A Novel Analysis of Accelerometry Data: Calibration and Flexible Association Modeling in an Epidemiological Study of Older Women

Sixuan Wu

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington

2020

Committee:

Chongzhi Di

Ying Qing Chen

Program Authorized to Offer Degree:
Biostatistics - Public Health

University of Washington

Abstract

A Novel Analysis of Accelerometry Data: Calibration and Flexible Association Modeling in an Epidemiological Study of Older Women

Sixuan Wu

Chair of the Supervisory Committee:
Chongzhi Di
Department of Biostatistics

Accelerometers have been widely deployed to objectively measure and monitor physical activity and sedentary behavior in large epidemiological studies. The traditional summary metric, known as counts, summarizes raw high-resolution acceleration signals for a pre-specified epoch length (e.g., 1 minute). However, its definition is proprietary to device manufacturers, making it difficult to compare studies that use different devices. Alternative summary metrics have been introduced in recent years. This master thesis conducted a novel analysis of accelerometry data based on the activity index, a novel and transparent way to summarize the high-dimensional accelerometry data, among older women within the Women's Health Initiative. We first built calibrating equations of activity index for estimating metabolic equivalents (METs) and derive cutpoints to classify epochs into distinct physical activity intensity categories. We then utilized single variable and isotemporal substitution models to investigate associations of one or more physical activity intensity category and health outcomes, such as cardiometabolic risk factors in the Objective Physical Activity and Cardiovascular Health (OPACH) Study. Further, we adopted a newly developed functional data analysis framework to quantify the dose-response relationships between continuous physical activity intensity and cardiometabolic risk factors in the OPACH Study.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, we give a brief overview of physical activity (PA) epidemiology, including its scientific relevance, measurement challenges, and rapid developments within the past decade with increasingly popular adoption of objective measurements through wearable devices. In particular, we highlight the usage of accelerometers, typical output data format including epoch-level summarized data and high-resolution raw acceleration data, and some novel metrics or processing methods based on raw acceleration. We then introduce the Women Health Initiative and its ancillary study called the Objective Physical Activity and Cardiovascular Health, which is the primary motivating study for this thesis. Finally, we describe an Activity Index, a novel metric of accelerometry data, which is the primary building block for analyses in this thesis.

1.1 Physical Activity Epidemiology

It is well known that being physically active improves overall health and prevents many adverse health outcomes among people of all ages. The health benefits of PA include reduced risk of mortality, many chronic diseases and mental disorders such as depression and cognitive function (Piercy et al., 2018). Based on the *Physical Activity Guidelines for Americans. 2nd ed* (US Dept of Health and Human Services, 2018), in order to have substantial health benefits, adults should do at least 150 minutes (2 hours and 30 minutes) per week of moderate-intensity PA, or 75 minutes (1 hour and 15 minutes) per week of

vigorous-intensity aerobic PA. Although it has been widely recognized that PA is beneficial for health, there lacks rigorous scientific evidence that quantifies the precise dose-response association of PA of various intensity, frequency and duration and health outcomes.

PA is a complex and multidimensional behavior that has many aspects, e.g., intensity, frequency, duration, types, context, as well as daily, weekly and seasonal variations. Intensity, frequency and duration are often viewed as the most important aspects in relation to health. PA intensity is often measured in terms of energy expenditure, in terms of calories or metabolic equivalents (METs). For example, when a person is sitting still with little movement, the corresponding activity intensity is 1 MET. An intenity of 2 METs for a specific activity means that a person's energy expenditure during this activity is twice as high as that of sitting still. METs are often used to divide PA into a few intensity categories. Specifically, activities with 0.9-1.5 METs, 1.6-2.9 METs, 3.0-5.9 METs and 6.0+ METs are categorized as sedentary behavior, light intensity, moderate intensity and vigorous intensity PA, respectively. In addition, moderate and vigorous intensity physical activities (MVPA) are often combined as a single category. With these categories defined, cumulative time spent in each intensity and its associated MET-hours are common summary measures of PA.

Despite the prominent impact of PA on the health outcome, it's challenging for researchers to employ a valid and reliable method to measure PA intensity. Common method involves questionnaires (Voorrips et al., 1991), diaries/logs, direct observations and wearable devices (BOUTEN et al., 1994; Troiano et al., 2008; Ward et al., 2005). Among them, self-report questionnaires (PAQ) are traditionally used before objective measurements become available. Generally speaking, it is designed to include questions about types, frequency and duration of a variety of PA types. Based on the 2011 Compendium of Physical Activities (Shephard, 2012), each activity recorded in the questionnaire will be assigned an intensity level in METs and then the total time spent in each intensity along with its total METs-hour will be further calculated. Although such subjective methods benefit from cost effectiveness and wide adoption among many studies in various population, it suffers from systematic under- or over-estimation PA due to recall, reporting bias and other measurement error (Dishman et al., 2001). In the past decade, wearable devices such as accelerometers became more widely adopted with their ability to measure PA objectively with the higher accuracy (Janz, 2006).

1.2 Accelerometry

An increasing number of studies have utilized objective measurements to record PA patterns, with accelerometers being the most popular wearable devices adopted in large epidemiological studies.

1.2.1 Accelerometers

Accelerometers are devices that measure body movements in terms of acceleration, and can be used to estimate the intensity of PA over time. The theoretical basis for accelerometry-based PA monitors is that the acceleration is directly proportional to muscular forces and thus it is reflective of the energy expenditure. In addition, signal data contains rich information which can be further processed into other signals such as speed and distance using the integration with respect to time (Chen and Bassett, 2005).

In practice, a piezoelectric sensor, consisted by a piezoelectric element and a seismic mass, is used in the accelerometer to measure the acceleration in one direction. Specifically, when there is a acceleration happening in the measured direction, the seismic mass will cause the piezoelectric element to bend, tighten or compressed. And a variable output voltage signal will then be generated corresponds to this deformation and hence the acceleration data is measured (Chen and Bassett, 2005). To measure acceleration in multiple directions, then multiple piezoelectric sensors are needed. For example, tri-axial accelerometer has three piezoelectric sensors to detect the accelerations in three orthogonal planes (antero-posterior, mediolateral, and vertical) (Chen and Bassett, 2005) and record the acceleration data into voltage signals.

One major limitation of piezoelectric sensor is that it won't be able to detect the static body postures such as sitting or standing and it is reliable only when detecting the dynamic activities.

1.2.2 Accelerometry Data Format

Current accelerometers collect high-resolution voltage signals as output data directly, but the most commonly used output data format is a summary measures of the voltage signals over a given time window (e.g., 1 minute).

Specifically, acceleration data is first sampled with the frequency rate determined by

the monitor computer. In order to ensure that the activity motion can be fully captured, the frequency rate need to be at least twice the frequency of the highest frequency of the movement, otherwise the motion with extreme frequency will be distorted. After having the sampled data, a band pass filter will be applied to attenuate the frequencies that are not between the specified low- and high-frequency limit. It is important to select a appropriate frequency range for the band pass filter (Chen and Bassett, 2005). Specifically, a filter with too wide bandwidth will end up including too much noise that is unrelated to the true movement. On the contrary, a great amount of data that represents the movement will be excluded if a too narrow bandwidth is used.

With the sampling and filtering, each sensor in the accelerometers will output a bidirectional voltage signal to the micro-processor in the device. Through proprietary algorithms developed by accelerometer manufacturers (e.g., Actigraph), the voltage signals will be further converted into a metric for each given time window called an epoch to summarize the multi-dimensional accelerometry data. Such metric, a 1D proxy of subjects' activity along time, is called activity count (AC).

However, there are several limitations when using AC. The algorithms that used to calculate AC differ from manufacturer to manufacturer and even differ with devices from the same manufacturer. In addition, there is not a clear formula available to help researchers understand which elements are used to compute the AC and what is the difference between these algorithms. So it makes it hard to compare AC device- or manufacturer-wise and the interpretation of the AC is not straightforward.

Due to the limitation of AC, efforts have been made to establish summary metrics from the raw data with explicit formulas that are available to public in order to allow comparison of summary metrics from different accelerometers, and clarify the interpretation of the results. Two notable summary metrics are Activity Intensity (AI) by Bai et al., 2016 and Euclidean Norm Minus One (ENMO) by Hees et al., 2013. Generally speaking, both of the metric aim to quantify the magnitude of acceleration during a given epoch, while the ENMO focuses on measuring the vector magnitude of raw accelerometry signal after removing one Earth standard gravitational unit and the AI focuses on measuring the amplitude of the raw accelerometry signal relative to its amplitude distribution at rest. More details of the AI will be discussed in section 1.4.

1.3 The Objective Physical Activity and Cardiovascular Health (OPACH) Study

This thesis is motivated by an ancillary study within the Women's Health Initiative (WHI), called the OPACH Study.

WHI was initiated by the U.S. National Institutes of Health (NIH) in 1991 to address major health issues in postmenopausal women aged between 50 to 79. It consisted of three clinical trials with 68,132 post-menopausal women and an observational study with 93,676 post-menopausal women from total 40 clinical sites in the US from 1993 to 1998 (Anderson et al., 2003). WHI main study ended in mid-2005 and was approved to extended the study through three Extensions Studies. These extensions are referred to as "Extension Study 1" (2005-2010), "Extension Study 2" (2010-2015), and the recently undertaken "Extension Study 3" (2015-2020).

During the second WHI Extension Study, the WHI Long Life Study was conducted with a subcohort of 7,875 WHI participants to studying factors associated with cardiovascular risk and aging. And the OPACH study was conducted within the second WHI Extension Study as a ancillary study between 2012-2014 with 7048 women aged 63–99. All the participants of the OPACH study was provided with an ActiGraph GT3X + (Pensacola, Florida) triaxial accelerometer, a sleep log, and an OPACH PA Questionnaire to collect their 7-day PA measurements.

Specifically, for collecting the accelerometer data, women were asked to wear a hip-worn accelerometer (Actigraph GT3X+) for 7 consecutive days during both waking and sleeping hours, except when bathing or swimming. And the sleep log was used to isolate the accelerometer data during sleeping time. The collected data obtained from 6,489 women who returned the device were processed into AC for every 15-second epoch. Other data including the baseline characteristics such as age, race/ethnicity, education and BMI was collected since their WHI enrollment. And biomarker data such as glucose, insulin and cholesterol was also collected through measuring the fasting blood draw of total 5,100 participants in the OPACH study (Lacroix et al., 2017).

Besides the main study, the OPACH study also conducted a substudy, a laboratory-based OPACH calibration study (Evenson et al., 2015), to determine intensity-specific accelerometer count cutpoints appropriate for women in older age. Total of 200 women aged

60 to 91 from WHI study centers were recruited in the calibration study and were asked to visit the study clinic site.

To collect the raw accelerometry data and AC, participants were provided with the same model of accelerometry that is used in the OPACH main study and placed at the same body location. The standard of PA intensity, energy expenditure in METs, were calculated through oxygen uptake (VO2) and heart rate measured continuously during the physical activity tasks using a portable calorimeter called Oxycon (Oxycon Mobile; CareFusion, Rolle, Switzerland). With the devices, participants were asked to performed activity tasks in the following order: watching DVD while sitting quietly (alias: DVD), assembling puzzle while sitting (alias: PUZZ), washing dishes while standing (alias: DISH), doing laundry while standing (alias: LAUD), 400-meter walking (alias: WALK), dust mopping while standing (alias: MOP), treadmill walking at 1.5mph (alias: TM15), and treadmill walking at higher speed, either 2.0mph (alias: TM20) or 2.5mph (alias: TM25). For each of the task, the duration was chosen in order to achieve steady rate metabolism for measurement of oxygen uptake.

Two selection criteria, maximizing the sum of sensitivity plus specificity and balancing the number of false positives and false negatives, were used to derive the cutpoints of sedantary, light low, light high and MVPA. For each of the criteria, 6 sets of cutpoins were reported corresponding to the different accelerometry filter frequencies and intensity thresholds definitions. The OPACH main study adopted the cutpoints derived from balancing the number of false positives and false negatives with normal filter where 1 MET = 3.0 mL/kg/min for their further analysis. Some other previous analysis (Bradley et al., 2019; Heath, 2019; Nicklas, 2018) also adopted the calibrated OPACH cutpoints to categorize PA into sedantery, light, moderate, and vigorous intensity in order to study the association between PA with mortality and health outcomes. By using the calibrated OPACH cutpoints, an inverse associations for light intensity PA with mortality and cardiovascular disease incidence have been found.

In addition to the calibration study which derived cutpoints to classify the PA based on the AC, machine learning based methods such as random forest (Rosenberg et al., 2017) and neural network (Lyden, KATE and Keadle, SARAH KOZEY and Staudenmayer, JOHN and Freedson, Patty S, 2014) were applied on the raw high-resolution accelerometer data directly to detect specific types of physical activities and sedentary behaviors.

1.4 An Activity Index (AI) Based on High-resolution Raw Acceleration Data

The activity index (AI), a new metric for summarizing raw tri-axial accelerometry data, was introduced by Bai et al., 2016. As mentioned in section 1.2.2, the purpose of the AI is to provide a transparent way to summarize raw accelerometry data given the limitation of AC.

The formula of the AI is stated as

$$AI_i(t,H) = \sqrt{max(\frac{1}{3}\{\sum_{m=1}^{3} \frac{\sigma_{im}^2(t;H) - \bar{\sigma}_i^2}{\bar{\sigma}_i^2}\}, 0)},$$

where $\sigma_{im}^2(t;H)$ denotes the variance of Participant i's acceleration signals along axis $m(m = 1,2,3)$ in the window of length H starting at t and $\bar{\sigma}_i^2$ denotes the systematic noise variance when the device is placed steady.

The formula is fairly straightforward and thus easy to implement efficiently in a large scale study. Besides the systematic noise variance, one can see that the variance of raw accelerometry data along the three axes are chosen as key components to construct the AI. The reason is that the variability of raw acceleration signals (standard deviation or variance) in short epochs can provide a summary measure of activity intensity with gravity from the measured acceleration removed. Moreover, the standard deviation can also capture the changed data pattern from one activity to another. For example, when a participant switched from walking to running, the standard deviations of the raw accelerometry data along the three axes can detect the increased variability of the signals, while the means may not change accordingly (Bai et al., 2016). In addition, the calculation of AI does not depend on the location or orientation of accelerometer since it summarizes the standard deviations from all three directions. It is also easy to calculate aggregated AI in any window length using the 1-second AI. Lastly, with the AI calculated in the relative scale respect to the systematic noise variance, a value of 1 is equivalent to the smallest amount of variability detectable by the device and the AI will end up in a range similar to AC, which might be more familiar to researchers who have used AC before.

Given the advantages of AI, it was proved in (Bai et al., 2016) that AI is better in distinguishing among various types of physical activities across different intensity levels and

also has better prediction performance of METs as well as of classifying an given epoch into various intensity categories.

1.5 Outline of Thesis

In this thesis, we will conduct a novel analysis of the OPACH data based on the AI among older women. In Chapter 2, we will derive calibrating equations for estimating METs, based on accelerometry data as well as personal characteristics. In Chapter 3, cutpoints for AI will be derived for classifying each epoch into one of the PA intensity categories. Chapter 4 presents association analysis between AI-based PA patterns and cardiometabolic risk factors in the OPACH. We will end with concluding remarks and discussion for future work in Chapter 5.

Chapter 2

Calibration Analysis: Building Calibration Equations

For any metric of accelerometry (including the AI) to be useful, it is important to calibrate it against the golden standard of PA intensity, which is energy expenditure in METs. In this chapter, we use data collected in the OPACH calibration study to investigate the relationship between AI and activity intensity in METs and aim to build calibration equations to predict METs from accelerometry data.

2.1 An Overview of the Activity Index (AI)

The AI, introduced by Bai et al., 2016, is a summary metric constructed by using the variance of raw accelerometry data along three axes as building blocks to quantify the magnitude of acceleration during a given epoch (see the diagram in Figure 1). Mathematically, AI of Participant i in an epoch of length H starting at time t is defined by

$$AI_i(t,H) = \sqrt{max(\frac{1}{3}\{\sum_{m=1}^{3} \frac{\sigma_{im}^2(t;H) - \bar{\sigma}_i^2}{\bar{\sigma}_i^2}\}, 0)},$$

where $\sigma_{im}^2(t;H)$ denotes the variance of Participant i's acceleration signals along axis $m(m = 1,2,3)$ and $\bar{\sigma}_i^2$ denotes the systematic noise variance which depends on the accuracy of the device.

The AI has three desirable properties mathematically. First, AI can be easily imple-

mented in a computationally efficient way since it is an open-source metric proposed with the explicit formula. Second, it is additive. That is, AI for a time period can be obtained by summing up all 1-second AI within that period, which makes it possible to compare studies that choose different epoch lengths for AI. Third, AI is rotational invariant, which implies that it captures the magnitude of movement over three axes regardless of orientation of the device. This property is crucial in real-world setting since it guarantees the same AI values if the accelerometer is rotated during an activity.

As compared to AC, the traditional metric for summarizing movement in an epoch, Bai et al., 2016 showed that AI is more sensitive to activity with low intensity while AC usually corresponds to 0. So AI outperforms AC when quantifying activities especially for sedentary and light activities.

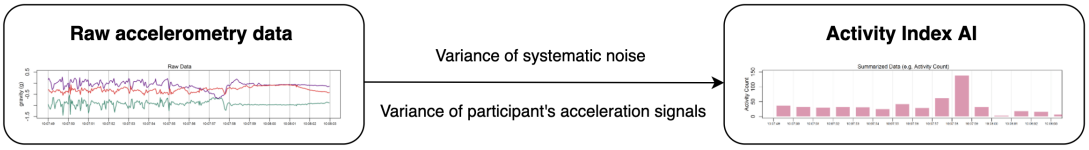

Figure 1: A general framework for accelerometry data transformation. The left panel illustrates the raw accelerometry data and right penal shows the summary metric AI.

2.2 OPACH Data Preprocessing and Exploratory Analysis

In the OPACH calibration study, there were 200 women in total and 193 women had complete raw data available, which were used in our analysis. Using formula in Bai et al. (2015), we first calculated AI per second for each subject. They were then aggregated into AI per 15-second. We also computed the standard deviation of each 15-second window of 1-sec AI to describe the variation of acceleration signals within the corresponding window.

The 193 women used in our analysis had mean age 75.48 years (standard deviation 7.70). Their ages distributed as follows: 60-69 years old ($n = 42$), 70-74 years old ($n = 55$), 75-81 years old ($n = 46$) and 82-91 years old ($n = 55$). For body mass index (BMI), there were 71 participants who had normal weight or underweight, 60 participants who were overweight and 62 who were obese.

Activity Type	Count	Min	Q1	Mean (SD*)	Median	Q3	Max
DVD	192	0.00	0.00	9.15 (27.63)	0.00	2.97	463.24
PUZZ	191	0.00	49.09	78.12 (45.16)	70.42	99.65	435.58
DISH	192	0.00	93.33	139.7 (72.13)	123.00	166.23	657.76
LAUD	192	0.00	152.68	203.28 (71.67)	195.00	243.70	633.09
MOP	191	38.06	290.16	383.72 (138.92)	363.45	453.64	1191.19
WALK	188	57.31	771.79	928.25 (257.96)	915.00	1077.98	2706.04

Table 1: Summary statistics of AI per 15 seconds of each activity in OPACH calibration study (N=200). *Abbreviation: standard deviation, SD.

Summary statistics for AI per 15-second are shown in Table 1 with following abbreviations: watching DVD while sitting quietly, DVD; assembling puzzle while sitting, PUZZ; washing dishes while standing, DISH; doing laundry while standing, LAUD; dust mopping while standing, MOP; 400-meter walking, WALK. As the energy cost of activities increase, the mean of AI per 15-second and its standard deviation increase.

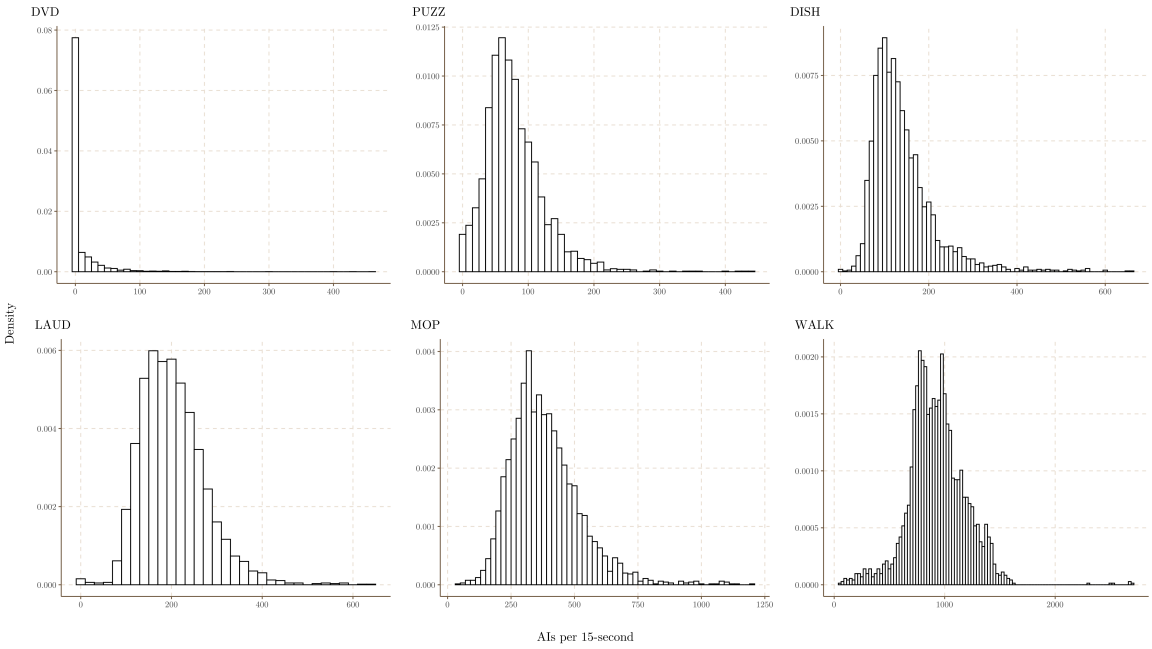

Figure 2: Histograms of AI per 15-second by activity type in OPACH calibration study (N=200).

Figure 2 displays the distribution for AI per 15-second for all participants by activity types. The overall shape of the distribution for each activity type is right-skewed, especially for the low energy cost of activities such as DVD. we found that the skewness tends to alleviate as the energy cost of activities increase.

Figure 3 shows distributions of AI per 15-second across three BMI groups. Notice that

AI per 15-second are often equal or close to 0 for DVD. Women with normal BMI and overweight women have higher proportions of 15-second epochs with exact zero AI during DVD, compared to obese women. For other activity types, AI distributions appeared to be similar among BMI groups.

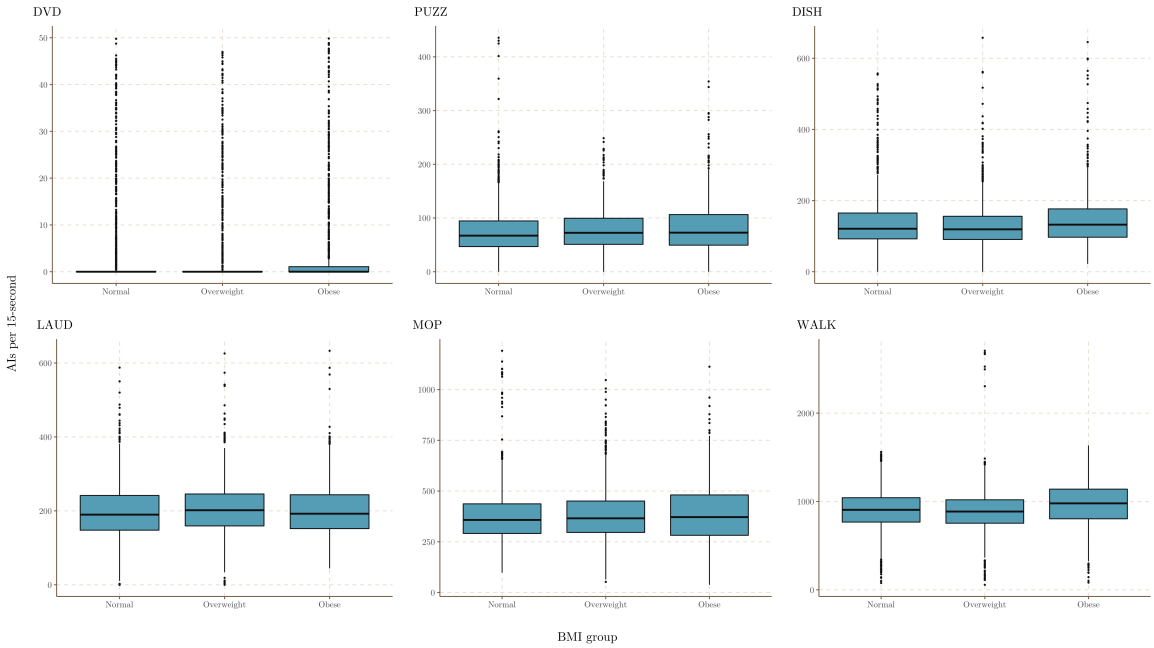

Figure 3: Boxplots of AI per 15-second and BMI group by activity type in OPACH calibration study (N=200).

Stratifying by age group (Figure 4), women younger than 70 years old tend to have slightly higher AI per 15-second in all activity types. For WALK, average AI per 15-seconds decrease with age. For other activities, the average AI per 15-second are similar among women aged 70 and older.

Summary statistics for METs are shown in Table 2. Similar to AI per 15-second, both mean and standard deviation of METs increase as the energy cost of activities increases. The overall distribution of each activity type has approximately normal shape (Figure 6).

Due to the skewness of AI, we tried two different transformations on AI per 15-second such as taking square root and natural logarithm. Figure 5 displays AI per 15-second on original scale, square root scale and natural log scale. Both transformations alleviate the right-skewness of the data to some extent. We then explored the relationships between METs and AI per 15-second, with and without the two transformations (Figure 7). One can see that METs are positively correlated with all three transformations, with coefficients of determination (R^2) values of 0.84, 0.86 and 0.71, respectively. However, different

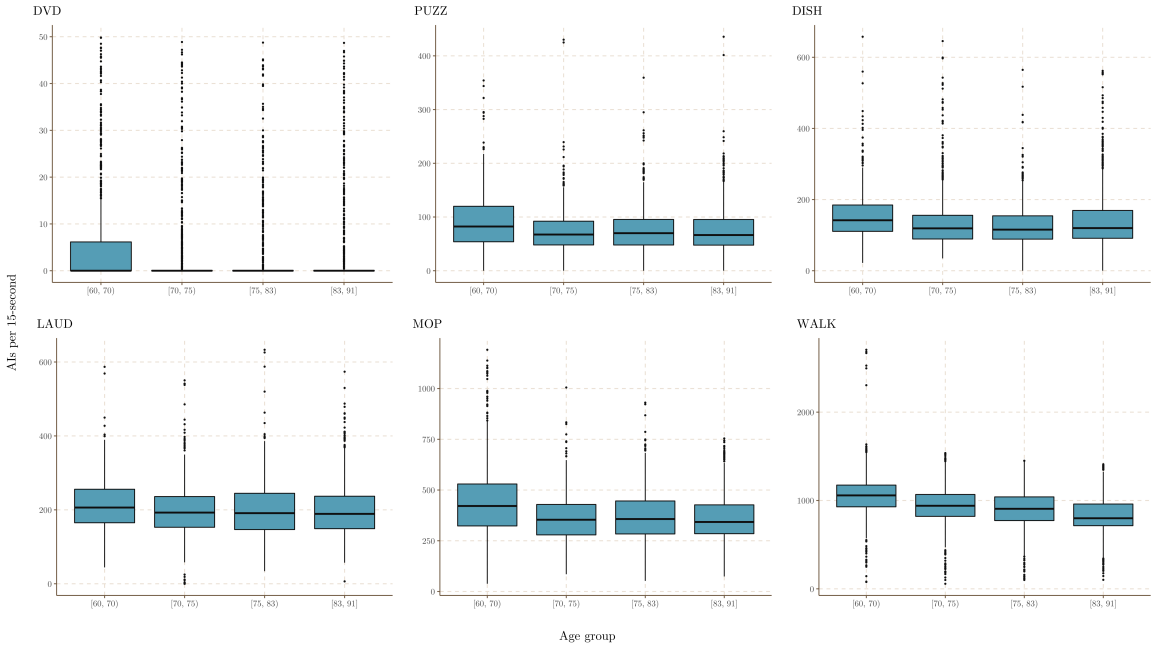

Figure 4: Boxplots of AI per 15-second and age group by activity type in OPACH calibration study (N=200).

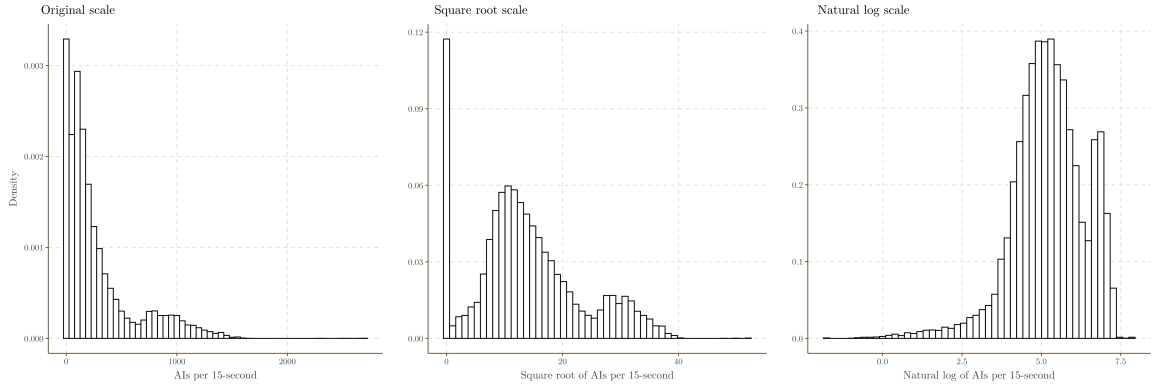

Figure 5: Histograms of AI per 15-second on original scale, square root scale and natural log scale.

Activity Type	Count	Min	Q1	Mean (SD)	Median	Q3	Max
DVD	192	0.45	0.85	1 (0.21)	0.98	1.13	1.88
PUZZ	191	0.54	1.15	1.34 (0.28)	1.35	1.50	2.40
DISH	192	0.79	1.53	1.76 (0.36)	1.77	1.98	2.89
LAUD	192	0.77	1.76	2.01 (0.4)	1.99	2.27	3.16
MOP	191	1.00	2.14	2.46 (0.56)	2.42	2.75	4.09
WALK	188	2.11	3.23	3.74 (0.69)	3.73	4.18	5.99

Table 2: Summary statistics of METs of each activity in OPACH calibration study (N=200).

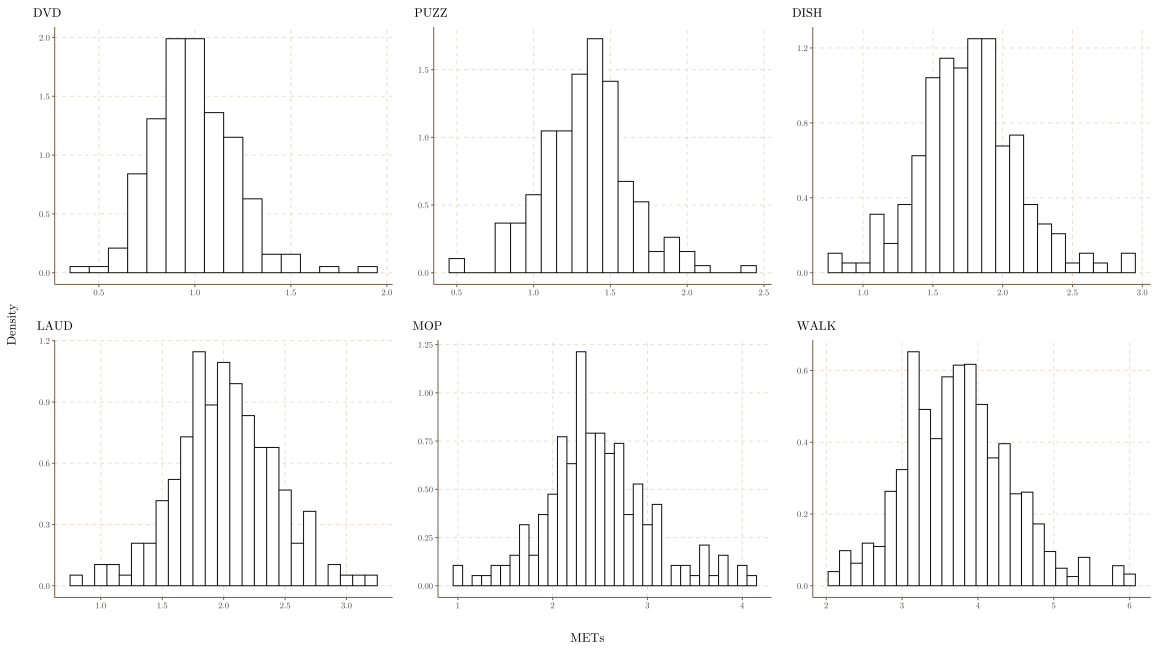

Figure 6: Histogram of METs by activity types in OPACH calibration study (N=200).

transformations reveal different patterns of the relationships especially among low intensity activities. In particular, data points for DVD, PUZZ and DISH tend to be clustered vertically on the original scale, while data points with low METs are more spread out with both square root and log transformations. The relationship between METs and square root of AI appear to be linear across the full range, while the relationship is clearly nonlinear for the log transformation.

2.3 Calibration by Linear Regression

Based on patterns observed from exploratory analysis, linear regression was first used to further study the relationship between AI and METs and to build calibration equations to predict METs.

When modeling the relationship between METs and AI per 15-second using linear regression, the fitted line is forced to go through the point $(0, 1)$ to reflect the scientific knowledge. Specifically, a zero AI per 15-second generally imply that a person is sitting still with little movement, so the corresponding activity intensity is exactly 1 MET.

We generated some diagnostic plots of linear regressions that model on AI per 15-second on the original, square root and logarithm scales to further determine the appropriate transformation used in the model (figure 8). The first row displays the residual plots of AI

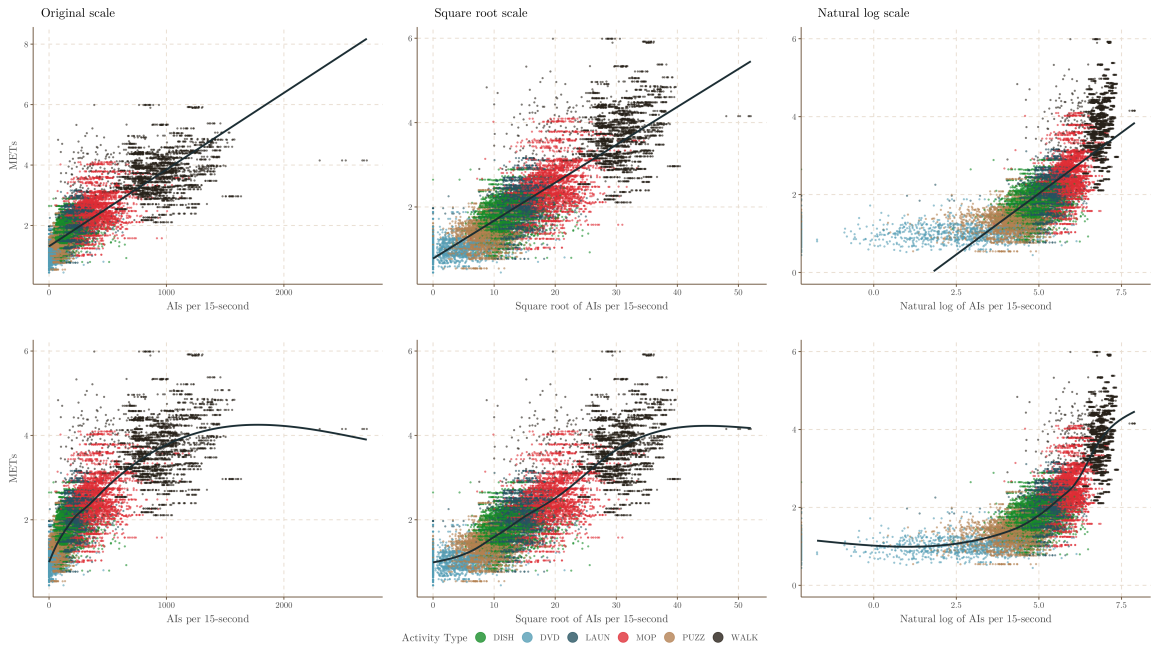

Figure 7: Scatterplots of METs and transformed AI per 15-second in OPACH calibration study (N=200). The top panels and the bottom panels show the scatterplots with METs and AI per 15-second on different scales with linear and nonlinear fit, respectively.

per 15-second on different scales with predicted METs values on the x-axis and residuals on the y-axis. For the original scale, we can see that data points tend to cluster at the beginning of the plot. Also, the difference between data points and the horizontal line at 0 gets larger as the as the fitted value gets larger, indicating a potential nonlinear relationship. For residual plot of AI per 15-second on square root scale, although such fanned pattern can also be observed, the data points are more symmetric and randomly dispersed around the horizontal line at 0 some extend. As for the log scale, a nonlinear pattern can be easily observed, which indicates that the relationship between METs and log AI per 15-second may be better described by a non-linear model. The second row shows the observed vs predicted plots of AI per 15-second in three different scales. We can see that when using the original scale, the observed METs centered around the predicted values, except underestimation at the right end. It implied that the prediction accuracy becomes worse as MET gets larger. For the square root scale, the prediction accuracy is clearly better especially for PA with moderate to higher METs. As for the log scale, the prediction performance is not pretty poor when using the linear model to describe the relation between METs and AI per 15-second on log scale.

Based on all the observations above, we conclude that linear regression is most appro-

priately used on the square root transformation of AI per 15-second, compared to the other two scales. Hence the square root transformation used for further analysis.

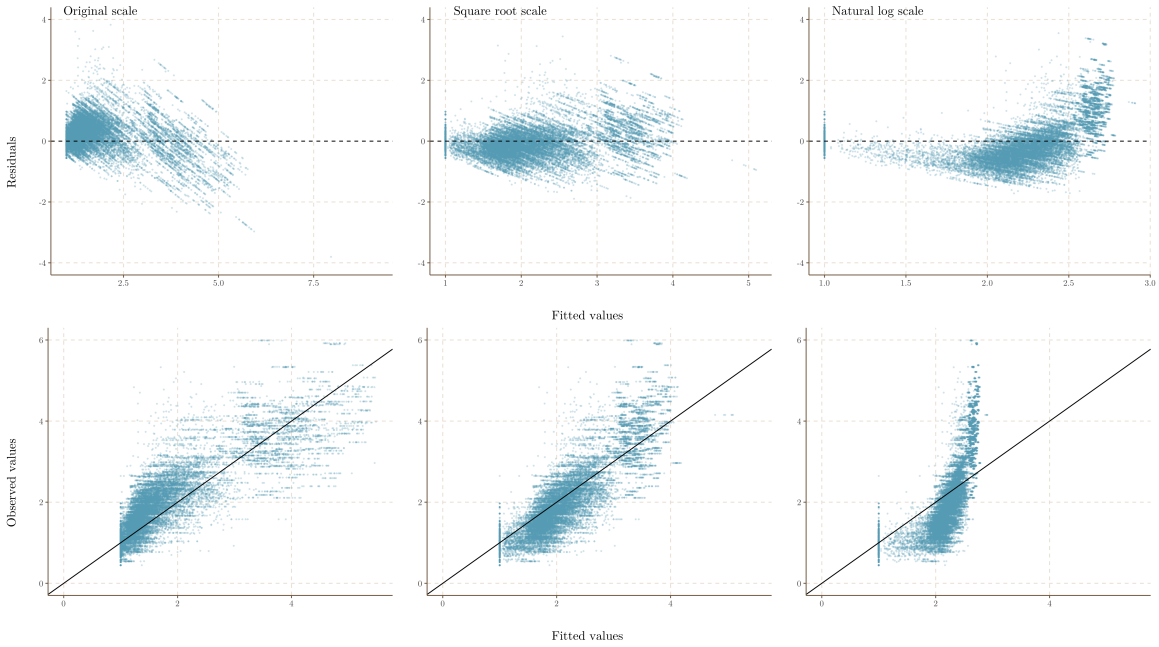

Figure 8: Diagnostic plots for linear regression lines using AI per 15-second on different scales. Figures in the top row display residual plots using AI per 15-second on original scale, square root scale and natural log scale, while the Bottom row displays observed values versus fitted values using AI per 15-second on original scale, square root scale and natural log scale.

Linear Regression Model	RMSE	R^2
1. $METs = 1 + \sqrt{AI15}$	0.5029	0.7472
2. $METs = 1 + \sqrt{AI15} + \sqrt{AI15} \times AgeGroup$	0.4820	0.7702
3. $METs = 1 + \sqrt{AI15} + \sqrt{AI15} \times BMIGroup$	0.4874	0.7638
4. $METs = 1 + \sqrt{AI15} + \sqrt{AI15_{SD}}$	0.4764	0.7592
5. $METs = 1 + \sqrt{AI15} + \sqrt{AI15_{SD}} + \sqrt{AI15} \times AgeGroup$	0.4550	0.7809
6. $METs = 1 + \sqrt{AI15} + \sqrt{AI15_{SD}} + \sqrt{AI15} \times BMIGroup$	0.4605	0.7753
7. $METs = 1 + \sqrt{AI15} + \sqrt{AI15} \times AgeGroup + \sqrt{AI15} \times BMIGroup$	0.4736	0.7786
8. $METs = 1 + \sqrt{AI15} + \sqrt{AI15_{SD}} + \sqrt{AI15} \times AgeGroup + \sqrt{AI15} \times BMIGroup$	0.4459	0.7897

Table 3: Performance summary of linear regression models.

Table 3 displays total 8 linear regression models we fit with square root of AI per 15-second (AI15) as the main predictor. All the models are forced to go through the point (0, 1) as we mentioned above. Meanwhile, interaction terms between square root of AI15 and age groups as well as BMI groups are included in some models to allow for more flexibility. Notice that we decided not to include the main term of age group and BMI group for the interaction to ensure that MET is 1 when AI15 is 0, among each age and BMI strata. Since

we aggregated the AI per 1-second into 15-second value directly, it's possible that some information may be lost within this 15-second window. So we also have models adding standard deviation of each 15-second window ($AI15_{SD}$) to evaluate whether $AI15_{SD}$ help improve the predicting of METs. We have scatterplot of METs and $AI15_{SD}$ displayed in appendix (figure 19).

Models are evaluated by root mean square error (RMSE) and R-squared (R^2) through 5-fold cross validation (Hastie et al., 2009). In 5-fold cross validation, the dataset D is randomly split into 5 mutually exclusive subsets with approximately equal size: D_1, D_2, D_3, D_4 and D_5. The model is trained and tested 5 times. Specifically, it is trained on the $D \setminus D_t$ (80%) and tested on D_t for $t \in 1, 2, 3, 4, 5$. For each round of cross validation, The RMSE and the R^2 is calculated. Finally, we averaged across these 5 RMSEs and R^2s to get the cross validation estimate of RMSEs and R^2s listed in table 3.

Model 1 is the baseline model with only square root of AI15 included. It has RMSE that equals to 0.5029 and R^2 that equals to 0.7472. When allowing the interaction with age group (model 2) and BMI group (model 3), RMSE is reduced to 0.4820 and 0.4874, while R^2 is increased to 0.7702 and 0.7638, respectively. Both of these two models yield better prediction accuracy of METs and model fit between METs and AI15. We also fit a model with only $AI15_{SD}$ adjusted (model 4). Compared to the baseline model, model 4 also achieved better RMSE and R^2. Although the RMSE of model 4 reduced more when compared to model 2 and model 3, the model fit does not improve as much as these two models do. For model 5 and model 6, we basically added $AI15_{SD}$ as a covariate into model 2 and model 3, respectively. Both of these models resulted in much lower RMSE and higher R^2. Among all these 8 models, the largest model (model 8) with $AI15_{SD}$ as well as interaction terms of age groups and BMI groups has the best performance which yields RMSE 0.4459 and R^2 0.7897.

Table 4 shows the coefficients, standard errors and p-values for model 1, model 2, model 3 and model 8. The coefficients in model 1 indicates that for every unit increase in measurement of square root of AI15, there is, on average, 0.079 ($p - value < 0.05$) kilocalorie per kilogram per hour more consumption in METs. Figure 9 also display this regression line with a positive trend.

The coefficients in model 2 represent the average METs change in corresponding age group. Specifically, the older the age group, the more the consumption on METs. Based

Variable	Coefficient	Std. Error	p-value
Model 1			
$\sqrt{AI15}$	0.079	0.0002	<0.001
Model 2			
$\sqrt{AI15}\times$ $[60, 70)$ years old	0.070	0.0004	<0.001
$\sqrt{AI15}\times$ $[70, 75)$ years old	0.073	0.0004	<0.001
$\sqrt{AI15}\times$ $[75, 82)$ years old	0.080	0.0005	<0.001
$\sqrt{AI15}\times$ $[82, 91]$ years old	0.091	0.0004	<0.001
Model 3			
$\sqrt{AI15}\times$ normal group	0.087	0.0004	<0.001
$\sqrt{AI15}\times$ overweight group	0.080	0.0004	<0.001
$\sqrt{AI15}\times$ obese group	0.069	0.0004	<0.001
Model 8			
$\sqrt{AI15}\times$ $[60, 70)$ years old	0.103	0.0035	<0.001
$\sqrt{AI15}\times$ $[70, 75)$ years old	0.102	0.0007	<0.001
$\sqrt{AI15}\times$ $[75, 82)$ years old	0.110	0.0006	<0.001
$\sqrt{AI15}\times$ $[82, 91]$ years old	0.119	0.0007	<0.001
$\sqrt{AI15}\times$ normal group	0.103	0.0007	<0.001
$\sqrt{AI15}\times$ overweight group	0.097	0.0007	<0.001
$\sqrt{AI15}\times$ obese group	0.089	0.0006	<0.001

Table 4: Coefficient summary of linear regression model 1, model 2 and model 3 and model 8 in table 3.

on corresponding graph on figure 9, we can see that the difference between older age groups ($[83, 91]$ and $[75, 83)$) is substantial, while the difference between younger age groups ($[70, 75)$ and $[60, 70)$) tend to be smaller.

Similarly, the average METs changes in different BMI groups are also shown in table 4. From the coefficients, we can see that the higher the BMI of the group, the less the consumption in METs. This trend can also be observed in the graph on figure 9.

As for model 8, the first four coefficients in table 4 represent the average change in METs for different age groups after adjusting for the standard deviation of AI15 in square root scale and BMI groups. Compared to the strictly increasing trend of the coefficients in model 2, the average change in METs of [70, 75) age groups is now greater than that of [60, 70) age group adjusted for the same BMI group and the same standard deviation of AI15. But the difference between these two age groups is still subtle. And the average consumption of METs in each age group gets larger.

For the average change in METs of different BMI groups with age groups and standard deviation of AI15 adjusted, we can see that the overall trend does not change, in other words, participants will still have less average consumption in METs for every unit increase in AI15 if they have higher BMI. But the change of METs gets larger in each BMI group compared to the coefficients in model 3.

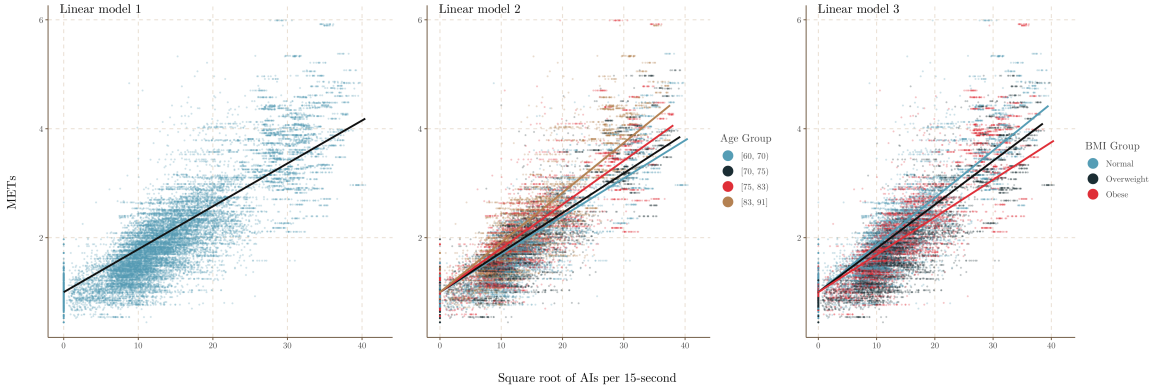

Figure 9: The first figure illustrate the scatterplot between METs and AI per 15-second. The second and third figure show the scatterplot between METs and AI per 15-second with respect to BMI group and age group, respectively. The linear regression lines in the figures correspond to model 1, model 2 and model 3 in table 3 and table 4.

2.4 Calibration by Nonlinear Regression

Given the results from linear regression, we found that a great amount of the data points are distributed around values with square root of AI15 smaller than 25 and METs smaller than 3 and the data points with larger AI15 and METs values are are more spread out. So it's possible that the linear regression may not be flexible enough to capture the trend of all data points due to the linear constrain. So we also fit nonlinear regression to model the association between METs and AI15.

For the nonlinear regression model, generalized additive models (S. N. Wood, 2011) implemented in package "mgcv" (S. Wood, 2006) is used. Specifically, we used options of cubic splines (S. N. Wood, 2003), penalized by the conventional intergrated square second derivative cubic spline penalty. Restricted maximum likelihood (REML) estimation is chosen as the approach to estimate the smoothing parameter (S. N. Wood, 2004). As for determining the number of knots (k), we chose k manually by grid searching. We fit the models repeatedly on a range of k from 15 to 35 and smoothness selection criterion was tracked for each of ks. We found that when k is around 30, there is no statistically important changes in the smoothness selection criterion. So we ended up using 30 knots in our models.

Nonlinear Regression Model	RMSE*	R^2
1. $METs = s(\sqrt{AI15})$	0.4705	0.7650
2. $METs = s(\sqrt{AI15}) + AgeGroup$	0.4526	0.7821
3. $METs = s(\sqrt{AI15}) + BMIGroup$	0.4504	0.7841

Table 5: Performance summary of generalized additive models. *Abbreviation: root mean square error, RMSE.

Table 5 displays the fitted nonlinear regression models. Similar to the linear regression, all the models are evaluated by root mean square error (RMSE) and R-squared (R^2) through 5-fold cross validation with 80% of the data as training set and 20% of the data as test set. We can see that the RMSE of the baseline nonlinear model is lower than the RMSE from baseline linear model (model 1 from table 3), which indicates better performance in terms of prediction. In addition, the baseline nonlinear model has higher R^2 compared to the baseline linear model. After adjusting for age groups (model 2), the performance tends to be better than the baseline model in terms of both RMSE and R^2 and it's also better than

the age-adjusted linear model (model 2 from table 3). Similarly, for BMI-adjusted model, it yields lower RMSE and higher R^2 compared to both baseline model and BMI-adjusted linear model (model 3 from table 3).

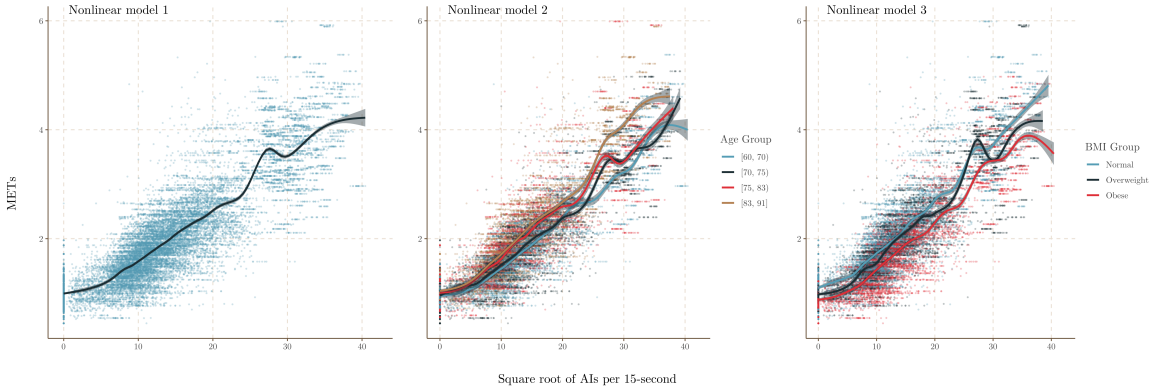

Figure 10: The first figure illustrate the scatterplot between METs and AI per 15-second. The second and third figure show the scatterplot between METs and AI per 15-second with respect to BMI group and age group, respectively. The nonlinear regression lines in the figures correspond to model 1, model 2 and model 3 in table 5.

Figure 10 displays the fitted nonlinear regression lines corresponding to model 1, model 2 and model 3 in table 5. Without the linear constrain, nonlinear regression line is more wiggly and can better interpolate the data points. For age-adjusted model, age group $[83, 91]$ still has the highest METs change compared to the other age groups given the same value change in AI15. But, unlike the pattern observed in age-adjusted linear regression models, the differences among the other three curves do not seem to be substantial due to some overlaps. For BMI-adjusted model, the overall trend among BMI groups does not change. The curve of obese group still lies below of the other two groups , and normal group tends to have larger METs change value compared to the overweight group given the same amount of change in AI15. However, in the area where AI15 is around 25, there is a sudden increase in overweight group. Specifically around this area, the METs change in overweight group exceeds the change in normal group.

Chapter 3

Calibration Analysis: Deriving Activity Intensity Cut-points

In this chapter, we determine cutpoints for classifying accelrometry counts into various PA intensity categories, including sedentary (SED), light (low and high) and moderate to vigorous physical activity (MVPA), using the data collected in the OPACH calibration study. Two approaches are adopted. The first one is based on the receiver operating characteristic (ROC) curve (Pepe, 2003) which is motivated by the cutpoints derivation for AC in the OPACH calibration study conducted by Evenson et al., 2015. The second one is based on regression models described in the previous chapter.

3.1 Classification-based Approaches

In the ROC curve analysis, the independent variable was participant's AI per 15-second value for each calibration activity and the dependent variable was a 0/1 binary indicator for the PA intensity category based on estimated METs value. Specifically, table 6 shows the METs values we used to define the intensity thresholds in the ROC analysis. For example, for sedentary intensity, activity was coded as sedentary (1) if its METs < 1.5 and non-sedentary (0) otherwise.

Once the ROC curve obtained, there are two methods for deriving the AI cutpoints. The first is maximized sum of true positive rate (sensitivity) and true negative rate (specificity), which is equivalent to choose a point on the ROC curve where the tangent slope is 1. The second is to balance the number of false positives and false negatives, which can be achieved

by calculating the difference between false positives and false negatives and then taking the threshold where the minimum absolute value of the difference occurred.

Intensity	Intensity thresholds definition based on MET values from each activity (where 1 MET = 3.5 mL/kg/min)
Sedentary (SED)	< 1.5
Light low (LL)	$[1.5, 2.25)$
Light high (LH)	$[2.25, 3)$
Moderate to vigorous (MVPA)	≥ 3

Table 6: Intensity thresholds definition based on MET values from each activity (N=200). Abbreviation: milliliters of oxygen per kilogram per minute, mL/kg/min.

The area under the intensity-specific ROC curve (AUC) (Pepe, 2003) was also calculated using logistic regression for each of the classifications as a measure of the accuracy to discriminate between two intensity classification, with values significantly greater than 0.5 indicating better discrimination than by chance alone.

Intensity	Cutpoint	Sensitivity	Specificity	Accuracy	Sensitivity+Specificity	AUC
Maximized sum of sensitivity and specificity						
SED vs LL	115.0	0.835	0.845	0.841	1.680	0.925
LL vs LH	219.3	0.872	0.845	0.854	1.717	0.938
LH vs MVPA	406.9	0.918	0.919	0.919	1.837	0.970
Balanced number of false positives and false negatives						
SED vs LL	100.7	0.787	0.882	0.849	1.670	0.925
LL vs LH	269.6	0.795	0.903	0.868	1.698	0.938
LH vs MVPA	572.9	0.828	0.968	0.946	1.796	0.970

Table 7: Hip-worn accelerometer cutpoints for AI derived from ROC-based approach (N=200); WHI OPACH Calibration Study, 2013.

The AI per 15-second cutpoints, sensitivity, specificity, accuracy, sum of sensitivity and specificity, and AUC are provided in table 7. Notice that accuracy is defined as the percentage of correctly classified instances $(TP+TN)/(TP+TN+FP+FN)$. where TP, FN, FP and TN represent the number of true positives, false negatives, false positives and true negatives, respectively.

We can see that the discrimination performance tends to be better as the intensity level increases for both ROC-based approaches. Comparing these two approaches, the cutpoint for classifying SED based on maximizing the sum of sensitivity and specificity is slightly higher than that based on balancing false positives and negatives, while the cutpoints for

classifying LH and MVPA based on the former approach are much lower those based on the latter. Since the difference between the cutpoints of SED from two approaches is not substantial, they yield similar sensitivity, specificity and accuracy. For the cutpoints of LH and MVPA, the sensitivity in the former approach are higher than that of the latter approach, which indicates slightly better performances of correctly identifying true LH and MVPA epochs. On the other hand, the latter approach yields higher specificity, implying that this approach is better at correctly rule out epochs that is not LH or MVPA.

3.2 Regression-based Approaches

In the regression-based approaches, linear baseline model (model 1 in table 3) and nonlinear baseline model (model 1 in 5) were used to derive cutpoints. Specifically, a range of AI per 15-second values were plugged into these two models to calculate METs, and the optimal cutpoints for each classification were taken where the computed MET was equaled or closest to corresponding intensity thresholds presented in table 6.

Intensity	Cutpoint	Sensitivity	Specificity	Accuracy	Sensitivity+Specificity	AUC
Linear regression						
SED vs LL	40.3	0.503	0.988	0.815	1.491	0.925
LL vs LH	251.7	0.821	0.884	0.864	1.706	0.938
LH vs MVPA	644.5	0.791	0.976	0.947	1.767	0.970
Nonlinear regression						
SED vs LL	82.6	0.716	0.927	0.852	1.643	0.925
LL vs LH	285.4	0.771	0.917	0.870	1.688	0.938
LH vs MVPA	615.8	0.804	0.973	0.947	1.777	0.970

Table 8: Hip-worn accelerometer cutpoints for AI derived from regression-based approach (N=200); WHI OPACH Calibration Study, 2013.

Table 8 displays the cutpoints results from regression-based approaches. When classifying SED and LH, linear regression yields smaller cutpoints. This is because the linear regression line lies above the nonlinear regression line when MET equals to 1.5 and 2.25 based on figure 11. However, from the graph, we can see that the cutpoint for MVPA from linear regression method is higher.

By comparing the evaluation metrics between these two approaches, although the specificity of SED from linear regression method is slightly higher, its sensitivity of SED from linear regression method is much lower. For the other two classifications, we don't see much

difference between the evaluation metrics.

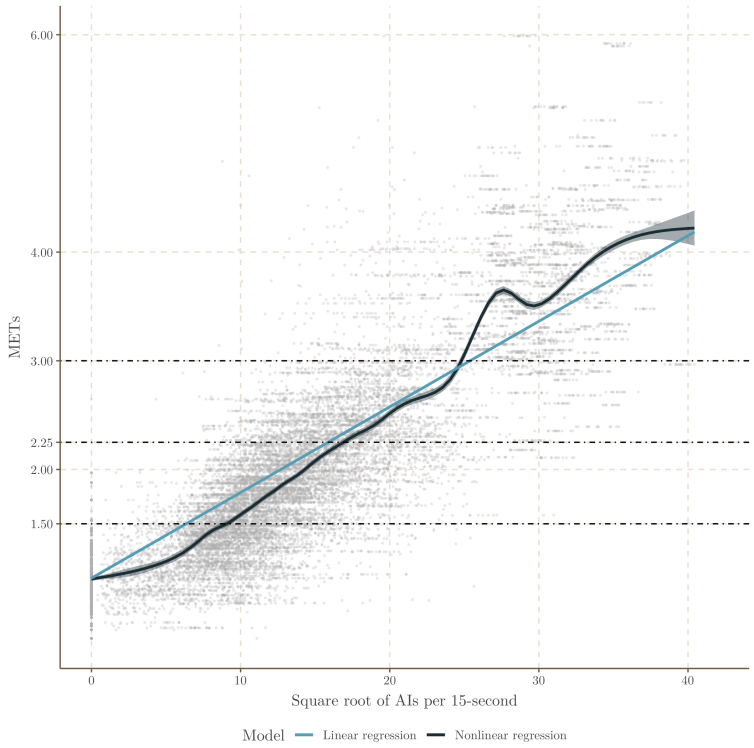

Figure 11: AI Cutpoints differences between linear regression criteria and nonlinear regression criteria. The dash lines from bottom to top represents the METs intensity thresholds for sedentary activity, light low activity, light high activity and moderate to vigorous activity.

Compared to the ROC-based approach (table 7), cutpoints from nonlinear regression tend to be closer to the ROC-based cutpoints, especially the balanced criteria.

Based on the approaches described above, BMI-specific AI cutpoints were also calculated due to the potential moderating effect of BMI on the classification. Figure 12 shows AUC values calculated from the logistic regression models with and without BMI as a covariate, and p-values on the plot indicate whether there is a significant difference between the AUC values from two models. Although p-value shows that AUC value from the model adjusted for BMI is statistically different with the model that does not adjust for BMI, the improvement in AUC is very small. Due to ease of implementation, we chose to report the uniform cutpoints in the main results, and BMI-specific results in the appendix table 22.

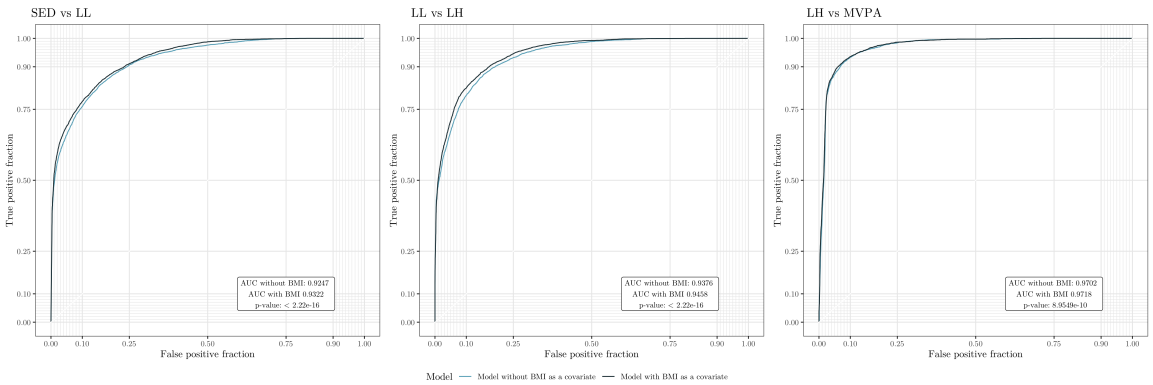

Figure 12: ROC curves for logistic regression models with- and without- adjusting for BMI groups. The corresponding AUC values are showed in the bottom right corner.

3.3 Sensitivity Analysis

Based on the cutpoint results above, we found that the specificity from all four criteria are generally very high, especially from the regression-based approaches. One possible reason is that, for each classification, there are some data points that will always be classified as negative case correctly. For example, data points with activity types such as MOP and WALK will usually have high METs and high AI values. As a result, when classifying between SED and LL, almost all of these data points will not be classified as SED and thus will be counted as true negatives, which will lead to high specificity. The proportion of these activity were fixed due to study design, which might have an undesirable consequence of artificially affecting sensitivity or specificity numbers. To evaluate the influence of study design on the derivation of cutpoints, we chose to conduct a sensitivity analysis to test the robustness of ROC-based approaches and regression-based approaches.

Specifically, we took three different subsets of the original data based on the activity types (table 9). For example, when classifying SED, we only used data points with activity types including DVD, PUZZ, LAUD and DISH to calculate the cutpoint, since MOP and WALK are almost always LH or MVPA and thus are not informative for deriving the SED/Light cutpoint. Both ROC-based approaches and regression-based approaches described above were applied on the new subset data in the sensitivity analysis to derive the cutpoints.

Intensity	Subset data based on activity types
SED vs LL	DVD, PUZZ, LAUD and DISH
LL vs LH	LAUD, DISH and MOP
LH vs MVPA	MOP and WALK

Table 9: Sensitivity analysis data for SED vs LL classification, LL vs LH classification and LH vs MVPA classification.

Table 23 and figure 13 shows the cutpoint results in and the AUC values calculated from the logistic regression models with and without BMI as a covariate in sensitivity analysis.

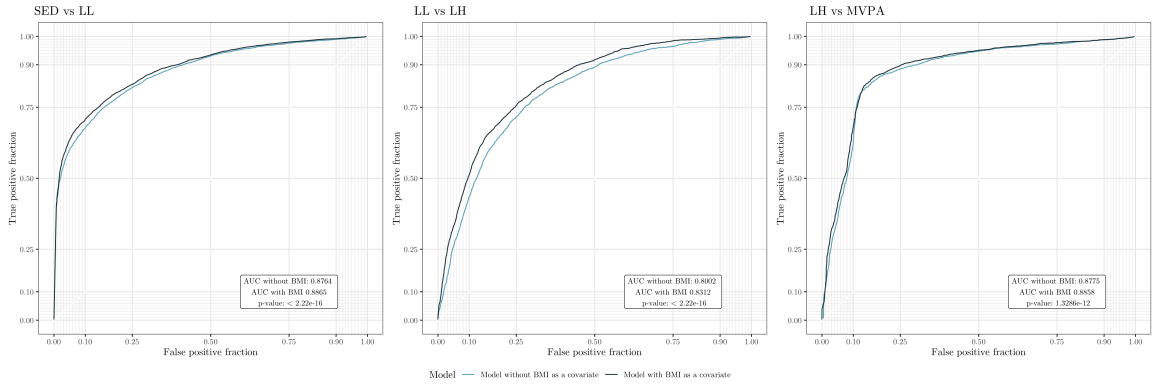

Figure 13: ROC curves for logistic regression models with- and without- adjusting for BMI groups in sensitivity analysis. The corresponding AUC values are showed in the bottom right corner.

We can see that, the cutpoints of each criteria indicate some changes in varying degree after excluding the non-informative data points. Specifically, the cutpoint for SED decreases from 115.0 to 98.5 when maximized sum of sensitivity and specificity, while the cutpoints for LH and MVPA become larger. For the cutpoints taken by balanced number of false positives and false negatives, SED cutpoint and LH cutpoint both drop slightly and MVPA cutpoint increases. For the cutpoints from regression-based approaches, although there are some great changes comparing between linear regression cutpoints before and after the sensitivity analysis, the difference for nonlinear regression cutpoints is subtle. As for the evaluation metrics, we can see the sensitivities, specificities and AUC scores in each approach become smaller.

In addition, after excluding the non-informative data points, the cutpoints of these four criteria become much closer, especially the balanced criteria and the nonlinear regression criteria.

Overall, among these four criteria, nonlinear regression criteria and balanced criteria is

relatively more robust since both of the criteria yield little change in cutpoints in sensitivity analysis.

Intensity	CutPoints	Sensitivity	Specificity	Accuracy	Sensitivity+Specificity	AUC
Maximized sum of sensitivity and specificity						
SED vs LL	98.5	0.795	0.785	0.790	1.581	0.876
LL vs LH	269.6	0.633	0.814	0.753	1.447	0.800
LH vs MVPA	578.6	0.834	0.844	0.839	1.677	0.878
Balanced number of false positives and false negatives						
SED vs LL	85.5	0.744	0.848	0.795	1.592	0.876
LL vs LH	227.4	0.752	0.723	0.733	1.475	0.800
LH vs MVPA	625.7	0.811	0.871	0.842	1.681	0.878
Linear regression						
SED vs LL	65.9	0.648	0.923	0.782	1.571	0.876
LL vs LH	302.2	0.541	0.861	0.753	1.402	0.800
LH vs MVPA	553.9	0.846	0.828	0.837	1.675	0.878
Nonlinear regression						
SED vs LL	82.8	0.730	0.860	0.794	1.590	0.876
LL vs LH	294.6	0.565	0.854	0.756	1.418	0.800
LH vs MVPA	607.3	0.819	0.861	0.841	1.680	0.878

Table 10: Hip-worn accelerometer cutpoints for AI derived from ROC-based approach and regression-based approach in sensitivity analysis (N=200); WHI OPACH Calibration Study, 2013.

Chapter 4

Association Analysis of Physical Activity and Cardiometabolic Health Biomarkers

In this chapter, we examined the association between accelerometer-measured PA intensity with various cardiovascular disease (CVD) risk factors (Kohl, 2001; Prentice et al., 2011) by using the AI cutpoints we derived from balanced criteria and nonlinear regression criteria derived in chapter 3 to classify intensity of PA.

4.1 OPACH Main Study Processing and Exploratory Analysis

In the OPACH main study, there were total 7048 women who were without known CVD recruited in the sample, and each woman was asked to wear the accelerometer for 7 days during both waking and sleeping hours, except when bathing or swimming. Among them, we calculated AI per-second and ACs per-second for 6078 women. Then we aggregated them into AI per 15-second and VM counts per 15-second, which were used for further analysis.

We defined a measurement day to be valid if the awake wear time in that day is greater than 10 hours. For each woman, we first calculated the total minutes of awake wear time, total minutes spent on SED, LL, LH activity and MVPA on all the valid days using the

aggregated AI and ACs with selected cutpoints (table 11) total epochs of ACs per 15-second and AI per 15-second. We then averaged them on the total number of valid days and the total number of epochs on the valid days that the woman had.

Specifically, for the cutpoint selection, we used the ACs cutpoints derived from OPACH calibration study introduced by Evenson et al., 2015. We chose the method that defined a MET as 3.0 ml/kg/min and with data processed using the normal frequency filter, where cutpoints were derived by balancing false positive and false negatives. For AI, we used two sets of the cutpoints that were derived by balancing false positive and false negatives, as well as nonlinear regression criteria in chapter 3.

Intensity	Cutpoint Value
Activity Count (AC) per 15-second	
Sedentary (SED)	≤ 18.5
Low light (LL)	$(18.5, 225.2]$
Low high (LH)	$(225.2, 518.5]$
Moderate to vigorous (MVPA)	≥ 518.5
Activity Index (AI) per 15-second	
Criteria: balancing false positive and false negatives	
Sedentary (SED)	≤ 100.7
Low light (LL)	$(100.7, 269.6]$
Low high (LH)	$(269.6, 572.9]$
Moderate to vigorous (MVPA)	≥ 572.9
Criteria: nonlinear regression	
Sedentary (SED)	≤ 82.6
Low light (LL)	$(82.6, 285.4]$
Low high (LH)	$(285.4, 615.8]$
Moderate to vigorous (MVPA)	≥ 615.8

Table 11: Categorization of physical activity intensity levels based on AC per 15-second cuptoints and AI per 15-second cutpoints.

The summary statistics of PA-related metrics for the cohort are shown in table 12. The mean awake wear time is 892.4 min/day in this cohort. Compared to the AC-based intensity time, the mean average LL time and the mean average LH time tend to be higher, while the mean average MVPA time is lower in the AI-based intensity time.

CVD risk factors, including cholesterol, low-density lipoprotein cholesterol (LDL cholesterol), high-density lipoprotein cholesterol (HDL cholesterol), triglyceride, glucose, insulin, and high-sensitivity C-reactive protein (CRP), were obtained through the fasting blood analysis performed in the University of Minnesota Fairview ARDL Laboratory. Other fac-

PA-related metric	Min	Q1	Mean	SD	Median	Q3	Max
Average awake wear time (min/d)	615.00	844.79	892.40	78.81	898.89	946.61	1158.96
Average AC per 15-second (epoch/d)	9.88	72.20	102.76	42.36	97.16	127.87	407.58
Average AI per 15-second (epoch/d)	31.14	107.16	140.49	46.56	135.30	168.50	436.22
AC-based intensity time							
Average SED time (min/d)	196.92	487.04	552.90	99.46	555.25	619.68	993.75
Average LL time (min/d)	29.25	153.45	188.74	50.02	186.58	219.79	401.00
Average LH time (min/d)	5.82	74.04	99.10	35.70	96.75	121.54	308.96
Average MVPA time (min/d)	0.64	25.38	51.66	34.84	44.75	70.33	361.18
AI-based intensity time							
Criteria: balancing false positive and false negatives							
Average SED time (min/d)	189.58	488.82	559.99	104.95	561.96	631.32	992.79
Average LL time (min/d)	28.62	141.62	180.10	53.65	177.71	215.25	503.18
Average LH time (min/d)	4.93	82.11	110.43	40.63	106.64	134.79	349.54
Average MVPA time (min/d)	0.12	19.64	41.89	29.57	36.04	57.46	277.96
Criteria: nonlinear regression							
Average SED time (min/d)	149.08	450.45	524.76	108.83	526.50	599.62	966.75
Average LL time (min/d)	36.62	178.39	225.16	64.86	222.14	268.06	538.18
Average LH time (min/d)	3.79	79.46	107.71	40.87	103.50	132.46	345.39
Average MVPA time (min/d)	0.12	14.71	34.78	26.54	29.00	48.35	251.00

Table 12: Summary statistics of PA-related metrics derived from AC per 15-second cutpoints and AI per 15-second cutpoints for the overall cohort (N=4688).

tors such as averaged diastolic blood pressure and averaged systolic blood pressure were also included. Table 14 displays the summary statistics of cardiovascular characteristics for the cohort and the corresponding histograms are shown in figure 14. We can see that the distributions of triglyceride, insulin and C-reactive protein are heavily right-skewed and we ended up taking the natural logarithm on these three cardiovascular measures for further modeling.

Cardiovascular Measure	Min	Q1	Mean	SD	Median	Q3	Max
Total cholesterol	82.0	169.00	197.37	39.40	195.00	222.00	355.00
LDL cholesterol	18.0	91.00	115.28	34.33	112.00	136.00	273.00
HDL cholesterol	25.0	50.00	60.45	14.96	59.00	69.00	138.00
Triglyceride	24.0	71.00	109.13	56.80	96.00	131.00	766.00
Systolic BP	80.0	117.00	125.73	14.21	124.50	134.00	200.00
Diastolic BP	40.0	67.00	72.35	8.72	72.00	79.00	107.50
Glucose	20.0	84.00	98.30	27.77	92.00	103.00	519.00
Insulin	2.0	38.00	93.39	144.52	60.00	99.00	5747.00
C-reactive protein	0.1	0.88	3.62	8.09	1.79	3.58	294.53

Table 13: Summary statistics of cardiovascular characteristics for the overall cohort (N=4688).

We ended up having 4688 women with both accelerometer data and complete CVD risk factor measures. Participant characteristics included age, race, education level, BMI

and waist measurement were described for the entire cohort (table 14), and then compared according to quartiles of total PA based on the AI cutpoints derived by balancing false positive and false negatives using linear regression for continuous variables and χ^2 tests for proportions. The mean age among all women is 78.9 years, 52.3% are white, and 79.7% have at least some college education. More than one third of women are obese ($BMI \geq 30 kg/m^2$). Compared with women in the lowest quartile of total PA, those in the highest quartile are younger, less likely to be white and obese ($p - value < 0.001$). Education were not associated with total PA levels.

		Total Physical Activity Quartiles (Min/d)				
Variables	Overall	< 262.2	$[262.2, 328.2)$	$[328.2, 399.8)$	≥ 399.9	Trend, p-value
Age, y	78.9±6.7	81.4±6.3	79.4±6.6	78.2±6.4	76.3±6.3	< 0.001
63 to 69	466 (9.9)	49 (4.2)	100 (8.6)	126 (10.7)	191 (16.3)	< 0.001
70 to 79	1825 (38.9)	337 (28.7)	429 (36.7)	494 (42)	565 (48.3)	
80 to 89	2197 (46.9)	686 (58.5)	589 (50.4)	528 (44.9)	394 (33.7)	
≥ 90	200 (4.3)	101 (8.6)	51 (4.4)	28 (2.4)	20 (1.7)	
Race-ethnicity						
White	2453 (52.3)	715 (61)	633 (54.1)	600 (51)	505 (43.2)	< 0.001
Black	1429 (30.5)	324 (27.6)	366 (31.3)	365 (31)	374 (32)	
Hispanic	806 (17.2)	134 (11.4)	170 (14.5)	211 (17.9)	291 (24.9)	
Education						
High school or less	946 (20.3)	236 (20.2)	245 (21.1)	224 (19.2)	241 (20.6)	0.271
Some college	1797 (38.5)	482 (41.3)	432 (37.2)	445 (38.1)	438 (37.5)	
College graduate	1921 (41.2)	448 (38.4)	483 (41.6)	500 (42.8)	490 (41.9)	
BMI	28±5.7	30±6.3	28.5±5.5	27.2±5.3	26.2±4.9	< 0.001
Normal	1554 (33.4)	263 (22.6)	320 (27.5)	434 (37.3)	537 (46.1)	< 0.001
Overweight	1401 (30.1)	522 (44.8)	383 (32.9)	286 (24.6)	210 (18)	
Obese	1701 (36.5)	379 (32.6)	462 (39.7)	443 (38.1)	417 (35.8)	
Waist	35.4±5.4	37.6±5.7	35.9±5.2	34.7±5.1	33.2 ±4.6	< 0.001

Table 14: Participant characteristics according to quartiles of total physical activity (N=4688). Data are mean±SD, or N (%).

Linear correlations between CVD risk factors and intensity measures were evaluated using Spearman correlations adjusted for average awake accelerometer wear time and age, and the correlations between age and CVD risk factors are unadjusted. The results are shown in Table 15. We can see that all CVD measures were correlated with each intensity measure in the expected direction with p-value$<$ 0.05. The result of linear correlation between CVD risk factors and intensity measures defined from nonlinear regression based

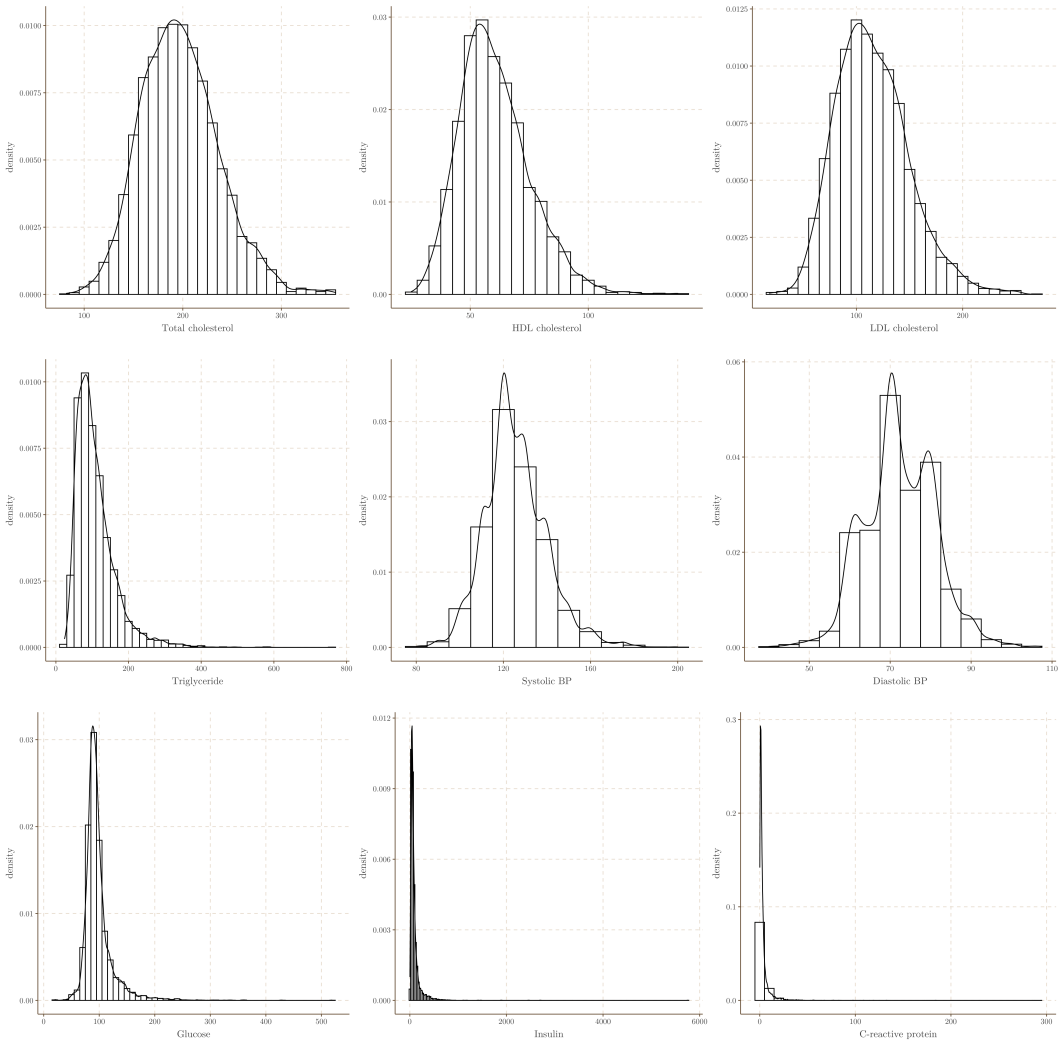

Figure 14: Histograms of cardiovascular measures for the cohort (N=4688).

AI can be found in appendix (table 24).

	Intensity Measure*				
Cardiovascular Measures	SED	LL	LH	MVPA	Age
Total cholesterol	-0.11	0.05	0.10	0.13	-0.05
LDL cholesterol	-0.09	0.04	0.08	0.11	-0.08
HDL cholesterol	-0.21	0.17	0.19	0.16	0.03
Triglyceride	0.17	-0.16	-0.14	-0.08	0.03
Systolic BP	0.10	-0.06	-0.09	-0.11	0.09
Diastolic BP	0.08	-0.07	-0.08	-0.05	-0.09
Glucose	0.12	-0.07	-0.12	-0.11	-0.01
Insulin	0.19	-0.12	-0.19	-0.18	-0.12
C-reactive protein	0.18	-0.09	-0.19	-0.17	-0.09
BMI	0.33	-0.21	-0.36	-0.29	-0.22
Waist girth	0.35	-0.23	-0.36	-0.32	-0.12
Age	0.28	-0.17	-0.18	-0.44	-

Table 15: Age- and average accelerometer wear-time adjusted Spearman correlations between activity and cardiovascular health measures (N=4688). Correlations between age and CVD risk factors are unadjusted. *Intensity measures are defined from AI cutpoints balancing false positive and false negatives.

4.2 Traditional Analysis

For a fixed amount of time engaged in activity in different intensity, activity choice may associate with CVD measures differently depending partly on other activities' displacement. Two different multiple linear regression models (single activity and substitution (Mekary et al., 2009; Stamatakis et al., 2015) were used to examine the associations between time changes in activity intensity per day and CVD measures. Both of the model adjusted for average awake wear time, that is, fixing the total amount of time engaged in activity per day, and confounders including age, race- ethnicity, education, waist measurements and BMI groups. Notice that when the outcome variable is BMI, we only adjust for average awake wear time, age, race- ethnicity, education and waist measurements. Triglyceride, insulin and C-reactive protein are modeled in their log scales.

4.2.1 Marginal Models

The single activity model adjusted for average awake wear time and other covariates is expressed as follows:

$$\text{Biomarker} = (\alpha_0) + (\alpha_1)\text{covariates} + (\beta_1)\text{average awake wear time} + (\beta_2)\text{activity}.$$

In this model, α_0 represents the intercept, α_1 represents the coefficients of adjusted covariates generally and β_1 represents the coefficient of average awake wear time variable. The coefficient of interest is β_2 which represents the effect of increasing the time spent on this intensity type of activity while holding the total awake wear time per day constant and adjusting for the other covariates. For example, if average time spent in SED activity is fit in the model, then the interpretation for the coefficient would be: every 1 min/day increment in SED activity is associated with β_1 change in the corresponding biomarker measurement in the model while holding the total awake wear time per day constant and adjusting for the other covariates.

Cardiovascular Measure	SED	LL	LH	MVPA
Total cholesterol				
Coefficient	-0.665	0.531	1.268	3.36
Std. Error	0.202	0.353	0.489	0.706
p-value	0.001	0.133	0.01	<0.001
LDL cholesterol				
Coefficient	-0.485	0.332	0.963	2.586
Std. Error	0.178	0.311	0.43	0.62
p-value	0.006	0.286	0.025	<0.001
HDL cholesterol				
Coefficient	-0.572	0.805	1.064	1.547
Std. Error	0.072	0.126	0.175	0.252
p-value	<0.001	<0.001	<0.001	<0.001
Triglyceride*				
Coefficient	1.018	0.973	0.967	0.966
Std. Error	0.002	0.004	0.005	0.008
p-value	<0.001	<0.001	<0.001	<0.001
Systolic BP				
Coefficient	0.153	-0.204	-0.22	-0.598
Std. Error	0.072	0.126	0.174	0.252
p-value	0.033	0.105	0.207	0.018

Cardiovascular Measure	SED	LL	LH	MVPA
Diastolic BP				
Coefficient	0.092	-0.223	-0.146	0.069
Std. Error	0.045	0.078	0.108	0.156
p-value	0.039	0.004	0.177	0.658
Glucose				
Coefficient	0.529	-0.372	-1.267	-2.337
Std. Error	0.141	0.246	0.341	0.493
p-value	<0.001	0.131	<0.001	<0.001
Insulin*				
Coefficient	1.015	0.984	0.975	0.941
Std. Error	0.004	0.006	0.009	0.012
p-value	<0.001	0.012	0.005	<0.001
C-reactive protein*				
Coefficient	1.027	0.979	0.93	0.912
Std. Error	0.005	0.009	0.011	0.016
p-value	<0.001	0.021	<0.001	<0.001
BMI$^+$				
Coefficient	0.133	-0.133	-0.385	-0.29
Std. Error	0.018	0.032	0.043	0.063
p-value	<0.001	<0.001	<0.001	<0.001

Table 16: Marginal effect of activity intensity, per 30-minute/day increase and cardiovascular measure changes. PA-related measures are classified by AI cutpoints from balancing the false positive and false negatives. $^+$Marginal effects between BMI and PA-related measures only adjust for average awake wear time, age, race-ethnicity, education and waist measurement. *The coefficients and std. errors for triglyceride, insulin and C-reactive protein are transformed from log scale to original scale by exponentiation and delta methods.

Table 16 displays the results of association between 30-minute/day increases in time spent in specific activity intensity categorized by AI cutpoints balancing the false positive and false negatives and cardiovascular measure change.

We can see that the marginal effect of SED activity for all measures are significant (p-value<0.05). For LL activity, the effects are significant (p-value<0.05) except for total cholesterol, LDL cholesterol, systolic BP and glucose. As for LH activity and MVPA, only the systolic BP and diastolic BP do not show significance on the coefficients, respectively.

In addition, there is an monotone pattern on the marginal effect with respect to the PA-related measures. Specifically, for most of the cardiovascular measures, the change is either strictly increasing or decreasing as the intensity levels increase. For example, the average amount of HDL cholesterol decreases by 0.572 for every 30-minute/day increases in time spent in SED activity while holding the total average awake wear time and other covariates as constant. Then the average amount of change starts to increase (0.805) when there is a 30-minute/day increase in the time spent in physical activities and the change becomes larger as the intensity level of physical activity increases. However, for other cardiovascular measures such as BMI, the average amount decreases the most (-0.385) when there is 30-minute/day increases in time spent in the light high activity among the PA-related measures.

Notice that the coefficients and std.errors for triglyceride, insulin and C-reactive protein are transformed back to their original scale. Specifically, we exponentiated the coefficients in log scale directly and applied the delta methods on the std. errors in log scale. The interpretation is then on the ratio level. Take triglyceride as an example, we expect to see about 1.8% increase in triglyceride when there is every 30-minute/day increase in time spent in SED activity.

The results for activity intensity categorized from nonlinear regression criteria are shown in table 25 of appendix. The differences between these two tables in terms of overall trend or absolute amount of change are not substantial.

4.2.2 Isotemporal Substitution Models

The isotemporal substitution model (Mekary et al., 2009; Stamatakis et al., 2015) with the following expression estimates the effect of replacing activities in one intensity with activities in another intensity for the same amount of time.

$$\text{Biomarker} = (\alpha_0) + (\alpha_1)\text{covariates} + (\beta_1)\text{average awake wear time} + (\beta_2)\text{activity}_a + (\beta_3)\text{activity}_b.$$

We aggregated the time spent in light low activity and light high activity together into light activity, and considered the substitution effect among SED activity, light activity and MVPA. Similarly, in this model, α_0 represents the intercept, α_1 represents the coefficients of adjusted covariates generally and β_1 represents the coefficient of average awake wear time variable. β_2 and β_3 are coefficients of respective activity in one type of intensity. For example, by eliminating one intensity from the model (e.g., MVPA), activity$_a$ stands for SED activity and activity$_b$ stands for light activity. In such case, β_2 represents the consequence of substituting 30 minutes/day of SED activity for MVPA while holding time spent in light activity constant and β_3 represents the consequence of substituting 30 minutes/day of light activity for MVPA while holding time spent in SED activity constant.

Cardiovascular Measure	SED to Light	SED to MVPA	Light to MVPA
Total cholesterol			
Coefficient	0.194	3.179	2.986
Std. Error	0.242	0.742	0.848
p-value	0.424	<0.001	<0.001
LDL cholesterol			
Coefficient	0.111	2.483	2.372
Std. Error	0.213	0.651	0.744
p-value	0.602	<0.001	0.001
HDL cholesterol			
Coefficient	0.471	1.108	0.637
Std. Error	0.086	0.264	0.302
p-value	<0.001	<0.001	0.035
Triglyceride*			
Coefficient	0.983	0.981	0.998
Std. Error	0.003	0.008	0.009
p-value	<0.001	0.018	0.869
Systolic BP			
Coefficient	-0.085	-0.518	-0.433
Std. Error	0.086	0.265	0.303
p-value	0.326	0.05	0.152
Diastolic BP			
Coefficient	-0.148	0.208	0.356
Std. Error	0.054	0.164	0.187
p-value	0.006	0.205	0.057
Glucose			
Coefficient	-0.231	-2.122	-1.891
Std. Error	0.169	0.518	0.591
p-value	0.172	<0.001	0.001
Insulin*			
Coefficient	0.993	0.947	0.954
Std. Error	0.004	0.012	0.014
p-value	0.106	<0.001	0.002
C-reactive protein*			
Coefficient	0.982	0.928	0.945
Std. Error	0.006	0.018	0.021
p-value	0.004	<0.001	0.008
BMI$^+$			
Coefficient	-0.126	-0.17	-0.045
Std. Error	0.022	0.066	0.076
p-value	<0.001	0.01	0.556

Table 17: Substitution effect of activity intensity, per 30-minute/day increase and cardiovascular measure changes. PA-related measures are classified by AI cutpoints from balancing the false positive and false negatives. $^+$Substitution effects between BMI and PA-related measures only adjust for average awake wear time, age, race-ethnicity, education and waist measurement. *The coefficients and std. errors for triglyceride, insulin and C-reactive protein are transformed from log scale to original scale by exponentiation and delta methods.

Table 17 displays the results for the substitution effects among SED activity, light activity and MVPA for all cardiovascular measures. The activity intensity is categorized by

AI cutpoints balancing the false positive and false negatives.

We can see that the substitution effects from SED activity to light activity are only statistically significant (p-value$<$ 0.05) for HDL cholesterol, triglyceride, diastolic BP, C-reactive protein and BMI. Take HDL cholesterol as an example, the substitution models suggest that substituting a 30-minute/day increase in SED activity for a 30-minute/day increase in light activity is associated with 0.471 amount of increase in HDL cholesterol. For the effect of 30-minute/day SED activity replacing 30-minute/day MVPA activity, only diastolic BP and systolic BP do not show significance. Moreover, the substitution effects from light activity to MVPA for these two measures are not significance, either. Such pattern on light activity to MVPA can also be observed in BMI.

Compared to the result of nonlinear regression criteria (table 26 in appendix), the difference is not substantial.

4.2.3 Comparison to AC

Given the results from marginal models and substitution models, we further compared the effect with the results of activity intensity derived from AC cutpoints. Table 18 and table 19 display the marginal effect and substitution effect, respectively.

Compared to the result of marginal effect on AI cutpoints, we can see that the effect of increasing 30-minute/day SED activity for systolic BP is now insignificant ($p - value = 0.189$). And the average change in BMI when increasing 30-minute/day MVPA is not significant, either. As for the effect pattern with respect to intensity, the overall change trends are very similar to the change trends observed in AI cutpoints result, except for HDL cholesterol, triglyceride, C-reactive protein and BMI. Take HDL cholesterol as an exmaple, the average change is the highest (0.972) among all the intensity when there is an 30-minute/day increase in light high activity. Within those cardiovascular measures that follow the trend, the overall difference of the absolute amount of change between the two results is not substantial (table 16).

As for the substitution effect, we don't see any difference in the effects of replacing 30-minute/day SED activity to equivalent amount of either light activity or MVPA in terms of significance. However, there are some differences for the substitution effect of 30-minute/day to 30-minute/day MVPA in HDL cholesterol, C-creative protein and BMI. Take BMI as an example, the AC substitution model suggests that substituting a 30-minute/day increase in

Cardiovascular Measure	SED	LL	LH	MVPA
Total cholesterol				
Coefficient	-0.626	0.049	1.361	2.975
Std. Error	0.214	0.386	0.564	0.572
p-value	0.003	0.899	0.016	<0.001
LDL cholesterol				
Coefficient	-0.455	-0.067	0.971	2.402
Std. Error	0.188	0.339	0.496	0.502
p-value	0.016	0.843	0.05	<0.001
HDL cholesterol				
Coefficient	-0.61	0.882	1.224	1.163
Std. Error	0.076	0.138	0.201	0.205
p-value	<0.001	<0.001	<0.001	<0.001
Triglyceride*				
Coefficient	1.019	0.968	0.964	0.972
Std. Error	0.002	0.004	0.006	0.006
p-value	<0.001	<0.001	<0.001	<0.001
Systolic BP				
Coefficient	0.142	-0.181	-0.2	-0.416
Std. Error	0.076	0.138	0.201	0.204
p-value	0.062	0.189	0.32	0.042

Cardiovascular Measure	SED	LL	LH	MVPA
Diastolic BP				
Coefficient	0.097	-0.263	-0.141	0.03
Std. Error	0.047	0.085	0.125	0.127
p-value	0.04	0.002	0.257	0.81
Glucose				
Coefficient	0.583	-0.414	-1.273	-1.953
Std. Error	0.149	0.269	0.393	0.399
p-value	<0.001	0.124	0.001	<0.001
Insulin*				
Coefficient	1.018	0.978	0.974	0.946
Std. Error	0.004	0.007	0.01	0.009
p-value	<0.001	0.002	0.012	<0.001
C-reactive protein*				
Coefficient	1.031	0.966	0.923	0.942
Std. Error	0.005	0.01	0.013	0.014
p-value	<0.001	<0.001	<0.001	<0.001
BMI$^+$				
Coefficient	0.133	-0.166	-0.488	-0.081
Std. Error	0.019	0.034	0.05	0.051
p-value	<0.001	<0.001	<0.001	0.115

Table 18: Marginal effect of activity intensity derived from AC cutpoints, per 30-minute/day increase and cardiovascular measure changes. $^+$Marginal effects between BMI and PA-related measures only adjust for average awake wear time, age, race-ethnicity, education and waist measurement. *The coefficients and std. errors for triglyceride, insulin and C-reactive protein are transformed from log scale to original scale by exponentiation and delta methods.

Cardiovascular Measure	SED to Light	SED to MVPA	Light to MVPA
Total cholesterol			
Coefficient	-0.079	3.025	3.103
Std. Error	0.269	0.596	0.72
p-value	0.77	<0.001	<0.001
LDL cholesterol			
Coefficient	-0.142	2.491	2.633
Std. Error	0.236	0.523	0.632
p-value	0.547	<0.001	<0.001
HDL cholesterol			
Coefficient	0.548	0.82	0.272
Std. Error	0.096	0.213	0.257
p-value	<0.001	<0.001	0.29
Triglyceride*			
Coefficient	0.979	0.986	1.006
Std. Error	0.003	0.006	0.008
p-value	<0.001	0.024	0.413
Systolic BP			
Coefficient	-0.076	-0.368	-0.292
Std. Error	0.096	0.213	0.257
p-value	0.428	0.084	0.256

Cardiovascular Measure	SED to Light	SED to MVPA	Light to MVPA
Diastolic BP			
Coefficient	-0.165	0.134	0.298
Std. Error	0.06	0.132	0.159
p-value	0.006	0.311	0.061
Glucose			
Coefficient	-0.222	-1.814	-1.592
Std. Error	0.188	0.416	0.502
p-value	0.238	<0.001	0.002
Insulin*			
Coefficient	0.991	0.952	0.961
Std. Error	0.005	0.01	0.012
p-value	0.063	<0.001	0.002
C-reactive protein*			
Coefficient	0.973	0.957	0.984
Std. Error	0.007	0.014	0.018
p-value	<0.001	0.004	0.369
BMI$^+$			
Coefficient	-0.181	0.034	0.216
Std. Error	0.024	0.053	0.064
p-value	<0.001	0.52	<0.001

Table 19: Substitution effect of activity intensity derived from AC cutpoints, per 30-minute/day increase and cardiovascular measure changes. $^+$Substitution effects between BMI and PA-related measures only adjust for average awake wear time, age, race-ethnicity, education and waist measurement. *The coefficients and std. errors for triglyceride, insulin and C-reactive protein are transformed from log scale to original scale by exponentiation and delta methods.

light activity for a 30-minute/day increase in MVPA is now associated with 0.216 amount of increase in BMI.

4.3 Functional Data Analysis

Traditional methods used above typically reduce continuous data on activity intensity into several categories using the cutpoints. However, such approach may lead to loss of statistical information embedded in the continuous data. In this section, we utilized a newly developed framework, proposed by Di et al. (2020) (Di et al., 2020), based on analysis of intensity as a continuous variable. Functional data analysis (FDA) (Ramsay and Silverman, 2005), as a key component in the framework, is a method that can analyze data with rich information embedded in curves with considering each sample element to be a function. Specifically, for accelerometry data, it allows for the use of the entire profile of daily AC/AI data that are collected over a 24-hour day, rather than summary values. An important technique in FDA called functional principal components analysis (FPCA) (Ramsay and Silverman, 2005) can examine the dominant modes of variation of the data as a method for understanding the major sources of data variability.

In this chapter, we utilized FDA on the same cohort data to study the association between volume of activity and cardiovascular measure, over the range of observed activity intensity.

4.3.1 Methods

Recently, Di et al. (2020) (Di et al., 2020) proposed an FDA-based approach to summarize PA intensity profile continuously and to flexibly model the dose-response relationship between PA accumulated at various intensity and health outcomes. More specifically, they first summarized up to a week of accelerometer data as a function—the complementary cumulative distribution function (CCDF).

To utilize the full pattern of physical activity intensity data for each participant, they proposed to use their entire activity intensity trajectory $z_i(t)$ in continuous fashion, instead of categorizing it based on cut points. Specifically, to describe the distribution of activity at different levels. They propose to use the complementary cumulative distribution function (CCDF): $S_i(z) = Pr(Z_i > z)$. While technically this function is a probability ("Pr")

function, note the value of the CCDF $S_i(z)$ at any intensity level z is interpretable as the proportion of time that participant i spends on activities more intense than the threshold z.

They then applied functional principal components analysis to the CCDF data, which identified major modes of variations among different women's PA profile. The analysis identifies distinct principal components, each of which represents a dimension along which functions vary from person to person. As a hypothetical example, one dimension could indicate how a given volume of physical activity is divided between lower versus higher intensity minutes. The order of principal components is determined by the amount of variation explained by each component, called eigenvalues. That is, the first component explains the most variations of patterns, while the second component explains the most of the remaining variation, etc. The main advantages of FPCA are two fold. First, it captures the major modes of variations of the complex data, which are helpful to interpret the high-dimensional functional data. Second, it also provides an effective data-adaptive tool to reduce dimensions for further analysis. In practice, it is common that the first few principal components explain most of variation of the data. Under such situations, principal component scores from these components provide a low dimensional summary and yet still preserve most information. Importantly, the scores are orthogonal (uncorrelated) to each other, making them very desirable to be used for further regression analysis.

Finally they proposed to use functional regression methods to estimate a coefficient function to describe the strength of association between volume of activity and a health outcome, over the range of observed activity intensity. Functional regression models are useful statistical tools to investigate the relationship between two variables, one or both of which involve functional data. The coefficient function describes the strength of association between minutes of activity and a health outcome, over the range of observed activity intensity. That is, suppose a research study derives a dose-response relationship between weekly minutes of physical activity at various intensity levels (counts or METs) and mortality risk. If it is a constant function, it means that every minute increase of physical activity has the same effect on health, whether it is at lower or higher intensity levels (e.g., 2-MET activity 5-MET activity). If it is a monotonically increasing function, it implies that MET accumulated at higher intensity levels have additional health benefits. Thus, inference on the coefficient function is highly relevant to understanding how the intensity of

physical activity influences its health benefits. This understanding is, in turn, important for public health physical activity guidelines. Currently, US public health guidelines for physical activity reflect that the health benefits of increasing activity within the intensity range of MVPA are well-established and can be quantified. In contrast, while growing evidence shows health benefits of increasing LPA/reducing SB, the evidence was deemed insufficient for quantitative guidelines.

4.3.2 Results

Applying the methods described above, we first calculated the estimated CCDF for each women based on their AI per 15-second data. Figure 15 illustrates the mean CCDF of all 4866 women and CCDF's for 10 randomly selected women. The value of a CCDF at the left end reflects the percent time spent in SED activity, light activity and MVPA of any intensity.

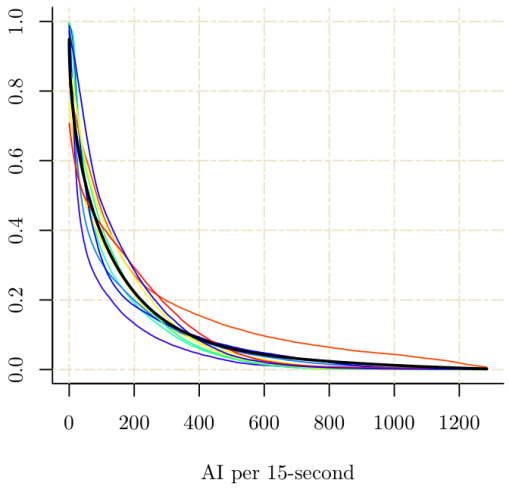

Figure 15: Mean CCDF* and CCDFs for 10 randomly selected women. *Mean CCDF is colored in black.

FPCA is then conducted to understand the major variability patterns in the CCDFs. Based on table 20, the first two principal components (PC) explained approximately 97% of the total variation (75% and 22% for the first and second PCs, respectively).

Figure 16 illustrates the modes of variation based on the two leading principal components. Specifically, the top panels show the shape of the two components with respect to

Principle component	Eigenvalue ($\times 10^{-4}$)	percentage of variance	Cumulative percentage of variance
PC1	14.8	75%	75%
PC2	4.4	22%	97%
PC3	0.5	3%	100%

Table 20: Summary of FPCA result with PC1, PC2 and PC3.

the AI per 15-second. The bottom panels display the type of variation described by the components, by showing the population mean curve as well as two colored lines with red line and green line corresponding to functional principal component curves added (+) and subtracted (-) from the mean function, respectively.

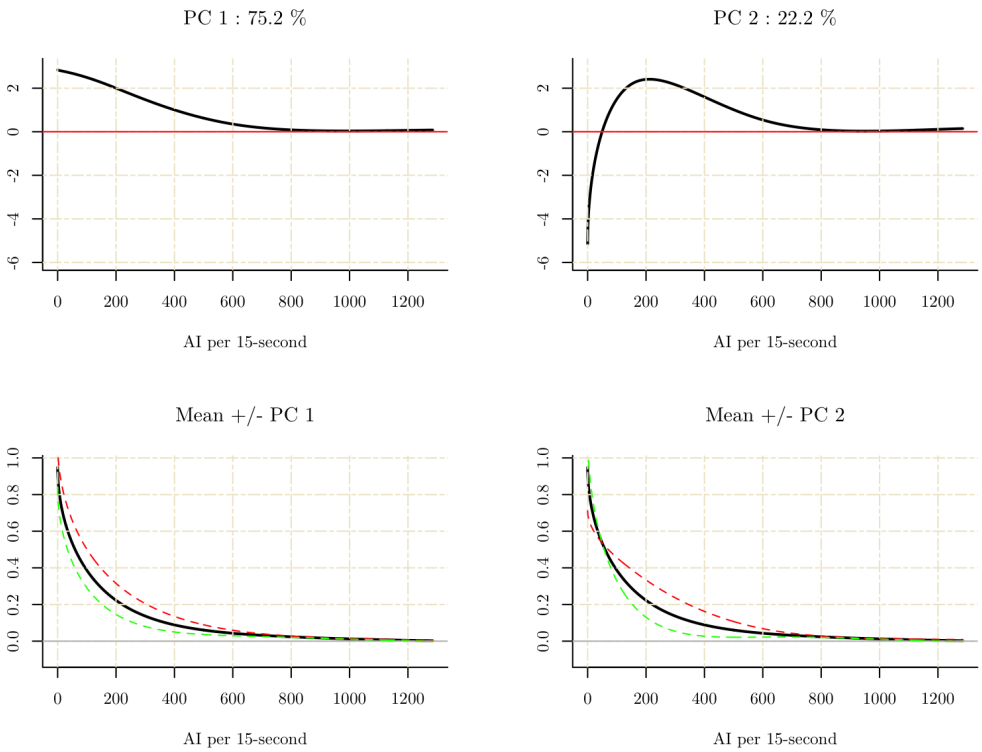

Figure 16: The two leading estimated functional principal components. The top panels show the shape of the first and second components with respect the AI per 15-second. In the bottom panels, the solid line represents the mean function, and the red line and green line are the functional principal component curves added (+) and subtracted (-) from the mean function.

The first PC clearly captures the variation in the volume of activity (Di et al., 2020), as the curve is shifted upwards (red dash line) overall in women with higher activity levels throught intensity levels; the curve shifts downward (green dash line) in less active women. For the second PC, it corresponds to the relative distribution of intensity levels (Di et al.,

2020). Given the volume of activity, one woman can do mainly LPA for a long time and little MVPA, while others can achieve the same volume of activity by doing less LPA but more MVPA. The shapes of the curve differ in the graph, steepness of the slope in particular, also conveys the same information that is captured by the second PC.

Variable	Coefficient	Std. Error	p-value
FPC Score 1* vs BMI groups			
Intercept	-0.2570	0.0247	<0.001
Overweight group	0.2480	0.0342	<0.001
Obese group	0.5510	0.0359	<0.001
FPC Score 1* vs Age groups			
Intercept	-0.4545	0.0450	<0.001
[70, 79]	0.2905	0.0504	<0.001
[80, 89]	0.6424	0.0495	<0.001
[90, 97]	0.9462	0.0821	<0.001

Variable	Coefficient	Std. Error	p-value
FPC Score 2* vs BMI groups			
Intercept	0.1895	0.0250	<0.001
Overweight group	-0.1675	0.0346	<0.001
Obese group	-0.4237	0.0363	<0.001
FPC Score 2* vs Age groups			
Intercept	0.2081	0.0456	<0.001
[70, 79]	-0.0543	0.0511	0.2879
[80, 89]	-0.3398	0.0502	<0.001
[90, 97]	-0.6495	0.0833	<0.001

Table 21: Summary of linear regression of FPC scores and BMI/age groups. *FPC score 1 and FPC score 2 are standardized by their means and standard deviations, respectively. One example interpretation of standardized coefficient for overweight group in FPC score 1 vs BMI groups is that there is 0.248 higher in standard deviation for FPC score 1 in overweight group.

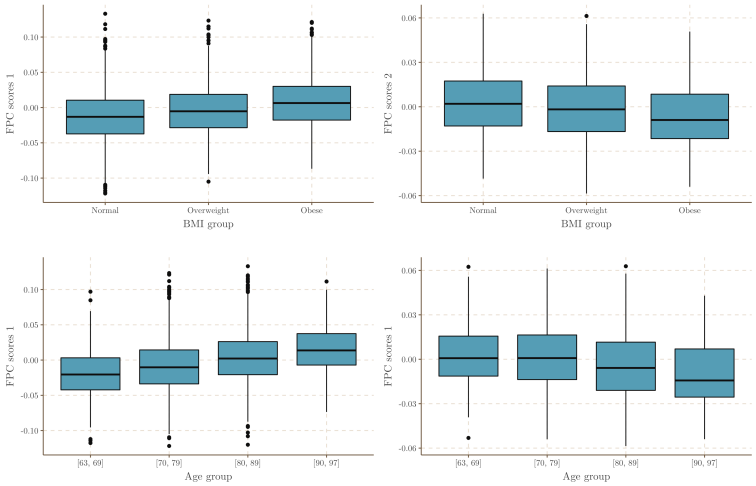

Figure 17: Boxplots of functional principal component (FPC) scores for the individuals and BMI groups/age groups.

We also compared if the FPC scores distribute differently according to participant characteristics such as BMI and age. Figure 17 displays the boxplots between two FPC scores and BMI/age groups and table 21 shows the corresponding linear regression results. Notice that in the linear regression, we standardized the FPC scores by their means and standard deviations. We can see that the coefficients are significant ($p-value < 0.05$), except for the age group [70, 79] in FPC score 2 model. Graphically, the difference between age group [63,

60] and [70, 79] with respect to FPC score 2 is subtle. Overall, both of the first and second FPC scores are significantly associated with BMI, which means that women with higher BMI are generally less active than those with lower BMI. Similarly, older women generally become less active as they age and they tend to shift their distribution of activities towards lighter intensity activities.

Functional regression models were then used to investigate the association between activity intensity with eight cardiovascular measures and BMI. For each cardiovascular measures, we evaluated the unadjusted effect and adjusted effect with average awake wear time, age, race-ethnicity, education and BMI. For BMI, we included above variables except for BMI to adjust the models.

Figure 18 displays the estimated regression coefficient function $\gamma(t)$, as well as its corresponding 95% confidence regions, with respect to METs from 1.5 to 6 and its corresponding AI per 15-second values. The coefficient function are adjusted for average awake wear time, age, race-ethnicity, education, BMI and waist measurement. While it is only adjusted for average awake wear time, age, race-ethnicity, education and waist measurement when BMI is evaluated as outcome.

We can see that the overall relationship is positive and monotonically increasing for total cholesterol, LDL cholesterol and HDL cholesterol. In other words, it suggests that women with higher activity levels have higher average total cholesterol, LDL cholesterol and HDL cholesterol. And the monotonically increasing shape indicates that the effect of activity on corresponding measures varies by activity intensity. Specifically, activity accumulated at higher intensity levels has larger effect on the measures while controlling for the total activity volume.

Similarly, for glucose and insulin in log scale, the overall relationship is mostly negative and monotonically decreasing, meaning that women with higher activity levels have lower corresponding outcomes. The monotonically decreasing shape indicates that activity accumulated at higher intensity levels has higher effect on the measures while controlling for the total activity volume.

There are also some gamma functions that are not monotone, that is, the effect on the measures varies with respect to the intensity while controlling for the total activity volume. For example, the gamma function of unadjusted BMI starts with a positive value, then decreases to around -0.5 and increases to some positive values at the end. One can see

that women can achieve the same effect on BMI by doing more light activity, and others by doing less MVPA. And after the intensity reaches some level (e.g. MET is 5), the effect on BMI starts to reverse from decreasing to increasing. While in the adjusted BMI result, the effect remains approximately constant after after 5 METs.

The unadjusted coefficients are presented in table 20 of appendix. For some of the biomarkers such as total cholesterol, HDL cholesterol, LDL cholesterol and glucose, the difference of coefficient change between adjusted model and unadjusted model is not substantial. But there are also some biomarkers' patterns that change substantially from unadjusted model to adjusted model. Such biomarkers includes log of triglyceride, diastolic BP and systolic BP. Overall, the effects tend to be attenuated after adjusted the covariates if there is a substantial pattern change comparing between adjusted- and unadjusted- model.

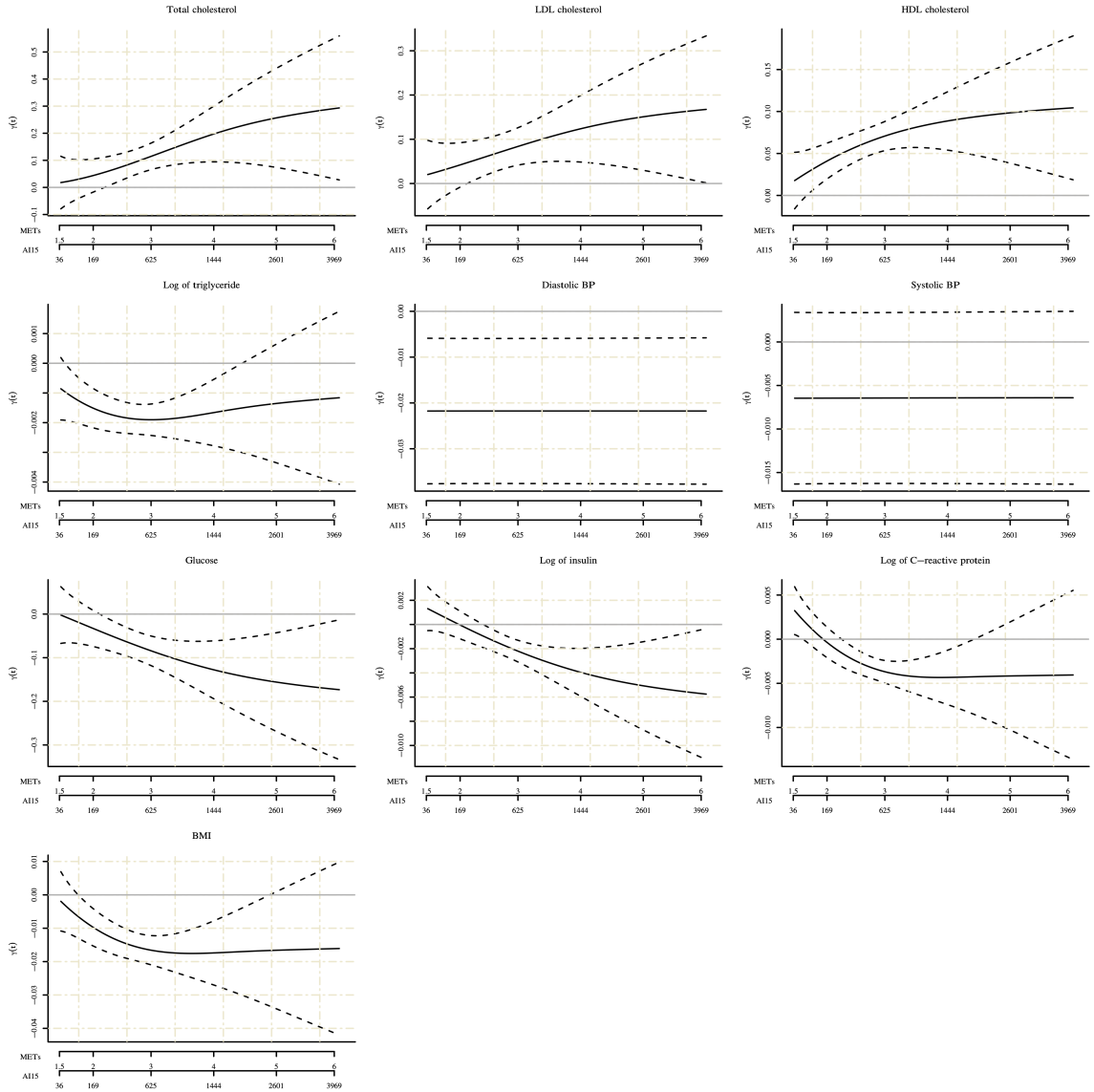

Figure 18: Functional regression analysis for cardiovascular measures with functional coefficients in black, 95% confidence intervals in dash lines, and zero demarcated in grey. The functional regression is adjusted for average awake wear time, age, race-ethnicity, education, BMI and waist measurement. Notice that the regression is only adjusted for average awake wear time, age, race-ethnicity, education and waist measurement when the outcome is BMI.

Chapter 5

Discussion

5.1 Conclusions

In this thesis, we derived calibrating equations for estimating METs based on AI summarized from raw accelerometry data of OPACH calibration study. Linear regression and nonlinear regression were used to model the association between AI and METs continuously. Covariates such as BMI, age and variation of AI within each 15 second window were adjusted in the models. The performance of each model were evaluated and compared based on RMSE and R^2. Overall, nonlinear regression models tend to have higher prediction accuracy and model fit compared to the linear models.

We then applied ROC-based approaches and regression-based approaches to derive the cutpoints of AI for PA intensity classification. Within the ROC-based approach, two different criteria including maximized sum of sensitivity and specificity, and balanced number of false positives and false negatives were applied. Within the regression-based approach, we derived the cutpoints through linear regression model and nonlinear regression model from calibration equation results. Evaluation metrics such as sensitivity, specificity and accuracy were used to compare the cutpoints derived from different criteria. We also conducted sensitivity analysis to evaluate the influence of study design on the derivation of cutpoints and to test the robustness of ROC-based approaches and regression-based approaches. By comparing the results, the differences among cutpoints from different criteria in sensitivity analysis are smaller compared to that in the normal analysis. Among these four criteria, nonlinear regression criteria and balanced criteria is relatively more robust since both of the results yield little change in cutpoints in sensitivity analysis. Hence cutpoints from

nonlinear regression criteria and balanced criteria were adopted in the association analysis.

To illustrate the potential usefulness of these cutpoints, we then analyzed the association of accelerometer measures summarized from corresponding cutpoints with ten health indicators: total cholesterol, LDL cholesterol, HDL cholesterol, triglyceride, diastolic BP, Diastolic BP, glucose, insulin, C-reactive protein and BMI. The analysis used accelerometry data from OPACH main study. As traditional analysis methods, marginal effect model and isotemporal substitution model were first used to study the association. Overall, the marginal model suggests that the increasing time spent in light high activity and MVPA contributes toward a more favorable cardiometabolic risk profile while holding the average awake wear time per day and other covariates as constant. Meanwhile, the substitution model suggests that the health risks are higher when replacing time spent in activity with higher intensity to activity with lower intensity (e.g. MVPA to SED). By comparing the results to the AC association results, AI-based summary measures are more strongly associated with heath indicators. In addition to the traditional analysis, we also adopted a newly developed framework that considers analysis of intensity as a continuous variable. Functional regression, as a key component in the frame, was used to estimate the coefficient function to describe the strength of association between volume of activity and a health outcome.

5.2 Future Work

There are several potential limitations and future work that can be done. For example, we only considered the general nonlinear regression model when constructing the calibration equation. However, such regression model dose not pose any constrain on the monotonicity. So nonlinear regression model with monotonicity assumption such as isotonic regression (Best and Chakravarti, 1990) could be conducted to calibrate the association and to derive the cutpoints. Another limitation is that we only investigate the difference between AI and AC in the functional regression framework. It remains an open problem to compare AI to the other summarized metrics such as ENMO (Hees et al., 2013). As many current studies have adopted summarized metrics other than AC, it is important to understand how AI differs with other summarized metrics. Finally, we only utilized the marginal model and isotemporal model to study the marginal effect and substitution effect of activities

on the biomarkers. However, a compositional data analysis framework can be applied to investigate the combined effect of time spent in physical activity and sedentary behaviors that together constitute a composite whole on cardiometabolic health markers (Carson et al., 2016; Chastin et al., 2015). Instead of considering each activity intensity as variable, compositional data analysis considers individuals' daily activity data as compositions, made up of mutually exclusive and exhaustive parts (times spent in sleep, sedentary behaviour, light physical activity (LPA) and MVPA) (Pedisic, 2014). In such framework, it is possible to detect asymmetrical estimates, that is, a quantum reallocation of one behaviour for another does not predict the exact inverse effect as the reverse reallocation (Dumuid et al., 2018).

Appendix

A.1

Figure 19, referenced in chapter 2.3, shows the scatterplot of METs and square root of AI per 15-second standard deviation.

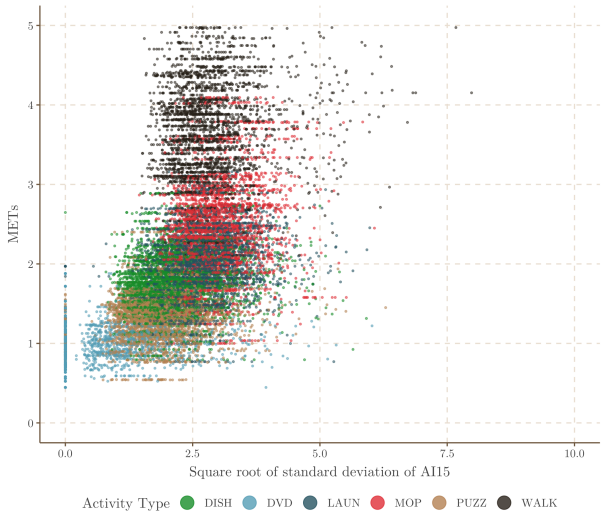

Figure 19: Scatterplot of METs and square root of AI per 15-second standard deviation (N=200)

A.2

Table 22 and table 23 (referenced in chapter 3.2 and chapter 3.3) shows the BMI-specific AI per 15-second cutpoints derived from four criteria for all participants in OPACH calibration study and for sensitivity analysis, respectively.

BMI Strata	Intensity	CutPoints	Sensitivity	Specificity	Accuracy	Sensitivity+Specificity	AUC
Maximized sum of sensitivity and specificity							
Normal	SED vs LL	98.4	0.828	0.846	0.840	1.674	0.924
	LL vs LH	217.4	0.825	0.887	0.864	1.712	0.934
	LH vs MVPA	378.2	0.909	0.924	0.921	1.832	0.972
Overweight	SED vs LL	115.2	0.892	0.854	0.866	1.746	0.951
	LL vs LH	220.0	0.896	0.848	0.864	1.744	0.947
	LH vs MVPA	485.4	0.924	0.953	0.948	1.877	0.974
Obese	SED vs LL	126.1	0.796	0.879	0.843	1.675	0.921
	LL vs LH	280.4	0.867	0.885	0.880	1.752	0.956
	LH vs MVPA	424.2	0.932	0.907	0.910	1.839	0.967
Balanced number of false positives and false negatives							
Normal	SED vs LL	80.5	0.764	0.896	0.856	1.660	0.924
	LL vs LH	219.6	0.819	0.890	0.863	1.709	0.934
	LH vs MVPA	475.9	0.825	0.964	0.940	1.789	0.972
Overweight	SED vs LL	94.3	0.806	0.903	0.870	1.709	0.951
	LL vs LH	271.8	0.814	0.910	0.878	1.724	0.947
	LH vs MVPA	634.0	0.866	0.976	0.959	1.841	0.974
Obese	SED vs LL	133.8	0.814	0.856	0.838	1.670	0.921
	LL vs LH	347.9	0.799	0.933	0.900	1.733	0.956
	LH vs MVPA	692.3	0.807	0.969	0.946	1.776	0.967
Linear regression							
Normal	SED vs LL	33.3	0.536	0.981	0.845	1.517	0.924
	LL vs LH	208.3	0.837	0.867	0.856	1.704	0.934
	LH vs MVPA	533.2	0.787	0.976	0.943	1.763	0.972
Overweight	SED vs LL	38.8	0.533	0.995	0.840	1.528	0.951
	LL vs LH	242.3	0.860	0.878	0.872	1.737	0.947
	LH vs MVPA	620.4	0.872	0.974	0.958	1.846	0.974
Obese	SED vs LL	52.9	0.483	0.997	0.773	1.479	0.921
	LL vs LH	330.4	0.813	0.925	0.897	1.738	0.956
	LH vs MVPA	845.7	0.579	0.977	0.922	1.557	0.967
Nonlinear regression							
Normal	SED vs LL	55.9	0.651	0.954	0.861	1.605	0.924
	LL vs LH	235.1	0.794	0.903	0.862	1.696	0.934
	LH vs MVPA	568.7	0.765	0.980	0.943	1.744	0.972
Overweight	SED vs LL	87.3	0.780	0.921	0.874	1.701	0.951
	LL vs LH	269.1	0.819	0.908	0.879	1.727	0.947
	LH vs MVPA	592.0	0.883	0.971	0.957	1.854	0.974
Obese	SED vs LL	115.4	0.764	0.905	0.844	1.669	0.921
	LL vs LH	388.0	0.765	0.950	0.904	1.715	0.956
	LH vs MVPA	693.0	0.806	0.969	0.946	1.775	0.967

Table 22: BMI-specific AI per 15-second cutpoints derived from four criteria for all participants in OPACH calibration study (N=200)

BMI Strata	Intensity	CutPoints	Sensitivity	Specificity	Accuracy	Sensitivity+Specificity	AUC
Maximized sum of sensitivity and specificity							
Normal	SED vs LL	70.7	0.724	0.857	0.798	1.581	0.866
	LL vs LH	217.4	0.725	0.760	0.745	1.486	0.908
	LH vs MVPA	536.2	0.798	0.867	0.830	1.665	0.887
Overweight	SED vs LL	86.5	0.779	0.854	0.817	1.633	0.796
	LL vs LH	220.5	0.817	0.715	0.751	1.532	0.842
	LH vs MVPA	556.6	0.919	0.848	0.882	1.768	0.820
Obese	SED vs LL	90.7	0.711	0.919	0.793	1.630	0.886
	LL vs LH	205.5	0.891	0.599	0.657	1.490	0.900
	LH vs MVPA	657.8	0.824	0.839	0.833	1.663	0.862
Balanced number of false positives and false negatives							
Normal	SED vs LL	80.5	0.764	0.811	0.790	1.575	0.866
	LL vs LH	221.0	0.712	0.768	0.743	1.480	0.908
	LH vs MVPA	477.7	0.837	0.810	0.825	1.648	0.887
Overweight	SED vs LL	94.6	0.808	0.813	0.810	1.621	0.796
	LL vs LH	268.3	0.683	0.829	0.777	1.511	0.842
	LH vs MVPA	652.1	0.868	0.884	0.876	1.752	0.820
Obese	SED vs LL	126.5	0.838	0.749	0.803	1.587	0.886
	LL vs LH	351.7	0.518	0.882	0.810	1.400	0.900
	LH vs MVPA	692.3	0.807	0.850	0.831	1.657	0.862
Linear regression							
Normal	SED vs LL	47.6	0.608	0.939	0.792	1.547	0.866
	LL vs LH	238.8	0.666	0.802	0.741	1.468	0.908
	LH vs MVPA	474.0	0.837	0.806	0.823	1.643	0.887
Overweight	SED vs LL	63.1	0.659	0.948	0.805	1.607	0.796
	LL vs LH	286.3	0.628	0.864	0.781	1.492	0.842
	LH vs MVPA	535.0	0.926	0.834	0.877	1.760	0.820
Obese	SED vs LL	108.5	0.783	0.835	0.803	1.617	0.886
	LL vs LH	433.8	0.358	0.935	0.822	1.293	0.900
	LH vs MVPA	690.3	0.810	0.850	0.832	1.659	0.862
Nonlinear regression							
Normal	SED vs LL	56.4	0.653	0.915	0.798	1.568	0.866
	LL vs LH	314.9	0.447	0.903	0.700	1.349	0.908
	LH vs MVPA	532.5	0.798	0.865	0.829	1.664	0.887
Overweight	SED vs LL	85.8	0.774	0.855	0.815	1.628	0.796
	LL vs LH	274.6	0.665	0.841	0.779	1.505	0.842
	LH vs MVPA	597.1	0.896	0.861	0.877	1.757	0.820
Obese	SED vs LL	117.0	0.808	0.798	0.804	1.606	0.886
	LL vs LH	404.1	0.425	0.921	0.824	1.346	0.900
	LH vs MVPA	699.7	0.801	0.850	0.829	1.652	0.862

Table 23: BMI-specific AI per 15-second cutpoints derived from four criteria in sensitivity analysis.

A.3

Table 24, referenced in chapter 4.1, shows the age- and average accelerometer wear-time adjusted Spearman correlations between activity and cardiovascular health measures. Notice that correlations between age and CVD risk factors are unadjusted. Intensity measures are defined from AI cutpoints based on nonlinear regression criteria.

	Intensity Measure*				
Cardiovascular Measures	SED	LL	LH	MVPA	Age
Total cholesterol	-0.11	0.05	0.10	0.13	-0.05
LDL cholesterol	-0.09	0.04	0.08	0.11	-0.08
HDL cholesterol	-0.21	0.17	0.19	0.16	0.03
Triglyceride	0.17	-0.17	-0.13	-0.08	0.03
Systolic BP	0.09	-0.06	-0.10	-0.10	0.09
Diastolic BP	0.08	-0.07	-0.07	-0.04	-0.09
Glucose	0.12	-0.07	-0.12	-0.11	-0.01
Insulin	0.18	-0.11	-0.18	-0.18	-0.12
C-reactive protein	0.17	-0.09	-0.19	-0.17	-0.09
BMI	0.32	-0.20	-0.36	-0.28	-0.22
Waist girth	0.34	-0.22	-0.36	-0.31	-0.12
Age	0.27	-0.15	-0.20	-0.45	-

Table 24: Age- and average accelerometer wear-time adjusted Spearman correlations between activity and cardiovascular health measures (N=4688). Correlations between age and CVD risk factors are unadjusted. *Intensity measures are defined from AI cutpoints based on nonlinear regression criteria.

Table 25, referenced in chapter 4.2, shows the marginal effect of activity intensity, per 30-minute/day increase and cardiovascular measure. Table 26, referenced in chapter 4.2, shows the substitution effect of activity intensity, per 30-minute/day increase and cardiovascular measure changes.

In both of the table 25 and table 26, PA-related measures are classified by AI cutpoints based on nonlinear regression criteria. The effects between BMI and PA-related measures only adjust for average awake wear time, age, race-ethnicity, education and waist measurement. The coefficients and std. errors for triglyceride, insulin and C-reactive protein are transformed from log scale to original scale by exponentiation and delta methods.

Cardiovascular Measure	SED	LL	LH	MVPA
Total cholesterol				
Coefficient	-0.596	0.42	1.33	3.727
Std. Error	0.19	0.292	0.488	0.784
p-value	0.002	0.151	0.006	<0.001
LDL cholesterol				
Coefficient	-0.427	0.249	1.01	2.873
Std. Error	0.167	0.257	0.428	0.688
p-value	0.011	0.332	0.018	<0.001
HDL cholesterol				
Coefficient	-0.538	0.662	1.063	1.674
Std. Error	0.068	0.104	0.174	0.28
p-value	<0.001	<0.001	<0.001	<0.001
Triglyceride				
Coefficient	1.017	0.978	0.967	0.964
Std. Error	0.002	0.003	0.005	0.008
p-value	<0.001	<0.001	<0.001	<0.001
Systolic BP				
Coefficient	0.147	-0.169	-0.255	-0.636
Std. Error	0.068	0.104	0.174	0.28
p-value	0.03	0.104	0.143	0.023

Cardiovascular Measure	SED	LL	LH	MVPA
Diastolic BP				
Coefficient	0.092	-0.184	-0.132	0.093
Std. Error	0.042	0.065	0.108	0.173
p-value	0.028	0.004	0.221	0.593
Glucose				
Coefficient	0.461	-0.272	-1.305	-2.543
Std. Error	0.132	0.204	0.34	0.547
p-value	<0.001	0.183	<0.001	<0.001
Insulin				
Coefficient	1.013	0.987	0.976	0.931
Std. Error	0.003	0.005	0.009	0.013
p-value	<0.001	0.014	0.006	<0.001
C-reactive protein				
Coefficient	1.023	0.986	0.93	0.908
Std. Error	0.005	0.007	0.011	0.018
p-value	<0.001	0.053	<0.001	<0.001
BMI				
Coefficient	0.115	-0.099	-0.371	-0.307
Std. Error	0.017	0.026	0.043	0.07
p-value	<0.001	<0.001	<0.001	<0.001

Table 25: Marginal effect of activity intensity, per 30-minute/day increase and cardiovascular measure changes. PA-related measures are classified by AI cutpoints based on nonlinear regression criteria. $^{+}$Marginal effects between BMI and PA-related measures only adjust for average awake wear time, age, race-ethnicity, education and waist measurement. *The coefficients and std. errors for triglyceride, insulin and C-reactive protein are transformed from log scale to original scale by exponentiation and delta methods.

Cardiovascular Measure	SED to Light	SED to MVPA	Light to MVPA
Total cholesterol			
Coefficient	0.215	3.517	3.302
Std. Error	0.216	0.811	0.892
p-value	0.318	<0.001	<0.001
LDL cholesterol			
Coefficient	0.124	2.754	2.63
Std. Error	0.19	0.712	0.782
p-value	0.514	<0.001	<0.001
HDL cholesterol			
Coefficient	0.447	1.238	0.792
Std. Error	0.077	0.289	0.318
p-value	<0.001	<0.001	0.013
Triglyceride			
Coefficient	0.984	0.978	0.994
Std. Error	0.002	0.009	0.01
p-value	<0.001	0.014	0.562
Systolic BP			
Coefficient	-0.095	-0.543	-0.448
Std. Error	0.077	0.29	0.318
p-value	0.215	0.061	0.16

Cardiovascular Measure	SED to Light	SED to MVPA	Light to MVPA
Diastolic BP			
Coefficient	-0.133	0.222	0.356
Std. Error	0.048	0.179	0.197
p-value	0.005	0.215	0.071
Glucose			
Coefficient	-0.217	-2.332	-2.115
Std. Error	0.15	0.566	0.622
p-value	0.149	<0.001	<0.001
Insulin			
Coefficient	0.993	0.938	0.944
Std. Error	0.004	0.014	0.015
p-value	0.089	<0.001	<0.001
C-reactive protein			
Coefficient	0.985	0.921	0.935
Std. Error	0.005	0.019	0.022
p-value	0.006	<0.001	0.003
BMI			
Coefficient	-0.104	-0.204	-0.1
Std. Error	0.019	0.073	0.08
p-value	<0.001	0.005	0.211

Table 26: Substitution effect of activity intensity, per 30-minute/day increase and cardiovascular measure changes. PA-related measures are classified by AI cutpoints based on nonlinear regression model. $^{+}$Substitution effects between BMI and PA-related measures only adjust for average awake wear time, age, race-ethnicity, education and waist measurement. *The coefficients and std. errors for triglyceride, insulin and C-reactive protein are transformed from log scale to original scale by exponentiation and delta methods.

Figure 20, referenced in chapter 4.3, shows the unadjusted functional regression analysis results for cardiovascular measures with functional coefficients in black, 95% confidence intervals in dash lines, and zero demarcated in grey.

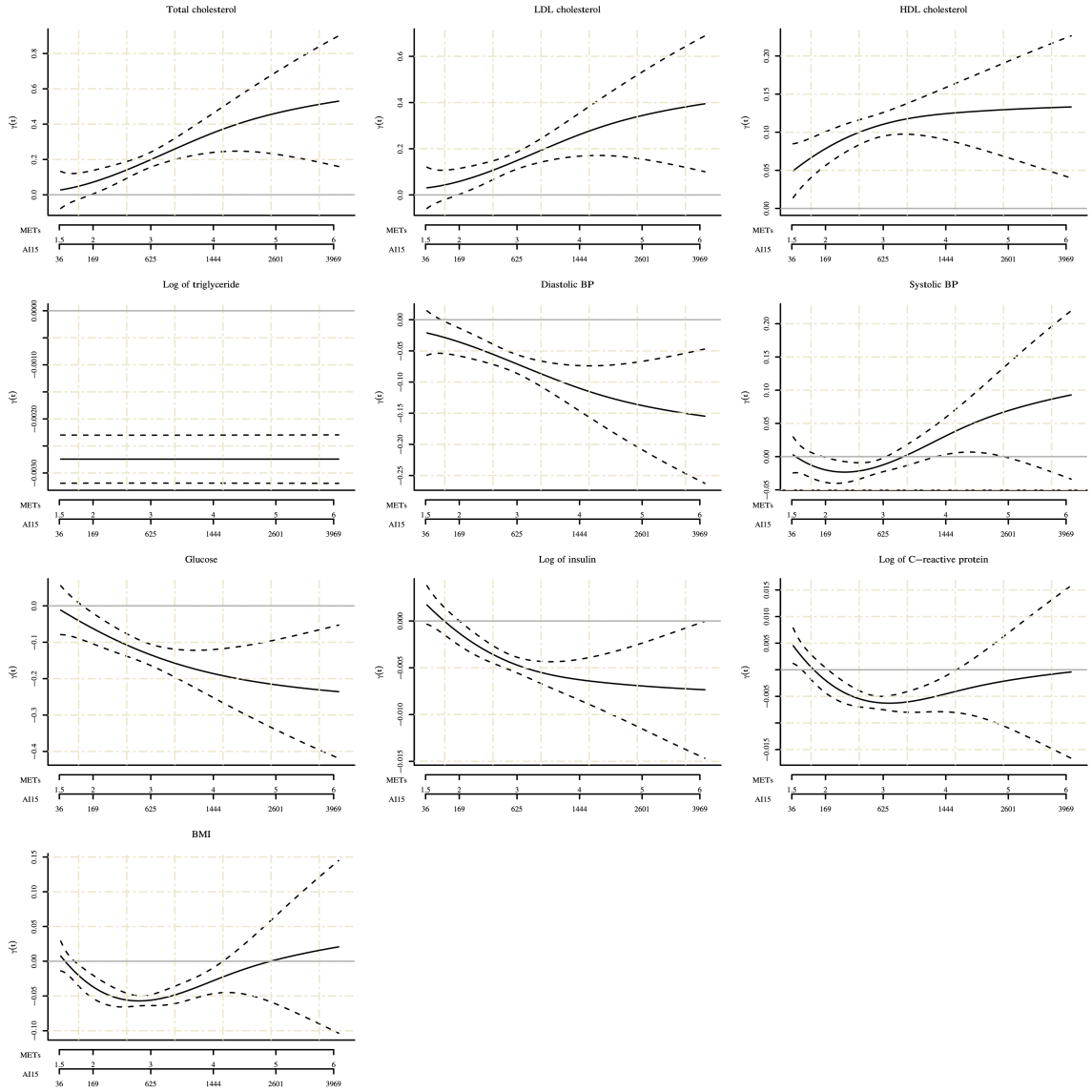

Figure 20: Functional regression analysis for cardiovascular measures with functional coefficients in black, 95% confidence intervals in dash lines, and zero demarcated in grey. The functional regression is unadjusted.

Bibliography

Anderson, G. L., Manson, J., Wallace, R., Lund, B., Hall, D., Davis, S., Shumaker, S., Wang, C.-Y., Stein, E., Prentice, R. L., & et al. (2003). Implementation of the womens health initiative study design. *Annals of Epidemiology*, *13*(9). https://doi.org/10.1016/s1047-2797(03)00043-7

Bai, J., Di, C., Xiao, L., Evenson, K. R., LaCroix, A. Z., Crainiceanu, C. M., & Buchner, D. M. (2016). An activity index for raw accelerometry data and its comparison with other activity metrics. *PLoS ONE*, *11*(8). https://doi.org/https://doi.org/10.1371/journal.pone.0160644

Best, M., & Chakravarti, N. (1990). Active set algorithms for isotonic regression; a unifying framework. *Math. Program.*, *47*, 425–439. https://doi.org/10.1007/BF01580873

BOUTEN, C., WESTERTERP, K., Verduin, M., & JANSSEN, J. (1994). Assessment of energy expenditure for physical activity using a triaxial accelerometer. *Medicine & Science in Sports & Exercise*, *26*(12), 1516–1523.

Bradley, S. M., Michos, E. D., & Miedema, M. D. (2019). Physical activity, fitness, and cardiovascular health. *JAMA Network Open*, *2*(8). https://doi.org/10.1001/jamanetworkopen.2019.8343

Carson, V., Tremblay, M. S., Chaput, J.-P., & Chastin, S. F. (2016). Associations between sleep duration, sedentary time, physical activity, and health indicators among canadian children and youth using compositional analyses [PMID: 27306435]. *Applied Physiology, Nutrition, and Metabolism*, *41*(6 (Suppl. 3)), https://doi.org/10.1139/apnm-2016-0026, S294–S302. https://doi.org/10.1139/apnm-2016-0026

Chastin, S. F. M., Palarea-Albaladejo, J., Dontje, M. L., & Skelton, D. A. (2015). Combined effects of time spent in physical activity, sedentary behaviors and sleep on obesity

and cardio-metabolic health markers: A novel compositional data analysis approach. *PLOS ONE*, *10*(10), 1–37. https://doi.org/10.1371/journal.pone.0139984

Chen, K. Y., & Bassett, D. R. (2005). The technology of accelerometry-based activity monitors: Current and future. *Medicine & Science in Sports & Exercise*, *37*(Supplement). https://doi.org/10.1249/01.mss.0000185571.49104.82

Di, C., Buchner, D., Liu, S., Kooperberg, C., Lacroix, A., & Prentice, R. (2020). *A functional data analysis framework for continuously modeling intensity using accelerometer-measured physical activity data* [Unpublished Manuscript]. Unpublished Manuscript.

Dishman, R. K., Washburn, R. A., & Schoeller, D. A. (2001). Measurement of physical activity. *Quest*, *53*(3), 295–309. https://doi.org/10.1080/00336297.2001.10491746

Dumuid, D., Stanford, T. E., Pedišić, Ž., Maher, C., Lewis, L. K., Martın-Fernández, J.-A., Katzmarzyk, P. T., Chaput, J.-P., Fogelholm, M., Standage, M., Et al. (2018). Adiposity and the isotemporal substitution of physical activity, sedentary time and sleep among school-aged children: A compositional data analysis approach. *BMC public health*, *18*(1), 311.

Evenson, K. R., Wen, F., Herring, A. H., Di, C., Lamonte, M. J., Tinker, L. F., Lee, I.-M., Rillamas-Sun, E., Lacroix, A. Z., Buchner, D. M., & et al. (2015). Calibrating physical activity intensity for hip-worn accelerometry in women age 60 to 91years: The womens health initiative opach calibration study. *Preventive Medicine Reports*, *2*, 750–756. https://doi.org/10.1016/j.pmedr.2015.08.021

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer. https://books.google.com/books?id=eBSgoAEACAAJ

Heath, G. W. (2019). Light physical activity and incident coronary heart disease and cardiovascular disease among older women—a call for action. *JAMA Network Open*, *2*(3). https://doi.org/10.1001/jamanetworkopen.2019.0405

Hees, V. T. V., Gorzelniak, L., León, E. C. D., Eder, M., Pias, M., Taherian, S., Ekelund, U., Renström, F., Franks, P. W., Horsch, A., & et al. (2013). Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0061691

Janz, K. F. (2006). Physical activity in epidemiology: Moving from questionnaire to objective measurement. *British Journal of Sports Medicine*, *40*(3), 191–192. https://doi.org/10.1136/bjsm.2005.023036

Kohl, H. W. (2001). Physical activity and cardiovascular disease: Evidence for a dose response [PMCID: N/A (precedes mandate)]. *Medicine and Science in Sports and Exercise*, *33*(6; SUPP), S472–S483.

Lacroix, A. Z., Rillamas-Sun, E., Buchner, D., Evenson, K. R., Di, C., Lee, I.-M., Marshall, S., Lamonte, M. J., Hunt, J., Tinker, L. F., & et al. (2017). The objective physical activity and cardiovascular disease health in older women (opach) study. *BMC Public Health*, *17*(1). https://doi.org/10.1186/s12889-017-4065-6

Lyden, KATE and Keadle, SARAH KOZEY and Staudenmayer, JOHN and Freedson, Patty S. (2014). A method to estimate free-living active and sedentary behavior from an accelerometer [PMCID: PMC4527685]. *Medicine and Science in Sports and Exercise*, *46*(2), 386–397.

Mekary, R. A., Willett, W. C., Hu, F. B., & Ding, E. L. (2009). Isotemporal Substitution Paradigm for Physical Activity Epidemiology and Weight Change. *American Journal of Epidemiology*, *170*(4), 519–527. https://doi.org/10.1093/aje/kwp163

Nicklas, B. J. (2018). No expiration date on the association between physical activity and mortality. *Journal of the American Geriatrics Society*, *66*(5), 850–852. https://doi.org/10.1111/jgs.15243

Pedisic, Z. (2014). Measurement issues and poor adjustments for physical activity and sleep undermine sedentary behaviour research—the focus should shift to the balance between sleep, sedentary behaviour, standing and activity. *Kinesiology*, *46*, 135–146.

Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction*. Oxford University Press. https://books.google.com/books?id=UHQoAgAAQBAJ

Piercy, K. L., Troiano, R. P., Ballard, R. M., Carlson, S. A., Fulton, J. E., Galuska, D. A., George, S. M., & Olson, R. D. (2018). The Physical Activity Guidelines for Americans. *JAMA*, *320*(19), 2020–2028. https://doi.org/10.1001/jama.2018.14854

Prentice, R. L., Huang, Y., Kuller, L. H., Tinker, L. F., Van Horn, L., Stefanick, M. L., Sarto, G., Ockene, J., & Johnson, K. C. (2011). Biomarker-calibrated Energy and Protein Consumption and Cardiovascular Disease Risk Among Postmenopausal Women [PMCID: PMC3033986]. *Epidemiology*, *22*(2), 170–179.

Ramsay, J. O., & Silverman, B. W. (2005). Functional Data Analysis. Springer, New York.

Rosenberg, D., Godbole, S., Ellis, K., Di, C., Lacroix, A., Natarajan, L., & Kerr, J. (2017). Classifiers for accelerometer-measured behaviors in older women. *Medicine & Science in Sports & Exercise*, *49*(3), 610–616. https://doi.org/10.1249/mss.0000000000001121

Shephard, R. (2012). 2011 compendium of physical activities: A second update of codes and met values. *Yearbook of Sports Medicine*, *2012*, 126–127. https://doi.org/10.1016/j.yspm.2011.08.057

Stamatakis, E., Rogers, K., Ding, D., Berrigan, D., Chau, J. Y., Hamer, M., & Bauman, A. (2015). All-cause mortality effects of replacing sedentary time with physical activity and sleeping using an isotemporal substitution model: a prospective study of 201,129 mid-aged and older adults. https://repository.lboro.ac.uk/articles/journal_contribution/All-cause_mortality_effects_of_replacing_sedentary_time_with_physical_activity_and_sleeping_using_an_isotemporal_substitution_model_a_prospective_study_of_201_129_mid-aged_and_older_adults/9629624

Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T., McDowell, M., Et al. (2008). Physical activity in the united states measured by accelerometer. *Medicine and science in sports and exercise*, *40*(1), 181.

US Dept of Health and Human Services. (2018). Physical activity guidelines for americans. 2nd ed. US Dept of Health; Human Services Washington, DC.

Voorrips, L. E., Ravelli, A., Petra, C., Dongelmans, A., Deurenberg, P., & van Staveren, W. A. (1991). A physical activity questionnaire for the elderly. *Diet and physical activity as determinants of nutritional status in elderly women*, 43.

Ward, D. S., Evenson, K. R., Vaughn, A., Rodgers, A. B., & Troiano, R. P. (2005). Accelerometer use in physical activity: Best practices and research recommendations. *Medicine and science in sports and exercise*, *37*(11 Suppl), S582–8.

Wood, S. N. (2003). Thin-plate regression splines. *Journal of the Royal Statistical Society (B)*, *65*(1), 95–114.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, *99*(467), 673–686.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, *73*(1), 3–36.

Wood, S. (2006). *Generalized additive models: An introduction with r*. Chapman; Hall/CRC.