TITLE PAGE

Article Title:

Actigraphy Informs Distinct Patient-Centered Outcomes in Pre-COPD

Running Head:

Actigraphy in assessment of pre-COPD

Author List:

Jianhong Chen, MS [1,2,3] *, Lemlem Weldemichael, MD [1,2,3] *, Brian Giang, BA [4], Jeroen Geerts, MD [5], Siyang Zeng, MS [3,6], Wendy Czerina Ching, BS [1,2,3], Melissa Nishihama, BS [1,2,3], Warren M Gold, MD [1,3], Mehrdad Arjomandi, MD [1,2,3]

[1]Division of Pulmonary, Critical Care, Allergy and Immunology, and Sleep Medicine, Department of Medicine, University of California, San Francisco, California, USA

[2]Division of Occupational and Environmental Medicine; Department of Medicine, University of California, San Francisco, California, USA

[3]San Francisco Veterans Affairs Medical Center; San Francisco; California, USA

[4]George Washington University, School of Medicine, Washington, DC, USA

[5]Radboud University Medical Center, Nijmegen, the Netherlands

[6]University of Washington, School of Public Health, Seattle, WA, USA

* These authors contributed equally to this work.

Email Addresses:

JC: jianhong.chen@ucsf.edu

LW: weldemichael.lemlem@gmail.com

BG: briangiang@gwmail.gwu.edu

JG: jeroengeerts@msn.com

SZ: Siyang.zeng@ucsf.edu

WCC: Wendy.Ching@ucsf.edu

MN: mynishihama@gmail.com

WMG: warren.gold@ucsf.edu

MA: mehrdad.arjomandi@ucsf.edu

Corresponding Author Information:

Mehrdad Arjomandi, MD

Division of Pulmonary and Critical Care Medicine

University of California San Francisco

San Francisco Veterans Affairs Medical Center

Bldg 203, Room 3A-128, Mailstop 111-D

4150 Clement Street, San Francisco, CA 94121

TEL: (415) 221-4810 x24393

FAX: (415) 379-5538

EMAIL: mehrdad.arjomandi@ucsf.edu

Keywords: Actigraphy; machine learning; secondhand tobacco smoke; COPD; pre-COPD; air trapping; cardiopulmonary exercise; pulmonary function testing.

Contributions:

Conceived and designed the experiments: WMG, MA.

Developed study protocols: WCC, WMG, MA.

Collected data: LW, JG, BG, WCC, MN, WMG, MA.

Analyzed and interpreted data: JC, LW, JG, BG, SZ, MA.

Prepared the manuscript: JC, LW, BG, SZ, MA.

Edited the manuscript: JC, LW, JG, BG, SZ, WCC, MN, WMG, MA.

Obtained funding: JG, WMG, MA.

Funding:

This work was supported by:

1. Flight Attendant Medical Research Institute (012500WG to WMG and CIA190001 to MA).

2. California Tobacco-Related Disease Research Program (TRDRP) (T29IR0715 to MA).
3. The Department of Veterans Affairs Airborne Hazard and Burn Pit Center of Excellence (to MA).
4. Radboud University School of Medicine Scholarship (to JG).

The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The statements and conclusions in this publication are those of the authors and not necessarily those of the funding agency. The mention of commercial products, their source, or their use in connection with the material reported herein is not to be construed as an actual or implied endorsement of such products.

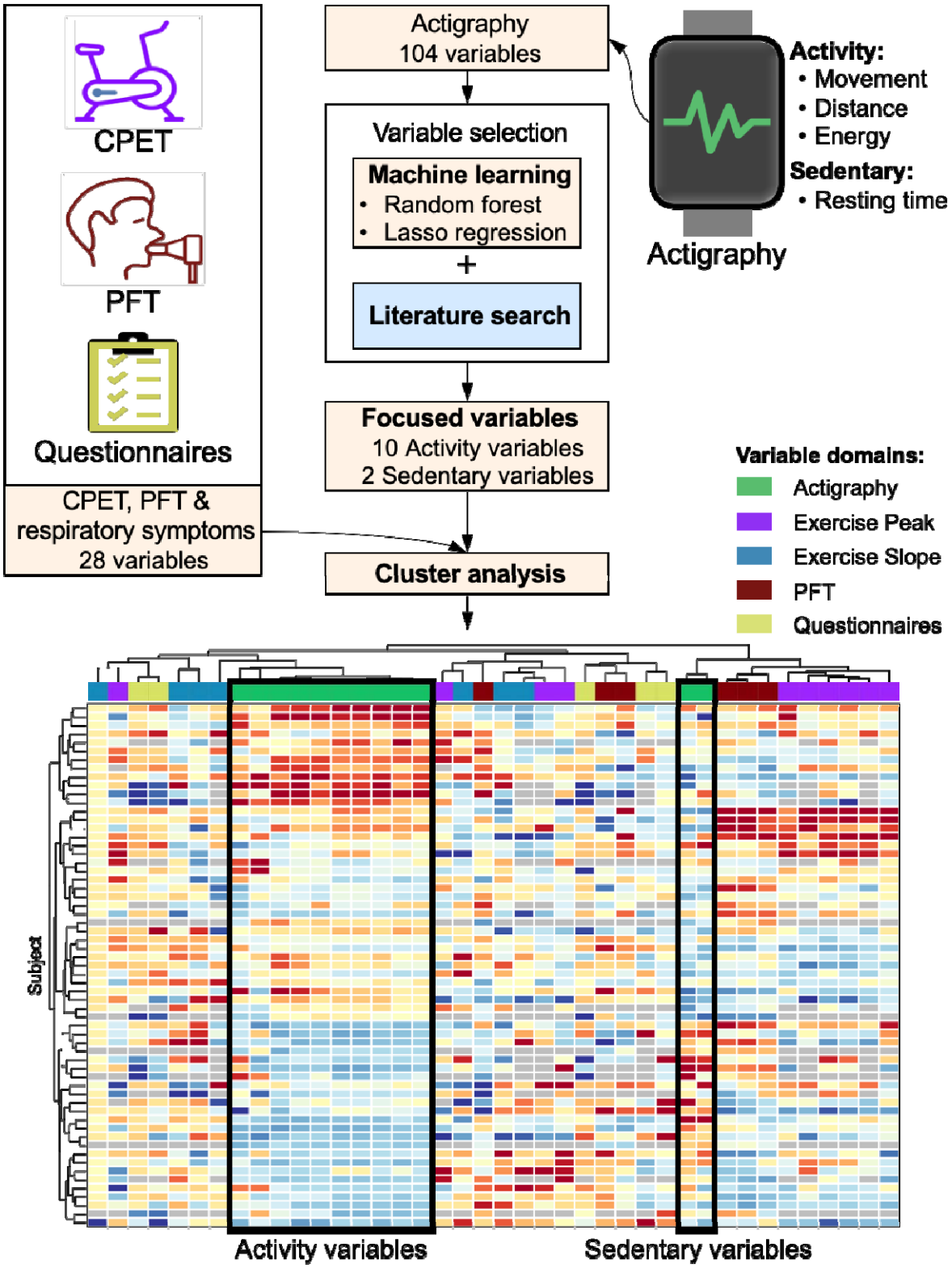
Competing Interests:

There are no competing interests for any author.

Acknowledgements:

We would like to thank Oliver Beech, Patricia Emerson-Healy, Emily Ghio, Liane Tolang, and Charlotte M Hunt for help with performance of cardiopulmonary exercise testing. We also would like to appreciate the contribution of the flight attendants who took time out of their busy schedules to participate as research subjects in this study.

VISUAL ABSTRACT



ABSTRACT

Background-Actigraphy can clarify useful patient-centered outcomes for quantification of physical activity in the "real-world" setting.

Methods-To characterize the relationship of actigraphy outputs with "in-laboratory" measures of cardiopulmonary function and respiratory symptoms in pre-COPD, we obtained actigraphy data for 8 hours/day for 5 consecutive days a week before in-laboratory administration of respiratory questionnaires, PFT, and CPET to a subgroup of subjects participating in the larger study of the health effects of exposure to secondhand tobacco smoke who had air trapping but no spirometric obstruction (pre-COPD). Using machine learning approaches, we identified the most relevant actigraphy predictors and examined their associations with symptoms, lung function, and exercise outcomes.

Results-Sixty-one subjects (age=66±7years; BMI=24±3kg/m^2; FEV_1/FVC=0.75±0.05; FEV_1=103±17%predicted) completed the nested study. In the hierarchical cluster analysis, the activity, distance, and energy domains of actigraphy, including moderate to vigorous physical activity, were closely correlated with each other, but were only loosely associated with spirometric and peak exercise measures of oxygen consumption, ventilation, oxygen-pulse, and anaerobic threshold (VO_{2AT}), and were divergent from symptom measures. Conversely, the sedentary domain clustered with respiratory symptoms, air trapping, airflow indices, and ventilatory efficiency. In Regression modeling, sedentary domain was inversely associated with baseline lung volumes and tidal breathing at peak exercise, while the activity domains were associated with VO_{2AT}. Respiratory symptoms and PFT data were not associated with actigraphy outcomes.

Discussion-Outpatient actigraphy can provide information for “real-world” patient-centered outcomes that are not captured by standardized respiratory questionnaires, lung function, or exercise testing. Actigraphy activity and sedentary domains inform of distinct outcomes.

INTRODUCTION

The ability to assess and quantify the physical activity levels of patients can offer valuable insights into their general health and day-to-day functioning that may not be captured by traditional objective measurements of disease status, such as interview- or questionnaire-based evaluations or physiological or laboratory testing.[1-3] Wearable activity monitors that measure "real-world" physical activity levels may thus serve as objective and reliable tools to assess health status and disease activity.[4-7] The current-generation wearable activity monitors, including pedometers and accelerometers, are transforming the field of biomedical research by their capacity to approximate free-living conditions and measure real-world physical activity in a continuous and longitudinal, yet objective manner.[8] However, the correlation between physical activity measurements obtained by actigraphy and the physical activity outcomes measured by physiologic and questionnaire-based tools has not been clearly identified.

In this study, we aimed to understand the relationship between physical activity measures obtained from a wearable activity monitor and those obtained from standardized respiratory questionnaires and "in-laboratory" lung function and exercise testing. We hypothesized that "real-world" actigraphy provides distinct outcomes that are not completely captured by standardized symptom questionnaires or traditional "in-laboratory" functional assessments. To assess this hypothesis, we examined the association of outpatient actigraphy measures with patient-reported respiratory physical activity and symptoms as well as lung function and exercise test results in a never-smoker cohort subgroup of a study on the health effects of exposure to secondhand tobacco smoke (SHS); the participants were at risk for COPD due to their prolonged occupational exposure to SHS but showed preserved spirometry.

METHODS

Study overview

This observational study was nested in a larger study examining the cardiopulmonary health effects of SHS exposure in a cohort of nonsmoking individuals with a range of occupational SHS exposure, as previously described.[9] Briefly, between July 2007 and March 2020, we recruited US airline flight crewmembers with a history of occupational exposure to SHS, along with nonsmoker controls without such occupational exposure, who were participating in a larger study of the cardiopulmonary health effects of prolonged exposure to SHS (ClinicalTrials.gov Identifier: NCT02797275). The participants underwent respiratory symptom questionnaire assessments, full pulmonary function testing (PFT), and maximum-effort cardiopulmonary exercise testing (CPET). For the actigraphy nested study, beginning February 2014, participants were asked to wear an activity monitor eight hours a day for five consecutive days during the week before they came in for in-laboratory evaluation and maintain a daily diary. The actigraphy data were then obtained and analyzed along with the respiratory questionnaire, PFT, and CPET data to examine its association with reported physical activity, symptoms, and in-laboratory measures of physical activity. The University of California San Francisco (UCSF) Institutional Review Board (IRB) and the San Francisco VA Medical Center Committee on Research and Development approved the study protocols. Full details of the methods are available in the **Supplemental Appendix**.

Study Population

US airline crewmembers, including flight attendants and pilots, were eligible to participate if they had worked for ≥5 years in an aircraft. A reference group of “sea-level” participants who lived in San Francisco Bay area and had never been employed as flight

crewmembers was also recruited. Participants were eligible if they were nonsmokers defined by never-smoking or, in ever smokers, no smoking for ≥20 years and a cumulative smoking history of <20 pack-years. Eligible participants were excluded if they had a known history of cardiopulmonary disease or recreational drug use, including marijuana consumption. All participants enrolled in the larger cohort were invited to participate in this nested study. For the nested actigraphy study, recruitment began in February 5, 2014 and continued through March 17, 2020.

Physical Activity Monitoring using Actigraphy

Physical activity was monitored using a triaxial accelerometer-based activity monitor (ActiGraph GT3X; Actigraph Corporations, Pensacola, FL). Technical details of the device can be found in the **Supplemental Appendix**. The ActiGraph monitor was initialized to continuously collect data over a period of 5 days. It was then mailed to participants along with a daily diary to keep a log of the time the monitor was worn and the activities the participants performed during that time. All participants received the ActiGraph monitors at least 7 days prior to the in-laboratory visit, during which respiratory symptom questionnaire, PFT, and CPET data were collected. Participants were instructed to wear the ActiGraph monitor on the waist upon awakening using a wrist band provided and to keep it on continuously for at least 8 hours for 5 consecutive days beginning the start of the first day of their work week. The 5-day monitoring period was chosen to allow for adequate weekday data collection[10] while avoiding recording of non-routine rest or activity periods that typically occur during weekends. All participants were carefully instructed on correctly positioning the device.

Actigraphy Data Description and Processing

Actigraphy data were processed using the ActiLife software program (Version 6.11.9; ActiGraph LLC) and saved in raw format as GT3X files. The ActiLife software generates a total of 52 variables in the distance, time (activity and sedentary), and energy domains. The list of variables and their definitions are shown in **Table S1**. The actigraphy data were matched against the diary to ascertain appropriate usage of the monitor, and the data were considered to be acceptable if the participants wore the monitor for a minimum of 3 days and for more than 5 hours (300 min) per day. The data was summarized into 5-day average values (weekly "mean" values) and the highest values of "maximum" values (maximum per epoch) across all 5 days (weekly "highest" values). The total amount of the time that the monitor was worn was included in the regression models as *total time worn*.

Actigraphy Data Variable Selection

Actigraphy generates a large number of variables, some of which are highly correlated. Given the low number of participants in our study, we decided to reduce the number of actigraphy variables to increase the robustness of the analysis. To achieve this, we pursued two approaches of variable selection: (1) machine learning approach, and (2) literature-guided approach (**Figure 1**).

In the machine learning approach, we built two models, random forest and lasso regression. Using random forest modeling, we generated a list of top-ranked actigraphy variables that were predictive of various questionnaire, PFT, and CPET outcomes by ranking variables based on minimizing prediction error at the splitting of the decision tree nodes. For lasso regression, we used a similar strategy to generate a list of important variables by utilizing the l_1-norm penalized terms to force unimportant variables to become zero. To rank the variables in lasso regression, we then compared the magnitude of variable coefficients in various models.

Finally, we summarized the actigraphy variables by including the top 10 variables from each model.

In the literature-guided approach, we reviewed the available literature on actigraphy and selected 9 variables to represent distance, energy, and the activity and sedentary time domains as described below (**Table S2**). The final set of variables was selected based on a combination of the machine learning and literature-guided approaches to provide meaningful variables of highest predictive value for our proposed analysis.

Pulmonary Function and Cardiopulmonary Exercise Testing

Details of our PFT and CPET procedures are presented in the **Supplemental Appendix** and have been previously described.[9,11,12]

Respiratory Questionnaires

Patient-reported respiratory symptoms, physical activity, and quality of life assessments were conducted using the COPD Assessment Test (CAT),[13] modified Medical Research Council (mMRC) Dyspnea Scale,[14] the Short Form 12-Item Health Survey (SF-12),[15] International Physical Activity Questionnaire (IPAQ),[16] and Airway Questionnaire 20 (AQ20).[17]

Data Analysis

Distributions of patients' actigraphy, respiratory symptom, lung function, and exercise data were visualized and inspected (**Figure S1**). Because most variables were not normally distributed, Spearman's rank correlation was used throughout the analysis. We then examined the association of actigraphy distance, energy, and activity and sedentary time domains with respiratory questionnaire, PFT, and CPET outputs after adjustment for age, sex, height, weight,

and time worn using hierarchical clustering with the Spearman correlation coefficient as the distance metric.

Even after selection of a focused set of actigraphy variables, the initial exploratory analyses showed two potential statistical challenges, namely, high dimensionality and high collinearity, for the application of these variables to ordinary linear regression modeling. To address these challenges, we employed partial principal component regression, which combined principal component analysis with adjustment for covariates to reduce the number of dimensions and transformed the original variables into orthogonal principal component (PC) axes. Next, we performed ordinary linear regression using the top six PC axes as the new predictive actigraphy variables within the model. We later computed the Pearson correlation coefficients between the original actigraphy variables versus the transformed PC axes to help interpret the representation of each PC axis. We used P-values of 0.05 as the cutoff for statistical significance and considered Spearman and Pearson correlation coefficients ±0.5 as indicating strong correlation (either positive or negative).

RESULTS

Participant Characteristics

Overall, 64 volunteers participated in the nested study and wore the actigraphy monitor during the week before participating in the in-laboratory assessments with respiratory questionnaires, PFT, and CPET. Three of the 64 subjects wore the actigraphy monitor for <3 consecutive days or <300 consecutive min/day and thus were excluded from the analysis. The characteristics of the 61 participants included in the analysis are shown in **Table 1**. The

participants were aged 65.7±10.7 years, predominantly female (57 [93%]), and never-smokers. All participants had a history of occupational exposure to SHS through their airline employment as flight crew for a median (interquartile range, [IQR]) period of 15.6 (6.0-25.0) years. All participants showed preserved spirometry, with forced expiratory volume in 1 s (FEV_1) and FEV_1/forced vital capacity (FVC) percent predicted values of 103±17 and 98±6, respectively. All of them showed mild air trapping, as defined by residual volume (RV)/total lung capacity (TLC) > 0.35 (0.38±0.06), and an overall reduced average diffusing capacity adjusted for hemoglobin of 67%±28% predicted. The percent predicted values for the measures of VO_2, V_E, and oxygen-pulse at the peak of exercise (VO_{2Peak}, V_{EPeak}, and O_2-PulsePeak) were 103.8±18.1, 61.1±14.3, and 109.4±20.7, respectively, and VO_2 at anaerobic threshold (VO_{2AT}) was 63.8%±16.7% of VO_{2Peak}. **Table 1** also shows the selected actigraphy data in the distance, energy, and activity and sedentary time domains. Overall, the participants wore the actigraphy monitor for a median (IQR) period of 5 (4-5) days and 38.9 (33.0-39.9) h (2,334 [1,980-2,396] min).

Combined Approach for Selection of Appropriate Actigraphy Variables

In the machine learning approach, random forest and lasso regression analyses were performed to identify the most relevant actigraphy variables that are predictive of the respiratory questionnaire, PFT, and CPET outcomes. The top-ranked variables for each machine learning approach, which are shown in **Figure S2 A and B**, were a mix of actigraphy domains, including the sedentary time domain. The ten top ranked variables from each approach were then matched against a list of most highly cited actigraphy variables in the literature, and a final set of actigraphy variables, which included variables from all domains, were then generated for further analysis (**Figure 1**). The final set of variables obtained by this combined approach is shown in **Table 2**.

The correlations among all and the final selection of actigraphy variables are shown in **Figure S1** and **Figure 2**, respectively. Variables from the distance, energy, and activity time domains were highly and directly correlated with each other, with the strongest correlation observed between the energy and activity time domains and the weakest between these domains and the distance domain. However, sedentary time domain variables were poorly and inversely correlated with the variables from the distance, energy, and activity time domains (**Figure 2**). Within the sedentary time domain, the total amount of time spent in the sedentary condition (Sedentary Time) was closely and directly associated with the number of times participants broke their sedentary condition (No. of Sedentary Breaks; $r=0.65$; $P<0.001$).

<u>Hierarchical Clustering Identified Two Clusters of Actigraphy Domains That Distinctly Group with Respiratory Questionnaire, PFT, and CPET Outcomes</u>

Hierarchical clustering analysis categorized actigraphy domains into two distinct clusters (**Figure S3** and **Figure 3**). Measures of the distance, energy, and activity time domains, including moderate-to-vigorous physical activity (MVPA), maximum number of steps taken per epoch (Max. Step Count), and average hourly energy expenditure (Avg. Kcal/hour), clustered together and were best correlated with the rate of increase in O_2-Pulse, tidal breathing, systolic blood pressure over workload ($O_2\text{-Pulse}_{Slope}$, SBP_{Slope}, and V_{TSlope},), and SF-12 and its "physical" component score, and to a lesser extent with air trapping (RV and RV/TLC), airflow obstruction (FEV_1/FVC), ventilatory efficiency at peak exercise (V_E/VCO_{2Peak}), and other respiratory questionnaires (AQ-20 and CAT) scores. Conversely, measures of the sedentary domain, including Sedentary Time and No. of Sedentary Breaks, clustered together with spirometric indices (FEV_1, FVC), TLC, and peak exercise measures of oxygen consumption (VO_{2Peak}), ventilation (V_{EPeak}), tidal volume (V_{TPeak}), and oxygen-pulse ($O_2\text{-Pulse}_{Peak}$).

Principal Component Analysis of a Focused Dataset of Actigraphy Variables Reproduced Contrasting Actigraphy Domains Represented by Two Principal Components

To address the high dimensionality and collinearity of actigraphy variables, we performed principal component analysis (**Table 3**) and regression of the final set of variables (**Table 4**). The first six principal components (PC) explained 87.8% of the variation in the dataset and were chosen for further analysis. The most contributing PC (PC1) showed high positive values for the following parameters: measures of energy expenditure (Avg. Kcal/hour=0.92, Highest MET=0.94, and Mean MET=0.96), maximum number of steps taken per epoch (Max. Step Count=0.56), average number of steps taken per minute (Avg. Steps/min=0.91), maximum number of steps counts on actigraphy axes 1 and 3, and as a vector magnitude across all 3 axes (Vector Magnitude Max. Count=0.4), time spent in MVPA (Highest Total MVPA=0.92; Avg. MVPA/hour=0.94). Therefore, PC1 was determined to be a reflection of the measures of activity time, distance, and energy domains of actigraphy. The second PC (PC2) was highly correlated with participant sex (0.77), height (0.84), and weight (0.84). The third PC (PC3) was highly correlated with the total time spent in the sedentary condition (Highest Sedentary Time=0.66), and thus was determined to be representative of the sedentary time domain of actigraphy. PC4 was only highly correlated with subject's age (0.77).

Results of the principal component regressions of the cardiopulmonary function and questionnaires outcomes over PC1 ("activity component) and PC3 ("sedentary component") are presented in **Table 4**. In the activity domains represented by PC1, a significant positive association was only observed with the volume of oxygen consumption at anaerobic threshold (VO_{2AT}) (P=0.026). In the actigraphy sedentary domain represented by PC3, significant negative associations were observed with FEV_1 (P=0.042) (marginally non-significance with FVC

[P=0.075] and TLC [P=0.057]) and the total and "mental" component of SF-12 score (P=0.042 and P=0.035, respectively). No other significant associations were present.

DISCUSSION

In this observational study of a population at risk for COPD due to prolonged occupational exposure to SHS but with no airflow obstruction ("pre-COPD"), we employed a combined machine learning and literature-based approach to tackle the high dimensionality of the actigraphy output variables and identified a set of "focused" variables that were most relevant to and best represented the daily physical activity pattern of subjects. We then pursued two different analytical strategies. In the first strategy, we used unsupervised cluster analysis to evaluate the general relationship among the in-laboratory physiological measures and questionnaire-based respiratory symptom scores with the focused set of actigraphy variables using an unadjusted approach and without any assumption of linearity to show how variables from different modalities may correlate with each other. Using this strategy, we found sedentary domain of actigraphy to be more closely grouped together with exercise variables representing lower peak exercise capacity and lack of anaerobic fitness, while distance, activity, and energy domains of actigraphy were more closely grouped together with variables representing presence of endurance and aerobic fitness (slope of variables over workload), better scores on respiratory questionnaires, and lung function indices of air trapping. In the second strategy, we used principal component regression to more exactly examine the associations between specific in-laboratory measures or specific respiratory symptom scores and focused actigraphy variables with adjustment for covariates in the models that assumed linear relationships. Using this strategy, we found only few significant associations. Although this overall lack of statistically

significant associations could be due to the sample size examined, the general trends were consistent with the unsupervised cluster analysis strategy, and suggested that, at least in this population with pre-COPD, actigraphy and in-laboratory physiological assessment and respiratory questionnaires are at best poorly correlated.

Overall, in this population at risk for COPD due to prolonged occupational exposure to SHS but with no airflow obstruction, actigraphy seems to provide distinct information about daily physical activity and functionality that are not captured by in-laboratory lung function, exercise testing, or standardized respiratory questionnaires. A possible explanation for this observation is that in-laboratory physical activity assessments may provide an estimate of the maximum functional *capacity*, while actigraphy may be more informative of the real-world level of physical activity that represents the functional *status*. Although functional capacity represents an individual's maximum capacity to perform physical activity, functional status may best represent the activities people actually do during the course of their daily lives,[18,19] and in this population with pre-COPD, while the functional capacity may be preserved, the functional status is reduced, resulting in poor correlation. In past studies of this cohort, we have demonstrated the cohort to have abnormal diffusing capacity that was associated with lower exercise tolerance despite having preserved spirometry.[11,12] In the same cohort and also another cohort of direct smokers with preserved spirometry (those with Global Initiative for Obstructive Lung Disease [GOLD] 0 category), we have shown that air trapping, as manifested by abnormal lung volumes (RV/TLC), could be present and is predictive of lower exercise tolerance.[9,20] On the other hand, patients with overt COPD have airflow obstruction in addition to any potential abnormal diffusing capacity or lung volumes (air trapping) that they may experience. This airflow obstruction may reduce exercise capacity more remarkably and affect functional status, causing

those with overt COPD to have a physical activity status that matches their maximum capacity. Corroborating this hypothesis are the reports that in overt COPD, activity monitor outputs have predictive value in the assessment of disease severity, exercise capacity, and healthcare utilization patients.[21-23] Thus, it is possible that the association of actigraphy with in-laboratory physiological assessment and respiratory symptoms is varied across COPD disease spectrum and improves as disease severity worsens. These findings also implicate measurement of physical activity status by actigraphy as a useful tool for detection of early disease in COPD when the maximum functional capacity, measured by in-laboratory testing, may be less affected. Further research involving larger cohorts and across the entire COPD disease spectrum would be needed to better understand potential diagnostic application of actigraphy for assessment of patients' real-world physical status.

Another intriguing finding in our study was that both unsupervised cluster analysis and principal component regression approaches showed that actigraphy variables to group into two distinct clusters of (1) distance, energy, and activity time domains and (2) sedentary time domain. Remarkably, these clusters did not show a simple inverse relationship, were rather poorly correlated, and grouped together differently with various measures of in-laboratory cardiopulmonary functional assessment and self-reported symptomatology, indicating that they are influenced by distinct physiological processes. These features of the clustering of the sedentary and activity domains of actigraphy indicate that actigraphy activity and sedentary domains are informative of distinct outcomes, which are only partially and incompletely captured by outcome measures of respiratory questionnaires and in-laboratory lung function and exercise testing. Our findings are consistent with the findings of other studies in which the

sedentary domain of actigraphy were predictive of metabolic outcomes and risk of mortality independent of the activity domain.[24,25]

Given that activity monitors can provide objective and reliable data for assessment of health status and disease activity by measuring the real-world physical activity level,[4-7,26,27] their use has grown significantly in various fields of medicine over the past several years. Actigraphy has become an essential tool in sleep medicine and is increasingly used in the clinical care of patients with sleep and circadian rhythm abnormalities.[28-30] Through objective measurement of sleep parameters (based on movement) in patient's home environment, actigraphy provides relevant data with greater external validity compared with self-report and even compared to in-laboratory polysomnography for certain sleep parameters.[31,32] Actigraphy has also been utilized in telerehabilitation as an alternative to the in-center pulmonary rehabilitation for patients with COPD,[33-35] or for other rehabilitation programs,[21,22,36,37] providing measures of physical activity data while the patients are educated and encouraged to participate in daily exercises such as jogging on the treadmill or bicycling. Actigraphy could also allow for high throughput evaluation of large numbers of patients/subjects in both clinical evaluation and clinical research including clinical trials,[38-40] particularly during the COVID-19 pandemic, which has presented major challenges with in-person and in-laboratory evaluation and testing. Actigraphy is also of particular interest because of the deployment of smartphone applications (App) for tracking daily physical activity, which has been popular due to the common possession of smartphones and, although may exist, they could in principle provide the same measurements as actigraphy does without the need of wearing an exact electronic device.[41,42] While smartphone App will gradually become more accurate and sustainable from the advancement in software engineering and accessibility of smartphones, actigraphy remains the current method of choice for scaling up

standardized physical activity data collection. Altogether, actigraphy could have diverse applications in various medical fields, and thus understanding how it is associated with the current physical activity assessment tools and its limitation are important. Further research with inclusion of larger number of subjects across the entire COPD disease spectrum would be needed to better understand actigraphy application in real-world assessment of physical status of patients.

Limitations

Our study had several limitations. First, this study had a relatively small sample of 61 participants, which may have diminished the learning power of our machine learning approach for variable selection as well as the statistical power of the final regression modeling. This study was nested in a larger cohort study of the cardiopulmonary health effects of prolonged occupational exposure to SHS, the inclusion and exclusion criteria of which were more stringent and limited our ability to expand our sample size. Nevertheless, our sample size was comparable if not larger relative to other published studies that have characterized physical activity using extensive examination including CPET,[43-45] and thus provides similar statistical power.

Second, actigraphy data were only collected for five days. In adults, a minimum of 3 to 5 days of accelerometer monitoring is usually considered appropriate to obtain reliable estimates of physical activity.[10] On the other hand, a longer measurement period of more than 7 days may be desirable to obtain reliable estimates of sedentary behavior since sedentary behavior is more difficult to capture as it might vary more on a day-by-day basis than other activities performed on a higher intensity-level.[46] However, we aimed to avoid the potential biases that usually occur with variations in level of activity, either increased or decreased, during the weekends by asking the participants to wear the activity monitor for 5 weekdays at the beginning of their working

week and the beginning of their day to capture their usual and customary level of activity. In fact, our findings demonstrate a greater number of significant associations for the sedentary domain, suggesting the adequacy of the timeframe within the population studied.

Third, a single activity monitor placed on the waist may not have detected all physical activity since accelerometers have a key limitation in that they are insensitive to certain types of movements, especially non-ambulatory physical activities with arms and/or limbs. However, studies employing multiple accelerometers to increase the accuracy of predicting energy expenditure reported only marginal improvements that would not be justifiable by the increased burden associated with wearing multiple accelerometers.[47,48]

Conclusion

In conclusion, in this study of a population with early obstructive lung disease or pre-COPD due to prolonged exposure to SHS, we showed that outpatient actigraphy provides "real-world" patient-centered outcomes that objectively inform of patients' physical activity status. We further showed that the activity and sedentary domains of actigraphy are divergent and provide distinct information that may likely represent the different physiologic processes such as anaerobic and aerobic fitness highlighted by those domains. Furthermore, in this population with pre-COPD, actigraphy measures are not associated with and could not be entirely explained by measures of standardized respiratory questionnaires, lung function, or exercise testing, and thus provide added physical activity measures that could be of value as objective patient-centered outcomes in observational and interventional research studies.

REFERENCES

References

1. Kocks JW, Asijee GM, Tsiligianni IG, Kerstjens HA, van der Molen T. Functional status measurement in COPD: a review of available methods and their feasibility in primary care. *Prim Care Respir J.* 2011;20(3):269-275.
2. Kim J, Lee CH, Lee MG, et al. Acute Exacerbation According to GOLD 2017 Categories in Patients with Chronic Obstructive Pulmonary Disease. *Arch Bronconeumol.* 2019;55(8):414-420.
3. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int J Behav Nutr Phys Act.* 2008;5:56.
4. Pevnick JM, Birkeland K, Zimmer R, Elad Y, Kedan I. Wearable technology for cardiology: An update and framework for the future. *Trends Cardiovasc Med.* 2018;28(2):144-150.
5. Chiauzzi E, Rodarte C, DasMahapatra P. Patient-centered activity monitoring in the self-management of chronic health conditions. *BMC Med.* 2015;13:77.
6. Ferriolli E, Skipworth RJ, Hendry P, et al. Physical activity monitoring: a responsive and meaningful patient-centered outcome for surgery, chemotherapy, or radiotherapy? *J Pain Symptom Manage.* 2012;43(6):1025-1035.
7. Gresham G, Schrack J, Gresham LM, et al. Wearable activity monitors in oncology trials: Current use of an emerging technology. *Contemp Clin Trials.* 2018;64:13-21.
8. Bunn JA, Navalta JW, Fountaine CJ, Reece JD. Current State of Commercial Wearable Technology in Physical Activity Monitoring 2015-2017. *Int J Exerc Sci.* 2018;11(7):503-515.
9. Arjomandi M, Zeng S, Geerts J, et al. Lung volumes identify an at-risk group in persons with prolonged secondhand tobacco smoke exposure but without overt airflow obstruction. *BMJ Open Respir Res.* 2018;5(1):e000284.
10. Trost SG, McIver KL, Pate RR. Conducting accelerometer-based activity assessments in field-based research. *Med Sci Sports Exerc.* 2005;37(11 Suppl):S531-543.
11. Arjomandi M, Haight T, Redberg R, Gold WM. Pulmonary function abnormalities in never-smoking flight attendants exposed to secondhand tobacco smoke in the aircraft cabin. *J Occup Environ Med.* 2009;51(6):639-646.
12. Arjomandi M, Haight T, Sadeghi N, Redberg R, Gold WM. Reduced exercise tolerance and pulmonary capillary recruitment with remote secondhand smoke exposure. *PLoS One.* 2012;7(4):e34393.
13. Jones PW, Harding G, Berry P, Wiklund I, Chen WH, Kline Leidy N. Development and first validation of the COPD Assessment Test. *Eur Respir J.* 2009;34(3):648-654.
14. Fletcher CM, Elmes PC, Fairbairn AS, Wood CH. The significance of respiratory symptoms and the diagnosis of chronic bronchitis in a working population. *Br Med J.* 1959;2(5147):257-266.
15. Jenkinson C, Layte R. Development and testing of the UK SF-12 (short form health survey). *J Health Serv Res Policy.* 1997;2(1):14-18.

16. Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc.* 2003;35(8):1381-1395.
17. Hajiro T, Nishimura K, Jones PW, et al. A novel, short, and simple questionnaire to measure health-related quality of life in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* 1999;159(6):1874-1878.
18. Leidy NK. Functional status and the forward progress of merry-go-rounds: toward a coherent analytical framework. *Nurs Res.* 1994;43(4):196-202.
19. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *Jama.* 1995;273(1):59-65.
20. Arjomandi M, Zeng S, Barjaktarevic I, et al. Radiographic Lung Volumes Predict Progression to COPD in Smokers with Preserved Spirometry in SPIROMICS. *Eur Respir J.* 2019.
21. Bauldoff GS, Ryan-Wenger NA, Diaz PT. Wrist Actigraphy Validation of Exercise Movement in COPD. *Western Journal of Nursing Research.* 2007;29(7):789-802.
22. Demeyer H, Burtin C, Van Remoortel H, et al. Standardizing the analysis of physical activity in patients with COPD following a pulmonary rehabilitation program. *Chest.* 2014;146(2):318-327.
23. Rabinovich RA, Louvaris Z, Raste Y, et al. Validity of physical activity monitors during daily life in patients with COPD. *Eur Respir J.* 2013;42(5):1205-1215.
24. 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.
25. Healy GN, Dunstan DW, Salmon J, Shaw JE, Zimmet PZ, Owen N. Television time and continuous metabolic risk in physically active adults. *Med Sci Sports Exerc.* 2008;40(4):639-645.
26. Rowe VT, Neville M. Measuring Reliability of Movement With Accelerometry: Fitbit(®) Versus ActiGraph(®). *Am J Occup Ther.* 2019;73(2):7302205150p7302205151-7302205150p7302205156.
27. Yang CC, Hsu YL. A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors (Basel).* 2010;10(8):7772-7788.
28. Morgenthaler T, Alessi C, Friedman L, et al. Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007. *Sleep.* 2007;30(4):519-529.
29. Smith MT, McCrae CS, Cheung J, et al. Use of Actigraphy for the Evaluation of Sleep Disorders and Circadian Rhythm Sleep-Wake Disorders: An American Academy of Sleep Medicine Clinical Practice Guideline. *Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine.* 2018;14(7):1231-1237.
30. Sadeh A, Acebo C. The role of actigraphy in sleep medicine. *Sleep Medicine Reviews.* 2002;6(2):113-124.
31. Kushida CA, Chang A, Gadkary C, Guilleminault C, Carrillo O, Dement WC. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Medicine.* 2001;2(5):389-396.
32. Martin JL, Hakim AD. Wrist actigraphy. *Chest.* 2011;139(6):1514-1527.

33. Vasilopoulou M, Papaioannou AI, Kaltsakas G, et al. Home-based maintenance tele-rehabilitation reduces the risk for acute exacerbations of COPD, hospitalisations and emergency department visits. *European Respiratory Journal.* 2017;49(5):1602129.
34. Tabak M, Vollenbroek-Hutten MMR, van der Valk PD, van der Palen J, Hermens HJ. A telerehabilitation intervention for patients with Chronic Obstructive Pulmonary Disease: a randomized controlled pilot trial. *Clinical Rehabilitation.* 2013;28(6):582-591.
35. Cox NS, Dal Corso S, Hansen H, et al. Telerehabilitation for chronic respiratory disease. *Cochrane Database of Systematic Reviews.* 2021(1).
36. Thorup C, Hansen J, Grønkjær M, et al. Cardiac Patients' Walking Activity Determined by a Step Counter in Cardiac Telerehabilitation: Data From the Intervention Arm of a Randomized Controlled Trial. *J Med Internet Res.* 2016;18(4):e69.
37. Kline PW, Melanson EL, Sullivan WJ, et al. Improving Physical Activity Through Adjunct Telerehabilitation Following Total Knee Arthroplasty: Randomized Controlled Trial Protocol. *Physical Therapy.* 2019;99(1):37-45.
38. Perry B, Herrington W, Goldsack JC, et al. Use of Mobile Devices to Measure Outcomes in Clinical Research, 2010-2016: A Systematic Literature Review. *Digit Biomark.* 2018;2(1):11-30.
39. Taibi DM, Vitiello MV, Barsness S, Elmer GW, Anderson GD, Landis CA. A randomized clinical trial of valerian fails to improve self-reported, polysomnographic, and actigraphic sleep in older women with insomnia. *Sleep Medicine.* 2009;10(3):319-328.
40. Peterson BT, Chiao P, Pickering E, et al. Comparison of actigraphy and polysomnography to assess effects of zolpidem in a clinical research unit. *Sleep Med.* 2012;13(4):419-424.
41. Dunton GF, Dzubur E, Kawabata K, Yanez B, Bo B, Intille S. Development of a smartphone application to measure physical activity using sensor-assisted self-report. *Front Public Health.* 2014;2:12-12.
42. Glynn LG, Hayes PS, Casey M, et al. Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. *Br J Gen Pract.* 2014;64(624):e384-391.
43. Van Remoortel H, Raste Y, Louvaris Z, et al. Validity of six activity monitors in chronic obstructive pulmonary disease: a comparison with indirect calorimetry. *PLoS One.* 2012;7(6):e39198.
44. Van Remoortel H, Giavedoni S, Raste Y, et al. Validity of activity monitors in health and chronic disease: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity.* 2012;9(1):84.
45. Demeyer H, Burtin C, Hornikx M, et al. The Minimal Important Difference in Physical Activity in Patients with COPD. *PLoS One.* 2016;11(4):e0154587.
46. Aadland E, Ylvisåker E. Reliability of Objectively Measured Sedentary Time and Physical Activity in Adults. *PLoS One.* 2015;10(7):e0133296.
47. Swartz AM, Strath SJ, Bassett DR, Jr., O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Med Sci Sports Exerc.* 2000;32(9 Suppl):S450-456.

48. Melanson EL, Jr., Freedson PS. Validity of the Computer Science and Applications, Inc. (CSA) activity monitor. *Med Sci Sports Exerc.* 1995;27(6):934-940.

FIGURE LEGENDS

Figure 1- Actigraphy variable selection. A schematic to show the process of variable selection by combining machine learning and literature guided approaches. Machine learning approach was done by implementing random forest and lasso regression. The top ranked 10 variables from each method were done used against a list of variables extracted from the available literature, and a final "focused" set of variables representing all actigraphy domains were selected. "Highest" and "Mean" for actigraphy variables represent the highest and mean of 5-day activity monitoring measurements.

Figure 2- Correlation among final set of actigraphy variables. Spearman correlation coefficient in clustering heat map of the final selection of actigraphy variables. Visualization shows that almost all of actigraphy variables are highly correlated.

Figure 3 – Hierarchical clustering analysis and heatmap representation of final selection of actigraphy variables with questionnaire, PFT, and CPET measures. We performed cluster analysis using Spearman correlation coefficients as the distance metric. The clustered relationships were present as the dendrograms on the top and left size of the heat map. Two distinct clusters of actigraphy variables were grouped together from the three domains (time, distance, energy). The two clusters were based on activity related (such as total energy spent in Kcals, number of steps walked, etc.) and sedentary related (such as sedentary time, break, etc.) respectively. Each row represents one subject. Grey box are missing values.

TABLES

Table 1- Subject characteristics.

Characteristics	N=61
Demographics	
Age (years)	65.7±7.5
Female sex [n(%)]	57(93%)
Height (cm)	165.5±7.5
Weight (kg)	65.5±10.7
BMI (kg/m^2)	23.9±3.2
SHS (years)	16.0 [6.0-25.0]
Smoking (pack-year)	0 [0-0]
Lung Function	
FEV_1 (L)	2.50±0.52
FEV_1 (% predicted)	103±17
FVC (L)	3.33±0.69
FVC (% predicted)	106±16
FEV_1/FVC [%]	75±5
FEV_1/FVC (% predicted)	98±6
TLC (L)	5.31±0.88
TLC (% predicted)	100±11
RV (L)	1.97±0.37
RV (% predicted)	91±15
RV/TLC [%]	38±6
RV/TLC (% predicted)	92±13
DcoSB adjusted for Hgb (% predicted)	67±28
Exercise	
VO_{2Peak} (L/min)	1.399±0.348
VO_{2Peak} (% predicted)	104±18
$Watts_{Peak}$ (watts)	120.9±34.2

VCO_{2Peak} (L/min)	1.7±0.4
O_2-$Pulse_{Peak}$ (L/beat)	9.6±2.6
V_{EPeak} (L/min)	56.30±13.86
V_{EPeak} (% predicted)	61±14
RER	33.7±8.5
RR_{Peak} (breaths/min)	33.7±8.5
V_{TPeak} (L)	1.70±0.42
V_{TPeak} (% predicted)	96±17
V_E/VO_{2Peak}	38.4±6.1
V_E/VCO_{2Peak}	32.4±4.3
HR_{Peak} (beats/min)	146±18
Systolic BP_{Peak} (mmHg)	189±19
Diastolic BP_{Peak} (mmHg)	85±10
VO_2 at Anaerobic Threshold (VO_{2AT}) (L/min)	1.16±0.33
V_E/VCO_2 at Anaerobic Threshold	31.8±4.1
O_2-Pulse at Anaerobic Threshold (L/beat)	9.9±2.8
Questionnaires	
AQ20	1±2
CAT	6.6±5.8
SF-12 [Total]	29.5±6.1
SF-12 (PCS)	37.1±5.2
SF-12 (MCS)	55.8±8.8
IPAQ Category 3: High [n (%)]	44 (72%)
Actigraphy Domains	
Time	
Activity Domain	
Highest Total MVPA (min)	50.4±37.79
Mean Avg. MVPA/hour (min/hr)	3.87±2.64
Sedentary Domain	

Highest Sedentary Time (min)	363.6±38.6
Highest No. of Sedentary Breaks	7.6±3.4
No. of Sedentary Breaks/Sedentary Time	0.02±0.01
No. of Sedentary Breaks/Avg. Step Count	0.66±0.50
Distance	
Highest Max. Step Count	24.1±3.5
Mean Avg. Step Count (steps/min)	15.5±8.5
Highest Vector Magnitude Max. Count	2,081±524
Energy	
Highest Total Kcals (Kcals)	354.4±213.8
Mean Total Kcals (Kcals)	242.4±126.1
Mean Avg. Kcal/hour (Kcals/hour)	10.2±5.3
Highest MET (METs)	1.39±0.32
Mean MET (METs)	1.24±0.17

Footnote: Demographics, lung function, exercise, questionnaire scores, and actigraphy indices in this cohort of subjects with preserved spirometry. Data is presented as mean±standard deviation or number of subjects with positive value for the variable (n) out of the total number of subjects (N) and percentage of subjects (%). “Highest” and “Mean” for actigraphy variables represent the highest and mean of 5-day activity monitoring measurements. Abbreviations- BMI: body mass index; FEV_1: forced expiratory volume in 1 second; FVC: forced vital capacity; TLC: total lung capacity; RV: residual volume; DcoSB: single-breath diffusing capacity of carbon monoxide; Hgb: hemoglobin; VO_{2Peak}: maximum oxygen uptake; $Watts_{Max}$: maximum work stage completed in watts; VCO_{2Peak}: maximum carbon dioxide production; O_2pulse_{Peak}: oxygen uptake per heartbeat at maximum exercise; V_{EPeak}: maximum minute ventilation value; RER: respiratory exchange ratio (VCO_2/ VO_2) at maximum exercise; RR_{Peak}: maximum respiratory rate; V_{TPeak}:

maximum tidal volume; AQ20: Airways Questionnaire 20; CAT: COPD Assessment Test; SF-12: 12-Item Short Form Health Survey; PCS-12: physical component summary score of SF-12; MCS-12: mental component summary score of SF-12; IPAQ: International Physical Activity Questionnaire; METs: metabolic equivalents; MVPA: moderate to vigorous physical activity; CPM: counts per minute.

Table 2- Definitions of final selection of actigraphy variables

Selected Variables	Definition and Accelerometer Cut Point
Time Domains	
Activity	
Total MVPA	PA that involves EE above 3.0 METs; time (in min) when CPM ≥ 1952
Avg. MVPA/hour	Average time spent in MVPA; total MVPA (in min)/total time worn (in hours)
Sedentary	
Sedentary Time	Time spent in PA that does not increase EE substantially above the resting level (1.0 - 1.5 METs); time (in min) when CPM < 100
No. of Sedentary Break	Activity count change from <100 to ≥ 100 counts. A break only occurs if the change in count is sustained for more than 2 min
Distance Domain	
Max. Step Count	Maximum of the daily maximum number of steps taken per epoch (10 sec)
Avg. Step Count	Average number of steps taken per min; total step count/total time worn (in min)
Vector Magnitude CPM	Vector summation of CPM across all 3 axes (vertical, horizontal and perpendicular)
Energy Domain	
Avg. Kcal/hour	Measure of EE; a kcal is the amount of energy required to raise the temperature of a liter of water one degree centigrade at sea level;

	total kcal/total time worn (in hours)
MET	Measure of EE; MET is the ratio of a person's working metabolic rate compared to their resting metabolic rate; A person sitting quietly would be considered 1 MET (1 MET is equivalent to a caloric consumption of 1kcal/kg/hour); sedentary behavior 1.0-1.5 METs, light- intensity PA 1.6-2.9 METs, moderate-intensity PA ≥3.0 to <6.0 METs, vigorous-intensity ≥6.0 to <9.0 METs and very vigorous PA is ≥9.0 METs

Footnote: Definitions and literature citations for actigraphy variables selected using the literature-guided approach. Abbreviations- MVPA: moderate to vigorous physical activity; MET: metabolic equivalent; CPM: counts per minute; PA: physical activity; EE: energy expenditure.

Table 3- Results of the principal component analysis.

	Principal Component					
Variable	**1**	**2**	**3**	**4**	**5**	**6**
Highest Total MVPA	**0.92**	-0.07	0.01	-0.08	0.06	-0.07
Mean Avg. MVPA/hour	**0.94**	-0.07	-0.02	-0.05	0.02	0
Highest Sedentary Time	**-0.54**	0.22	**0.66**	-0.26	0.14	-0.09
Highest No. of Sedentary Breaks	-0.38	0.34	**0.67**	-0.20	0.20	0.15
Highest Max. Step Counts	**0.56**	-0.33	0.38	-0.08	-0.39	-0.10
Mean Avg. Step Count	**0.91**	-0.15	-0.10	0.01	0.05	-0.09
Highest Vector Magnitude Max. Counts	0.40	-0.09	0.26	0.15	-0.7	0.45
Highest Total Kcals	**0.91**	0.25	0.09	-0.11	0.13	0.02
Mean Total Kcals	**0.92**	0.27	-0.02	-0.04	0.10	0.08
Mean Avg. Kcal/hour	**0.92**	0.29	-0.01	-0.03	0.10	0.08
Highest MET	**0.94**	0.11	0.10	-0.15	0.11	-0.03
Mean MET	**0.96**	0.15	0.03	-0.11	0.09	0.05
Age (years)	0.15	-0.19	0.23	**0.77**	0.43	0.18
Sex	0.02	**0.77**	-0.11	0.30	-0.25	0.22
Height (cm)	-0.16	**0.84**	-0.10	-0.01	-0.17	-0.25
Weight (kg)	-0.04	0.4	0.03	0.21	0.01	0.28
Total Time Worn (min)	0.41	-0.17	0.45	0.40	-0.13	-0.49

Footnote: This table shows the value of Pearson's correlation coefficient for each variable in our focused dataset. Bold type indicates values <0.5. Abbreviations- MET: metabolic equivalents; MVPA: moderate to vigorous physical activity; CPM: counts per minute.

Table 4- Results of principal component regression analysis.

	PC1 (activity component)				PC3 (sedentary component)				Model Error
Outcomes	Estimate	Lower CI	Upper CI	P-value	Estimate	Lower CI	Upper CI	P-value	RMSE
Lung function									
FEV_1 (L)	0.005	-0.032	0.042	0.808	**-0.094**	**-0.183**	**-0.006**	**0.042**	0.39
FVC (L)	-0.002	-0.050	0.046	0.924	-0.107	-0.222	0.009	0.075	0.51
FEV_1/FVC [%]	0.002	-0.002	0.007	0.284	-0.004	-0.014	0.006	0.458	0.05
RV (L)	-0.005	-0.037	0.028	0.781	-0.045	-0.131	0.04	0.305	0.32
TLC (L)	-0.016	-0.077	0.044	0.597	-0.143	-0.288	0.001	0.057	0.32
RV/TLC [%]	-0.001	-0.006	0.004	0.796	0.003	-0.010	0.017	0.617	0.05
Exercise									
VO_{2Peak} (L/min)	0.002	-0.020	0.025	0.838	0.016	-0.039	0.070	0.575	0.23
O_2-$Pulse_{Peak}$ (L/beat)	0.179	-0.021	0.379	0.088	0.131	-0.327	0.590	0.578	1.52
V_{EPeak} (L/min)	0.133	-1.217	1.482	0.848	0.231	-3.144	3.607	0.894	11.25
RR_{Peak} (breaths/min)	-0.611	-1.390	0.167	0.130	1.741	-0.171	3.654	0.081	7.92
V_{TPeak} (L)	0.006	-0.022	0.035	0.660	-0.062	-0.131	0.008	0.087	0.29
V_E/VCO_{2Peak}	-0.147	-0.625	0.331	0.551	0.314	-0.882	1.509	0.610	3.99
$Watts_{Peak}$ (watts)	1.067	-0.939	3.072	0.302	-0.104	-5.029	4.821	0.967	20.40
Systolic BP_{Peak} (mmHg)	0.112	-1.639	1.863	0.901	-0.376	-4.696	3.945	0.865	17.51
Diastolic BP_{Peak} (mmHg)	0.515	-0.488	1.518	0.320	0.042	-2.434	2.518	0.974	10.04

VO_2 at Anaerobic Threshold (L/min)	**0.032**	**0.005**	**0.059**	**0.026**	0.005	-0.065	0.076	0.885	0.25
VCO_{2Slope} (L/min)	-0.005	-0.017	0.007	0.405	-0.002	-0.031	0.028	0.917	0.12
O_2-$Pulse_{Slope}$ (L/beat)	0.071	-0.016	0.158	0.118	0.127	-0.090	0.344	0.259	0.73
V_{ESlope} (L/min)	-0.516	-1.163	0.132	0.127	0.185	-1.435	1.805	0.824	5.40
RR_{Slope} (breaths/min)	-0.183	-0.714	0.349	0.503	0.128	-1.177	1.1434	0.848	5.41
V_{TSlope} (L)	-0.016	-0.046	0.015	0.315	-0.025	-0.099	0.049	0.507	0.31
Systolic BP_{Slope} (mmHg)	8.095	-9.451	25.641	0.370	18.368	-24.726	61.462	0.408	178.45
Diastolic BP_{Slope} (mmHg)	0.205	-0.768	1.177	0.682	-0.936	-3.325	1.453	0.446	9.89
Questionnaires									
AQ20	-0.150	-0.336	0.036	0.120	0.261	-0.193	0.715	0.266	1.86
CAT	-0.332	-0.980	0.316	0.322	-0.101	-1.697	1.496	0.902	5.31
SF-12 [Total]	-0.319	-0.880	0.241	0.270	**-1.471**	**-2.840**	**-0.103**	**0.041**	5.60
SF-12 (PCS)	-0.026	-0.513	0.462	0.918	-0.451	-1.641	0.738	0.461	4.87
SF-12 (MCS)	-0.791	-1.568	-0.014	0.052	**-2.102**	**-3.999**	**-0.204**	**0.035**	7.76

FIGURES

Figure 1- Actigraphy variable selection. A schematic to show the process of variable selection by combining machine learning and literature guided approaches. Machine learning approach was done by implementing random forest and lasso regression. The top ranked 10 variables from each method were done used against a list of variables extracted from the available literature, and a final "focused" set of variables representing all actigraphy domains were selected. "Highest" and "Mean" for actigraphy variables represent the highest and mean of 5-day activity monitoring measurements.

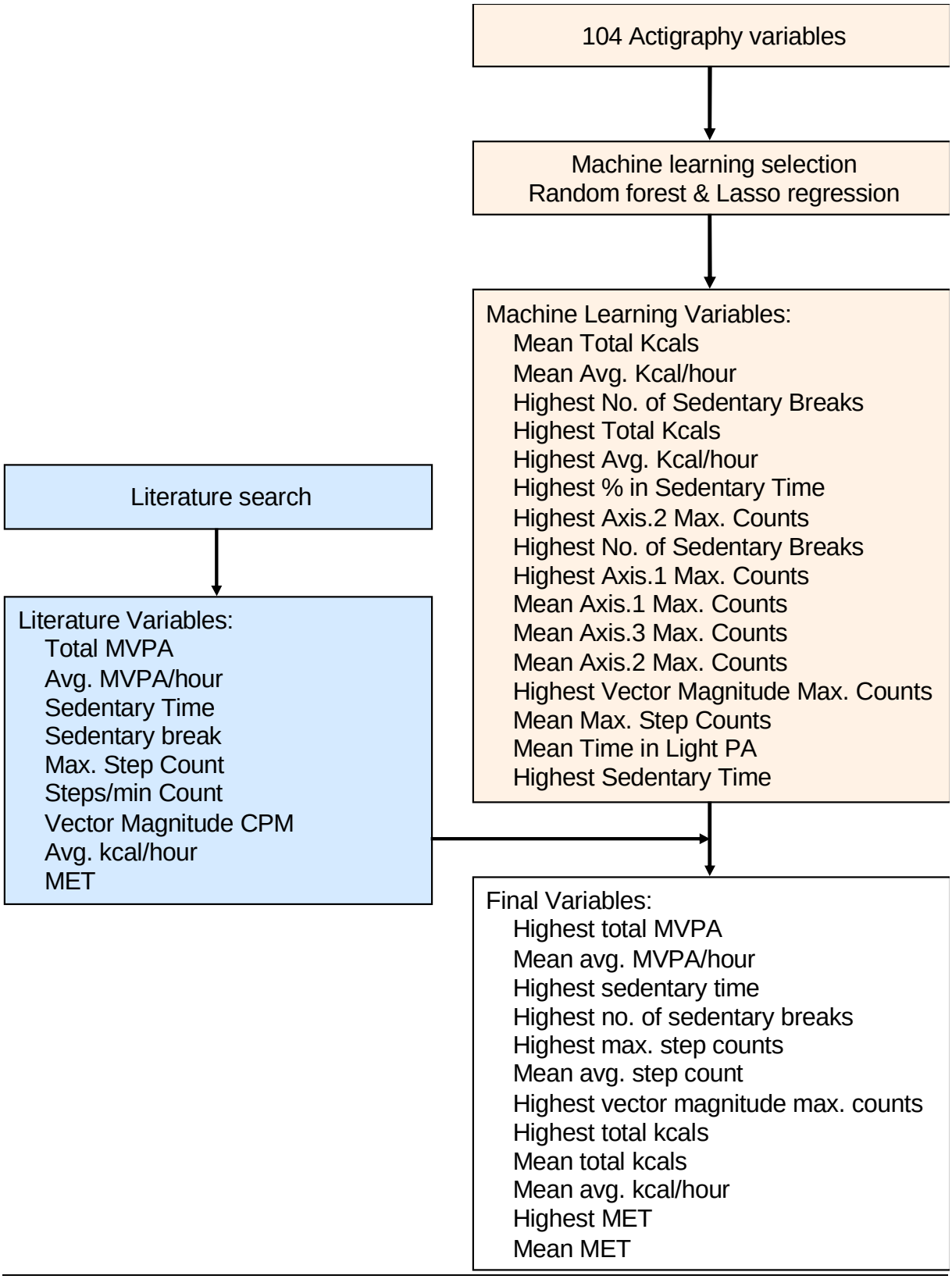

104 Actigraphy variables
Machine learning selection
Random forest & Lasso regression
Machine Learning Variables:
Mean Total Kcals
Mean Avg. Kcal/hour
Highest No. of Sedentary Breaks
Highest Total Kcals
Highest Avg. Kcal/hour
Highest % in Sedentary Time
Highest Axis.2 Max. Counts
Highest No. of Sedentary Breaks
Highest Axis.1 Max. Counts
Mean Axis.1 Max. Counts
Mean Axis.3 Max. Counts
Mean Axis.2 Max. Counts
Highest Vector Magnitude Max. Counts
Mean Max. Step Counts
Mean Time in Light PA
Highest Sedentary Time
Literature search
Literature Variables:
Total MVPA
Avg. MVPA/hour
Sedentary Time
Sedentary break
Max. Step Count
Steps/min Count
Vector Magnitude CPM
Avg. kcal/hour
MET
Final Variables:
Highest total MVPA
Mean avg. MVPA/hour
Highest sedentary time
Highest no. of sedentary breaks
Highest max. step counts
Mean avg. step count
Highest vector magnitude max. counts
Highest total kcals
Mean total kcals
Mean avg. kcal/hour
Highest MET
Mean MET

Figure 2- Correlation among final set of actigraphy variables. Spearman correlation coefficient in clustering heat map of the final selection of actigraphy variables. Visualization shows that almost all of actigraphy variables are highly correlated.

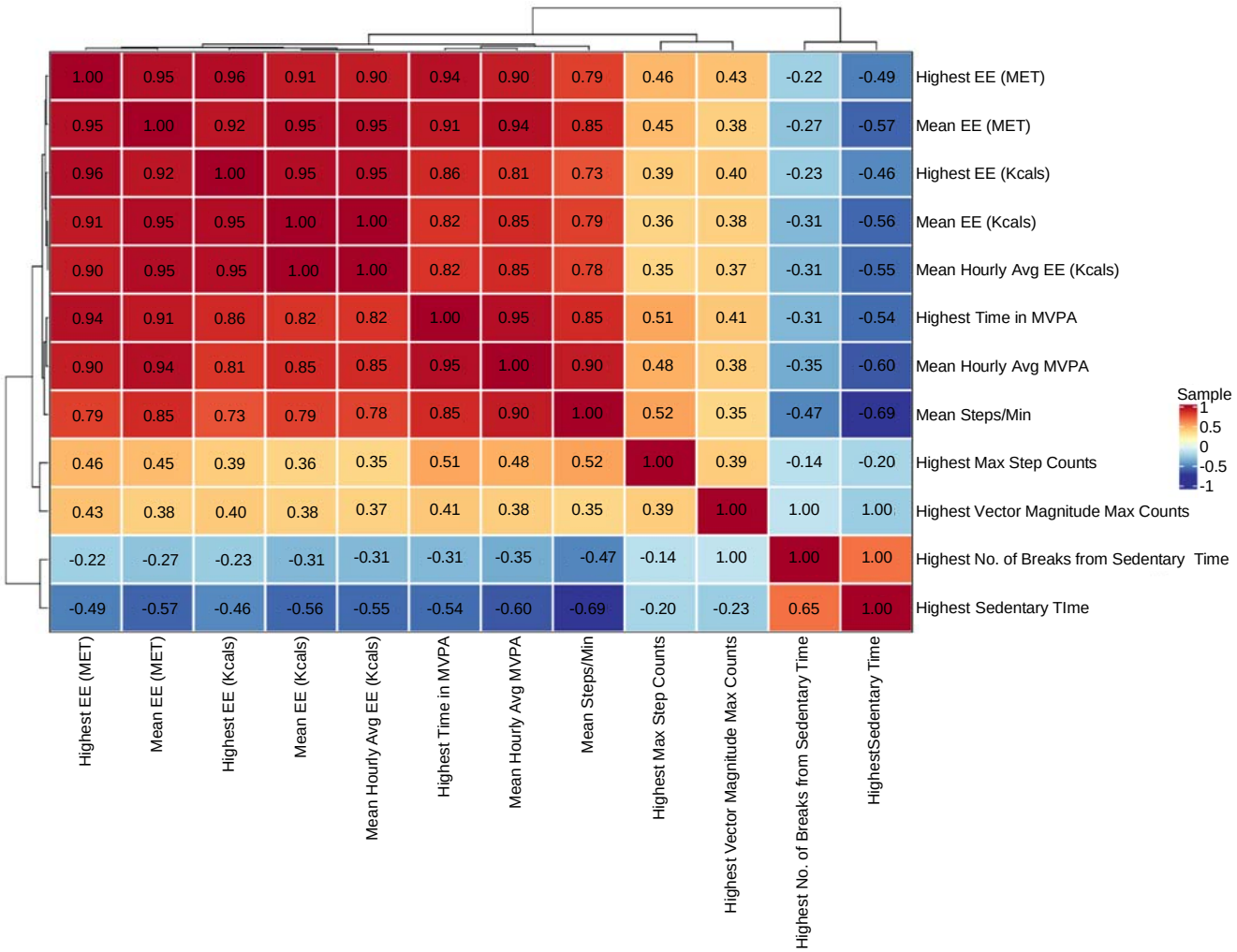
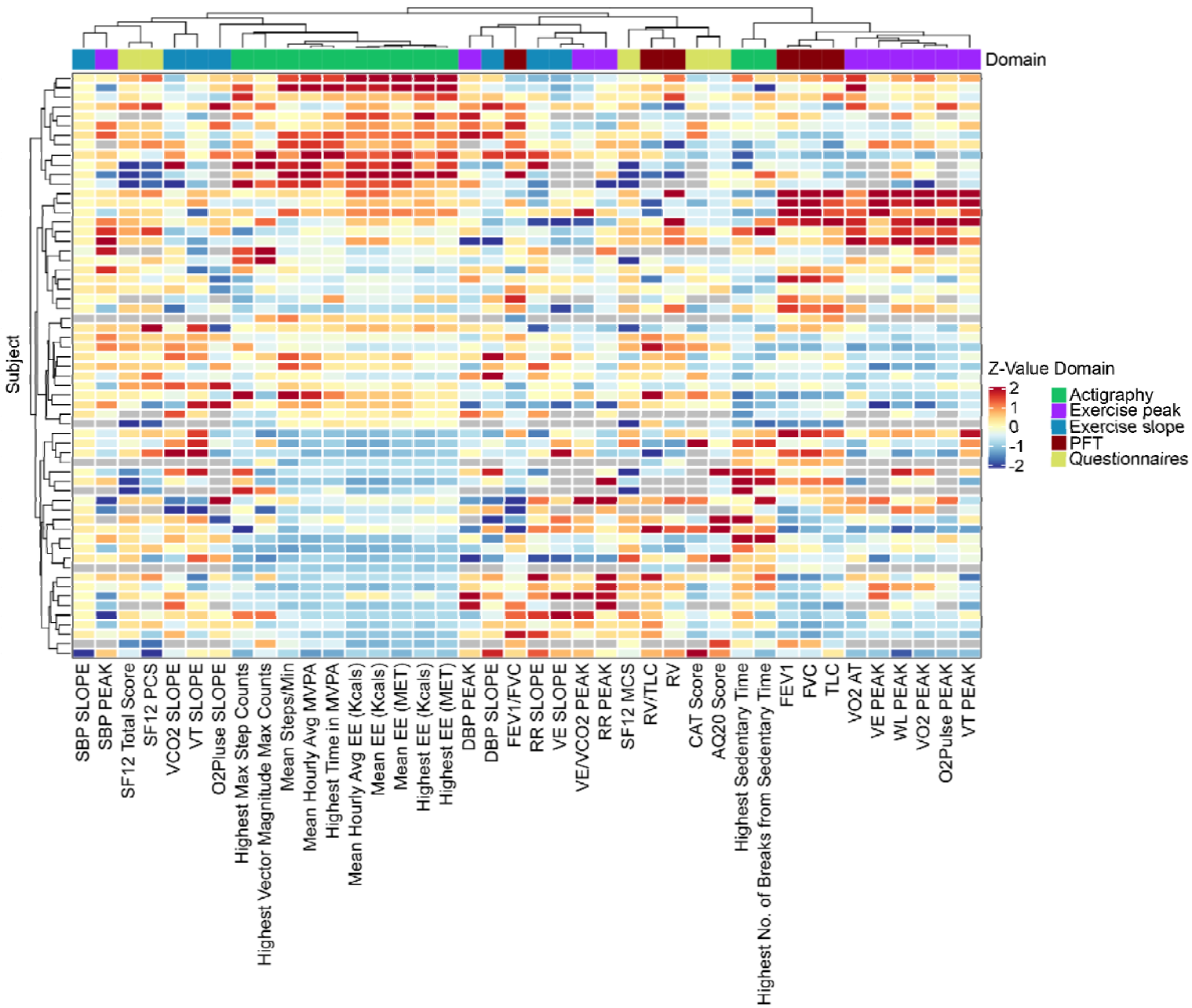


Figure 3- Hierarchical clustering analysis and heatmap representation of final selection of actigraphy variables with questionnaire, PFT, and CPET measures. We performed cluster analysis using Spearman correlation coefficients as the distance metric. The clustered relationships were present as the dendrograms on the top and left size of the heat map. Two distinct clusters of actigraphy variables were grouped together from the three domains (time, distance, energy). The two clusters were based on activity related (such as total energy spent in Kcals, number of steps walked, etc.) and sedentary related (such as sedentary time, break, etc.) respectively. Each row represents one subject. Grey box are missing values.


Domain
Subject
Z-Value Domain
2
1
0
-1
-2
Actigraphy
Exercise peak
Exercise slope
PFT
Questionnaires
SBP SLOPE
SBP PEAK
SF12 Total Score
SF12 PCS
VCO2 SLOPE
VT SLOPE
O2Pluse SLOPE
Highest Max Step Counts
Highest Vector Magnitude Max Counts
Mean Steps/Min
Mean Hourly Avg MVPA
Highest Time in MVPA
Mean Hourly Avg EE (Kcals)
Mean EE (Kcals)
Mean EE (MET)
Highest EE (Kcals)
Highest EE (MET)
DBP PEAK
DBP SLOPE
FEV1/FVC
RR SLOPE
VE SLOPE
VE/VCO2 PEAK
RR PEAK
SF12 MCS
RV/TLC
RV
CAT Score
AQ20 Score
Highest Sedentary Time
Highest No. of Breaks from Sedentary Time
FEV1
FVC
TLC
VO2 AT
VE PEAK
WL PEAK
VO2 PEAK
O2Pulse PEAK
VT PEAK