Functional Data Analysis Methods for Analyzing Accelerometry Data in Mobile Health

by

Margaret M. Banker

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in the University of Michigan
2023

Doctoral Committee:

Professor Peter X.K. Song, Chair
Assistant Professor Walter Dempsey
Professor Jian Kang
Professor Karen E. Peterson

Margaret M. Banker
mbanker@umich.edu
ORCID iD: 0009-0000-8653-5118

DEDICATION

To my friends and family, the architects of my being. You have been an unwavering source of support, inspiration, and joy throughout this journey.
Thank you for believing in me.

ACKNOWLEDGEMENTS

"Gratitude is the memory of the heart"

First, I thank my committee and co-advisors for their guidance and support throughout this journey. A special acknowledgment goes to my Dissertation Chair and Thesis Advisor, Dr. Peter X.K. Song; I am profoundly grateful for your unwavering support, expert guidance in statistical knowledge, and seemingly indefatigable energy. Your mentorship has been indispensable in shaping my academic journey. I extend heartfelt thanks to Dr. Karen Peterson, my unofficial "GSRA Advisor" and mentor in collaborative research. In addition to providing insights into the scientific side of my research, you warmly welcomed me to the ELEMENT research group and fostered an exceptional collaborative research environment. You have taught me how to be an engaged leader and an enthusiastic researcher, and have been an invaluable source of scientific and personal growth. Special thanks to my additional committee members: Dr. Jian Kang, your insightful comments and methodological guidance have been invaluable, and Dr. Walter Dempsey, your expertise in wearable device methodologies has been instrumental.

I would like to extend my gratitude to the numerous researchers, mentors, professors, and staff who have played a pivotal role in shaping my academic journey. The DMMC and ELEMENT groups have been paramount in my growth as a collaborative statistician, particularly Dr. Karen Peterson, Dr. Erica Jansen, Dr. Jackie Goodrich, Dr. Jen LaBarre Meijer, and Laura Arboleda Merino, who patiently explained scientific concepts and helped me make clinical sense of statistical results. Thank you to the Professors within the University of Michigan Department of Biostatistics for cultivating an environment of scholarship, mentorship, and leadership that greatly contributed to my development. Thank you to the administration staff in the Department of Biostatistics office for all the incredible support. A special thank you to Nicole Fenech, who has fielded my many questions from Day 1, and always had time for a friendly chat. And to Dan Barker, the computing guru - my simulations never stood a chance without you. Thanks for your patience and friendship.

My friends and family have also provided invaluable support over the years and this dissertation could not have been completed with them. My first biostat buddies, Emily

Morris and Emma Enache - I am so grateful to have had you by my side for homework nights and study sessions. To the friends I made in my final PhD years, Holly Hartman and Nicky Wakim, your joyful energies made the journey much more enjoyable, and helped me rediscover joy in research. My Ann Arbor friends, Martha Siegmund, Sarah Watkins, Cathy Smith, Erin Loomas - thanks for all the laughs and adventures! To my long-distance friends, Bridget Lynch, Eliza Fradkin, Heather Petroccia, and Jacqui Bloch - the FaceTimes and group texts and visits are a constant source of happiness in my life.

Lastly, my family. Thank you to my parents, Mike and Mary Ellen Banker, for always supporting me and providing endless emotional support on this graduate school journey. You have helped me to develop the grit and perseverance necessary to succeed in all things. Thank you to my siblings Phil, Rachel, Kate, Frank, and Sarah, who have always been available for last-minute stress-calls and de-stressing adventures. Thank you to my godmother Ruth Olech, who, through a lifetime of example, has taught me to say ”yes!” to hard things. To my Ann Arbor Olechs, Steve, Laura, Claire, Ben, and Wes, who have made Ann Arbor a home these past eight years - thank you for welcoming into your hearts (and around your dinner table). I am truly fortunate to have such a wonderful family by my side.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF FIGURES viii

LIST OF TABLES xi

LIST OF APPENDICES xiii

LIST OF ACRONYMS xiv

ABSTRACT xv

CHAPTER

1 Introduction 1

1.1 Motivation 1
1.1.1 Physical Activity and Health 1
1.1.2 Mobile Health and Wearable Devices 2
1.1.3 Functional Data Analysis 7
1.2 Contribution 8
1.3 Organization of this Dissertation 11

2 Supervised Learning of Physical Activity Features from Functional Accelerometer Data 12

2.1 Introduction 12
2.2 Motivating Study Data 16
2.2.1 ELEMENT Actigraphy Data 16
2.2.2 Epigenetic Age 17
2.3 Occupation Time Curves: A Functional Predictor 18
2.4 Method 20
2.4.1 Scalar-on-Function Regression and Changepoint Detection 20
2.4.2 FRACT Methodology 22
2.5 FRACT Implementation 25
2.5.1 Initial Clustering 25
2.5.2 Tuning K 25

2.5.3 Refinement Learning of Cutpoints 26

2.6 Simulation Experiments 28

2.6.1 Simulation Setup 28

2.6.2 Simulation Performance 30

2.7 Data Analysis 34

2.8 Discussion and Conclusion 37

3 Regularized One-Step Estimation of Changepoint and Functional Parameter in Functional Accelerometer Data Analysis 40

3.1 Introduction 40

3.1.1 Biological Aging and Epigenetic Age 40

3.1.2 Wearable Devices and Accelerometer Data 41

3.1.3 Study Objectives 42

3.2 Motivating Cohort Study 43

3.2.1 Occupation Time Curves (OTCs) 43

3.3 Model Formulations 44

3.3.1 Scalar-on-Function Analysis with Occupation Time Curve (OTC) Functional Predictor 44

3.3.2 Existing L_1 Regularization Approaches 46

3.3.3 Integer Programming and L_0 Penalization 47

3.4 Mixed Integer Optimization 48

3.4.1 Proposed Fusion-Adapted MIO Formulation 49

3.4.2 MIO Implementation 50

3.5 Theoretical Guarantees 50

3.6 Simulation Experiments 52

3.6.1 Simulation Setup 52

3.6.2 Simulation Results 53

3.6.3 Computation Time 54

3.7 Data Analysis 56

3.8 Discussion 60

4 Supervised Fusion Learning of Physical Activity Features from Longitudinal Functional Accelerometer Data 62

4.1 Introduction 62

4.1.1 Longitudinal Functional Data 62

4.1.2 Motivating Longitudinal Cohort Data 63

4.1.3 Longitudinal Occupation Time Curves 64

4.2 Longitudinal Functional Data Analysis Model 66

4.2.1 Longitudinal Scalar-on-Function Model 66

4.3 Constrained Quasi-Estimation 68

4.3.1 Quadratic Inference Function (QIF) 68

4.3.2 Mixed Integer Optimization Formulation (MIO) for QIF 70

4.3.3 Implementation of MIO QIF model 71

4.4 Theoretical Guarantees 74

4.5 Simulation Experiments 76

4.5.1 Simulation Setup 76
4.5.2 Simulation Implementation 78
4.5.3 Simulation Performance 78
4.5.4 Computational Requirements 79
4.6 Data Analysis 85
4.6.1 Full Data Analysis 86
4.6.2 Sex-Stratified Data Analysis 89
4.7 Discussion 93

5 Summary and Future Work 96

5.1 Summary 96
5.2 Future Work 98

APPENDICES 101

BIBLIOGRAPHY 105

LIST OF FIGURES

FIGURE

1.1 24 hours of accelerometer data from a tri-axial wrist-worn device, collected from a single subject from the motivating data set. Each panel deptics the time-series data from a different axis, or direction of acceleration, with Vector Magnitude representing a summary of the three-dimensional axis data. Each time-series demonstrates slightly different activity patterns, illustrating that different activity information is captured by the unique measurements. 4

1.2 24 hours of accelerometer data summarized by 1-minute epoch of Vector Magnitude counts. This data was collected from a wrist-worn accelerometer device adorned by an adolescent from the motivating data set. The horizontal lines represent two different sets of previously published cutoffs, both of which have been validated for wrist-worn accelerometer devices for children. The blue horizontal lines depict the Chandler cutoffs [Chandler et al., 2016], while the red horizontal lines depict the Crouter cutoffs [Crouter et al., 2015]. Each set of activity cutoffs classify the data into 4 activity categories: Sedentary, Light, Moderate, and Vigorous. 5

1.3 24 hours of accelerometer data summarized by 1-minute epoch of Axis 1 counts. This data was collected from a wrist-worn accelerometer device adorned by an adolescent from the motivating data set. The horizontal lines represent two different sets of previously published cutoffs validated against a specific population, neither of which align with the motivating dataset. The blue horizontal lines depict the Freedson Child [Freedson et al., 2005] and the red horizontal lines depict the Troiano Adult cutoffs [Troiano et al., 2008], both of which were validated with Axis 1 Counts on waist-worn devices. Each set of activity cutoffs classify the data into 4 activity categories: Sedentary, Light, Moderate, and Vigorous. . 6

2.1 Accelerometer Data for an individual over a 24 hour period. The horizontal lines indicate PA categorization cutoffs based on Chandler Vector Magnitude cutoffs for 1-minute epochs [Chandler et al., 2016]. These cutoffs are pre-determined by in-lab supervised research, and now applied to a free-living subject. 14

2.2 (a) The construction of an OTC from a time-series of VM counts. The left panel represents accelerometer data, with an increasing bar of count cutoff indicated by the horizontal lines with varying point-shapes. The corresponding proportion of time spent at or above that level of activity is shown in the indicated point on the right panel with the corresponding point shape. The grey continuous curve is the realized OTC for this individual (b) Comparison of OTC shapes for More vs Less Active Individuals, with VM count summarized over 1-min epochs varying over 0 to 30000. The distinctive shapes of the curves represent the subject's activity pattern. For example,curves of less active people steeply drop in the beginning, signifying that a small percentage of their time is spent in even mid-active regions. 19

2.3 Comparison of initial estimates obtained from a randomly selected simulation dataset of $\beta_j,\ j = 1, \cdots, J$ with/without the MCP Penalty and with/without augmentation by 5 intervals (J=300, 60 respectively). The true $\beta(c)$ is shown as the piece-wise horizontal function, whose visually flat pattern in (a) is due to the scale of the beta values. 31

2.4 OTCs for 354 ELEMENT subjects, representing functional covariates. The vertical lines represent Chandler's cutoffs (Chandler et al., 2016) for activity levels (Sedentary, Light, Moderate, Vigorous). The relative shape of these OTCs reflect the subject's activity profile. 34

3.1 Left panel: Two realized OTCs with vertical dashed lines representing the cutoffs for activity window ranges of interest. Right panel: Example of a non-constant functional $\beta-$ parameter as a step-function to estimate the association of Physical Activity (PA) in selected activity-count windows with a health outcome of interest 45

3.2 Estimates of θ_j coefficients from two different standard analysis approaches including multiple linear regression (left) and an L_1 Regularization approach of Fused Lasso (right). The respective model parameter estimates are represented by the black circles, while the true θ_j values are represented as the red step function. 47

3.3 OTCs for 354 ELEMENT subjects stratified by boys and girls. The vertical lines represent Chandler's cutoffs [Chandler et al., 2016] for prefixed activity levels (Sedentary, Light, Moderate, Vigorous). The relative shape of OTC reflects the subject's activity profile. 56

4.1 OTCs for three individuals from the motivating ELEMENT cohort at both Time point 1 (T1) and Time point 2 (T2), collected approximately two years apart. The OTC curves from both T1 and T2 represent PA during the weekend between 4:00PM-10:00PM. Subjects A and B demonstrate similar PA patterns across the two time points, whereas Subject C demonstrates a changing PA profile from T1 to T2. 65

4.2 OTCs for 429 ELEMENT subjects stratified by longitudinal time point of accelerometer data collection. The relative shape of OTC reflects the subject's activity profile, with the red-dotted line representing the mean proportion of time spent above each activity level across the individuals. The first vertical line, colored blue, represents the first detected cutpoint, indicating the changepoint between critical activity Windows 1 and 2. Likewise, the second vertical line in green represents the detected cutpoint 2, signalling the changepoint from critical activity Windows 2 and 3. 87

4.3 OTCs for 197 male ELEMENT subjects stratified by longitudinal time point of accelerometer data collection. The relative shape of OTC reflects the subject's activity profile, with the red-dotted line representing the mean proportion of time spent above each activity level across the individuals. The first vertical line, colored blue, represents the first detected cutpoint, indicating the changepoint between critical activity Windows 1 and 2. Likewise, the second vertical line in green represents the detected cutpoint 2, signalling the changepoint from critical activity Windows 2 and 3. 90

4.4 OTCs for 232 female ELEMENT subjects stratified by longitudinal time point of accelerometer data collection. The relative shape of OTC reflects the subject's activity profile, with the red-dotted line representing the mean proportion of time spent above each activity level across the individuals. The vertical blue, represents the first detected cutpoint, indicating the changepoint between critical activity Windows 1 and 2. 92

B.1 An illustration of the AUC Ratio calculation for Window 3. The shaded green regions represent the AUC for subject i at time t for $t = c(1,2)$ within the detected window, or A_{i3}. The outlined rectangle represents the area of the full rectangle created by the cutpoints ($c_2 = 145, c_3 = 300$) of Window 3, deemed R_3, which represents the hypothetical subject spending 100% of time at or above this activity level. The AUC Ratio for this first window is then calculated by AUC Ratio$_{i3} = \frac{A_{i3}}{R_3}$. The interpretation of this ratio depends on the location of the window. For the K^{th} window, the interpretation of the AUC Ratio$_{iK}$ represents the percent of time the individual spends within that window compared to the hypothetically most active person. In this case, a higher AUC Ratio$_{i3}$ value represents higher PA within the window. We can see that the AUC Ratio$_{i3}$ at T1 is lower than the AUC Ratio$_{i3}$at T2, suggesting a higher outcome at T2 versus T1. In reality, we see an SSST outcome of 12.0mm at T1, increasing to 12.5mm at T2. 104

LIST OF TABLES

TABLE

2.1 Simulation Results of 3-Group Model to Evaluate FRACT 32

2.2 Sensitivity Simulation Results for 3-Group Model 33

2.3 FRACT Data Analysis Results 36

3.1 Simulation Results of a 3-group Model with $N = 500$ and number of intervals $J = 60, 300$ demonstrate the performance of Fused Lasso, summarized over 500 replicates. Results include average estimate (Mean), median estimate (Med.), and empirical standard error (ESE). Cutpoint values are represented as VM/100. 48

3.2 Simulation Results of the 3-group model with number of micro-intervals $J = 300$ and sample size of $N = 500$ summarized over 500 replicates, including average estimate (L_0 Mean), empirical standard error (L_0 ESE), and average estimate from an L_1 Fused Lasso analysis (FL Mean) using R package `glasso`. Cutpoint values are represented as VM/100. 53

3.3 Simulation Results of the 3-group model with number of micro-intervals $J = 300$ and sample size of $N \in \{500, 250, 100\}$ summarized over 500 replicates, including average estimate (L_0 Mean), empirical standard error (L_0 ESE), and average estimate from an L_1 Fused Lasso analysis (FL Mean) using R package `glasso`. Cutpoint values are represented as VM/100. Sensitivity for selecting 3-group model based on goodness-of-fit comparisons was greater than 99% in all scenarios. 55

3.4 Data Analysis Results obtained by the fusion-adapted L_0 method, where J indicates the number of micro-intervals and K is a prefixed number of activity windows. Significance is measures in two different ways; 'p-val' represents the p-values from linear regression, whereas 'perm' represents empirical p-value when assessing 'p-val' to the distribution of p-values attained through 1000 permutations. Cutpoint values are represented as VM/100. 58

4.1 Step 1 MIO Results for J=300 80

4.2 Step 1 MIO Results for J=60 81

4.3 Step 2 QIF Results for J=300 82

4.4 Step 2 QIF Results for $J = 60$ with longitudinal correlation parameter $\rho = 0.5$. 83

4.5 Step 2 QIF Results for $J = 60$ with longitudinal correlation parameter $\rho = 0.2$. 84

4.6 Full Population Data Analysis Results obtained by the fusion-adapted L_0 QIF method, where J indicates the number of micro-intervals and K is a prefixed number of activity windows. Significance is indicated by 'p-val', representing the p-values from the final Quadratic Inference Function (QIF) model. Cutpoint values are represented as VM/100. 88

4.7 Sex-Stratified Data Analysis Results for Male subjects obtained by the fusion-adapted L_0 QIF method, where J indicates the number of micro-intervals and K is a prefixed number of activity windows. Significance is indicated by 'p-val', representing the p-values from final QIF model. Cutpoint values are represented as VM/100. 91

4.8 Sex-Stratified Data Analysis Results for Girls obtained by the fusion-adapted L_0 QIF method, where J indicates the number of micro-intervals and K is a prefixed number of activity windows. Significance is indicated by 'p-val', representing the p-values from final QIF model. Cutpoint values are represented as VM/100. . . 93

A.1 Simulation Results of 4-group Model with number of original intervals $J = 300$ to evaluate the adapted L_0-fusion analytic under a environment with severe multicollinearity. Results are summarized over 500 replicates and include average estimate (L_0 Mean), empirical standard error (L_0 ESE), and Mean from a standard Fused Lasso analysis (FL Mean) for a sample size of $N = 500$. Cutpoint values are represented as VM/100 for ease of visualization. 101

A.2 Simulation Results of 4-group Model with number of intervals $J = 60$ to evaluate the adapted L_0-fusion analytic, summarized over 500 replicates. Results include average estimate (L_0 Mean), empirical standard error (L_0 ESE), and Mean from a standard Fused Lasso analysis (FL Mean) as comparison. Cutpoint values are represented as VM/100 for ease of visualization. Sensitivity for selecting 3-group model based on goodness-of-fit comparisons was greater than 99% in all scenarios.102

LIST OF APPENDICES

A Chapter 3 Appendix 101

B Chapter 4 Appendix 103

LIST OF ACRONYMS

AUC Area Under the Curve

CGM Continuous Glucose Monitoring

EBIC Extended BIC

ELEMENT Early Life Exposure in Mexico to ENvironmental Toxicants

FCPA Functional Principal Component Analysis

FDA Functional Data Analysis

FRACT Functional Regularized Adaptive Changepoint-detection Technique

GEE Generalized Estimating Equations

HMM Hidden Markov Model

MCP Miniax Convex Penalty

MIO Mixed Integer Optimization

MIP Mixed Integer Programming

MIQO Mixed Integer Quadratic Optimization

MSE Mean Squared Error

MVPA Moderate-to-Vigorous Activity

OTC Occupation Time Curve

PA Physical Activity

QIF Quadratic Inference Function

SFL Supervised Fusion Learning

SSST Subscapular Skin Thickness

SVD Single Value Decomposition

VM Vector Magnitude

ABSTRACT

Accelerometry data collected by high-capacity sensors present a primary data type in smart mobile health. Such data enable scientists to extract personal digital features that are useful for precision health decision making. Existing methods in accelerometry data analysis typically begin with discretizing summary single-axis counts by certain fixed cutoffs into several activity categories, such as Vigorous, Moderate, Light, and Sedentary. One well-known limitation is that the chosen cutoffs have often been validated under restricted settings, and thus they cannot be generalizable across populations, devices, or studies. Motivated by the Early Life Exposure in Mexico to ENvironmental Toxicants (ELEMENT) research cohort, in this dissertation I develop data-driven approaches to overcome this bottleneck in the analysis of physical activity data.

In Chapter 2, I propose to holistically summarize an individual subject's activity profile using Occupation Time curves (OTCs). Being a functional predictor, OTCs describe the percentage of time spent at or above a continuum of activity count levels. The resulting functional curve is informative to capture time-course individual variability of physical activities. I develop a multi-step adaptive learning algorithm, termed FRACT (Functional Regularized Adaptive Changepoint-detection Technique), to perform supervised learning via scalar-on-function regression modeling that involves OTC as the functional predictor of interest as well as other scalar covariates. This learning analytic first incorporates a hybrid approach of fused lasso for clustering and Hidden Markov Model for change-point detection, and then executes a few refinement procedures to determine activity windows of interest. Through extensive simulation experiments I show the proposed FRACT performs well in both changepoint detection and regression coefficient estimation. In application of this method on real world data, I analyze 354 adolescent subjects from the ELEMENT cohort to assess the influence of physical activity on two different biological aging outcomes. I find that the different biological aging outcomes are each associated with different activity window of interest, demonstrating the flexibility of the method to determine data-driven associations based both on the underlying functional variables of interest, as well as the specific health outcomes.

In Chapter 3, I investigate functional analytics under an L_0 regularization approach that

enables the handling of highly correlated micro-activity windows that serve as predictors in the scalar-on-function regression model proposed in Chapter 2. Relatively recent advances in L_0 regularization and discrete optimization have promoted this powerful optimization paradigm making it computationally viable. Utilizing such recent algorithmic and numeric capabilities, I develop a new one-step method that can simultaneously conduct fusion via change-point detection and parameter estimation through a new L_0 constraint formulation. This new approach is not only computationally efficient but also avoids propagation of subjective errors incurred in a multi-stage analytic. I implement a new algorithm via GUROBI, a modern optimization solver that provides a fast one-stage analytic for both parameter fusion and changepoint detection. I evaluate and illustrate the performance of the proposed learning analytics through simulation experiments and a reanalysis of the relationship between physical activity and biological aging.

In Chapter 4, I extend the previous L_0 regularization framework of Chapter 3 to a longitudinal functional framework with repeated wearable data to understand the influence of serially measured functional accelerometer data on longitudinal health outcomes. The statistical methodological extension invokes the means of Quadratic Inference Functions (QIF), with an aim to detect physical activity intensity windows and assess their population-average effects on children health outcomes. I consider a population-average effects model, and develop a regularized QIF via mixed integer optimization to carry out longitudinal data analysis. In contrast to the previous chapters, which considered the physical activity data during a seven-day period, with the repeated measurements taken approximately two years after the first, I focus on a longitudinal study of physical activity patterns from late-adolescence into early adulthood on sub-scapular skin thickness (SSST). SSST is a measure of truncal fat distribution; changes in SSST diverge dramatically in boys and girls as they undergo puberty. SSST is among the measures of body composition that can be influenced by PA behaviors, which decline and vary in adolescents. To our knowledge, this is the first study to consider a longitudinal functional measure of PA in relation to changes in SSST in male and female adolescents.

CHAPTER 1

Introduction

1.1 Motivation

Physical Activity (PA) plays a crucial role in promoting overall health and well-being [Soares-Miranda et al., 2016, Shook et al., 2015, Hallal et al., 2006, Wu et al., 2020, Aljahdali et al., 2022]. In the modern era of mobile health, advancements in technology have allowed us to monitor PA more comprehensively and effortlessly [Trost, 2001, Troiano et al., 2014, Liu et al., 2021]. Advances in wearable devices have revolutionized the way we monitor and quantify PA in individuals. These devices, such as fitness trackers and smartwatches, collect vast amounts of data on movement patterns and energy expenditure [Freedson et al., 2005, Crouter et al., 2015, Chandler et al., 2016]. In recent years, Functional Data Analysis (FDA) techniques have emerged as a powerful tool to analyze and interpret these data streams. By treating PA data as functional data, researchers can explore the dynamics and patterns of movement over time, gaining insights into activity profiles and fluctuations [Ramsay, 2004, Goldsmith et al., 2012]. This deeper understanding of activity patterns can help identify optimal exercise regimes, track changes in health-related behaviors, and detect early signs of health issues. Integrating PA, mobile health, and functional data analysis opens up new avenues for promoting healthier lifestyles, facilitating personalized interventions, and advancing our understanding of the complex relationship between PA and health. Here we introduce and explore these concepts in greater detail.

1.1.1 Physical Activity and Health

Physical Activity has an undeniable impact on overall health; numerous studies [Soares-Miranda et al., 2016, Shook et al., 2015, Hallal et al., 2006, Wu et al., 2020, Aljahdali et al., 2022] have pointed to the strong and varied relationships between the two. Among the myriad of findings, researchers have shown that increases in physical inactivity and sedentary behavior increases the risk of cardiovascular disease and strokes in older adults

[Soares-Miranda et al., 2016], as well as the risks of over 25 chronic conditions and overall mortality [Warburton and Bredin, 2016]. In a similar vein, additional research has suggested that low levels of PA are a risk factor for fat mass gain in young adults [Shook et al., 2015], and that increased PA in adolescence has both short- and long-term benefits for physical and mental health [Hallal et al., 2006]. Thus overall, increased PA has shown to reduce the risk of diseases, promote healthy weight management, and improve mental health.

While researchers agree that PA has an important impact on health, important questions remain regarding the importance of specific intensity, types, and bouts of activity [Warburton and Bredin, 2016]. Some studies suggest that the cumulative effect of low-to-moderate activity levels have beneficial health impacts, while others promote shorter bouts of high-intensity activity [Jakicic et al., 2019]. Another field of research investigates whether the timing PA is important, assessing if the health benefits of PA differ based on when during the day it was performed [Janssen et al., 2022]. Previously, the field of PA and health research has relied on self-reported data; however, the more recent rise in popularity and reliability of wearable devices are providing new opportunities to explore these important questions.

1.1.2 Mobile Health and Wearable Devices

Wearable technologies are devices worn over continuous time-periods, which are designed to collect subjects' personal data. Notably, these devices can conduct automatic data collection in high frequency and track physiological variables and clinical symptoms outside of clinical environments. In providing this high-frequency, personalized time-series data, wearable devices are promising technologies to promote smart-Health care management and precision medicine. Additionally, the data collection can be relatively cheap, convenient, and flexible in variable environments, which increases their popularity in both research and personal use.

The increased use of wearable devices in both research and personal use has given rise to objective large-scale time-series datasets, providing unique opportunities and challenges in data analytics. Now, researchers have the opportunity to collect repeated measurements of functional digital features from wearable devices, and must determine how best to leverage such data to answer scientific questions of interest. While their popularity and potential usefulness are growing quickly, the ability to efficiently and effectively glean statistically-robust information from wearable devices is slower to catch up. The data retrieved from these technologies present challenges in data analysis, due to their inherent noisy nature, the non-generalizability of methodologies, and high computational requirements. These challenges motivate the need for statistical innovations in data analysis, machine learning, and data science, to enable the wide-scale use of such wearable smart-Health devices in research amd clinical practice.

Accelerometers are a type of sensor available in wearable devices that captures continuous PA and movement data, providing real-time, large-scale, personalized information on an individual's PA patterns. Accelerometer sensors measure how speed changes over time through electrical signals representing the volume and intensity of movement. Such data are recorded in high resolutions of sampling frequencies (Hz), and then processed via proprietary software, such as ActiLife LLC. With the ability to provide high-frequency measurements (example frequency: 60Hz, or 60 data points per second), the technical challenge becomes how to retrieve useful information from this high-frequency time-series data type. For example, in a tri-axial accelerometer, data is collected along three orthogonal axes, deemed Axis 1, Axis 2, and Axis 3. Each axis provides different information on PA, though the measurements are correlated. Typically, the processed data describing PA levels at each axis are expressed as activity "counts" over specific periods of time known as "epochs" [Chen and Bassett, 2005]. While some researchers [Naiman and Song, 2022] have recently proposed methods to analyze the three-dimensional time-series counts directly, many researchers often summarize the three-dimensional count information at each time-point into a one-dimensional summary value of Vector Magnitude (VM), with $VM = \sqrt{axis1^2 + axis2^2 + axis3^2}$ [Liu et al., 2021].

Figure 1.1 depicts 24 hours of tri-axial accelerometer data from a single individual, with the collected time-series data from each of the three different axes, as well as the summarized VM. We can see that, while each time-series follows similar macro patterns, the different measurements provide different micro patterns and magnitudes over time. From such count level data, researchers need to decide not only which measurement to analyze, but also the means of summarizing the high-frequency time-series data into a feasible summarized format for analysis. The latter is deemed for dimension reduction in order to apply existing data analytics to the analysis.

Previous research has emphasized easy-to-understand summary statistics in order to obtain dimension reduction and reduce data complexity. These measurement include classification of raw accelerometry into PA intensity levels based on pre-defined cutpoints, as well as daily/weekly activity count averages or totals. In the case of the latter, the summarizing suffers from a significant loss of information and the deterioration of the functional nature of the data [Lin et al., 2023]. In the case of the former, a researcher must rely on pre-determined categorization thresholds. Specifically, a popular method of analyzing accelerometer data involves specifying activity level "cutoffs" to discretize activity counts into categories, such as Sedentary, Light, and Moderate-to-Vigorous [Freedson et al., 2005, Crouter et al., 2015, Chandler et al., 2016, Troiano et al., 2008]. While categorization is a useful approach, the cutoffs must be pre-specified by the researcher. There are many potential cutoffs published

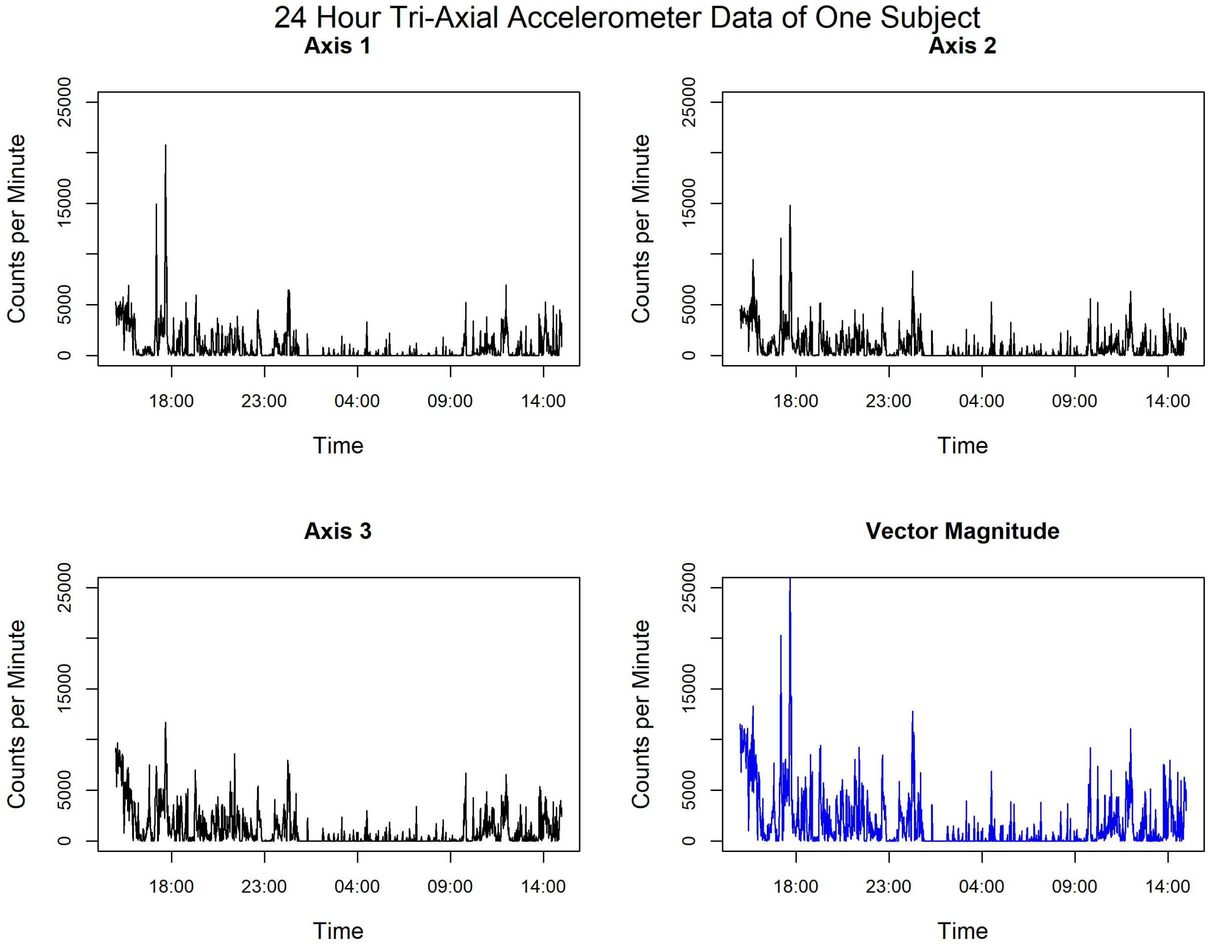


Figure 1.1: 24 hours of accelerometer data from a tri-axial wrist-worn device, collected from a single subject from the motivating data set. Each panel deptics the time-series data from a different axis, or direction of acceleration, with Vector Magnitude representing a summary of the three-dimensional axis data. Each time-series demonstrates slightly different activity patterns, illustrating that different activity information is captured by the unique measurements.

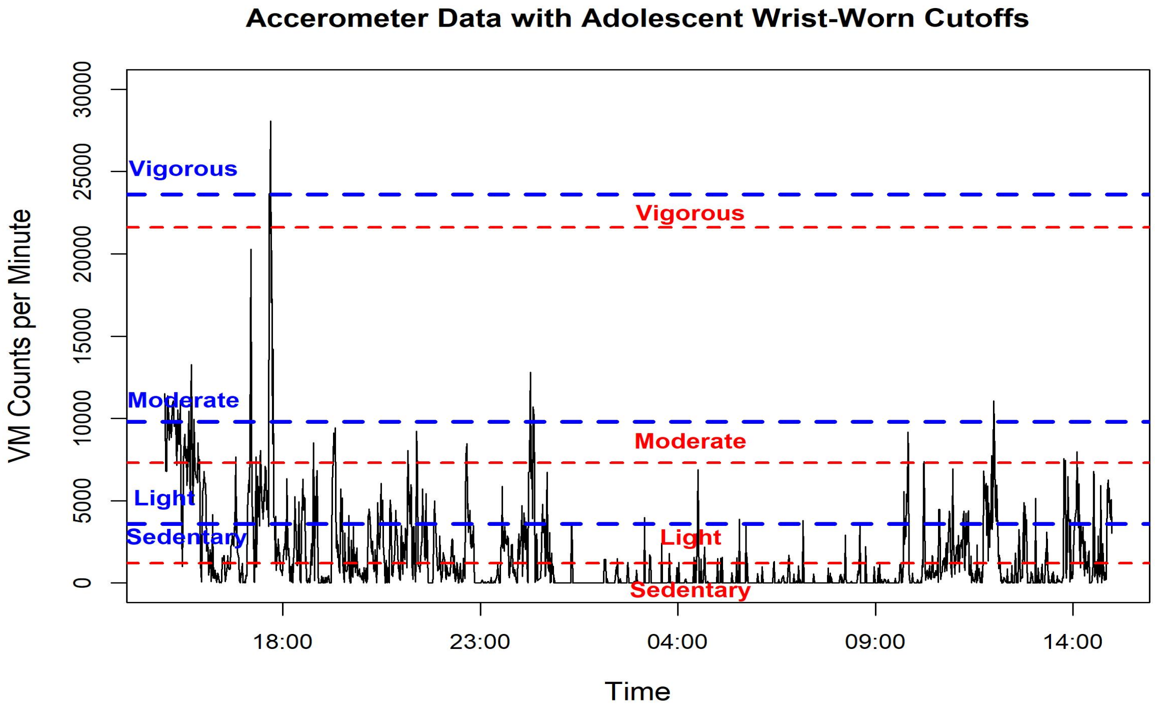


Figure 1.2: 24 hours of accelerometer data summarized by 1-minute epoch of Vector Magnitude counts. This data was collected from a wrist-worn accelerometer device adorned by an adolescent from the motivating data set. The horizontal lines represent two different sets of previously published cutoffs, both of which have been validated for wrist-worn accelerometer devices for children. The blue horizontal lines depict the Chandler cutoffs [Chandler et al., 2016], while the red horizontal lines depict the Crouter cutoffs [Crouter et al., 2015]. Each set of activity cutoffs classify the data into 4 activity categories: Sedentary, Light, Moderate, and Vigorous.

in the literature, each validated against different narrowly focused studies with small subgroup populations. These cutoffs may be affected by many personal-level variables, including what device was used (e.g. Actigraph GT3X, Actigraph GT9X, Fitbit), the placement of the device (e.g. hip, wrist, ankle), or characteristics of the study population (e.g. age, sex, race) [Freedson et al., 2005].

Figures 1.2 and 1.3 depict 24 hours of accelerometry data with different set of cutoff thresholds applied. While each set of cutoffs classifies the data into the four physical activity categories of Sedentary, Light, Moderate, and Vigorous activity, the sets of cutoffs do not always agree on the specific classification at each time epoch. Figure 1.2 demonstrates cutpoint thresholds from the Chandler [Chandler et al., 2016] and Crouter [Crouter et al., 2015] thresholds, which were validated against a similar underlying population as our moti-

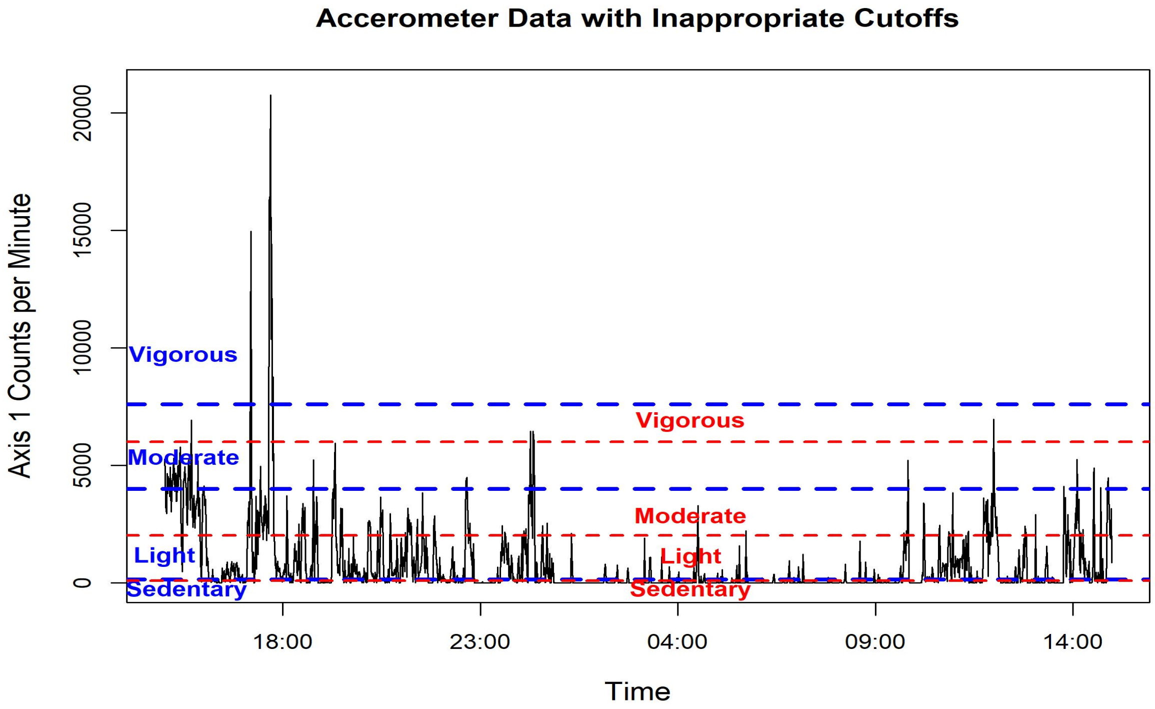


Figure 1.3: 24 hours of accelerometer data summarized by 1-minute epoch of Axis 1 counts. This data was collected from a wrist-worn accelerometer device adorned by an adolescent from the motivating data set. The horizontal lines represent two different sets of previously published cutoffs validated against a specific population, neither of which align with the motivating dataset. The blue horizontal lines depict the Freedson Child [Freedson et al., 2005] and the red horizontal lines depict the Troiano Adult cutoffs [Troiano et al., 2008], both of which were validated with Axis 1 Counts on waist-worn devices. Each set of activity cutoffs classify the data into 4 activity categories: Sedentary, Light, Moderate, and Vigorous.

vating dataset; that is, the Chandler and Crouter cutpoints were created for accelerometer data summarized into one-minute epoch Vector Magnitude counts collected from adolescent subjects with wrist-worn tri-axial accelerometer devices. Even in this “ideal” case, in which the motivating data set aligns with the cutoff validation data set, it is clear that the choice between the Chandler and Crouter cutoffs impacts PA classification and subsequent data analyses. Figure 1.3, where the Freedson Child [Freedson et al., 2005] and Troiano Adult [Troiano et al., 2008] cutoffs are depicted, demonstrates the more problematic scenario in which the applied cutoffs were validated with underlying characteristics that do not align with the motivating dataset. These cutoffs were validated for Axis 1 accelerometer counts collected from waist-worn devices, and clearly do not provide appropriate distinction between the time-series accelerometer data. In summary, a clear bottleneck in the analysis of accelerometer data pertains to the utility of prefixed cutoffs that may produce misleading results.

Given the multitude benefits of PA on health as described in Section 1.1.1, and the new wealth of data researchers are now able to collect from wearable devices, it is becoming increasingly vital to further develop novel statistical models that are both robust and informative. These models can be leveraged to explore the relationship between varying aspects of PA with health outcomes of interest. Motivated by the Early Life Exposure in Mexico to ENvironmental Toxicants (ELEMENT) research cohort, this dissertaion plans to develop data-driven approaches within the functional data framework to over-come the bottleneck in the analysis of PA data from wearable devices. While this dissertation focuses on PA collected from wearable accelerometer devices, the proposed methodologies can be applied to alternative types of wearable devices that collect high-frequency time-series data, such as continuous glucose monitoring (CGM) devices, heart rate monitors, and toxicant sensors, among others.

1.1.3 Functional Data Analysis

Functional data analysis (FDA) is a statistical analysis framework that addresses the challenges of analyzing functional data, which itself consists of a set of curves, surfaces, or other types of data observed over a continuum, such as time or space [Ramsay, 2004]. By analyzing functional data collected from wearable devices, researchers can gain valuable insights into how an individual's physiological patterns affect health outcomes, monitor health conditions, track changes over time, and make informed decisions regarding lifestyle, fitness, and healthcare management. At a high level, characteristics of functional data include being (i) high-dimensional (ii) temporal or structural in nature (iii) recorded over a continuous domain [Ramsay, 2005]. Compared to traditional statistics, where data is typically represented as a set of discrete observations, functional data considers the function as a whole as the primary unit of analysis. In our case, the PA serves as a functional predictor in the analysis. Considering the data in this way enables the extraction of valuable information regarding the overall shape, trends, and patterns present in the data [Chen and Müller, 2012]. As a relatively new and growing field, comprehensive reviews of Functional Data and their uses are provided by, Ramsay & Silverman (2005) [ram, 2005], Ferraty & Vieu (2006) [fer, 2006], and Horvath & Kokoszka (2012) [Horváth and Kokoszka, 2012], among others.

As discussed above, wearable devices often include sensors that can capture various physiological signals continuously over time, so the collected data are good candidates for Functional Data Analysis (FDA); in particular, they are used as functional predictors in this dissertation. The signals captured from the devices can be represented as functions that describe the changes in physiological measurements over a continuous domain, such as time. That is, each observation represents a function that describes the values of a variable at different time points. An example of functional data from a wearable device is that of Continuous Glucose Monitoring (CGM). The CGM device records glucose readings at regular intervals, capturing an individual's glucose levels and fluctuations throughout the day. Each observation in this hypothetical dataset would consist of a function that represents the glucose levels over the course of a single day. That is, the observation for individual i over continuous time t for $t \in (0, T)$, represented as $G_i(t)$, could be sampled at discrete time points, such that: $G_i(t) = (g_{i1}, g_{i2}, \ldots, g_{it}, \ldots, g_{iT})$. The observation $G_i(t)$ represents the glucose function recorded from individual i wearing the continuous glucose monitoring device. The time points $t = 1, 2, \ldots, T$ represent the discrete time instances at which the glucose level was measured in high frequency, and the values $g_{i1}, g_{i2}, \ldots, g_{iT}$ represent the glucose concentration at each time point.

There are many FDA techniques that can be used to explore various aspects of such functional data sets. Some analytic methods within this framework include: functional regression [Ramsay, 2005, Reiss et al., 2017], which can be used to model the relationship between the functional data and other variables; functional principal component analysis (FPCA) [Ramsay, 2004, 2005, Goldsmith et al., 2015, Yao et al., 2005, Nwanaji-Enwerem et al., 2021, Chen and Müller, 2012], which can help in dimension reduction and identifying dominant patterns variation; and functional classification and clustering [Heinzl and Tutz, 2014], which involve grouping functional data based on similarity or dissimilarity measures.

This dissertation will focus functional linear regression techniques with functional predictors. These techniques extend the concept of linear regression to functional predictors and responses. More specifically, it allows for modeling the relationship between functional variables, such as predicting a response curve based on a set of predictor curves. In the following chapters we will focus on a subset of functional linear regression, deemed scalar-on-function regression, in which the relationship between scalar outcomes and functional predictors is assessed.

1.2 Contribution

In the first project, I develop a more holistic, generalized, functional-focused approach to analyze PA data. Specifically, the proposed approach aims to free the dependence on subjective choices of pre-determined PA categorizations, and instead allow the data to adapatively determine the change-points and different activity ranges of interest. In this way, I utilize the supervised learning paradigm to assess the association of PA ranges with health outcomes of interest. I propose to holistically summarize an individual subject's activity profile using Occupation Time curves (OTCs). Being a functional predictor, OTCs describe the percentage of time spent at or above a continuum of activity count levels. The resulting functional curve is informative to capture time-course individual variability of PA. Utilizing the OTCs as a functional variable in the supervised learning paradigm leads us to a FDA approach, specifically a scalar-on-function model. I develop a multi-step adaptive learning algorithm, termed Functional Regularized Adaptive Changepoint-detection Technique (FRACT), to perform supervised learning via a scalar-on-function regression model. FRACT involves OTCs as the functional predictor of interest as well as other scalar covariates.

This learning analytic first incorporates a hybrid approach of fused lasso for clustering and Hidden Markov Model for change-point detection, and then executes a few refinement procedures to determine activity windows of interest. The multi-step nature of the proposed analytic provides some benefits versus one-step estimation alternatives, such as integer programming approaches proposed in Chapters 3 and 4. With its multiple estimation iterations, FRACT mines various relevant features in the data, providing useful insights into the intermediary steps of the data-learning process as well as the data quality and data structure. This attribute is particularly important when collaborating with non-statisticians who have limited training in data analytics and want to understand and cross-validate the analytic steps. Thus, the proposed methodology in Chapter 2 is more digestible to practitioners who may then prefer a deep dive into complex data and take a more understandable approach when conducting their scientific studies.

In the second project, I investigate functional associations between health outcomes and PA under an L_0 regularization approach. L_0 regularization is a type of regularization technique used in machine learning and optimization to add a penalty for the number of non-zero coefficients or variables in a model [Bertsimas et al., 2016]. It is typically applied in the context of linear regression or other linear models to promote sparsity; that is, it encourages the model to use fewer features or variables. Until relatively recently, L_0 regularization and discrete optimization has been less of a focus verses the L_1-related continuous optimization approaches as it was deemed computationally impractical. However, with recent advances in algorithmic and numeric capabilities, discrete optimization is a feasible and powerful tool [Bertsimas et al., 2016, Bertsimas and Shioda, 2009, Bertsimas et al., 2020].

I implement the modern optimization methods to functional analysis, by means of Mixed Integer Optimization (MIO) and Mixed Integer Programming (MIP)[Wolsey, 2008], to detect critical activity windows of interest again leveraging the information inherent in OTCs. In MIP, the objective is to optimize a linear or nonlinear function subject to a combination of continuous and discrete variables, with the goal of finding optimal values for all variables that satisfy the given constraints [Wolsey, 2008]. By formulating an L_0 regularization problem using binary variables, or by introducing additional binary variables to represent the sparsity pattern, one can incorporate L_0 regularization into MIP. In this formulation, the L_0 regularization penalty is transformed into a linear combination of binary variables and continuous variables, which makes it compatible with MIP solvers [Bertsimas et al., 2016]. In this case, rather than considering sparsity in terms of the number of individual variables, we consider the sparsity of differences between sequential variables, thereby proposing a fusion-adapted L_0 regularized learning method with MIP and MIO. To the best of my knowledge, I am the first to consider MIO methodologies both to conduct fusion as well as in a functional data framework. I will demonstrate that this method is computationally feasible and scalable to practically-sized problems of interest.

In the third project, I extend the previous L_0 framework to a functional framework with repeated measures to understand the population-average influence of repeatedly measured functional accelerometer data on health outcomes. Such an informatics toolbox can be applied to analyze the relationship of longitudinal functional digital features with longitudinal continuous outcomes; that is, this extension takes the form of longitudinal analysis with repeat measurements in both PA and outcome. Specifically, I propose a methodology that detects changepoints in serially measured functional accelerometer data to define critical windows of activity intensity that impacts longitudinal health outcomes, while also accounting for covariates of interest. In my data analysis example, with the repeat measurement taken approximately two years after the first, I study PA patterns from late-adolescence into early adulthood, and their longitudinal association with certain health outcomes of interest.

This statistical methodological extension incorporates the framework of Quadratic Inference Functions (QIF), where longitudinal dependence is modeled by estimation of covariance matrices [Song et al., 2009]. QIF offer several benefits in longitudinal data analysis. Importantly, QIF provides a flexible and efficient framework to handle correlated data often observed in longitudinal studies and can also accommodate different covariance structures, allowing researchers to model complex dependencies between repeated measurements accurately. Furthermore, QIF is robust to the mis-specification of the working covariance structure, making it a reliable option when the true covariance model is unknown. Overall, QIF contributes to more accurate and robust longitudinal data analysis, enhancing the validity of statistical inferences [Song et al., 2009]. Notably, the QIF framework does not require subject-level detail to conduct consistent estimation. Rather, one can use summary statistics from each time-point (or each correlated cluster) to conduct this analysis. This aspect opens the possibility for federated learning extensions, in which summary statistics from different data sources (such as hospitals or research groups) are combined for analysis, while still maintaining data privacy.

1.3 Organization of this Dissertation

This dissertation is organized into three distinct projects presented in Chapters 2, 3, and 4, respectively. Each of these Chapters, outlined in Section 1.2 above, proposes a distinct FDA framework methodology related to analysis of functional accelerometer data and follows a consistent high-level structure comprising three key components: method development, simulation experiments, and data analysis. For the proposed frameworks, the new methodologies are developed and justified based on their relevance to the specific problem, with theoretical guarantees introduced where appropriate. Simulation experiments are then presented to evaluate the performance and effectiveness of the proposed method under various scenarios. Data generated from these simulations are meticulously analyzed to draw meaningful conclusions and highlight the strengths and limitations of the approach. The proposed validated methodologies are then utilized in a real data analysis to assess the functional association between PA and health outcomes of interest, thereby examining the practical implications of the proposed methods and drawing conclusions from the application to real data.

Finally, Chapter 5 includes a comprehensive Summary and Future Work section, where the findings of each project are synthesized and compared. This section provides a cohesive analysis of the overall research outcomes, highlighting the strengths and limitations of the methodologies. Additionally, it identifies potential avenues for future research, discussing novel extensions or refinements to the proposed approaches to further enhance their utility and applicability. This culminating section provides a comprehensive reflection on the dissertation's contributions to the field, paving the way for further advancements and applications in the subject area.

CHAPTER 2

Supervised Learning of Physical Activity Features from Functional Accelerometer Data

2.1 Introduction

Physical Activity (PA) is of ubiquitous interest in smart-Health related research. One question of great interest is whether a more physically active person would be biologically "younger" than a less active person. In clinical lab settings researchers can directly observe and measure PA by well-designed experiments and facilities. However, measuring PA levels is more difficult to conduct in free-living populations outside of the lab setting. In the past, PA for these populations was often measured via subjective methods such as self-reported PA diaries. More recently, AI-guided sensors such as accelerometers have been utilized as objective measures to provide continuous high-frequency PA data [Trost, 2001, Troiano et al., 2014, Liu et al., 2021], giving rise to new technical needs and challenges in data analyses.

Accelerometer devices capture how speed changes over time through electrical signals representing the volume and intensity of movement. Such data are recorded in high resolutions of sampling frequencies (Hz), and then processed via proprietary software, such as ActiLife LLC. Typically, the processed data describing PA levels are expressed as activity "counts" over specific periods of time known as "epochs" [Chen and Bassett, 2005]. The count levels reflect the relative intensity of activity, with higher values indicating more intense exertion. For tri-axial accelerometers, the three-dimensional count information at each time point is often summarized into a one-dimensional summary value of Vector Magnitude (VM), with $VM = \sqrt{axis1^2 + axis2^2 + axis3^2}$ [Liu et al., 2021].

With the ability to provide high-frequency measurements (example frequency: 60Hz, or 60 data points per second), the technical challenge becomes how to retrieve useful information from this high-frequency time-series data type. A popular method of analyzing accelerometer data involves specifying activity "cutoffs" to discretize activity counts into categories, such as Sedentary, Light, and Moderate-to-Vigorous [Freedson et al., 2005, Crouter et al., 2015, Chandler et al., 2016]. Figure 2.1 illustrates an example of pre-specified cutoffs (the horizontal lines) applied to 24-hours of accelerometer data for a subject from our motivating dataset.

While the use of accelerometers provides a multitude of benefits, including reducing reporting bias found in subjective measures (e.g. self-reported PA surveys), and providing a continuous account of activity over a wear-time period, their use in studies does present some analytic challenges in data analysis [Troiano et al., 2014]. First, while categorization is a useful approach, the cutoffs must be pre-specified by the researcher. There are many potential cutoffs published in the literature, each validated against different narrowly focused studies with small subgroup populations [Crouter et al., 2015, Chandler et al., 2016, Troiano et al., 2014]. These cutoffs may be affected by many variables, including what device was used (e.g. Actigraph GT3X, Fitbit), the placement of the device (e.g. hip, wrist, ankle), or characteristics of the study population (e.g. age, sex, race) [Freedson et al., 2005]. For example, in the software ActiLife, which is used to analyze actigraphy data from Actigraph devices, there are over fifteen cutoff options. In addition, a researcher can choose to input their own cutoffs in ActiLife (ActiLife software, v6.13.3), leading to subjectivity in data processing. Such flexibility exposes analyses to the risk of applying pre-set cutoffs that do not align with a specific study population, potentially resulting in incorrect or biased activity classifications. Thus, it becomes important to call a new algorithm to adaptively choose appropriate cutoffs tailored to different studies.

The goal of this chapter is to develop a more holistic, generalized, functional-focused approach to analyze PA data. Specifically, the proposed approach aims to free the dependence on subjective choices of pre-determined PA categorizations, and instead allow the data to adaptively determine the changepoints and different activity ranges of interest. In this way, we utilize the supervised learning paradigm to assess the association of PA ranges with health outcomes of interest.

To this end, we consider actigraphy data under the purview of Occupation Time Curves (OTCs). This method of analyzing PA data involves a summary curve which describes the proportion of time an individual spends at or above successive activity levels [Bogachev and Ratanov, 2011]. The OTCs retain key features of the activity profile while greatly reducing the background noise inherent to accelerometer devices. OTCs compute the empirical proportional activity across possible activity levels for each individual, defined mathematically by $\mathbb{P}(VM(t) \geq c)$, where $VM(t)$ is the time-series of Vector Magnitude counts (as defined above), c represents the sequential moving activity levels, and $\mathbb{P}$ denotes an empirical probability measure defined by the proportion: $\frac{duration\ of\ \{t:VM(t)\geq c\}}{total\ duration\ of\ VM(t)}$ [Chang and McKeague, 2022]. Figure 2.2a illustrates the construction of an OTC, and Figure 2.2b illustrates four

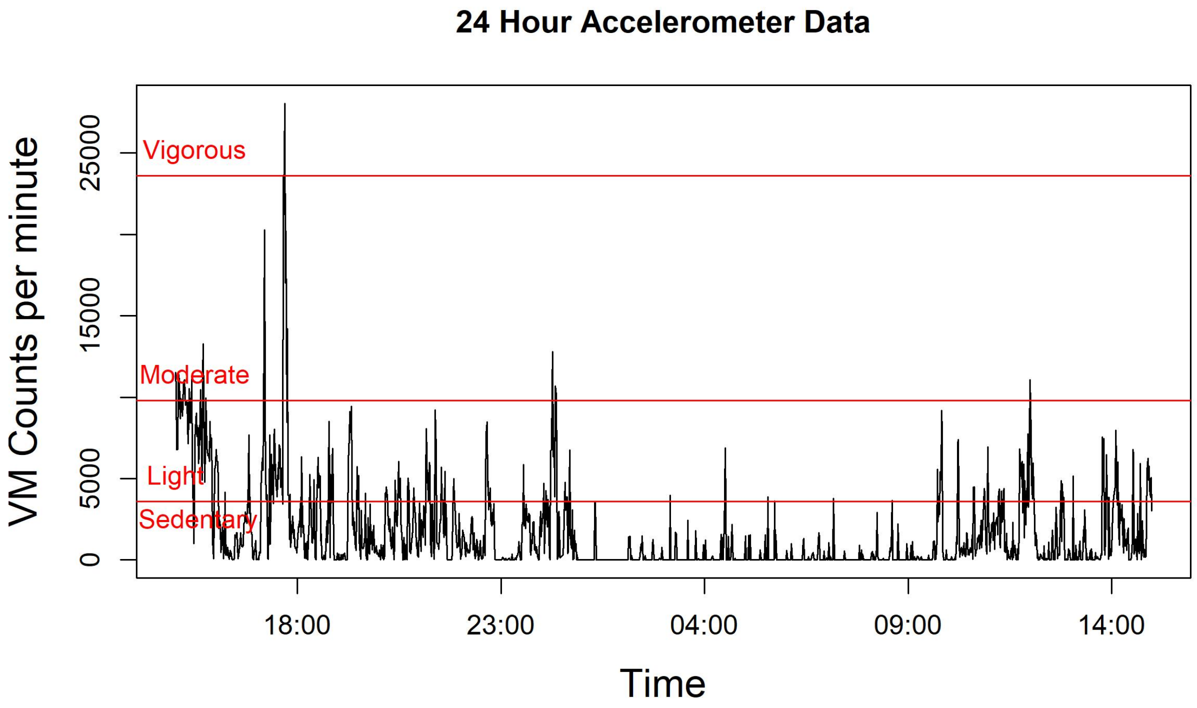


Figure 2.1: Accelerometer Data for an individual over a 24 hour period. The horizontal lines indicate PA categorization cutoffs based on Chandler Vector Magnitude cutoffs for 1-minute epochs [Chandler et al., 2016]. These cutoffs are pre-determined by in-lab supervised research, and now applied to a free-living subject.

OTC plots over VM counts varying from 0 to 30,000. Notably, Figure 2.2b includes OTCs for both more-active and less-active individuals, illustrating that the curves for more-active people show a distinct shape from those of less-active people. We aim to utilize the features inherent in the curves to assess the influence of PA profiles on a certain health outcome of interest (e.g. biological aging). Refer to Section 2.2 for more details concerning the motivating data and scientific research questions.

Utilizing the OTCs as a functional variable in the supervised learning paradigm leads us to a Functional Data Analysis (FDA) approach. For the ease of exposition, suppressing other covariates for the time being, we consider the following scalar-on-function regression model:

$$Y =< X, \beta > + \epsilon = \int_C X(c)\beta(c)dc + \epsilon, \tag{2.1}$$

where Y is a scalar health outcome of interest, $X(c)$ is the functional OTC defined on $C \subset \mathbb{R}$ and ϵ is the error term. Here, $< a, b >$ depicts the inner product of two square-integrable functions, namely $\int_C a(c)b(c)dc$ with $\int_C a^2(c)dc < \infty$ and $\int_C b^2(c)dc < \infty$. Categorization is of specific interest in this field for interpretability. Thus, we aim to develop a changepoint detection method that searches for the best segmentation of $X(c)$ by adaptively determining both the number and location of cutoffs that align with the PA intensity patterns. In this way, data-driven cutoffs are not only determined in a supervised fashion by the outcome of interest, but are also tailored to a study population under investigation. The rationale behind the goal of activity categorization is that not all PA ranges would impact a health outcome, and influential windows of activity, if they exist, should be appealing for the sake of interpretation. To address these technical needs, we develop a supervised learning analytic that incorporates multi-step, adaptive, learning procedures to estimate the functional parameter $\beta(c)$ with possible jump points representing activity ranges associated with the scalar health outcome.

The supervised learning aspect of this proposed methodology is due to the changepoint detection and functional parameter estimation being outcome dependent, as is required to address the scientific need of generalizability. The proposed method provides great flexibility to study similar scientific questions in other populations with various underlying characteristics and devices. This use of functional regression is notably different from current methods of analyzing accelerometer activity and investigating windows of activity associated with health outcomes. Unlike methods establishing fixed cutoff values regardless of specific outcomes under investigation, our analysis takes a new supervised learning approach in which changepoints and activity ranges are determined by the specific outcome of interest labelling in the model, which may take different forms in different applications.

The multi-step nature of the proposed analytic is an additional benefit versus one-step estimation alternatives, such as integer programming. With its multiple estimation iterations, the proposed method mines various relevant features in the data, providing useful insights into the intermediary steps of the data-learning process as well as the data quality and data structure. This is particularly important when collaborating with non-statisticians who have limited training in data analytics and want to understand and cross-validate the analytic steps. Thus, the proposed methodology is more digestible to practitioners who may then prefer a deep dive into complex data and take a more understandable approach when conducting their scientific studies.

This chapter is organized as follows. Section 2.2 introduces our motivating dataset from the Early Life Exposure in Mexico to ENvironmental Toxicants (ELEMENT) study, where epigenetic age (a scalar outcome that reflects biological aging) is described, while Section 2.3 introduces Occupation Time Curves (OTCs). Section 2.4 concerns the development of a multi-step supervised adaptive learning analytic that enables changepoint detection of important activity ranges, whose implementation is detailed in Section 2.5 and performance is evaluated and demonstrated through simulation experiments in Section 2.6. In Section 2.7, we apply our proposed method to assess the functional association between PA and epigenetic age. Section 2.8 includes a few concluding remarks.

2.2 Motivating Study Data

This work is motivated by the Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) cohort, which is a longitudinal birth cohort study involving mother/child dyads from Mexico City. Details of this study have been discussed in a previously published review paper [Perng et al., 2019], with details relevant to this research described below.

2.2.1 ELEMENT Actigraphy Data

As a part of this 2015 ELEMENT follow-up study, researchers collected actigraphy data from 539 children (258 boys and 281 girls) with mean (SD) ages of 13.9 (2.2), ranging from 9 to 18 years old. The participants were provided a wrist-worn, tri-axial Actigraph GT3X+ (Actigraph LLC), which was worn for seven consecutive days with no interruption. The Actigraph GT3X+ has an acceleration range of $\pm 6g (g = 9.81 m/s^2)$ with a default sampling frequency of 30 Hz corresponding to a collection of 30 measurements per second. The raw tri-axial data was processed and summarized into epochs of various lengths (i.e. 10 sec, 30 sec, 1 min). In this chapter, we focus on Vector Magnitude (VM) activity counts over one-minute epochs, which is widely used in practice. This Actigraph device is water-resistant and can be removed only when physically cut off. This warranted both high study compliance and limited non-wear time during the consecutive seven days of actigraphy data collection.

In addition to PA, the ELEMENT cohort also collected DNA methylation data from EPIC array (850K) that was used to calculated epigenetic age. In our study, we also consider covariates, including chronological age, sex, lead exposure [Wu et al., 2019], and pubertal status measured by a five-category ordinal variable of Tanner staging, and others.

2.2.2 Epigenetic Age

Biological aging rates are of great interest, but not well understood. There is significant variation in how people visibly age or are affected by age-related disease. By quantifying this characteristic, the biological aging rate can act as a biomarker of the overall state of health, and allow for personalized or pre-emptive health interventions [Marioni et al., 2015]. Epigenetic Age is one such quantitative way to represent a person's biological aging, and is the outcome of interest in this chapter. Epigenetic Age Calculators hosted online [Horvath, 2013] receive inputs of DNA methylation (DNAm) alterations along different areas of the genome and deliver an output of predicted Epigenetic Age; see Horvath (2013), among others.

Research in fetal origins of health and disease suggests that early life exposures could form the foundation of health issues experienced by individuals in adulthood. After birth, studies show that children and adolescents (age 0-18) undergo the fastest and most dynamic rate of growth and DNAm changes [Wu et al., 2019, McEwen et al., 2020]. These childhood environmental and experiential factors can be observed in changes in the the DNA methalome and thus reflected in epigenetic age. An important investigation of scientific interest is to assess the association of epigenetic age with objectively measured functional PA (e.g. OTCs shown in Figure 2.4).

Research has demonstrated the relationship between health and epigenetic age may not be monotonic, motivating the use of different types of epigenetic ages to study the influence of PA on different biological aging processes. Here, we focus on two biological ages for the adolescents aged 9-18: Horvath's Skin and Blood clock (DNAm AgeSkinBlood) [Horvath et al., 2018] and Levine's Phenotypic Age clock (DNAm PhenoAge) [Levine et al., 2018], both calculated from the epigenetic age calculators [Horvath, 2013]. The former reflects aging in skin and blood cells while the latter pertains to age-related disease/phenotypes.

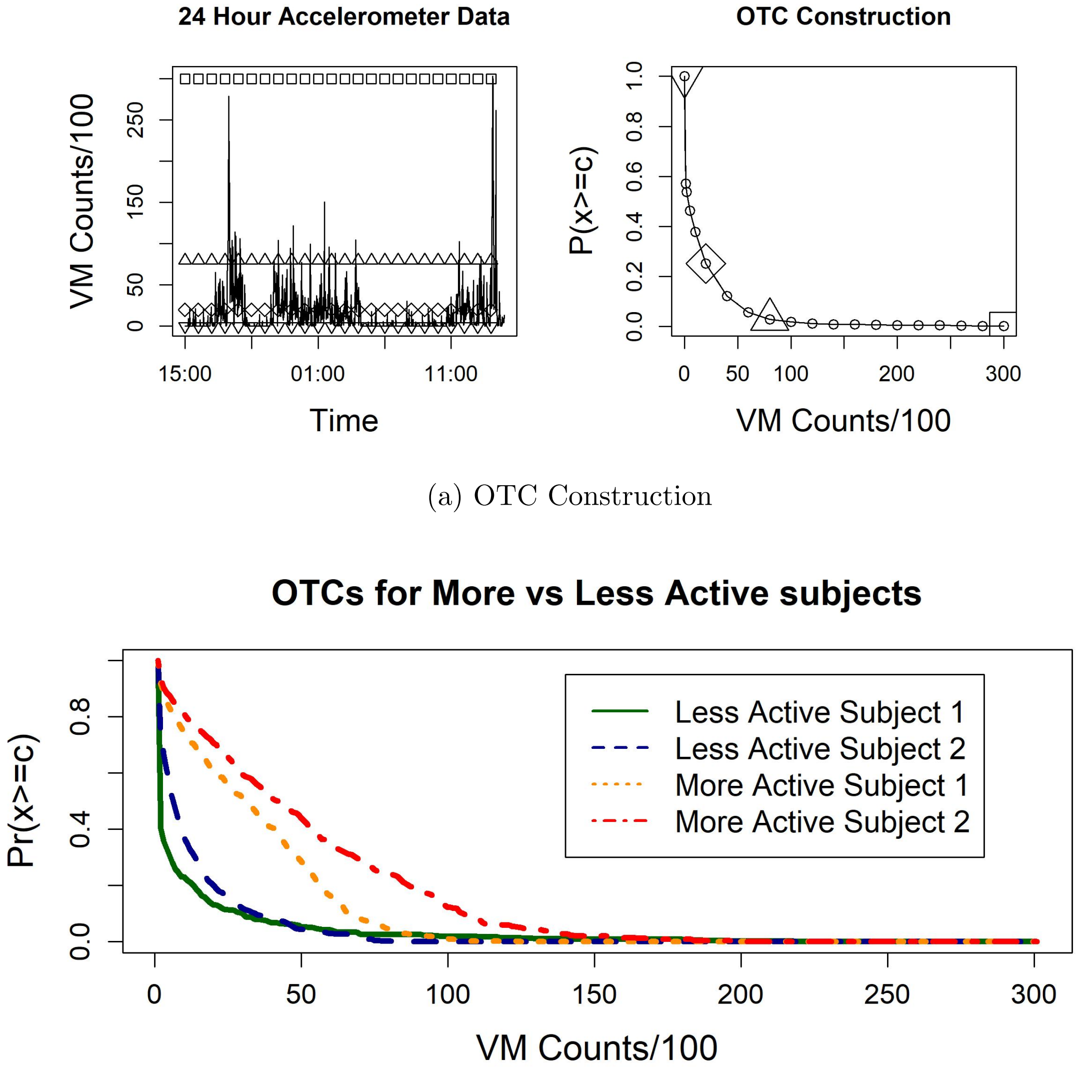
2.3 Occupation Time Curves: A Functional Predictor

The OTC [Bogachev and Ratanov, 2011] provides a useful way to summarize PA patterns and represent data as an informative functional curve. OTCs summarize high-frequency time-series accelerometer data by representing the empirical proportion of time an individual spends at successive activity count levels. For a vector of VM time-series data $VM(t)$, an OTC can be calculated over a domain of count values $\mathcal{C}$ by: $OTC(c) = \mathbb{P}(VM(t) \geq c)$ for $c \in \mathcal{C}$, where c represents the sequential moving activity levels, and $\mathbb{P}$ denotes an empirical probability measure defined by the proportion: $\frac{duration\ of\ \{t: VM(t) \geq c\}}{total\ duration\ of\ VM(t)}$.

Calculating an OTC is computationally straightforward. Take an example of a time-series of Vector Magnitude (VM) counts $VM(t)$. To yield the first data point of OTC, we calculate the empirical proportion of time spent at or above count level of 0, (i.e. $c = 0$), or $OTC(0) = \mathbb{P}(VM(t) \geq 0)$, which is clearly 100%. We then vary c in ascending order, such as say $c = 100$ and $c = 200$, and calculate $OTC(100) = \mathbb{P}(VM(t) \geq 100)$ and $OTC(200) = \mathbb{P}(VM(t) \geq 200)$, and so on. This calculation continues up to $c = 30,000$, or 300×10^2, which appears as the largest ordered VM count in the data. Figure 2.2a illustrates this numerical OTC construction procedure. The resulting OTC is denoted as $X(c),\ c \in \mathcal{C} = [0, 300 \times 10^2]$ throughout the rest of this chapter.

Figure 2.2a illustrates the construction of an OTC, with the successive increase in threshold c shown on the left panel, and the respective $\mathbb{P}(VM(t) \geq c)$ shown in the resulting continuous curve on the right panel. The shape of the OTCs reflect the relative amounts of time an individual spends in different activity levels. For an inactive person, who spends the majority of time in low-activity counts, their OTC curve would decay quickly, representing a high proportion of time in low activity levels and a small proportion of time in high activity levels. However, the OTC of an individual with higher proportions of time spent in high activity levels would appear more linear in nature at its start, before eventually flattening. These differences are illustrated in Figure 2.2b. Thus, the OTCs reflect inherent PA characteristics of each individual.

These OTCs provide a more flexible and generalizable PA summary variable than using the standard "minutes per activity category" from continuous accelerometer data as shown in Figure 2.1. In order to ensure comparable summary measures among subjects when using the previously described standard approach, the data from each subject should reflect non-missing continuous data over the same length of time. As subjects generally have different lengths of "awake" (i.e. non-sleep) time, as well as different patterns of non-wear time (i.e. missing data), these requirements are not often met in practice. In contrast, OTCs scale the PA measures to the duration of time under consideration, providing more apt comparison



(a) OTC Construction

(b) OTC Comparison

Figure 2.2: (a) The construction of an OTC from a time-series of VM counts. The left panel represents accelerometer data, with an increasing bar of count cutoff indicated by the horizontal lines with varying point-shapes. The corresponding proportion of time spent at or above that level of activity is shown in the indicated point on the right panel with the corresponding point shape. The grey continuous curve is the realized OTC for this individual (b) Comparison of OTC shapes for More vs Less Active Individuals, with VM count summarized over 1-min epochs varying over 0 to 30000. The distinctive shapes of the curves represent the subject's activity pattern. For example,curves of less active people steeply drop in the beginning, signifying that a small percentage of their time is spent in even mid-active regions.

between individuals who have different lengths of time of continuous accelerometer data.

Utilizing the functional OTC curve also requires a different approach to estimating the parameters of effect between PA and specific health outcomes of interest. While the standard analysis approach illustrated in Figure 2.1 incorporates fixed coefficients relating Total Minutes in each pre-fixed activity window, the OTC requires a non-parametric coefficient. We model the OTC as a functional covariate in a scalar-on-function regression model (described further in Section 2.4.1) in which the goal is to estimate the non-constant β parameter as a function of activity count, and more specifically as a step-function. For example, Figure 3.1 illustrates a continuous β estimation as a step-function of activity count, which reflects specific activity windows in the OTC. This functional β suggests that the proportion of time spent in the three different segments of the OTC have different impact on the health outcome of interest. We will develop this model formulation in Section 2.4.

2.4 Method

In this chapter we develop a holistic multi-step supervised learning approach to analyze accelerometer data in that both changepoints and PA ranges are adaptively determined via scalar-on-function regression model. A key aspect of the proposed method is to detect PA ranges that impact the association between health outcomes of interest (e.g. epigenetic age) and a functional covariate, OTC, adjusting for other variables. This proposed methodology is considered *supervised* learning as the goal is to estimate changepoints that are *outcome dependent*. That is, the activity windows are determined by the specific outcome labelling in the model; this supervised-learning aspect is intrinsic to the generalizability of the proposed method to a wide range of scientific problems that may use different wearable devices or means of data collection for different study purposes.

2.4.1 Scalar-on-Function Regression and Changepoint Detection

It is natural to utilize such inherent variability in the curves to study the association between PA and our outcome of interest, epigenetic age. Since OTCs present distinct informative functional shapes on individual's PA profile, it is desirable to take OTC as a functional covariate in the association analysis through a scalar-on-function regression model within a functional data analysis framework. Identifying PA ranges pertains to detection of changepoints or cutoffs on OTC $X(c)$. This sets up a different analytic goal than that of standard PA analyses performed in literature. For the ease of exposition, we begin with a simple scalar-on-function linear model with no covariates as described in Equation (2.1), in which the error terms ϵ_i's are assumed to be independent and identically normal distributed with mean 0 and variance σ^2.

Our analytic goal is to estimate the functional parameter $\beta(c)$ with certain jump points, which describes a piece-wise varying effect of the OTC $X(c)$ on epigenetic age Y. Here, the changepoints define the windows of PA, similar to the practice of activity categorization widely considered in the literature for scientific interpretation. In other words, our proposed approach focuses on changepoint detection, or grouping like activity count ranges with similar effects on the outcome to gain better insights and interpretations for the functional association. Our key idea is rooted in the utility of fused regularization technique that enables the identification of jump points of functional parameter $\beta(c)$. This analytic task is technically challenging as it involves both clustering and estimation of functional parameters. To proceed, we first discretize each OTC into many small segments so the integral $\int_C \beta(c)X(c)dc$ can be approximated by a step-function over many small pieces. That is, we divide interval C into J-many small successive intervals with a grid $c_0 = 0, c_1, \cdots, c_J = 30,000$, namely $C = [0, c_1] \cup_{j=2}^{J} (c_{j-1}, c_j]$, and assume $\beta(c)$ takes one parameter within one small interval. Precisely, on the j^{th} interval $(c_{j-1}, c_j]$, we set constant parameters β_j, $j = \{1, \cdots, J\}$ resulting in $\beta(c) \approx \sum_{j=1}^{J} \beta_j I\left(X(c) \in (c_{j-1}, c_j]\right)$. Consequently, we have

$$\begin{aligned} \int_C \beta(c)X(c)\ dc &= \sum_{j=1}^{J} \int_{c_{j-1}}^{c_j} \beta(c)X(c)\ dc \\ &\approx \sum_{j=1}^{J} \beta_j \int_{c_{j-1}}^{c_j} X(c)\ dc := \sum \beta_j A_j, \end{aligned} \tag{2.2}$$

where A_j denotes the Area Under the Curve (AUC) over interval $(c_{j-1}, c_j]$ or $A_j = \int_{c_{j-1}}^{c_j} X(c)\ dc$. Of note, while $X(c)$ is monotonically decreasing due to nature or OTCs, there is no restriction of monotonicity on $\beta(c)$. In preparation of regularized estimation, we normalize individual $A_j's$ to mean 0 and variance 1, respectively. One key methodological goal is to fuse similar β_j's into bigger segments to identify appropriate activity windows affecting the outcome of interest. One challenge arising from the discretization strategy is that $\beta_j's$ in Equation (2.2) may be high-dimensional, inevitably requiring a regularization method (e.g. fused lasso). Unfortunately, the resulting A_j variables appear highly correlated, essentially challenging existing high-dimensional regularization methods.

There are existing methods applicable to carry out the parameter fusion on β_j, among which fused lasso [Tibshirani et al., 2005] and Hidden Markov Model (HMM) [Rabiner and Juang, 1986] are popular. However, these existent approaches do not perform well due to the high correlation of A_j's. As shown in Figure 2.3a from a simulation model, the regularized estimates of β_j's are bifurcate, leading to a clearly poor parameter fusion on these estimates and an inaccurate determination of changepoints.

To address the issue of high correlation, we propose a supervised learning analytic: the Functional Regularized Adaptive Changepoint-detection Technique (FRACT), which provides more effective strategies via adaptive, multi-step learning algorithms. FRACT consists of the following procedures: (i) Tuning J, the number of intervals, (ii) Initialization of β_j's, (iii) Changepoint Detection, and (iv) Refinement Learning.

2.4.2 FRACT Methodology

Here we present the details of FRACT. The multi-step learning analytic encompasses two penalization themes; strategies 1 and 2 deal with the first regularized estimation to generate initial estimates of β_j, while strategies 3 and 4 aim to group the initial β_j estimates to form activity windows whose edges determine jump points. Algorithm 1 outlines these procedures, with further detail in the FRACT Implementation Section 2.5.

Strategy 1: Tuning the number of intervals, J. The first strategy of FRACT is to alleviate the high correlation among $A_j's$. This is achieved through tuning the number of intervals J via a trade-off between pair-wise correlation levels and minimal loss of signal strength. Our experience from simulation experiments suggests that selecting a J-partition of C such that $cor(A_j, A_{j+1}) < 0.98$, $cor(A_j, A_{j+5}) < 0.90$, and $cor(A_j, A_{j+10}) < 0.80$ is reasonable. Future applications of this methodology should select the maximum J such that the correlation parameters remain below this suggested threshold. Such a selection minimizes the trade-off between signal strength and multi-collinearity, resulting in a more stable and reliable data analyses. Note that this tuning step is mostly responsible for ensuring the quality of initial estimates $\hat{\beta}_j$'s; that is, to avoid the bifurcate initial estimates in Figure 2.3a. However, this tuning step may not be necessary if an algorithm more suited to high-correlation is used.

Strategy 2: β_j Initialization. This step is to generate initial regularized estimates of $\beta_j's$ in Equation (2.2). To gain numeric stability we adopt the Minimax Convex Penalty (MCP) that takes the form: $p_\lambda(\beta) = \lambda|\beta| - \frac{\beta^2}{2a}$ if $|\beta| \leq a\lambda,$ and $\frac{a\lambda^2}{2}$otherwise, with $\lambda \geq 0$, $a > 1$, and enjoys lower estimation bias [Zhang, 2010]. With this choice in parameter estimation, and with tuning parameters λ and a selected via cross-validation, for model $Y = \sum \beta_i A_i + Z^T\alpha + \epsilon$ we obtain initial estimates by $(\hat{\beta}, \hat{\alpha}) = argmin_{(\beta,\alpha)} \frac{1}{2}||Y - A\beta - Z\alpha||_2^2 + p_\lambda(\beta)$, where Y is the outcome of interest, and A, Z are the relevant variables, covariates with respective parameters β, α.

Strategy 3: Changepoint Detection. The initial estimates $\hat{\beta}_j$'s are then processed under

Strategy 3 to perform a clustering analysis by means of Hidden Markov Modelling (HMM) initialized by the results obtained by Fused Lasso. HMM contains a latent process that helps group similar $\hat{\beta}_j$ values, leading to the detection of jump points. An important element of HMM is the transition matrix, which determines the probabilities of transitioning between latent states, and which plays a two-fold role in the FRACT analytic. First, supplying HMM with a smart initial estimated transition matrix provides useful "warm starting points" to improve numeric stability of the EM algorithm used in HMM. These "warm starting points" are estimated from initial clustering estimates obtained by fusing the initial $\hat{\beta}_j$'s via Fused Lasso. See Algorithm 2. Second, an evaluation of the final transition matrix from the fitted HMM model facilitates tuning the ideal number of latent clusters K (or, the number of activity windows). Both the use of Fused Lasso-estimated "warm starting points" of the initial transition matrix, and the evaluation of the fitted HMM transition matrix, improve the FRACT analytic's ability to correctly identify changepoints versus applying a one-step algorithm alone, as discussed in detail in Section 2.5.

Strategy 4: Refinement Learning of Cutpoints. The final step in FRACT pertains to a Refinement Learning procedure to fine-tune the changepoint detection selections via supervised learning techniques. This strategy allows us to systematically evaluate if micro-modifications of changepoints in the fitted model result in improved goodness of fit, and addresses concerns of potential over or under-fitting by comparing the current size K model with those of size $K-1$ and $K+1$. See implementation details in Algorithms 3 and 4 in Section 2.5.

Notes on Regularization. In this analytic, the worry of over-fitting pertains to the number of activity windows K; we do not want to estimate a higher K than necessary. To control this, we use the regularization methods of Fused Lasso, HMM, and goodness-of-fit measures to learn, test, and calibrate the value K. Fused Lasso provides initial feature fusion as a smoothing technique. These results provide some initial demonstration of group structures as the so-called "warm starting points" of HMM's initial transition matrix, thereby reducing the sensitivity to individual data points and outliers. Within the HMM framework, regularization focuses on controlling the complexity of the model via limiting the number of hidden states to range of K (between $2-4$ here), and further data evidence is generated to test and calibrate the group structure. Lastly, the goodness-of-fit measure of EBIC (Extended Bayesian Information Criterion) is used as a regularization method to make a choice of parsimonious models. EBIC is an extension of BIC that incorporates an additional penalty term, the complexity parameter, which further controls the trade-off between model fit and complexity. Thus, using EBIC for final-model selection helps to prevent overfitting and find a parsimonious model that balances goodness of fit and complexity.

Algorithm 1 FRACT for Changepoint and Activity Window Detection

Input: Time series accelerometer data of Vector Magnitude counts $VM(t)$ with continuous range of activity counts $\mathcal{C}$

Output: Estimated activity window cutpoints, and estimated association parameters

```
 1: procedure CALCULATE OTCS AND AUCS
 2:     Set $\mathcal{C} = (0, c_{max})$, where $c_{max}$ = the maximum VM count observed
 3:     With $VM(t)$, calculate OTC $X(c),\ c \in \mathcal{C}$
 4:     Given partition $c_0 = 0, c_1, \cdots, c_J$, calculate the AUCs $A_j = \int_{c_{j-1}}^{c_j} X(c)\, dc$, $j \in 1, \cdots, J$
 5:     Normalize individual $A_j$'s to have mean 0 and variance 1, respectively
 6: end procedure
 7: procedure STRATEGY 1: TUNE J
 8:     Alleviate the high correlation among the AUCs $A_j$, $j \in 1, \cdots, J$
 9:     Combine successive normalized integrals to reduce pairwise correlations so that
        the resulting partition satisfies: $cor(A_j, A_{j+1}) < 0.98$, $cor(A_j, A_{j+5}) < 0.90$, and
        $cor(A_j, A_{j+10}) < 0.80$
10:     Note: If an using algorithm suited to correlated $A'_j s$, Strategy 1 may be skipped
11: end procedure
12: procedure STRATEGY 2: INITIALIZE ESTIMATES OF $\beta'_j s$
13:     Conduct high-dimensional linear regression with Minimax Convex Penalty (MCP)
14:     Run linear Model: $Y \sim A_1 + \cdots + A_J + Z$, where Z is a vector of scalar covariates
15: end procedure
16: procedure STRATEGY 3: DETECT CHANGEPOINTS
17: Set: $K$ = number of groups, with $K \geq 2$
18:     Run $K$-size fused lasso for the initial $\hat{\beta}'_j s$ obtained by MCP in Strategy 2 (lines 13-14)
19:     Calculate an initial transition matrix from the group labels determined by fused lasso
        ▷ See Algorithm 2
20:     repeat
21:         Fit $K$-state HMM with initial transition matrix
22:         Calculate Extended BIC (EBIC)
23:         Assess updated transition matrix to determine the convergence of the HMM
24:     until Repeat HMM fit (lines 21-23) $m$ (say, 10) times for numeric stability
25:     return Best model among the $m$ fitted models selected based on EBIC and updated
        transition matrix at convergence of HMM
26:     Fit piecewise linear model for association parameter estimation and inference under
        the HMM defined cutpoints from line 25, as well as vector of covariates $Z$,
27:     Continue to Strategy 4 to select final $K$
28: end procedure
29: procedure STRATEGY 4: PERFORM REFINEMENT LEARNING
30:     Edge-swapping                                              ▷ See Algorithm 3
31:     Merging                                                    ▷ See Algorithm 3
32:     Breaking                                                   ▷ See Algorithm 4
33: end procedure
```

2.5 FRACT Implementation

2.5.1 Initial Clustering

All the line numbers in this section refer to Algorithm 1 unless specified otherwise. The FRACT methodology begins with clustering the initial estimates $\hat{\beta}_1, \cdots, \hat{\beta}_J$ obtained from the MCP regularized linear regression, where J is tuned in advance, if necessary, to satisfy the correlation constraints among $A_j's$. This clustering analysis involves a step of fused lasso to generate "warm starting points" for initial changepoints (Lines 18-19), followed by a HMM fit to settle cutpoints and group membership (Lines 21-25). The "warm starting points" seem useful to improve the numeric stability of the EM algorithm used in the subsequent HMM, as well as provide the initial transition matrix in the HMM fit. With a given K (the number of categories in the latent process), a transition probability is estimated by the relative length of a fused range of initial $\hat{\beta}_j$'s (Line 19). Algorithm 2 gives an example calculation of initial transition probability matrix. In Lines 21-25, clusters, cluster membership, and cutpoints are determined by means of the HMM analysis in that one latent state corresponds to one cluster of the initial estimates $\hat{\beta}_j$.

2.5.2 Tuning K

For each scenario of K clusters (e.g. the number of latent states in HMM), the HMM is executed m times (*say*, 10) to ensure numeric stability; among m fitted HMMs, the one with the smallest EBIC is selected. With the selected best HMM, the corresponding transition probabilities are then used to assess viability of clusters. Here, a clustering result is deemed "viable" if the last state in the latent process gets stabilized with no chance of jumping between states, as judged by the diagonals of the estimated transition matrix at convergence. If the probability of remaining in the final $k = K$ state, defined as the diagonal element p_{KK} in the estimated transition matrix, is 1 then the HMM is said to have "settled" in its final latent state. This implies stable clustering results with reliable cutpoints and cluster membership.

In Line 26 of Algorithm 1, with clustering results from the K-state HMM, we fit a linear model with a piece-wise mean function based on the estimated cutoffs of $c_1, \cdots c_K$:

$$Y = \beta_1 \tilde{A}_1 + \beta_2 \tilde{A}_2 + \cdots + \beta_K \tilde{A}_K + Z^T \alpha + \epsilon \tag{2.3}$$

where $\tilde{A}_j's$ denote the AUCs over the updated partition intervals via merging smaller intervals into a few big intervals or windows. Using EBIC again, we can determine the K by directly comparing the different size models, such as 2-state, 3-state, and 4-state models. The choice

Algorithm 2 Example Calculation of Initial Transition Probabilities with $K = 3$

Input: $K = 3$ groups with cluster membership and cutpoints derived from fused lasso (Line 18 of Algorithm 1), and group cardinality J_1, J_2, J_3 such that $J_1 + J_2 + J_3 = J$. The groups $(1, 2, 3)$ are determined by the ordering of A_j

Output: 3×3 initial Transition Probability Matrix

Note: Transitions between groups is allowed only between neighboring groups in a forward direction. This results in a K-band transition matrix.

```
procedure CALCULATE INITIAL PROBABILITY MATRIX(3 × 3 matrix)
    for k=(1,2) do
        Prob. of remaining in Group k = $p_{kk} = \frac{J_k - 1}{J_k}$
        Prob. of transitioning from Group $k$ to Group $k+1 = p_{k,k+1} = \frac{1}{J_k}$
        Note: $p_{kk} + p_{k,k+1} = 1$
    end for
    Prob. of transitioning from Group 3 = $p_{33} = 1$
    3 × 3 initial probability matrix:
    $\begin{bmatrix} p_{11} & p_{12} & 0 \\ 0 & p_{22} & p_{23} \\ 0 & 0 & 1 \end{bmatrix}$
end procedure
```

of potential Ks is motivated by prior scientific knowledge; in this case, PA is classified into $2-4$ intensity levels in literature. Considering larger potential K classifications risks presenting results with statistical, but not clinical or scientific, significance.

2.5.3 Refinement Learning of Cutpoints

To further improve the analysis we propose a Refinement Learning procedure that refines the cutoff selection via supervised learning techniques, summarized in Algorithms 3 and 4. We suppose that cutoffs $c_1, \cdots, c_K$ and estimates $\hat{\beta}_1, \cdots, \hat{\beta}_K$ are available from previous steps.

Edge-Swapping and Merging Technique. Algorithm 3 aims to refine the cutoffs $c_1, \cdots, c_K$ and minimize over-fitting. For fixed K, the edge-swapping technique begins with identifying the narrowest gap in adjacent estimates $\hat{\beta}_k$ and associated intervals: $k_{min} = argmin_k|\hat{\beta}_{k+1} - \hat{\beta}_k|$, $k \in \{1, 2, \cdots, K-1\}$. These two identified intervals k_{min} and $k_{min}+1$ are considered for edge-swapping; we systematically swap their window edge-points until the intervals become completely merged. Each swap gives rise to a different K-group partition with different group cardinality as well as different $\tilde{A}_j, j = k_{min}, k_{min}+1$. As a result of this edge-swapping, we create many K-group partitions as well as one $K-1$ partition under the two intervals being fully merged. The EBIC from each linear regression model (2.3) is compared to make model selection and thus further select a desirable partition.

Algorithm 3 Implementation: Edge-Swapping and Merging

```
Input: $K \geq 2$ number of intervals, with cutpoints $(c_{k-1}, c_k)$, and estimates $\hat{\beta}'_k s$, $k = 1, \cdots, K$
Calculate $k_{min} = argmin_k|\hat{\beta}_{k+1} - \hat{\beta}_k|, k \in \{1, 2, \cdots, K-1\}$ and identify the left and right intervals    ▷ Determine minimum gap in adjacent $\beta'_k s$
procedure EDGE-SWAPPING
    for $c^* \in (c_{k_{min}-1} + 1, c_{k_{min}})$ do    ▷ Swap edge-points between the two windows
        Identify the left $L(c^*) = (c_{k_{min}}, c^*)$ and the right window $R(c^*) = (c^*, c_{k_{min}+1})$
        Fit a piecewise linear model (2.3) with the resulting partition from $c^*$
        return EBIC
    end for
    return Best $K$-group model with the smallest EBIC among all models, including the original K-group model and all K-group models with edge-swapping by $c^*$
end procedure
procedure MERGING
    Set merged window = $(c_{k_{min}-1}, c_{k_{min}+1})$
    Fit a piecewise linear model with new partition
    return EBIC of $(K-1)$-group model
end procedure
```

Breaking by Exploration and Confirmation Steps. To address potential under-fitting, or a current K being smaller than desirable, we implement a "Breaking" strategy described in Algorithm 4. This strategy attempts to create one more interval, forming a new piecewise mean function in (2.3) in which one of the original windows (or intervals) is split into two smaller windows. This splitting consists of two steps: *Exploration* and *Confirmation.* In the *Exploration* step, we randomly choose a locality at which the target window is split into two sub-windows, increasing the number of groups from K to $K+1$. This splitting is repeated s times (*say*, 10) at different random localities. Of the total number of s breaks, the number of break locations in interval$k$, s_k,is determined by the proportional cardinality, J_k, of that window versus the total cardinality, J, of all the windows. Given an s_k for a specific interval k, the exact break locations are determined randomly under a uniform distribution over that interval. We use EBIC to determine if any of the resulting $(K+1)$-group models has a better fit than the original K-group model. If so, we moved to the subsequent *Confirmation* step.

In the *Confirmation* step, we consider d (*say*, 10) surrounding points as alternative breakpoints. Based on these d new cuptoints, we refit the $(K+1)$-group model and calculate EBIC. If over 50% of these $(K+1)$-group models (2.3) have better fit than the original K-group model, we accept the new partition of $(K+1)$-groups and choose the best $(K+1)$-group model with the smallest EBIC. This *Confirmation* step is run with multiple flanking points to ensure the augmented $(K+1)$-group model is superior over the smaller K-group model, i.e. the lower EBIC value is not simply a result of chance.

Up to now, a fixed K is preset. The final choice of K is determined in the final step. We run the above algorithms $1-4$ over $K = 2, 3, 4,$ and more if needed, and use EBIC to select the final model with the best goodness of fit among the candidates. At its end, the FRACT analytic delivers both estimates of activity window cutpoints and their associated parameter estimates, as well as the estimates of covariate effects for the included covariates of interest.

2.6 Simulation Experiments

We assessed the performance of the proposed FRACT analytic through extensive simulation experiments under various effect sizes, window lengths, and number of windows (or cutpoints). We focus on the performance of FRACT in determining the number of activity windows K, estimating the cut-points $(c_1, \cdots, c_K)$ and effect sizes $(\beta_1, \cdots, \beta_K)$.

2.6.1 Simulation Setup

To simulate the functional OTC, we first simulated 6-hour time-series of VM counts by linking many consecutive 10-minute intervals of accelerometer data from the 539 ELEMENT children, whose individual 6-hour time-series of VM counts are divided into non-overlapping 10-minute segments. Each 10-minute interval is randomly drawn from a pool of 539 10-minute candidate segments. To ensure that the variability in the simulated OTC curves reflected the variability in the motivating dataset (Figure 2.4), we first classified the subjects into three groups with low, medium, and high levels of PA respectively, as defined by tertiles of "Moderate-to-Vigorous" VM counts using the pre-set Chandler cutoffs. We then simulated the time-series data within each tertile. With the simulated VM counts, OTC curves were calculated as described in Section 2.3.

Given true cutoffs $c_0^* = 0, c_1^*, \cdots, c_{K^*-1}^*, c_{K^*} = 30,000$, the successive integrals over C were specified, and the K^*-element vector of AUCs, $(A_1, \cdots, A_K^*)^T$, for each subject was calculated. Outcome Y was simulated from a linear model $Y = \sum_{k=1}^{K^*} \beta_k^* A_k^* + Z\alpha^* + \epsilon$, with true effect sizes $\beta_1^*, \cdots, \beta_K^*$ and α^*, where single continuous covariate $Z \sim N(0, 1)$ and $\epsilon \sim N(0, 10)$. The

Algorithm 4 Implementation: Breaking via Exploration and Confirmation

Input: Let K = number of activity windows, with window length $J_k = |c_{k-1} - c_k|$, end cutpoint c_k, and current estimates $\hat{\beta}'_k s$, $K \geq 2$, $k \in \{1, \cdots, K\}$, $J_1 + \cdots + J_K = J$

Define: $EBIC(K)$ =EBIC of K-group model

```
procedure EXPLORATION STEP
Set: $s$ = total number of break locations
    for $k \in (1, \cdots, K)$ do
        Calculate $s_k = J_k/J \times s$ = proportional number of potential breakpoints $s$ allocated to Window k, $(c_{k-1}, c_k)$
        $\mathcal{B}_k = \{b_1, \cdots, b_{s_k}\}$ = Randomly selected $s_k$ break points from interval $(c_{k-1}, c_k)$
        With each element $b \in \mathcal{B}_k$, fit a piecewise linear model (2.3) with new partitions
        return $EBIC(K+1, b)$ = EBIC of $(K+1)$-group model at break $b$, $b = b_1, \cdots, b_{s_k}$
    end for
    if $argmin_{b \in \mathcal{B}} EBIC(K+1, b) < EBIC(K)$, with $\mathcal{B} = \{\mathcal{B}_1, \cdots, \mathcal{B}_K\}$ and $b \in \mathcal{B}$ then
        Proceed to Confirmation step at a partition with $b_{min} = argmin_{b \in \mathcal{B}} EBIC(K+1, b)$
    else
        No breaking, and remain at $K$-group model
    end if
end procedure
procedure CONFIRMATION STEP
    For the $b_{min}$ partition select $d$ ($say$, 10) breakpoints around $b_{min}$ denoted by $\mathcal{B}(b_{min})$
    for b $\in \mathcal{B}(b_{min})$ do
        Create a new partition with $b$, and fit $K+1$ model
        return $EBIC(K+1, b)$ = EBIC of $K+1$ model with breakpoint $b$
    end for
    if $\sum_{b \in \mathcal{B}(b_{min})} I[EBIC(K+1, b) < EBIC(K)]/d \geq 50\%$ then
        Keep the $(K+1)$ model with the partition at $b_{min}$
    else
        Reject breaking
    end if
end procedure
```

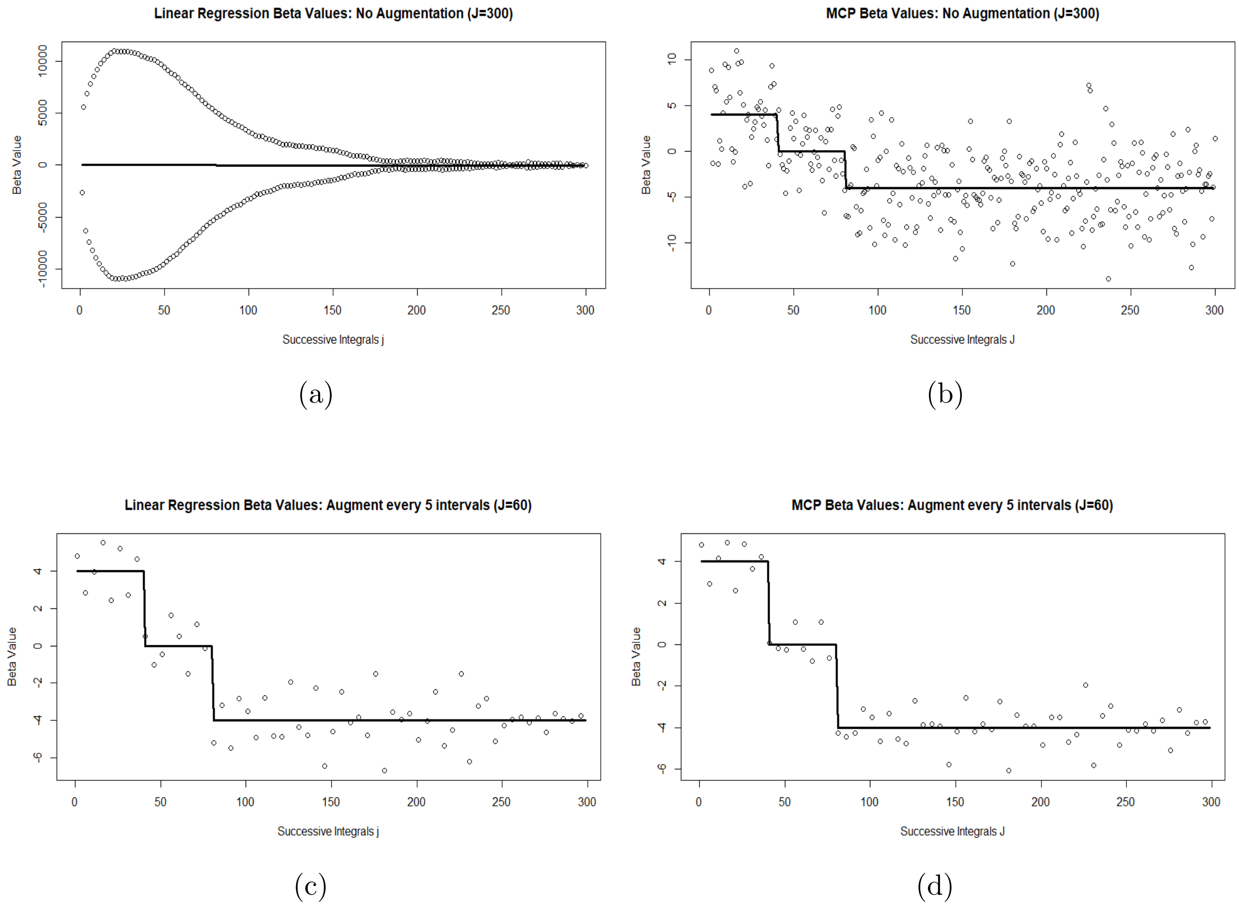
true model we specified as a 3-group model ($K^* = 3$) with $(c_1^*, c_2^*, c_3^*) = (4000, 8000, 30000)$ and effect sizes: $(\beta_1^*, \beta_2^*, \beta_3^*) \in \{(4, 0, -4),\ (2, 0, -2),\ (4, 0, -2)\}$. These simulations were conducted for three different sample sizes $N \in \{100, 250, 500\}$, with results from 500 rounds of simulations summarized in Table 2.1.

2.6.2 Simulation Performance

The proposed FRACT analytic and algorithms have shown both high sensitivity in selecting the correct number of windows ($K^* = 3$) and small bias in the estimation of cutpoint and effect size. For example, in considering Scenario 1 with model parameters $(\beta_1^*, \beta_2^*, \beta_3^*) = (4, 0, -4)$, FRACT selected a 3-group model 96% (N=500), 91% (N=250), and 88% (N=100) of the time over 500 simulations. Additionally, the bias of all the effect size estimates $\hat{\beta}_k' s$ and cutpoints $\hat{c}_k' s$ were low, with respective mean estimates of 3.99, 0.02, and -4.0 and cutpoints 49.92, 79.87, and 300 for the N=250 scenario (and similar strong results for N=500 and N=100). Estimates $\hat{\beta}_2$, however, do have increased variability around the truth in comparison to estimates $\hat{\beta}_1$ and $\hat{\beta}_3$. In Table 2.1, the Empirical Standard Error (ESE) is reported to reflect the precision and stability of the estimation. Average Standard Error (ASE) is not reported since estimates $\hat{\beta}_k$ are dependent on $\tilde{A}_k$, which are related to estimates $\hat{c}_k$'s. Because of this, $\tilde{A}_k$'s are moving across the 500 simulations and thus so are $\hat{\beta}_k$'s.

In each of the simulation scenarios, tuning J was evaluated as part of the strategy to address the high-correlation among AUC $A_j' s$. With initial $J = 300$, there were high mean pairwise correlations: $cor(A_j, A_{j+1}) = 0.998$, $cor(A_j, A_{j+5}) = 0.985$, $cor(A_j, A_{j+10}) = 0.967$. This severe multi-collinearity impaired standard linear regression analysis, producing bifurcate initial estimates $\hat{\beta}_j,\ j = 1, \cdots, J$ as shown in Figure 2.3a. By merging every five successive $A_j' s$, the augmented $\check{A}_j$ variables gave reduced mean pairwise correlations $cor(\check{A}_j, \check{A}_{j+1}) < 0.98$, $cor(\check{A}_j, \check{A}_{j+5}) < 0.90$, $cor(\check{A}_j, \check{A}_{j+10}) < 0.80$. The resultant initial estimates $\hat{\beta}_j$ unveil cleaner patterns than those from non-augmented J. Refer to Figures 2.3b, 2.3d, which suggests that the penalized regression improves separating the $\beta(c)$ function into pieces. The utilization of the MCP penalty seems to help the lowering of estimation bias in the case of multicollinearity.

The Refinement Learning steps, i.e. Edge-Swapping, Merging, and Breaking, were evaluated in the simulation as the sample size N decreases (Table 2.2). In the simulation scenario of $(\beta_1^*, \beta_2^*, \beta_3^*) = (4, 0, -2)$ with N=500, FRACT selected a final 3-group model in 94% of the 500 replicates. This high sensitivity was achieved by the successive improvements given by the sequential multi-step process: first, 58.6% sensitivity in the original 3-group model, then adding 27.4% by Edge-Swapping, 0.6% by Merging, and 7.8% by Breaking. When the sample size decreased to N=250, the sensitivity of the original 3-group model dropped



(a)

(b)

(c)

(d)

Figure 2.3: Comparison of initial estimates obtained from a randomly selected simulation dataset of $\beta_j,\ j = 1, \cdots, J$ with/without the MCP Penalty and with/without augmentation by 5 intervals (J=300, 60 respectively). The true $\beta(c)$ is shown as the piece-wise horizontal function, whose visually flat pattern in (a) is due to the scale of the beta values.

Table 2.1: Simulation Results of 3-Group Model to Evaluate FRACT

Results summarized over 500 replicates including average (Mean) and median (Med.) estimate, empirical standard error (ESE), and percent of correctly selected 3-group models (Sensitivity). Cutoff c_3 is not estimated but included for completeness. Cutpoint values are shown as VM/100 for ease of visualization.

Scenario 1		N = 500			N=250			N=100		
	Truth	Mean	Med.	ESE	Mean	Med.	ESE	Mean	Med.	ESE
β_1	4	4.00	4.00	0.03	3.99	3.99	0.04	3.98	3.99	0.11
β_2	0	0.01	0.01	0.05	0.02	0.00	0.14	0.17	0.03	0.58
β_3	-4	-4.00	-4.00	0.00	-4.00	-4.00	0.00	-3.99	-3.99	0.01
c_1	40	39.98	40.00	0.32	39.92	40.00	0.84	39.27	40.00	3.66
c_2	80	79.99	80.00	0.23	79.87	80.00	0.8	79.14	80.00	3.96
c_3	300	300	300	0.00	300	300	0.00	300	300	0.00
Sensitivity:		96%			91%			88%		
Scenario 2										
β_1	2	2.00	2.00	0.04	1.99	1.99	0.08	1.94	1.94	0.19
β_2	0	0.06	0.01	0.20	0.10	0.03	0.42	-0.12	0.00	0.73
β_3	-2	-2.00	-2.00	0.00	-2.00	-2.00	0.00	-2.00	-2.00	0.01
c_1	40	39.53	40.00	2.53	39.03	40.00	5.51	42.05	40.00	10.48
c_2	80	79.42	80.00	2.25	78.88	80.00	5.01	82.12	80.00	11.05
c_3	300	300	300	0.00	300	300	0.00	300	300	0.00
Sensitivity:		92%			93%			85%		
Scenario 3										
β_1	4	4.00	4.00	0.04	3.99	3.99	0.08	3.95	3.95	0.16
β_2	0	0.02	0.01	0.13	0.01	0.00	0.30	-0.09	-0.03	0.62
β_3	-2	-2.00	-2.00	0.00	-2.00	-2.00	0.01	2.00	-2.00	0.01
c_1	40	39.96	40.00	0.92	40.20	40.00	2.12	40.97	40.00	4.51
c_2	80	79.85	80.00	1.29	80.34	80.00	4.28	82.11	80.00	8.37
c_3	300	300	300	0.00	300	300	0.00	300	300	0.00
Sensitivity:		94%			91%			91%		

dramatically to 29.4%, but remarkably, Edge-Swapping added 31.6% sensitivity, Merging contributed 1.6% sensitivity, and Breaking strikingly boosted sensitivity 28.0%, reaching the final 91% sensitivity. The Refinement Learning steps increase the sensitivity of correctly selecting the activity windows and determining the $c_1^*, \cdots, c_K^*$ cutpoints.

Table 2.2: Sensitivity Simulation Results for 3-Group Model

Results summarized over 500 replicates and signify the importance of FRACT's multiple steps. The percentages represent the added sensitivity based on the corresponding Refinement Learning process. E.g., the row "3 Gps: Swapping" represents the percent of 3-group models selected after undergoing systematic swapping of the cutpoints, as described in Algorithm 3 in Section 2.4.

	β = (4, 0,-4)			β = (2, 0,-2)			β = (4, 0,-2)		
Selection	500	250	100	500	250	100	500	250	100
2 Groups	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
2 Gps: Swapping	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2 Gps: Merging	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3 Groups	79.6	58.8	22.6	36.4	18.2	3.6	58.6	29.4	2.8
3 Gps: Swapping	13.4	23.6	41.2	37.2	44.0	23.8	27.4	31.6	25.4
3 Gps: Merging	0.0	0.6	1.2	0.2	1.4	2.0	0.6	1.6	0.4
3 Gps: Breaking	3.4	7.6	22.6	17.8	29.4	55.8	7.8	28.0	62.2
4 Groups	0.2	0.8	0.2	0.4	0.4	0.2	0.4	0.4	0.2
4 Gps: Swapping	0.2	0.6	3.4	2.2	1.0	1.4	1.4	2.4	1.6
4 Gps: Merging	0.2	0.0	0.4	0.2	0.4	0.2	0.0	0.4	0.4
4 Gps: Breaking	3	7.6	8.2	5.6	5.2	11.8	3.8	6.2	6.8
5 Groups	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5 Gps: Swapping	0.0	0.4	0.2	0.0	0.0	0.2	0.0	0.0	0.2
5 Gps: Breaking	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

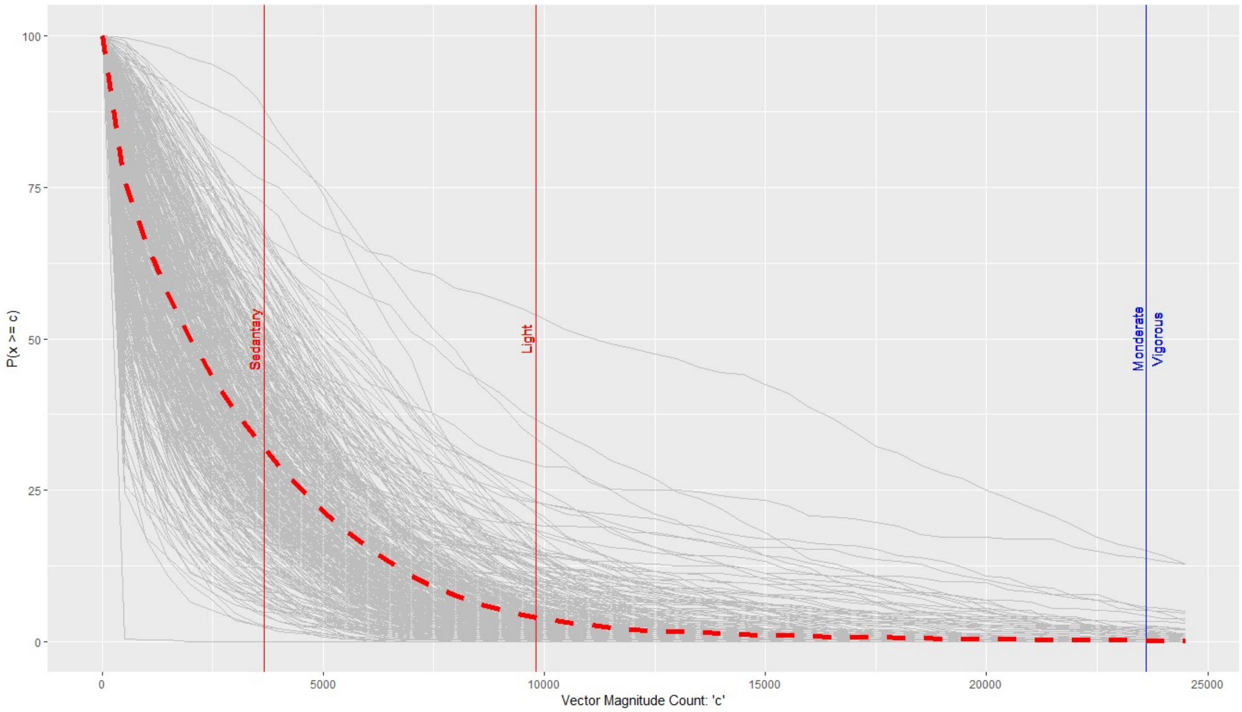


Figure 2.4: OTCs for 354 ELEMENT subjects, representing functional covariates. The vertical lines represent Chandler's cutoffs (Chandler et al., 2016) for activity levels (Sedentary, Light, Moderate, Vigorous). The relative shape of these OTCs reflect the subject's activity profile.

2.7 Data Analysis

We now apply the proposed FRACT methodology to investigate the functional association between PA and epigenetic age (DNAm). As discussed in Section **??**, this analysis incorporated a vector of covariates (Z) with centered chronological age, sex, pubertal status (based on Tanner staging) and lead exposure (measured in micrograms of lead per deciliter of blood, or $\mu g/dL$) for each subject. We had complete accelerometry and covariate data for 354 ELEMENT subjects (172 male, 182 female), with mean(SD) age of 13.7(1.9) years and mean(SD) lead exposure of 3.17(3.33) $\mu g/dL$.The majority, 332, of the subjects completed puberty based on Tanner staging standards. Figure 2.4 shows OTCs for the 354 subjects, representing PA profiles during weekends between 4:00PM - 10:00PM, selected with the rationale that this period reflects a block of time they have more control over their activities. Among multiple epigenetic age clocks [Horvath, 2013] we are interested in two specific ones: Horvath's AgeSkinBlood Clock [Horvath et al., 2018], and Levine's PhenoAge Clock [Levine et al., 2018]. For each of these outcomes, we fit a scalar-on-function regression model with scalar covariates.

The FRACT analytic began the analysis by setting $J = 300$, each interval covering 100 VM counts. We detected two activity windows of interest for each of the epigenetic age outcomes, albeit with different cutpoints and effect sizes. Association parameters, standard errors, and p-values are estimated from the final linear model. Of note, the p-values are conditioned on the cutoffs found by the FRACT analytic. Table 2.3 summarizes the data analysis results. In considering Levine's PhenoAge epigenetic clock, we detected activity window VM > 8000 count to have a significantly negative association with epigenetic age; the higher the AUC in the VM > 8000 range, the lower the PhenoAge. As higher AUC at the end of an OTC represents more activity within that range, this finding may be interpreted as: more activity in VM > 8000 range is related to younger epigenetic age, i.e. more time in higher PA levels is related to slower biological aging.

In considering Horvarth's AgeSkinBlood clock, our method identified other important activity windows. Specifically, Table 2.3 suggests that FRACT determined a window of activity VM ≤ 1500 counts to be positively associated with this epigenetic age. As larger AUCs at the beginning of the OTC reflect more time spent above that activity level, this positive association implies that: less time spent in VM range ≤ 1500 counts is associated with a higher epigenetic age. Correspondingly, more time in the low-activity time window is related to lower epigenetic age, or slower biological aging. As the epigenetic age clocks are calculated from different sets of methylation variables, reflecting different aspects of biological aging, these differences are biologically meaningful and not surprising.

Encouragingly, as shown in Tables 2.3(a)-(c), there are agreements between our detected cutoffs and the previously published Crouter [Crouter et al., 2015] and Chandler [Chandler et al., 2016] cutoffs. The changepoint for "Moderate" activity using Chandler or Crouter cutoffs are 9805 and 7320 respectively; notably, our PhenoAge detected changepoint of 8000 is within this range. The SkinBloodAge activity range of VM counts ≤ 1500 is similar to Crouter's cutpoint for Sedentary behavior (VM ≤ 1200).

While the FRACT-identified cutpoints enjoyed some agreement with the Chandler/Crouter pre-determined cutpoints, the proposed FRACT methodology did result in stronger statistical results. As seen in Tables 2.3(b)-(c), the Chandler cutpoints detected no significant associations with Levine's PhenoAge, nor Crouter with Horvath's AgeBloodSkin, at a significance level of $p = .05$. In the instances where the pre-determined cutpoints do detect significant associations, the FRACT-determined activity windows demonstrate more statistically significant associations than their Crouter/Chandler counterparts. For example, while Crouter's "Moderate-to-Vigorous" activity window, with VM count interval (7400,30000) was negatively associated with PhenoAge ($p = .04$), the FRACT-determined window covering VM counts ($8100 - 30000$) was more strongly negatively associated with

Table 2.3: FRACT Data Analysis Results

(a) FRACT results in the scalar-on-function regression model. (b-c) Results if use published adolescent cutoffs from Chandler and Crouter. Standard analyses with these cutoffs consider 3 activity windows for "Sedentary" Physical Activity (PA), "Light" PA, and Moderate-to-Vigorous PA "MVPA".

(a) FRACT

PhenoAge	Cutpoints/100	Coef	SE	p-value
Window 1	(0-80)	0.86	1.42	0.55
Window 2	(81-300)	−1.23	0.54	0.02
Sex (Base: Male)		296.20	185.22	0.11
Chronological Age		1.67	0.14	<.001
Lead Exposure		25.54	27.17	0.35
Pubertal Status		154.53	335.71	0.66
AgeBloodSkin				
Window 1	(0-15)	4.73	1.90	0.01
Window 2	(16-300)	−0.11	0.12	0.33
Sex (Base: Male)		6.67	51.40	0.90
Chronological Age		0.86	0.04	<.001
Lead Exposure		−6.32	7.54	0.40
Pubertal Status		−96.39	92.94	0.30

(b) Chandler Pre-set Cutoffs

PhenoAge	Cutpoints/100	Coef	SE	p-value
Window 1 (Sedentary)	(0-36)	1.22	4.32	0.77
Window 2 (Light)	(37-98)	0.06	2.96	0.98
Window 3 (MVPA)	(99-300)	−1.23	0.65	0.06
Sex (Base: Male)		297.30	186.75	0.11
Chronological Age		1.67	0.14	<.001
Lead Exposure		25.63	27.27	0.35
Pubertal Status		153.81	337.44	0.65
AgeBloodSkin				
Window 1 (Sedentary)	(0-36)	2.74	1.19	0.02
Window 2 (Light)	(37-98)	−0.94	0.28	0.25
Window 3 (MVPA)	(99-300)	−0.01	0.18	0.94
Sex (Base: Male)		3.39	51.69	0.93
Chronological Age		0.86	0.04	<.001
Lead Exposure		−6.34	7.55	0.40
Pubertal Status		−89.42	93.41	0.34

(c) Crouter Pre-set Cutoffs

PhenoAge	Cutpoints/100	Coef	SE	p-value
Window 1 (Sedentary)	(0-12)	5.36	12.75	0.67
Window 2 (Light)	(13-73)	0.14	2.87	0.96
Window 3 (MVPA)	(74-300)	−1.14	0.55	0.04
Sex (Base: Male)		291.89	185.90	0.12
Chronological Age		1.67	0.14	<.001
Lead Exposure		26.33	27.28	0.33
Pubertal Status		161.11	336.40	0.63
AgeBloodSkin				
Window 1 (Sedentary)	(0-12)	6.64	3.53	0.06
Window 2 (Light)	(13-73)	−0.35	0.79	0.65
Window 3 (MVPA)	(74-300)	−0.07	0.15	0.60
Sex (Base: Male)		8.77	51.48	0.86
Chronological Age		0.86	0.04	<.001
Lead Exposure		−6.21	7.55	0.41
Pubertal Status		−95.13	93.15	0.31

PhenoAge ($p = .02$).

To assess the stability of this supervised learning, we conducted 5-fold cross-validation. To achieve this, we split the ELEMENT data into five equally sized subsets. We trained the cutpoints with four subsets and then fit the associated step-wise model with the remaining testing subset. Both epigenetic age clocks demonstrated stable results, with the results for the AgeBloodSkin clock discussed here. In this case, all training models detected a 2-group model, with the end cutpoint of Window 1 at $(25, 15, 15, 15, 15)$ demonstrating strong stability in changepoint detection. In the associated parameter estimations of Window 1 and Window 2 in the testing datasets, the mean (sd) parameter estimates are 5.19(1.45) and $-1.13(2.40)$ respectively. This assessment demonstrates that the significant association between Window 1 and the AgeBloodSkin clock shown in Table 2.3 is stable.

2.8 Discussion and Conclusion

In this chapter we developed the Functional Regularized Adaptive Changepoint-detection Technique (FRACT), to transform functional accelerometer data collected from wearable devices into knowledge on PA's effect on human biological aging. This learning analytic detects changepoints to define critical windows of activity, while accounting for covariates of interest. Such an informatics toolbox can be applied to analyze the relationship of functional digital features with outcomes.

It is worth highlighting a key technical advance FRACT's supervised learning: unlike methods discretizing functional features via existing cutoffs regardless of specific outcomes, FRACT provides a simultaneous operation of supervised changepoint detection and functional association parameter estimation. Thus, this data analytic is adaptive to data collections and populations under investigation. In the investigation for the influence of PA on biological aging, when applied to different study populations (e.g. adolescent or adult) or to different wearable devices (e.g. Actigraph or Fitbits), FRACT provides data-driven solutions tailored to characteristics of the study. This avoids potential bias in data processing and data analyses by applying some pre-set cutoffs (e.g. Chandler's values for children) to different populations. Additionally, this flexibility demonstrates the value of the FRACT analytic in the analysis of future wearable devices. In the ever-evolving world of wearable devices, there are constantly new devices or sensors available. The applications of the proposed FRACT methodology are not restricted to accelerometer sensors; rather it can easily be applied to other devices including biomedical/smart-Health devices and be used as a decision process in biomedical engineering devices. In such a role, it can help translate data collected from existent/future sensors into decision-making knowledge. As such physiological sensors can have a great potential impact the future of health-monitoring and intervention, the translational role FRACT plays in turning high-frequency time-series data into decision-making knowledge is invaluable to practitioners.

FRACT has demonstrated flexibility and reliability in identifying changepoints/critical windows, and estimating functional association parameters. This greatly benefits our analysis of detecting critical changepoints of PA related to epigenetic age. When Levine's PhenoAge clock was used as the age outcome, we found that an increase in mid-range PA is associated with younger epigenetic age. This epigenetic clock was specifically created to reflect age and disease-related phenotypes, such as inflammation and physical functioning [Levine et al., 2018]. In this, the direction of association of increased PA and lower epigenetic age makes intuitive sense. On the other hand, when the AgeSkinBlood clock was used in the analysis, we detected the benefit of sedentary PA, which could reflect more time spent indoors versus outdoors in the sun, which in a warm geographical area such as Mexico City could have a beneficial effect on skin aging. This latter analysis focuses on epigenetic variables in skin/blood cells including fibroblasts, which deal with the structural components of skin cells. It is interesting to reach an agreement between the FRACT-identified cutoffs and the Chandler/Crouter adolescents cutoffs. As Chandler, Crouter and FRACT investigated populations with similar underlying characteristics, the agreement among identified cutpoints signify FRACT's reliability.

FRACT requires careful tuning steps in order to achieve optimal performance. Assessing the level of multi-collinearity when calculating initial parameter estimates is critical to overall performance. Our experience with simulation experiments demonstrated that ignoring the high correlation among AUCs in initialization steps has a detrimental effect on overall performance. However, when J is tuned such that the correlations are below the threshold provided in Section 2.4, the issues with multi-collinearity are mitigated. These simulations also highlighted the importance of the Refinement Learning steps, particularly Edge-Swapping, especially with decreasing sample size. While the Refinement Learning strategies of *Edge-Swapping*, *Merging*, and *Breaking* can be conducted in any order, we recommend first focusing on Edge-Swapping, leading to the most favorable increase in FRACT's sensitivity. From a computational standpoint, FRACT is fast. The most time-consuming step is conducting the initial changepoint estimation via HMM due to the need of repeated fitting to aid numeric stability. We found 10 repetitions of HMM fitting to be sufficient, though if multi-collinearity is less of an issue in a specific analysis the number of repetitions may be reduced. In the data analysis, computation time for determining the final model using 1 CPU was less than 15 seconds, with a sample size of $N = 354$ and potential model sizes of $K = (2, 3, 4)$.

While wearable accelerometer devices provide many advantages to measuring objective PA, particularly versus self-reported data, there are some inherent limitations. First, though single-sensor accelerometer devices are informative on PA intensities, they are less equipped to differentiate between specific activity types than multi-sensor devices. Such multi-sensor devices can provide valuable additional measurements such as skin temperature, heart rate, and blood volume. If using a device with these additional variables, FRACT could either utilize a new functional covariate or include the data as covariates of interest. However, these devices, while useful, are more expensive and can be barriers in scientific study and personal use. Additionally, while wrist-worn devices are often used in studies of PA due to high-compliance and feasibility, they have some disadvantages versus devices worn on other body location, such as ankle or hip. For example, wrist-worn devices can capture arm movement during sedentary activities as PA, whereas it could fail to identify activity such as biking as PA due to the stationary placement of wrists during this activity [Liu et al., 2021, Gao et al., 2021]. Lastly, some studies that aim to classify PA will validate their method with in-lab studies, aligning wearable device readings with observed PA intensities. [Sevil et al., 2020, van Loo et al., 2017]. Future research in this FRACT methodology could aim to further validate the analytic by such a clinical setting.

A limitation in our current application of FRACT is that it focused on PA over a single time frame. Some research has shown that the timing of activity, not just the magnitude, can impact certain health outcomes [Minaeva et al., 2020]. Future work may extend this analysis by considering time-specific OTCs and adaptively detecting corresponding activity ranges associated with the health outcome of interest. Additionally, time of year may impact the effect of functional PA levels on health outcomes. For example, in geographic areas with significant seasonal weather differences, it may be important to account for these changes [Garriga et al., 2021]. While this study considered subjects living in a limited weather variability area, for studies with variable weather patterns a researcher can include a covariate or interaction term between PA and season to address the different functional relationships to account for seasonal impact. Another extension could involve a functional longitudinal framework to understand the influence of repeatedly measured functional accelerometer data on longitudinal health outcomes. The previous methodological advancements considered PA data in a seven-day period; with repeated measurements we can study how changing PA patterns from late-adolescence into early adulthood affects certain longitudinal health outcomes of interest. While this chapter focused on linear relationship, FRACT could be extended to non-normal and non-linear models, such as logistic regression with binary outcomes, and Cox regressions with time-to-event outcomes.

CHAPTER 3

Regularized One-Step Estimation of Changepoint and Functional Parameter in Functional Accelerometer Data Analysis

3.1 Introduction

3.1.1 Biological Aging and Epigenetic Age

Biological aging is a growing area of research that seeks to understand the variation in how people age biologically, as opposed to chronologically, or are affected by age-related diseases. Epigenetic age is a biological concept that refers to the biological age of an individual, as determined by the epigenetic modifications that occur on their DNA [Horvath, 2013]. The term "epigenetic" refers to modifications that occur on the DNA molecules that do not change the actual DNA code sequence, but can alter the way genes are expressed. This is an emerging field of research that has gained much attention in recent years due to its potential to provide insight into the aging process and the development of age-related diseases. Thus, epigenetic age can act as a useful biomarker of an individual's overall state of health and allow for personalized or preemptive health interventions [Marioni et al., 2015]. Various epigenetic age calculators consider different groups of DNA methylation (DNAm) alterations along different areas of the genome to deliver a predicted epigenetic age, and are hosted online [Horvath, 2013]; see Horvath, 2013, among others.

While much of the research into epigenetic age has focused on adults, there is also interest in studying epigenetic age in children. This is because epigenetic modifications can be influenced by a range of environmental factors, including prenatal and early life experiences, which may impact later-life health outcomes. As these childhood environmental and experiential factors can be observed in changes in the the DNA methalome, they are thus reflected in epigenetic age. Studies show that children and adolescents (age 0-18) undergo the fastest and most dynamic rate of growth and DNAm changes [Wu et al., 2019, McEwen et al., 2020]. As these childhood environmental and experiential factors can be observed in changes in the the DNA methalome, they are thus reflected in epigenetic age. One study [Wiklund et al., 2019] found that maternal smoking during pregnancy was associated with accelerated epigenetic aging in offspring. In this study, data from five prospective birth cohorts were used to examine the relationship between maternal smoking during pregnancy and DNA methylation patterns in offspring, and children whose mothers smoked during pregnancy were found to have their DNA methylation patterns consistent with accelerated epigenetic aging. Research into epigenetic age in children has also shown that it may be a useful tool for predicting future health outcomes. For example, [Huang et al., 2019] found that epigenetic age acceleration in adolescents was associated with risk of cardiovascular disease in middle-age.

By better understanding the relationship between epigenetic modifications and childhood experiences, researchers may be able to develop interventions to prevent or mitigate the negative health effects of early life stressors. However, further research is needed to fully understand the complex relationship between genetics, epigenetics, and environmental factors in shaping health outcomes across the lifespan. An important investigation of scientific interest is to assess the relationship between the experiential determinant of PA in adolescence with biological aging. Research and conventional wisdom suggest that increased PA may slow epigenetic aging [Kankaanpää et al., 2022, Quach et al., 2017]. By promoting PA in children, we may be able to improve not only their current health outcomes but also their long-term health outcomes by slowing down the aging process. The focus of this chapter is to investigate the association of epigenetic age with objectively measured functional PA as captured by wearable devices.

3.1.2 Wearable Devices and Accelerometer Data

Wearable technologies use sensors worn over continuous time-periods to collect subjects' personal data. Notably, these devices can conduct automatic real-time data collection in high frequency and track physiological variables and clinical symptoms outside of clinical environments. In providing this high-frequency, personalized time-series data, wearable devices are promising technologies to promote smart-Health care management and precision medicine. Additionally, the data collection can be relatively cheap, convenient, and flexible in variable environments, which increases their popularity in both research and personal use.

While their popularity and potential usefulness are growing quickly, the ability to efficiently and effectively glean statistically-robust information from wearable devices is slower to catch up. The data retrieved from these technologies present challenges in data analysis, due to their inherent noisy nature, the non-generalizability of methodologies, and high computational requirements. These challenges motivate the need for statistical innovations to enable the wide-scale use of such wearable smart-Health devices in research related to improving quality of life.

Accelerometers are a type of wearable device that measures continuous PA and movement data, providing real-time, large-scale, personalized information on an individual's PA patterns. They capture raw gravitational acceleration data that are then processed into activity "counts" over specific "epochs", or lengths of time [Chen and Bassett, 2005]. The count levels reflect the relative intensity of activity, with higher values indicating more intense exertion. For tri-axial accelerometers, the three-dimensional count information at each time point is often summarized into a one-dimensional summary value of Vector Magnitude (VM), with $VM = \sqrt{axis1^2 + axis2^2 + axis3^2}$. Figure 2.1 depicts continuous time-series VM count data for an individual from our motivating data detailed in Section 2.2. A typical analysis may then categorize these count values into activity levels of interest, such as Sedentary, Light, and Moderate-to-Vigorous Activity (MVPA), based on certain pre-specified activity thresholds, and assess the association between amount of time spent in each activity level and a health outcome of interest [Freedson et al., 2005, Crouter et al., 2015, Chandler et al., 2016]. As these types of analyses are dependent on pre-specified threshold values, which in turn need to be validated for each specific device type (e.g. Fitbit, Apple Watch) and underlying population characteristics (e.g. age, sex), they suffer from a lack of generalizability and flexibility. Such shortcomings make them impractical within a smart-Health setting. Thus, it is beneficial to conduct feature extraction from PA accelerometer data using a more generalizabale approach to relax or even eliminate the dependence on pre-fixed cutoffs.

3.1.3 Study Objectives

This need of data-adaptive cutoffs motivates the statistical objective of this chapter: to develop a generalized, functional-focused approach to analyze PA data with the aim to free the dependence on pre-determined PA categorizations. Our new approach will allow the data to adaptively determine the change-points and different PA ranges of interest together with our primary task of assessing the association of detected PA ranges with health outcomes of interest. To this end, we consider actigraphy data under the purview of OTCs. This method of analyzing activity data involves a summary curve which describes the proportion of time an individual spends at or above successive activity levels [Bogachev and Ratanov, 2011]. Introduced in Section 2.3, these OTCs act as functional predictors in a Functional

Data Analysis (FDA) paradigm.

We consider a supervised learning framework of scalar-on-function models in which we develop a simultaneous operation of estimation and changepoint detection (or clustering). The scalar-on-function regression allows us to investigate functional associations between health outcomes, specifically epigenetic age, and OTCs adjusted by confounding factors. In particular, we propose an L_0 regularization approach to determine cutoff points adaptively. Until relatively recently, L_0 regularization and discrete optimization has been less of a focus verses the L_1-related continuous optimization approaches as it was deemed computationally impractical. However, with recent advances in algorithmic and numeric capabilities, discrete optimization is a feasible and powerful tool [Bertsimas et al., 2016]. We implement the modern optimization methods to functional analysis, by means of Mixed Integer Optimization (MIO), to accurately detect critical activity windows of interest, conducting regularization in a supervised learning framework. This MIO-based optimization is demonstrated to be computationally feasible and scalable to practically-sized problems of interest.

The organization of this chapter is as follows. We review the motivating cohort study and functional OTC variables in Section 3.2, while Section 3.3 compares existing and proposed functional model formulations. Section 3.4 introduces MIO and presents its formulation for our scalar-on-function statistical analysis, with a discussion of theoretical guarantees in Section 3.5. In Section 3.6 we explore numeric experiments illustrating the capabilities of this approach, while Section 3.7 provides a detailed analysis with our motivating data, exploring the functional associations between OTCs and epigenetic age. Lastly, we discuss the merits, limitations, and potential extensions of this discrete optimization approach in Section 3.8. Some additional numerical results are included in Appendix A.

3.2 Motivating Cohort Study

This work is motivated by the Early Life Exposures in Mexico to Environmental Toxicants (ELEMENT) longitudinal birth cohort study involving mother/child dyads in Mexico City. Refer to a review paper by [Perng et al., 2019] for details regarding the study as a whole, and Chapter 2 Section 2.2 for specific details related to accelerometry and epigenetic age.

3.2.1 Occupation Time Curves (OTCs)

The subject's PA profile may be summarized by an OTC that entirely translates the high-frequency time-series data to a single functional curve. OTCs greatly reduce the device's inherent noise while retaining key features of activity. A detailed description of OTCs and their construction is included in Chapter 2, with related Figure 2.2a illustrating the numerical OTC construction procedure. The resulting OTC is denoted as $X(c),\ c \in C = [0, 300 \times 10^2]$ throughout the rest of this chapter.

As shown in Figure 2.2b, the OTC's shape provides key information of variability on the subjects' PA profiles, with the curves of more and less active individuals taking distinct shapes. Here, "more" or "less" active is determined by relative levels of high activity counts. Figure 3.3 illustrates OTC variability in its representation of 354 OTCs from ELEMENT.

3.3 Model Formulations

3.3.1 Scalar-on-Function Analysis with OTC Functional Predictor

Motivated by the inherent variability of the OTCs, it is natural to analyze the features under the auspices of FDA, considering the OTCs as a functional covariate with a varying association on a health outcome of interest. The goal of such an analysis is to understand the functional association between OTC and a health outcome, particularly identifying critical changepoints and critical activity windows. For a certain scalar outcome Y, the standard scalar-on-function model is expressed as:

$$Y =< X, \theta > + Z^T\alpha + \epsilon = \int_C \theta(c)X(c)\ dc + Z^T\alpha +\ \epsilon, \tag{3.1}$$

where $Y \in \mathbb{R}^{n\times 1}$; $X(c)$ is the functional OTC defined on $C \subset \mathbb{R}$; Z is a q-dimensional vector of confounders with corresponding parameter vector α; ϵ is the error term with mean 0 and variance σ^2; and $< a, b >$ depicts the inner product of two square-integrable functions, namely $\int_C a(c)b(c)dc$ with $\int_C a^2(c)dc < \infty$ and $\int_C b^2(c)dc < \infty$.

The goal of a scalar-on-function model as defined above is to estimate the functional parameter $\theta(c)$. In this case, our goal is to discretize the continuous functional parameter estimate $\theta(c)$ as a piece-wise function with jump points, thereby effectively defining windows of PA levels by fusing the $\theta(c)$ of adjacent count ranges with similar effects on the outcome. Figure 3.1 illustrates this goal.

In other words, for a given K number of PA windows (e.g. $K = 3$), we aim to estimate both the jump points $c_1, \cdots, c_{K-1}$ and step-function parameter values simultaneously. In this way, we reparamaterize the functional parameter $\theta(c)$ into a step-function parameter represented with $\beta_1, \cdots, \beta_K$ as the respective coefficients for activity windows $[0, c_1], (c_1 + 1, c_2], \cdots, (c_{K-1}, c_{max}]$. Here, c_{max} denoted the maximum activity count considered in the analysis.

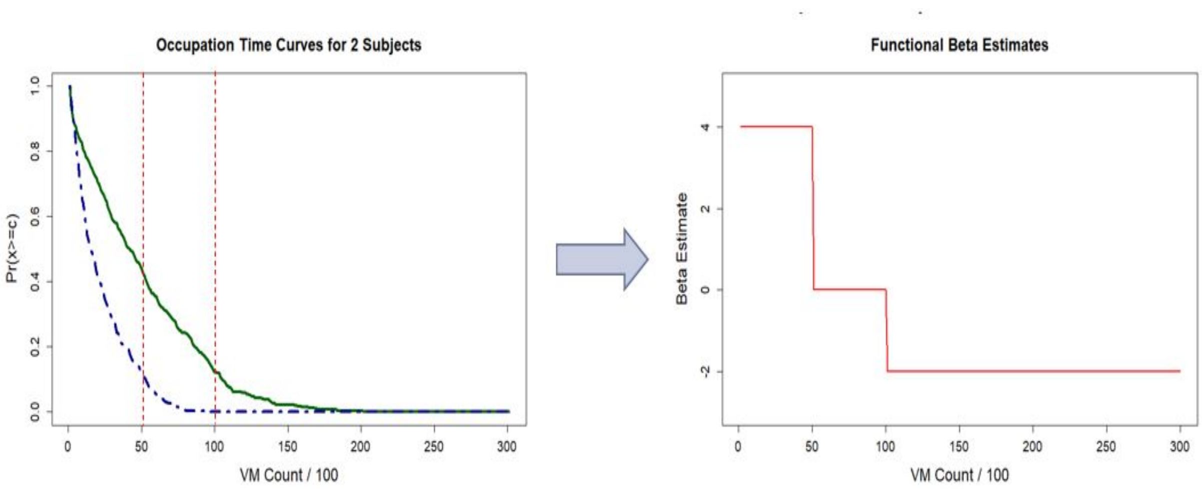


Figure 3.1: Left panel: Two realized OTCs with vertical dashed lines representing the cutoffs for activity window ranges of interest. Right panel: Example of a non-constant functional $\beta-$ parameter as a step-function to estimate the association of PA in selected activity-count windows with a health outcome of interest

To achieve this goal, we first discretize each OTC into many small segments by dividing the interval C into J-many small successive intervals with a grid $c_0 = 0, c_1, \cdots, c_J = 30,000$, with $C = [0, c_1] \cup_{j=2}^{J} (c_{j-1}, c_j]$. Within each interval j, we treat $\theta(c)$ as a constant parameter θ_j, which leads to Equation (3.2) given as follows:

$$\begin{aligned}\int_C \theta(c) X(c)\, dc + Z^T \alpha &= \sum_{j=1}^{J} \int_{c_{j-1}}^{c_j} \theta(c) X(c)\, dc + Z^T \alpha \\ &\approx \sum_{j=1}^{J} \theta_j \int_{c_{j-1}}^{c_j} X(c)\, dc + Z^T \alpha \\ &:= \sum_{j=1}^{J} \theta_j A_j + Z^T \alpha,\end{aligned} \tag{3.2}$$

where $X(c)$ is defined as above; A_j denotes the Area Under the Curve (AUC) over interval $(c_{j-1}, c_j]$ or $A_j = \int_{c_{j-1}}^{c_j} X(c)\, dc$; and Z is a q-dimensional vector of confounders with corresponding parameter vector α. Of note, while $X(c)$ is monotonically decreasing due to the inherent structure of OTCs, there is no restriction of monotonicity on $\theta(c)$. Unlike the conventional functional regression analysis, in the same spirit of categorization shown in Figure 3.1, our analytic aim is to fuse similar adjacent parameter θ_j's together in order to estimate a K-group sized step function with parameters β_k for $k = 1, \cdots, K$. This results in a final estimate model: $\sum_{k=1}^{K} \beta_k A_k + Z^T \alpha$, with A_k denoting AUC over interval (c_{k-1}, c_k) or $A_k = \int_{c_{k-1}}^{c_k} X(c)dc$. The resulting step function for $\theta(c)$ is deemed for desirable results of scientific interest, including both critical activity windows and assessing their influence on the outcome, as well as and their interpretability.

3.3.2 Existing L_1 Regularization Approaches

There are existing methods applicable to carry out the parameter fusion on θ_j, among which Fused Lasso [Tibshirani et al., 2005] and Hidden Markov Model (HMM) [Rabiner and Juang, 1986] are of great popularity. However, such an L_1 penalization approach, like Fused Lasso, have known computational issues especially when faced with high multi-collinearity. L_1 penalization is known to induce bias in the estimation due to its nature of penalizing larger coefficients more than smaller coefficients [Bertsimas et al., 2016]. While this bias often can be controlled via various correction methods (such as adaptive lasso) [Candès et al., 2008, Candès and Plan, 2009, Zou, 2006], when there is severe multi-collinearity among predictors the bias can become out of control and may produce misleading results. Indeed, with the OTCs, the A_j variables experience severe multi-collinearity; for example, the mean pairwise correlations between AUC variables $A_j's$ under $J = 300$ from our motivating data were: $cor(A_j, A_{j+1}) = 0.998$, $cor(A_j, A_{j+5}) = 0.985$, $cor(A_j, A_{j+10}) = 0.967$. This unduly high multi-collinearity presents a great challenge to the Fused Lasso approach and introduces mis-specifications in both changepoint detection and parameter estimation.

To demonstrate the inability of the L_1 regularization approach, which fails to accurately conduct changepoint and parameter estimation, we consider a simulation experiment with functional OTC variables with $J = 300$ and correlation patterns as described above, analyzed under a scalar-on-function linear model as described in Equation (3.2) with a null covariate matrix Z. This simulation experiment used three activity windows (A_1^*, A_2^*, A_3^*) with corresponding end cutpoints $(c_1^*, c_2^*, c_3^*)) = (40, 80, 300)$ and parameters $(\beta_1^*, \beta_2^*, \beta_3^*) = (4, 0, -4)$, as well as normally distributed error term with mean 0 and variance 1. Given this relatively easy case (i.e. big between-window gaps), when conducting changepoint detection and parameter estimation of the piece-wise functional $\theta(c)$ under the fused lasso approach using the R package `glasso`, both estimates of the changepoints and the associated parameters are severely biased, as shown in Figure 3.2 (the right panel) and Table 3.1. We obtain similar poor results even when reducing the collinearity to $cor(A_j, A_{j+1}) = 0.98$, $cor(A_j, A_{j+5}) = 0.90$, and $cor(A_j, A_{j+10}) = 0.80$ by setting $J = 60$; see Table 3.1. In this example, it is clear that we have undesirable results; the cutpoint $\hat{c}_1$ is over-estimated, leading to mis-specified cardinalities of activity intervals $\hat{A}_1$ and $\hat{A}_2$ as well as negatively-biased estimates of both

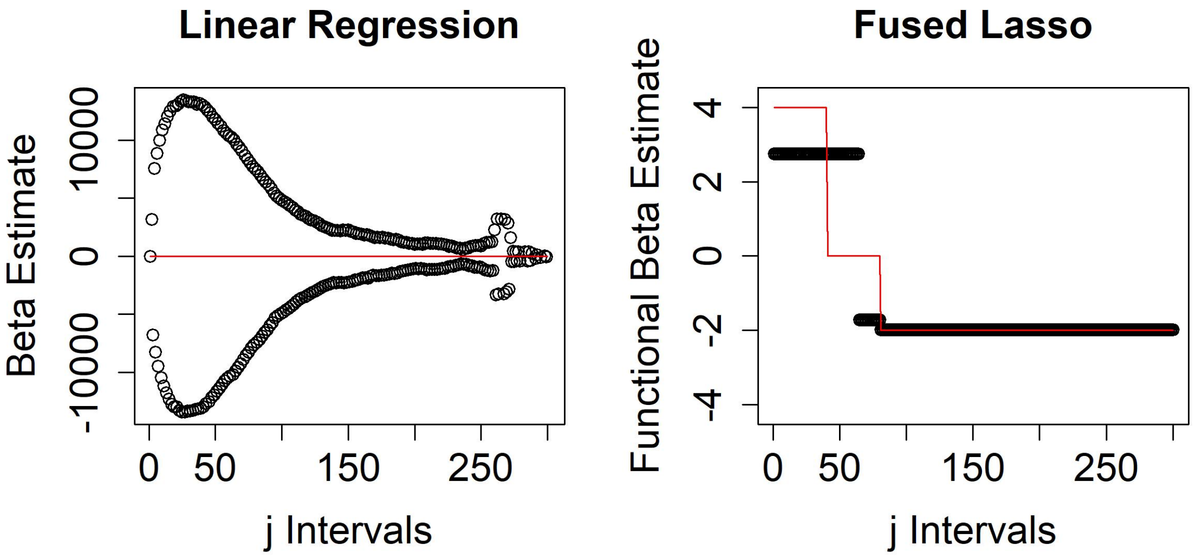


Figure 3.2: Estimates of θ_j coefficients from two different standard analysis approaches including multiple linear regression (left) and an L_1 Regularization approach of Fused Lasso (right). The respective model parameter estimates are represented by the black circles, while the true θ_j values are represented as the red step function.

$\hat{\beta}_1$ and $\hat{\beta}_2$. Additionally, the left panel of Figure 3.2 shows the performance multiple linear regression using the R package `lm` where unduly large discrepancies from the true values are apparent due to the curse of highly correlated predictors. Note that in this case we a large sample size with $N > J$ so we are able to conduct the multiple regression analysis for this comparison. This motivates us to consider an alternative solution, and after analyzing the same model using an L_0 penalization approach, we find that such bias can be reduced to almost zero. The detail is included in Section 3.4.

3.3.3 Integer Programming and L_0 Penalization

Under a modified L_0 optimization strategy, we can simultaneously conduct fusion via changepoint detection and parameter estimation in a one-step approach. Based on a repertoire of literature, the L_0 approach has been shown to be robust against bias and multi-collinearity [Bertsimas et al., 2016, Bertsimas and Shioda, 2009, Bertsimas et al., 2020]. The standard L_0 penalization with constraints on the number of non-zero parameters is not flexible enough to solve our dual analytic goals of changepoint detection and parameter fusion in our analysis; rather, we propose a modified L_0-fusion method for constrained optimization.

A straightforward explanation of standard discrete optimization using L_0 penalization

Table 3.1: Simulation Results of a 3-group Model with $N = 500$ and number of intervals $J = 60, 300$ demonstrate the performance of Fused Lasso, summarized over 500 replicates. Results include average estimate (Mean), median estimate (Med.), and empirical standard error (ESE). Cutpoint values are represented as VM/100.

		J = 60			J = 300		
	Truth	Mean	Med.	ESE	Mean	Med.	ESE
β_1	4	2.97	2.83	0.46	2.87	2.77	0.41
β_2	0	−0.58	−0.31	0.98	−0.79	−0.41	1.19
β_3	−4	−3.98	−3.99	0.01	−3.98	−3.98	0.01
c_1	40	57.75	60.00	9.50	59.47	60.00	9.04
c_2	80	78.40	80.00	2.45	78.47	79.00	2.22

is by means of the the best subset problem. Suppose we have a linear regression model: $\boldsymbol{y = A\theta + Z\alpha + \epsilon}$ where $\boldsymbol{y}$ is an $n \times 1$ response vector, $\boldsymbol{A}$ is an $n \times J$ design matrix $\in \mathbb{R}^{n\times J}$, and $\boldsymbol{\theta}$ is a $J \times 1$ vector of regression coefficients $\in \mathbb{R}^{J\times 1}$. It is often advantageous, particularly in cases of $J > n$, to estimate a sparse parameter vector $\boldsymbol{\theta}$. The best subset problem constrains the level of sparsity by restricting the set of non-zero regression estimates to a maximum cardinality, of say k [Miller, 2002]. This can be expressed as:

$$\min_{\boldsymbol{\theta}} \|\boldsymbol{Y} - \boldsymbol{A\theta} - \boldsymbol{Z\alpha}\|_2^2, \quad \text{subject to} \quad \|\boldsymbol{\theta}\|_0 \leq \boldsymbol{k}, \tag{3.3}$$

where $\|\boldsymbol{\theta}\|_0 = \sum_{i=1}^{J} 1(\theta_i) \neq 0$, or the L_0-norm of $\boldsymbol{\theta}$, with $1(\cdot)$ representing an indicator function. Thus, $\|\boldsymbol{\theta}\|_0$ effectively counts the number of non-zero regression coefficients, and is constrained to maximum cardinality k. As this formulation with discrete constraints has historically been considered computationally intractable in standard approaches [Natarajan, 1995], the best subset problem is often estimated via continuous constraint surrogates, such as Tibshirani's Lasso [Tibshirani, 1996].

3.4 Mixed Integer Optimization

This section details the utility of mixed integer optimization (MIO) to achieve the following analytic goals by one-step operation in a supervised learning paradigm: (i) Fusion (or clustering) and (ii) estimation. Its application in our study results in critical windows of physical activity.

3.4.1 Proposed Fusion-Adapted MIO Formulation

Bertismas et al [Bertsimas et al., 2016] offered an MIO formulated-solution to address the best subset problem in Equation (3.3) using Specially Ordered Sets of Type 1 (SOS-1). In this chapter, we propose an adaptation of this MIO framework with new L_0 constraint formulations to conduct concurrent parameter fusion and changepoint detection to analyze Equation (3.2). The number of groupings is controlled by setting the number of desired clusters K, which is tuned by goodness-of-fit measures such as BIC. Before formalizing the MIO constraints, we first introduce variable $\boldsymbol{\zeta}$ identifying group membership such that:

$$\boldsymbol{\zeta}_k = (\zeta_k^1, \zeta_k^2, \cdots, \zeta_k^J) \in \{0,1\}^{J\times 1},\ k = 1, \cdots, K \tag{3.4}$$

where $\zeta_k^j = 1$ corresponds to the case of β_j belonging in activity window k. Given cutoffs or edges of windows, $c_1, \ldots, c_K$ with $c_K = J$, such binary group labels take values:

$$\zeta_1^j = \begin{cases} 1, & j = 1, \cdots, c_1 \\ 0, & otherwise \end{cases}, \zeta_k^j = \begin{cases} 1, & j = c_{k-1}, \cdots, c_k \\ 0, & otherwise \end{cases}, \cdots, \zeta_K^j = \begin{cases} 1, & j = c_{K-1}+1, \cdots, J \\ 0, & otherwise \end{cases}. \tag{3.5}$$

For a K-group model, a fusion-adapted L_0 constrained optimization with J original intervals and q covariates is represented as:

$$\begin{aligned} \min_{\boldsymbol{\theta},\boldsymbol{\zeta},\boldsymbol{\beta},\boldsymbol{c},\boldsymbol{\alpha}} \quad & \|\boldsymbol{Y} - \boldsymbol{A\theta} - \boldsymbol{Z\alpha}\|_2^2 \\ \text{subject to} \quad & \boldsymbol{\theta} = (\theta_1, \cdots, \theta_J)^T \in \mathbb{R}^{J\times 1},\ \boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_q)^T \in \mathbb{R}^{q\times 1}; \\ & \mathbf{c} = (c_1, \cdots, c_{K-1}) \in \mathbb{N}^{1\times(K-1)} \\ & c_1 \geq 1,\ c_k \geq c_{k-1} + 1,\ c_{K-1} \leq J - 1, k = 1, \cdots K - 1; \\ & \boldsymbol{\zeta} = (\zeta_j^k)_{J\times K} \in \mathbb{R}^{K\times J} \\ & \theta_j - \beta_1 = 0,\ j = 1, \cdots, c_1; \\ & \theta_j - \beta_2 = 0,\ j = c_1 + 1, \cdots, c_2; \\ & \vdots \\ & \theta_j - \beta_K = 0,\ j = c_{K-1} + 1, \cdots, J, \end{aligned} \tag{3.6}$$

where $\boldsymbol{Y} \in \mathbb{R}^{n\times 1}$, $\boldsymbol{A} \in \mathbb{R}^{n\times J}$, and $\boldsymbol{Z} \in \mathbb{R}^{n\times q}$. The above optimization is operated via augmented parameters where labels $\boldsymbol{\zeta}$ and cutpoints $\boldsymbol{c} = (c_1, \ldots, c_{K-1})^T$ do not exist in the original model (3.2) but are added for parameter fusion. Obviously, group labels $\boldsymbol{\zeta}$ and cutpoints $\boldsymbol{c}$ are determined in a one-to-one correspondence fashion, which will be enforced via adequate constraints given below in Section 3.4.2. As mentioned in Section 3.3, there is no restriction of monotonicity on $\theta(c)$; however, if researchers were interested in a monotonic structure for association parameter $\theta(c)$, this could be enforced via the inclusion of additional constraints relating the relative values of parameters $\beta_1, \ldots, \beta_K$.

3.4.2 MIO Implementation

This MIO model can be solved via numerical software such as GUROBI under a system of constraints. These constraints are set to minimize the objective function by optimizing cut-points $c_1, \cdots, c_{K-1}$ and thus the cluster labels represented by the variable $\zeta_k,$, $k = 1, \cdots, K$ defined in Equations (4.8) and (4.9). This set of linear constraints for the K-group model is specified as follows:

$$\begin{aligned}
&\zeta_k^j(\theta_j - \beta_k) = 0, \ \ j = 1, \cdots, J, \ k = 1, \cdots, K \text{ (SOS-1 constraints)};\\
&\sum_{k=1}^{K} \zeta_k^j = 1, \ j = 1, \cdots, J;\\
&c_0 = 0, c_1 \geq 1; \ c_k \geq c_{k-1} + 1; \ \text{ and } c_{K-1} \leq J - 1, \ \text{for } k = 2, \cdots, K-1;\\
&\frac{c_k - j}{J} \leq 1 - \zeta_{k+1}^j, \ j = 1, \cdots, J, \ \text{for } k = 0, \cdots, K-1;\\
&\frac{c_{k+1} - j}{J} \times \frac{(j - c_k)}{J} \leq \zeta_{k+1}, \ j = 1, \cdots, J, \ \text{for } k = 0, \cdots, K-1;\\
&\frac{j - c_{k+1} + 1}{J} \leq 1 - \zeta_{k+1}, \ j = 1, \cdots, J, \ \text{for } k = 0, \cdots, K-1;
\end{aligned} \tag{3.7}$$

These constraints determine the locality of changepoints and grouping in the fusion-adapted MIO formulation for the L_0-type analysis of a K-group model. In this chapter, the constraints are implemented in GUROBI numerical solver package in Python. In a recent paper [Wang et al., 2022] show that the MIO GUROBI optimization solvers provide the global optimal solutions for a similar homogeneity fusion problem.

3.5 Theoretical Guarantees

Here we discuss the selection consistency of the MIO estimator of the parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_J)^T$ obtained by the constrained optimization given in Equations (3.6) and (3.7) under some mild regularity conditions. This paves the theoretical basis for large-sample statistical inference. Wang et al. [Wang et al., 2022] considered a more general version of an MIO optimization problem than ours, where their group parameters $\theta_1, \ldots, \theta_J$ are not sequentially ordered. In other words, the MIO optimization given in Equations (3.6) and (3.7) is a special case of the setting studied by [Wang et al., 2022], and thus, we can establish relevant theoretical guarantees by arguments given in [Wang et al., 2022].

To present the sufficient conditions for selection consistency, we first introduce the oracle estimators that represent the parameter estimates under the true number of clusters K^* and cutpoints $c^* = (c_1^*, \cdots, c_K^*)$. We denote the oracle estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ as $\hat{\boldsymbol{\beta}}^{ol}$ and $\hat{\boldsymbol{\alpha}}^{ol}$ respectively, which are obtained through the ordinary least squares (LS) estimation:

$$(\hat{\boldsymbol{\beta}}^{ol}, \hat{\boldsymbol{\alpha}}^{ol}) := \underset{\boldsymbol{\beta},\boldsymbol{\alpha}}{\operatorname{argmin}} \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{\alpha}\|_2^2. \tag{3.8}$$

When these cutpoints are unknown, we propose to use the MIO approach to obtain consistent estimators of the model parameters and the cutoff values in one step. In order to achieve this, we aim to minimize our constrained objective function in Equations (3.6) and (3.7) where cutpoints $\boldsymbol{c}$ are determined when J individual-level parameters θ_j's are reduced to K group-level parameters β_k via suitable constraints. To quantify the sensitivity of the model to the precision of clustering, we follow [Zhu et al., 2013]'s work with simultaneous grouping and feature selection and adopt a measure of Mean Squared Error (MSE) sensitivity deemed c_{min}. This measurement quantifies the minimum increase of MSE due to an inaccurately determined set of cutpoints $\boldsymbol{c}$. That is,

$$\begin{aligned} c_{min} \equiv c_{min}(\boldsymbol{\xi}^*, \boldsymbol{A}, \boldsymbol{Z}) = \min_{\xi} \frac{\|\boldsymbol{A}(\boldsymbol{\theta} - \boldsymbol{\beta}^*) + \boldsymbol{Z}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)\|_2^2}{n \max(d(\boldsymbol{\theta}, \boldsymbol{\beta}^*), 1)}, \\ \text{subject to Equations (3.6) and (3.7).} \end{aligned} \tag{3.9}$$

with the true values $\boldsymbol{\xi}^* = (\boldsymbol{\beta}^{*\top}, \boldsymbol{\alpha}^{*\top}, \boldsymbol{c}^*)^\top \in \mathbb{R}^{2K-1+q}$, $\boldsymbol{\xi} = (\boldsymbol{\theta}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{c})^\top \in \mathbb{R}^{J+K+q-1}$, and $d(\boldsymbol{\theta}, \boldsymbol{\beta}^*)$ represents a grouping incongruity measure reflective of the accuracy of the cutpoint estimation. See more details in [Wang et al., 2022]. In addition, we assume that errors $\boldsymbol{\epsilon}$ in the scalar-on-function model are normally distributed with mean 0 and variance σ^2.

To present selection consistency, for any MIO estimator $\hat{\boldsymbol{\xi}}^{\boldsymbol{MIO}} = (\hat{\boldsymbol{\beta}}^{MIO^\top}, \hat{\boldsymbol{\alpha}}^{MIO^\top})^\top$ of $\boldsymbol{\xi}^*$ with true number of windows K^*, and associated estimated cutpoints $\hat{\boldsymbol{c}}^{MIO}$ of $\boldsymbol{c}^*$, we define a loss function $L(\hat{\boldsymbol{\xi}}^{\boldsymbol{MIO}}; \boldsymbol{\xi}^*)$ as the grouping risk associated with inaccurate grouping and estimates of $\boldsymbol{c}^*$ in the form of $L(\hat{\boldsymbol{\xi}}^{\boldsymbol{MIO}}; \boldsymbol{\xi}^*) \equiv \mathbb{P}(\hat{\boldsymbol{\xi}}^{\boldsymbol{MIO}} \neq \boldsymbol{\xi}^*)$. Under the condition of random errors $\boldsymbol{\epsilon}$ that are independent sub-Gaussian with mean zero and variance $\sigma^2 < \infty$, when $K = K^* < \infty$ and for any J, we show the finite sample error bound is given by:

$$L(\hat{\boldsymbol{\xi}}^{\boldsymbol{MIO}}; \boldsymbol{\xi}^*) \leq 4exp\left[-\frac{3N}{200\sigma^2} \left\{ c_{min} - \frac{\sigma^2}{N} (134 \log(JK^*) + 220) \right\} \right]. \tag{3.10}$$

This bound in (3.10) implies that when $c_{min} > \frac{\sigma^2}{N}(134\log(JK^*) + 220)$, $\hat{\boldsymbol{\xi}}^{\boldsymbol{MIO}}$ consistently reconstructs $\boldsymbol{\xi}^*$ because $N, J \longrightarrow \infty$, $\mathbb{P}(\hat{\boldsymbol{\xi}}^{\boldsymbol{MIO}} \neq \boldsymbol{\xi}^*) \longrightarrow 0$. Thus for any finite fixed J, or for $J \longrightarrow \infty$, $\hat{\boldsymbol{\xi}}^{\boldsymbol{MIO}}$ consistently reconstructs $\boldsymbol{\xi}^*$ as $N \longrightarrow \infty$. In particular, in the former case of fixed J, the above condition for c_{min} holds automatically as $N \longrightarrow \infty$, and therefore the selection consistency is warranted under very mild regularity conditions. The proof of this sufficient condition result can be carried out by following the lines of arguments given in the proof of Theorem 3.4 in [Wang et al., 2022] and thus is omitted in this chapter.

3.6 Simulation Experiments

Simulation experiments demonstrate robust, reliable performance of the proposed MIO paradigm. Here we discuss the setup of the conducted numerical experiments, report on their results, and comment on computational performance comparing to Fused Lasso.

3.6.1 Simulation Setup

We first simulated 6-hour time-series of VM counts by linking many consecutive 10-minute intervals of the ELEMENT accelerometer data. To achieve this, the individual 6-hour time-series of VM counts for the 539 subjects from the ELEMENT dataset were divided into non-overlapping 10-minute segments. Each 10-minute interval was randomly drawn from a pool of 539 10-minute candidate segments. To ensure that the variability in the simulated PA reflected the variability of the ELEMENT dataset as shown in Figure 3.3, we first classified these 539 subjects into three groups with low, medium, and high levels of PA respectively, as defined by tertiles of "Moderate-to-Vigorous" VM counts using the pre-set Chandler cutoffs [Chandler et al., 2016]. We then simulated the time-series data within each tertile. With the simulated VM counts, OTC curves were calculated as described in Section 2.3. For the 500 simulated OTCs, we calculated the $J = 300$ successive integrals (i.e. AUCs) over domain $\mathcal{C} = (0, 30, 000)$, with each interval covering 100 VM counts: $(c_0 = 0, c_1 = 100, \cdots, c_J = 30, 000)$. For ease of exposition, we will refer to the $VM/100$ values, i.e. $c = (0/100, \cdots, 30000/100)^\top$ or $c = (0, \cdots, 300)^\top$.

To assess the fusion-adapted L_0 approach's ability to detect the true cutoffs and parameter estimates, we specified $K^* = 3$ groups and corresponding true cutoffs (c_1^*, c_2^*, c_3^*) in addition to $c_0^* = 0$, and calculated the vector of AUCs, $(A_1^*, A_2^* A_3^*)^T$ with $A_k^* = \int_{c_{k-1}}^{c_k} OTC(c)dc$. Finally, we generated outcome Y from the zero-intercept linear model $Y = \sum_{k=1}^{3} A_k^* \beta_k^* + Z\alpha^* + \epsilon$, with true effect sizes $(\beta_1^*, \beta_2^*, \beta_3^*)$ and α^*, where single continuous covariate $Z \sim N(0, 1)$ and $\epsilon \sim N(0, 10)$. We specified various 3-group models $(K^* = 3)$ with VM count changepoints $(c_1^*, c_2^*, c_3^*) \in \{(40, 80, 300), (20, 120, 300)\}$ to evaluate the performance under various window sizes. Here, we specified two different scenarios of effect size $(\beta_1^*, \beta_2^*, \beta_3^*) \in \{(4, 0, -4),\ (1, 0, -1)\}$. These simulations were conducted for two different specifications of J, the number of intervals to fuse over, $J \in \{60, 300\}$, representing two different levels of multi-collinearity with $J = 300$ encompassing the most severe multi-collinearity among the A_j's. Additionally, we conducted scenarios with three different sample sizes $N \in \{100, 250, 500\}$ with $J = 60$, and $N = 500$ when $J = 300$. Note that when $J = 300$, the method is limited to scenarios with $J < N$ as the fusion-adapted L_0 formulation does not introduce the true sparsity into the model that allows for $J > N$. With these simulated 3-group models, we applied the new fusion-adapted L_0 constraint method using GUROBI to fit models with $K = 2, 3, 4$, and used BIC to select the final model with the best goodness of fit among the candidates in order to determine the method's sensitivity in selecting the right-sized model. Additionally, we conducted simulation experiments for the $K^* = 4-$group model with corresponding true cutoffs $(c_1^*, c_2^*, c_3^*, c_4^*)$ and true effect sizes $(\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*)$ (refer to Appendix A).

3.6.2 Simulation Results

The simulation results produced by the fusion-adapted L_0 constraint model demonstrated that this new approach has high sensitivity to select the right-sized model, produces reliable change-point detection and parameter estimation, and is robust to handle highly correlated AUCs. Tables 3.2 and 3.3 summarize the results from 500 rounds of simulations of the 3−group model for the $J = 300$ and $J = 60$ settings, respectively.

Table 3.2: Simulation Results of the 3-group model with number of micro-intervals $J = 300$ and sample size of $N = 500$ summarized over 500 replicates, including average estimate (L_0 Mean), empirical standard error (L_0 ESE), and average estimate from an L_1 Fused Lasso analysis (FL Mean) using R package `glasso`. Cutpoint values are represented as VM/100.

	Scenario A				Scenario B				Scenario C			
	Truth	L_0 Mean	L_0 ESE	FL Mean	Truth	L_0 Mean	L_0 ESE	FL Mean	Truth	L_0 Mean	L_0 ESE	FL Mean
β_1	4	4.00	0.04	2.87	1	1.00	0.04	0.71	1	1.01	0.09	0.31
β_2	0	−0.01	0.11	−0.79	0	−0.01	0.11	−0.22	0	< 0.01	0.02	−0.17
β_3	−4	−4.00	0.00	−3.98	−1	−1.00	0.00	−0.99	−1	−1.00	0.01	−0.98
c_1	40	40.04	0.75	59.47	40	39.95	3.49	60.45	20	19.80	2.10	76.83
c_2	80	80.05	0.63	78.47	80	80.28	2.65	79.50	120	119.92	1.38	118.65
α	1	0.98	0.44	0.94	1	0.98	0.44	1.20	1	0.98	0.44	1.16

This method demonstrated high sensitivity, selecting the correct sized model in over 99%

of simulations in all Scenarios A, B, C, for both $J = 300$ and 60. Among these correctly specified models, the fusion-adapted constraint model correctly identified the changepoints $(c_1^*, c_2^*, c_3^*) = \{(40, 80, 300), (20, 120, 300)\}$ and estimated the β parameters $(\beta_1^*, \beta_2^*, \beta_3^*) \in \{(4, 0, -4),\ (1, 0, -1)\}$ with minimal bias.

The method maintained its ability to reliably identify changepoints and estimate parameters as the sample size N decreased from $N = 500$ to $N = 250$ and even $N = 100$. For Scenario B with $(\beta_1^*, \beta_2^*, \beta_3^*) = (1, 0, -1)$ and $N = 250$, the mean (ESE) estimates of β_1^*, β_2^*, and β_3^* from this L_0 constrained approach are 1.00(0.06), −0.02(0.17), and −1.00(0.01). Similar strong results are repeated in the second window size scenario of $(c_1^*, c_2^*, c_3^*) = (20, 120, 300)$.

In contrast, the L_1 fused lasso approach via the R package `glasso` had undesirable sensitivity, ranging from 0-30% across the different Scenarios and sample size combinations. Furthermore, even if the number of windows is correctly specified in advance, namely $K^* = 3$, the performance of Fused Lasso analysis exhibited high bias in both coefficient and changepoint detection, as shown in the "FL Mean" columns of Tables 3.2 and 3.3. The proposed MIO formulation can produce desirable results even in scenarios of severe multicollinearity. In fact, in the $J = 300$ setting, the pairwise correlation was extremely high with $cor(A_j, A_{j+1}) = 0.998$, $cor(A_j, A_{j+5}) = 0.985$, and $cor(A_j, A_{j+10}) = 0.967$. Even in this very challenging scenario, the parameter and changepoint estimates have been estimated well with remarkably low bias and variance. Results for the $K^* = 4$ simulation experiments were similarly strong for the proposed MIO approach, and weak for an L_1 Fused Lasso approach, as shown in the Appendix A Tables.

3.6.3 Computation Time

The fused-adapted MIO solver via GUROBI is also computationally efficient. A 3-group simulation model with $N = 500, J = 60$ computes in 10 seconds, with $J = 300$ scenario completing in 10 minutes. The method is scalable to a reasonable number of windows, with computation time for a 4-group model taking 30 seconds and 30 minutes for $J = 60, 300$ scenarios respectively. These simulation scenarios represent instances of defined signal for the $K = 3, 4$ number of groups. However, it is possible that in scenarios of low signal or inappropriate number of groups K that the computation would take longer. Thus in the simulation and data analysis, we implement a computation budget of 20 hours to control the run time. If the analysis does not complete within this time frame, the MIO model is terminated and the combination of (J, K) deemed an inappropriate model representation.

Table 3.3: Simulation Results of the 3-group model with number of micro-intervals $J = 300$ and sample size of $N \in \{500, 250, 100\}$ summarized over 500 replicates, including average estimate (L_0 Mean), empirical standard error (L_0 ESE), and average estimate from an L_1 Fused Lasso analysis (FL Mean) using R package `glasso`. Cutpoint values are represented as VM/100. Sensitivity for selecting 3-group model based on goodness-of-fit comparisons was greater than 99% in all scenarios.

		N = 500			N=250			N=100		
	Truth	L_0 Mean	L_0 ESE	FL Mean	L_0 Mean	L_0 ESE	FLMean	L_0 Mean	L_0 ESE	FL Mean
Scenario A										
β_1	4	4.00	0.03	2.97	3.99	0.04	2.87	4.00	0.09	2.79
β_2	0	0.00	0.05	−0.58	0.00	0.06	−0.69	−0.01	0.22	−0.96
β_3	−4	−4.00	0.00	−3.98	−4.00	0.00	−3.98	−4.00	0.01	−3.97
c_1	40	40.01	0.22	57.75	40.06	0.54	59.44	40.02	1.64	60.83
c_2	80	80.01	0.22	78.40	80.01	0.22	78.44	80.13	1.28	80.12
α	1	0.98	0.44	0.98	1.04	0.66	0.92	1.01	1.04	1.07
Scenario B										
β_1	1	1.00	0.04	0.69	1.00	0.06	0.69	1.03	0.19	0.68
β_2	0	−0.01	0.12	−0.18	−0.02	0.17	−0.21	−0.06	0.30	−0.28
β_3	−1	−1.00	0.00	−0.99	−1.00	0.01	−0.99	−1.00	0.02	−0.96
c_1	40	39.94	3.59	61.63	39.98	5.38	61.51	39.73	9.75	61.89
c_2	80	80.28	2.85	78.24	80.84	5.08	80.27	84.93	21.07	84.73
α	1	0.98	0.44	1.20	1.00	0.67	1.11	1.00	1.04	1.46
Scenario C										
β_1	1	1.00	0.10	0.29	1.01	0.14	0.28	1.05	0.28	0.27
β_2	0	−0.001	0.02	−0.16	0.001	0.03	−0.19	−0.002	0.05	−0.25
β_3	−1	−1.00	0.005	−0.98	−1.00	0.01	−0.98	−1.00	0.01	−0.97
c_1	20	20.14	2.21	79.36	20.09	3.50	81.25	20.22	5.98	80.47
c_2	120	120.02	1.27	118.27	119.94	2.33	118.73	120.31	3.60	120.88
α	1	0.98	0.44	1.15	1.00	0.66	1.09	1.00	1.02	1.41

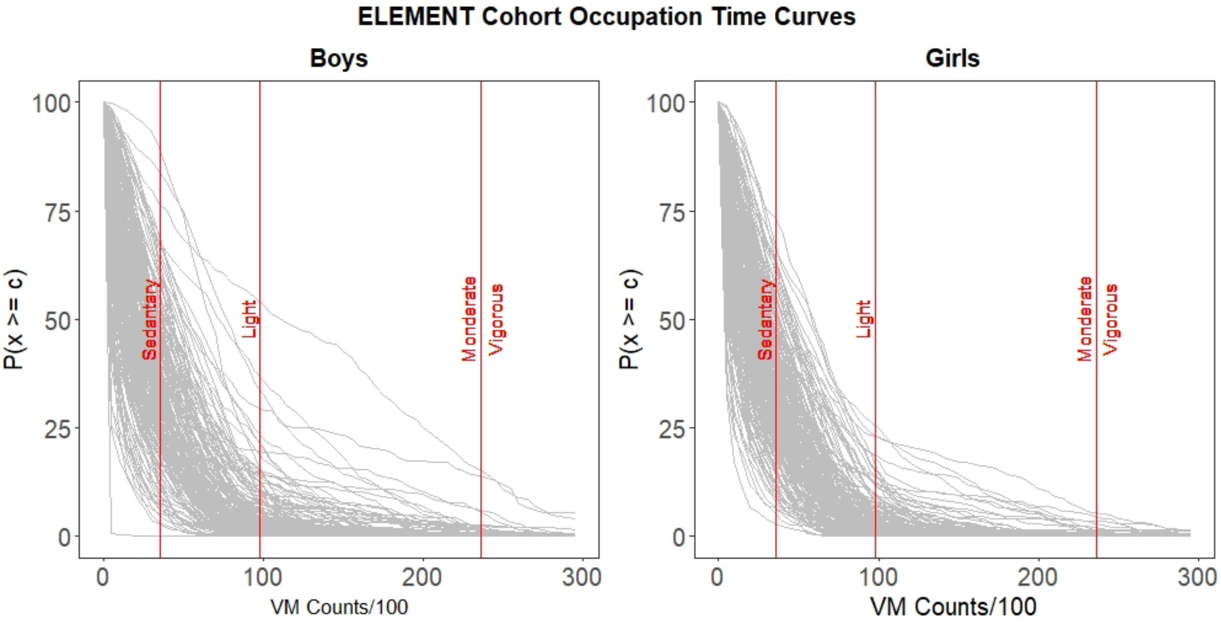


Figure 3.3: OTCs for 354 ELEMENT subjects stratified by boys and girls. The vertical lines represent Chandler's cutoffs [Chandler et al., 2016] for prefixed activity levels (Sedentary, Light, Moderate, Vigorous). The relative shape of OTC reflects the subject's activity profile.

3.7 Data Analysis

The primary objective of this data analysis was to investigate whether physically more active individuals are biologically younger or older. To do this, we focused on assessing the functional relationship between PA and biological aging through a scalar-on-function regression model. Introduced in Section 2.2, we had complete accelerometry and covariate data for 354 subjects from our motivating dataset (172 male, 182 female), with mean(SD) age of 13.7(1.9) years and mean(SD) lead exposure of 3.17(3.33) $\mu g/dL$. The majority (332) of subjects had completed puberty in terms of Tanner staging standards. Figure 3.3 illustrates the functional predictors of OTCs representing the subjects' activity profiles fom 4:00PM - 10:00PM on weekends; this block was chosen with the rationale that it is reflective of time when the children have more control over their activities. Our choice of outcome was Horvath's AgeSkinBlood Clock [Horvath et al., 2018] that primarily targets DNAm changes in skin and blood cells that undergo rapid changes during adolescence, including fibroblasts that help with the structural components of skin.

To use the fusion-adapted L_0 analytic in Section 3.4, we began setting $J = 300$, with each interval covering 100 VM counts, followed by an augmentation scenario of $J = 60$ by summing every five successive intervals. For ease of interpretation, we considered $K \in \{2, 3, 4\}$ PA windows, in which the final selection of K was determined by BIC. Each setting was given a budget of 20 hours runtime; if the search did not converge within this time, the attempt was terminated and the respective combination of (J, K) disregarded from reporting.

Table 3.4 shows the results, among which the 3-group model demonstrated the best fit in both the $J = 300$ and the $J = 60$ scenarios according to BIC. In the scenario $J = 300, K = 4$ the MIO search did not complete within the budgeted 20 hours, and was thus terminated. As the scenario of $K = 4, J = 60$ was inferior over the scenario of $K = 3, J = 60$, the chance of $K = 4, J = 300$ scenario being the best seemed to be rather low and thus the decision of termination was not concerning. P-values and BIC are determined by fitting a resulting linear model with the detected cutpoints.

To assess the validity of the p-values used in the above discovery we conducted a permutation analysis. To do so, we randomly permuted the epigenetic age outcomes to be misaligned with the original covariates and extracted the p-values from the refit linear model. Using 1000 permutations, we established a null distribution, and then compared the analytic p-values with those determined by the permutation-derived null distribution, termed as "permuted p-value". We found that the permuted p-values follow an approximately uniform distribution on $(0, 1)$, and are remarkably similar to the original analytic p-values, as evident in Table 3.4. The uniform distribution of the permuted p-values indicates that the L_0 fusion-adapted model furnishes an adequate approximation of the functional relationship between epigenetic age and functional OTCs. If this functional model were not an adequate approximation, the error term, ϵ, would carry a substantial proportion of the relationship between the outcome and covariates, thus resulting in a non-uniform distribution of p-values in the permutation analysis.

Under the chosen $K = 3$ model, the estimated activity windows reflect that (i) more time in the low PA window $c = [0, 20]$ is associated with younger AgeBloodSkin ($\hat{\beta} = 4.17$, p-value 0.004), and (ii) more time in the extreme window $c = (290, 300]$ is associated with older AgeBloodSkin ($\hat{\beta} = 13.0$, p-value 0.012). Such findings suggest that more PA is associated with faster biological aging of blood cells and skin in adolescents.

To facilitate a clinically understandable interpretation of the analysis results in Table 3.4, we propose an AUC Ratio metric that measures the amount of time an individual spends within a PA window relative to the maximum amount of time they could spend above that PA level. That is, it represents a relative activity level of the individual within the detected window compared to the hypothetical most active person. Computationally, as illustrated in Figure 3.7, the AUC Ratio is a ratio of the individual's AUC in the detected activity window versus the area of the full activity window rectangle, with the latter representing the PA of an individual who spends all of his or her time above the PA level of this window. The

Table 3.4: Data Analysis Results obtained by the fusion-adapted L_0 method, where J indicates the number of micro-intervals and K is a prefixed number of activity windows. Significance is measures in two different ways; 'p-val' represents the p-values from linear regression, whereas 'perm' represents empirical p-value when assessing 'p-val' to the distribution of p-values attained through 1000 permutations. Cutpoint values are represented as VM/100.

	J = 300						J = 60								
	K=2			K=3			K=2			K=3			K=4		
Parameters	Est	p-val	perm.	Est	p-val	perm.	Est	p-val	perm.	Est	p-val	perm.	Est	p-val	perm.
β_1	19.50	.03	.04	4.17	.01	.01	4.73	.01	.02	4.21	.01	.01	0.13	.53	.52
β_2	−0.07	.53	.54	−0.36	.02	.02	−0.12	.34	.34	−0.37	.02	.02	−126.26	$< .01$	$< .01$
β_3	–	–	–	13.00	.01	.02	–	–	–	8.56	.01	.02	62.91	$< .01$	$< .01$
β_4	–	–	–	–	–	–	–	–	–	–	–	–	−0.54	.70	.70
c_1	3	–	–	20	–	–	15	–	–	20	–	–	240	–	–
c_2	–	–	–	293	–	–	–	–	–	290	–	–	245	–	–
c_3	–	–	–	–	–	–	–	–	–	–	–	–	255	–	–
Sex (Male)	15.27	.76	.77	−3.33	.94	.95	6.68	.89	.91	−3.24	.94	0.95	23.50	.64	.64
Chron. Age	0.86	$< .01$	$< .01$	0.86	$< .01$	$< .01$	0.86	$< .01$	$< .01$	0.86	$< .01$	$< .01$	0.85	$< .01$	$< .01$
Lead	−6.44	.39	0.44	−6.39	.39	.43	−6.32	.40	.45	−6.32	.39	.44	−7.03	.34	.37
Puberty	−97.10	.29	0.30	−88.14	.34	.34	−96.39	.30	.31	−88.25	.34	.34	−86.86	.35	.35
BIC	5395.9			5393.6			5394.2			5393.7			5394.7		

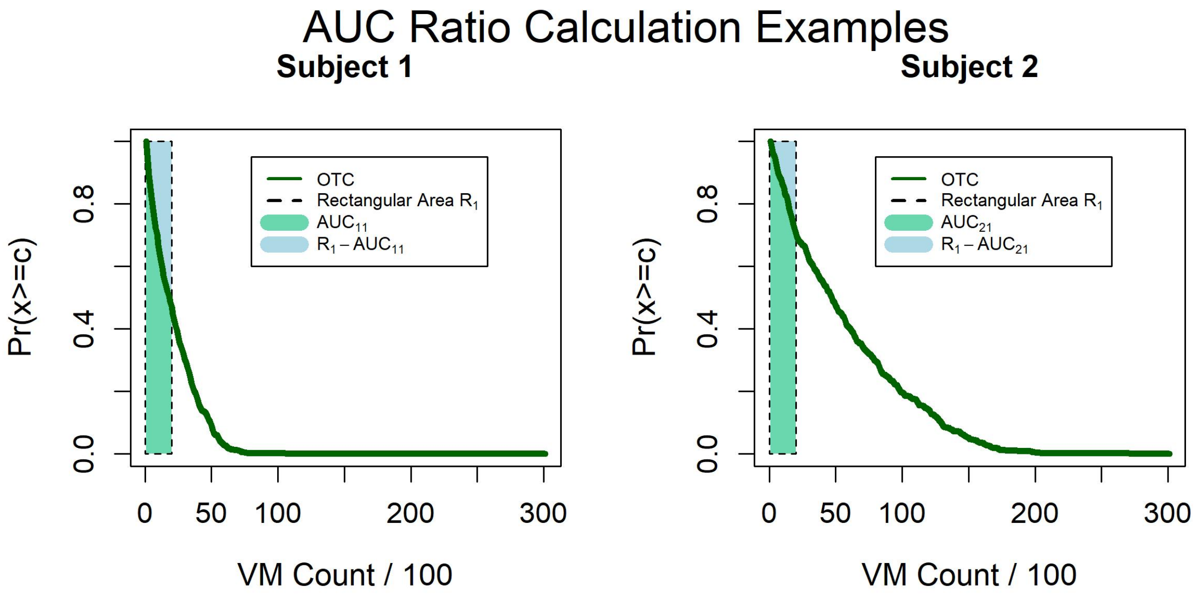


Figure 3.4: An illustration of the AUC Ratio calculation for Window 1. The shaded green regions represent the AUC for subject i for $i = c(1,2)$ within the detected window, or A_{i1}. The outlined rectangle represents the area of the full rectangle created by the cutpoints ($c_0 = 0, c_1 = 20$) of Window 1, deemed R_1, which represents the hypothetical subject spending 100% of time above this activity level. The AUC Ratio for this first window is then calculated by AUC Ratio$_{i1}$ = $\frac{A_{i1}}{R_1}$. The interpretation of this ratio depends on the location of the window. For all but the last sequential window, i.e. for windows $1, \cdots, K-1$, the value 1−AUC Ratio represents what percentage the individual is Less Active than the hypothetical most active person in that window. For example, in the above figure Subject 1 has a smaller AUC (green shaded region) than Subject 2, representing that Subject 1 spends more time within the cutpoints (c_0, c_1) than Subject 2. The value 1−AUC Ratio$_{11}$, or $\frac{R_1 - A_{11}}{R_1}$, represents the percentage Subject 1 is less active than the hypothetical most active person within the first window. This value is greater 1−AUC Ratio$_{21}$, or $\frac{R_1 - A_{21}}{R_1}$, as can be visualized by the area of the blue shaded regions. For the K^{th} window, the interpretation of the AUC Ratio$_{iK}$ represents the percent of time the individual spends within that window compared the hypothetically most active person. In this case, a higher AUC Ratio$_{iK}$ value represents higher PA within the window.

interpretation of this AUC Ratio depends on its sequential location. For an example of the first Window 1, a lower AUC Ratio represents more time spent within the specific window, and less time spent above the window. In contrast, for the last window K, a higher AUC Ratio represents more time spent within the specific window.

Let us interpret the results in Table 3.4 for the scenario of $K = 3, J = 300$ under model $y \sim \beta_1 A_1 + \beta_2 A_2 + \beta_3 A_3 + Z^T \alpha$, where AUC $A_k = \int_{c_{k-1}}^{c_k} OTC(c)dc, k = 1, 2, 3$. Here Window 1 has estimated cutpoints $[c_0, c_1] = [0, 20]$ with $\hat{\beta}_1 = 4.17$ for predictor A_1. For subject i, the area of the Window 1 rectangle $R_1 = (c_1 - c_0) \times (1 - 0) = 20$, and AUC Ratio of Window 1 is $\frac{A_{i1}}{20}$. Correspondingly, the parameter estimate $\hat{\beta}_1$ may be adjusted by $\hat{\beta}_{1Ratio} = 20\hat{\beta}_1$ for the interpertability. In the case of Window 1, a lower AUC Ratio reflects *more* time spent within the activity cutpoints $[0, 20]$ than the hypothetical "most active individual" who spends all his or her time above the cutpoint range $[0, 20]$. Thus, for a subject who is 1% more active in the activity range of Window 1 compared to the hypothetical "most active individual", as reflected by a smaller AUC Ration, this subject's BloodSkin epigentic age decreases approximately 80 days. See Figure 3.7 for a schematic of this calculation.

3.8 Discussion

In this chapter we utilize a scalar-on-function model to assess the influence of PA on biological age using a proposed methodology of fusion-adapted L_0 regularization. This scalar-on-function regression naturally accommodates a functional accelerometer predictor with great flexibility to study similar scientific questions in other populations with various underlying characteristics and devices. We adopt a mixed integer optimization (MIO) analytic that can simultaneously detect key cutpoints to define critical windows of activity and estimate discretized functional association parameters, while accounting for important covariates of interest.

One advantage of the MIO technique lies on the fully data-driven simultaneous operation in both cutpoint detection and parameter estimation. This use of functional regression is notably different from current methods of analyzing accelerometer activity and investigating windows of activity associated with health outcomes. Unlike methods establishing fixed cutoff values regardless of specific outcomes under investigation, our analysis takes a new supervised learning approach that involves the outcome of interest to detect different change points, which are adaptive to different outcomes of interest and study populations under investigation. For example, if applied to different age populations, or to analyze data collected from a different accelerometer device (e.g. Fitbit or iWatch), the functional OTC predictor would likely form a robust functional PA profile despite the different activity count ranges that may be recorded by different populations and devices. Thus, as shown in our numerical analyses, our MIO based optimization approach can deliver reliable and reproducible findings on activity windows of importance and functional associations in the study of the influence of PA on human health outcomes. In contrast, existing approaches that apply pre-set child-specific cutpoints (e.g. Chandler) to an adult population could potentially lead to biased or even contradicting results.

We perform extensive simulation experiments to numerically demonstrate the high stability and accuracy of the MIO technique, including a useful finding that the strength of the results is not overly sensitive to the choice of J, the starting number of correlated intervals. Simulation results for $J = 60$ and $J = 300$ were very similar, with computation time slightly longer for the larger number of intervals. Investigators can choose the number of J intervals based on factors of sample size and data availability without concern that the tuning choice of J will significantly affect the analysis. Such desirable numerical performance confirms the selection consistency property for the MIO solution under mild regularity conditions.

The proposed framework gives rise to a data analytic toolbox enabling researchers to explore various questions of interest related to the effect of functional PA features on health outcomes. For example, some researchers hypothesize that the timing of PA, not only the relative intensity, is related to specific health outcomes. Through application of this MIO technique focusing on PA during different time periods of the day, such as morning versus evening, researchers can investigate if the activity intensity changepoints are dependent on time of day. Additionally, future extensions can include multiple functional covariates to assess the longitudinal affect of functional PA profiles on health outcomes, and even longitudinal effects with both repeated outcomes and functional exposures to PA using longitudinal functional data analysis models.

While this chapter focused on time-series of PA counts from wearable accelerometer devices, the use of Occupation Time Curves (OTCs) to summarize such high-frequency time-series data can be extended to a myriad of applications. Other forms of data from objective high-frequency measurements, such as ambulatory blood pressure or glucose level monitoring, can be represented as functional OTCs. In this way, important windows of the blood pressure or glucose levels to a health outcome of interest can be identified and assessed for statistical significance and scientific importance. The MIO technique is also flexible to accommodate different forms and number of covariates with an extension from the current formulation via little effort. Currently, our analysis of biological age focuses on a continuous outcome, though future work could extend this data analytic to non-normal and non-linear models, such as logistic regression with binary outcomes, and Cox regressions with time-to-event outcomes.

CHAPTER 4

Supervised Fusion Learning of Physical Activity Features from Longitudinal Functional Accelerometer Data

4.1 Introduction

Thus far in this dissertation, we have discussed proposed methodologies for independent functional data in which a single curve is observed for each of the subjects in the study. In this chapter we now consider repeated functional data, in which multiple functional curves are observed from the subjects under consideration. While this chapter focuses on physical activity (PA) collected from wearable accelerometer devices, and associated longitudinal outcomes, the proposed methodologies can be applied to alternative types of wearable devices that collect high-frequency time-series data, such as continuous glucose monitoring (CGM) devices, heart rate monitors, and toxicant sensors, among others.

4.1.1 Longitudinal Functional Data

In recent years, researchers have begun to consider the methodological challenges related to longitudinal functional data. Overall, this area of study focuses on analyzing the patterns and trends in the longitudinal data that are collected repeatedly over time, and can be classified into a few general categories. Some of the longitudinal FDA techniques consider Functional Principal Component Analysis (FCPA) and longitudinal FCPA, in which a major goal is dimension reduction and identifying patterns that contribute to variation across curves [Ramsay, 2004, 2005, Goldsmith et al., 2015, Yao et al., 2005, Nwanaji-Enwerem et al., 2021, Lin et al., 2023, Chen and Müller, 2012]. Another focus of longitudinal FDA research is Functional Data Clustering, which aims to identify groups or clusters of similar functional curves [Heinzl and Tutz, 2014], while additional research considers Longitudinal Functional

Regression, which extends the standard regression models to functional data [Reiss et al., 2017]

In this chapter, we focus on the realm of Longitudinal Functional Regression. In general, functional regression models can be classified into three types: (i) scalar-on-function regression models, which include scalar responses and functional predictors, (ii) function-on-scalar regression models, which consider functional responses and scalar predictors, and (iii) function-on-function regression models, which assess the relationship between functional responses and functional predictors [Ramsay, 2005, Reiss et al., 2017]. We consider the longitudinal extension of scalar-on-function models. This new setting extends from Chapter 3, in which we develop a longitudinal fusion method to assess population-average effects of functional predictors on longitudinal continuous scalar outcomes. While other researchers have studied longitudinal functional fusion methodologies, existing methods are not appropriate to answer our scientific question. These existing methods encompass different approaches: one subset of methods focus on the fusion of functional parameters in order to predict categorical outcomes [Adhikari et al., 2019], while others employ fusion techniques to estimate time-varying effects from non-repeated functional covariates [Yu and Zhong, 2021]. In contrast, our aim is to detect population-average critical activity-intensity windows from repeated functional accelerometer data and their population-average association with longitudinal outcomes of interest.

We propose a longitudinal fusion learning which extends the regularized MIO approach from Chapter 3 to a longitudinal functional framework with repeated wearable data in order to understand the influence of serially measured functional accelerometer data on longitudinal health outcomes. This proposed longitudinal approach invokes the means of Quadratic Inference Functions (QIF) [Song et al., 2009], with an aim to detect critical PA intensity windows and assess their population-average effects on children health outcomes. Discussed in detail in Section 4.3.1, QIF is a powerful tool in statistical modeling, particularly in the context of estimating complex models with clustered or correlated data. Specifically, it produces consistent estimators, and is robust to model mis-specfications, even producing more efficient estimators than competing methods when the model is misspecified. I consider a population-average effects model, and develop a regularized QIF via Mixed Integer Optimization (MIO) to carry out longitudinal data analysis withing the FDA framework.

4.1.2 Motivating Longitudinal Cohort Data

The motivating longitudinal birth cohort, ELEMENT, was previously introduced in Section 2.2, with details on a single time point discussed. Now we introduce the longitudinal data of a second visit. These two time points of data collection are referred to as “T1” and “T2” respectively, and discussed in detail in review paper [Perng et al., 2019]. Briefly, at T1 researchers collected actigraphy data from 539 children (258 boys and 281 girls) with mean (SD) ages of 13.9 (2.2), ranging from 9 to 18 years old. This actigraphy data collection was then repeated approximately two years from the same group of participants. The second visit data was collected from 496 subject (230 boys, 266 girls), now with mean(sd) age of 16.42(2.11), ranging from 12.45 to 20.68. At both T1 and T2, the participants were directed to wear a wrist-worn, tri-axial Actigraph GT3X+ (Actigraph LLC) for seven consecutive days. The tri-axial high-frequency time-series data was processed and summarized from the raw 30Hz data into epochs of various lengths (e.g. 30 sec, 1 min). In this chapter, we focus on analyzing repeatedly measured Vector Magnitude activity counts over one-minute epochs.

In addition to PA, the ELEMENT cohort also collected extensive anthropometric measurements to assess the subject’s health status. These measurements include Subscapular Skin Thickness (SSST), a measure of truncal fat distribution that changes markedly in males and females during puberty. This study focuses on longitudinal SSST as the health outcome of interest, and investigates its population-average association with PA. We consider confounding covariates, including chronological age, sex, and adult-status based on completing puberty as determined by a five-category ordinal variable of Tanner staging. Briefly, Tanner stages reflect pubertal status and progression based observation of secondary sexual characteristics, with Tanner stage 1 indicating pre-pubertal status, T2 pubertal onset and T5 post-puberty, e.g. adult status.

4.1.3 Longitudinal Occupation Time Curves

As in Chapters 2 and 3, we consider PA under the purview of Occupation Time Curves (OTCs). Discussed in detail in Section 2.3, the OTC provides insights into the behavior and characteristics of an individual’s PA profile. Here, we introduce a longitudinal aspect, incorporating a second time point of PA as measured from accelerometers and summarized by OTCs. Figure 4.1 illustrates the repeated OTCs from three individuals from our motivating cohort data. This figure demonstrates that some subjects follow very similar PA patterns at both Time point 1 and Time point 2 (Subjects A and B), whereas others demonstrate a changing PA profile over time (Subject C). These repeated OTC measurements represent functional covariates in the longitudinal scalar-on-function regression model.

This chapter is organized as follows: In Section 4.2 we introduce Longitudinal Functional Data Analysis Models, with Section 4.3 focusing on Constrained Quasi-Estimation methods, including the proposed MIO formulation of QIF. Relevant theoretical arguments are then

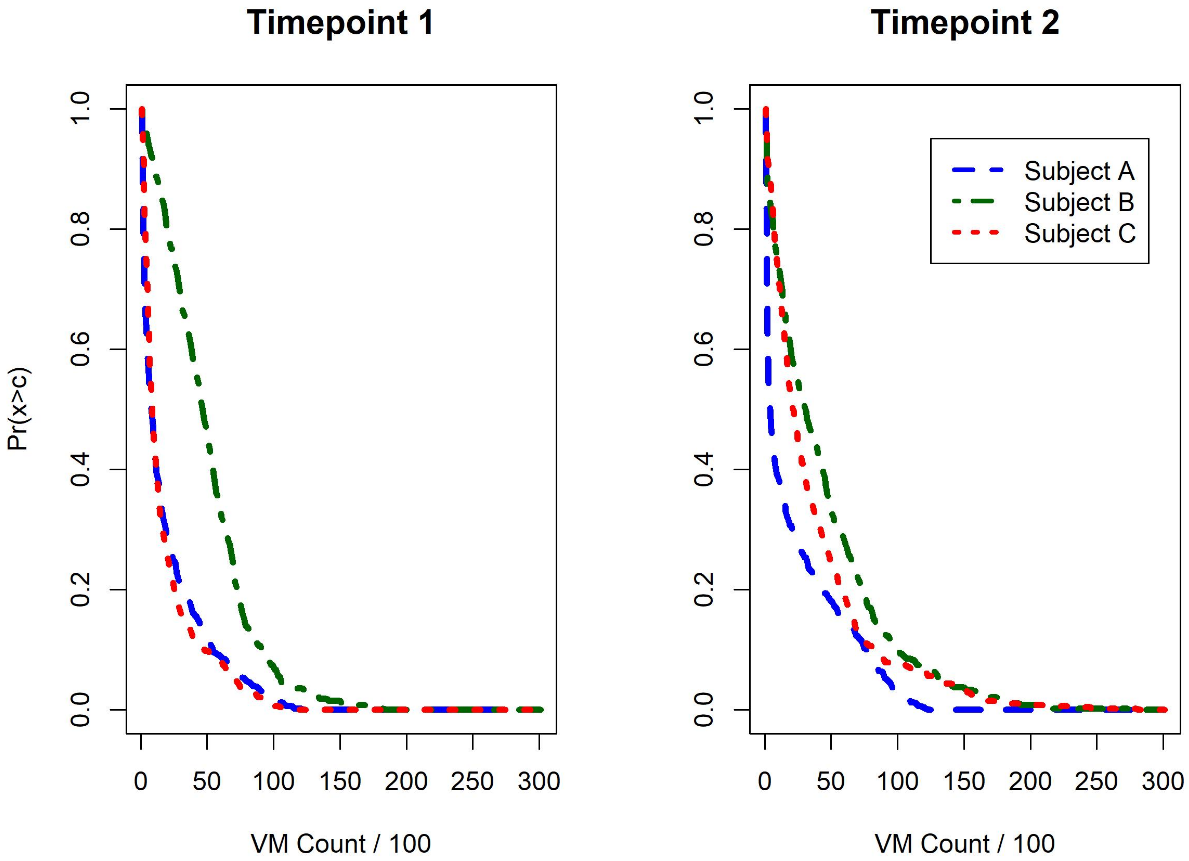


Figure 4.1: OTCs for three individuals from the motivating ELEMENT cohort at both Time point 1 (T1) and Time point 2 (T2), collected approximately two years apart. The OTC curves from both T1 and T2 represent PA during the weekend between 4:00PM-10:00PM. Subjects A and B demonstrate similar PA patterns across the two time points, whereas Subject C demonstrates a changing PA profile from T1 to T2.

introduced in Section 4.4. Section 4.5 discusses the numerical experiments demonstrating the proposed longitudinal fusion-learning methodology, while Section 4.6 provides a data analysis with our motivating cohort data, investigating the longitudinal population-average association between functional OTCs and scalar SSST measures. We end in Section 4.7 with a discussion on the merits, limitations, and potential extensions of this proposed model.

4.2 Longitudinal Functional Data Analysis Model

To assess the longitudinal functional association between functional PA and our health outcomes of interest, we consider the marginal model for population-average effect. Of note, while our study focuses on two repeated measurements at T1 and T2, this framework can be extended to m repeated measures.

4.2.1 Longitudinal Scalar-on-Function Model

In this longitudinal scalar-on-function model, we assume a longitudinal scalar outcome with a mix of functional and non-functional predictors. More specifically, for an individual subject i, we have an m-sized vector of scalar outcomes $y_i = (y_{i1}, y_{i2}, \dots, y_{im})^\top$, with y_{it} representing the subject's outcome at time t for $t \in \{1, 2, \dots, m\}$. Additionally, for subject i we consider a set of m functional covariates. Here, the functional covariate of interest is the OTC as defined in Section 2.3, denoted $OTC(c)$. Thus we invoke the longitudinal FDA model in Equation (4.1).

$$y_{it} = \int_C \theta(c) OTC(c)_{it} dc + \boldsymbol{z}_{it}^\top \boldsymbol{\alpha} + \epsilon_{it} \text{ for } t \in \{1, 2, \dots, m\} \text{ and } i \in \{1, \dots, N\}, \tag{4.1}$$

where $y_{it} \in \mathbb{R}^{1\times1}$ is the scalar outcome of interest; $OTC(c)_{it}$ is the functional OTC variable at time t as defined above with functional parameter $\theta(c)$; $\boldsymbol{z}_{it} \in \mathbb{R}^q$ with $\boldsymbol{z}_{it}^\top$ a $1 \times q$ row vector represents the relevant scalar confounders at time t, with associated parameter vector $\boldsymbol{\alpha} \in \mathbb{R}^q$. Here $c \in C \in \mathbb{R}$ represents the domain of Vector Magnitude activity counts. Note that in model (4.1) we assume the same population-average functional parameter $\theta(c)$ associated with the repeated functional variable $OTC(c)$, as well as time-invariant confounder parameter $\boldsymbol{\alpha}$ over time. The error terms, $\epsilon_{it}, t \in 1, \dots, m$, are assumed to be multivariate normal with $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{im})^\top \sim MVN(0, \sigma^2 R)$ where R is an $m \times m$ correlation matrix.

As we aim to estimate PA intensity windows of interest with distinct changepoints, the goal of the longitudinal scalar-on-function model as defined in (4.1) is to estimate functional parameter $\theta(c)$ as a step-function. In this way, we obtain non-overlapping activity windows, each with a constant parameter of population-average association with the health outcome of interest. To achieve this, we first discretize the functional term $\int_C \theta(c)OTC(c)_{it}dc$ into J-many equal-sized, successive, non-overlapping intervals at cut locations $c_1, \cdots, c_{J-1}$, where $c_0 = 0$ and $c_J = max(c)$ (here, $max(c) = 30,000$) where J is pre-fixed, such as J=300. Within each small interval $j \in J$, we assume the associated function $\theta_j(c)$ takes a constant value θ_j. That is, we reach the following approximation of the functional parameter leading to Equation (4.2):

$$\begin{aligned} y_{it} &= \sum_{j=1}^{J} \int_{c_{j-1}}^{c_j} \theta(c)OTC(c)_{it}dc + \boldsymbol{z}_{it}^{\top}\boldsymbol{\alpha} + \epsilon_{it} \\ &\approx \sum_{j=1}^{J} \theta_j \int_{c_{j-1}}^{c_j} OTC(c)_{it}dc + \boldsymbol{z}_{it}^{\top}\boldsymbol{\alpha} + \epsilon_{it} \quad , \\ &:= \sum_{j=1}^{J} \theta_j A_{itj} + \boldsymbol{z}_{it}^{\top}\boldsymbol{\alpha} + \epsilon_{it} \end{aligned} \tag{4.2}$$

where $OTC(c)_{it}$, $\boldsymbol{z}$, $\boldsymbol{\alpha}$ are defined as above and A_{itj} denotes the AUC over interval $(c_{j-1}, c_j]$ or $A_{itj} = \int_{c_{j-1}}^{c_j} OTC(c)_{it}\ dc$, for individual i at time t. Notably, the mean model in Equation (4.2) takes the form:

$$\begin{aligned} \mu_{it} &= \sum_{j=1}^{J} \theta_j A_{itj} + \boldsymbol{z}_{it}^{\top}\boldsymbol{\alpha} \\ &:= \left(\mu_{i1}, \ldots, \mu_{im}\right)^{\top}. \end{aligned} \tag{4.3}$$

With the parameters $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_J)^{\top} \in \mathbb{R}^J$ and $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_q)^{\top} \in \mathbb{R}^q$, we define the set of association parameters as $\boldsymbol{\eta} = (\boldsymbol{\theta}^{\top}, \boldsymbol{\alpha}^{\top})^{\top} \in \mathbb{R}^{J+q}$.

In the spirit of distinct activity window detection, our analytic aim in this functional longitudinal regression is to fuse similar adjacent parameter θ_j's together in order to estimate a K-group sized step function with group-level parameters β_k for $k = 1, \cdots, K$. Here, β_k is the association parameter related to the AUC of the k^{th} activity window over interval $(c_{k-1}, c_k]$. This results in a final estimate model: $y_{it} \sim \sum_{k=1}^{K} \beta_k A_{itk} + \boldsymbol{z}_{it}^{\top}\boldsymbol{\alpha}$, with A_{itk} denoting the AUC over interval $(c_{k-1}, c_k]$ or $A_{itk} = \int_{c_{k-1}}^{c_k} OTC_{it}(c)dc$ for individual i at time t, with i and t as defined above. Note that $(c_{k-1}, c_k]$ is resulting from merging many $(c_{j-1}, c_j]$ intervals by the fusion technique based on MIO.

4.3 Constrained Quasi-Estimation

We invoke the means of QIF [Song et al., 2009] with proper fusion constraints to estimate model parameters in that we treat the serial correlation as nuisance and thus do not estimate it. The competing population-average effect longitudinal methodology of Generalized Estimating Equations (GEE) does require the estimation of this serial correlation, and is thus not the appropriate quasi-estimation model in our setting. The constrained QIF enables us to establish our Supervised Fusion Learning (SFL) that is deemed statistically consistent and computationally efficient.

4.3.1 Quadratic Inference Function (QIF)

Quasi-likelihood models are often used to estimate the marginal population-average effect in longitudinal models, including GEE [Zeger et al., 1988] and QIF [Song et al., 2009]. While both GEE and QIF can be leveraged to provide estimation and inference in longitudinal models, QIF has several advantages in comparison to GEE. First, QIF estimators are more efficient than GEE estimators when the working correlation structure is unknown or misspecfied. The QIF framework also provides a goodness-of-fit BIC-type measurement for the mean-model specification, which is particularly necessary when comparing models with different number of K activity windows. Additionally, only the QIF framework, and not the GEE framework, incorporates the minimization of a quadratic objection function, which is integral to the Mixed Integer Optimization (MIO) formulation (discussed further in Section 4.3.2). Furthermore, QIF does not need to estimate the correlation R matrix in Equation (4.1) explicitly.

The QIF model has been previously described in detail [Song et al., 2009]. Briefly, the methodology is based on minimizing the QIF objective function, analogous to twice the log-likelihood, defined by:

$$Q_n(\eta) = n\bar{g}_n(\eta)^\top \bar{C}_n^{-1}(\eta)\bar{g}_n(\eta) \tag{4.4}$$

,

where

$$\bar{g}_n(\eta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\eta) \approx \frac{1}{n}\begin{pmatrix} \sum_{i=1}^{n}(\dot{\mu}_i)^\top V_i(y_i - \mu_i) \\ \sum_{i=1}^{n}(\dot{\mu}_i)^\top V_i^{\frac{1}{2}} M_1 V_i^{\frac{1}{2}}(y_i - \mu_i) \\ \vdots \\ \sum_{i=1}^{n}(\dot{\mu}_i)^\top V_i^{\frac{1}{2}} M_B V_i^{\frac{1}{2}}(y_i - \mu_i) \end{pmatrix} \tag{4.5}$$

.

The QIF estimator $Q_n(\eta) = \underset{\eta}{argmin}\ Q_n(\eta)$. Here, $(\dot{\mu}_i) = \frac{\partial \mu_i}{\partial \eta}$, where $\mu_i = (\mu_{i1}, \ldots, \mu_{im})$ represents the marginal mean function of Equation (4.3), and $M_1, \cdots, M_B$ are known basis matrices with 0 or 1 as the components, while V_i is the diagonal matrix of the marginal variances, $var(y_{it})$, which equals to σ^2 in the case of normal errors. Additionally, $\bar{C}_n^{-1}(\eta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\eta) g_i^\top(\eta)$, the sample covariance matrix of $g_i(\eta)'s$. Note that the nuisance parameter R is not present in the quadratic objective function (4.4).

In the scenario of two time points, $t \in \{1, 2\}$, under the assumption of compound symmetry, we have $M_b \in \{M_0, M_1\}$ with $M_0 = I_{2\times 2}$ and $M_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Under the additional assumption of constant variance σ^2 in normal distribution, σ^2 does not affect the minimization of $Q_n(\eta)$ in Equation (4.4), so we can set $V_i = I_{(2\times 2)}$. Moreover, we have the following extended score vector:

$$g_i(\eta) = \begin{pmatrix} (\dot{\mu}_i)^\top M_0 (y_i - \mu_i) \\ (\dot{\mu}_i)^\top M_1 (y_i - \mu_i) \end{pmatrix} = \begin{pmatrix} (\dot{\mu}_i)^\top \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} (y_i - \mu_i) \\ (\dot{\mu}_i)^\top \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} (y_i - \mu_i) \end{pmatrix}, \tag{4.6}$$

where $(y_i - \mu_i) = \begin{pmatrix} y_{i1} - \mu_{i1} \\ y_{i2} - \mu_{i2} \end{pmatrix}$ and $(\dot{\mu}_i) = \frac{\partial \mu_i}{\partial \boldsymbol{\eta}} = \begin{pmatrix} \frac{\partial \mu_{i1}}{\partial \boldsymbol{\eta}} \\ \frac{\partial \mu_{i2}}{\partial \boldsymbol{\eta}} \end{pmatrix}$. Specifically, the longitudinal marginal mean function μ_{it} from Equation (4.3) takes the form:

$$\begin{aligned} \mu_{it} &= \sum_{j=1}^{J} \theta_j A_{itj} + \boldsymbol{z}_{it}^\top \boldsymbol{\alpha} = \begin{bmatrix} A_{it1} & A_{it2} & \cdots & A_{itJ} & z_{it1} & \cdots & z_{itq} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_J \\ \alpha_1 \\ \vdots \\ \alpha_q \end{bmatrix} \\ &= \left(\boldsymbol{A}_{it}^\top, \boldsymbol{z}_{it}^\top\right) \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\alpha} \end{pmatrix} \\ &:= \boldsymbol{x}_{it}^\top \boldsymbol{\eta}, \text{ with } \boldsymbol{x}_{it} \in \mathbb{R}^{J+q}, \boldsymbol{\eta} \in \mathbb{R}^{J+q}. \end{aligned}$$

Thus, $\frac{\partial \mu_{it}}{\partial \boldsymbol{\eta}} = \boldsymbol{x}_{it}^\top$, a $(J + q)$-dimensional vector.

With $m = 2$, we have:

$$\mu_i = \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_{i1}^\top \boldsymbol{\eta} \\ \boldsymbol{x}_{i2}^\top \boldsymbol{\eta} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_{i1}^\top \\ \boldsymbol{x}_{i2}^\top \end{pmatrix} \boldsymbol{\eta} := \boldsymbol{X}_i^\top \boldsymbol{\eta},$$

where $\boldsymbol{X}_i^\top$ has dimension $2\times(J+q)$ and $\dot{\mu}_i = \frac{\partial \mu_i}{\partial \boldsymbol{\eta}} = \boldsymbol{X}_i^\top$, which is a $2\times(J+q)$ dimensional matrix. Thus, the extended score vector of the QIF objective function (4.4) can be represented as the $2(J+q) \times 1$ vector:

$$\begin{aligned} \bar{g}_n(\boldsymbol{\eta}) &= \frac{1}{n} \sum_{i=1}^{n} g_i(\boldsymbol{\eta}) \\ &= \frac{1}{n} \begin{pmatrix} \sum_{i=1}^{n} (\boldsymbol{X}_i) \begin{pmatrix} y_{i1} - \mu_{i1} \\ y_{i2} - \mu_{i2} \end{pmatrix} \\ \sum_{i=1}^{n} (\boldsymbol{X}_i) \begin{pmatrix} y_{i2} - \mu_{i2} \\ y_{i1} - \mu_{i1} \end{pmatrix} \end{pmatrix}. \end{aligned} \tag{4.7}$$

4.3.2 Mixed Integer Optimization Formulation (MIO) for QIF

Here we propose an adaptation of the Bertismas MIO framework [Bertsimas et al., 2016] for a fusion-adapted longitudinal L_0 constraint formulation. This aim of this proposed framework is to conduct concurrent parameter fusion and changepoint detection to analyze the mean model (4.3). The number of groupings is controlled by fixing the number of desired clusters K, which is tuned by a BIC-type goodness-of-fit measures for QIF models. As an important element of the MIO constraints, we first consider the window labeling variables $\boldsymbol{\zeta}$, which identify group (or window) membership in each activity window such that:

$$\boldsymbol{\zeta}_k = (\zeta_k^1, \zeta_k^2, \cdots, \zeta_k^J)^\top \in \{0, 1\}^{J\times 1}, \; k = 1, \cdots, K, \tag{4.8}$$

where $\zeta_k^j = 1$ corresponds to the case of β_j belonging in activity window k. Given cutoffs or edges of windows, $c_1, \ldots, c_K$ with $c_K = J$, such binary group labels take values:

$$\zeta_1^j = \begin{cases} 1, & j = 1, \cdots, c_1 \\ 0, & otherwise \end{cases}, \zeta_k^j = \begin{cases} 1, & j = c_{k-1}, \cdots, c_k \\ 0, & otherwise \end{cases}, \cdots, \zeta_K^j = \begin{cases} 1, & j = c_{K-1} + 1, \cdots, J \\ 0, & otherwise \end{cases}. \tag{4.9}$$

For a K-group model, a fusion-adapted L_0 constrained optimization with J original intervals and q covariates is represented as:

$$
\begin{aligned}
\min_{\boldsymbol{\eta},\boldsymbol{\zeta},\boldsymbol{\beta},\boldsymbol{c}} \quad & Q_n(\boldsymbol{\eta}) = n\bar{g}_n(\boldsymbol{\eta})^\top \bar{C}_n^{-1}(\boldsymbol{\eta})\bar{g}_n(\boldsymbol{\eta}) \\
\text{subject to} \quad & \boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\alpha} \end{pmatrix}; \\
& \mathbf{c} = (c_1, \cdots, c_{K-1}) \in \mathbb{N}^{1\times(K-1)} \\
& c_1 \geq 1,\ c_k \geq c_{k-1} + 1,\ c_{K-1} \leq J - 1, k = 1, \cdots K - 1; \\
& \boldsymbol{\zeta} = (\zeta_j^k)_{J\times K} \in \mathbb{R}^{K\times J} \\
& \theta_j - \beta_1 = 0,\ j = 1, \cdots, c_1; \\
& \theta_j - \beta_2 = 0,\ j = c_1 + 1, \cdots, c_2; \\
& \vdots \\
& \theta_j - \beta_K = 0,\ j = c_{K-1} + 1, \cdots, J
\end{aligned}
\tag{4.10}
$$

where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_J)^\top \in \mathbb{R}^{J\times1}$, $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_q)^\top \in \mathbb{R}^{q\times1}$, and $\bar{g}_n(\boldsymbol{\eta})$ is as defined in Equation (4.7). The above optimization is operated via augmented parameters where labels $\boldsymbol{\zeta}$ and cutpoints $\boldsymbol{c} = (c_1, \ldots, c_{K-1})^\top$ do not exist in the original model (4.2) but are added for parameter fusion. Obviously, group labels $\boldsymbol{\zeta}$ and cutpoints $\boldsymbol{c}$ are determined in a one-to-one correspondence fashion, which will be enforced via adequate constraints given below in (4.11).

4.3.3 Implementation of MIO QIF model

This MIO model can be solved via numerical software such as GUROBI under a system of constraints. These constraints are set to minimize the objective function by optimizing cutpoints $c_1, \cdots, c_{K-1}$ and thus the cluster labels represented by the label variable $\zeta_k,$, $k = 1, \cdots, K$ defined in (4.8) and (4.9). This set of linear constraints for the K-group model is specified as follows:

$$
\begin{aligned}
&\zeta_k^j(\theta_j - \beta_k) = 0,\ \ j = 1, \cdots, J,\ k = 1, \cdots, K \text{ (SOS-1 constraints);}\\
&\sum_{k=1}^{K} \zeta_k^j = 1,\ j = 1, \cdots, J;\\
&c_0 = 0, c_1 \geq 1;\ c_k \geq c_{k-1} + 1;\ \text{and } c_{K-1} \leq J - 1,\ \text{for } k = 2, \cdots, K-1;\\
&\frac{c_k - j}{J} \leq 1 - \zeta_{k+1}^j,\ j = 1, \cdots, J,\ \text{for } k = 0, \cdots, K-1;\\
&\frac{c_{k+1} - j}{J} \times \frac{(j - c_k)}{J} \leq \zeta_{k+1},\ j = 1, \cdots, J,\ \text{for } k = 0, \cdots, K-1;\\
&\frac{j - c_{k+1} + 1}{J} \leq 1 - \zeta_{k+1},\ j = 1, \cdots, J,\ \text{for } k = 0, \cdots, K-1.
\end{aligned}
\tag{4.11}
$$

These constraints determine the locality of changepoints and grouping in the QIF-fusion-adapted MIO formulation of a K-group model. In this chapter, the constraints are implemented in the GUROBI numerical solver package in Python. A recent paper [Wang et al., 2022] showed that the MIO GUROBI optimization solvers provide the global optimal solutions for a similar homogeneity fusion problem.

The calculation of the inverse of weighting matrix $\bar{C}_n(\boldsymbol{\eta})$, the sample covariance matrix of matrix $g_i(\boldsymbol{\eta})$, is an important step in the QIF formulation. Here, we will estimate the sample covariance matrix based on initial estimates of $\boldsymbol{\eta}$, or $\boldsymbol{\eta}^{(0)}$, as estimated from a QIF formulation under working independence correlation, namely the Identity Matrix $\bar{C}_n^{(0)} = I$. Then, we estimate: $\widehat{\bar{C}_n(\boldsymbol{\eta}^{(0)})} = \widehat{Var(g_i(\boldsymbol{\eta}^{(\mathbf{0})})} = \frac{1}{n}\sum_{i=1}^{n} g_i(\boldsymbol{\eta}^{(0)}) g_i(\boldsymbol{\eta}^{(0)})^\top$, a matrix with dimension $2(J + q) \times 2(J + q)$. It is important to note that the initial estimator $\boldsymbol{\eta}^{(0)}$ is consistent and QIF does not require a correctly specified sample covariance matrix, though a more appropriate covariance estimate results in increased efficiency of QIF estimation. That is, an incorrectly specified covariance matrix does not affect estimation consistency, though does affect estimation efficiency.

Lastly, to ensure the appropriate format for the GUROBI minimization software, which requires the minimization of a quadratic objective function, we translate the QIF problem below into a quadratic form via the means of Single Value Decomposition (SVD). That is, we have the QIF objective function of the form:

$$
\begin{aligned}
Q_n(\boldsymbol{\eta}) &= n \bar{g}_n(\boldsymbol{\eta})^\top \bar{C}_n^{-1}(\boldsymbol{\eta}) \bar{g}_n(\boldsymbol{\eta})\\
&= n \left(\frac{1}{n} \begin{pmatrix} \sum_{i=1}^{n} (\boldsymbol{X}_i) \begin{pmatrix} y_{i1} - \mu_{i1} \\ y_{i2} - \mu_{i2} \end{pmatrix} \\ \sum_{i=1}^{n} (\boldsymbol{X}_i) \begin{pmatrix} y_{i2} - \mu_{i2} \\ y_{i1} - \mu_{i1} \end{pmatrix} \end{pmatrix} \right)^\top Var^{-1}(g_i(\boldsymbol{\eta}^{(0)}) \left(\frac{1}{n} \begin{pmatrix} \sum_{i=1}^{n} (\boldsymbol{X}_i) \begin{pmatrix} y_{i1} - \mu_{i1} \\ y_{i2} - \mu_{i2} \end{pmatrix} \\ \sum_{i=1}^{n} (\boldsymbol{X}_i) \begin{pmatrix} y_{i2} - \mu_{i2} \\ y_{i1} - \mu_{i1} \end{pmatrix} \end{pmatrix} \right).
\end{aligned}
\tag{4.12}
$$

.

By SVD we have $\bar{C}(\boldsymbol{\eta}^{(0))}) = \boldsymbol{P}\Lambda\boldsymbol{P}^{\top}$, where $\boldsymbol{P}$ is a square orthogonal matrix of dimension $2(J+q) \times 2(J+q)$, with $P^{\top}P = I$ and $\Lambda = diag(\lambda_1, \ldots, \lambda_{2(J+q)})$. Thus we have $\bar{C}^{-1}(\boldsymbol{\eta}^{(0))}) = \boldsymbol{P}^{\top}\Lambda^{-1}\boldsymbol{P}$. Consequently, we re-express QIF objective function (4.12) as the following quadratic form:

$$\begin{aligned} Q_n(\boldsymbol{\eta}) &= n\bar{g_n}(\boldsymbol{\eta})^{\top}\boldsymbol{P}^{\top}\Lambda^{-1}\boldsymbol{P}\bar{g_n}(\boldsymbol{\eta}) \\ &= n\{\boldsymbol{P}\bar{g_n}(\boldsymbol{\eta})\}^{\top}\Lambda^{-1}\{\boldsymbol{P}\bar{g_n}(\boldsymbol{\eta})\} \\ &= n\sum_{l=1}^{2(J+q)} \lambda_l^{-1}(\boldsymbol{P}\bar{g_n}(\boldsymbol{\eta}))_l^2 \end{aligned} \tag{4.13}$$

where $\boldsymbol{P}\bar{g_n}(\boldsymbol{\eta}))_l$ is the l^{th} element of the vector $(\boldsymbol{P}\bar{g_n}(\boldsymbol{\eta}))$ for $l = 1, \cdots, 2(J+q)$.

In the simplest case with $\bar{C}(\boldsymbol{\eta}^{(0))}) = I_{2(J+q)\times 2(J+q)}$, we have $Q_n(\boldsymbol{\eta}) = n\sum_{l=1}^{2(J+q)}(\bar{g_n}(\boldsymbol{\eta}))_l^2$. This simplified version is leveraged to obtain initial consistent estimates $\boldsymbol{\eta}^{(0)}$, thus generating an estimate of the weighting matrix $Var^{-1}(g_i(\boldsymbol{\eta}^{(0)})$. Of interest, this simplified version can be re-written using only summary statistics. That is, the expression $\sum_{i=1}^{n}(\boldsymbol{X}_i)\begin{pmatrix} y_{i1} - \mu_{i1} \\ y_{i2} - \mu_{i2} \end{pmatrix}$ can be re-expressed in matrix form as:

$$\boldsymbol{X}(\boldsymbol{Y} - \mu) = \boldsymbol{X}(\boldsymbol{Y} - \boldsymbol{X}^{\top}\boldsymbol{\eta}) = \boldsymbol{X}\boldsymbol{Y} - \boldsymbol{X}^{\top}\boldsymbol{X}\boldsymbol{\eta}. \tag{4.14}$$

Importantly, both the terms $\boldsymbol{X}\boldsymbol{Y}$ and $\boldsymbol{X}^{\top}\boldsymbol{X}$ in Equation (4.14) only require summary statistics from the sample data. Thus, the initial parameter matrix $\boldsymbol{\eta}^{(0)}$ can be estimated without individual level data.

The fusion-adapted L_0 QIF model with associated constraints is utilized to finalize the the changepoints $c_1, \cdots, c_K$. With these changepoint estimates $\hat{c}_1, \cdots, \hat{c}_K$, we further fit the QIF model $y_{it} \sim \sum_{k=1}^{K}\beta_k A_{itk} + \boldsymbol{z}_{it}^{\top}\boldsymbol{\alpha}$ using the R Package "qif" [Song et al., 2009] to determine the final parameter estimations and conduct inference. It is important to note that the weighting matrix $Var^{-1}(g_i(\boldsymbol{\eta}^{(0)})$ needs to be positive definite in order for optimal QIF model fit. However, in some cases, such as in the case of compound symmetry or AR-1 correlation structure, the estimate weighting matrix $Var^{-1}(g_i(\boldsymbol{\eta}^{(0)})$ may be singular. Thus, strategies as proposed by Hu [Hu and Song, 2012] and Han [Han and Song, 2011] may be employed to overcome this non-invertible nature in the estimation process and provide a consistent and asymptotically optimal solution. With an appropriate weighting matrix employed, the QIF package provides inference measurements for both individual parameters of interest as well as overall model goodness-of-fit, the latter of which the competing method of GEE does not provide. Thus, we are able to both determine the model's compatibility with data, and conduct inference on the association of specified PA windows on the health outcome of interest.

4.4 Theoretical Guarantees

Here we discuss the selection consistency of the MIO estimator of parameter $\boldsymbol{\theta} = (\theta_1, \dots, \theta_J)^T$ obtained by the constrained QIF optimization given in Equations (4.10) and (4.11) under some mild regularity conditions. This paves the theoretical basis for large-sample statistical inference. We follow a similar approach as in Chapter 3 with a few important updates to account for the serial dependence in the repeated measurement data. The theoretical guarantees are again based on arguments given in [Wang et al., 2022].

To present the sufficient conditions for selection consistency, we first introduce the oracle estimators that represent the parameter estimates under the true number of clusters K^* and cutpoints $c^* = (c_1^*, \cdots, c_K^*)$. We denote the oracle estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ as $\hat{\boldsymbol{\beta}}^{ol}$ and $\hat{\boldsymbol{\alpha}}^{ol}$ respectively, which are obtained through the minimizing of the quadratic inference function:

$$\hat{\boldsymbol{\eta}}^{ol} = (\hat{\boldsymbol{\beta}}^{ol}, \hat{\boldsymbol{\alpha}}^{ol}) := \underset{\boldsymbol{\beta},\boldsymbol{\alpha}}{\operatorname{argmin}}\ Q_n(\boldsymbol{\eta}) = n\bar{g}_n(\boldsymbol{\eta})^\top \bar{C}_n^{-1}(\boldsymbol{\eta})\bar{g}_n(\boldsymbol{\eta}),$$

where $\boldsymbol{\eta} = (\boldsymbol{\theta}^\top, \boldsymbol{\alpha}^\top)^\top \in \mathbb{R}^{J+q}$ as defined in Section 4.3.

When these cutpoints are unknown, we propose to use the QIF fusion-adapted MIO approach to obtain consistent estimators of both the model parameters and cutoff values. In order to achieve this, we aim to minimize our constrained objective function in Equations (4.10) and (4.11) in order to estimate cutpoints $\boldsymbol{c}$, thereby reducing the J individual-level parameters θ_j's into K group-level parameters β_k's via suitable constraints. Using a similar theoretical argument as in Chapter 3, we again adopt a measure of Mean Squared Error (MSE) sensitivity deemed c_{min} to quantify the sensitivity of the model to the precision of clustering, thereby quantifying the minimum increase of MSE due to an inaccurately-determined set of cutpoints $\boldsymbol{c}$. That is, we define:

$$c_{min} \equiv c_{min}(\boldsymbol{\xi}^*, \boldsymbol{A}, \boldsymbol{z}) = \min_{\boldsymbol{\xi}} \frac{\|\boldsymbol{A}(\boldsymbol{\theta} - \boldsymbol{\beta}^*) + \boldsymbol{z}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)\|_2^2}{n \max(d(\boldsymbol{\theta}, \boldsymbol{\beta}^*), 1)}, \quad \text{subject to Equations (4.10) and (4.11).} \tag{4.15}$$

with the true values $\boldsymbol{\xi}^* = (\boldsymbol{\beta}^{*\top}, \boldsymbol{\alpha}^{*\top}, \boldsymbol{c}^*)^\top \in \mathbb{R}^{2K-1+q}$, estimate $\boldsymbol{\xi} = (\boldsymbol{\theta}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{c})^\top \in \mathbb{R}^{J+K+q-1}$, and $d(\boldsymbol{\theta}, \boldsymbol{\beta}^*)$ represents a grouping incongruity measure reflecting the accuracy of the cutpoint estimation. See more details in [Wang et al., 2022]. In the repeated measures framework with data collected at m time-points, we extend an n-dimensional vector of $i.i.d.$ univariate normal error terms as in Chapter 3 to an n $i.i.d.$ m-element normally distributed random vectors with zero mean vector and covariance matrix $\Sigma = \sigma^2 R$ with $R \in \mathbb{R}^{m \times m}$ being a correlation matrix and variance $\sigma^2 < \infty$. Here we assume a well-defined covariance matrix Σ whose minimal and maximal eigenvalues satisfy $0 < \lambda_{min}(R) \leq \lambda_{max}(R) < \infty$.

Similar to Chapter 3, we present the selection consistency by first defining a QIF MIO estimator $\hat{\boldsymbol{\xi}}^{\boldsymbol{MIO}} = (\hat{\boldsymbol{\beta}}^{MIO^\top}, \hat{\boldsymbol{\alpha}}^{MIO^\top})^\top$ of $\boldsymbol{\xi}^*$ with true number of windows K^* and associated estimated cutpoints $\hat{\boldsymbol{c}}^{MIO}$ of $\boldsymbol{c}^*$. We also define a loss function $L(\hat{\boldsymbol{\xi}}^{\boldsymbol{MIO}}; \boldsymbol{\xi}^*)$ as the grouping risk associated with inaccurate grouping and estimates of $\boldsymbol{c}^*$ as $L(\hat{\boldsymbol{\xi}}^{\boldsymbol{MIO}}; \boldsymbol{\xi}^*) \equiv \mathbb{P}(\hat{\boldsymbol{\xi}}^{\boldsymbol{MIO}} \neq \boldsymbol{\xi}^*)$. Under the conditions for covariance matrix Σ with bounded eigenvalues, when $K = K^* < \infty$ and for any J, we can establish a finite sample error bound given by:

$$L(\hat{\boldsymbol{\xi}}^{\boldsymbol{MIO}}; \boldsymbol{\xi}^*) \leq (4m)\, exp\left[-\frac{3N}{200\sigma^2} \left\{ c_{min} - \frac{\sigma^2}{N} (134 \log(JK^*) + 220) \right\} \right]. \tag{4.16}$$

This bound in (4.16) implies that when $c_{min} > \frac{\sigma^2}{N}(134 \log(JK^*) + 220)$, $\hat{\boldsymbol{\xi}}^{\boldsymbol{MIO}}$ consistently reconstructs $\boldsymbol{\xi}^*$ because $N, J \longrightarrow \infty$, $\mathbb{P}(\hat{\boldsymbol{\xi}}^{\boldsymbol{MIO}} \neq \boldsymbol{\xi}^*) \longrightarrow 0$. Thus for any finite fixed J, or for $J \longrightarrow \infty$, $\hat{\boldsymbol{\xi}}^{\boldsymbol{MIO}}$ consistently reconstructs $\boldsymbol{\xi}^*$ as $N \longrightarrow \infty$. Note that in the former case of fixed J, the above condition for c_{min} holds automatically as $N \longrightarrow \infty$, and thus the selection consistency can be yielded under very mild conditions. The proof of this sufficient condition result can be carried out by following the lines of arguments given in the proof of Theorem 3.4 in [Wang et al., 2022], with some minor adaptations to accommodate the repeated measurements using Bonferroni inequity. These adaptations are described below.

Proof Adaptations: Let $\boldsymbol{\eta} \in \boldsymbol{\Theta}(k)$ denote the set of all k-group models. For any MIO estimate grouping $\mathbb{G}(\boldsymbol{\theta})$ such that $\boldsymbol{\eta} = (\boldsymbol{\theta}^\top, \boldsymbol{\alpha}^\top)^\top \in \Theta(K^*)$, define $\boldsymbol{P}_{\mathbb{G}(\theta)}$ as the projection matrix onto the column space of $(\boldsymbol{A}, \boldsymbol{z})$. The error matrix is assumed to be $n \times m$ with row-wise independence. Let ϵ_t denote the n-dimensional vector of $i.i.d.$ cross-sectional normal error terms at time t. For any $\boldsymbol{\eta}$ satisfying $\boldsymbol{\Theta} \in (K^*)$ and $\mathbb{G}(\boldsymbol{\theta}) \neq \mathbb{G}(\boldsymbol{\theta}^*)$, we have by Bonferroni inequality:

$$
\begin{aligned}
&\mathbb{P}\Big(\min_{\tilde{\boldsymbol{\eta}}\in\Theta(K^*),\mathbb{G}(\tilde{\boldsymbol{\theta}})=\mathbb{G}(\boldsymbol{\theta})} \|\boldsymbol{Y}-(\boldsymbol{A},\boldsymbol{z})\tilde{\boldsymbol{\eta}}\|_2^2 < \|\boldsymbol{Y}-(\boldsymbol{A},\boldsymbol{z})\boldsymbol{\eta}^{\hat{ol}}\|_2^2\Big) \\
=&\mathbb{P}\Big(2\sum_{t=1}^{m}\boldsymbol{\epsilon}_t^\top(\boldsymbol{I}-\boldsymbol{P}_{\mathbb{G}(\theta)}(\boldsymbol{A},\boldsymbol{z})\boldsymbol{\theta}^* + \|(\boldsymbol{I}-\boldsymbol{P}_{\mathbb{G}(\theta)}(\boldsymbol{A},\boldsymbol{z})\boldsymbol{\theta}^*\|_2^2 - \sum_{t=1}^{m}\boldsymbol{\epsilon}_t^\top(\boldsymbol{P}_{\mathbb{G}(\theta)}-\boldsymbol{P}_{\mathbb{G}(\theta^*)})\boldsymbol{\epsilon}_t < 0\Big) \\
\le&\sum_{t=1}^{m}\mathbb{P}\Big(2\boldsymbol{\epsilon}_t^\top(\boldsymbol{I}-\boldsymbol{P}_{\mathbb{G}(\theta)}(\boldsymbol{A},\boldsymbol{z})\boldsymbol{\theta}^* + \frac{1}{m}\|(\boldsymbol{I}-\boldsymbol{P}_{\mathbb{G}(\theta)}(\boldsymbol{A},\boldsymbol{z})\boldsymbol{\theta}^*\|_2^2 - \boldsymbol{\epsilon}_t^\top(\boldsymbol{P}_{\mathbb{G}(\theta)}-\boldsymbol{P}_{\mathbb{G}(\theta^*)})\boldsymbol{\epsilon}_t < 0\Big) \\
:=&\sum_{t=1}^{m}P_t
\end{aligned}
\tag{P.1}
$$

For each $t \in \{1,\ldots,m\}$, we evaluate each P_t term. For any $0 < \delta_t < 1$, we have:

$$
\begin{aligned}
P_t \text{ in Equation (P.1)} \le&\mathbb{P}\Big(2\boldsymbol{\epsilon}_t^\top(\boldsymbol{I}-\boldsymbol{P}_{\mathbb{G}(\theta)}(\boldsymbol{A},\boldsymbol{z})\boldsymbol{\theta}^* + \frac{\delta_t}{m}\|(\boldsymbol{I}-\boldsymbol{P}_{\mathbb{G}(\theta)}(\boldsymbol{A},\boldsymbol{z})\boldsymbol{\theta}^*\|_2^2 < 0\Big) + \\
&\mathbb{P}\Big((1-\frac{\delta_t}{m})\|(\boldsymbol{I}-\boldsymbol{P}_{\mathbb{G}(\theta)}(\boldsymbol{A},\boldsymbol{z})\boldsymbol{\theta}^*\|_2^2 - \boldsymbol{\epsilon}_t^\top(\boldsymbol{P}_{\mathbb{G}(\theta)}-\boldsymbol{P}_{\mathbb{G}(\theta^*)})\boldsymbol{\epsilon}_t < 0\Big)
\end{aligned}
\tag{P.2}
$$

From the point of (P.2), and by setting suitable δ_t, the proof of this sufficient condition result can be carried out by following the lines of arguments given in the proof of Theorem 3.4 in [Wang et al., 2022] and thus is omitted here.

4.5 Simulation Experiments

Simulation experiments demonstrate robust, reliable performance of the proposed fusion-adapted QIF MIO paradigm. In this section we discuss the setup of the conducted numerical experiments, report on their results, and discuss the computational requirements of this constrained QIF methodology.

4.5.1 Simulation Setup

4.5.1.1 OTC Data Generation

In order to conduct numerical experiments on our proposed longitudinal QIF MIO functional framework with OTCs, we first simulated repeated accelerometer measures. More specifically, we simulated 6-hour time-series of VM counts for 500 individuals, each at two different time points. To achieve this, we created permuted accelerometer time-series data using the subjects from the ELEMENT dataset, with the simulated data for the first time point created from permutations of the ELEMENT Time 1 data (539 subjects), and the simulated data for second time point created from permutations of the ELEMENT Time 2 data (496 subjects). For each of the $t = 1, 2$ time points, we first divided the time stamped accelerometer data into non-overlapping 10-minute segments before randomly drawing each 10-minute interval from the pool of candidate segments.

In order to ensure the variability of the simulated datasets reflected the variability of our motivating ELEMENT dataset, we permuted the time-series VM counts for low, medium, and high levels of PA. That is, at each T1 and T2 separately, we first classified the subjects into three groups with low, medium, and high levels of PA respectively, as defined by tertiles of "Moderate-to-Vigorous" VM counts using the pre-set Chandler cutoffs [Chandler et al., 2016] before simulating the time-series data within each tertile.

Lastly, with the simulated VM counts for 500 subjects at two different time points, OTC curves were calculated as described in Section 2.3. For each of the OTCs, we calculated the $J = 300$ successive integrals (i.e. AUCs) over domain $\mathcal{C} = (0, 30000)$, with each interval covering 100 VM counts: $(c_0 = 0, c_1 = 100, \cdots, c_J = 30000)$. In this chapter, we will refer to the $VM/100$ values, i.e. $\boldsymbol{c} = (0/100, \cdots, 30000/100)^\top$ or $\boldsymbol{c} = (0, \cdots, 300)^\top$.

4.5.1.2 Longitudinal Model Generation

In the simulation experiments, we assessed the proposed fusion-adapted L_0 approach's ability to detect the true cutoffs and parameter estimates within a QIF framework. We first specified the true number of $K^* = 3$ groups and corresponding true cutoffs (c_1^*, c_2^*, c_3^*) in addition to $c_0^* = 0$, and calculated the vector of AUCs, $(A_{it1}^*, A_{it2}^* A_{it3}^*)^\top$ for each individual i with $A_{itk}^* = \int_{c_{k-1}}^{c_k} OTC_{it}(c)dc$ at both time points $t \in \{1, 2\}$ and scaled the variables to normal distribution of mean 0 and variance 1. Finally, we generated longitudinal outcome Y_{it} from the zero-intercept longitudinal linear model $Y_{it} = \sum_{k=1}^{3} A_{itk}^* \beta_k^* + z_{it}\alpha^* + \epsilon_{it}$, with true effect sizes $(\beta_1^*, \beta_2^*, \beta_3^*)$ and α^* as well as a single continuous covariate $z_{it} \overset{iid}{\sim} N(0, 1)$. To simulate the correlated error terms in the longitudinal framework, ϵ_{it} were simulated from a bivariate normal distribution with $\boldsymbol{\epsilon}_i = \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{pmatrix} \sim$ BivariateNormal $\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$, with $\sigma = 10$ and varying ρ.

To conduct this assessment under a variety of scenarios, we specified various 3-group models ($K^* = 3$) under a combination of different VM count changepoints (c_1^*, c_2^*, c_3^*), to achieve different window sizes, as well as various effect sizes $(\beta_1^*, \beta_2^*, \beta_3^*)$. Specifically, we specified three scenarios: (A) VM count changepoints $(c_1^*, c_2^*, c_3^*) = (40, 80, 300)$ with activity window parameters $(\beta_1^*, \beta_2^*, \beta_3^*) = (4, 0, -4)$, (B) VM count changepoints $(c_1^*, c_2^*, c_3^*) = (40, 80, 300)$ with activity window parameters $(\beta_1^*, \beta_2^*, \beta_3^*) = (1, 0, -1)$, and (C) VM count changepoints

$(c_1^*, c_2^*, c_3^*) = (40, 120, 300)$ with activity window parameters $(\beta_1^*, \beta_2^*, \beta_3^*) = (4, 0, -2)$. Additionally, these three scenarios were assessed under different assumptions of correlated longitudinal error terms, with $\rho \in \{0.2, 0.5\}$.

4.5.2 Simulation Implementation

With the generated data, the simulation experiments were implemented in a 2-step process. In Step 1, we apply the MIO-adapted QIF formulation using GUROBI to achieve estimation of cutpoints c_1^*, c_2^*, c_3^* with initial point-estimates of the parameters. This process involves both an "Initial MIO Estimate" and a "Final MIO Estimate". The "Initial MIO Estimate" assumes identity weighting matrix for $\bar{C}_n(\boldsymbol{\eta}) = I$ and garners initial cutpoint estimates $(\hat{c}_1^{(0)}, \hat{c}_2^{(0)}, \hat{c}_3^{(0)})$ and consistent parameter estimates $\boldsymbol{\eta}^{(0)}$ leading to the estimated sample variance matrix $\bar{C}(\boldsymbol{\eta}^{(0)})$. The "Final MIO Estimate" then leverages $\bar{C}(\boldsymbol{\eta}^{(0)})$ to achieve a more efficient estimation of the cuptoints, i.e. the final estimates $(\hat{c}_1, \hat{c}_2, \hat{c}_3)$. In Step 2, we fit the final QIF model $y_{it} \sim \sum_{k=1}^{K} \beta_k A_{itk} + \boldsymbol{z}_{it}^\top \boldsymbol{\alpha}$ with the estimated cutpoints from Step 1. To assess the increased efficiency of the "Final MIO Estimate" versus the "Initial MIO Estimate", we considered QIF models with both the initial cutpoint estimates $(\hat{c}_1^{(0)}, \hat{c}_2^{(0)}, \hat{c}_3^{(0)})$ and the final cutpoint estimates $(\hat{c}_1, \hat{c}_2, \hat{c}_3)$.

The numerical experiments were conducted for two different specifications of J, the number of intervals to fuse over, $J \in \{60, 300\}$. The varying J represent two different levels of multicollinearity among the A_{itj} variables with $J = 300$ encompassing the most severe multicollinearity. The $J = 300$ scenario represents the original A_{itj} from the simulated OTC curves, with $J = 60$ calculated by merging every five successive A_{itj}'s. These two specifications of J were analyzed to assess the proposed methodology's sensitivity to the choice of original J intervals. Additionally, the simulations were conducted with three different sample sizes $N \in \{100, 250, 500\}$ with $J = 60$, and $N = 500$ when $J = 300$. Note that when $J = 300$, the method is limited to scenarios with $J < N$ as the fusion-adapted L_0 formulation does not introduce the true sparsity into the model that allows for $J > N$.

4.5.3 Simulation Performance

The simulation results produced by the fusion-adapted L_0 constraint QIF model demonstrated that this proposed approach provides reliable and efficient change-point detection and parameter estimation. It has also demonstrated robustness in handling highly correlated AUCs.

4.5.3.1 Step 1: MIO QIF Method

Tables 4.1 and 4.2 summarize the stability of the change-point detection of Step 1 from 500 rounds of simulations of the 3−group model for the $J = 300$ and $J = 60$ settings, respectively. These tables additionally report on the stability of the initial parameter estimations based on these estimated cutpoints via the mean and empircal standard error (ESE) of the point estimates.

The method maintained its ability to reliably identify changepoints as the sample size N decreased from $N = 500$ to $N = 250$, though we observed increased ESE when $N = 100$. For Scenario C with starting number of intervals $J = 60$, true cutpoints $(c_1^*, c_2^*, c_3^*) = (40, 120, 300)$ and $N = 250$, the mean (ESE) estimates of c_1^* and c_2^* from the Final MIO estimate of this L_0 constrained QIF approach are 40.05(1.75) and 120.10(1.40) when $\rho = 0.5$. As the longitudinal correlation parameter ρ decreases to 0.2, we observe similarly strong accuracy and efficiency of changepoint detection. Similar strong results are repeated in Scenarios A and B.

4.5.3.2 Step 2: Final QIF Method

Tables 4.3 - 4.5 summarize the final QIF model results, when utilizing the activity window cutpoint estimates of c_1^*, c_2^*, c_3^* from Step 1. In addition to reporting mean values for the point estimates and their empirical standard errors (ESE), as determined by calculating the standard error of the point estimates achieved from each simulation, Tables 4.3, 4.4, and 4.5 also report the inference measurements of mean p-values and the average standard error (ASE). These inference parameters were achieved from the R package "qif" using the the matrix inverse option of "generalized inverse" to estimate the weighting matrix.

In Step 2, we observe very low bias and variability in parameter estimation, both for the estimates of β_k as well as α. As expected, the efficiency of the estimates improves slightly between the Initial QIF estimates, conducted with initial cutpoint estimates $(\hat{c}_1^{(0)}, \hat{c}_2^{(0)}, \hat{c}_3^{(0)})$ and the Final QIF Estimates, conducted with the final cutpoint estimates $(\hat{c}_1, \hat{c}_2, \hat{c}_3)$.

4.5.4 Computational Requirements

The proposed method is computationally efficient. In Step 1, in which the fusion adapted QIF model is solved via GUROBI's MIO process, a 3-group simulation model with $N = 500, J = 60$ computes in ten minutes with the $J = 300$ scenario completing in two hours. The method is scalable to a reasonable number of windows. When the true 3-group model was assessed as a 4-group model, the computation times increased to one hour and eight hours for the $J = 60, 300$ scenarios respectively. Once the estimated changepoints are determined from Step 1, fitting the final QIF model in Step 2 requires trivial time.

Table 4.1: Step 1 MIO Results for J=300

		$\rho = 0.5$; $N = 500$		$\rho = 0.2$; $N = 500$	
	Truth	Mean	ESE	Mean	ESE
Scenario A					
Initial MIO Estimate					
β_1	4	4.00	0.06	4.12	1.55
β_2	0	−0.002	0.29	0.02	0.37
β_3	−4	−4.00	0.003	−4.00	0.003
c_1	40	39.91	2.15	39.65	3.74
c_2	80	80.05	1.38	79.97	1.67
α	3	2.97	0.55	2.98	0.53
Final MIO Estimate					
β_1	4	4.01	0.07	4.09	1.46
β_2	0	−0.002	0.30	0.01	0.34
β_3	−4	−4.00	0.003	−4.00	0.003
c_1	40	39.88	2.31	39.69	3.34
c_2	80	80.06	1.42	79.99	1.55
α	3	2.97	0.55	2.98	0.53
Scenario B					
Initial MIO Estimate					
β_1	1	1.07	0.59	1.07	0.52
β_2	0	−0.01	0.27	−0.005	0.27
β_3	−1	−1.00	0.003	−1.00	0.003
c_1	40	38.50	8.96	38.41	9.02
c_2	80	81.14	7.49	81.06	6.96
α	3	2.97	0.54	2.98	0.52
Final MIO Estimate					
β_1	1	1.11	0.74	1.10	0.73
β_2	0	−0.002	0.28	−0.01	0.27
β_3	−1.00	−1.00	0.004	−1.00	0.003
c_1	40	38.08	9.79	38.31	9.36
c_2	80	81.47	12.09	81.10	7.08
α	3	2.97	0.54	2.97	0.52
Scenario C					
Initial MIO Estimate					
β_1	4	4.00	0.05	4.08	1.33
β_2	0	0.001	0.04	0.01	0.10
β_3	−2	−2.00	0.004	−2.00	0.004
c_1	40	39.95	0.88	39.81	2.47
c_2	120	120	0.96	119.91	1.54
α	3	2.97	0.56	2.98	0.56
Final MIO Estimate					
β_1	4	4.22	3.46	4.41	4.74
β_2	0	0.01	0.10	0.01	0.14
β_3	−2	−2.00	0.004	−2.00	0.005
c_1	40	39.81	2.63	39.63	3.58
c_2	120	119.93	1.62	119.82	2.12
α	3	2.96	0.60	2.95	0.72

Table 4.2: Step 1 MIO Results for J=60

		ρ = 0.5						ρ = 0.2					
		N = 500		N=250		N=100		N = 500		N=250		N=100	
	Truth	Mean	ESE	Mean	ESE	Mean	ESE	Mean	ESE	Mean	ESE	Mean	ESE
Scenario A													
Initial MIO Estimate													
β_1	4	4.00	0.04	4.00	0.09	4.02	0.16	4.00	0.05	4.00	0.09	4.02	0.16
β_2	0	−0.03	0.20	−0.04	0.37	−0.04	0.56	−0.03	0.22	−0.04	0.40	−0.01	0.59
β_3	−4	−4.00	0.002	−4.00	0.004	−4.00	0.01	−4.00	0.003	−4.00	0.004	−4.00	0.01
c_1	40	40.15	1.30	41.00	2.80	39.85	4.25	40.10	1.55	40.10	2.95	39.65	4.55
c_2	80	80.15	1.05	80.25	1.85	80.35	2.80	80.15	1.15	80.30	2.00	80.25	3.00
α	3	2.93	1.60	2.78	2.15	2.96	3.06	2.94	1.58	2.77	2.12	2.96	3.04
Final MIO Estimate													
β_1	4	4.00	0.10	4.01	0.17	4.04	0.34	4.00	0.09	4.02	0.17	4.05	0.34
β_2	0	−0.03	0.34	−0.05	0.48	−0.06	0.68	−0.02	0.33	−0.04	0.54	−0.02	0.69
β_3	−4	−4.00	0.003	−4.00	0.004	−4.00	0.01	−4.00	0.003	−4.00	0.004	−4.00	0.01
c_1	40	40.00	3.25	40.00	4.00	39.70	5.85	40.00	2.55	39.85	4.50	39.40	6.05
c_2	80	80.20	1.65	80.35	2.30	80.50	3.45	80.15	1.65	80.35	2.60	80.35	3.40
α	3	2.98	1.03	2.93	1.58	4.04	0.34	2.99	0.97	2.92	1.57	4.05	0.34
Scenario B													
Initial MIO Estimate													
β_1	1	1.02	0.10	1.04	0.17	1.09	0.34	1.02	0.10	1.05	0.22	1.10	0.37
β_2	0	−0.02	0.26	−0.06	0.34	−0.11	0.53	−0.03	0.27	−0.05	0.35	−0.10	0.56
β_3	−1	−1.00	0.003	−1.00	0.005	−1.01	0.05	−1.00	0.003	−1.00	0.005	−1.01	0.05
c_1	40	39.20	8.10	39.35	10.8	39.30	14.6	39.35	8.15	38.85	11.4	38.8	15.10
c_2	80	81.25	6.70	84.05	16.10	93	40.75	81.45	6.90	82.40	16.70	92.75	40.35
α	3	2.94	1.50	2.79	2.04	2.96	2.93	2.96	1.48	2.78	2.02	2.94	2.93
Final MIO Estimate													
β_1	1	1.06	0.25	1.09	0.31	1.18	0.52	1.04	0.17	1.08	0.32	1.15	0.49
β_2	0	−0.04	0.32	−0.05	0.38	−0.08	0.79	−0.04	0.30	−0.09	0.40	−0.12	0.81
β_3	−1	−1.00	0.01	−1.00	0.01	−1.00	0.05	−1.00	0.004	−1.00	0.005	−1.00	0.05
c_1	40	38.6	11.00	38.15	13.1	38.00	16.80	39.00	10.15	39.05	13.15	38.85	17.05
c_2	80	83.50	18.00	85.15	21.85	94.05	44.35	82.90	14.85	85.87	19.60	93.95	43.35
α	3	2.93	0.94	2.87	1.56	1.18	0.52	3.02	0.94	2.81	1.52	1.15	0.49
Scenario C													
Initial MIO Estimate													
β_1	4	4.00	0.03	4.00	0.05	4.00	0.13	4.00	0.04	4.00	0.06	4.01	0.13
β_2	0	−0.001	0.03	0.001	0.05	0.003	0.11	−0.001	0.03	0.00	0.05	0.005	0.11
β_3	−2	−2.00	0.003	−2.00	0.004	−2.00	0.01	−2.00	0.003	−2.00	0.005	−2.00	0.01
c_1	40	40.05	0.50	40.00	0.95	39.95	2.25	40.05	0.60	40.00	1.05	39.90	2.30
c_2	120	120.05	0.50	120.00	1.00	120.00	2.35	120.05	0.55	120.05	1.10	119.95	2.40
α	3	2.94	1.64	2.77	2.23	2.97	3.20	2.96	1.61	2.79	2.21	2.93	3.18
Final MIO Estimate													
β_1	4	4.00	0.05	4.00	0.10	4.01	0.18	4.00	0.05	4.00	0.12	4.00	0.19
β_2	0	−0.001	0.04	−0.002	0.07	0.004	0.13	−0.002	0.04	−0.001	0.08	−0.001	0.14
β_3	−2	−2.00	0.003	−2.00	0.005	−2.00	0.01	−2.00	0.003	−2.00	0.005	−2.00	0.01
c_1	40	40.00	0.85	40.05	1.75	39.90	2.85	40.04	0.85	40.05	1.85	40.05	3.10
c_2	120	120.05	0.80	120.10	1.40	120.05	2.65	120.05	0.75	120.05	1.55	120.10	2.90
α	3	2.96	0.98	2.83	1.77	4.01	0.18	2.97	1.04	2.86	1.65	4.00	0.19

Table 4.3: Step 2 QIF Results for J=300

		ρ = 0.5; N = 500				ρ = 0.2; N = 500			
	Truth	Mean	ESE	ASE	p	Mean	ESE	ASE	p
Scenario A									
Initial MIO Estimate									
β_1	4	4.00	0.07	0.02	0.00	4.05	0.53	0.02	0.00
β_2	0	0.00	0.28	0.02	0.04	0.02	0.36	0.02	0.05
β_3	-4	-4.00	0.00	0.00	0.00	-4.00	0.00	0.00	0.00
α	3	2.98	0.38	0.39	0.00	2.99	0.34	0.35	0.00
Final MIO Estimate									
β_1	4	4.01	0.08	0.02	0.00	4.04	0.56	0.02	0.00
β_2	0	0.00	0.29	0.02	0.04	0.02	0.33	0.02	0.05
β_3	-4	-4.00	0.00	0.00	0.00	-4.00	0.00	0.00	0.00
α	3	2.99	0.38	0.39	0.00	2.99	0.34	0.35	0.00
Scenario B									
Initial MIO Estimate									
β_1	1	1.04	0.22	0.02	0.00	1.04	0.21	0.02	0.00
β_2	0	0.00	0.26	0.02	0.05	0.00	0.26	0.02	0.05
β_3	-1	-1.00	0.00	0.00	0.00	-1.00	0.00	0.00	0.00
α	3	2.98	0.38	0.39	0.00	2.98	0.34	0.35	0.00
Final MIO Estimate									
β_1	1	1.05	0.27	0.02	0.00	1.05	0.26	0.02	0.00
β_2	0	0.00	0.27	0.02	0.04	0.00	0.26	0.02	0.05
β_3	-1	-1.00	0.00	0.00	0.00	-1.00	0.00	0.00	0.00
α	3	2.98	0.38	0.39	0.00	2.98	0.34	0.35	0.00
Scenario C									
Initial MIO Estimate									
β_1	4	4.00	0.05	0.01	0.00	4.05	0.85	0.02	0.00
β_2	0	0.00	0.04	0.01	0.15	0.01	0.10	0.01	0.14
β_3	-2	-2.00	0.00	0.00	0.00	-2.00	0.00	0.00	0.00
α	3	2.98	0.38	0.39	0.00	2.98	0.34	0.35	0.00
Final MIO Estimate									
β_1	4	4.07	1.01	0.02	0.00	4.14	1.47	0.02	0.00
β_2	0	0.01	0.11	0.01	0.14	0.02	0.15	0.01	0.14
β_3	-2	-2.00	0.00	0.00	0.00	-2.00	0.00	0.00	0.00
α	3	2.98	0.38	0.39	0.00	2.98	0.37	0.35	0.00

Table 4.4: Step 2 QIF Results for $J = 60$ with longitudinal correlation parameter $\rho = 0.5$

		$\rho = 0.5$											
		N = 500				N=250				N=100			
	Truth	Mean	ESE	ASE	pval	Mean	ESE	ASE	pval	Mean	ESE	ASE	pval
Scenario A													
Initial MIO Estimate													
β_1	4	3.995	0.041	0.018	0.000	3.999	0.094	0.026	0.000	4.010	0.156	0.041	0.000
β_2	0	−0.029	0.191	0.020	0.469	−0.037	0.354	0.028	0.403	−0.030	0.539	0.045	0.278
β_3	−4	−4.000	0.003	0.002	0.000	−4.000	0.005	0.003	0.000	− 4.001	0.007	0.005	0.000
α	3	2.981	0.379	0.387	0.000	3.003	0.538	0.551	0.001	2.977	0.920	0.876	0.022
Final MIO Estimate													
β_1	4	4.002	0.097	0.016	0.000	4.008	0.150	0.023	0.000	4.021	0.223	0.038	0.000
β_2	0	−0.024	0.326	0.017	0.427	−0.042	0.459	0.025	0.333	−0.043	0.664	0.039	0.233
β_3	−4	−4.000	0.003	0.002	0.000	−4.001	0.005	0.003	0.000	−4.001	0.008	0.005	0.000
α	3	2.986	0.382	0.388	0.000	3.004	0.556	0.551	0.001	2.976	0.965	0.870	0.025
Scenario B													
Initial MIO Estimate													
β_1	1	1.012	0.075	0.019	0.000	1.019	0.113	0.028	0.000	1.037	0.178	0.047	0.000
β_2	0	-0.017	0.251	0.019	0.089	-0.052	0.330	0.026	0.079	-0.097	0.502	0.040	0.067
β_3	-1	-1.000	0.003	0.002	0.000	-1.001	0.005	0.003	0.000	-1.005	0.036	0.007	0.000
α	3	2.982	0.376	0.387	0.000	3.000	0.535	0.549	0.001	2.969	0.916	0.863	0.021
Final MIO Estimate													
β_1	1	1.029	0.126	0.018	0.000	1.043	0.158	0.026	0.000	1.063	0.215	0.044	0.000
β_2	0	-0.029	0.310	0.016	0.069	-0.042	0.371	0.022	0.053	-0.082	0.659	0.039	0.048
β_3	-1	-1.001	0.005	0.002	0.000	-1.002	0.009	0.003	0.000	-1.005	0.038	0.007	0.000
α	3	2.983	0.378	0.387	0.000	3.009	0.546	0.546	0.001	2.968	0.946	0.851	0.022
Scenario C													
Initial MIO Estimate													
β_1	4	3.998	0.028	0.015	0.000	4.000	0.052	0.021	0.000	4.004	0.127	0.034	0.000
β_2	0	-0.001	0.025	0.009	0.480	0.000	0.046	0.013	0.474	0.004	0.105	0.020	0.369
β_3	-2	-2.000	0.003	0.003	0.000	-2.000	0.005	0.004	0.000	-2.000	0.010	0.006	0.000
α	3	2.982	0.374	0.386	0.000	3.004	0.528	0.546	0.001	2.981	0.912	0.863	0.021
Final MIO Estimate													
β_1	4	4.000	0.047	0.013	0.000	3.999	0.098	0.018	0.000	4.011	0.163	0.030	0.000
β_2	0	-0.001	0.039	0.008	0.468	-0.002	0.072	0.011	0.432	0.005	0.123	0.018	0.301
β_3	-2	-2.000	0.003	0.002	0.000	-2.000	0.005	0.003	0.000	-2.000	0.010	0.005	0.000
α	3	2.984	0.375	0.385	0.000	3.008	0.544	0.543	0.001	2.976	0.935	0.854	0.022

Table 4.5: Step 2 QIF Results for $J = 60$ with longitudinal correlation parameter $\rho = 0.2$

		$\rho = 0.2$											
		N = 500				N=250				N=100			
	Truth	Mean	ESE	ASE	pval	Mean	ESE	ASE	pval	Mean	ESE	ASE	pval
Scenario A													
Initial MIO Estimate													
β_1	4	3.998	0.050	0.018	0.000	4.000	0.097	0.026	0.000	4.015	0.157	0.042	0.000
β_2	0	-0.025	0.209	0.019	0.452	-0.039	0.378	0.028	0.386	-0.008	0.565	0.045	0.262
β_3	-4	-4.000	0.003	0.002	0.000	-4.000	0.005	0.003	0.000	-4.000	0.008	0.005	0.000
α	3	2.982	0.338	0.347	0.000	3.001	0.484	0.495	0.000	2.978	0.825	0.792	0.012
Final MIO Estimate													
β_1	4	4.002	0.084	0.018	0.000	4.012	0.160	0.026	0.000	4.031	0.233	0.042	0.000
β_2	0	-0.019	0.313	0.019	0.423	-0.034	0.521	0.028	0.323	-0.010	0.667	0.043	0.215
β_3	-4	-4.000	0.004	0.002	0.000	-4.001	0.005	0.003	0.000	-4.001	0.009	0.005	0.000
α	3	2.987	0.340	0.348	0.000	3.010	0.509	0.498	0.000	2.985	0.864	0.790	0.013
Scenario B													
Initial MIO Estimate													
β_1	1	1.012	0.074	0.019	0.000	1.026	0.122	0.029	0.000	1.042	0.182	0.048	0.000
β_2	0	-0.023	0.256	0.019	0.095	-0.045	0.339	0.025	0.072	-0.088	0.519	0.039	0.060
β_3	-1	-1.000	0.003	0.002	0.000	-1.001	0.005	0.003	0.000	-1.006	0.037	0.007	0.000
α	3	2.983	0.337	0.347	0.000	3.000	0.481	0.492	0.000	2.972	0.822	0.777	0.011
Final MIO Estimate													
β_1	1	1.021	0.109	0.019	0.000	1.035	0.158	0.029	0.000	1.052	0.222	0.048	0.000
β_2	0	-0.029	0.296	0.018	0.076	-0.076	0.390	0.025	0.069	-0.109	0.707	0.045	0.056
β_3	-1	-1.001	0.004	0.002	0.000	-1.002	0.006	0.003	0.000	-1.005	0.046	0.007	0.000
α	3	2.986	0.340	0.347	0.000	3.005	0.490	0.491	0.000	2.961	0.855	0.769	0.013
Scenario C													
Initial MIO Estimate													
β_1	4	3.998	0.034	0.015	0.000	4.000	0.058	0.021	0.000	4.007	0.131	0.033	0.000
β_2	0	-0.001	0.027	0.009	0.477	-0.001	0.051	0.013	0.466	0.006	0.106	0.020	0.355
β_3	-2	-2.000	0.003	0.003	0.000	-2.000	0.005	0.004	0.000	-2.000	0.010	0.006	0.000
α	3	2.984	0.335	0.345	0.000	3.003	0.474	0.489	0.000	2.987	0.817	0.777	0.012
Final MIO Estimate													
β_1	4	3.997	0.046	0.014	0.000	3.999	0.109	0.020	0.000	4.001	0.181	0.032	0.000
β_2	0	-0.002	0.036	0.009	0.468	-0.002	0.077	0.012	0.422	-0.001	0.135	0.020	0.289
β_3	-2	-2.000	0.003	0.003	0.000	-2.000	0.006	0.004	0.000	-2.000	0.011	0.006	0.000
α	3	2.986	0.338	0.344	0.000	3.007	0.487	0.488	0.000	2.973	0.843	0.772	0.013

4.6 Data Analysis

The primary objective of this data analysis was to investigate the population-average association between PA levels and body fat, as measured longitudinally by Sub-Scapular Skin Thickness (SSST). Research has suggested that higher level of PA may be associated with lower SSST, though guidelines are less clear on what PA intensity levels have the most significant effect on this anthropometric outcome, and how this association varies in males and females in adolescence. Thus, this analysis focused on assessing the functional relationship between PA and SSST via longitudinal scalar-on-function regression model to detect distinct activity intensity windows of interest, estimate the association parameter, and perform inference. I conducted this assessment considering the full sample population, as well as in a sex-stratified manner.

To use the fusion-adapted L_0 QIF analytic in Section 4.3.2, I began by setting $J = 300$, with each interval covering 100 VM counts, followed by an augmentation scenario of $J = 60$ by summing every five successive intervals. For ease of interpretation, I considered $K \in \{2, 3, 4\}$ critical windows of PA intensity. In *Step 1: MIO QIF Model*, each setting was given a budget of 20 hours run-time. If the search did not converge within this time, the attempt was terminated and the respective combination of (J, K) disregarded from reporting. Additionally, if the search completed, but the GUROBI software reported a "sub-optimal" convergence, the combination of (J, K) was disregarded. In this data analysis example, I report the successful results for the $J = 60$ scenario. The cutpoints $c_1, \cdots, c_K$ estimated from *Step 1* were used to calculate the AUCs for each activity window K, A_{itk}, which were then each scaled to mean 0, variance 1, before performing a final QIF analysis with the zero-intercept model $y_{it} \sim \sum_{k=1}^{K} \beta_k A_{itk} + \boldsymbol{z}_{it}^{\top}\boldsymbol{\alpha}$ to achieve parameter estimation and inference.

Model selection was conducted using a combination of BIC goodness-of-fit, coupled with scientific significance. For example, in the $J = 60$ results in Table 4.6, the BIC measurement ranks the models with $K = 2$ size model as the best fit, followed by $K = 3$. However, when assessing the cutpoints of the $K = 2$-group model. we see that the windows are: Window 1: (0,59] and Window 2: (59,60]. As this model only differentiates between the last potential cutpoint, encompassing the extreme tail end of the OTCs, it is unlikely to be driven by true scientific differences, but more likely by statistical anomalies such as outliers. Thus, the $K = 3$ model is selected based on the combination of statistical and scientific significance.

The interpretation of the parameter estimates for each of the detected critical activity windows depends on the activity window's sequential location. For the first activity windows, a lower AUC represents more time spent within the specific window, and less time spent above the window. Thus, a positive parameter estimate would indicate that more time spent within the specific window is associated with a smaller outcome value. In contrast, for the last window K, a higher AUC represents more time spent within the specific window. In this case, a positive parameter estimates suggests that more time spent within the final activity window is associated with a larger outcome value.

The AUC Ratio metric was introduced in Chapter 3.7 in order to facilitate a clinically understandable interpretation of analysis results. Briefly, this AUC Ratio metric measures the amount of time a subject spends within a PA window relative to the maximum amount of possible time, thereby representing a relative level of activity for the individual within the detected window compared to the hypothetical most active person. As with the interpretation of the estimated parameter coefficients β_k's, the interpretation of the AUC Ratio also depends on its sequential location. For example, with the first Window 1, a lower AUC Ratio represents more time spent within the specific window, and less time spent above the window. In contrast, for the last window K, a higher AUC Ratio represents more time spent within the specific window. I will again interpret the data analysis results under the purview of this metric, with the relevant computational figures included in Appendix B.

4.6.1 Full Data Analysis

Section 4.1.2 introduced the longitudinal ELEMENT cohort data with repeated measurements at T1 and T2, collected approximately two years apart. From this data, there was complete accelerometry and covariate data for 429 subjects (197 boys, 232 girls) at the two distinct time points. At T1, the subjects were of mean(sd) age 14.35(2.09) years, with 18% having achieved “adult” Tanner staging status. At T2, these time-varying covariates were: mean(sd) age of 16.30(2.08) years and 48% having completed puberty, e.g. achieved Tanner stage 5. The mean (sd) increase in age from T1 to T2 was 1.95(0.36) years, with the age range increasing from (10.72, 18.06) to (12.45, 20.54). These two time points represent a vital stage in human development, as the subjects transition from adolescence to early-adulthood. The functional OTCs for the full population at time points T1 and T2 are shown in Figure 4.2. Visually, we can see no significant change between the mean activity profiles between the first and second time points. This suggests that the same PA windows at T1 and T2 are acceptable.

Table 4.6 summarizes the results for the full dataset. Based on the combination of scientific and statistical significance, the $K = 3$ model was selected. In this model, we observe a significantly negative association with Window 3 and SSST outcome; this result suggests that as individuals spend more time in the activity range (145,300], they have lower SSST measurement, suggesting lower body fat and a healthier outcome.

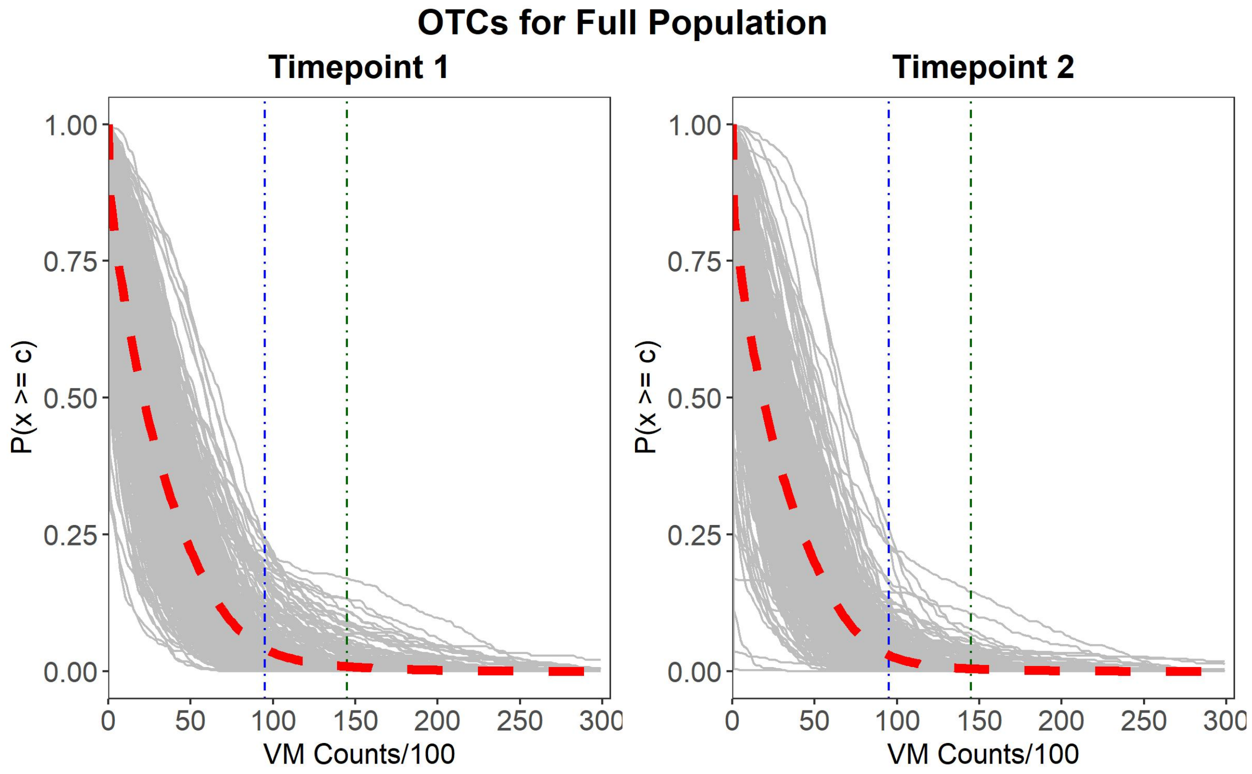


Figure 4.2: OTCs for 429 ELEMENT subjects stratified by longitudinal time point of accelerometer data collection. The relative shape of OTC reflects the subject's activity profile, with the red-dotted line representing the mean proportion of time spent above each activity level across the individuals. The first vertical line, colored blue, represents the first detected cutpoint, indicating the changepoint between critical activity Windows 1 and 2. Likewise, the second vertical line in green represents the detected cutpoint 2, signalling the changepoint from critical activity Windows 2 and 3.

Table 4.6: Full Population Data Analysis Results obtained by the fusion-adapted L_0 QIF method, where J indicates the number of micro-intervals and K is a prefixed number of activity windows. Significance is indicated by 'p-val', representing the p-values from the final QIF model. Cutpoint values are represented as VM/100.

	J = 60					
	K=2		K=3		K=4	
Parameters	Est	p-val	Est	p-val	Est	p-val
β_1	0.001	0.154	0.002	0.367	−0.079	0.644
β_2	−0.111	0.113	0.015	0.189	0.078	0.678
β_3	–	–	−0.006	0.009	0.003	0.095
β_4	–	–	–	–	−0.006	0.028
c_1	295	–	95	–	5	–
c_2	–	–	145	–	10	–
c_3	–	–	–	–	190	–
Sex (Male)	4.391	< 0.001	4.415	< 0.001	4.471	< 0.001
Age	0.052	0.791	0.099	0.614	0.112	0.566
Adult	2.133	< 0.001	2.105	< 0.001	2.094	< 0.001
BIC	37.75		53.64		59.00	
AIC	17.04		25.21		26.51	
GOF pval	0.53		0.13		0.23	

Let us interpret the results in Table 4.6 for the scenario of $K = 3, J = 60$ under model $y_{it} \sim \beta_1 A_{it1} + \beta_2 A_{it2} + \beta_{it3} A_3 + z_{it}^{\top} \alpha$, where AUC $A_{itk} = \int_{c_{k-1}}^{c_k} OTC_{it}(c)dc, k = 1, 2, 3$. Here Window 3 has estimated cutpoints $(c_2, c_3] = (145, 300]$ with $\hat{\beta}_1 = -0.006$ for predictor A_{it3}. For subject i, the area of the Window 3 rectangle $R_3 = (c_3 - c_2) \times (1 - 0) = 155$, and AUC Ratio of Window 3 is $\frac{A_{it3}}{155}$. Correspondingly, the parameter estimate $\hat{\beta}_3$ may be adjusted by $\hat{\beta}_{3Ratio} = 155\hat{\beta}_3$ for interpretability. In the case of Window 3, a lower AUC Ratio reflects *less* time spent within the activity cutpoints $(145, 300]$ than the hypothetical "most active individual" who spends all his or her time at or above the cutpoint range $(145, 300]$. Thus, for a subject who is 1% less active in the activity range of Window 3 compared to the hypothetical "most active individual", as reflected by a smaller AUC Ration, this subject's body fat as measured by SSST increases approximately approximately 0.93mm. See Figure B.1 in Appendix B for a schematic of this calculation.

4.6.2 Sex-Stratified Data Analysis

The proposed approach was also applied to the sex-stratified ELEMENT data set to determine if there were sex-specific PA intensity windows associated with the health outcome of Sub-Scapular Skin Thickness (SSST). In these sex-stratified analyses, we again considered the covariates of age and "adult" status based on Tanner staging, e.g., had completed Tanner stage 5. Tables 4.7 and 4.8 summarize the sex-stratified data analysis results for male and female subjects, respectively.

Male Adolescents: In the male subgroup, 18% of subjects had achieved "adult" status at T1 with mean(sd) age of 14.28(2.06) years. These variables increased to 49% classified as "adult" at T2, with mean(sd) age of 16.23(2.02) years. The longitudinal functional OTCs for the males in this study are shown in Figure 4.3, including the activity curves at both T1 and T2. Visually, we can see a slight pattern shift in activity profiles for male adolescents between T1 and T2. The OTC at T1 have slightly higher values after VM Count of 100 versus T2. This pattern shift demonstrates that the male subjects spend more time in higher levels of activity at T1 versus T2. This can also be visualized by the T2 curves decaying faster than the T1 curves, demonstrating that at T2 the male subjects spend a higher proportion of their time within the less-active activity ranges, and a lower proportion of their time at more-active activity ranges.

When considering the male-specific model, the results reflect that of the full-population, with a $K = 3$-group model demonstrating the strongest combination of statistical and scientific significance. Refer to Table 4.7. Here, the Window 3 is similar to that of the $K = 3$-group full-population model, with cutpoints (190-300]. The negative $\hat{\beta}_3$ value again suggests that

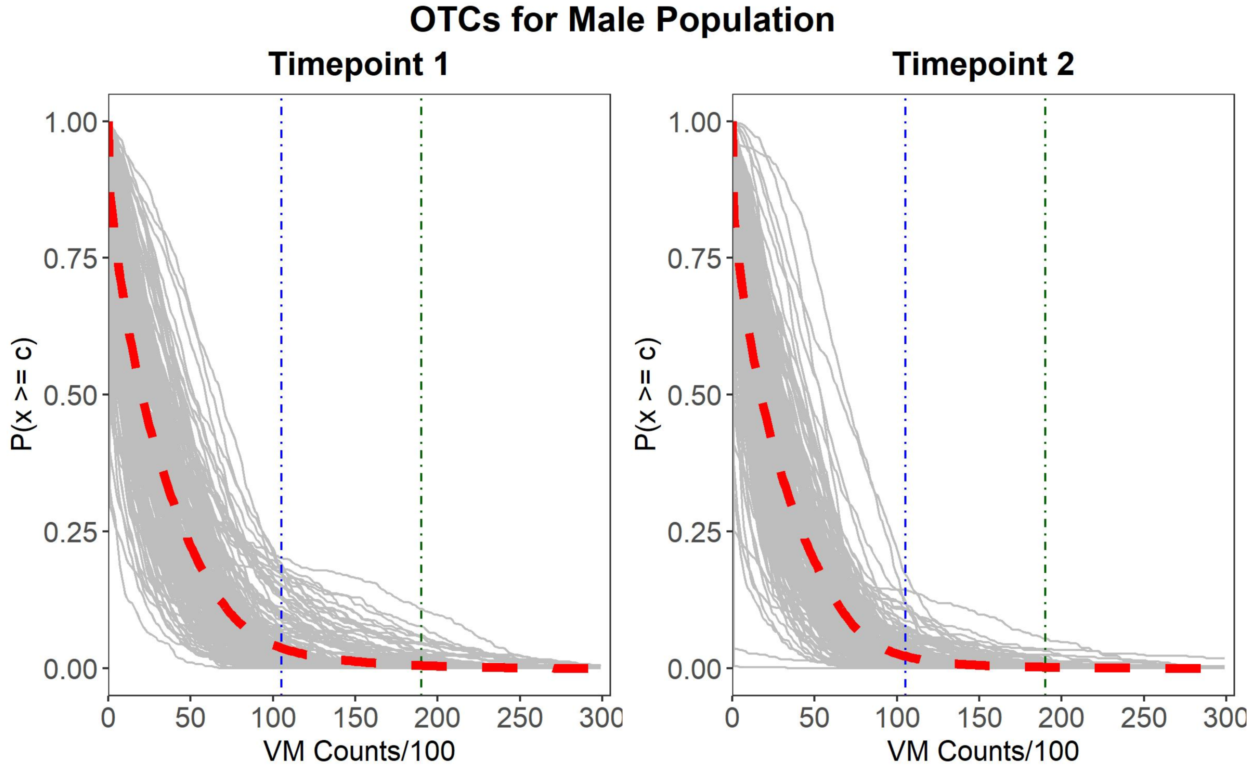


Figure 4.3: OTCs for 197 male ELEMENT subjects stratified by longitudinal time point of accelerometer data collection. The relative shape of OTC reflects the subject's activity profile, with the red-dotted line representing the mean proportion of time spent above each activity level across the individuals. The first vertical line, colored blue, represents the first detected cutpoint, indicating the changepoint between critical activity Windows 1 and 2. Likewise, the second vertical line in green represents the detected cutpoint 2, signalling the changepoint from critical activity Windows 2 and 3.

increased time spent in the third activity window is associated with decreased SSST. This finding demonstrates that more activity in this "moderate-to-vigorous" intensity level has a beneficial impact on body fat levels. Figure 4.3 illustrates the Window 3 changepoint; individuals with higher values on their OTC beyond this detected cutpoints spend a higher proportion of time in this critical activity window.

Table 4.7: Sex-Stratified Data Analysis Results for Male subjects obtained by the fusion-adapted L_0 QIF method, where J indicates the number of micro-intervals and K is a prefixed number of activity windows. Significance is indicated by 'p-val', representing the p-values from final QIF model. Cutpoint values are represented as VM/100.

	J = 60					
	K=2		K=3		K=4	
Parameters	Est	p-val	Est	p-val	Est	p-val
β_1	0.0009	0.453	0.003	0.368	0.006	0.104
β_2	−0.134	0.013	0.005	0.345	0.527	0.089
β_3	–	–	−0.013	0.006	1.207	0.055
β_4	–	–	–	–	−0.015	0.004
c_1	290	–	105	–	140	–
c_2	–	–	190	–	150	–
c_3	–	–	–	–	155	–
Age	−0.021	0.94	0.059	0.831	0.044	0.871
Adult	2.018	0.001	1.987	0.001	1.881	0.003
BIC	29.99		41.00		46.08	
AIC	13.58		21.31		23.10	
GOF pval	0.61		0.16		0.25	

Let us interpret the results in Table 4.7 in terms of AUC Ration. Consider the scenario of $K = 3, J = 60$ under model $y_{it} \sim \beta_1 A_{it1} + \beta_2 A_{it2} + \beta_{it3} A_3 + z_{it}^{\top}\alpha$, where AUC $A_{itk} = \int_{c_{k-1}}^{c_k} OTC_{it}(c)dc, k = 1, 2, 3$. Here Window 3 has estimated cutpoints $(c_2, c_3] = (190, 300]$ with $\hat{\beta}_3 = -0.013$ for predictor A_{it3} at both time points $t \in \{1, 2\}$. For subject i at time t, the area of the Window 3 rectangle $R_3 = (c_3 - c_2) \times (1 - 0) = 110$, and the AUC Ratio of Window 3 is $\frac{A_{it3}}{110}$. Correspondingly, the parameter estimate $\hat{\beta}_3$ may be adjusted by $\hat{\beta}_{3Ratio} = 110\hat{\beta}_3$ for interpretability. In the case of Window 3, a lower AUC Ratio reflects *less* time spent within the activity cutpoints $(190, 300]$ than the hypothetical "most active individual" who spends all his or her time within or above the cutpoint range (190, 300]. Thus, for a subject who is 1% less active in the activity range of Window 3 compared to the hypothetical "most active individual", as reflected by a smaller AUC Ration, this subject's body fat as measured by SSST increases approximately approximately 1.43mm. See Figure B.1 in Appendix B for a schematic of a similar calculation.

Female Adolescents: In the female subgroup, the mean(sd) age at T1 was 14.41(2.13) years, with 19% of the 232 female subjects having completed puberty, e.g. achieved Tanner stage 5. At the second time point, the mean(sd) age increased to 16.36(2.14) years, with 47% of the adolescent females reaching adult status. The repeated functional OTCs for the females in this study are shown in Figure 4.4, including the activity curves at both T1 and T2.

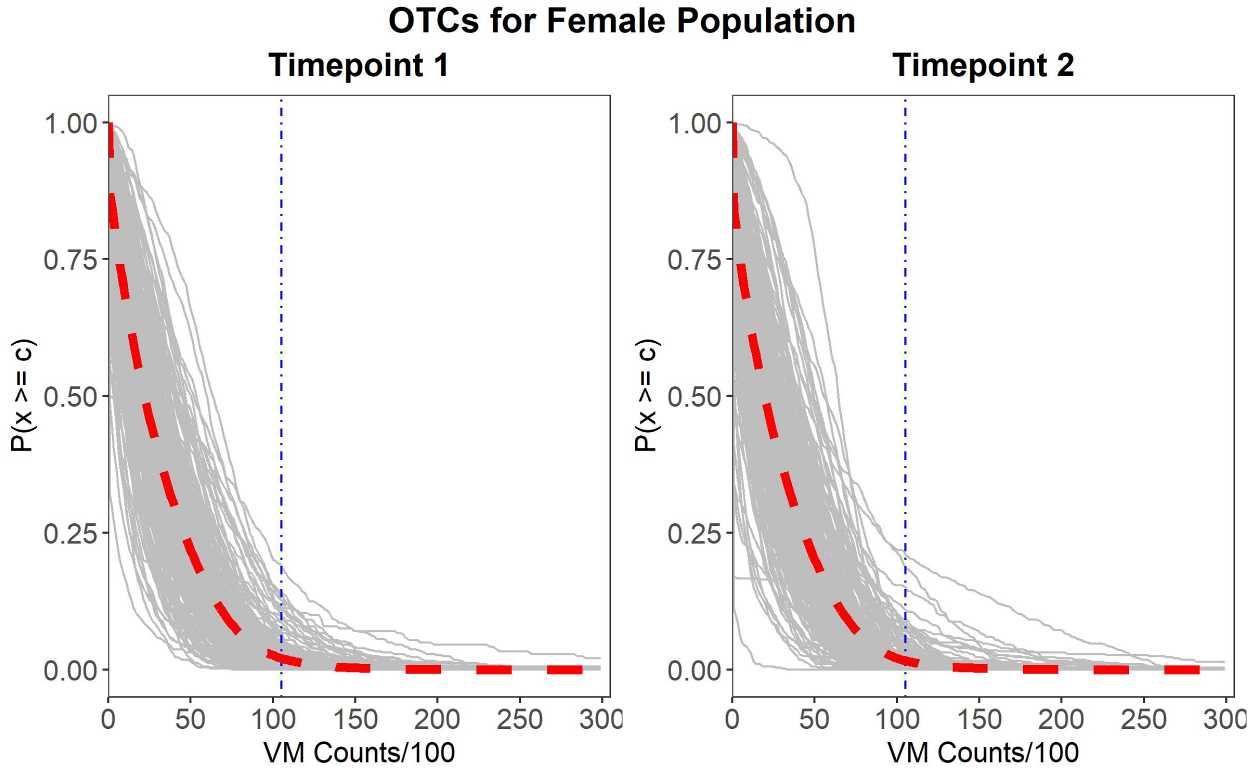


Figure 4.4: OTCs for 232 female ELEMENT subjects stratified by longitudinal time point of accelerometer data collection. The relative shape of OTC reflects the subject's activity profile, with the red-dotted line representing the mean proportion of time spent above each activity level across the individuals. The vertical blue, represents the first detected cutpoint, indicating the changepoint between critical activity Windows 1 and 2.

The data analysis for the female-specific model demonstrated less significant results than either the male-specific or full population models. Here, the $K = 2$-group model achieves the best goodness of fit, as evidenced by BIC and AIC measures, though neither of the PA windows present a statistically significant association with SSST. Furthermore, the small coefficient sizes are not indicative of clinical significance. As the female-specific results did not demonstrate statistical or clinical significance in the longitudinal association between functional PA and SSST, the AUC Ration interpretation is not included here. However, the calculations would be similar to those detailed in Sections 3.7, 4.1.2, as well as Figure B.1.

Table 4.8: Sex-Stratified Data Analysis Results for Girls obtained by the fusion-adapted L_0 QIF method, where J indicates the number of micro-intervals and K is a prefixed number of activity windows. Significance is indicated by 'p-val', representing the p-values from final QIF model. Cutpoint values are represented as VM/100.

	J = 60					
	K=2		K=3		K=4	
Parameters	Est	p-val	Est	p-val	Est	p-val
β_1	< 0.001	0.112	−0.037	0.554	0.003	0.072
β_2	< 0.001	0.554	0.241	0.109	−1.921	0.001
β_3	–	–	0.001	0.383	2.960	< 0.001
β_4	–	–	–	–	−0.216	0.002
c_1	105	–	30	–	265	–
c_2	–	–	35	–	270	–
c_3	–	–	–	–	275	–
Age	0.006	0.808	0.027	0.916	0.044	0.871
Adult	2.183	0.0002	2.209	0.002	2.209	0.0002
BIC	34.28		39.97		45.59	
AIC	17.05		19,29		21.46	
GOF pval	0.22		0.29		0.38	

4.7 Discussion

In this chapter, I extend the L_0 fusion-adapted framework from Chapter 3 to a longitudinal functional framework with repeated measures captured from wearable devices. To achieve this statistical methodological extension, I employ the Quadratic Inference Function (QIF) framework, which considers a population-average effects model, and develop a regularized QIF via mixed integer optimization (MIO). This methodology detects changepoints in serially measured functional accelerometer data to define critical windows of activity intensities that impact longitudinal health outcomes, while also accounting for covariates of interest. Such an informatics toolbox can be applied to analyze the relationship of functional digital features with continuous outcomes. In the data analysis application, I employ the developed framework to detect critical PA windows and assess their population-average effects on Sub-Scapular Skin Thickness (SSST) on adolescents from the Early Life Exposure in Mexico to ENvironmental Toxicants (ELEMENT) cohort. I find that both the full dataset, as well as the male-stratified dataset, have a negative association between increased levels of high-intensity activity and SSST.

Through extensive simulation experiments, this chapter numerically demonstrates the high stability and accuracy of the longitudinal MIO technique via QIF. As in the single-time point framework of Chapter 3, we again find that the strength of the results is not overly sensitive to the choice of J, the starting number of correlated intervals. Simulation results for $J = 60$ and $J = 300$ were very similar, though the computation time for a larger number of starting intervals does significantly increase. Investigators can choose the number of J intervals based on factors of sample size and data availability without concern that the tuning choice of J will significantly affect the analysis. In practice, larger choices of J do not always provide optimal solutions, as indicated by the MIO software declaring a "non-optimal" solution path, as in the case of the $J = 300$ scenarios from our data analysis. In these instances, which are clearly communicated by software results, researchers should decrease their choice of J.

While this chapter focused on a data example with two repeated functional measures and outcomes, the longitudinal L_0 fusion-adapted MIO framework can be easily extended to more than two time points, with trivial adjustments in the implementation of the QIF-based objective function by the MIO optimization solver. The methodology can also be extended to allow for time-varying association parameter estimates. That is, this extension would allow for the detection and estimation of different critical activity windows at the repeated time points. Such an analysis would enable researchers to investigate if the effect of the functional covariate of interest changed over time when the functional predictors vary strongly over time. For example, perhaps the effect of PA on mental health is more important as subjects age. To achieve this model, one would need to update the objective function in a less trivial manner than when increasing the number of repeated measurements. To accommodate the time-varying effects, a larger η parameter vector is necessary, with associated updates to the design matrix X. In the case of two repeated measures with time-varying effects on both the functional variable discretized into J intervals as well as q number of covariates, the η parameter vector would increase from size $J + q$ to $2(J + q)$. This change would necessitate additional adjustment to the QIF objective function Q_n. Lastly, future work could extend the proposed methodology to non-linear outcomes, such a binomial outcomes via logistic regression or count data via Poisson modelling. As the QIF framework is capable of fitting such binomial and Poisson models, among others, this extension would involve changes to the Q_n objective function via the mean function μ_{it}.

Notably, the proposed methodology also presents the opportunity to extend into Federated Learning framework, in which summary statistics from different data sources (such as hospitals or research groups) are combined for analysis, while still maintaining data privacy. As pointed out in Section 4.3.3, this is possible as the QIF framework does not require subject-level detail in the optimization to conduct QIF estimation. Rather, one can use summary statistics from each time-point (or each correlated cluster) to conduct this analysis. As demonstrated in Equation (4.14), if we assume the simplest QIF case where $\bar{C}(\boldsymbol{\eta}^{(0)}) = I$, then the Q_n objective function relies on sample-level summary statistics of $\boldsymbol{XY}$ and $\boldsymbol{X}^\top \boldsymbol{X}$, rather than individual level data. This simple-case could be used to detect critical activity windows and MIO-derived-estimates of the parameter vector $\boldsymbol{\eta}$. Though our data analysis and simulation experiments implemented this simple-case for the initial $\boldsymbol{\eta}$ estimates that were ultimately updated using an empirical estimate $\bar{C}(\boldsymbol{\eta})$, the simulation results did demonstrate that the initial MIO-based parameter estimates could have relatively satisfactory results with low bias and high efficiency (refer to Tables 4.1,4.2). In this case, there would be a trade-off between efficiency and privacy, which could be determined by the specific research and security needs.

CHAPTER 5

Summary and Future Work

In this dissertation, I have presented novel supervised learning frameworks using both L_1 and L_0 regularization methods to conduct functional data analysis (FDA). This work was motivated by high-frequency time-series data collected from wearable devices, specifically from accelerometer devices that provide objective measures of physical activity (PA). Overall, the proposed analytics conduct supervised learning via scalar-on-function regression models that involve Occupation Time Curves (OTCs) as the functional predictors and assess their association with specific scalar health outcomes of interest. These proposed approaches free the dependence on subjective choices of pre-determined PA categorizations in analysis of high-frequency time-series data from accelerometers, and instead allow the data to adaptively detect changepoints to define critical windows of activity intensities.

5.1 Summary

In Chapter 2, I present Occupation Time Curves (OTCs), which describe the percentage of time spent at or above a continuum of activity count levels. The resulting functional curve is informative to capture time-course individual variability of PA. I introduce the multi-step adaptive learning algorithm, termed FRACT (Functional Regularized Adaptive Changepoint-detection Technique), to perform supervised learning via a scalar-on-function regression model that involves OTC as the functional predictor with the ability to include other scalar covariates of interest. FRACT focuses on L_1 regularization approaches to determine activity windows of interest, incorporating a hybrid approach of fused lasso for clustering, Hidden Markov Model (HMM) for change-point detection, and refinement procedures and goodness-of-fit measures for final model selection. I show that FRACT has flexibility and reliability in identifying changepoints/critical windows and can effectively transform functional accelerometer data collected from wearable devices into knowledge on PA's effect on human biological aging. In addition, I demonstrated that different sets of changepoints and associated parameters are detected for different epigenetic aging outcomes, highlighting FRACT's adaptive ability across different settings. This generalizable role in translating accelerometer data into scientific knowledge is important to researchers and practitioners alike.

In Chapter 3, I move from L_1 regularization approaches to those in the L_0 framework. In this capacity, I investigate functional associations between health outcomes and PA under an L_0 regularization approach by formulating and implementing modern optimization methods to functional analysis by means of Mixed Integer Optimization (MIO). In this chapter, I show via numerical experiments and theoretical arguments that the proposed method is able to detect critical activity intensity windows of interest and provide parameter estimation in a one-step process, as opposed to the multi-step learning algorithm, FRACT, introduced in Chapter 2. To the best of my knowledge, I am the first to consider MIO methodologies both to conduct fusion as well as in a functional data framework.

In Chapter 4, I extend the fusion-adapted MIO methodology to a longitudinal framework. Here, I consider repeated wearable data to understand the influence of serially measured functional accelerometer data on longitudinal health outcomes again leveraging L_0 regularization techniques. Through a combination of numerical experiments and theoretical arguments, I demonstrate that the proposed method, which leverages Quadratic Inference Functions (QIF) to consider a population-average effects model via MIO, produces consistent results in a time-efficient manner. To the best of my knowledge I am the first to conduct longitudinal fusion in this manner. To showcase the utility of the method in a real-world environment, I focus on a longitudinal study of PA patterns from late-adolescence into early adulthood on Sub-Scapular Skin Thickness (SSST), a measure of truncal fat distribution that is among the measures of body composition that can be influenced by PA behaviors. This data analysis determined a moderate-to-high intensity activity window that was associated with lower SSST measurements in both the full population, and in sex-specific manner. Notably, the full population and sex-stratified analyses identified unique activity windows and parameters of association, demonstrating the methodology's adaptive ability to detect unique critical windows across different populations.

Importantly, the flexibility of the methodologies proposed in this dissertation demonstrate their value to the analysis of future wearable devices. In the ever-evolving world of wearable devices, there are constantly new devices or sensors available. The applications of the proposed methodologies are not restricted to accelerometer sensors; rather they can easily be applied to other devices including biomedical/smart health devices and environmental toxicant sensors, among others. In such a role, the methodologies proposed in this dissertation can help translate data collected from existent/future sensors into decision-making knowledge. As such physiological and environmental sensors can have a great potential impact on the future of health-monitoring and intervention, the translational role these proposed analytics play in turning high-frequency time-series data into decision-making knowledge is invaluable.

5.2 Future Work

The supervised learning fusion-adapted FDA frameworks I developed provide useful toolboxes for analyzing wearable device data. As I continue in this field of research, I am interested in pursuing this research in multiple different directions. These future projects are described below.

Future Project I: In this dissertation, I focused on OTCs that summarized the adolescent's PA profiles on weekends afternoons. The rationale behind this choice was that during this time, adolescents have greater control over their activities as they are less constrained by school or home responsibilities. However, it is important to consider how the choice of time period for constructing OTCs can impact the analysis. For instance, some adolescents may be highly active in school sports on weekday afternoons and then use the weekends to rest. In such cases, the weekend-based OTCs might only reflect their lower levels of activity during the weekends and fail to capture their high activity levels during the week.

To address this limitation in future research, I intend to explore alternative approaches. One option is to analyze OTCs summarizing the subject's PA over an entire week, thereby capturing their PA data for the entire week. However, this approach may be more sensitive to periods of missing data. Another alternative could involve identifying each subject's most active window of time and using that as the unit of analysis. Determining the most active window would require careful consideration, and potential measures could include the 4-hour window with the highest cumulative count level or the windows with the highest intensity count values. By considering these alternative strategies, I aim to refine the analysis and ensure a more comprehensive understanding of adolescents' PA patterns in different contexts. This will contribute to a more robust interpretation of the data and better inform future interventions and policies aimed at promoting healthier lifestyles among adolescents.

Future Project II: I am interested in exploring whether the timing of PA, not only the intensity, is associated with health outcomes. To achieve this, I plan to extend the single-timepoint fusion-adapted L_0 approach of Chapter 3 to handle multiple functional covariates. Through application of this extended MIO technique focusing on PA during different time periods of the day, such as morning versus evening, one can investigate if the activity intensity changepoints are dependent on time of day. This is distinct from the longitudinal framework introduced in Chapter 4 in that it incorporates multiple functional covariates with complex correlation structures, and not correlated repeated outcomes.

Future Project III: In the future, I plan to extend the fusion-adapted longitudinal model introduced in Chapter 4 to allow for time-varying association parameter estimates. This extension would allow for the detection and estimation of different critical activity windows at the repeated time points when functional predictors vary significantly over time. Such an analysis would enable researchers to investigate whether the effect of PA on a scalar outcome changes over time. This assessment would be particularly interesting in our motivating ELEMENT cohort that follows the subjects from adolescence to early adulthood. In our data analysis example, it is feasible that the association between PA and body composition, as measured by SSST, changes as the subjects transition out of childhood into adulthood. These associations could then be further studied as the subjects continue to age, with research considering the longitudinal effect from childhood to mature adulthood and beyond.

Future Project IV: The proposed longitudinal functional data model from Chapter 4 presents the opportunity to extend into the Federated Learning framework, in which summary statistics from different data sources (such as hospitals or research groups) are combined for analysis. Rather than incorporating multiple functional predictors and associated scalar outcomes from different time points, one could incorporate these repeated measurements as coming from various correlated sources, such as different hospitals in similar geographic area. As the proposed QIF framework does not require subject-level detail to conduct consistent estimation, these datasets could be combined while still maintaining data privacy by leveraging summary statistics from each correlated cluster. I plan on pursuing this extension in the near future.

Future Project V: The methods proposed in these dissertation chapters give rise to data analytic toolboxes enabling the exploration of various questions of interest related to the effect of functional PA features on health outcomes. While this dissertation focuses on linear relationships between functional OTCs and scalar outcomes, the proposed methodologies could be extended to non-normal and non-linear models, such as logistic regression with binary outcomes, and Cox regressions with time-to-event outcomes. For example, with such extensions, one could investigate questions such as what critical PA intensities impact time to Cardiovascular Events in patients with diagnosed cardiovascular complications.

Software Development: The dissemination of statistical methodologies through software development plays a pivotal role in advancing research, decision-making, and problem-solving. Software packages that implement statistical methods offer a user-friendly and efficient way to analyze data, making complex statistical techniques accessible to a broader audience. Furthermore, software development facilitates the reproducibility and transparency of statistical analyses as well as enhancing the methodology's efficiency and scalability. As such, I plan on disseminating my novel supervised learning frameworks not only via publication, but also by the development and publication of related software. For the L_1 regularization method presented in Chapter 2, I will share my code and relevant examples on my github page. For the L_0 regularized learning frameworks presented in Chapters 3 and 4, I plan on developing both R and Python packages for other researchers to easily access.

APPENDIX A

Chapter 3 Appendix

Table A.1: Simulation Results of 4-group Model with number of original intervals $J = 300$ to evaluate the adapted L_0-fusion analytic under a environment with severe multi-collinearity. Results are summarized over 500 replicates and include average estimate (L_0 Mean), empirical standard error (L_0 ESE), and Mean from a standard Fused Lasso analysis (FL Mean) for a sample size of $N = 500$. Cutpoint values are represented as VM/100 for ease of visualization.

	Scenario D				Scenario E				Scenario F			
	Truth	L_0 Mean	L_0 ESE	FL Mean	Truth	L_0 Mean	L_0 ESE	FL Mean	Truth	L_0 Mean	L_0 ESE	FL Mean
β_1	2	2.00	0.04	1.52	4	4.00	0.04	3.02	8	7.99	0.04	6.01
β_2	1	0.98	0.11	0.30	2	2.01	0.12	0.57	4	3.99	0.10	0.99
β_3	0	−0.01	0.12	−0.34	0	−0.01	0.10	−0.65	0	−0.01	0.10	−1.35
β_4	−1	−1.00	0.01	−0.98	−2	−2.00	0.01	−1.95	−4	−4.00	0.01	−3.91
c_1	40	39.82	3.61	77.17	40	40.05	1.59	77.78	40	40.01	0.75	78.39
c_2	80	80.39	4.52	97.94	80	80.17	1.90	98.43	80	80.02	0.91	99.28
c_3	120	120.31	2.79	117.68	120	120.11	1.24	117.28	120	120.02	0.63	117.22
α	1	1.00	0.43	1.08	1.00	0.98	0.43	0.87	1.00	0.97	0.43	0.45

Table A.2: Simulation Results of 4-group Model with number of intervals $J = 60$ to evaluate the adapted L_0-fusion analytic, summarized over 500 replicates. Results include average estimate (L_0 Mean), empirical standard error (L_0 ESE), and Mean from a standard Fused Lasso analysis (FL Mean) as comparison. Cutpoint values are represented as VM/100 for ease of visualization. Sensitivity for selecting 3-group model based on goodness-of-fit comparisons was greater than 99% in all scenarios.

		N = 500			N=250			N=100		
	Truth	L_0 Mean	L_0 ESE	FL Mean	L_0 Mean	L_0 ESE	FLMean	L_0 Mean	L_0 ESE	FL Mean
Scenario D										
β_1	2	2.00	0.04	1.51	2.01	0.07	1.52	2.04	0.22	1.55
β_2	1	1.01	0.13	0.23	1.02	0.25	0.30	1.06	0.68	0.33
β_3	0	−0.01	0.13	−0.39	−0.03	0.23	0.37	−0.21	0.69	−0.41
β_4	−1	−1.00	0.01	−0.97	−1.00	0.01	−0.97	−1.01	0.07	−0.98
c_1	40	39.74	3.63	78.12	39.21	6.21	76.93	37.69	11.91	74.05
c_2	80	80.13	4.54	100.61	80.43	8.19	99.06	83.01	16.86	100.28
c_3	120	120.29	3.01	118.01	121.71	10.54	118.51	135.33	39.14	121.94
age	1	0.98	0.44	1.08	0.99	0.67	1.05	1.01	1.06	1.36
Scenario E										
β_1	4	3.99	0.04	2.99	4.00	0.06	3.03	4.01	0.10	3.06
β_2	2	1.99	0.11	0.43	2.01	0.17	0.47	2.02	0.32	0.63
β_3	0	−0.01	0.08	−0.79	−0.01	0.17	−0.72	−0.06	0.35	−0.77
β_4	−2	−2.00	0.01	−1.95	−2.00	0.01	−1.94	−2.00	0.02	−1.95
c_1	40	40.05	1.40	78.90	39.88	2.45	77.93	39.56	4.60	75.83
c_2	80	80.25	1.84	101.80	80.16	2.98	100.37	80.42	6.18	99.31
c_3	120	120.08	0.95	117.50	120.05	2.01	118.01	121.73	11.96	119.91
age	1	0.98	0.44	0.86	1.01	0.67	0.91	1.02	1.06	1.09
Scenario F										
β_1	8	7.99	0.03	5.99	7.99	0.04	6.03	8.00	0.09	6.14
β_2	4	4.00	0.04	0.75	4.00	0.09	0.91	4.01	0.27	1.31
β_3	0	−0.00	0.03	−1.67	−0.004	0.08	−1.52	−0.03	0.27	−1.51
β_4	−4	−4.00	0.00	−3.90	−4.00	0.01	−3.89	−4.00	0.01	−3.89
c_1	40	40.00	0.00	78.76	40.04	0.55	78.34	39.87	1.96	75.63
c_2	80	80.00	0.00	102.57	80.03	0.74	100.61	80.14	2.51	98.51
c_3	120	120.00	0.00	117.73	120.01	0.39	117.80	120.22	1.57	119.73
age	1	0.98	0.44	0.45	1.01	0.66	0.59	1.01	1.06	0.52

APPENDIX B

Chapter 4 Appendix

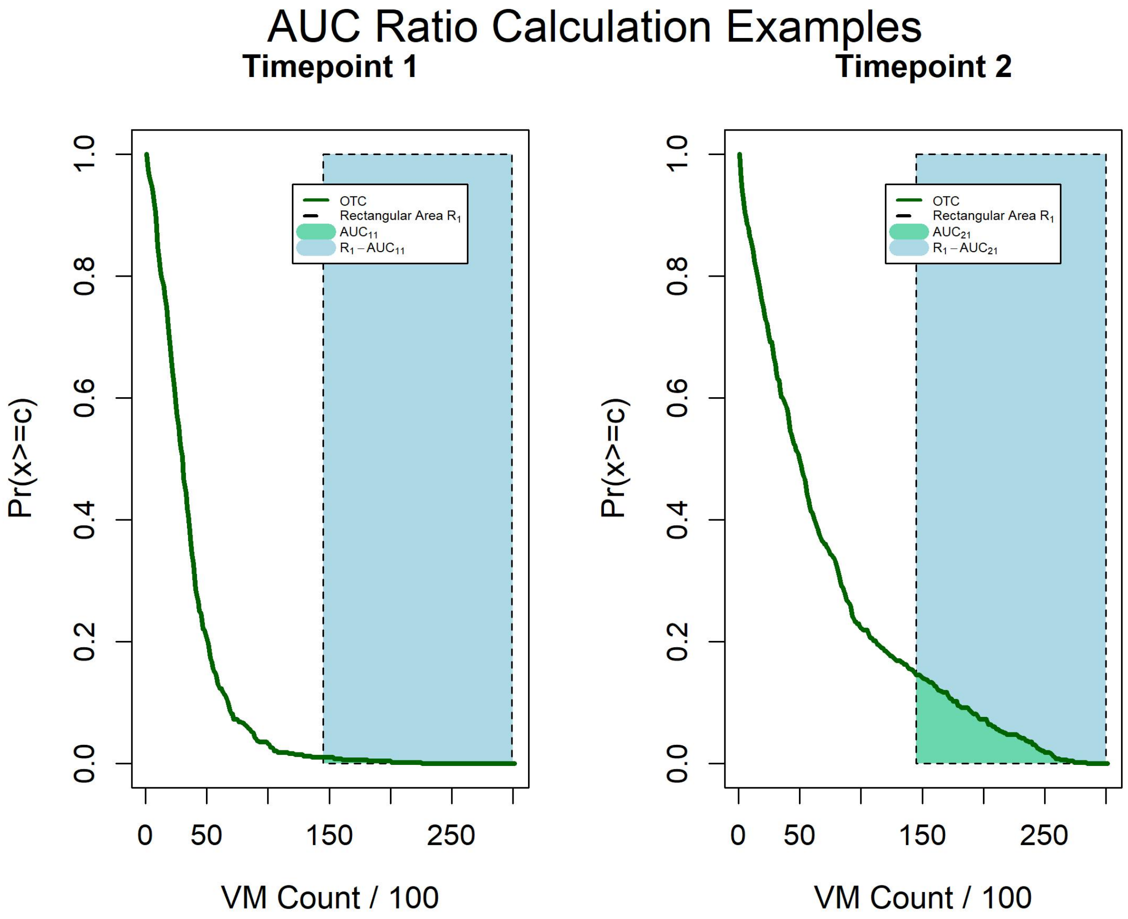


Figure B.1: An illustration of the AUC Ratio calculation for Window 3. The shaded green regions represent the AUC for subject i at time t for $t = c(1, 2)$ within the detected window, or A_{i3}. The outlined rectangle represents the area of the full rectangle created by the cutpoints ($c_2 = 145, c_3 = 300$) of Window 3, deemed R_3, which represents the hypothetical subject spending 100% of time at or above this activity level. The AUC Ratio for this first window is then calculated by AUC Ratio$_{i3} = \frac{A_{i3}}{R_3}$. The interpretation of this ratio depends on the location of the window. For the K^{th} window, the interpretation of the AUC Ratio$_{iK}$ represents the percent of time the individual spends within that window compared to the hypothetically most active person. In this case, a higher AUC Ratio$_{i3}$ value represents higher PA within the window. We can see that the AUC Ratio$_{i3}$ at T1 is lower than the AUC Ratio$_{i3}$at T2, suggesting a higher outcome at T2 versus T1. In reality, we see an SSST outcome of 12.0mm at T1, increasing to 12.5mm at T2.

BIBLIOGRAPHY

Introduction. In J. O. Ramsay and B. W. Silverman, editors, *Functional Data Analysis*, Springer Series in Statistics, pages 1–18. Springer, New York, NY, 2005. ISBN 978-0-387-22751-1. doi: 10.1007/0-387-22751-2_1. URL https://doi.org/10.1007/0-387-22751-2_1.

Introduction to Functional Nonparametric Statistics. In Frédéric Ferraty and Philippe Vieu, editors, *Nonparametric Functional Data Analysis: Theory and Practice*, Springer Series in Statistics, pages 5–10. Springer, New York, NY, 2006. ISBN 978-0-387-36620-3. doi: 10.1007/0-387-36620-2_1. URL https://doi.org/10.1007/0-387-36620-2_1.

Samrachana Adhikari, Fabrizio Lecci, James T. Becker, Brian W. Junker, Lewis H. Kuller, Oscar L. Lopez, and Ryan J. Tibshirani. High-dimensional longitudinal classification with the multinomial fused lasso. *Statistics in Medicine*, 38(12):2184–2205, 2019. ISSN 1097-0258. doi: 10.1002/sim.8100. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8100. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.8100.

Abeer A. Aljahdali, Ana Baylin, Edward A. Ruiz-Narvaez, Hyungjin Myra Kim, Alejandra Cantoral, Martha M. Tellez-Rojo, Margaret Banker, and Karen E. Peterson. Sedentary patterns and cardiometabolic risk factors in Mexican children and adolescents: analysis of longitudinal data. *International Journal of Behavioral Nutrition and Physical Activity*, 19(1):143, December 2022. ISSN 1479-5868. doi: 10.1186/s12966-022-01375-0. URL https://doi.org/10.1186/s12966-022-01375-0.

Dimitris Bertsimas and Romy Shioda. Algorithm for cardinality-constrained quadratic optimization. *Computational Optimization and Applications*, 43(1):1–22, May 2009. ISSN 1573-2894. doi: 10.1007/s10589-007-9126-9. URL https://doi.org/10.1007/s10589-007-9126-9.

Dimitris Bertsimas, Angela King, and Rahul Mazumder. Best subset selection via a modern optimization lens. *The Annals of Statistics*, 44(2):813–852, April 2016. ISSN 0090-5364, 2168-8966. doi: 10.1214/15-AOS1388. URL https://projecteuclid.org/journals/annals-of-statistics/volume-44/issue-2/Best-subset-selection-via-a-modern-optimization-lens/10.1214/15-AOS1388.full. Publisher: Institute of Mathematical Statistics.

Dimitris Bertsimas, Jean Pauphilet, and Bart Van Parys. Sparse Regression: Scalable algorithms and empirical performance. *Statistical Science*, 35(4), November 2020.

ISSN 0883-4237. doi: 10.1214/19-STS701. URL http://arxiv.org/abs/1902.06547. arXiv:1902.06547 [stat].

Leonid Bogachev and Nikita Ratanov. Occupation time distributions for the telegraph process. *Stochastic Processes and their Applications*, 121(8):1816–1844, August 2011. ISSN 0304-4149. doi: 10.1016/j.spa.2011.03.016. URL https://www.sciencedirect.com/science/article/pii/S0304414911000755.

Emmanuel J. Candès and Yaniv Plan. Near-ideal model selection by 1 minimization. *The Annals of Statistics*, 37(5A):2145–2177, October 2009. ISSN 0090-5364, 2168-8966. doi: 10.1214/08-AOS653. URL https://projecteuclid.org/journals/annals-of-statistics/volume-37/issue-5A/Near-ideal-model-selection-by-%e2%84%931-minimization/10.1214/08-AOS653.full. Publisher: Institute of Mathematical Statistics.

Emmanuel J. Candès, Michael B. Wakin, and Stephen P. Boyd. Enhancing Sparsity by Reweighted 1 Minimization. *Journal of Fourier Analysis and Applications*, 14(5):877–905, December 2008. ISSN 1531-5851. doi: 10.1007/s00041-008-9045-x. URL https://doi.org/10.1007/s00041-008-9045-x.

J. L. Chandler, K. Brazendale, M. W. Beets, and B. A. Mealing. Classification of physical activity intensities using a wrist-worn accelerometer in 8–12-year-old children. *Pediatric Obesity*, 11(2):120–127, 2016. ISSN 2047-6310. doi: 10.1111/ijpo.12033. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ijpo.12033. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ijpo.12033.

Hsin-wen Chang and Ian W. McKeague. Empirical likelihood-based inference for functional means with application to wearable device data. *Journal of the Royal Statistical Society Series B*, 84(5):1947–1968, 2022. URL https://ideas.repec.org//a/bla/jorssb/v84y2022i5p1947-1968.html. Publisher: Royal Statistical Society.

Kehui Chen and Hans-Georg Müller. Modeling Repeated Functional Observations. *Journal of the American Statistical Association*, 107(500):1599–1609, December 2012. ISSN 0162-1459. doi: 10.1080/01621459.2012.734196. URL https://doi.org/10.1080/01621459.2012.734196. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2012.734196.

Kong Y. Chen and David R. Bassett. The technology of accelerometry-based activity monitors: current and future. *Medicine and Science in Sports and Exercise*, 37(11 Suppl): S490–500, November 2005. ISSN 0195-9131. doi: 10.1249/01.mss.0000185571.49104.82.

Scott E. Crouter, Jennifer I. Flynn, and David R. Bassett. Estimating Physical Activity in Youth Using a Wrist Accelerometer. *Medicine and science in sports and exercise*, 47 (5):944–951, May 2015. ISSN 0195-9131. doi: 10.1249/MSS.0000000000000502. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4362848/.

Patty Freedson, David Pober, and Kathleen F. Janz. Calibration of accelerometer output for children. *Medicine and Science in Sports and Exercise*, 37(11 Suppl):S523–530, November 2005. ISSN 0195-9131. doi: 10.1249/01.mss.0000185658.28284.ba.

Zan Gao, Wenxi Liu, Daniel J. McDonough, Nan Zeng, and Jung Eun Lee. The Dilemma of Analyzing Physical Activity and Sedentary Behavior with Wrist Accelerometer Data: Challenges and Opportunities. *Journal of Clinical Medicine*, 10(24):5951, January 2021. ISSN 2077-0383. doi: 10.3390/jcm10245951. URL https://www.mdpi.com/2077-0383/10/24/5951. Number: 24 Publisher: Multidisciplinary Digital Publishing Institute.

Antonio Garriga, Nuria Sempere-Rubio, María José Molina-Prados, and Raquel Faubel. Impact of Seasonality on Physical Activity: A Systematic Review. *International Journal of Environmental Research and Public Health*, 19(1):2, December 2021. ISSN 1661-7827. doi: 10.3390/ijerph19010002. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8751121/.

Jeff Goldsmith, Ciprian M. Crainiceanu, Brian Caffo, and Daniel Reich. Longitudinal Penalized Functional Regression for Cognitive Outcomes on Neuronal Tract Measurements. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, 61(3):453–469, May 2012. ISSN 0035-9254. doi: 10.1111/j.1467-9876.2011.01031.x.

Jeff Goldsmith, Vadim Zipunnikov, and Jennifer Schrack. Generalized multilevel function-on-scalar regression and principal component analysis. *Biometrics*, 71(2):344–353, June 2015. ISSN 1541-0420. doi: 10.1111/biom.12278.

Pedro C. Hallal, Cesar G. Victora, Mario R. Azevedo, and Jonathan C. K. Wells. Adolescent Physical Activity and Health. *Sports Medicine*, 36(12):1019–1030, December 2006. ISSN 1179-2035. doi: 10.2165/00007256-200636120-00003. URL https://doi.org/10.2165/00007256-200636120-00003.

Peisong Han and Peter X. K. Song. A note on improving quadratic inference functions using a linear shrinkage approach. *Statistics & Probability Letters*, 81(3):438–445, March 2011. ISSN 0167-7152. doi: 10.1016/j.spl.2010.12.010. URL https://www.sciencedirect.com/science/article/pii/S0167715210003500.

Felix Heinzl and Gerhard Tutz. Clustering in linear-mixed models with a group fused lasso penalty. *Biometrical Journal*, 56(1):44–68, 2014. ISSN 1521-4036. doi: 10.1002/bimj.201200111. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.201200111. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bimj.201200111.

Steve Horvath. DNA methylation age of human tissues and cell types. *Genome Biology*, 14(10):3156, December 2013. ISSN 1474-760X. doi: 10.1186/gb-2013-14-10-r115. URL https://doi.org/10.1186/gb-2013-14-10-r115.

Steve Horvath, Junko Oshima, George M. Martin, Ake T. Lu, Austin Quach, Howard Cohen, Sarah Felton, Mieko Matsuyama, Donna Lowe, Sylwia Kabacik, James G. Wilson, Alex P. Reiner, Anna Maierhofer, Julia Flunkert, Abraham Aviv, Lifang Hou, Andrea A. Baccarelli, Yun Li, James D. Stewart, Eric A. Whitsel, Luigi Ferrucci, Shigemi Matsuyama, and Kenneth Raj. Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging (Albany NY)*, 10(7):1758–1775, July 2018. ISSN 1945-4589. doi: 10.18632/aging.101508. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6075434/.

Lajos Horváth and Piotr Kokoszka. Functional data structures. In Lajos Horváth and Piotr Kokoszka, editors, *Inference for Functional Data with Applications*, Springer Series in Statistics, pages 1–17. Springer, New York, NY, 2012. ISBN 978-1-4614-3655-3. doi: 10.1007/978-1-4614-3655-3_1. URL https://doi.org/10.1007/978-1-4614-3655-3_1.

Youna Hu and Peter X.-K. Song. Sample size determination for quadratic inference functions in longitudinal design with dichotomous outcomes. *Statistics in Medicine*, 31(8):787–800, 2012. ISSN 1097-0258. doi: 10.1002/sim.4458. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4458. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.4458.

Rae-Chi Huang, Karen A Lillycrop, Lawrence J Beilin, Keith M Godfrey, Denise Anderson, Trevor A Mori, Sebastian Rauschert, Jeffrey M Craig, Wendy H Oddy, Oyekoya T Ayonrinde, Craig E Pennell, Joanna D Holbrook, and Phillip E Melton. Epigenetic Age Acceleration in Adolescence Associates With BMI, Inflammation, and Risk Score for Middle Age Cardiovascular Disease. *The Journal of Clinical Endocrinology & Metabolism*, 104(7):3012–3024, July 2019. ISSN 0021-972X. doi: 10.1210/jc.2018-02076. URL https://doi.org/10.1210/jc.2018-02076.

John M. Jakicic, William E. Kraus, Kenneth E. Powell, Wayne W. Campbell, Kathleen F. Janz, Richard P. Troiano, Kyle Sprow, Andrea Torres, Katrina L. Piercy, and 2018 Physical Activity Guidelines Advisory Committee. Association between Bout Duration of Physical Activity and Health: Systematic Review. *Medicine and science in sports and exercise*, 51(6):1213, June 2019. doi: 10.1249/MSS.0000000000001933. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6527142/. Publisher: NIH Public Access.

Ian Janssen, Julie E. Campbell, Samah Zahran, Travis J. Saunders, Jennifer R. Tomasone, and Jean-Philippe Chaput. Timing of physical activity within the 24-hour day and its influence on health: a systematic review. *Health Promotion and Chronic Disease Prevention in Canada : Research, Policy and Practice*, 42(4):129–138, April 2022. ISSN 2368-738X. doi: 10.24095/hpcdp.42.4.02. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9116725/.

Anna Kankaanpää, Asko Tolvanen, Aino Heikkinen, Jaakko Kaprio, Miina Ollikainen, and Elina Sillanpää. The role of adolescent lifestyle habits in biological aging: A prospective twin study. *eLife*, 11:e80729, November 2022. ISSN 2050-084X. doi: 10.7554/eLife.80729.

Morgan E. Levine, Ake T. Lu, Austin Quach, Brian H. Chen, Themistocles L. Assimes, Stefania Bandinelli, Lifang Hou, Andrea A. Baccarelli, James D. Stewart, Yun Li, Eric A. Whitsel, James G Wilson, Alex P Reiner, Abraham Aviv, Kurt Lohman, Yongmei Liu, Luigi Ferrucci, and Steve Horvath. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)*, 10(4):573–591, April 2018. ISSN 1945-4589. doi: 10.18632/aging.101414. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5940111/.

Wenyi Lin, Jingjing Zou, Chongzhi Di, Dorothy D. Sears, Cheryl L. Rock, and Loki Natarajan. Longitudinal Associations Between Timing of Physical Activity Accumulation and Health: Application of Functional Data Methods. *Statistics in Biosciences*, 15(2):309–329, July 2023. ISSN 1867-1772. doi: 10.1007/s12561-022-09359-1. URL https://doi.org/10.1007/s12561-022-09359-1.

Fangyu Liu, Amal A Wanigatunga, and Jennifer A Schrack. Assessment of Physical Activity in Adults Using Wrist Accelerometers. *Epidemiologic Reviews*, 43(1):65–93, December 2021. ISSN 1478-6729. doi: 10.1093/epirev/mxab004. URL https://doi.org/10.1093/epirev/mxab004.

Riccardo E Marioni, Sonia Shah, Allan F McRae, Stuart J Ritchie, Graciela Muniz-Terrera, Sarah E Harris, Jude Gibson, Paul Redmond, Simon R Cox, Alison Pattie, Janie Corley, Adele Taylor, Lee Murphy, John M Starr, Steve Horvath, Peter M Visscher, Naomi R Wray, and Ian J Deary. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *International Journal of Epidemiology*, 44(4): 1388–1396, August 2015. ISSN 0300-5771. doi: 10.1093/ije/dyu277. URL https://doi.org/10.1093/ije/dyu277.

Lisa M. McEwen, Kieran J. O'Donnell, Megan G. McGill, Rachel D. Edgar, Meaghan J. Jones, Julia L. MacIsaac, David Tse Shen Lin, Katia Ramadori, Alexander Morin, Nicole Gladish, Elika Garg, Eva Unternaehrer, Irina Pokhvisneva, Neerja Karnani, Michelle Z. L. Kee, Torsten Klengel, Nancy E. Adler, Ronald G. Barr, Nicole Letourneau, Gerald F. Giesbrecht, James N. Reynolds, Darina Czamara, Jeffrey M. Armstrong, Marilyn J. Essex, Carolina de Weerth, Roseriet Beijers, Marieke S. Tollenaar, Bekh Bradley, Tanja Jovanovic, Kerry J. Ressler, Meir Steiner, Sonja Entringer, Pathik D. Wadhwa, Claudia Buss, Nicole R. Bush, Elisabeth B. Binder, W. Thomas Boyce, Michael J. Meaney, Steve Horvath, and Michael S. Kobor. The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. *Proceedings of the National Academy of Sciences*, 117(38):23329–23335, September 2020. doi: 10.1073/pnas.1820843116. URL https://www.pnas.org/doi/full/10.1073/pnas.1820843116. Publisher: Proceedings of the National Academy of Sciences.

Alan Miller. *Subset Selection in Regression*. Chapman and Hall/CRC, 0 edition, April 2002. ISBN 978-0-429-11918-7. doi: 10.1201/9781420035933. URL https://www.taylorfrancis.com/books/9781420035933.

Olga Minaeva, Sanne H. Booij, Femke Lamers, Niki Antypa, Robert A. Schoevers, Marieke Wichers, and Harriëtte Riese. Level and timing of physical activity during normal daily life in depressed and non-depressed individuals. *Translational Psychiatry*, 10(1):1–11, July 2020. ISSN 2158-3188. doi: 10.1038/s41398-020-00952-w. URL https://www.nature.com/articles/s41398-020-00952-w. Number: 1 Publisher: Nature Publishing Group.

Joseph Naiman and Peter Xuekun Song. Multivariate Functional Kernel Machine Regression and Sparse Functional Feature Selection. *Entropy*, 24(2):203, January 2022. ISSN 1099-4300. doi: 10.3390/e24020203. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8871497/.

B. K. Natarajan. Sparse Approximate Solutions to Linear Systems. *SIAM Journal on Computing*, 24(2):227–234, April 1995. ISSN 0097-5397. doi: 10.1137/S0097539792240406. URL https://epubs.siam.org/doi/abs/10.1137/S0097539792240406. Publisher: Society for Industrial and Applied Mathematics.

Jamaji C. Nwanaji-Enwerem, Lars Van Der Laan, Katherine Kogut, Brenda Eskenazi, Nina Holland, Julianna Deardorff, and Andres Cardenas. Maternal adverse childhood experiences before pregnancy are associated with epigenetic aging changes in their children. *Aging*, 13(24):25653–25669, December 2021. ISSN 1945-4589. doi: 10.18632/aging.203776. URL https://www.aging-us.com/article/203776/text.

Wei Perng, Marcela Tamayo-Ortiz, Lu Tang, Brisa N. Sánchez, Alejandra Cantoral, John D. Meeker, Dana C. Dolinoy, Elizabeth F. Roberts, Esperanza Angeles Martinez-Mier, Hector Lamadrid-Figueroa, Peter X. K. Song, Adrienne S. Ettinger, Robert Wright, Manish Arora, Lourdes Schnaas, Deborah J. Watkins, Jaclyn M. Goodrich, Robin C. Garcia, Maritsa Solano-Gonzalez, Luis F. Bautista-Arredondo, Adriana Mercado-Garcia, Howard Hu, Mauricio Hernandez-Avila, Martha Maria Tellez-Rojo, and Karen E. Peterson. Early Life Exposure in Mexico to ENvironmental Toxicants (ELEMENT) Project. *BMJ Open*, 9(8): e030427, August 2019. ISSN 2044-6055, 2044-6055. doi: 10.1136/bmjopen-2019-030427. URL https://bmjopen.bmj.com/content/9/8/e030427. Publisher: British Medical Journal Publishing Group Section: Public health.

Austin Quach, Morgan E. Levine, Toshiko Tanaka, Ake T. Lu, Brian H. Chen, Luigi Ferrucci, Beate Ritz, Stefania Bandinelli, Marian L. Neuhouser, Jeannette M. Beasley, Linda Snetselaar, Robert B. Wallace, Philip S. Tsao, Devin Absher, Themistocles L. Assimes, James D. Stewart, Yun Li, Lifang Hou, Andrea A. Baccarelli, Eric A. Whitsel, and Steve Horvath. Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging (Albany NY)*, 9(2):419–437, February 2017. ISSN 1945-4589. doi: 10.18632/aging.101168. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5361673/.

L. Rabiner and B. Juang. An introduction to hidden Markov models. *IEEE ASSP Magazine*, 3(1):4–16, January 1986. ISSN 1558-1284. doi: 10.1109/MASSP.1986.1165342. Conference Name: IEEE ASSP Magazine.

J. O. Ramsay. Functional Data Analysis. In *Encyclopedia of Statistical Sciences*. John Wiley & Sons, Ltd, 2004. ISBN 978-0-471-66719-3. doi: 10.1002/0471667196.ess0646. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/0471667196.ess0646. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/0471667196.ess0646.

James Ramsay. Functional Data Analysis. In *Encyclopedia of Statistics in Behavioral Science*. John Wiley & Sons, Ltd, 2005. ISBN 978-0-470-01319-9. doi: 10.1002/0470013192.bsa239. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/0470013192.bsa239. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/0470013192.bsa239.

Philip T. Reiss, Jeff Goldsmith, Han Lin Shang, and R. Todd Ogden. Methods for Scalar-on-Function Regression. *International Statistical Review*, 85(2):228–249, 2017. ISSN 1751-5823. doi: 10.1111/insr.12163. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/insr.12163. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/insr.12163.

Mert Sevil, Mudassir Rashid, Zacharie Maloney, Iman Hajizadeh, Sediqeh Samadi, Mohammad Reza Askari, Nicole Hobbs, Rachel Brandt, Minsun Park, Laurie Quinn, and Ali Cinar. Determining Physical Activity Characteristics from Wristband Data for Use in Automated Insulin Delivery Systems. *IEEE sensors journal*, 20(21):12859–12870, November 2020. ISSN 1530-437X. doi: 10.1109/jsen.2020.3000772.

Robin P Shook, Gregory A Hand, Clemens Drenowatz, James R Hebert, Amanda E Paluch, John E Blundell, James O Hill, Peter T Katzmarzyk, Timothy S Church, and Steven N Blair. Low levels of physical activity are associated with dysregulation of energy intake and fat mass gain over 1 year12. *The American Journal of Clinical Nutrition*, 102(6): 1332–1338, December 2015. ISSN 0002-9165. doi: 10.3945/ajcn.115.115360. URL https://www.sciencedirect.com/science/article/pii/S0002916523272056.

Luisa Soares-Miranda, David S. Siscovick, Bruce M. Psaty, W. T. Longstreth, and Dariush Mozaffarian. Physical Activity and Risk of Coronary Heart Disease and Stroke in Older Adults. *Circulation*, 133(2):147–155, January 2016. doi: 10.1161/CIRCULATIONAHA.115.018323. URL https://www.ahajournals.org/doi/full/10.1161/CIRCULATIONAHA.115.018323. Publisher: American Heart Association.

Peter X.-K. Song, Zhichang Jiang, Eunjoo Park, and Annie Qu. Quadratic inference functions in marginal models for longitudinal data. *Statistics in Medicine*, 28(29):3683–3696, 2009. ISSN 1097-0258. doi: 10.1002/sim.3719. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.3719. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.3719.

Robert Tibshirani. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996. ISSN 2517-6161. doi: 10.1111/j.2517-6161.1996.tb02080.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1996.tb02080.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1996.tb02080.x.

Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108, 2005. ISSN 1467-9868. doi: 10.1111/j.1467-9868.2005.00490.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2005.00490.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2005.00490.x.

Richard P. Troiano, David Berrigan, Kevin W. Dodd, Louise C. Mâsse, Timothy Tilert, and Margaret McDowell. Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise*, 40(1):181–188, January 2008. ISSN 0195-9131. doi: 10.1249/mss.0b013e31815a51b3.

Richard P. Troiano, James J. McClain, Robert J. Brychta, and Kong Y. Chen. Evolution of accelerometer methods for physical activity research. *British Journal of Sports Medicine*, 48(13):1019–1023, July 2014. ISSN 1473-0480. doi: 10.1136/bjsports-2014-093546.

Stewart G. Trost. Objective Measurement of Physical Activity in Youth: Current Issues, Future Directions. *Exercise and Sport Sciences Reviews*, 29(1):32–36, January 2001. ISSN 0091-6331. URL https://journals.lww.com/acsm-essr/Fulltext/2001/01000/Objective_Measurement_of_Physical_Activity_in.7.aspx.

Christiana M.T. van Loo, Anthony D. Okely, Marijka J. Batterham, Trina Hinkley, Ulf Ekelund, Søren Brage, John J. Reilly, Stewart G. Trost, Rachel A. Jones, Xanne Janssen, and Dylan P. Cliff. Wrist Accelerometer Cut-points for Classifying Sedentary Behavior in Children. *Medicine and science in sports and exercise*, 49(4):813–822, April 2017. ISSN 0195-9131. doi: 10.1249/MSS.0000000000001158. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5332065/.

Wen Wang, Shihao Wu, Ziwei Zhu, Ling Zhou, and Peter X. K. Song. Supervised homogeneity fusion: a combinatorial approach, 2022.

Darren E. R. Warburton and Shannon S. D. Bredin. Reflections on Physical Activity and Health: What Should We Recommend? *Canadian Journal of Cardiology*, 32(4):495–504, April 2016. ISSN 0828-282X. doi: 10.1016/j.cjca.2016.01.024. URL https://www.sciencedirect.com/science/article/pii/S0828282X16000647.

Petri Wiklund, Ville Karhunen, Rebecca C. Richmond, Priyanka Parmar, Alina Rodriguez, Maneka De Silva, Matthias Wielscher, Faisal I. Rezwan, Tom G. Richardson, Juha Veijola, Karl-Heinz Herzig, John W. Holloway, Caroline L. Relton, Sylvain Sebert, and Marjo-Riitta Järvelin. DNA methylation links prenatal smoking exposure to later life health outcomes in offspring. *Clinical Epigenetics*, 11(1):97, July 2019. ISSN 1868-7083. doi: 10.1186/s13148-019-0683-4. URL https://doi.org/10.1186/s13148-019-0683-4.

Laurence A. Wolsey. Mixed Integer Programming. In *Wiley Encyclopedia of Computer Science and Engineering*, pages 1–10. John Wiley & Sons, Ltd, 2008. ISBN 978-0-470-05011-8. doi: 10.1002/9780470050118.ecse244. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470050118.ecse244. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470050118.ecse244.

Xiaohui Wu, Weidan Chen, Fangqin Lin, Qingsheng Huang, Jiayong Zhong, Huan Gao, Yanyan Song, and Huiying Liang. DNA methylation profile is a quantitative measure of biological aging in children. *Aging*, 11(22):10031–10051, November 2019. ISSN 1945-4589. doi: 10.18632/aging.102399. URL https://www.aging-us.com/article/102399/text.

Yue Wu, Jaclyn M. Goodrich, Dana C. Dolinoy, Brisa N. Sánchez, Edward A. Ruiz-Narváez, Margaret Banker, Alejandra Cantoral, Adriana Mercado-Garcia, Martha M. Téllez-Rojo, and Karen E. Peterson. Accelerometer-measured Physical Activity, Reproductive Hormones, and DNA Methylation. *Medicine and Science in Sports and Exercise*, 52(3):598–607, March 2020. ISSN 1530-0315. doi: 10.1249/MSS.0000000000002175.

Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association*, 100(470):577–590, 2005. ISSN 0162-1459. URL https://www.jstor.org/stable/27590579. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

Jaehong Yu and Hua Zhong. Time Varying Mixed Effects Model with Fused Lasso Regularization. *Journal of Applied Statistics*, 48(8):1513–1526, 2021. ISSN 0266-4763. doi: 10.1080/02664763.2020.1791805.

Scott L. Zeger, Kung-Yee Liang, and Paul S. Albert. Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrics*, 44(4):1049–1060, 1988. ISSN 0006-341X. doi: 10.2307/2531734. URL https://www.jstor.org/stable/2531734. Publisher: [Wiley, International Biometric Society].

Cun-Hui Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894–942, April 2010. ISSN 0090-5364, 2168-8966. doi: 10.1214/09-AOS729. URL https://projecteuclid.org/journals/annals-of-statistics/volume-38/issue-2/Nearly-unbiased-variable-selection-under-minimax-concave-penalty/10.1214/09-AOS729.full. Publisher: Institute of Mathematical Statistics.

Yunzhang Zhu, Xiaotong Shen, and Wei Pan. Simultaneous Grouping Pursuit and Feature Selection Over an Undirected Graph. *Journal of the American Statistical Association*, 108 (502):713–725, 2013. ISSN 0162-1459. URL https://www.jstor.org/stable/24246476. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

Hui Zou. The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101(476):1418–1429, December 2006. ISSN 0162-1459. doi: 10.1198/016214506000000735. URL https://doi.org/10.1198/016214506000000735. Publisher: Taylor & Francis _eprint: https://doi.org/10.1198/016214506000000735.