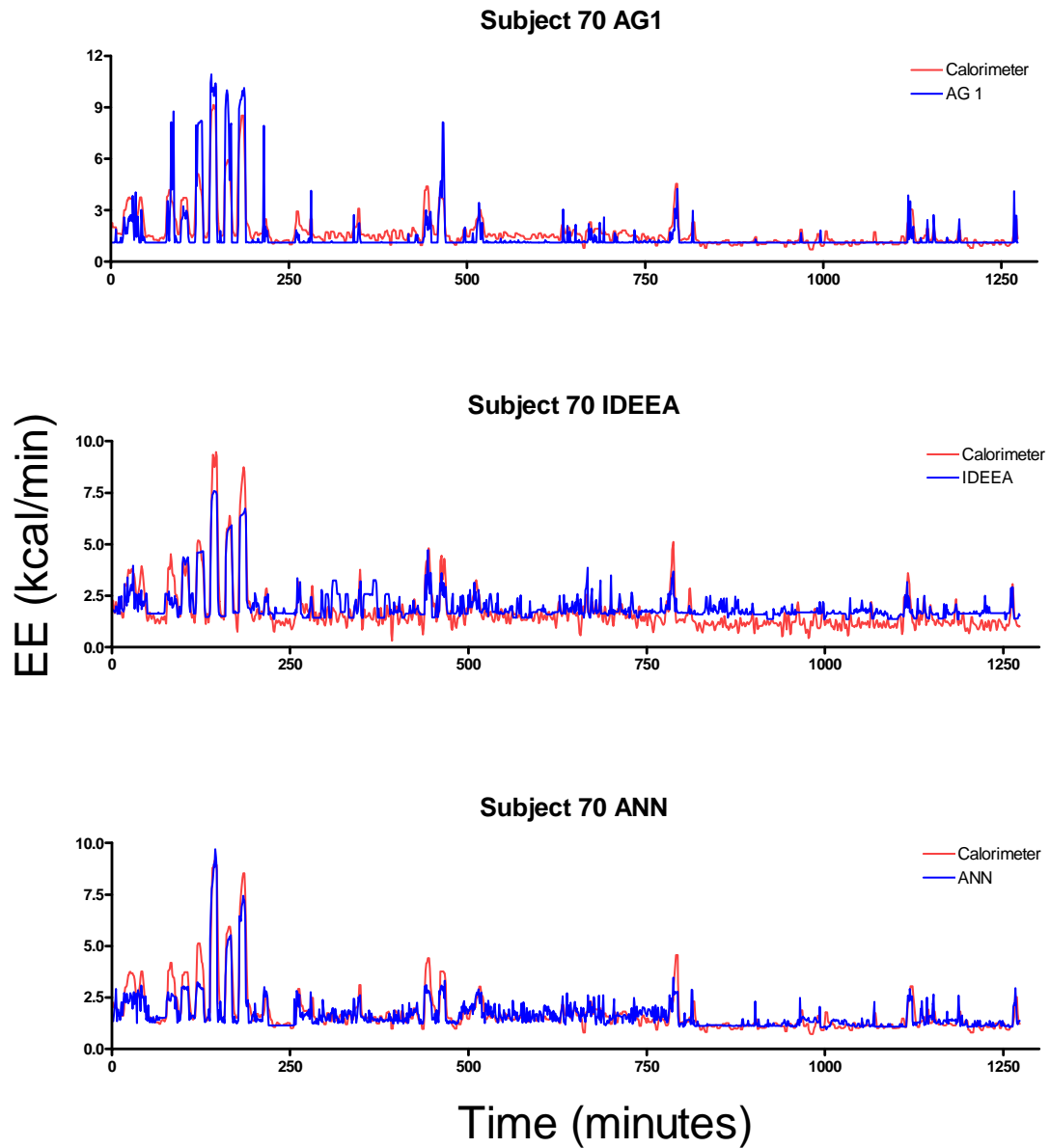
Figure D18: Minute-by-minute EE prediction using ActiGraph (AG 1), IDEEA, and ANN. The top row of graphs shows the entire study visit, while the bottom shows the first 600 minutes of collected data. This subject was a 48 year old female (height – 1.57m; weight – 96.5 kg; BMI – 38.9 kg/m^2). Using BMI, this individual is characterized as obese.
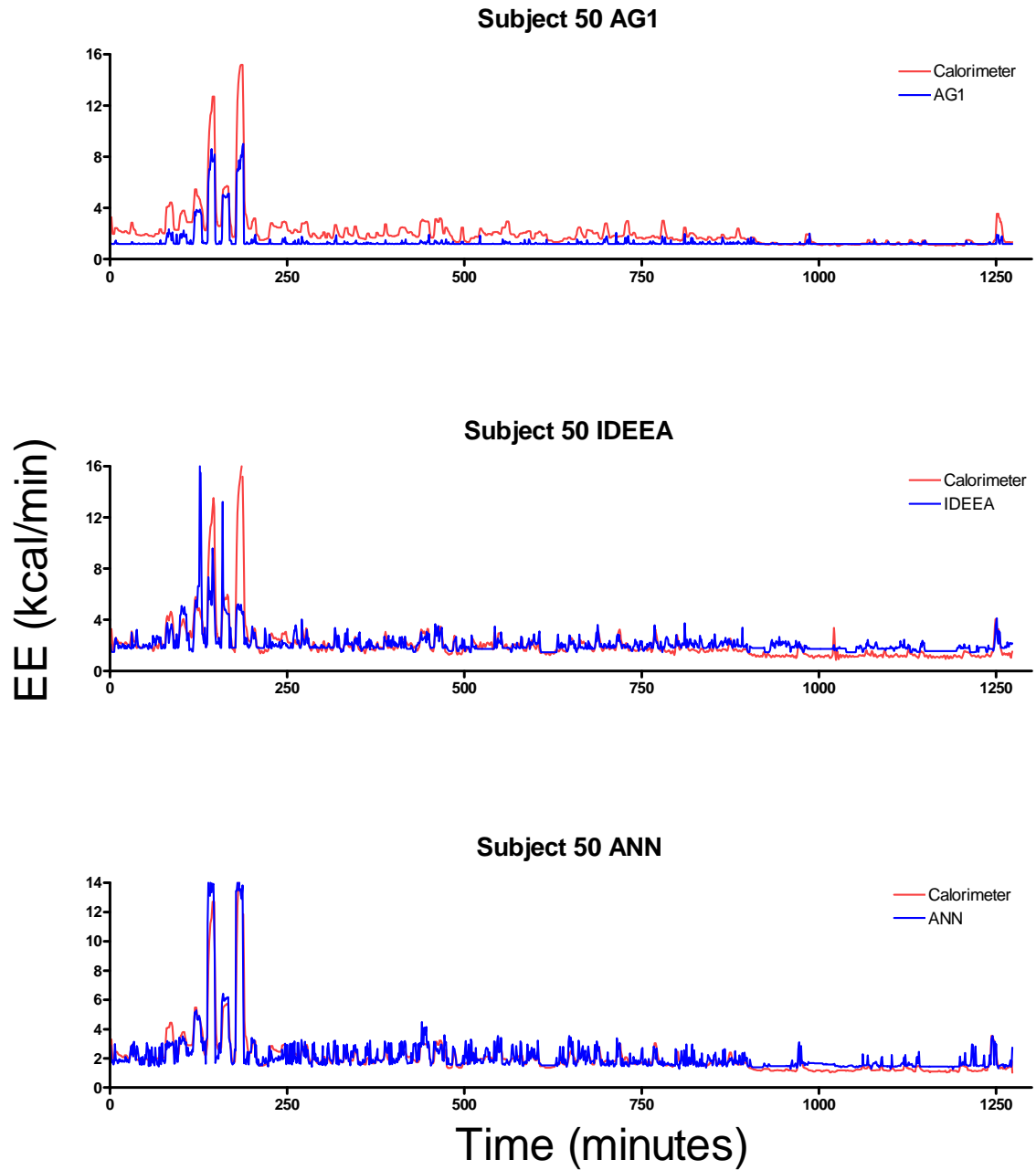
Figure D19: Minute-by-minute EE prediction using ActiGraph (AG 1), IDEEA, and ANN. The top row of graphs shows the entire study visit, while the bottom shows the first 600 minutes of collected data. This subject was a 57 year old male (height – 1.79m; weight – 98.2 kg; BMI – 30.6 kg/m^2). Using BMI, this individual is characterized as borderline overweight/obese.

Subject 54 AG1

Subject 54 IDEEA

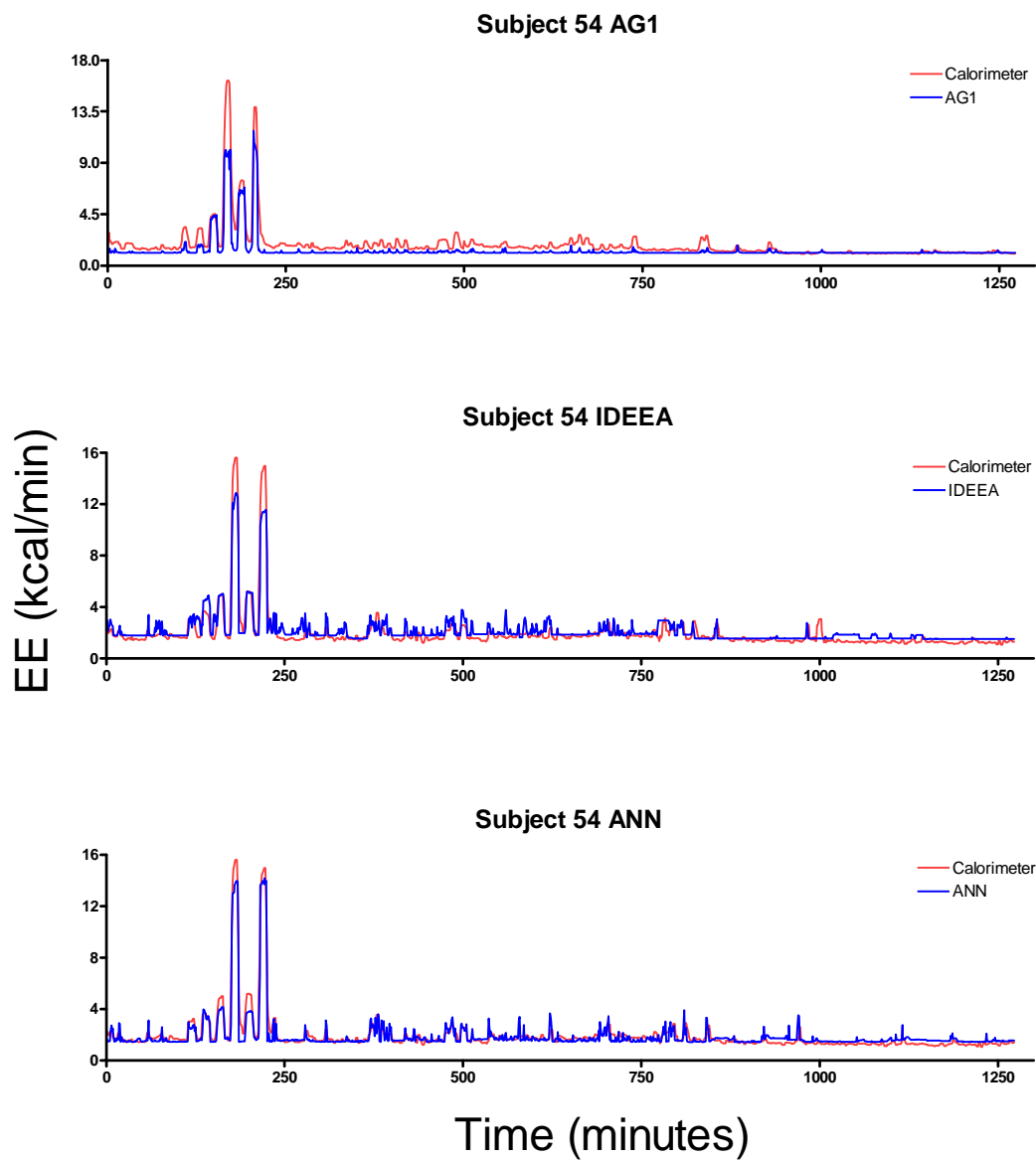**Subject 54 ANN**

EE (kcal/min)

Time (minutes)

Figure D20: Minute-by-minute EE prediction using ActiGraph (AG 1), IDEEA, and ANN. The top row of graphs shows the entire study visit, while the bottom shows the first 600 minutes of collected data. This subject was a 40 year old male (height – 1.75m; weight – 101.5 kg; BMI – 33.1 kg/m^2). Using BMI, this individual is characterized as obese.

Summary

The work in this appendix addresses technical issues associated with the ANN model development which were not presented in Chapter V. These include description of the data features extracted from the raw acceleration data. The input values in the order they were presented to the model were presented along with Hinton plots showing the sign and magnitude of the network weighting coefficients and a brief analysis of the sensitivity of the model solution when random noise was applied to the optimized weights. Expanded results which show that when minute-by-minute EE was stratified into four categories and the means of the measured intensities within each category were compared to the model predicted means, each model showed a specific pattern of over and under-prediction of EE. As in our other experiments, the ANN showed a lower standard deviation term outside the sedentary intensity range. An acceleration only ANN model was developed, and it was determined that while the variance in EE predictions was similar to the ActiGraph, the MAE and MSE were significantly reduced. This model was comparable to the IDEEA model, which had more acceleration data and subject characteristics. Finally, minute-by-minute EE predictions for ten subjects were shown for the ActiGraph model, IDEEA, and ANN. These results demonstrate graphically the improvements in detection of EE intensity using the ANN. While these results are promising, they do indicate that more physical activity intensities, subject characteristics, or alterations to the network architecture may be required to optimally predict minute-by-minute EE.

APPENDIX E

COMPARING INDIVIDUAL AND GROUP ANN MODELS

In order to determine the capability of the ANN approach to develop robust individual models for EE prediction, a subject-specific hip-only model was developed for each of the subjects used in the development of the general ANN model, using a similar feed-forward back-propagation approach as described in Chapter V. Subject characteristics were no longer used as model inputs since each subject is unique and the impact these features make on metabolic response should be predictable through model training without direct presentation of this information. Due to the reduction in the number of input features from 17 to 10, the architecture of the model was reduced to a 5-10-1 neuron structure (the original model used a 12-20-1). The learning rate, functional form of basis functions, and convergence criteria were maintained from the generalized model.

Two aspects of training were altered for the individual model. First, the training data was not biased towards high intensity examples. Each data point from the training set was presented once. This was due to the fact that within a subject there is less diversity in the metabolic response given a particular acceleration response than there is over the population. The second change was the validation procedure. For this modeling approach we performed a leave 200 minutes out approach, repeated 20 times for each subject. The mean EE prediction error over the 20 holdout iterations are reported. Two hundred data points represent approximately 15% of the total study visit and should

represent a diverse sample of activity types and intensities. However, the 200 data points

included in each hold out set were chosen randomly and we did not study the intensity

composition of each hold out set, which may be an important factor in the outcome of the

model.

Using a Wilcoxon rank sum test, we compared the mean results from the 20

holdout sets derived from each subject's individual model to the validation error for their

respective holdout set from the generalized model. Significant reduction

($p < 0.001$) was observed in the MAE, MSE, and percent difference in measured TEE

(Figure E1). A significant increase in the correlation coefficient (r^2) was also observed.

The results suggest that there may be some benefit to individualized models with respect
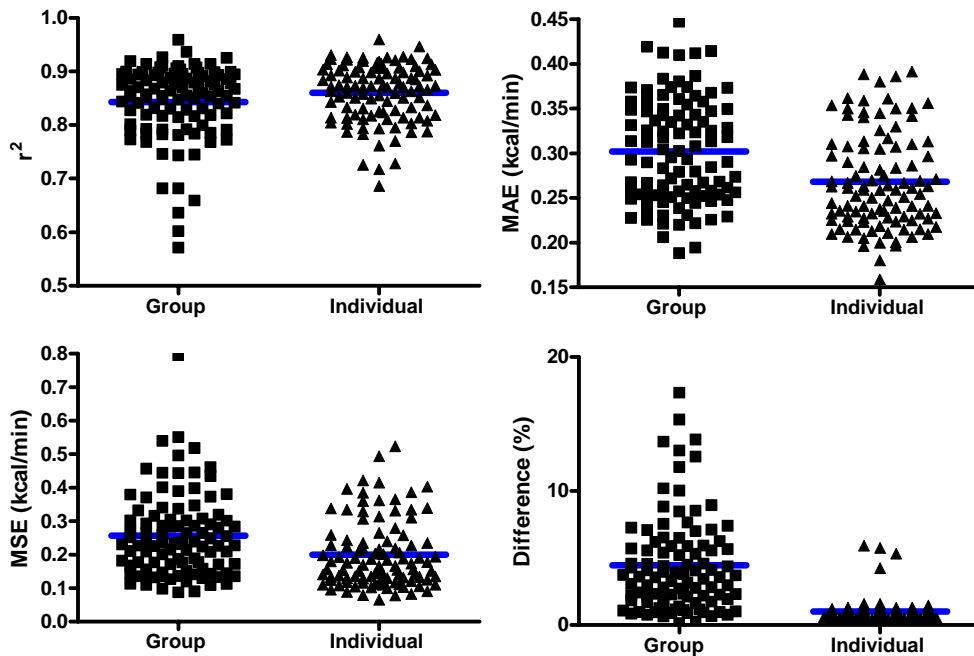
to validation errors.

Figure E1: Comparison of the performance of Group ANN models with individual models.

To represent trends in under and over prediction of total EE for the 200 minute validation sets used in the individual model development, Bland-Altman plots were made for eight randomly selected subjects (Figure E2). Each plot reflects the results of the 20 network optimizations performed for the subject.

The mean differences in total and predicted EE for the 200 minute validation set ranged from -2.5 kcal to 0.92 kcal. No non-zero slope trends ($p > 0.05$) were detected for the Bland-Altman plots for these subjects. The differences in EE between the measurement and the prediction derived from the individual models are smaller than those in the group model. These improvements could be due to improved model fit or could result from the validation set containing fewer data points with less measurement diversity.
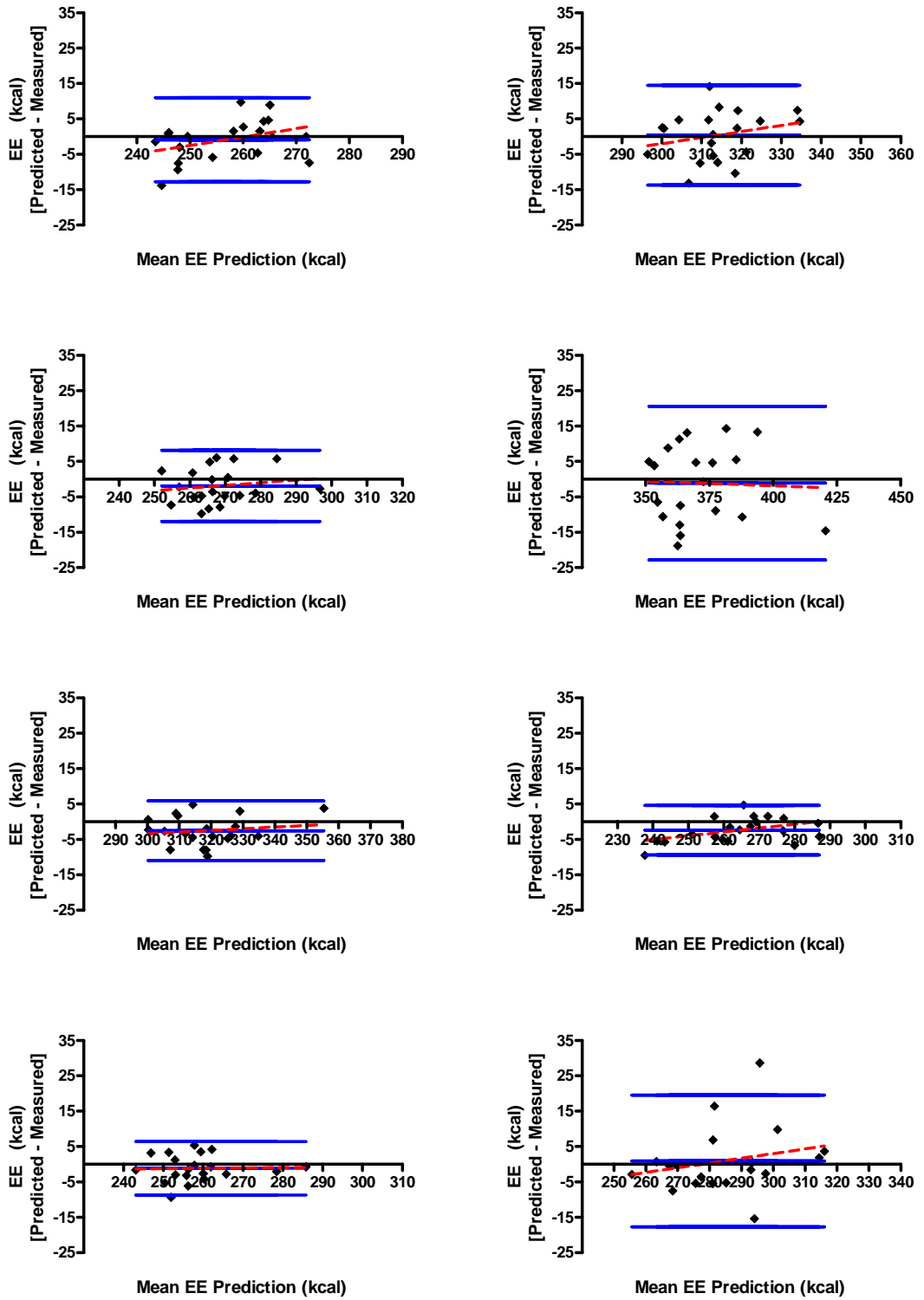
Figure E2: Individual Bland-Altman plots for eight randomly chosen subjects.
Each point represents an optimization of the network.

Performing individualized modeling, while certainly less time consuming and potentially more accurate than generalized modeling, suffers from several important limitations. First of all, subject specific training data is required for the model development. This means that in order for the model to be developed, each subject must be studied for some period of time in a laboratory setting, where simultaneous acquisition of accelerometer and calorimeter data is performed. Ideally, this study period would be long enough for each subject to perform many of the activity types that are common in their daily life. This is more of a burden on the subject than simply wearing the activity monitor and having the investigator apply a general model to the data. This is an especially important consideration for large studies, where time and resources can be saved if subjects are not required to come to the research facility.

Another limitation is that the individual model for a subject adapts to the EE of each subject at the time of model development. This means that changes in their personal characteristics or physical fitness could cause an individual model to be invalid at future time points. This would be especially true in interventional studies where the outcome goal involves making changes in either the subject's body composition or exercise habits. In the case of a general model, these new characteristics would simply be inputted into the model and given enough training data was used for the generalized model development, a prediction that is more accurate for the current attributes of the subject would be realized without retraining.

This work serves as a conceptual demonstration that individual ANN models can be developed using hip-only acceleration data. These models show a reduction in error when compared to generalized models developed using the same acceleration signals.

The solutions were also derived on a much smaller network architecture, which diminished the computational time for model training. The networks also showed nearly perfect match between in the total EE predicted for the 200 minute validation set. Though the error was reduced in the individual models relative to the generalized model presented in Chapter V, it is important to remember that individualized modeling requires each subject to come into the lab so that the investigator can obtain valid training data, and the model is only valid as long as the subjects metabolic response is not altered by changes in their personal attributes or lifestyle.