medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

TITLE PAGE

Article Title:

Actigraphy Informs Distinct Patient-Centered Outcomes in Pre-COPD

Running Head:

Actigraphy in assessment of pre-COPD

Author List:

Lemlem Weldemichael, MD [1,2,3] *, Jianhong Chen, MS [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

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

* These authors contributed equally to this work.

Email Addresses:

LW: weldemichael.lemlem@gmail.com

JC: jianhong.chen@ucsf.edu

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

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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: LW, JC, JG, BG, SZ, MA.

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

Edited the manuscript: LW, JC, 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).

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

What is the key question?

How do real-world activity monitoring outputs correlate with in-laboratory measures of cardiopulmonary function or respiratory symptoms?

What is the bottom line?

Actigraphy provided distinct information about functionality and daily physical activity that was not captured by in-laboratory lung function, exercise testing, or standardized respiratory questionnaires.

Why read on?

Although in-laboratory functional assessments are helpful in evaluating functional *capacity*, actigraphy may be more informative of the real-world functional *status*, and may serve as an additional objective patient-centered outcome in observational and interventional studies aiming to assess functionality.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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,

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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**.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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-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-Pulse$_{Peak}$).

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[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, actigraphy variables were grouped 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. Despite their grouping patterns, the actigraphy variables were significantly associated with only a few in-laboratory measures of cardiopulmonary function or symptoms. For example, the principal component representing activity domains was only significantly associated with the anaerobic threshold (VO_{2AT}), and the principal component representing sedentary domain was only significantly associated with lower airflow (FEV_1) and possibly smaller lung volumes (TLC and FVC), and worse symptoms and quality of life score by one questionnaire (SF-12). Overall, actigraphy provided distinct information about functionality and daily physical activity that was not captured by in-laboratory lung function, exercise testing, or standardized respiratory questionnaires. Thus, while in-laboratory functional assessments may help evaluate functional *capacity*, actigraphy may be more informative of the real-world functional *status*, and may serve as an additional objective patient-centered technique in studies aiming to assess functionality.

Wearable activity monitors that measure the real-world physical activity level in individuals provide objective and reliable data for assessment of health status and disease

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

activity.[4-7] The advantage of using wearable technology for functional assessment stems from its niche ability to provide invaluable data pertaining to the daily functioning and overall health of patients that is otherwise inaccessible and unobtainable through traditional clinical assessments.[18,19] For example, in COPD, activity monitor outputs may have predictive value in the assessment of disease severity as measured by exercise capacity and healthcare utilization.[20,21] A multicenter trial in which several activity monitors were validated against directly measured energy expenditure showed certain accelerometers, including ActiGraph GT3X used in our study, provide great correlations with most measures of exercise capacity in COPD.[22] However, the utility of these activity monitors has not been sufficiently studied in those at risk for COPD but without airflow obstruction (early disease with preserved spirometry or pre-COPD). Similarly, their relationships with other laboratory or clinical outcome modalities of disease assessment such as lung function and exercise performance are not clear. Our study used the literature and machine learning approaches to examine the associations of various domains of actigraphy with in-laboratory and self-reported measures of cardiopulmonary function in people at risk for COPD but with preserved spirometry.

Given their practicality, low cost, and general acceptance, standardized questionnaires are frequently used for assessing self-reports of symptoms and quality of life as well as physical activity. Questionnaires can evaluate a wide range of physical activities, including stationary activities such as weight lifting, over long timeframes. However, they are subjective in nature, suffer from social desirability and recall bias,[23,24] and often tend to over- or under-estimate true physical activity, energy expenditure, and sedentary behavior.[3,25,26] Activity monitors could provide more objective assessments of physical activity. Physical activity scores measured by questionnaires and accelerometry show a weak-to-moderate correlation.[27-29] In our study, the

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

outcome measures of respiratory questionnaires (especially SF-12 and its physical component) were loosely clustered together with the distance, energy, and activity domains of actigraphy, while in PCA regression, SF-12 and its mental component were negatively associated with the sedentary domain of actigraphy. These features of the clustering of the sedentary and activity domains of actigraphy indicates that actigraphy activity and sedentary domains are informative of distinct outcomes, which are only partially and incompletely captured by outcome measures of respiratory questionnaires, consistent with the findings of other studies in which the sedentary domain of actigraphy could predict metabolic outcomes and risk of mortality independent of the activity domain.[30,31]

CPET is a relatively noninvasive in-laboratory physiological test that objectively measures the cardiopulmonary response to exercise and is used to diagnose early cardiopulmonary and metabolic diseases and monitor their response to treatment. However, CPET may be less feasible in clinical applications because of its cost and technical requirements. Activity monitors provide a less labor-intensive and yet real-world representative assessment of the physical activity state of patients. In healthy youth, a positive relationship between accelerometer-measured intensity of physical activity and cardiorespiratory fitness as measured by maximum volume of oxygen consumption (VO_{2MAX}) has been established.[32,33] Similarly, in healthy adults, the physical activity variables of the time spent in MVPA or vigorous physical activity were shown to be related to the VO_{2MAX}.[34] In the present study of older adults at risk for COPD but with preserved spirometry, the actigraphy activity domains (including MVPA) did not group together with CPET outcomes at peak exercise (e.g. VO_{2Peak}) in cluster analysis, nor were they significantly associated with these outcomes in PCA regression. Interestingly, while they were not statistically significant in PCA regression, actigraphy activity domains were closely

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

associated with the rate of increase in some of the cardiovascular responses to exercise, including O2-Pulse$_{Slope}$ and SBP$_{Slope}$, suggesting a possible correlation between the activity domains of actigraphy and cardiovascular performance. Accordingly, the association of the activity domains of actigraphy with VO_{2AT} likely represents the level of subjects' fitness, with the more physically active individuals reaching their anaerobic threshold at a higher VO_2.

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,[35-37] 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.[38] 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

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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.[39,40]

Conclusion

In conclusion, in this study of a population with early obstructive lung disease or pre-COPD showed that (1) outpatient actigraphy provides “real-world” patient-centered outcomes that objectively inform of patients’ physical activity status, and (2) the activity and sedentary domains of actigraphy are divergent and provide distinct information likely representing the different physiologic processes highlighted by those domains. Furthermore, actigraphy measures are not entirely explained by measures of standardized respiratory questionnaires, lung function, or exercise testing, and thus provide added physical activity measures that are valuable as objective patient-centered outcomes in observational and interventional research studies.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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. Rowe VT, Neville M. Measuring Reliability of Movement With Accelerometry: Fitbit(®) Versus ActiGraph(®). *Am J Occup Ther.* 2019;73(2):7302205150p7302205151-7302205150p7302205156.

19. Yang CC, Hsu YL. A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors (Basel).* 2010;10(8):7772-7788.

20. Rahman MJ, Nemati E, Rahman M, Vatanparvar K, Nathan V, Kuang J. Toward Early Severity Assessment of Obstructive Lung Disease Using Multi-Modal Wearable Sensor Data Fusion During Walking. *Annu Int Conf IEEE Eng Med Biol Soc.* 2020;2020:5935-5938.

21. Lin WY, Verma VK, Lee MY, Lin HC, Lai CS. Prediction of 30-Day Readmission for COPD Patients Using Accelerometer-Based Activity Monitoring. *Sensors (Basel).* 2019;20(1).

22. 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.

23. Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. *Res Q Exerc Sport.* 2000;71(2 Suppl):S1-14.

24. Adams SA, Matthews CE, Ebbeling CB, et al. The effect of social desirability and social approval on self-reports of physical activity. *Am J Epidemiol.* 2005;161(4):389-398.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

25. Adamo KB, Prince SA, Tricco AC, Connor-Gorber S, Tremblay M. A comparison of indirect versus direct measures for assessing physical activity in the pediatric population: a systematic review. *Int J Pediatr Obes.* 2009;4(1):2-27.

26. Kowalski K, Rhodes R, Naylor PJ, Tuokko H, MacDonald S. Direct and indirect measurement of physical activity in older adults: a systematic review of the literature. *Int J Behav Nutr Phys Act.* 2012;9:148.

27. Helmerhorst HJ, Brage S, Warren J, Besson H, Ekelund U. A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. *Int J Behav Nutr Phys Act.* 2012;9:103.

28. Skender S, Ose J, Chang-Claude J, et al. Accelerometry and physical activity questionnaires - a systematic review. *BMC Public Health.* 2016;16:515.

29. van Poppel MN, Chinapaw MJ, Mokkink LB, van Mechelen W, Terwee CB. Physical activity questionnaires for adults: a systematic review of measurement properties. *Sports Med.* 2010;40(7):565-600.

30. 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.

31. 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.

32. Dencker M, Thorsson O, Karlsson MK, et al. Daily physical activity and its relation to aerobic fitness in children aged 8-11 years. *Eur J Appl Physiol.* 2006;96(5):587-592.

33. Gutin B, Yin Z, Humphries MC, Barbeau P. Relations of moderate and vigorous physical activity to fitness and fatness in adolescents. *Am J Clin Nutr.* 2005;81(4):746-750.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

34. Cao ZB, Miyatake N, Higuchi M, Miyachi M, Ishikawa-Takata K, Tabata I. Predicting VO2max with an objectively measured physical activity in Japanese women. *Med Sci Sports Exerc*. 2010;42(1):179-186.

35. 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.

36. 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.

37. 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.

38. Aadland E, Ylvisåker E. Reliability of Objectively Measured Sedentary Time and Physical Activity in Adults. *PLoS One.* 2015;10(7):e0133296.

39. 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.

40. 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.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

566

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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	

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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}:

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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;

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

	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.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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_2pulse_{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

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

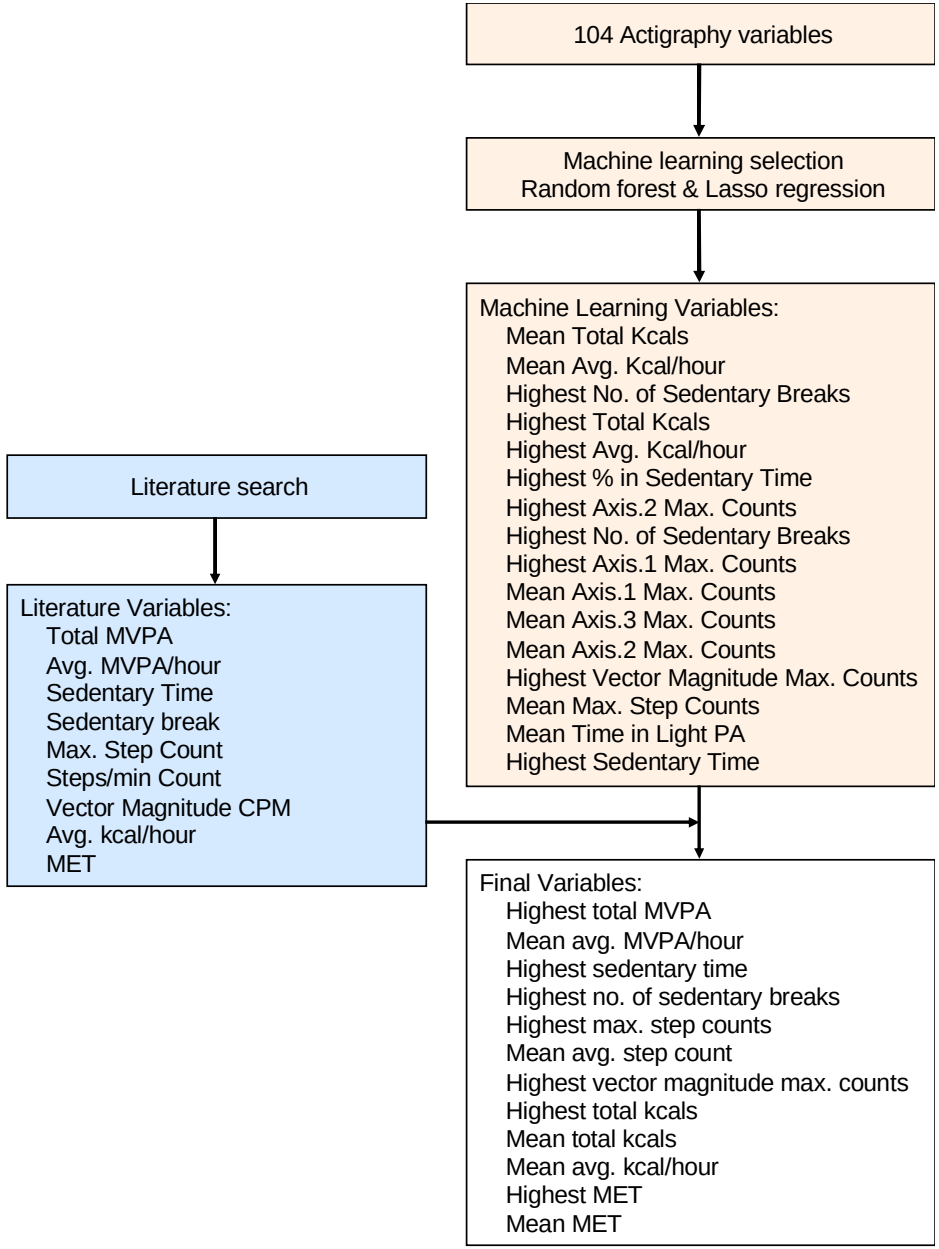
VO_2 at Anaerobic Threshold (L/min)	**0.032**	**0.005**	**0.059**	**0.026**	0.005	-0.065	0.076	0.885	0.25
VCO_{2Slope} (L/min)	-0.005	-0.017	0.007	0.405	-0.002	-0.031	0.028	0.917	0.12
O_2-$Pulse_{Slope}$ (L/beat)	0.071	-0.016	0.158	0.118	0.127	-0.090	0.344	0.259	0.73
V_{ESlope} (L/min)	-0.516	-1.163	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

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

FIGURES

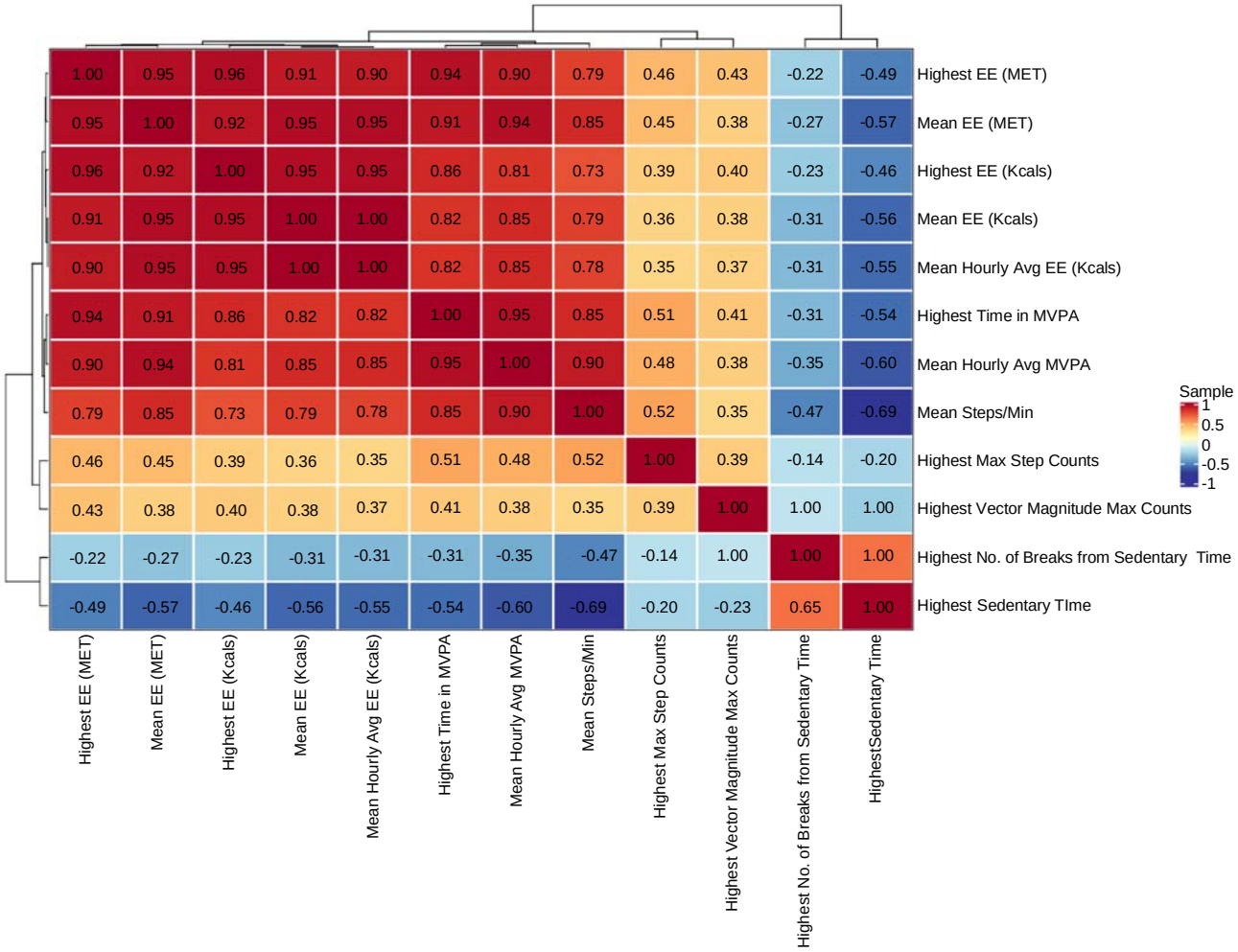
Figure 1- Actigraphy variable selection. A schematic to show the process of variable selection by combining machine learning and literature guided approaches. Machine learning approach was done by implementing random forest and lasso regression. The top ranked 10 variables from each method were done 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.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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.



medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.01.21256454; this version posted May 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.