Advance Access publication September 30, 2020

OXFORD

Research Article

# Quantifying the Predictive Performance of Objectively Measured Physical Activity on Mortality in the UK Biobank

Andrew Leroux, PhD,<sup>1,3,\*</sup> Shiyao Xu, ScM,<sup>1</sup> Prosenjit Kundu, PhD,<sup>1</sup> John Muschelli, PhD,<sup>1</sup> Ekaterina Smirnova, PhD,<sup>2</sup> Nilanjan Chatterjee, PhD,<sup>1</sup> and Ciprian Crainiceanu, PhD<sup>1</sup>

<sup>1</sup>Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland. <sup>2</sup>Department of Biostatistics, School of Medicine, Virginia Commonwealth University, Richmond. <sup>3</sup>Department of Biostatistics & Informatics, Colorado School of Public Health, University of Colorado, Aurora.

\*Address correspondence to: Andrew Leroux, ScM, Department of Biostatistics & Informatics, Colorado School of Public Health, Fitzsimons Building, 4th Floor, 13001 E. 17th Place, Mail Stop B119, Aurora, CO 80045. E-mail: andrew.leroux@cuanschutz.edu

Received: April 12, 2020; Editorial Decision Date: September 16, 2020

Decision Editor: Anne B. Newman, MD, MPH, FGSA

#### **Abstract**

**Background:** Objective measures of physical activity (PA) derived from wrist-worn accelerometers are compared with traditional risk factors in terms of mortality prediction performance in the UK Biobank.

**Method:** A subset of participants in the UK Biobank study wore a tri-axial wrist-worn accelerometer in a free-living environment for up to 7 days. A total of 82 304 individuals over the age of 50 (439 707 person-years of follow-up, 1959 deaths) had both accelerometry data that met specified quality criteria and complete data on a set of traditional mortality risk factors. Predictive performance was assessed using cross-validated Concordance (*C*) for Cox regression models. Forward selection was used to obtain a set of best predictors of mortality.

Results: In univariate Cox regression, age was the best predictor of all-cause mortality (C = 0.681) followed by 12 PA predictors, led by minutes of moderate-to-vigorous PA (C = 0.661) and total acceleration (C = 0.661). Overall, 16 of the top 20 predictors were objective PA measures (C = 0.578-0.661). Using a threshold of 0.001 improvement in Concordance, the Concordance for the best model that did not include PA measures was 0.735 (9 covariates) compared with 0.748 (12 covariates) for the best model with PA variables (p-value < .001).

Conclusions: Objective measures of PA derived from accelerometry outperform traditional predictors of all-cause mortality in the UK Biobank except age and substantially improve the prediction performance of mortality models based on traditional risk factors. Results confirm and complement previous findings in the National Health and Nutrition Examination Survey (NHANES).

Keywords: Exercise, Longevity, Physical activity

Objective physical activity (PA) measurements obtained from accelerometry have been identified as major mortality risk factors (1–8). The recent shift in measuring PA from using questionnaires to accelerometers in health studies is due to the substantial recall, social desirability, and cognitive and psycho-social biases (9,10) of PA questionnaires. Importantly, PA is one of the few known risk factors for mortality that is modifiable for a large number of individuals. Studying objectively measured PA in large cohorts in free-living conditions is crucial to the development of public health guidelines surrounding the amount and frequency of PA.

Recently, Smirnova et al. (8) showed that objective PA measures obtained from hip-worn accelerometers are among the strongest predictors of 5-year all-cause mortality in the National Health and

Nutrition Examination Survey (NHANES) 2003–2004 and 2005–2006, a nationally representative subset of the U.S. population. The authors included 2978 participants over the age of 50 who wore a hip-worn accelerometer in a free-living environment for up to 7 days. The number of all-cause mortality events was 297 in the first 5 years of follow-up.

Here, we investigate the individual, relative, and combined predictive performance of objective PA measures obtained from wristworn accelerometers for time to death in the UK Biobank study (11). A subset of participants in the UK Biobank study wore a tri-axial wrist-worn accelerometer in a free-living environment for up to 7 days. A total of 82 304 individuals over the age of 50 (439 707 person-years of follow-up, 1959 deaths) had both accelerometry data

that met specified quality criteria and complete data on a set of traditional mortality risk factors. There are many differences between the 2 studies including the sampling design, type and placement of the device, unit of measurement, wear protocol, and some variable definitions and collection timing. Moreover, the analysis conducted in NHANES was a logistic regression with 5-year mortality as outcome, whereas the analysis of the UK Biobank data is based on Cox models with time to mortality as outcome. This was done because, as of writing, the follow-up time from the accelerometry substudy for a sizable number of participants in the UK Biobank is less than 5 years (25 300) and there is a high degree of heterogeneity in censoring times, though all participants have at least 4 years of follow-up.

In spite of the many differences between NHANES and UK Biobank data and analytic procedures, the key results are remarkably similar and contribute to the body of evidence suggesting that objective measures of PA obtained from accelerometers are among the strongest predictors of mortality.

#### Method

#### Study Population

The UK Biobank is a large prospective cohort study which enrolled more than 500 000 individuals aged from 37 to 70 years with approximately 88% having British ancestry. The UK Biobank collected an exceptional breadth and depth of information on various factors including sociodemographic, lifestyle, environment, accelerometry, imaging, and genetics (11–13). The current research is conducted under the UK Biobank Resource Application 17712 and the mortality data are obtained from death registries released in the Data Showcase on June 13, 2020.

## UK Biobank: Accelerometry Substudy

A subset of UK Biobank participants was invited to participate in the accelerometry substudy, where they were asked to wear a triaxial wrist-worn accelerometer for up to 7 days. Of the 502 536 participants in the UK Biobank study who had not withdrawn consent as of October 16, 2018, 236 488 individuals were invited to participate in the accelerometer substudy. Invitations were sent out on a rolling basis starting in May 2013 on 112 unique dates, which were generally 7 days apart in batches of either 2000 or 3000 invitations. Individuals received these invitations an average (SD) of 5.65 (1.10) years after their baseline visits, with 3022 individuals missing data on date of invitation to the substudy. Thus, accelerometry data and baseline characteristics are not collected at the same time. Of the 236 488 participants who were invited to participate in the accelerometer substudy, 132 788 either refused or did not respond to the invitation, resulting in 103 700 individuals who had at least some accelerometer data. According to the UK Biobank data processing procedure (11), 7164 of these participants were identified as having poor quality data, not enough time wearing the device to estimate an average activity profile, or uncalibrated devices. This resulted in a total of 96 536 participants with "good" accelerometer data as determined by the criteria provided by the UK Biobank team. Individuals who were invited were, on average, younger and self-reported better overall health, lower incidence of comorbidities, lower rates of cigarette smoking, and had a lower body mass index (BMI). Individuals who were invited but declined or did not respond had higher rates of obesity, being overweight, cigarette smoking, worse self-reported health, and worse rates of diabetes (for more details, see Supplementary Material). Despite

these differences, a recent study has shown that the healthy volunteer effect does not distort risk factor association for most major causes of mortality compared to what have been observed for more representative studies (14).

For our analysis, we applied additional exclusion criteria. Specifically, we excluded study participants younger than 50, who have less than 3 calendar days of accelerometry data with an estimated wear time of at least 1368 minutes (95% of the day), or do not have complete data for all demographic, lifestyle, health, and physical function variables presented in Table 1. The variables

**Table 1.** Population Characteristics of UK Biobank Participants Who Had Any Accelerometer Data and Were Aged 50 or Older at Accelerometer Wear Time Stratified by Whether They Met the Inclusion Criteria ("Included") or Were Excluded From Our Analysis ("Excluded")

	Included	Excluded	
	N = 82 304	N = 12 562	
Age at recruitment	57.98 (6.83)	57.33 (7.06)	
Age at accelerometer wear	63.69 (6.81)	63.16 (6.99)	
Race (% non-white)	2155 (2.6)	513 (4.2)	
Body mass index (%)			
Underweight	450 (0.5)	77 (0.6)	
Normal	31 591 (38.4)	4437 (36.0)	
Overweight	34 373 (41.8)	5043 (40.9)	
Obese	15 890 (19.3)	2784 (22.6)	
Self-reported overall health (%)			
Good	49 606 (60.3)	7139 (57.9)	
Excellent	18 230 (22.1)	2189 (17.7)	
Fair	12 478 (15.2)	2529 (20.5)	
Poor	1990 (2.4)	480 (3.9)	
Cigarette smoker (%)	, ,	, ,	
Never	46 603 (56.6)	6740 (54.8)	
Previous	30 318 (36.8)	4559 (37.0)	
Current	5383 (6.5)	1008 (8.2)	
Alcohol (%)	( , , , , , , , , , , , , , , , , , , ,	,	
Daily or almost daily	19 625 (23.8)	2917 (23.4)	
3–4 times a week	21 674 (26.3)	3125 (25.0)	
Once or twice a week	20 254 (24.6)	3057 (24.5)	
1–3 times a month	8585 (10.4)	1311 (10.5)	
Special occasions only	7600 (9.2)	1281 (10.3)	
Never	4566 (5.5)	794 (6.4)	
Heart attack (% yes)	1357 (1.6)	206 (1.7)	
Stroke (% yes)	805 (1.0)	181 (1.5)	
High blood pressure (% yes)	19 762 (24.0)	3157 (25.4)	
Diabetes (% yes)	2965 (3.6)	485 (3.9)	
Cancer (% yes)	6304 (7.7)	1070 (8.7)	
Right-hand grip strength	32.15 (10.94)	31.38 (11.11)	
Gait speed (%)	, , ,	,	
Steady average pace	40 023 (48.6)	6066 (49.5)	
Slow pace	3883 (4.7)	799 (6.5)	
Brisk pace	38 398 (46.7)	5395 (44.0)	
Illness or injury in the past	6657 (8.1)	1224 (10.4)	
2 years (% yes)			
Longstanding illness or	23 835 (29.0)	3548 (33.2)	
disability (% yes)	()	()	
Mortality (% deceased)	1959 (2.4)	366 (2.9)	

*Note*: For continuous variables, group-specific averages and *SD*s (in parentheses) are shown. For categorical variables, the number of individuals within each category and percentages (in parentheses) are shown. Exclusion was based on missing data in any variable while summaries in the "Excluded" column are based on all available data for each variable.

presented in Table 1 were chosen based on either their known association with mortality or their being highly predictive of 5-year mortality in the whole UK Biobank study (15). Of the 103 700 individuals with at least some accelerometry data, 8834 participants were excluded from the analysis based on age and an additional 12 562 participants based on insufficient and/or missing data. Of those individuals excluded for a reason other than age, roughly 70% (N = 8760) were excluded for insufficient and/or poor quality accelerometry data. Thus, the final analytic sample size was 82 304 individuals with a combined 258 364 personyears of follow-up time (average 5.4 years, maximum 6.8 years of follow-up time). We present a flowchart detailing the exclusion criteria in Supplementary Material. Table 1 compares summary statistics for a sampling of demographic, lifestyle, and physical function variables for the individuals who were over age 50 at the time of accelerometer wear stratified by their inclusion status. Study participants included in the analysis tended to be healthier in terms of self-reported health, rates of comorbidity, and lifestyle factors despite being, on average, slightly older than the excluded group (63.7 vs 63.2 years old). The mortality rate was also higher among the excluded participants. With the exception of a pilot assessment center location (Stockport), all recruitment centers are represented in our analytic sample in roughly equal proportions to the whole UK Biobank cohort.

#### Variables/Measures

### Traditional mortality predictors

The UK Biobank collects a large number of sociodemographic, lifestyle, and comorbidity variables which are obtained at a baseline visit via touchscreen and in-person questionnaires. We focus on a large number of established predictors of mortality, common confounders in epidemiological studies, and the top 5 predictors of 5-year all-cause mortality among all UK Biobank participants identified by Ganna and Ingelsson (15). Specifically, we consider race (white vs non-white), BMI (categorized into underweight, normal, overweight, and obese), self-reported overall health (poor, fair, good, and excellent), cigarette smoking (current, previous, and never), alcohol consumption (daily or almost daily, 3–4 times a week, once or twice a week, 1-3 times a month, special occasions only, and never), heart attack, stroke, high blood pressure, diabetes, cancer, right-hand grip strength, self-reported gait speed (slow pace, steady average pace, and brisk pace), illness or injury to self within the past 2 years, and longstanding illness or disability.

#### Accelerometry-derived predictors

The UK Biobank provides accelerometry data at multiple resolutions including: (i) raw tri-axial acceleration signals; (ii) average acceleration (milli-gs, or thousands of earth gravitation units, g=9.81 m/s²) in 5-second epochs; and (iii) subject-level summaries. We use data summarized in 5-second intervals, which is sufficient for distinguishing between sedentary, light, moderate, and vigorous behaviors (16). The 5-second summary measure provided by the UK Biobank is a modified version of the Euclidean norm minus one (ENMO) (17). We consider several different, but correlated, summaries of PA obtained from the UK Biobank ENMO data at the 5-second resolution.

The subject-specific UK Biobank 5-second level data files contain 5-second average ENMO and an indicator of whether that 5-second ENMO was imputed. Imputation was performed whenever an

individual was determined to not be wearing the device (see Doherty et al. (11) for more details). Data were transformed into the 1440+ memory efficient analytic format suggested by Leroux et al. (18). Even with this reformatting, the full data are very large when loaded into working memory. At over 400 Gigabytes of RAM, the computational cost for analyzing such data becomes challenging even on high performance computing clusters. Thus, data were further compressed by averaging the 5-second level data within every minute. Using the 1-minute aggregated data, the following exclusion criteria were applied: (i) excluding participants identified by UK Biobank as not having enough usable data, or having data obtained from uncalibrated devices, or uncalibrated data; (ii) excluding days with less than 95% estimated wear time; and (iii) excluding participants with less than 3 days of accelerometry data that met the previous 2 criteria. Total daily wear time was estimated based on the missing data imputation indicators. Supplementary Material contains code that provide a reproducible description of all exclusion and data processing steps.

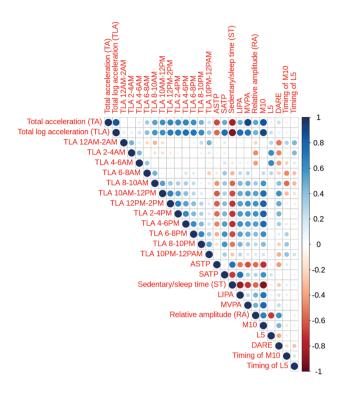
After aggregating data at the minute level, the accelerometry data were further summarized in the following features: (i) total acceleration (TA); (ii) total log(1 + acceleration), labeled total log acceleration (TLA); (iii) TLA in 2-hour windows; (iv) total sedentary time (ST), where sedentary is defined for each minute if the average milli-g in a particular minute is below a given threshold; and (v) total minutes of moderate/vigorous physical activity (MVPA), where moderate to vigorous is defined for each minute if the average milli-g is above another given threshold. The choice of these thresholds is subject to some debate and we provide our choices in the next paragraph. In addition to the commonly used features, we also considered 2 measures of fragmentation (8,19): sedentary to active transition probability (SATP) and active to sedentary transition probability (ASTP). An additional 6 features associated with circadian rhythms were considered: average log acceleration during the most active 10 hours (M10) and least active 5 hours (L5), the timing of L5 and M10, the relative amplitude (RA) defined as RA = (M10 - L5)/(M10 + L5), and the daytime activity ratio estimate (DARE) defined as the TLA between 8 AM and 8 PM divided by the TLA during the day (20). The exact formulas used to derive each of these features are presented in Supplementary Table 1.

For this analysis, sleep time was not estimated. Thus, ST and measures derived from it (ASTP, SATP), combine sleep and sedentary behaviors. This avoids the non-trivial problem of estimating sleep, has a very clear technical definition, but combines sleep periods with sedentary while awake periods. A sedentary minute for a wrist-worn accelerometer is defined as a minute with an average ENMO below 30 milli-g (17,21,22). An MVPA minute is defined as a minute with an average ENMO higher than 100 milli-g (23). All other minutes (with an average ENMO between 30 milli-g to 100 milli-g) are defined as light intensity physical activity (LIPA). Given the complexity of threshold choices, better notation is probably necessary. For example, for MVPA, it is much more precise to say

 $MVPA_{w,E,m,100} = MVPA(wrist, ENMO, minute, threshold = 100),$ 

to indicate that MVPA is calculated from wrist data, based on the ENMO metric, averaged at the minute level, and thresholded at 100 milli-g. This could substantially reduce the current confusion about the choice of thresholds at the expense of additional notation.

Figure 1 displays the pairwise correlations between the 25 PA predictors obtained from wrist accelerometry. The first 14 variables correspond to measures of volume of PA accumulated at different



**Figure 1.** Correlation plot of all 25 PA variables derived from wrist accelerometry. Color, transparency, and size of circles indicates the strength correlation between variables. In the online version of this manuscript, color indicates positive positive (blue) or negative (red) correlation between variables.

times of the day. The next 2 variables are fragmentation measures, followed by time spent in each of sedentary/sleep, LIPA, and MVPA. The last 6 variables are measures associated with circadian rhythms. Many features are highly correlated, though L5, timing of L5 and M10, DARE, and PA accumulated during early morning exhibit lower correlations with the other features.

There are many features which can be used to describe aspects of an individual's PA. Our choice of specific features was driven by which variables have been found to be predictive of, or associated with, all-cause mortality in other populations and include features that are proxies for 3 different domains of PA: volume, fragmentation, and circadian rhythm. Measures of volume of PA based on thresholding (eg, ST, LIPA, MVPA) are interpretable and highly predictive of mortality (3,7,24), but are sensitive to the choice of thresholds. In contrast, measures of volume which are not based on thresholds (eg, TA, TLA) are less interpretable, but do not lose information due to thresholding. Previous work by Varma et al. (25) has shown that measures of total activity and total log transformed activity are strongly correlated with light PA and MVPA, respectively. These results are based on activity counts and not milli-gs. We include both threshold and non-threshold-based measures of PA because: (i) the thresholds for levels of PA have not been extensively validated in the UK Biobank; and (ii) non-threshold based features were more predictive of mortality in NHANES (8).

Fragmentation measures the stability of an individual's activity patterns. The measures we consider here, ASTP and SATP, have natural interpretations as probabilities. One measures the probability that an individual will exit an active state (ASTP) while the other measures the probability that an individual will exit a sedentary

state (SATP). Interestingly, Wanigatunga et al. (26) reported that fragmentation, as measured by ASTP, was significantly associated with all-cause mortality in well-functioning adults over the age of 65 while volume of PA was not. In contrast, Smirnova et al. (8) found ASTP, SATP, and volume of PA to be highly associated with all-cause mortality in a nationally representative sample of U.S. adults aged 50–84. These results suggest that different features of PA may be predictive of mortality in different populations. Since there is no prior information about which specific features are predictive of mortality in the UK Biobank population, we use a large number of accelerometry-derived PA predictors.

In addition to features quantifying the volume and fragmentation of PA, features quantifying circadian rhythms have been associated with all-cause mortality (27-29). However, many of these features require estimation of various sleep parameters, which we do not consider in this analysis. In contrast, M10, L5, RA, and DARE are features that are related to different aspects of the circadian rhythm and are estimable without the need for estimating sleep periods. These measures have been shown to be predictive of several health outcomes and prevalent comorbidities (20,30,31). M10 and L5 measure how much activity an individual accumulates during the most active 10 hours and least active 5 hours of the day, respectively. For a fixed PA volume, a higher M10 indicates that PA is more concentrated during a 10-hour window as opposed to being more evenly distributed over the course of the day. In contrast, higher L5 indicates that an individual may have more activity during periods of least activity, which typically correspond to sleep. RA is defined as (M10 - L5)/ (M10 + L5), which measures the difference in activity during the most and least active periods of the day. That is, a larger RA is typically an indicator of a more pronounced circadian rhythm of activity. Finally, DARE measures the proportion of PA accumulated between the hours of 8 AM and 8 PM, which uses a pre-specified portion of the day compared to RA, which adapts to the individual- and dayspecific activity patterns.

# Mortality prediction models

Following Smirnova et al. (8), our aims are to: (i) rank predictors in terms of their predictive performance for time to death; and (ii) identify the most predictive combination of mortality predictors. These aims were addressed by fitting Cox regression models (32) using time from accelerometer wear to mortality or censoring. Predictive performance was assessed using 10-fold cross-validated Concordance (33). For the first aim, variables were ranked from the most to the least predictive based on the estimated Concordance from univariate models. For the second aim, we selected the best set of predictors using forward selection. Variables are included sequentially based on the net change in the 10-fold cross-validated Concordance associated with their addition to the set of model predictors; interactions, non-linear effects, or time-varying effects were not considered. We used 2 stopping rules for adding predictors to the model: an improvement in cross-validated Concordance of at least 0.01 and at least 0.001. A larger threshold in Concordance improvements results in more parsimonious models as fewer variables can improve the Concordance with a more stringent threshold. Finally, the predictive performance of PA variables in addition to traditional predictors was studied. This was done using a 2-stage forward selection process. The first stage included only traditional predictors using the same 2 stopping rules described above. The second step added PA variables using the same 2 stopping rules. A partial likelihood ratio test was used to compare models with and without PA predictors. To assess the potential optimism bias associated with forward selection, data were partitioned into 2 equal sized subsets, models were fit on the first subset, and the estimated Concordance was obtained by applying the model fit on the first subset to the second subset. Finally, we assessed potential changes in the univariate predictive value of each variable over time using time-dependent incident/dynamic area under the receiver operating characteristic curve (AUC), which has been shown to have a direct relationship to Concordance (34). Specifically, Concordance is a weighted average of time-dependent incident/dynamic AUC. We obtain estimates of time-dependent incident/dynamic AUC at each unique non-censored event time and smooth the estimates using penalized regression splines.

#### Results

#### Best Single Predictors of Mortality

Table 2 presents the top 30 predictors of mortality as measured by cross-validated Concordance in univariate Cox regression models. Also shown are summary statistics stratified by whether or not the individual was deceased at the end of the follow-up period. Variables are ordered in terms of cross-validated Concordance in single variable regression from the most predictive (age rank 1) to least predictive (TLA between 4 AM and 6 AM). Overall, with the exception of age, measures derived from accelerometers outperform established predictors of mortality including longstanding disability or illness (rank 14), gender (rank 16), and self-reported overall health (rank 17). The difference in Concordance is also substantial from 0.661 for MVPA and TA to 0.590 for gender. Among the top 20 predictors of mortality, 16 are derived from accelerometry. The top 5 PA predictors capture the duration and frequency of higher-intensity activities (TA, MVPA), strength of the circadian rhythm (RA), TLA during the most active 10 hours (M10), and amount of low/light PA accumulated during late afternoon (TLA 4-6 PM). Although these PA predictors measure different aspects of an individual's PA, they are highly correlated (eg, see Figure 1) with the lowest correlation between MVPA and RA (0.47) and highest between TA and MVPA (0.91).

Similar results were reported by Smirnova et al. (8) using NHANES 2003–2006 data, though in the UK Biobank accelerometer-derived PA measures underperform age while in NHANES age and volume of higher-intensity activities have similar predictive performance. Smirnova et al. (8) also reported that total activity and fragmentation perform similarly, whereas in UK Biobank, TA and minutes MVPA substantially outperform fragmentation measures (ASTP, SATP). There are many potential explanations for some of the differences. In particular, the follow-up time is different in the UK Biobank, the PA variables were derived from different devices placed at different locations on the body, and the UK Biobank population analyzed here is younger and healthier than the NHANES study population (8).

An earlier version of this analysis which used mortality data available through 2018 showed the top 5 PA predictors reported here were more predictive of mortality than age. This phenomenon can be explained by considering the estimated time-dependent incident dynamic AUC. In Supplementary Figure 2, we plot smooth estimates of the time-dependent incident/dynamic AUC for age and the top 5 PA predictors. The PA features are more predictive of mortality than age in the early follow-up period (1–2 years) by a considerable margin. However, their predictive performance decreases over time

and 2 years after accelerometer wear, age becomes more predictive than all 5 PA variables. These results suggest that PA-derived variables are more predictive of mortality immediately after taking the measurement, which may be consistent with a time-varying association with mortality risk. This is a potentially interesting area of future research.

# Best Subset of Predictors of Mortality

Table 3 presents the results of forward selection when we consider both PA and non-PA variables as candidates for the model at each step of the procedure. Variables were accepted into the model one at a time according to which increased Concordance the most. Two stopping rules were used: when the Concordance did not increase by more than 0.01 or by more than 0.001. Variables are ordered according to their inclusion order. Coefficient estimates and 95% confidence intervals are shown for each model associated with a stopping rule, respectively. For the more conservative stopping rule (change in Concordance of less than 0.01 and labeled  $\delta C$  $\geq$  0.01), only 2 variables are selected: age and relative amplitude. For the less conservative stopping rule (change in Concordance of less than 0.001 and labeled  $\delta C \ge 0.001$ ), 3 additional PA predictors are selected: TA, ST, and L5. Further, 6 additional established predictors of mortality are included (15): self-reported overall health, cigarette smoking, gender, history of cancer, longstanding illness/ disability, and injury/illness within past 2 years. The PA variables selected consist of measures of volume (TA, ST, L5) and circadian rhythms (RA, L5). No measures of fragmentation were selected using this procedure.

Table 4 presents the results of the 2-step forward selection procedure designed to quantify the added predictive performance of accelerometer-derived PA variables above and beyond traditional predictors of mortality. The difference from the previous procedure is that we now first select the best set of traditional predictors of mortality and then add the accelerometry-derived PA measures. The structure of Table 4 mirrors that of Table 3. In the traditional plus PA variables model, RA is included using the more conservative stopping rule  $(\delta C = 0.01)$  and RA, TA, and ST are included using the less conservative stopping rule ( $\delta C = 0.001$ ). The non-PA variables selected for inclusion overlap almost perfectly with those identified in Table 3 with the exception of BMI and alcohol consumption, which are included in the less conservative 2-stage model results presented in Table 4. In the more parsimonious model ( $\delta C \ge 0.01$ ), including relative amplitude (RA) increased the cross-validated Concordance from 0.717 to 0.733. In the less parsimonious model ( $\delta C \ge 0.001$ ), adding RA, TA, and ST increased the cross-validated concordance from 0.735 to 0.748. In both models, adding the PA variables resulted in highly statistically significant improvements in model fit (p < .001 for both).

Our sensitivity analysis showed that, overall, the results of forward selection were largely unchanged in terms of variables selected. However, we did find that using a threshold of  $\delta C \geq 0.001$  resulted in the inclusion of some variables which did not improve Concordance in the test data. This suggests that a threshold of 0.001 may be overly optimistic in our application. See Supplementary Tables 4 and 5 for more details.

## **Discussion**

The results presented here provide important new evidence supporting previous findings, which identified objective PA features obtained from wearable accelerometers as some of the strongest

Table 2. Population Characteristics Stratified by Mortality Status at Follow-up Ordered by Estimated Concordance From Univariate Cox Regressions

Rank	Variable	Concordance	Alive	Deceased
1	Age	0.681	63.59 (6.80)	67.79 (5.76)
2	MVPA	0.661	96.95 (52.62)	70.38 (50.36)
3	TA	0.661	40526.15 (11854.44)	34224.63 (11657.22
4	RA	0.644	0.54 (0.05)	0.51 (0.07)
5	M10	0.641	3.32 (0.33)	3.13 (0.40)
6	TLA 4–6 PM	0.64	369.55 (53.01)	341.86 (59.38)
7	TLA 6–8 PM	0.629	350.31 (53.34)	326.02 (55.91)
3	TLA	0.628	3427.01 (317.42)	3272.91 (368.70)
9	ST	0.627	1037.25 (100.00)	1084.39 (114.42)
10	TLA 2–4 PM	0.617	379.61 (52.55)	355.91 (59.42)
11	TLA 12–2 PM	0.609	393.21 (47.80)	372.61 (54.77)
12	ASTP	0.599	0.20 (0.05)	0.22 (0.08)
13	SATP	0.595	0.07 (0.02)	0.07 (0.02)
14	Longstanding illness/disability (% yes)	0.59	22 925 (28.5)	910 (46.5)
15	DARE	0.589	0.66 (0.03)	0.65 (0.04)
16	Gender (% male)	0.587	35 253 (43.9)	1208 (61.7)
17	Self-reported overall health (%)	0.585	, ,	,
17	Good	****	48 561 (60.4)	1045 (53.3)
	Poor		1854 (2.3)	136 (6.9)
	Fair		12 004 (14.9)	474 (24.2)
	Excellent		17 926 (22.3)	304 (15.5)
18	TLA 10 Am-12 PM	0.585	399.66 (53.58)	381.83 (60.92)
19	TLA 8–10 PM	0.584	292.74 (53.48)	277.63 (53.88)
20	LIPA	0.578	305.80 (72.85)	285.23 (84.34)
21	Smoking (%)	0.577	303.80 (72.83)	203.23 (04.34)
-1	Never	0.377	45 759 (57.0)	844 (43.1)
	Former		29 436 (36.6)	882 (45.0)
	Current		5150 (6.4)	233 (11.9)
22	TLA 8–10 Am	0.575	367.76 (69.64)	348.77 (75.72)
23	Usual walking speed (%)	0.574	367.76 (62.64)	346.// (/3./2)
23	Steady average pace	0.374	38 977 (48.5)	1046 (53.4)
			3661 (4.6)	
	Slow pace		, ,	222 (11.3)
1.4	Brisk pace Body mass index	0.565	37 707 (46.9)	691 (35.3)
24	Normal	0.363	21.019 (29.6)	572 (20.2)
			31 018 (38.6)	573 (29.2)
	Underweight		437 (0.5)	13 (0.7)
	Overweight		33 546 (41.8)	827 (42.2)
	Obese	0.562	15 344 (19.1)	546 (27.9)
25	L5	0.563	0.99 (0.11)	1.02 (0.13)
26	High blood pressure (% yes)	0.562	19 056 (23.7)	706 (36.0)
27	TLA 2-4 AM	0.557	131.09 (23.82)	135.37 (24.85)
28	TLA 6-8 AM	0.549	221.65 (79.10)	208.94 (74.46)
29	Alcohol consumption (%)	0.545	10.075 (22.7)	550 (20.4)
	Daily or almost daily		19 075 (23.7)	550 (28.1)
	3–4 times a week		21 258 (26.5)	416 (21.2)
	Once or twice a week		19 797 (24.6)	457 (23.3)
	1–3 times a month		8405 (10.5)	180 (9.2)
	Special occasions only		7399 (9.2)	201 (10.3)
	Never		4411 (5.5)	155 (7.9)
30	TLA 4–6 AM	0.543	136.17 (30.65)	138.96 (29.03)

Note: ASTP = active to sedentary transition probability; DARE = daytime activity ratio estimate; L5 = average log acceleration during the least active 5 h of the day; LIPA = light intensity physical activity; M10 = average log acceleration during the most active 10 h of the day; MVPA = moderate-to-vigorous physical activity; RA = relative amplitude; SATP = sedentary to active transition probability; ST = sedentary/sleep time; TA = total acceleration; TLA = total log acceleration. Only the top 30 single predictors are listed here. Continuous measures are reported as the average (SD) separately for alive and deceased participants. Categorical variables are reported as the number of participants within each category (% frequency).

predictors of all-cause mortality, including age, gender, and other traditional risk factors. Moreover, even after adjusting for age and standard predictors of mortality, we have found that inclusion of predictors derived from accelerometry significantly improved the

predictive performance. A crucial limitation of the UK Biobank study is that many of the non-PA variables are not measured contemporaneously with PA. However, these variables seem to retain much of their predictive power relative to that seen in other studies.

**Table 3.** Results of Forward Selection Using a Set of Variables Which Includes Physical Activity Predictors Using 2 Different Stopping Rules Based on Improvement in Concordance ( $\delta C$ )

Stopping Rule: $\delta C \ge 0.01$			
	Cumulative		
Variable	Concordance	$\delta C$	$\hat{eta} \pm 2SE(\hat{eta})$
Age	0.681	0.681	0.096 (0.088, 0.104)
RA	0.720	0.038	-0.446 (-0.481, -0.411)
Stopping Rule: $\delta C \ge 0.001$			
Age	0.681	0.681	0.086 (0.078, 0.094)
RA	0.720	0.038	-0.497 (-0.620, -0.374)
Self-reported overall health	0.729	0.009	
Poor			0.581 (0.384, 0.778)
Fair			0.240 (0.123, 0.357)
Good			Ref.
Excellent			0.010 (-0.121, 0.141)
Cigarette smoking	0.734	0.005	
Never			Ref.
Former			0.196 (0.100, 0.292)
Current			0.654 (0.506, 0.802)
Gender (male)	0.738	0.003	0.394 (0.297, 0.492)
Cancer	0.741	0.003	0.394 (0.263, 0.524)
Longstanding illness/disability	0.744	0.002	0.256 (0.153, 0.359)
TA	0.746	0.002	-0.328 (-0.453, -0.203)
Sedentary time, min	0.747	0.001	-0.003 (-0.005, -0.002)
Injury/illness within past 2 years	0.749	0.001	0.316 (0.185, 0.447)
L5	0.750	0.001	-0.200 (-0.304, -0.097)

Note: L5 = average log acceleration during the least active 5 h of the day; RA = relative amplitude; TA = total acceleration. Variables are listed in order of their inclusion into the model and the 10-fold cross-validated Concordance from a model which includes all variables up to the current variable is presented in the second column. The third column presents the improvement in cross-validated Concordance from including the current variable in the model. The fourth column presents the estimated coefficient and 95% confidence intervals obtained from the final model fit. Coefficients for total acceleration, relative amplitude, and L5 have been standardized such that interpretation is in units of SDs. Sedentary time is expressed in units of minutes.

In terms of validating the results from previous studies, we found that the most predictive mortality features in the UK Biobank are similar to those identified in a representative sample of the adult U.S. population. However, these measures are not identical because the 2 studies used different devices, data are summarized using different algorithms (activity count in NHANES and milli gravitational units in UK Biobank), placed at different locations on the body (hip for NHANES and wrist for UK Biobank).

The exact predictive performance reported in the UK Biobank and NHANES is different. This can be due to the different composition and context of the populations, differences in location of the device and data summarization approaches, sampling mechanisms, and synchronization of baseline and accelerometer data collection. The fact that age outperforms all PA features as a univariate predictor of mortality (which was not the case in the NHANES) may be attributable to the study population. Specifically, NHANES is a U.S. nationally representative sample while the UK Biobank accelerometry study is drawn from the UK population and is subject to substantial bias (the analyzed data are from individuals who are healthier than the UK Biobank study participants of the same age). In addition, while the measures of activity fragmentation are strong predictors of mortality in the UK Biobank, they do not perform as well as in NHANES. Some potential reasons for this discrepancy could be: (i) the UK Biobank protocol required continuous monitoring of activity, whereas NHANES participants were instructed

to remove the accelerometer during sleep; (ii) the wrist-worn device captures a qualitatively different fragmentation pattern than the hip-worn accelerometer; (iii) the threshold of 30 milli-gs for discriminating between active and sedentary behaviors is either too high or too low. Measures of circadian rhythms were not used in the analysis of the NHANES data and thus no direct comparisons can be made. The reason is that the NHANES study protocol required study participants to remove the device during sleep, which would introduce substantial bias and measurement error in the calculation of these variables.

Although measures of PA volume are the most predictive of mortality in univariate models among accelerometry-derived PA measures, they are also highly correlated with age. Thus, much of their predictive performance may be attributable to the natural decline in age-related volume of PA. In contrast, measures of circadian rhythmicity (RA) are less correlated with age and add substantial predictive performance to traditional mortality risk factors. Measures of fragmentation appear to be less predictive of mortality in the UK Biobank study as compared to NHANES. This may indicate that different domains of PA are more predictive of mortality in specific populations. While results do not target causal relationships between PA and mortality, they suggest that it worth investigating interventions that simultaneously target: (i) increasing volume of PA when possible; (ii) modifying timing of PA to align with an individuals' specific circadian profile; and (iii) improving endurance as a means of reducing fragmentation of PA.

Table 4. Results of a 2-Stage Forward Selection Procedure

Stopping Rule: $\delta C \ge 0.01$					
Variable	Cumulative Concordance	δC	$\hat{eta} \pm 2 SE(\hat{eta})$		
Age	0.681	0.681	0.097 (0.089, 0.105)		
Self-reported overall health	0.705	0.023			
Poor			0.995 (0.812, 1.179)		
Fair			0.454 (0.344, 0.565)		
Good			Ref.		
Excellent			-0.125 (-0.253, 0.003)		
Gender (male)	0.717	0.012	0.358 (0.263, 0.452)		
RA	0.733	0.016	-0.351 (-0.389, -0.313)		
Stopping Rule: $\delta C \ge 0.001$					
Age	0.681	0.681	0.085 (0.077, 0.094)		
Self-reported overall health	0.705	0.023			
Poor			0.546 (0.348, 0.744)		
Fair			0.220 (0.102, 0.338)		
Good			Ref.		
Excellent			0.019 (-0.112, 0.150)		
Gender (male)	0.717	0.012	0.412 (0.312, 0.512)		
Cigarette smoking	0.724	0.007			
Never			Ref.		
Former			0.198 (0.100, 0.295)		
Current			0.656 (0.507, 0.805)		
Longstanding illness/disability	0.728	0.004	0.246 (0.142, 0.349)		
Cancer	0.731	0.003	0.397 (0.267, 0.528)		
BMI	0.733	0.002			
Underweight			0.509 (-0.042, 1.060)		
Normal			Ref.		
Overweight			-0.045 (-0.155, 0.064)		
Obese	0.704	0.004	0.079 (-0.047, 0.206)		
Injury/illness within past 2 years	0.734	0.001	0.312 (0.182, 0.443)		
Alcohol consumption	0.735	0.001	D (		
Daily or almost daily			Ref.		
3 or 4 times a week			-0.163 (-0.291, -0.035)		
Once or twice a week			-0.000 (-0.126, 0.125)		
1–3 times a month			-0.087 (-0.258, 0.085)		
Special occasions only			0.023 (-0.144, 0.189)		
Never	0.545	0.000	0.155 (-0.028, 0.339)		
RA	0.745	0.009	-0.264 (-0.312, -0.217)		
TA	0.746	0.001	-0.378 (-0.502, -0.255)		
ST, min	0.748	0.001	-0.002 (-0.003, -0.001)		

Note: BMI = body mass index; RA = relative amplitude; ST = sedentary time; TA = total acceleration. The first phase uses variables that exclude physical activity predictors, while the second phase adds PA variables. Results are presented for 2 different stopping rules based on improvement in Concordance (δC). Variables are listed in order of their inclusion into the model and the 10-fold cross-validated Concordance from a model which includes all variables up to the current variable is presented in the second column. The third column presents the improvement in cross-validated Concordance from including the current variable in the model. The fourth column presents the estimated coefficient and 95% confidence intervals obtained from the final model fit. Coefficient for relative amplitude has been standardized such that interpretation is in units of SDs.

As with any cross-sectional study, there is a concern regarding about the potential for reverse causality. More precisely, the observed PA variables may be caused by the inherent risk of mortality associated with an individual's health status rather than causing mortality. Another concern is that the causal effect may be a bidirectional. This possibility is partially supported by the decreased predictive performance of PA variables over time. However, several PA variables have very strong prediction performance for many years up to 6 years after the baseline measurement, as quantified by the time-dependent incident/dynamic AUC. The concerns regarding reverse causality could partially be addressed by conducting a longitudinal study, though increased follow-up time may also provide additional insight.

Despite these limitations, our study offers novel insights into the ability of objectively measured PA to predict mortality. Importantly, PA is one of the few modifiable risk factors for mortality. We found features of PA to be predictive of mortality above and beyond traditional predictors such as age, suggesting possible targets for future PA interventions. Specifically, measures of volume of MVPA and strength of circadian rhythm appear to be particularly appealing targets for interventions aimed at reducing the risk of all-cause mortality.

# **Supplementary Material**

Supplementary data are available at *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences* online.

# **Funding**

This work was supported by the National Institute of Neurological Disorders at the National Institutes of Health (R01NS060910) and the National Institute on Aging Training Grant (T32AG000247).

# **Conflict of Interest**

C.C. is consulting with Bayer and Johnson and Johnson on methods development for wearable devices in clinical trials. The details of the contracts are disclosed through the Johns Hopkins University eDisclose system and have no direct or apparent relationship with this manuscript. All other authors declare no conflicts of interest.

## **Author Contributions**

Study concept and design: A.L., E.S., N.C., and C.C. Acquisition of the data: N.C. Analysis and interpretation of the data: A.L., S.X., P.K, J.M, N.C, and C.C. Preparation of the manuscript: A.L., S.X., P.K., J.M., E.S., N.C., and C.C.

## References

- Chudasama YV, Khunti KK, Zaccardi F, et al. Physical activity, multimorbidity, and life expectancy: a UK Biobank longitudinal study. BMC Med. 2019;17(1):108. doi:10.1186/s12916-019-1339-0
- Fishman EI, Steeves JA, Zipunnikov V, et al. Association between objectively measured physical activity and mortality in NHANES. *Med Sci Sports Exerc*. 2016;48(7):1303–1311. doi:10.1249/MSS.00000000000000885
- Koster A, Caserotti P, Patel KV, et al. Association of sedentary time with mortality independent of moderate to vigorous physical activity. PLoS ONE. 2012;7(6):e37696. doi:10.1371/journal.pone.0037696
- Loprinzi PD. Accelerometer-determined physical activity and mortality in a national prospective cohort study: considerations by hearing sensitivity. *Am J Audiol*. 2015;24(4):569–572. doi:10.1044/2015\_AJA-15-0044
- Loprinzi PD. Accelerometer-determined physical activity and all-cause mortality in a national prospective cohort study of hypertensive adults. J Hypertens. 2016;34(5):848–852. doi:10.1097/HJH.00000000000000869
- Loprinzi PD, Walker JF. Increased daily movement associates with reduced mortality among COPD patients having systemic inflammation. *Int J Clin Pract*. 2016;70(3):286–291. doi:10.1111/ijcp.12778
- Matthews CE, Keadle SK, Troiano RP, et al. Accelerometer-measured doseresponse for physical activity, sedentary time, and mortality in US adults. Am J Clin Nutr. 2016;104(5):1424–1432. doi:10.3945/ajcn.116.135129
- Smirnova E, Leroux A, Cao Q, et al. The predictive performance of objective measures of physical activity derived from accelerometry data for 5-year all-cause mortality in older adults: National Health and Nutritional Examination Survey 2003–2006. J Gerontol A Biol Sci Med Sci. 2019;75(9):1779–1785. doi:10.1093/gerona/glz193
- Washburn RA. Assessment of physical activity in older adults. Res Q Exercise Sport. 2000;71(Suppl. 2):79–87. doi:10.1080/02701367.200 0.11082790
- Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. Res Q Exercise Sport. 2000;71(Suppl. 2):1–14. doi:10.1080/02701367.2000.11082780
- Doherty A, Jackson D, Hammerla N, et al. Large scale population assessment of physical activity using wrist worn accelerometers: the UK Biobank study. PLoS ONE. 2017;12(2):e0169649. doi:10.1371/journal.pone.0169649
- Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203–209. doi:10.1038/s41586-018-0579-z
- Elliott LT, Sharp K, Alfaro-Almagro F, et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature*. 2018;562(7726):210–216. doi:10.1038/s41586-018-0571-7
- 14. Batty GD, Gale CR, Kivimäki M, Deary IJ, Bell S. Comparison of risk factor associations in UK Biobank against representative, general population

- based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. *Br Med J.* 2020;368:m131. doi:10.1136/bmj.m131
- Ganna A, Ingelsson E. 5 year mortality predictors in 498,103 UK Biobank participants: a prospective population-based study. *Lancet*. 2015;386(9993):533–540. doi:10.1016/S0140-6736(15)60175-1
- Bai J, Di C, Xiao L, et al. An activity index for raw accelerometry data and its comparison with other activity metrics. *PLoS ONE*. 2016;11(8):e0160644. doi:10.1371/journal.pone.0160644
- van Hees VT, Gorzelniak L, Dean León EC, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PLoS ONE. 2013;8(4):1–10. doi:10.1371/journal.pone.0061691
- Leroux A, Di J, Smirnova E, et al. Organizing and analyzing the activity data in NHANES. *Stat Biosci*. 2019;11:262–287. doi:10.1007/s12561-018-09229-9
- 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.doi:10.1101/182337
- Kudchadkar SR, Aljohani O, Johns J, et al. Day-night activity in hospitalized children after major surgery: an analysis of 2271 hospital days. *J Pediatr*. 2019;209:190–197.e1. doi:10.1016/j.jpeds.2019.01.054
- Hees van V. Accelerometer Data Processing with Ggir. 2019.https://cran.rproject.org/web/packages/GGIR/vignettes/GGIR.html. Accessed February 18, 2020.
- Sabia S, Hees van V, Shipley M, et al. Association between questionnaireand accelerometer-assessed physical activity: the role of sociodemographic factors. Am J Epidemiol. 2014;179(6):781–790. doi:10.1093/aje/kwt330
- Hildebrand M, Van Hees V, Hansen B, Ekelund U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. Med Sci Sports Exerc. 2014;46(9):1816–1824. doi:10.1249/MSS.0000000000000000089
- LaMonte MJ, Buchner DM, Rillamas-Sun E, et al. Accelerometer-measured physical activity and mortality in women aged 63 to 99. J Am Geriatr Soc. 2018;66(5):886–894. doi:10.1111/jgs.15201
- Varma VR, Dey D, Leroux A, et al. Total volume of physical activity: TAC, TLAC or TAC(λ). Prev Med. 2018;106:233–235. doi:10.1016/j. vpmed.2017.10.028
- Wanigatunga AA, Di J, Zipunnikov V, et al. Association of total daily physical activity and fragmented physical activity with mortality in older adults. *JAMA Netw Open*. 2019;2(10):e1912352. doi:10.1001/ jamanetworkopen.2019.12352
- 27. Tranah GJ, Blackwell T, Ancoli-Israel S, et al. Circadian activity rhythms and mortality: the study of osteoporotic fractures. *J Am Geriatr Soc.* 2010;58(2):282–291. doi:10.1111/j.1532-5415.2009.02674.x
- 28. Smagula SF, Stone KL, Redline S, et al. Actigraphy- and polysomnographymeasured sleep disturbances, inflammation, and mortality among older men. *Psychosom Med.* 2016;78(6):686–696. doi:10.1097/ PSY.0000000000000312
- Zeitzer JM, Blackwell T, Hoffman AR, Cummings S, Ancoli-Israel S, Stone K. Daily patterns of accelerometer activity predict changes in sleep, cognition, and mortality in older men. J Gerontol A Biol Sci Med Sci. 2018;73(5):682–687. doi:10.1093/gerona/glw250
- van Someren EJ, Hagebeuk EE, Lijzenga C, et al. Circadian restactivity rhythm disturbances in Alzheimer's disease. *Biol Psychiatry*. 1996;40(4):259–270. doi:10.1016/0006-3223(95)00370-3
- 31. Lyall LM, Wyse CA, Graham N, et al. Association of disrupted circadian rhythmicity with mood disorders, subjective wellbeing, and cognitive function: a cross-sectional study of 91 105 participants from the UK Biobank. *Lancet Psychiatry*. 2018;5(6):507–514. doi:10.1016/ S2215-0366(18)30139-1
- 32. Cox DR. Regression models and life-tables. *J Roy Stat Soc B (Methodol)*. 1972;34(2):187–220. doi:10.1111/j.2517–6161.1972.tb00899.x
- Harrell J, Frank E, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *J Am Med Assoc.* 1982;247(18):2543–2546. doi:10.1001/jama.1982.03320430047030
- 34. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. Biometrics. 2005;61(1):92–105. doi:10.1111/j.0006-341X.2005.030814.x