

REGULARIZED SCALAR-ON-FUNCTION REGRESSION ANALYSIS TO ASSESS FUNCTIONAL ASSOCIATION OF CRITICAL PHYSICAL ACTIVITY WINDOW WITH BIOLOGICAL AGE

BY MARGARET BANKER^{1,a},
LEYAO ZHANG^{1,b} AND PETER X.K. SONG^{1,c}

¹University of Michigan, Department of Biostatistics, ^ambanker@umich.edu; ^bleyaozh@umich.edu; ^cpxsong@umich.edu

Accelerometry data enables scientists to extract personal digital features useful in precision health decision making. Existing analytic methods often begin with discretizing Physical Activity (PA) counts into activity categories via fixed cutoffs; however, the cutoffs are validated under restricted settings and cannot be generalized across studies. Here, we develop a data-driven approach to overcome this bottleneck in the analysis of PA data, in which we holistically summarize an individual's PA profile using Occupation-Time Curves that describe the percentage of time spent at or above a continuum of activity levels. The resulting functional curve is informative to capture time-course individual variability of PA. We investigate functional analytics under an L_0 regularization approach, which handles highly correlated micro-activity windows that serve as predictors in a scalar-on-function regression model. We develop a new one-step method that simultaneously conducts fusion via change-point detection and parameter estimation through a new L_0 constraint formulation, which is evaluated via simulation experiments and a data analysis assessing the influence of PA on biological aging.

1. Introduction.

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. 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).

Keywords and phrases: Functional Data Analysis, Fusion Regularization, Occupation Time Curve, Scalar-on-Function Regression, Wearable Device.

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 physical activity (PA) in adolescence with biological aging. Research and conventiional wisdom suggest that increased physical activity may slow epigenetic aging (Kankaanpää et al., 2022; Quach et al., 2017). By promoting physical activity 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 paper is to investigate the association of epigenetic age with objectively measured functional PA as captured by wearable devices.

1.2. Wearable Devices and Accelerometer Data. Wearable technologies use devices 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 1 depicts continuous time-series VM count data for an individual from our motivating data detailed in Section 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

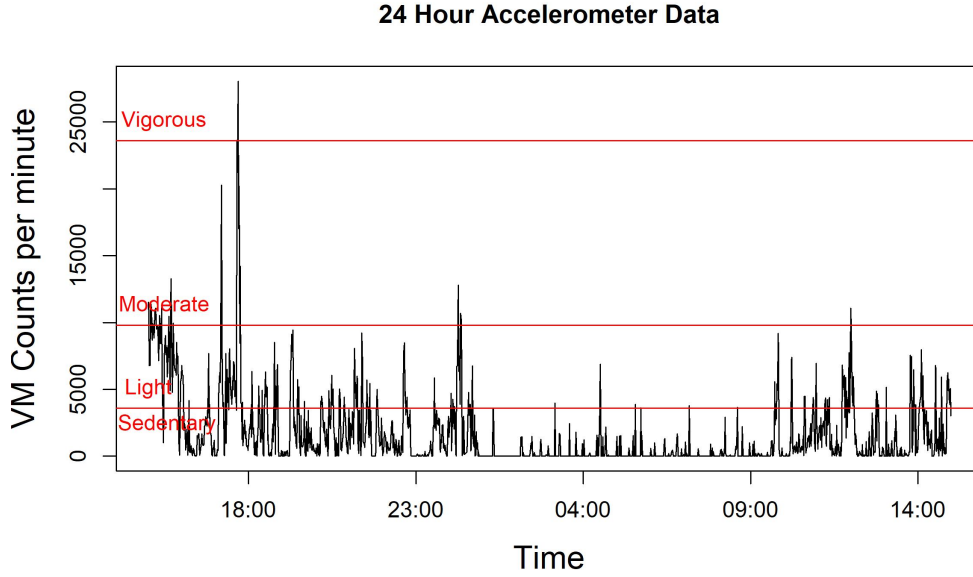


Figure 1: Accelerometer Data for an individual over a 24 hour period. The horizontal lines indicate activity categorization cutoffs based on Chandler Vector Magnitude cutoffs for 1-minute epochs (Chandler et al., 2016). These cutoffs are pre-determined by in-lab supervised research and then applied to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

and a health outcome of interest (Freedson, Pober and Janz, 2005; Crouter, Flynn and Bassett, 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 generalizable approach to relax or even eliminate the dependence on pre-fixed cutoffs.

1.3. Wearable Devices and Functional Data Analysis. 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.

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 and Silverman (2005), Ferraty and Vieu (2006), and Horváth and Kokoszka (2012), among others.

1.4. Study Objectives. This need of data-adaptive cutoffs motivates the statistical objective of this paper: to develop a generalized, functional-focused approach to analyze PA data with the aim to free the dependence on subjective choices of 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. Of note, by a more “generalizable” approach, we refer to methodological and algorithmic generalizability in terms of the ability to adaptively determine cutpoints when analyzing data from various devices. In this way, practitioners will not have to rely on some pre-established cutpoints, which need to be validated for every new device/device placement/population of interest. Here we consider actigraphy data under the purview of Occupation Time Curves (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). Section 3 introduces these OTCs 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 versus 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, King and Mazumder, 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 paper is as follows. We introduce the motivating cohort study in Section 2. Section 3 concerns the functional OTC variables, while Section 4 compares existing and proposed model formulations. Section 5 introduces MIO and presents its formulation for our scalar-on-function statistical analysis, with a discussion of theoretical guarantees in Section 6. In Section 7 we explore numeric experiments illustrating the capabilities of this approach, while Section 8 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 9. Some additional numerical results are included in the Supplementary Material.

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. Briefly, in 2015 researchers collected actigraphy data from 549 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 directed to wear a wrist-worn, tri-axial Actigraph GT3X+ (Actigraph LLC) for

seven consecutive days with no interruptions. As this Actigraph device is water-resistant and can be removed only when physically cut off, the study warranted high compliance and limited non-wear time during data collection. While the utilized Actigraph GT3X+ collects 30 measurements per second (30 Hz), this raw tri-axial high-frequency time-series data was processed and summarized into epochs of various lengths (e.g. 30 sec, 1 min). In this paper, we focus on analyzing activity counts over one-minute epochs, which is widely used in practice.

In addition to PA, the ELEMENT cohort also collected DNA methylation data from EPIC array (850K) that was used to calculate 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.

3. Occupation Time Curves: A Functional Predictor. The Occupation Time Curve (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: $\frac{\text{duration of } \{t: VM(t) \geq c\}}{\text{total duration of } VM(t)}$.

Figure 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 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 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 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 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 4) 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 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 4.

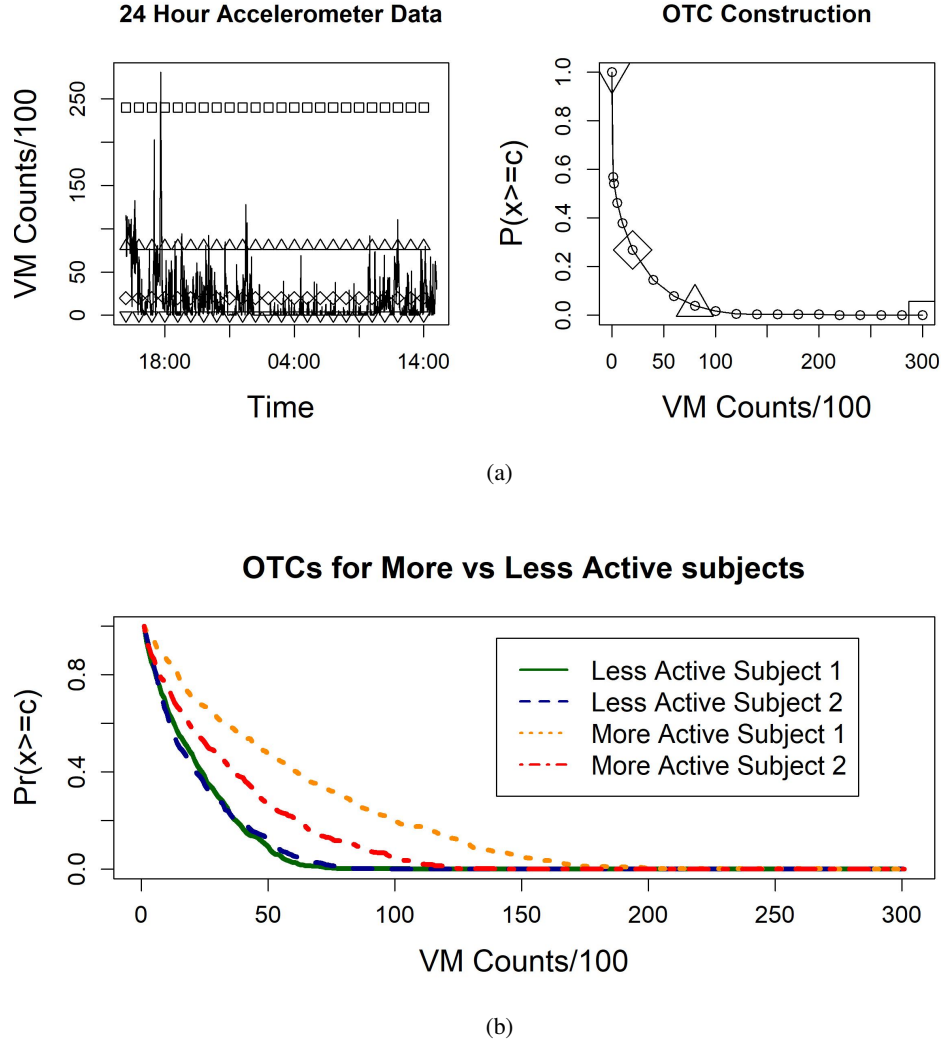


Figure 2: (a) An illustration of the construction of an OTC from a time-series of VM counts. The left panel represents accelerometer data, with a rising bar of “count” cutoff indicated by the horizontal lines with varying point-shapes. The corresponding proportion of time spent at or above that level of activity is indicated in the right panel by the corresponding point shape. The continuous curve is the realized OTC for this individual. (b) Comparison of OTC shapes for More Active vs Less Active individuals, with VM count summarized over 1-minute epochs varying over 0 to 30000. The distinctive shapes of the OTCs represent the subject’s activity pattern. For example, curves for less active people decay quickly in the beginning, signifying that a small percentage of their time is spent in even mid-active regions.

4. Model Formulations. There are many Functional Data Analysis (FDA) techniques that can be used to analyze various aspects of these functional data sets. Some analytical methods within this framework include: functional regression (Ramsay, 2005; Reiss et al., 2017), which can be used to model the relationship between functional data and other variables; functional principal component analysis (FPCA) (Ramsay, 2004, 2005; Goldsmith, Zipunnikov and Schrack, 2015; Yao, Müller and Wang, 2005; Nwanaji-Enwerem et al., 2021;

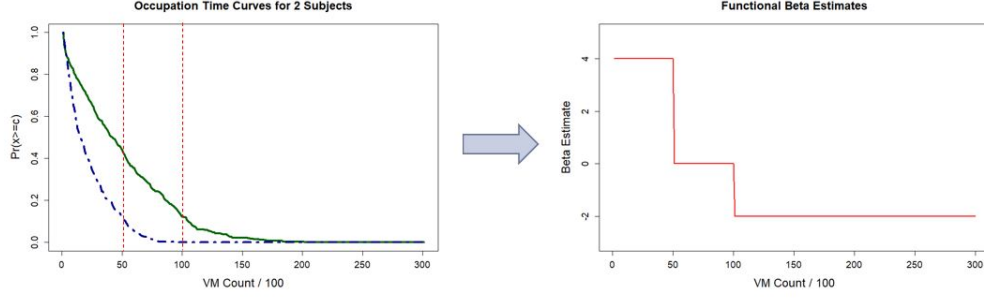


Figure 3: Left panel: Two realized OTCs with vertical dashed lines representing the cutoffs for activity window ranges of interest. Right panel: Example of a non-constant functional β -parameter as a step-function to estimate the association of PA in selected activity-count windows with a health outcome of interest

Chen and Müller, 2012), which can help in dimension reduction and identifying dominant patterns variation; and functional classification and clustering (Heinzel and Tutz, 2014), which involve grouping functional data based on measures of similarity or dissimilarity. This paper will focus on functional linear regression techniques with functional predictors. These techniques extend the concept of linear regression to functional predictors and responses. More specifically, it allows one to model the relationship between functional variables, such as predicting a response curve based on a set of predictor curves. Here, 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.

4.1. 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:

$$(1) \quad Y = \int_{\mathcal{C}} \theta(c) X(c) dc + \mathbf{Z}^T \boldsymbol{\alpha} + \epsilon,$$

where $\mathbf{Y} \in \mathbb{R}^{n \times 1}$; $X(c)$ is the functional OTC defined on $\mathcal{C} \subset \mathbb{R}$; $\mathbf{Z}$ is a q -dimensional vector of confounders with corresponding parameter vector $\boldsymbol{\alpha}$; ϵ is the vector of error term with mean 0 and variance σ^2 ;

The goal of a scalar-on-function model as defined above is to estimate the functional parameter $\theta(c)$. In practice, researchers are typically interested in estimating a smooth curve as functional parameter $\theta(c)$. In these cases, a popular choice is to estimate $\theta(c)$ via basis expansions, such as B-splines, natural cubic splines, and tensors, among others (Goldsmith, Zipunnikov and Schrack, 2015; Goldsmith et al., 2012). In this case, however, our goal is to discretize the continuous functional parameter estimate $\theta(c)$ as a piece-wise function with change points, thereby effectively defining windows of PA levels by fusing the $\theta(c)$ of adjacent count ranges with similar effects on the outcome. In other words, for a given K number of PA windows (e.g. $K = 3$), we aim to estimate both the change points $c_1, \dots, c_{K-1}$ and step-function parameter values simultaneously. In this way, we reparameterize the functional

parameter $\theta(c)$ into a step-function parameter represented with $\beta_1, \dots, \beta_K$ as the respective coefficients for activity windows $[0, c_1], (c_1 + 1, c_2], \dots, (c_{K-1}, c_{max}]$. Here, c_{max} denoted the maximum activity count considered in the analysis. Thus, the changepoints impose an abrupt change in the functional coefficient of physical activity, which influences the mean function of the outcome. Such categorization in the stepwise function is reflective of the fact that not all physical activity ranges nor every small PA changes would impact health outcomes; rather, critical influential windows of activity, if they exist, are appealing for the sake of clinical interpretation and subsequent translational research. Thus, a piece-wise linear functional coefficient would provide distinct activity intensity ranges that are associated with the health outcome of interest, as well as estimates of association magnitude.

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

$$\begin{aligned}
 \int_{\mathcal{C}} \theta(c) X(c) dc + \mathbf{Z}^T \boldsymbol{\alpha} &= \sum_{j=1}^J \int_{c_{j-1}}^{c_j} \theta(c) X(c) dc + \mathbf{Z}^T \boldsymbol{\alpha} \\
 (2) \quad &\approx \sum_{j=1}^J \theta_j \int_{c_{j-1}}^{c_j} X(c) dc + \mathbf{Z}^T \boldsymbol{\alpha} \\
 &:= \sum_{j=1}^J \theta_j A_j + \mathbf{Z}^T \boldsymbol{\alpha},
 \end{aligned}$$

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 $\mathbf{Z}$ is a q -dimensional vector of confounders with corresponding parameter vector $\boldsymbol{\alpha}$. Unlike the conventional functional regression analysis, in the same spirit of categorization shown in Figure 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 $k = 1, \dots, K$. This results in a final estimate model: $\sum_{k=1}^K \beta_k A_k + \mathbf{Z}^T \boldsymbol{\alpha}$, 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 window and its influence on the outcome, as well as and their interpretability.

4.2. Existing L_1 Regularization Approaches. 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 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, King and Mazumder, 2016). While this bias often can be controlled via various correction methods (such as adaptive lasso) (Candès, Wakin and Boyd, 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.

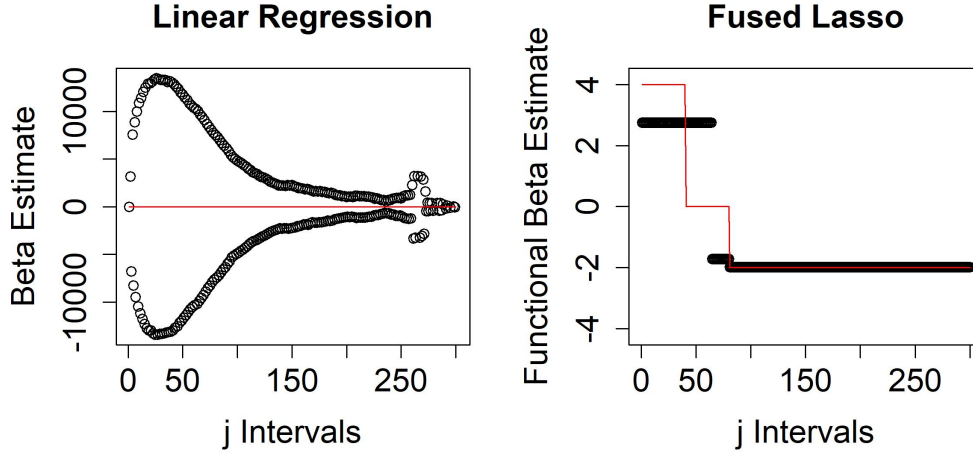


Figure 4: Estimates of β_j coefficients from two different standard analysis approaches including multiple linear regression (left) and an L_1 Regularization approach of Fused Lasso (right). The respective model parameter estimates are represented by the black circles, while the true β_j values are represented as the step function.

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 (2) with a null covariate matrix $\mathbf{Z}$. 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 $\beta(c)$ 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 4 (the right panel) and Table 1. We obtain similar poor results even when reducing the collinearity to $\text{cor}(A_j, A_{j+1}) = 0.98$, $\text{cor}(A_j, A_{j+5}) = 0.90$, and $\text{cor}(A_j, A_{j+10}) = 0.80$ by setting $J = 60$; see Table 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 $\hat{\beta}_1$ and $\hat{\beta}_2$. Additionally, the left panel of Figure 4 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 have 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 5.

4.3. Integer Programming and L_0 Penalization. Under a modified L_0 optimization strategy, we can simultaneously conduct fusion via change-point 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, King and Mazumder, 2016; Bertsimas and Shioda, 2009; Bertsimas, Pauphilet and Van Parys, 2020). The standard L_0 penalization with constraints on the number of non-zero parameters is not flexible

TABLE 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.

	Truth	J = 60			J = 300		
		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

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 is by means of the the best subset problem. Suppose we have a linear regression model: $\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$ where $\mathbf{y}$ is an $n \times 1$ response vector, $\mathbf{A}$ 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:

$$(3) \quad \min_{\boldsymbol{\theta}} \|\mathbf{Y} - \mathbf{A}\boldsymbol{\theta} - \mathbf{Z}\boldsymbol{\alpha}\|_2^2, \quad \text{subject to } \|\boldsymbol{\theta}\|_0 \leq k,$$

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).

5. 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.

5.1. Proposed Fusion-Adapted MIO Formulation. Bertsimas et al (Bertsimas, King and Mazumder, 2016) offered an MIO formulated-solution to address the best subset problem in Equation (3) using Specially Ordered Sets of Type 1 (SOS-1). In this paper, 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 (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 $\boldsymbol{\eta}$ identifying group membership such that:

$$(4) \quad \boldsymbol{\eta}_k = (\eta_k^1, \eta_k^2, \dots, \eta_k^J) \in \{0, 1\}^{J \times 1}, \quad k = 1, \dots, K$$

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

$$(5) \quad \eta_1^j = \begin{cases} 1, & j = 1, \dots, c_1 \\ 0, & \text{otherwise} \end{cases}, \eta_k^j = \begin{cases} 1, & j = c_{k-1}, \dots, c_k \\ 0, & \text{otherwise} \end{cases}, \dots, \eta_K^j = \begin{cases} 1, & j = c_{K-1} + 1, \dots, J \\ 0, & \text{otherwise} \end{cases}.$$

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

$$(6) \quad \begin{aligned} \min_{\boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{c}, \boldsymbol{\alpha}} \quad & \|\mathbf{Y} - \mathbf{A}\boldsymbol{\theta} - \mathbf{Z}\boldsymbol{\alpha}\|_2^2 \\ \text{subject to} \quad & \boldsymbol{\theta} = (\theta_1, \dots, \theta_J)^T \in \mathbb{R}^{J \times 1}, \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_q)^T \in \mathbb{R}^{q \times 1}; \\ & \mathbf{c} = (c_1, \dots, c_{K-1}) \in \mathbb{N}^{1 \times (K-1)} \\ & c_1 \geq 1, c_k \geq c_{k-1} + 1, c_{K-1} \leq J - 1, k = 1, \dots, K - 1; \\ & \boldsymbol{\eta} = (\eta_j^k)_{J \times K} \in \mathbb{R}^{K \times J} \\ & \theta_j - \beta_1 = 0, j = 1, \dots, c_1; \\ & \theta_j - \beta_2 = 0, j = c_1 + 1, \dots, c_2; \\ & \vdots \\ & \theta_j - \beta_K = 0, j = c_{K-1} + 1, \dots, J, \end{aligned}$$

where $\mathbf{Y} \in \mathbb{R}^{n \times 1}$, $\mathbf{A} \in \mathbb{R}^{n \times J}$, and $\mathbf{Z} \in \mathbb{R}^{n \times q}$. The above optimization is operated via augmented parameters where labels $\boldsymbol{\eta}$ and cutpoints $\mathbf{c} = (c_1, \dots, c_{K-1})^T$ do not exist in the original model (2) but are added for parameter fusion. Obviously, group labels $\boldsymbol{\eta}$ and cutpoints $\mathbf{c}$ are determined in a one-to-one correspondence fashion, which will be enforced via adequate constrains given below in Section 5.2.

5.2. MIO Implementation. This MIO model can be solved via numerical software such as GUROBI under a system of constraints. These constraints set to minimize the objective function by optimizing cutpoints $c_1, \dots, c_{K-1}$ and thus the cluster labels represented by the variable η_k , $k = 1, \dots, K$ defined in Equations (4) and (5). This set of linear constraints for the K -group model is specified as follows:

$$(7) \quad \begin{aligned} & \eta_k^j (\theta_j - \beta_k) = 0, \quad j = 1, \dots, J, k = 1, \dots, K \text{ (SOS-1 constraints);} \\ & \sum_{k=1}^K \eta_k^j = 1, j = 1, \dots, J; \\ & c_0 = 0, c_1 \geq 1; c_k \geq c_{k-1} + 1; \text{ and } c_{K-1} \leq J - 1, \text{ for } k = 2, \dots, K - 1; \\ & \frac{c_k - j}{J} \leq 1 - \eta_{k+1}^j, j = 1, \dots, J, \text{ for } k = 0, \dots, K - 1; \\ & \frac{c_{k+1} - j}{J} \times \frac{(j - c_k)}{J} \leq \eta_{k+1}^j, j = 1, \dots, J, \text{ for } k = 0, \dots, K - 1; \\ & \frac{j - c_{k+1} + 1}{J} \leq 1 - \eta_{k+1}^j, j = 1, \dots, J, \text{ for } k = 0, \dots, K - 1; \end{aligned}$$

These constraints determine the locality of changepoints and grouping in the fused-adaption MIO formulation for the L_0 -type analysis of a K -group model. In this paper, 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.

6. Theoretical Guarantees. Here we discuss the selection consistency of the MIO estimator of the parameter $\boldsymbol{\theta} = (\theta_1, \dots, \theta_J)^T$ obtained by the constrained optimization given in Equations (6) and (7) under some mild regularity conditions. This paves the theoretical basis for large-sample statistical inference. Wang et al. (2022) considered a more general version of an MIO optimization problem than ours, where their group parameters $\theta_1, \dots, \theta_J$ are not sequentially ordered. In other words, the MIO optimization given in Equations (6) and (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 $\mathbf{c}^* = (c_1^*, \dots, c_K^*)$. 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:

$$(8) \quad (\hat{\boldsymbol{\beta}}^{ol}, \hat{\boldsymbol{\alpha}}^{ol}) := \underset{\boldsymbol{\beta}, \boldsymbol{\alpha}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{A}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\alpha}\|_2^2.$$

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 (6) and (7) where cutpoints $\mathbf{c}$ 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, Shen and Pan (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 $\mathbf{c}$. That is,

$$(9) \quad c_{min} \equiv c_{min}(\boldsymbol{\zeta}^*, \mathbf{A}, \mathbf{Z}) = \min_{\boldsymbol{\zeta}} \frac{\|\mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\beta}^*) + \mathbf{Z}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)\|_2^2}{n \max(d(\boldsymbol{\theta}, \boldsymbol{\beta}^*), 1)},$$

subject to Equations (6) and (7).

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

To present selection consistency, for any MIO estimator $\hat{\boldsymbol{\zeta}}^{MIO} = (\hat{\boldsymbol{\beta}}^{MIO\top}, \hat{\boldsymbol{\alpha}}^{MIO\top})^\top$ of $\boldsymbol{\zeta}^*$ with estimated cutpoints $\hat{\mathbf{c}}$ of $\mathbf{c}^*$, we define a loss function $L(\hat{\boldsymbol{\zeta}}^{MIO}; \boldsymbol{\zeta}^*)$ as the grouping risk associated with inaccurate grouping and estimates of $\mathbf{c}^*$ in the form of $L(\hat{\boldsymbol{\zeta}}^{MIO}; \boldsymbol{\zeta}^*) \equiv \mathbb{P}(\hat{\boldsymbol{\zeta}}^{MIO} \neq \boldsymbol{\zeta}^*)$. With both the given number of intervals J and given number of activity windows K , we show that the finite sample error bound is given by:

$$(10) \quad L(\hat{\boldsymbol{\zeta}}^{MIO}; \boldsymbol{\zeta}^*) \leq 4\exp \left[-\frac{3N}{200\sigma^2} \left\{ c_{min} - \frac{\sigma^2}{N} (134 \log(JK) + 220) \right\} \right].$$

This bound in (10) implies that when $c_{min} > \frac{\sigma^2}{n} (134 \log(JK) + 220)$, $\hat{\boldsymbol{\zeta}}^{MIO}$ consistently reconstructs $\boldsymbol{\zeta}^*$ because $N, J \rightarrow \infty$, $\mathbb{P}(\hat{\boldsymbol{\zeta}}^{MIO} \neq \boldsymbol{\zeta}^*) \rightarrow 0$. The proof of this sufficient condition result can be carried out by following the lines of arguments given in the proof of Theorem 2.4 in Wang et al. (2022) and thus is omitted in this paper.

7. 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.

7.1. 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 549 subjects from the ELEMENT dataset were divided into non-overlapping 10-minute segments. Each 10-minute interval was randomly drawn from a pool of 549 10-minute candidate segments. To ensure that the variability in the simulated PA reflected the variability of the ELEMENT dataset as shown in Figure 5, we first classified these 549 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 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: $(c_0 = 0, c_1 = 100, \dots, c_J = 30,000)$. For ease of exposition, we will refer to the $VM/100$ values, i.e. $c = (0/100, \dots, 30000/100)^\top$ or $c = (0, \dots, 300)^\top$. **We normalize the OTC values at each j -interval to mean zero and variance one prior to the MIO operation.**

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, $(A_1^*, A_2^*, A_3^*)^\top$ with $A_k^* = \int_{c_{k-1}}^{c_k} VM(c)dc$. 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 four 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 candidate in order to determine the method’s sensitivity in selecting the right-sized model. **BIC was chosen as the goodness-of-fit (GoF) criterion as it is deemed best suited for discovery in connection to the goal of uncovering the true signals for the underlying model. In many occasions, BIC has been theoretically proved to be selection consistent for the true underlying data generating process. Other measures, such as AIC, are regarded as being more suited for the evaluation of prediction, which is not aligned with the scope of this analytic interest.** (Kass and Raftery, 1995; Vrieze, 2012). 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 the Supplemental Material).

7.2. 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

TABLE 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
Sensitivity:	> 99%				> 99%				> 99%			

handle highly correlated AUCs. Tables 2 and 3 summarize the results from 500 rounds of simulations of the 3-group model for the $J = 300$ and $J = 60$ settings, respectively.

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. As discussed previously, we employed the GoF measure BIC to select the model size as it is best served in these studies of discovering the true signals (Kass and Raftery, 1995). However, similar strong selection accuracy was observed when using the less (more) conservative measure of AIC (EBIC). 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 2 and 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 $\text{cor}(A_j, A_{j+1}) = 0.998$, $\text{cor}(A_j, A_{j+5}) = 0.985$, and $\text{cor}(A_j, A_{j+10}) = 0.967$. 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 Supplemental Tables.

7.3. Comparison to standard FDA approaches. Current FDA approaches aim to estimate functional parameter $\theta(c)$ as a smooth curve. A popular approach to estimate $\theta(c)$ is the cubic B-splines technique or as such. Here we compare the quality of fit for both scalar outcome and coefficient function via measures of Mean Square Error (MSE) and Integrated Mean Square Error (IMSE), respectively, when leveraging this approach to analyze the functional OTCs in a scalar-on-function regression model. To estimate functional parameter $\theta(c)$ we utilize cubic b-splines (R package: `fdac.usc` (Bande et al., 2022)) in two scenarios: (i) Setting the

TABLE 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		
Sensitivity:	> 99%				> 99%				> 99%			
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		
Sensitivity:	> 99%				> 99%				> 99%			
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		
Sensitivity:	> 99%				> 99%				> 99%			

knot-points to the oracle cutpoints of the simulation scenario; and (ii) selecting the number of equally-spaced knot-points via BIC. Table 4 summarizes the performance of the functional estimation as well as quality of fit for the “easy” simulation scenarios of $J = 60, N = 500$.

We also explored to employ existent FDA methodologies to estimate a piece-wise linear coefficient function to better address the need of identifying the underlying critical PA window given by the piece-wise linear coefficient function. We found that two popular functional data R packages `fda` and `fda.usc` cannot produce a stepwise coefficient function with more than one change point (or 2 or more steps). These software packages do not aim at changepoint detection as part of their solution. In contrast, the FDA package `refund` (Goldsmith et al., 2024) allows one to estimate a piece-wise linear coefficient function with multiple breaks; however, such R function has been largely restricted on the use of equally-spaced knots with little flexibility of knot selection. Nevertheless, we compared the performance of the `refund` package with our MIO method in terms of MSE, IMSE and sensitivity of detecting the true group number; see Table 4 for the detail.

These limitations discussed above are major driving factors for inferior performances over our MIO method in terms of MSE, IMSE and sensitivity of detecting the true segments. Using these criteria, we found that in all Scenarios A, B, and C (with varying cutpoints and

TABLE 4

Simulation results for competing FDA methods to estimate a 3-segment functional parameter $\theta(c)$ with number of micro-intervals $J = 60$ and sample size of $N = 500$ summarized over 500 replicates. ‘Oracle knots’ refers to the use of the true cutpoints as knots in the splines smoothing, while ‘GoF knots’ refers to the BIC-based selection of knots from equally-spaced knot points.

Metric	Method	Simulation Scenario		
		A	B	C
MSE	MIO	100.61	98.71	98.82
	Cubic B-splines: Oracle Knots	175.78	102.98	112.52
	Cubic B-splines: GoF knots	98.16	101.95	99.92
	Linear splines: GoF knots	116.83	100.68	100.53
IMSE	MIO	0.002	0.01	0.01
	Cubic b-splines: Oracle knots	0.51	0.03	0.12
	Cubic b-splines: GoF knots	10.20	0.04	0.04
	Linear splines: GoF knots	0.71	0.03	0.05
Segment Number ‘K’	MIO	3	3	3
	Cubic b-splines: Oracle knots	3	3	3
	Cubic b-splines: GoF knots	17.44	7.71	7.36
	Linear splines: GoF knots	21.92	6.34	11.98

effect sizes), the MIO approach clearly outperformed the existent FDA methodologies. In the two cubic B-splines methods, the inferior IMSE is not surprising, as the methods estimate a smoothed curve of $\theta(c)$, rather than a piece-wise linear function. In terms of MSE, the MIO approach demonstrates very similar, or better, results in all scenarios. The cubic B-spline method that utilizes BIC to determine the number of knot points performs the most similar to our “MIO” method; however, it over-selected knot points, resulting in over-fitting. It is evident from this simulation experiment that these existent FDA methods do not reach a comparable level of the quality of fit in comparison to the proposed MIO method. Moreover, the detection of the critical windows of PA, as determined by the “K” selected knot points, are vital. With the existing methods, it is hard to identify satisfactorily the locations of these knot points and thus establish the critical windows.

7.4. 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.

8. 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, 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) $\mu\text{g}/\text{dL}$. The

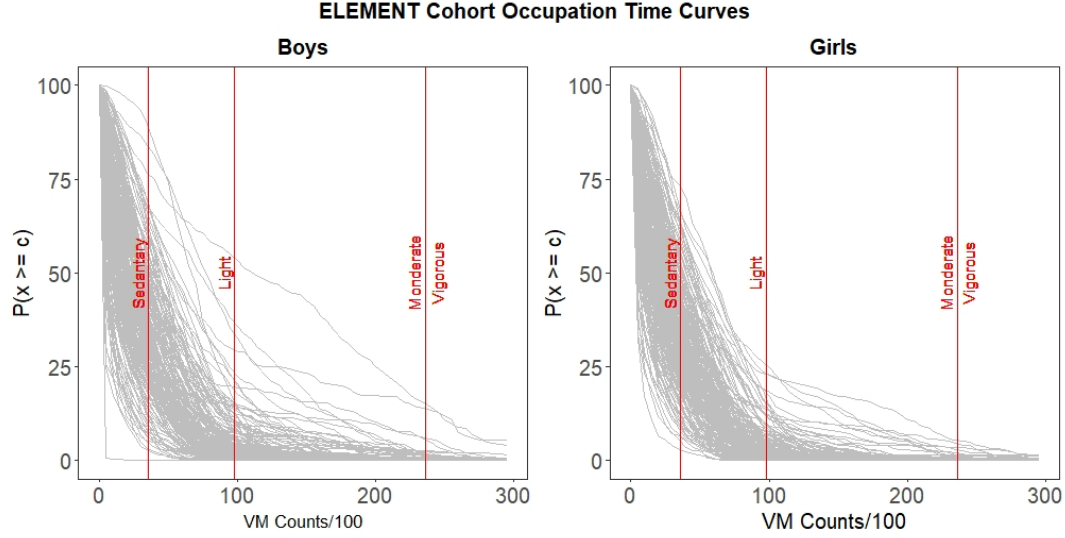


Figure 5: OTCs for 354 ELEMENT subjects stratified by boys and girls. The vertical lines represent Chandler’s cutoffs (Chandler et al., 2016) for prefixed activity levels (Sedentary, Light, Moderate, Vigorous). The relative shape of OTC reflects the subject’s activity profile.

reduction to 354 subjects was mainly driven by availability of the covariates of interest, which were assessed to be Missing Completely at Random (the detail omitted). The majority (332) of subjects had completed puberty in terms of Tanner staging standards. Figure 5 illustrates the functional predictors of OTCs representing the subjects’ activity profiles from 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 5, we began setting $J = 300$, with each interval covering 100 VM counts, followed by an augmentation scenario of $J = 60$ by summing every five. For ease of interpretation, we considered $K \in \{2, 3, 4\}$ PA windows, in which 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 5 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 follows approximately uniform distribution on $(0, 1)$, and remarkably similar to the original analytic p -values, as evident in Table 5. 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.

TABLE 5

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.

Parameters	J = 300						J = 60								
	K=2			K=3			K=2			K=3			K=4		
	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		

To facilitate a clinically understandable interpretation of analysis results in Table 5, we propose an AUC Ratio metric that measures the amount of time they spend within a PA window relative to the maximum amount of time they could spend above their 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 6, the AUC Ratio is a ratio of the individual's AUC in the detected activity window k (A_{ik}) versus that the area of the rectangle (R_k), with the latter representing PA of an individual who spends all of his or her time above the PA level of this window. In Figure 6, the AUC Ratio for the first window is calculated by $\text{AUC Ratio}_{i1} = \frac{A_{i1}}{R_1}$. The interpretation of this 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.

In general, for all but the last sequential window, i.e. for windows $1, \dots, K-1$, the value $(1 - \text{AUC Ratio})$ represents a percentage that the individual is Less Active than the hypothetical most active person in that window. For example, in Figure 6, Subject 1 has a smaller AUC (dark grey shaded region) than Subject 2, representing that Subject 1 spends more time within the cutpoints (c_0, c_1) than Subject 2. The value $(1 - \text{AUC Ratio}_{i1})$, or $\frac{R_1 - A_{i1}}{R_1}$, represents the

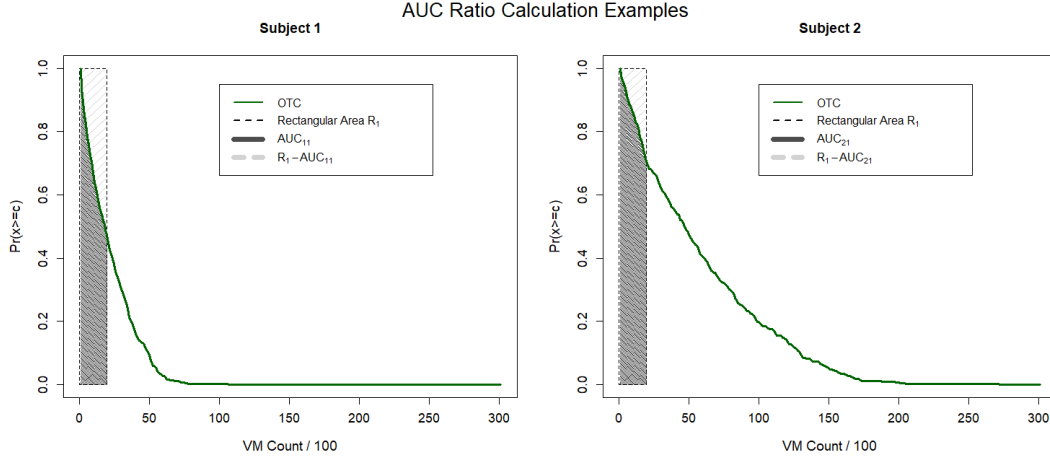


Figure 6: An illustration of the AUC Ratio calculation for Window 1. The dark grey shaded regions below the OTC curves represent the AUC for subjects 1 and 2, or A_{i1} and A_{i2} , respectively. The outlined rectangle represents the area of the full rectangle with cutpoints ($c_0 = 0, c_1 = 20$), deemed R_1 representing the hypothetical subject spending 100% of time above this activity level.

percentage Subject 1 is less active than the hypothetical most active person within the first window. This value is greater than $(1 - \text{AUC Ratio}_{21})$, or $\frac{R_1 - A_{21}}{R_1}$, as can be visualized by the area of the hashed shaded regions above the OTC curves. For the K^{th} window (i.e. the last one), 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_{iK} value represents higher PA within the window.

Let us interpret the results in Table 5 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 $\text{AUC } A_k = \int_{c_{k-1}}^{c_k} \text{OTC}(c) dc, k = 1, 2, 3$. 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 $\hat{\beta}_{1\text{Ratio}} = 20\hat{\beta}_1$ for the interpretability. 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 Ratio, this subject’s BloodSkin epigenetic age decreases approximately 80 days. See Figure 6 for a schematic of this calculation.

9. Discussion. In this paper we utilize a scalar-on-function model to assess the influence of physical activity on biological age using a 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 detects key cutpoints to define critical windows of activity and estimates 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), our functional OTC predictor would likely form a robust functional physical activity 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 physical activity 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.

A potential limitation of this adaptive data-driven cutpoint detection mechanism is that the activity windows are not yet tied to specific known activity types or METs (metabolic equivalent of task). This calls for an extension of our method, which however requires relevant information on device-specific characterization of PA patterns and METs. Thus, a device under investigation must be validated for populations of interest. This would lead to a major bottle-neck in analysis of PA patterns as more and more different devices are being used on populations of varying characteristics. Our approach unveils a marginal relationship between activity intensity levels and the health outcome, which may be applied prior to any validation studies for device/population of interest. In this case, however, the ability to make cross-study comparisons could be limited due to the adaptive nature of detected cutpoints.

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.

Our analysis gives rise to a data analytic toolbox enabling to explore various questions of interest related to the effect of functional physical activity features on health outcomes. For example, some researchers hypothesize that the timing of physical activity, not only they relative intensity, is related to specific health outcomes. Through application of this MIO technique focusing on physical activity during different time periods of the day, such as morning versus evening, researchers can investigate if the activity intensity change points are dependent on time of day. Additionally, future extensions can include multiple functional covariates to assess the longitudinal affect of functional physical activity profiles on health outcomes, and even longitudinal effects with both repeated outcomes and functional exposures to physical activity using longitudinal functional data analysis models. This proposed fusion MIO method can also easily be extended to include variable selection procedures for the selection of important covariates. For example, one could include additional L_0 regularization constraints for the real-values covariables of interest, akin to Bertsimas' L_0 'best-subset' solution (Bertsimas, King and Mazumder, 2016). We have implemented this extension (not shown in this paper), in which the functional coefficient can be estimated via the fusion constraints, while the additional covariables estimated via the best-subset constraints.

While this paper focused on time series of physical activity counts from wearable accelerometer devices, the use of Occupation-Time Curves 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.

10. Acknowledgements. The authors would like to thank Editor, Associate Editor, and two anonymous reviewers for their constructive comments that have helped improve the manuscript greatly. They want also to thank Dr. Karen Peterson and Laura Arboleda Merino for access to ELEMENT data, as well as Dr. Andrew Pitchford for guidance on physical activity accelerometer data. This work is supported by R24ES028502 and NSFDMS2113564.

REFERENCES

- BANDE, M. F., FUENTE, M. O. D. L., GALEANO, P., NIETO, A. and GARCIA-PORTUGUES, E. (2022). *fda.usc: Functional Data Analysis and Utilities for Statistical Computing*.
- BERTSIMAS, D., KING, A. and MAZUMDER, R. (2016). Best subset selection via a modern optimization lens. *The Annals of Statistics* **44** 813–852.
- BERTSIMAS, D., PAUPHILET, J. and VAN PARYS, B. (2020). Sparse Regression: Scalable algorithms and empirical performance. *Statistical Science* **35**.
- BERTSIMAS, D. and SHIODA, R. (2009). Algorithm for cardinality-constrained quadratic optimization. *Computational Optimization and Applications* **43** 1–22.
- BOGACHEV, L. and RATANOV, N. (2011). Occupation time distributions for the telegraph process. *Stochastic Processes and their Applications* **121** 1816–1844.
- CANDÈS, E. J. and PLAN, Y. (2009). Near-ideal model selection by ℓ_1 minimization. *The Annals of Statistics* **37** 2145–2177.
- CANDÈS, E. J., WAKIN, M. B. and BOYD, S. P. (2008). Enhancing Sparsity by Reweighted ℓ_1 Minimization. *Journal of Fourier Analysis and Applications* **14** 877–905.
- CHANDLER, J. L., BRAZENDALE, K., BEETS, M. W. and MEALING, B. A. (2016). Classification of physical activity intensities using a wrist-worn accelerometer in 8–12-year-old children. *Pediatric Obesity* **11** 120–127.
- CHEN, K. Y. and BASSETT, D. R. (2005). The technology of accelerometry-based activity monitors: current and future. *Medicine and Science in Sports and Exercise* **37** S490–500.
- CHEN, K. and MÜLLER, H.-G. (2012). Modeling Repeated Functional Observations. *Journal of the American Statistical Association* **107** 1599–1609.
- CROUTER, S. E., FLYNN, J. I. and BASSETT, D. R. (2015). Estimating Physical Activity in Youth Using a Wrist Accelerometer. *Medicine and science in sports and exercise* **47** 944–951.
- (2006). Introduction to Functional Nonparametric Statistics. In *Nonparametric Functional Data Analysis: Theory and Practice*, (F. FERRATY and P. VIEU, eds.). *Springer Series in Statistics* 5–10. Springer, New York, NY.
- FREEDSON, P., POBER, D. and JANZ, K. F. (2005). Calibration of accelerometer output for children. *Medicine and Science in Sports and Exercise* **37** S523–530.
- GOLDSMITH, J., ZIPUNNIKOV, V. and SCHRACK, J. (2015). Generalized multilevel function-on-scalar regression and principal component analysis. *Biometrics* **71** 344–353.
- GOLDSMITH, J., CRAINICEANU, C. M., CAFFO, B. and REICH, D. (2012). Longitudinal Penalized Functional Regression for Cognitive Outcomes on Neuronal Tract Measurements. *Journal of the Royal Statistical Society. Series C, Applied Statistics* **61** 453–469.
- GOLDSMITH, J., SCHEIPL, F., HUANG, L., WROBEL, J., DI, C., GELLAR, J., HAREZLAK, J., MCLEAN, M. W., SWIHART, B., XIAO, L., CRAINICEANU, C., REISS, P. T., CHEN, Y., GREVEN, S., HUO, L., KUNDU, M. G., PARK, S. Y., MILLER, D. L., STAIKU, A.-M., CUI, E., LI, R. and LI, Z. (2024). *refund: Regression with Functional Data*.
- HEINZL, F. and TUTZ, G. (2014). Clustering in linear-mixed models with a group fused lasso penalty. *Biometrical Journal* **56** 44–68.
- HORVATH, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology* **14** 3156.

- HORVATH, S., OSHIMA, J., MARTIN, G. M., LU, A. T., QUACH, A., COHEN, H., FELTON, S., MATSUYAMA, M., LOWE, D., KABACIK, S., WILSON, J. G., REINER, A. P., MAIERHOFER, A., FLUNKERT, J., AVIV, A., HOU, L., BACCARELLI, A. A., LI, Y., STEWART, J. D., WHITSEL, E. A., FERRUCCI, L., MATSUYAMA, S. and RAJ, K. (2018). Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging (Albany NY)* **10** 1758–1775.
- HORVÁTH, L. and KOKOSZKA, P. (2012). Functional data structures. In *Inference for Functional Data with Applications*, (L. Horváth and P. Kokoszka, eds.). *Springer Series in Statistics* 1–17. Springer, New York, NY.
- HUANG, R.-C., LILLYCROP, K. A., BEILIN, L. J., GODFREY, K. M., ANDERSON, D., MORI, T. A., RAUSCHERT, S., CRAIG, J. M., ODDY, W. H., AYONRINDE, O. T., PENNELL, C. E., HOLBROOK, J. D. and MELTON, P. E. (2019). Epigenetic Age Acceleration in Adolescence Associates With BMI, Inflammation, and Risk Score for Middle Age Cardiovascular Disease. *The Journal of Clinical Endocrinology & Metabolism* **104** 3012–3024.
- KANKAANPÄÄ, A., TOLVANEN, A., HEIKKINEN, A., KAPRIO, J., OLLIKAINEN, M. and SILLANPÄÄ, E. (2022). The role of adolescent lifestyle habits in biological aging: A prospective twin study. *eLife* **11** e80729.
- KASS, R. E. and RAFTERY, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association* **90** 773–795.
- MARIONI, R. E., SHAH, S., MCRAE, A. F., RITCHIE, S. J., MUNIZ-TERRERA, G., HARRIS, S. E., GIBSON, J., REDMOND, P., COX, S. R., PATTIE, A., CORLEY, J., TAYLOR, A., MURPHY, L., STARR, J. M., HORVATH, S., VISSCHER, P. M., WRAY, N. R. and DEARY, I. J. (2015). The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *International Journal of Epidemiology* **44** 1388–1396.
- MC EWEN, L. M., O'DONNELL, K. J., MCGILL, M. G., EDGAR, R. D., JONES, M. J., MACISAAC, J. L., LIN, D. T. S., RAMADORI, K., MORIN, A., GLADISH, N., GARG, E., UNTERNAEHRER, E., POKHVISNEVA, I., KARNANI, N., KEE, M. Z. L., KLENGEL, T., ADLER, N. E., BARR, R. G., LETOURNEAU, N., GIESBRECHT, G. F., REYNOLDS, J. N., CZAMARA, D., ARMSTRONG, J. M., ESSEX, M. J., DE WEERTH, C., BEIJERS, R., TOLLENAAR, M. S., BRADLEY, B., JOVANOVIĆ, T., RESSLER, K. J., STEINER, M., ENTRINGER, S., WADHWA, P. D., BUSS, C., BUSH, N. R., BINDER, E. B., BOYCE, W. T., MEANEY, M. J., HORVATH, S. and KOBOR, M. S. (2020). The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. *Proceedings of the National Academy of Sciences* **117** 23329–23335.
- MILLER, A. (2002). *Subset Selection in Regression*, 0 ed. Chapman and Hall/CRC.
- NATARAJAN, B. K. (1995). Sparse Approximate Solutions to Linear Systems. *SIAM Journal on Computing* **24** 227–234.
- NWANAJI-ENWEREM, J. C., LAAN, L. V. D., KOGUT, K., ESKENAZI, B., HOLLAND, N., DEARDORFF, J. and CARDENAS, A. (2021). Maternal adverse childhood experiences before pregnancy are associated with epigenetic aging changes in their children. *Aging* **13** 25653–25669.
- PERNG, W., TAMAYO-ORTIZ, M., TANG, L., SÁNCHEZ, B. N., CANTORAL, A., MEEKER, J. D., DOLINOY, D. C., ROBERTS, E. F., MARTINEZ-MIER, E. A., LAMADRID-FIGUEROA, H., SONG, P. X. K., ETTINGER, A. S., WRIGHT, R., ARORA, M., SCHNAAS, L., WATKINS, D. J., GOODRICH, J. M., GARCIA, R. C., SOLANO-GONZALEZ, M., BAUTISTA-ARREDONDO, L. F., MERCADO-GARCIA, A., HU, H., HERNANDEZ-AVILA, M., TELLEZ-ROJO, M. M. and PETERSON, K. E. (2019). Early Life Exposure in Mexico to ENvironmental Toxicants (ELEMENT) Project. *BMJ Open* **9** e030427.
- QUACH, A., LEVINE, M. E., TANAKA, T., LU, A. T., CHEN, B. H., FERRUCCI, L., RITZ, B., BANDINELLI, S., NEUHOUSER, M. L., BEASLEY, J. M., SNETSELAAR, L., WALLACE, R. B., TSAO, P. S., ABSHER, D., ASSIMES, T. L., STEWART, J. D., LI, Y., HOU, L., BACCARELLI, A. A., WHITSEL, E. A. and HORVATH, S. (2017). Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging (Albany NY)* **9** 419–437.
- RABINER, L. and JUANG, B. (1986). An introduction to hidden Markov models. *IEEE ASSP Magazine* **3** 4–16.
- RAMSAY, J. O. (2004). Functional Data Analysis. In *Encyclopedia of Statistical Sciences* John Wiley & Sons, Ltd _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/0471667196.ess0646>.
- RAMSAY, J. (2005). Functional Data Analysis. In *Encyclopedia of Statistics in Behavioral Science* John Wiley & Sons, Ltd _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/0470013192.bsa239>.
- (2005). Introduction. In *Functional Data Analysis*, (J. O. RAMSAY and B. W. SILVERMAN, eds.). *Springer Series in Statistics* 1–18. Springer, New York, NY.
- REISS, P. T., GOLDSMITH, J., SHANG, H. L. and OGDEN, R. T. (2017). Methods for Scalar-on-Function Regression. *International Statistical Review* **85** 228–249.
- TIBSHIRANI, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58** 267–288.
- TIBSHIRANI, R., SAUNDERS, M., ROSSET, S., ZHU, J. and KNIGHT, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67** 91–108.

- VRIEZE, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). *Psychological Methods* **17** 228–243.
- WANG, W., WU, S., ZHU, Z., ZHOU, L. and SONG, P. X. K. (2022). Supervised Homogeneity Fusion: a Combinatorial Approach.
- WIKLUND, P., KARHUNEN, V., RICHMOND, R. C., PARMAR, P., RODRIGUEZ, A., DE SILVA, M., WIELSCHER, M., REZWAN, F. I., RICHARDSON, T. G., VEIJOLA, J., HERZIG, K.-H., HOLLOWAY, J. W., RELTON, C. L., SEBERT, S. and JÄRVELIN, M.-R. (2019). DNA methylation links prenatal smoking exposure to later life health outcomes in offspring. *Clinical Epigenetics* **11** 97.
- WU, X., CHEN, W., LIN, F., HUANG, Q., ZHONG, J., GAO, H., SONG, Y. and LIANG, H. (2019). DNA methylation profile is a quantitative measure of biological aging in children. *Aging* **11** 10031–10051.
- YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association* **100** 577–590.
- ZHU, Y., SHEN, X. and PAN, W. (2013). Simultaneous Grouping Pursuit and Feature Selection Over an Undirected Graph. *Journal of the American Statistical Association* **108** 713–725.
- ZOU, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association* **101** 1418–1429.