**The predictive performance of objective measures of physical activity derived from accelerometry data for 5-year all-cause mortality in older adults: NHANES 2003-2006**

**Supplementary Materials**

Ekaterina Smirnova PhD1,\*, Andrew Leroux MS2, Quy Cao MS3, Lucia Tabacu PhD4, Vadim Zipunnikov PhD2, Ciprian Crainiceanu PhD2 , Jacek Urbanek PhD5.

1 Department of Biostatistics, School of Medicine, Virginia Commonwealth University, Richmond, VA

2 Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD

3 Department of Mathematical Sciences, College of Humanities and Sciences, University of Montana, Missoula, MT

4Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA

5Department of Geriatric Medicine and Gerontology, Johns Hopkins University, Baltimore, MD

**\*Corresponding author**. Email: [ekaterina.smirnova@vcuhealth.org](mailto:ekaterina.smirnova@vcuhealth.org)

**METHODS**

*Accelerometry derived predictors*

According to the NHANES protocol, the minute-by-minute activity data was recorded using a hip-worn ActiGraph AM-7164 (formerly the CSA/MTI AM-7164) accelerometer. Each participant was instructed to wear the device for a period of 7 consecutive days from the day of NHANES examination and remove the device during sleep and water-related activity, such as swimming and bathing. The device was returned to the CDC by mail in postage-paid padded envelopes. Not every study participant wore the device for the full 7-day period.

The high volume of minute-level activity measurements is challenging, which is why the current practice is to take summary measures. Popular PA summaries based on actigraphy include: (1) total activity count (TAC); (2) total log(1+activity count), referred to as total log activity count (TLAC); and (3) total minutes of moderate/vigorous physical activity (MVPA), where MVPA is defined as the total time with more than 2020 counts per minute. While informative, these summaries do not reflect the full complexity of daily activity patterns and may miss important information that could be associated with health and functional status. To evaluate the effect of daily PA patterns on mortality we introduce 12 additional summary variables (TLAC 12AM-2AM, TLAC 2AM-4AM, …, TLAC 10PM-12AM), where each variable corresponds to the total log(1+activity count) in a 2-hour interval. For example, TLAC 12AM-2AM is the total log activity between 12AM and 2AM. We also used two measures of activity fragmentation: transition probabilities from sedentary to active (SATP) and active to sedentary (ASTP) (1). The sedentary to active transition probability (SATP) is defined as , where is the total number of sedentary bouts (periods where the activity count is less tan 100) and is the total sedentary time (2). SATP is inversely proportional to the average length of the inactivity bout, but has better statistical properties (e.g., symmetric distribution with normal tails across study participants)

Similarly, the active to sedentary transition probability (ASTP) is defined as , where is the total number of active bouts and is the total activity time. Larger values of transition probabilities correspond to shorter average bout duration and more frequent switching between states and more fragmented PA. In NHANES there is substantial heterogeneity in non-wear time (due to study protocol and non-compliance), which makes it difficult to differentiate between sleep, sedentary, and non-wear time. Therefore, we consider transitions between active and the combination of sedentary/sleep/non-wear periods, which we denote as ASTPsl/nw and SATPsl/nw, respectively. We also include the total accelerometer wear time (Wear Time in minutes) and total sedentary/sleep/non-wear time. Each participant had 3 – 7 days of available accelerometry data. For each accelerometry derived summary measure we calculated the measure for each day (i.e., TAC\_day1, TAC\_day2, …, TAC\_day7) and then averaged across available days.

We also propose to use principal component analysis (PCA) to derive additional predictors. This is aprincipled, widely accepted, and fast approach to addressing whether something has been missed by simple summaries of the data. PCs were obtained as follows: (1) transform minute-level activity count data as *x* → log (1 + *x*), where *x* is the minute level activity count; (2) arrange all activity trajectories into a 18373 by 1440 dimensional matrix, X, where each row corresponds to a subject/day and each column corresponds to a specific minute (time) of the day; (3) conduct functional PCA (fPCA) on the matrix X using the **fpca.face()** function(3) in the **refund** package(4) in **R**; (4) retain the first 6 PCs, which explain ~57% variability in the activity data; and (5) obtain the score for each day on each PC and calculate the mean and standard deviation of these scores for each subject across days.To see what PCA-derived variables to retain we started with a model containing the standard demographic, behavioral and comorbidity variables and conducted forward selection on the means and standard deviations of the scores on the 6 PCs (a total of 12 variables).Using this procedure,the average scores for the first (mi1; OR= 1.014, CI: (1.004, 1.026); p = 0.008), and the standard deviation of the sixth PCs (si6; OR = 0.926, CI: (0.888, 0.965); p < 0.001) were found to be statistically associated with 5-year all-cause mortality.

PCA is a widely used, but is often criticized for the lack of intuition and transportability potential across studies. To overcome this problem, we have inspected the PCs and replaced them with surrogate variables that are intuitive and can be calculated directly from the observed data. These surrogate measures are summarized in Table 3 in the **rnhanesdata** vignette (5). To ensure full reproducibility of the calculation of all PA summary statistics, the supplementary R vignette available as a part of **rnhanesdata** package contains all the necessary information.

REFERENCES

1. Di J, Leroux A, Urbanek J, et al. Patterns of sedentary and active time accumulation are associated with mortality in US adults: The NHANES study. *bioRxiv*. 2017; 182337. doi**:** 10.1101/182337.
2. Troiano RP, Berrigan D, Dodd KW, et al. Physical activity in the United States measured by accelerometer. *Medicine and science in sports and exercise.* 2008; 40(1): 181. doi: [10.1249/mss.0b013e31815a51b3](https://doi.org/10.1249/mss.0b013e31815a51b3)
3. Xiao L, Zipunnikov V, Ruppert D, et al. Fast covariance estimation for high-dimensional functional data. *Statistics and Computing.* 2016; 26(1): 409–421. doi: [10.1007/s11222-014-9485-x](https://doi.org/10.1007/s11222-014-9485-x)
4. Crainiceanu CM, Reiss PT, Goldsmith JA, et al. refund: regression with functional data R package version 0.1-6. 2012; Accessed June 1, 2018.
5. Leroux A, Crainiceanu C, Smirnova E, Cao Q.: rnhanesdata: NHANES Accelerometry Data Pipeline. [https://github.com/andrew-leroux/rnhanesdata.](https://github.com/andrew-leroux/rnhanesdata) R package version 1.0. Accessed June 1, 2018.