Supervised Learning of Physical Activity Features From Functional Accelerometer Data

Margaret Banker and Peter X. K. Song

Abstract*—*Objective: **We propose a new health informatics framework to analyze physical activity (PA) from accelerometer devices. Accelerometry data enables scientists to extract personal digital features useful for precision health decision making. Existing methods in accelerometry data analysis typically begin with discretizing summary counts by certain fixed cutoffs into activity categories. One well-known limitation is that the chosen cutoffs are often validated under restricted settings, and cannot be generalizable across populations, devices, or studies.** ***Methods:*** **We develop a data-driven approach to overcome this bottleneck in PA data analysis, in which we holistically summarize a subject's activity profile using Occupation-Time curves (OTCs), which describe the percentage of time spent at or above a continuum of activity count levels. We develop multi-step adaptive learning algorithms to perform supervised learning via a scalar-on-function model that involves OTC as the functional predictor of interest as well as other scalar covariates. Our learning analytic first incorporates a hybrid approach of fused lasso for clustering and Hidden Markov Model for changepoint detection, then executes refinement procedures to determine activity windows of interest.** ***Results:*** **We evaluate and illustrate the performance of the proposed learning analytic through simulation experiments and real-world data analyses to assess the influence of PA on biological aging. Our findings indicate a different directional relationship between biological age and PA depending on the specific outcome of interest.** ***Significance:*** **Our bioinformatics methodology involves the biomedical outcome of interest to detect different critical points, and is thus adaptive to the specific data, study population, and health outcome under investigation.**

***Index Terms*—Actigraphy, changepoint detection, epigenetic ageing, fused lasso, scalar-on-function regression.**

I. 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 [1], [2], [3], 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" [4]. 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 $VM = \sqrt{axis1^2 + axis2^2 + axis3^2}$ [3].

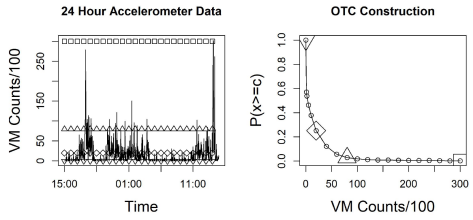
With the ability to provide high-frequency measurements (example frequency: 60 Hz, or 60 data points per second), the technical challenge becomes how to retrieve useful information from this high-frequency time-series data type. A popular method of analyzing accelerometer data involves specifying activity "cutoffs" to discretize activity counts into categories, such as Sedentary, Light, and Moderate-to-Vigorous [5], [6], [7]. Supplementary Fig. S1 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 [2]. 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 [2], [6], [7]. 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) [5]. 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

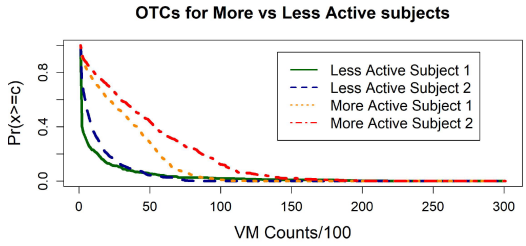
Manuscript received 12 January 2023; revised 6 July 2023; accepted 12 September 2023. Date of publication 22 September 2023; date of current version 6 December 2023. This work was supported by R24ES028502 and NSFDMS2113564. *(Corresponding author: Peter X. K Song.)*

The authors are with the Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI 48109 USA (e-mail: mbanker@umich.edu; pxsong@umich.edu).

Digital Object Identifier 10.1109/JBHI.2023.3318205



(a) OTC Construction



(b) OTC Comparison

Fig. 1. (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.

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 article 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 adapatively 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 activity data involves a summary curve which describes the proportion of time an individual spends at or above successive activity levels [8]. 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 the sequential moving activity levels, and $\mathbb{P}$ denotes an empirical probability measure defined by the proportion: $\frac{duration\ of\ \{t:VM(t)\geq c\}}{total\ duration\ of\ VM(t)}$ [9]. Fig. 1(a) illustrates the construction of an OTC, and Fig. 1(b) illustrates four OTC plots over VM varying from 0 to 30,000. Notably, Fig. 1(b) includes OTCs for more-active people whose curves 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 II 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:

$$Y = <X, \beta> + \epsilon = \int_{\mathcal{C}} X(c)\beta(c)dc + \epsilon, \qquad (1)$$

where Y is a scalar health outcome of interest, $X(c)$ is the functional OTC defined on $\mathcal{C} \subset \mathbb{R} = (-\infty, \infty)$ and ϵ is the error term. Here, $<a, b>$ depicts the inner product of two square-integrable functions, namely $\int_{\mathcal{C}} a(c)b(c)dc$ with $\int_{\mathcal{C}} a^2(c)dc < \infty$ and $\int_{\mathcal{C}} b^2(c)dc < \infty$. 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 activity 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 article is organized as follows. Section II introduces our motivating dataset from the Early Life Exposure in Mexico to ENvironmental Toxicants (ELEMENT) study, where both OTCs and epigenetic age (a scalar outcome that reflects biological aging) are described. Section III and IV concern the development and implementation of a multi-step supervised adaptive learning analytic that enables changepoint detection of important activity ranges, whose performance is evaluated and demonstrated through simulation experiments in Section V. In Section VI, we apply our proposed method to assess the functional association between PA and epigenetic age. Section VII includes a few concluding remarks.

II. 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. The study is discussed in detail in a previously published review article [10]. Briefly, the ELEMENT study consists of three birth cohorts (cohort 1 in 1994–1997, cohort 2 in 1997–2000, and cohort 3 in 2001–2005), with a 2015 follow-up study consisting of some members from the three original groups.

A. ELEMENT Actigraphy Data

As a part of this 2015 ELEMENT follow-up study, 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 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 $\pm 6\, g (g = 9.81\ \text{m/s}^2)$ 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 article, 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.

Besides age and sex, our analysis includes two additional covariates: lead exposure, shown to be associated with biological aging acceleration in children [11], and pubertal status measured by a five-category ordinal variable of Tanner staging.

B. 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 preemptive health interventions [12]. Epigenetic Age is one such quantitative way to represent a person's biological aging, and is the outcome of interest in this article. Epigenetic Age Calculators hosted online [13] 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.

Studies show that children and adolescents (age 0–18) undergo the fastest and most dynamic rate of growth and DNAm changes [11], [14]. Furthermore, research in fetal origins of adult disorders suggest that childhood experiences and exposures could form the foundation of health issues experienced by individuals in adulthood and beyond. 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 Fig. 2).

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) [15] and Levine's Phenotypic Age clock (DNAm PhenoAge) [16], both calculated from the epigenetic age calculators [13]. The former reflects aging in skin and blood cells while the latter pertains to age-related disease/phenotypes.

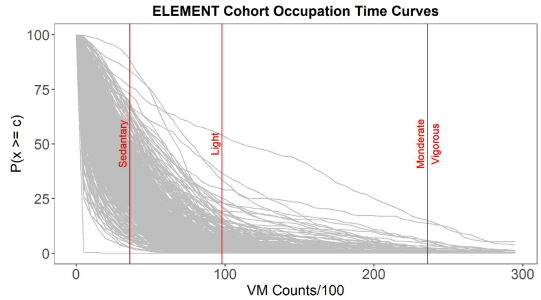


Fig. 2. OTCs for 354 ELEMENT subjects, representing functional covariates. The vertical lines represent Chandler's cutoffs (Chandler et al., 2016) for activity levels (Sedentary, Light, Moderate, Vigorous). The relative shape of these OTCs reflect the subject's activity profile.

C. 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. 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. Fig. 1(a) 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 article.

As shown in Fig. 1(b), 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. Fig. 2 illustrates OTC variability in its representation of 354 OTCs from ELEMENT.

III. Method

In this article 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.

A. 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 (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 $\int_{\mathcal{C}} \beta(c)X(c)dc$ can be approximated by a step-function over many small pieces. That is, we divide interval $\mathcal{C}$ into J-many small successive intervals with a grid $c_0 = 0, c_1, \ldots, c_J = 30,000$, namely $\mathcal{C} = [0, c_1] \cup_{j=2}^{J} (c_{j-1}, c_j]$, and assume $\beta(c)$ takes one parameter within one small interval. Precisely, on the jth interval $(c_{j-1}, c_j]$, we set constant parameters β_j, $j = \{1, \ldots, J\}$ resulting in $\beta(c) \approx \sum_{j=1}^{J} \beta_j I(X(c) \in (c_{j-1}, c_j])$. Consequently, we have

$$\int_{\mathcal{C}} \beta(c)X(c)\,dc = \sum_{j=1}^{J} \int_{c_{j-1}}^{c_j} \beta(c)X(c)\,dc \approx \sum_{j=1}^{J} \beta_j \int_{c_{j-1}}^{c_j} X(c)\,dc := \sum \beta_j A_j, \quad (2)$$

where A_j denotes the Area Under the Curve (AUC) over interval $(c_{j-1}, c_j]$ or $A_j = \int_{c_{j-1}}^{c_j} X(c)\,dc$. In preparation of regularized estimation, we normalize individual $A_j's$ to mean 0 and variance 1. 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 (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 [17] and Hidden Markov Model (HMM) [18] are popular. However, these existent approaches do not perform well due to the high correlation of A_j's. As shown in Supplementary Fig. S2 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.

B. 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 $\beta(c)$, while strategies 3 and 4 aim to group the initial β estimates to form activity windows whose edges determine jump points. Algorithm 1 outlines these procedures, with further detail in the FRACT Implementation Section IV.

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 tradeoff between pair-wise correlation levels and minimal loss of signal strength. Our experience from simulation experiments suggests that selecting a J-partition of $\mathcal{C}$ 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 Supplementary Fig. S2. 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 (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 [19]. 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 IV.

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 IV.

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.

IV. FRACT Implementation

A. 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 $\hat{\beta}_1, \ldots, \hat{\beta}_J$ 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 $\hat{\beta}_j$.

B. 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"

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 $c_0 = 0, c_1, \ldots, c_J$, calculate the AUCs $A_j = \int_{c_{j-1}}^{c_j} X(c)\, dc$, $j \in 1, \ldots, J$
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, $j \in 1, \ldots, J$
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**

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.

1: **procedure** CALCULATE INITIAL PROBABILITY MATRIX(3×3 matrix)
2: **for** k = (1, 2) **do**
3: Prob. of remaining in Group k $= p_{kk} = \frac{J_k - 1}{J_k}$
4: Prob. of transitioning from Group k to Group $k+1 = p_{k,k+1} = \frac{1}{J_k}$
5: Note: $p_{kk} + p_{k,k+1} = 1$
6: **end for**
7: Prob. of transitioning from Group $3 = p_{33} = 1$
8: 3×3 **initial probability matrix:**

$$\begin{bmatrix} p_{11} & p_{12} & 0 \\ 0 & p_{22} & p_{23} \\ 0 & 0 & 1 \end{bmatrix}$$

9: **end procedure**

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 $c_1, \cdots c_K$:

$$Y = \beta_1 \tilde{A}_1 + \beta_2 \tilde{A}_2 + \cdots + \beta_K \tilde{A}_K + Z^T \alpha + \epsilon \qquad (3)$$

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 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.

C. 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.

Algorithm 3: Implementation: Edge-Swapping and Merging.

Input: $K \geq 2$ number of intervals, with cutpoints (c_{k-1}, c_k), and estimates $\hat{\beta}'_k s$, $k = 1, \ldots, K$

Calculate $k_{\min} = argmin_k |\hat{\beta}_{k+1} - \hat{\beta}_k|, k \in \{1, 2, \ldots, K-1\}$ 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 (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

We suppose that cutoffs $c_1, \ldots, c_K$ and estimates $\hat{\beta}_1, \ldots, \hat{\beta}_K$ are available from previous steps.

Edge-Swapping and Merging Technique: Algorithm 3 aims to refine the cutoffs $c_1, \ldots, c_K$ 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: $k_{\min} = argmin_k |\hat{\beta}_{k+1} - \hat{\beta}_k|$, $k \in \{1, 2, \ldots, K-1\}$. 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 $\bar{A}_j, j = k_{\min}, k_{\min} + 1$. 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 (3) is compared to make model selection and thus further select a desirable partition.

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 (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 break-points. 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 (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.

V. 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 $(c_1, \ldots, c_K)$ and effect sizes $(\beta_1, \ldots, \beta_K)$.

A. 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 549 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 549 10-minute candidate segments. To ensure that the variability in the simulated OTC curves reflected the variability in the motivating dataset (Fig. 2), 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 II-C.

Given true cutoffs $c_0^* = 0, c_1^*, \ldots, c_{K^*-1}^*, c_{K^*} = 30,000$, the successive integrals over $\mathcal{C}$ were specified, and the K^*-element

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 $\hat{\beta}_k's$, $K \geq 2$, $k \in \{1, \ldots, K\}$, $J_1 + \cdots + J_K = J$

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

procedure EXPLORATION STEP
Set: s = total number of break locations
  for $k \in (1, \ldots, K)$ **do**
    Calculate $s_k = J_k/J \times s$ = proportional number of potential breakpoints s allocated to Window k, (c_{k-1}, c_k)
    $\mathcal{B}_k = \{b_1, \ldots, b_{s_k}\}$ = 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 (3) with new partitions
    return $EBIC(K+1, b)$ = EBIC of $(K+1)$-group model at break b, $b = b_1, \ldots, b_{s_k}$
  end for
  if $argmin_{b \in \mathcal{B}} EBIC(K+1, b) < EBIC(K)$, with $\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_K\}$ 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

vector of AUCs, $(A_1^*, \ldots, A_{K^*}^*)^T$, 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 $\beta_1^*, \ldots, \beta_K^*$ and α^*, where single continuous covariate $Z \sim N(0,1)$ and $\epsilon \sim N(0, 10)$. The 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 I.

TABLE I
SIMULATION RESULTS OF 3-GROUP MODEL TO EVALUATE FRACT

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%		

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.

B. 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 I, 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 $\hat{\beta}_j$, $j = 1, \ldots, J$ as shown in Supplementary Fig. S2. By merging every five successive $A_j's$, the augmented $\breve{A}_j$ variables gave reduced mean pairwise correlations $cor(\breve{A}_j, \breve{A}_{j+1}) < 0.98$, $cor(\breve{A}_j, \breve{A}_{j+5}) <$

TABLE II
SENSITIVITY SIMULATION RESULTS FOR 3-GROUP MODEL

	$\beta = (4, 0, -4)$			$\beta = (2, 0, -2)$			$\beta = (4, 0, -2)$		
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

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 Sect. III.

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 Supplementary Fig. S2, 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 II). 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 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 $c_1^*, \ldots, c_K^*$ cutpoints.

VI. DATA ANALYSIS

We now apply the proposed FRACT methodology to investigate the functional association between PA and epigenetic age. As discussed in Section II, 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. Fig. 2 shows OTCs for the 354 subjects, representing PA profiles during weekends between 4:00 PM–10:00 PM, selected with the rationale that this period reflects a block of time they have more control over their activities. Among multiple epigenetic age clocks [13] we are interested in two specific ones: Horvath's AgeSkinBlood Clock [15], and Levine's PhenoAge Clock [16]. 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 III 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 III suggests that FRACT determined a window of activity VM $\leq$ 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 $\leq$ 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 Table III(a)–(c), there are agreements between our detected cutoffs and the previously published Crouter [6] and Chandler [7] 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 $\leq$ 1500 is similar to Crouter's cutpoint for Sedentary behavior (VM $\leq$ 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 Table III(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"

TABLE III
DATA ANALYSIS RESULTS

(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

(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".

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 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 III is stable.

VII. DISCUSSION AND CONCLUSION

In this article we developed FRACT, the Functional Regularized Adaptive Changepoint-detection Technique, 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 of FRACT's supervised learning: unlike methods discretizing functional features via existing cutoffs regardless of specific outcomes, FRACT offers a simultaneous operation of supervised changepoint detection and functional association parameter estimation. Thus, it 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/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 may greatly 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 [16], and the direction of association of increased PA and lower epigenetic age makes intuitive sense. On the other hand, when the AgeSkinBlood clock was considered, which focuses on epigenetic variables in

skin/blood cells including fibroblasts that deal with the structural components of skin cells, we detected the benefit of sedentary PA. This 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. 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 III, 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 pulse. If using a device with these additional variables, FRACT could either utilize a new functional covariate or include as covariates of interest. However, these devices, while useful, are more expensive and can be barriers in scientific study and personal use. Similarly, as the proposed methodology considers a summary of activity over a pre-defined time window, it does not consider the transition from aerobic to anaerobic activity stage. However, if one can leverage data from the multi-sensor devices to classify the activity data into these stages, relative time in each activity stage could be considered as covariate in analysis. 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 [3], [20]. Lastly, some studies that aim to classify PA will validate their method with in-lab studies, aligning wearable device readings with observed physical activity intensities [21], [22]. 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 [23]. 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 [24]. 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. Our methodology considers PA data in a seven-day period; with repeated measurements we can study how changing PA patterns from late-adolescence into early adulthood affects longitudinal health outcomes. While this article 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.

ACKNOWLEDGMENT

The authors would like to thank Dr. K Peterson and L Arboleda Merino for access to ELEMENT data, as well as Dr. A Pitchford for guidance on PA accelerometer data.

REFERENCES

[1] S. G. Trost, "Objective measurement of physical activity in youth: Current issues, future directions," *Exercise Sport Sci. Rev.*, vol. 29, no. 1, pp. 32–36, Jan. 2001.

[2] R. P. Troiano et al., "Evolution of accelerometer methods for physical activity research," *Brit. J. Sports Med.*, vol. 48, no. 13, pp. 1019–1023, 2014.

[3] F. Liu, A. A. Wanigatunga, and J. A. Schrack, "Assessment of physical activity in adults using wrist accelerometers," *Epidemiologic Rev.*, vol. 43, no. 1, pp. 65–93, Dec. 2021.

[4] K. Y. Chen and D. Bassett, "The technology of accelerometry-based activity monitors: Current and future," *Med. Sci. Sports Exercise*, vol. 37, no. 11 Suppl, pp. S490–500, Nov. 2005.

[5] P. Freedson, D. Pober, and K. F. Janz, "Calibration of accelerometer output for children," *Med. Sci. Sports Exercise*, vol. 37, no. 11 Suppl, pp. S523–530, Nov. 2005.

[6] S. E. Crouter, J. I. Flynn, and D. R. Bassett, "Estimating physical activity in youth using a wrist accelerometer," *Med. Sci. Sports Exercise*, vol. 47, no. 5, pp. 944–951, May 2015.

[7] J. L. Chandler et al., "Classification of physical activity intensities using a wrist-worn accelerometer in 8–12-year-old children," *Pediatr. Obesity*, vol. 11, no. 2, pp. 120–127, 2016.

[8] L. Bogachev and N. Ratanov, "Occupation time distributions for the telegraph process," *Stochastic Processes Their Appl.*, vol. 121, no. 8, pp. 1816–1844, Aug. 2011.
[9] H.-W. Chang and I. W. McKeague, "Empirical likelihood-based inference for functional means with application to wearable device data," *J. Roy. Stat. Soc. Ser. B*, vol. 84, no. 5, pp. 1947–1968, 2022.
[10] W. Perng et al., "Early life exposure in Mexico to environmental toxicants (ELEMENT) project," *BMJ Open*, vol. 9, no. 8, Aug. 2019, Art. no. e030427.
[11] X. Wu et al., "DNA methylation profile is a quantitative measure of biological aging in children," *Aging*, vol. 11, no. 22, pp. 10031–10051, Nov. 2019.
[12] R. E. Marioni et al., "The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936," *Int. J. Epidemiol.*, vol. 44, no. 4, pp. 1388–1396, Aug. 2015.
[13] S. Horvath, "DNA methylation age of human tissues and cell types," *Genome Biol.*, vol. 14, no. 10, Dec. 2013, Art. no. 3156.
[14] L. M. McEwen et al., "The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells," in *Proc. Nat. Acad. Sci. USA*, vol. 117, no. 38, pp. 23329–23335, Sep. 2020.
[15] S. Horvath et al., "Epigenetic clock for skin and blood cells applied to hutchinson gilford progeria syndrome and ex vivo studies," *Aging (Albany NY)*, vol. 10, no. 7, pp. 1758–1775, Jul. 2018.
[16] M. E. Levine et al., "An epigenetic biomarker of aging for lifespan and healthspan," *Aging (Albany NY)*, vol. 10, no. 4, pp. 573–591, Apr. 2018.
[17] R. Tibshirani et al., "Sparsity and smoothness via the fused lasso," *J. Roy. Stat. Soc.: Ser. B. (Statist. Methodol.)*, vol. 67, no. 1, pp. 91–108, 2005.
[18] L. Rabiner and B. Juang, "An introduction to hidden Markov models," *IEEE ASSP Mag.*, vol. 3, no. 1, pp. 4–16, Jan. 1986.
[19] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Statist.*, vol. 38, no. 2, pp. 894–942, Apr. 2010.
[20] Z. Gao et al., "The dilemma of analyzing physical activity and sedentary behavior with wrist accelerometer data: Challenges and opportunities," *J. Clin. Med.*, vol. 10, no. 24, Jan. 2021, Art. no. 5951.
[21] M. Sevil et al., "Determining physical activity characteristics from wristband data for use in automated insulin delivery systems," *IEEE Sensors J.*, vol. 20, no. 21, pp. 12859–12870, Nov. 2020.
[22] C. M. v. Loo et al., "Wrist accelerometer cut-points for classifying sedentary behavior in children," *Med. Sci. Sports Exercise*, vol. 49, no. 4, pp. 813–822, Apr. 2017.
[23] O. Minaeva et al., "Level and timing of physical activity during normal daily life in depressed and non-depressed individuals," *Transl. Psychiatry*, vol. 10, no. 1, pp. 1–11, Jul. 2020.
[24] A. Garriga et al., "Impact of seasonality on physical activity: A systematic review," *Int. J. Environ. Res. Public Health*, vol. 19, no. 1, 2021, Art. no. 2.