



OPEN ACCESS

EDITED BY

Raeanne Moore,
University of California San Diego, United States

REVIEWED BY

Sarah Anne Graham,
Lark Health, United States
Matthew O'Connell,
King's College London, United Kingdom

*CORRESPONDENCE

Kaitlin B. Casaletto
kaitlin.casaletto@ucsf.edu

SPECIALTY SECTION

This article was submitted to Human Factors and Digital Health, a section of the journal Frontiers in Digital Health

RECEIVED 05 February 2022

ACCEPTED 08 August 2022

PUBLISHED 31 August 2022

CITATION

VandeBunte A, Gontrum E, Goldberger L, Fonseca C, Djukic N, You M, Kramer JH and Casaletto KB (2022) Physical activity measurement in older adults: Wearables versus self-report.
Front. Digit. Health 4:869790.
doi: 10.3389/fdgth.2022.869790

COPYRIGHT

© 2022 VandeBunte, Gontrum, Goldberger, Fonseca, Djukic, You, Kramer and Casaletto. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Physical activity measurement in older adults: Wearables versus self-report

Anna VandeBunte^{1,2}, Eva Gontrum¹, Lauren Goldberger¹, Corrina Fonseca¹, Nina Djukic¹, Michelle You¹, Joel H. Kramer¹ and Kaitlin B. Casaletto^{1*}

¹Department of Neurology, Memory and Aging Center, University of California, San Francisco, San Francisco, CA, United States, ²Department of Psychology, Palo Alto University, Palo Alto, CA, United States

Physical activity (PA) is associated with preserved age-related body and brain health. However, PA quantification can vary. Commercial-grade wearable monitors are objective, low burden tools to capture PA but are less well validated in older adults. Self-report PA questionnaires are widely accepted and more frequently used but carry inherent limitations. We aimed to compare these commonly used PA measures against one another and examine their convergent validity with a host of relevant outcomes. We also examined the factors that drive differences in PA self-reporting styles in older adults. 179 older adults completed 30-day Fitbit Flex2™ monitoring and reported PA levels via two widely used PA questionnaires: PASE and CHAMPS-METs (metabolic expenditure calories burned). Participants also completed measures of cardiometabolic (hypertension diagnosis, resting heart rate, A1C levels), cognitive (memory, processing speed, executive functioning), and brain MRI (medial temporal lobe volume) outcomes. The discrepancy between objective Fitbit monitoring and self-reported PA was evaluated using a sample-based z difference score. There were only modest relationships across all PA metrics. Fitbit step count demonstrated a stronger association with the PASE, whereas Fitbit calories burned was more strongly associated with CHAMPS-MET. Fitbit outcomes had more consistent convergence with relevant outcomes of interest (e.g., cardiometabolic and brain health indices) when compared to subjective measures; however, considerable heterogeneity within these associations was observed. A higher degree of overreporting was associated with worse memory and executive performances, as well as hypertension diagnoses. We build on prior findings that wearable, digital health indicators of PA demonstrate greater construct validity than self-report in older adults. We further show important clinical features (e.g., poorer cognitive status) of older adults that could contribute to a higher level of overreporting on self-report measures. Characterization of what PA measures truly operationalize will help elucidate relationships between most relevant facets of PA and outcomes of interest.

KEYWORDS

healthy aging, fitbit, actigraphy, CHAMPS, PASE

1. Introduction

Physical activity (PA) is associated with preserved vascular health, brain structure, and cognition with age (1–3). In the context of neurodegenerative diseases, active lifestyles are also linked with less functional decline and reduced risk of dementia (2, 4, 5). However, physical activity is a broad construct, and its operationalization can vary widely depending on the measures employed, possibly leading to imprecision and/or inconsistencies in the literature linking physical activity to brain health.

The most commonly utilized measures of PA may capture several aspects of activity. Actigraphy monitors are useful for objective assessment of free-living movement levels, ranging from everyday chores to structured exercise. Given their relative low cost and user friendly interfaces, commercially available wearables (like Fitbit actigraphy monitors) have garnered increasing attention and validation of metrics for reliable measurement of PA (6, 7). For instance, Fitbit monitors have demonstrated inter-device reliability with other actigraphy monitors (i.e., Actigraph GT3X+ accelerometer) and positive correlations with observed step count and gait speed on a treadmill (8–10). Within the Fitbit suite of outcomes, average steps has demonstrated the most robust validity though there is less validity for energy expenditure as a measurement of PA (11). However, few studies have been conducted in older adults who may have varying experience using wearable devices. As digital health tools expand, understanding their utility in populations that may be most vulnerable (e.g., older adults) is needed. Moreover, considering the importance of engaging in PA for healthy aging, it is important to understand the convergence across measures, how measures of PA may differ, and what are predictors of discrepancies between measures.

In the absence of actigraphy metrics, standard self-report measures of PA are used in older adults to capture more structured activities and exercise routines (e.g., duration, type of exercise). The Community Healthy Activities Model Program for Seniors (CHAMPS) and Physical Activity Scale for the Elderly (PASE) are two such widely used self-report questionnaires of PA in older adults. However, these subjective measures may carry inherent limitations and capture only segments of PA. For instance, the PASE evaluates the level of activity (e.g., frequency and duration), whereas the primary outcome for CHAMPS is metabolic expenditure. When considering utility, each is relatively quick, well-validated, and inexpensive to use. PASE has previously been associated with portable accelerometer readings, walking steps, and energy expenditure (12, 13). Higher PASE scores (indicating a greater degree of PA) have also been correlated with reduced likelihood of cardiometabolic, neurological, and psychological health conditions (14, 15). PASE has also demonstrated meaningful associations with health status and

physiologic measures, such as heart rate and static balance (13, 16). Similarly, the CHAMPS has been shown to relate to a host of relevant outcomes, such as physical functioning and psychological health (17). Similar to the PASE, the CHAMPS has demonstrated positive associations with total minutes of movement measured by an accelerometer and corresponding intensity, as well as measures of fitness capacity (i.e., the 6-min walk test) and lower body physical functioning (i.e., Short Physical Performance Battery) (17–19). However, such self-report tools may be limited by scope and subjectivity. For instance, self-report measures typically underestimate sedentary time compared to real-time digital health measures, such as an accelerometer or inclinometer (20). This is particularly important when using subjective measurements to assess older adults with cognitive difficulties, as there is greater risk of recall bias (21). Furthermore, individuals who have been encouraged to engage in exercise (e.g., many older adults by their physicians) have demonstrated the tendency to engage in more overreporting, perhaps related to well-known effects of social desirability bias (22).

To date, there is a gap in the literature directly comparing and evaluating PA as assessed across multiple standardly employed measures. Studies have not pragmatically demonstrated how metrics within the Fitbit suite compare against widely used self-report measures of PA in older adults. Similarly, research has yet to compare commonly used self-report measures (i.e., CHAMPS and PASE) alongside a comprehensive panel of relevant neurologically relevant aging outcomes (e.g., cognition, MRI outcomes) to characterize their convergent validity for use in brain aging studies. Furthermore, it is uncertain whether there are discrepancies in reporting styles across self-report measures, and if particular participant characteristics systematically predict older adults who over or under report. The current study will begin to elucidate some of these relationships to contribute to our understanding PA measurement tools in older adults.

In the current study, we aimed to (1) determine the comparability across commonly used self-reported measures of physical activity (PA) and Fitbit-based actigraphy metrics, (2) examine the convergent validity of Fitbit, PASE, and CHAMPS using a comprehensive panel of demographic, cardiometabolic, cognitive, and brain structural outcomes, and (3) examine the person-specific factors that characterize overreporting on self-report measures. We hypothesized that objective measures of PA *via* Fitbit would demonstrate the best construct validity and that the degree of overreporting would relate to poorer neurobehavioral status.

2. Materials and methods

One hundred seventy-nine older adults enrolled in the UCSF Memory and Aging Center's Longitudinal Aging Study

TABLE 1 Descriptive statistics.

	<i>n</i>	% or <i>M</i> (<i>SD</i>)
Sex, % female	105	58.66%
Race		
White	153	85.47%
Black	2	1.12%
Asian	19	10.61%
Other	5	2.79%
Age (years)	179	73.50 (8.23)
Education (years)	179	17.57 (1.85)
Fitbit steps (daily average)	179	7840.77 (3365.11)
Fitbit calories (daily average)	179	1862.27 (426.51)
PASE (possible range 0 to >500)	105	126.10 (60.66)
CHAMPS-MET (max calories burned in a week)	85	4062.76 (2275.75)
Hypertension, % yes	116	37.93%
Resting heart rate (bpm)	165	66.62 (9.51)
Hemoglobin A1C (%; normal range 4.3–5.6)	97	5.47 (0.33)
Technology Familiarity Questionnaire (Q8) ^d	128	4.50 (0.68)
Technology Familiarity Questionnaire (Q9) ^e	128	4.86 (0.41)
Memory (<i>z</i> -score) ^a	124	−0.07 (0.87)
Executive functioning (<i>z</i> -score) ^c	132	0.77 (0.58)
Processing speed (<i>z</i> -score) ^b	126	−2.60 (1.58)
Medial temporal lobe volume (voxels, 1 cm ³)	72	9.80 (1.05)

Note. *N* = 175.

^a*z*-scores on these tests represent performances compared to the larger Hillblom Aging cohort of older adults.

^b*z*-score represents performance compared to young adults (20–30 years old).

^c*z*-score derived from EXAMINER normative study group (adults aged 18–80+).

^dQuestion 8: "How much difficulty do you have using computers?" (Range 1–5, 1 = extreme difficulty, 5 = no difficulty).

^eQuestion 9: "How anxious (or nervous) do you typically feel when using a computer, tablet, or smartphone?" (Range 1–5, 1 = extremely anxious, 5 = not anxious).

who choose to participate in 30-day Fitbit monitoring (average daily steps and calories burned), and who completed at least one measure of self-reported physical activity levels (PASE, *n* = 105; CHAMPS-MET, *n* = 85) were included in the study (see **Table 1**). Participants completed comprehensive neurological and neuropsychological evaluations, as well as structural neuroimaging, cardiometabolic measures, and a study partner interview. Following evaluations, participants were reviewed at a case conference with board certified neurologists and neuropsychologists. Inclusion criteria for enrollment consisted of: (1) no current evidence of a memory or neurological condition (e.g., stroke, epilepsy), (2) no functional decline as operationalized as a Clinical Dementia Rating (CDR) scale of 0–0.5 *via* study partner interviews, (3) no history or evidence of DSM-5 major psychiatric disorders, active substance abuse, hepatitis C, blindness, deafness, HIV, and syphilis. Participants had very minimal cardiovascular medical histories (Myocardial Infarction, *n* = 4, Cerebrovascular Accident, *n* = 2, Transient Ischemic Attack, *n* = 2; note, all cardiovascular

events occurred >5 years prior to study participation). The study was approved by the institutional review board of the University of California, San Francisco and is conducted in accordance with the latest Declaration of Helsinki, including written informed consent from all participants.

2.1. Procedure

At the in-person visit, participants neuropsychological evaluations, brain MRI, self-reported measured, and Fitbit set-up. Participants then completed 30 days of subsequent Fitbit monitoring before mailing their device back to study personnel. All data from Fitbit devices were synced to the Fitabase platform and data were downloaded for quality control, cleaning, and analysis.

2.2. Actigraphy monitoring

The FitBit Flex2™ (Fitbit Inc., San Francisco, CA, USA; <https://www.fitbit.com>) recorded average daily steps and calories burned. The FitBit Flex2™ is a thin, flexible, Bluetooth fitness tracker with no visible record of physical activity measurements, with a three-axis acceleration sensor, and with the capability to store 7 days of detailed motion data. Participants were blinded to all notifications and indication of the duration of exercise for the 30-day time period. They were instructed to wear the Fitbit during all waking hours, and to synchronize nightly with their smartphone *via* Bluetooth 4.0 before charging at night. In cases where the participant did not have a Fitbit-compatible smartphone, the Fitbit was synchronized to an iPad and aggregate daily physical activity data was collected at the completion of the 30-day period. Fitbit accounts for each participant were connected to Fitabase, a platform specifically tailored for wearable research data management.

Average daily steps and average daily calories burned were selected as the primary objective outcomes of interest. Fitbit averages were calculated by taking the average daily steps and calories burned for the first 20 days of available monitoring data. Individual days with fewer than 100 steps were removed from the analyses to control for nonadherence. Participants were only included if they had at least 14 days of available monitoring data, and the first 20 days of the 30-day monitoring data were used in the analyses.

2.3. Physical activity questionnaires

2.3.1. Physical activity scale for the elderly

Self-reported physical activity was measured using the Physical Activity Scale for the Elderly (PASE), a widely

validated measure of self-reported activity levels for older adults. Participants were asked to rate the frequency, duration, and intensity of activity in three domains (leisure, household, and work-related activity) over the past seven days with the 11-item questionnaire. Utilizing the scoring manual, activity scores were computed by multiplying activity frequencies by the task-specific weights (16). Activity scores were then summed to obtain a total score representing overall physical activity level, with higher values indicating greater activity.

2.3.2. CHAMPS-MET

The Community Healthy Activities Model Program for Seniors physical activity questionnaire (CHAMPS) was administered to assess the variety of physical activities that older adult participants may engage in, from less intensive forms such as walking or stretching to more vigorous exercise routines (17). The questionnaire includes 41 items to evaluate the frequency and duration of light, moderate, and vigorous activities that were performed weekly over the last four weeks. Participants reported whether they participated in an activity during the four-week period and then selected the hours per week spent participating in the activity, rating the duration on a six-point scale from less than 1 to 9 or more hours. Each activity corresponds to a metabolic weight or MET value. Estimated caloric expenditure was calculated by multiplying the estimated duration of each activity by the corresponding MET value, in alignment with published guidelines (17).

2.4. Cognitive outcomes

Participants completed a neuropsychological battery assessing cognitive outcomes hypothesized to be associated with physical activity (23, 24). This brief standardized battery has been previously described and validated to be neuroanatomically sensitive to age-related neurodegeneration (25, 26).

2.4.1. Episodic memory

Verbal episodic memory was measured by the California Verbal Learning Test (CVLT-II) and a modified version of the Benson Figure Memory test. The CVLT-II includes a 16-item list presented over five learning trials, followed by free and cued recall of the list after an interference trial, and then again after a 20-min delay. Following the long delay, participants were given a list of 44 words and asked to discriminate between the target word and a distractor item (recognition trial). Outcome metrics included words correctly recalled after delays and recognition discrimination performance.

To assess visual memory, participants were asked to draw the modified Benson figure from memory after a 10-min delay. Recall of the figure was scored on a 17-point scale (25).

Sample-based z-scores were created for outcomes on both measures and averaged together to create an episodic memory composite.

2.4.2. Processing speed

Processing speed was assessed through five computerized tests of reaction time to different visual stimuli (dots, lines, search, shapes, abstract matching 1, abstract matching 2) (27). All tasks included a practice trial period where the participant had to perform at greater than 70% accuracy in order to continue to the test trials. Sample-based z-scores were created for each of the five tasks to calculate a processing speed composite score.

2.4.3. Executive functioning

Executive functions were measured by the NIH EXAMINER (28). NIH EXAMINER includes a composite score of five computer-based tests of working memory (dot counting, 1-back, 2-back), response inhibition (enclosed flanker), and set shifting (set shifting), and two verbally mediated tests of generativity (D-word and animal fluency). All computerized tasks included at least three practice trials.

2.5. Cardiometabolic outcomes

2.5.1. Hemoglobin A1C

Whole blood and serum samples were collected and stored in 0.5 ml aliquots at -80°C following baseline 12 h fasting blood draws, until used for biochemical processing. All laboratory analyses were performed by UCSF Clinical Laboratories, a CLIA-certified, CAP-accredited laboratory at UCSF Mission Bay Hospital. Hemoglobin A1C (HbA1C) levels were determined from whole blood by means of an Abbott Architect c8000 enzymatic immunoassay.

2.5.2. Resting heart rate

Participant resting heart rates were measured by a clinician or study staff. A normal resting heart rate for adults ranges from 60 to 100 beats per minute (29). Elevated heart rate is a risk factor for cardiovascular morbidity and mortality (30).

2.5.3. Hypertension

Personal medical history, including presence of hypertension, was collected *via* self-report during the clinical history gathering with a neurologist. If the participant did not report a history of hypertension, but indicated taking an antihypertensive drug, she/he was asked by the clinician or study staff if they were prescribed the medication for their blood pressure. If answered yes, the participant was marked as having a history of hypertension.

2.6. Brain MRI

2.6.1. Structural neuroimaging

Magnetic Resonance Imaging (MRI) scans were performed at the UCSF Neuroscience Imaging Center using a Siemens Prisma fit 3 T scanner. Magnetization prepared rapid gradient-echo (MPRAGE) sequences were used to obtain whole brain T1-weighted images (TR/TE/TI = 2300/2.9/900 ms, $\alpha = 9^\circ$). The field of view was 240 mm \times 256 mm, with 1 mm \times 1 mm in-plane resolution and 1 mm slice thickness with a sagittal orientation.

Before processing, all T1-weighted images were visually inspected for quality control and those with excessive motion or image artifact were excluded. Magnetic field bias was corrected using the N3 algorithm (31). Tissue segmentation was performed using unified segmentation in SPM12 (32). Each participant's gray matter segmentation was warped to create a study-specific template using Diffeomorphic Anatomical Registration using Exponentiated Lie algebra (DARTEL) (33). Participants' native space gray and white matter segmentations were then normalized and modulated to study-specific template space using nonlinear and rigid-body transformations. Images were smoothed using a Gaussian kernel of 4-mm full width half maximum. Each participant's segmentation was carefully inspected to ensure the robustness of the process.

For statistical purposes, linear and nonlinear transformations between DARTEL's space and ICBM space were applied (34). Quantification of volumes in specific brain regions was accomplished by transforming a standard parcellation atlas into International Consortium for Brain Mapping (ICBM) space and summing all modulated gray matter within each parcellated region of interest (ROI) (35). Total intracranial volume was calculated for each participant as the sum of the gray matter, white matter, and cerebrospinal fluid segmentations. Medial temporal lobe volume was selected as our brain MRI outcome, as exercise engagement in older adults has previously been associated with greater volume in this particular region (36). For the purpose of this study, medial temporal lobe volumes included the following bilateral regions: hippocampus, entorhinal cortex, and the parahippocampal gyrus.

2.7. Statistical analyses

First, we examined associations among the four PA measures (Fitbit steps, Fitbit calories burned, PASE, CHAMPS-MET) with Spearman's rank correlations to evaluate comparability. Next, we evaluated relationships between PA measures with demographic variables of interest *via* independent samples *t*-tests, Spearman's rank correlations, and ANOVA. We tested construct validity by evaluating the relationship between each PA measure and the cardiometabolic and cognitive outcomes of interest *via*

Spearman's rank correlations or independent samples *t*-tests (e.g., HTN diagnosis), as appropriate. Lastly, we examined relationships between each PA measure and medial temporal volume outcomes *via* linear regression modeling adjusting for total intracranial volume. We reported effect sizes as Spearman's correlations, Cohen's *d*, or standardized betas, as necessary.

In order to identify the characteristics of older adult participants who were over or under reporting, we performed a discrepancy score analysis. Given our data indicated that Fitbit total steps demonstrated the best construct validity out of all PA measures examined, we utilized Fitbit total steps as our "gold standard" metric in these analyses. First, within participants that completed all three PA metrics ($n = 75$), we computed sample-based z-scores for each PA measure (Fitbit steps, PASE, and CHAMPS-MET). We computed individual discrepancy scores separately by subtracting Fitbit total steps z-scores from each self-report PA measure z-score (e.g., PASE or CHAMPS-MET). Distribution of discrepancy scores approximated normality (Figure 1). In this manner, higher discrepancy scores indicated greater overreporting compared to Fitbit. Next, we evaluated relationships between each discrepancy score with demographic (i.e., gender, age, education), cardiometabolic, cognitive, and brain volume outcomes. We tested relationships and group differences *via* independent samples *t*-tests, linear regression, and Spearman's rank correlations. Interpretation of effect sizes was in alignment with Cohen's *Statistical Power Analysis for the Behavioral Sciences*: coefficients of 0.10 "small," .30 "medium," and those of 0.50 "large" in terms of the magnitude (37).

3. Results

On average, participants were 74 years old, 58% female, and took 7,841 average daily steps during the monitoring period. Older age was associated with less physical activity across metrics, but only reached significance for Fitbit outcomes (Table 2). Females had lower levels of quantified physical activity for Fitbit daily steps [$t(177) = 2.43, p = 0.02$], Fitbit daily calories [$t(177) = 15.77, p < 0.001$], and lower reported physical activity for CHAMPS-MET [$t(83) = 2.98, p < 0.001$], but not PASE [$t(103) = 0.72, p = 0.47$]. Education did not meaningfully associate with the physical activity metrics (Table 2).

3.1. Associations among physical activity measures

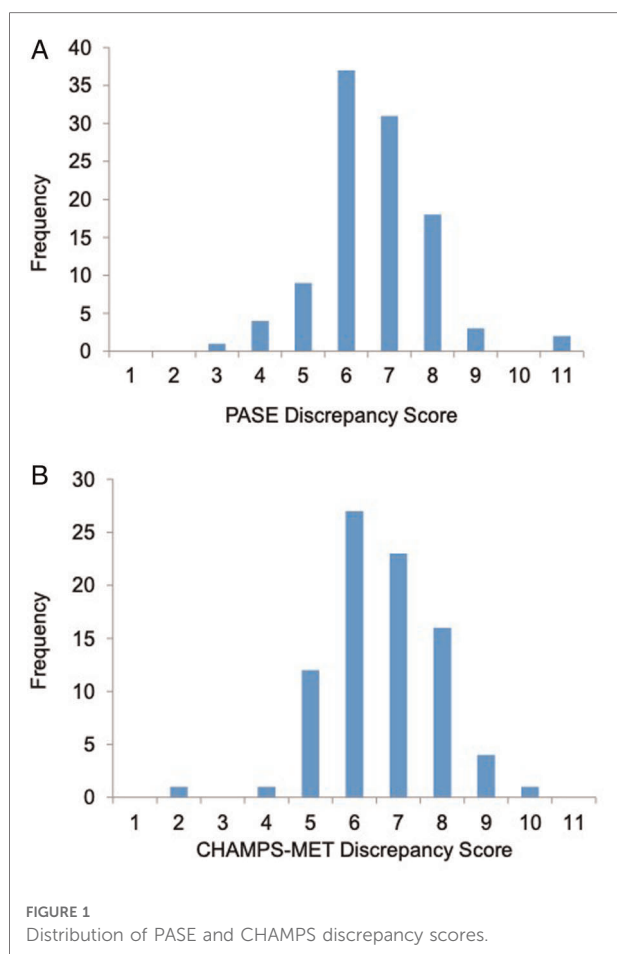
All of the PA metrics were positively correlated, though demonstrated only small-to-medium effect sizes (Table 2). Greater Fitbit step count was more strongly associated with PASE compared to CHAMPS-MET, whereas Fitbit calories burned was more strongly associated with CHAMPS-MET

compared to PASE scores. Self-reported CHAMPS-MET and PASE scores demonstrated a medium effect size correlation with one another.

3.2. Cognitive outcomes

Higher Fitbit step count [$\rho(132) = 0.28, p < 0.001$] and calories burned [$\rho(132) = 0.23, p = 0.01$], but not PASE [$\rho(78) = -0.10, p = 0.39$] were associated with better

performances on measures of executive functioning. Unexpectedly, CHAMPS-MET demonstrated an inverse relationship with executive functioning [$\rho(63) = -0.41, p < 0.01$]. Only Fitbit measures demonstrated expected positive associations with processing speed, though effect sizes were small (ρ range = 0.05–0.07, $ps < 0.43$). Fitbit step count [$\rho(124) = 0.20, p = 0.03$], but not Fitbit calories burned [$\rho(124) = -0.08, p = 0.38$], PASE [$\rho(76) = -0.26, p = 0.02$], or CHAMPS-MET [$\rho(62) = -0.31, p = 0.02$] was associated with better scores on tests assessing memory.



3.3. Cardiometabolic outcomes

Lower resting heart rate was significantly associated with greater daily calories burned ($\rho = -0.29, p < 0.001$), but less strongly with daily steps ($\rho = -0.14, p = 0.08$) and showed minimal associations with reported PA as measured by the PASE ($\rho = 0.02, p = 0.85$) or CHAMPS-MET ($\rho = -0.03, p = 0.80$) (Figure 2). Greater Fitbit step count ($t = 3.058, p < 0.001$), but not Fitbit calories burned ($t = -0.37, p = 0.71$), PASE ($t = 1.43, p = 0.16$) or CHAMPS-MET ($t = -1.75, p = 0.09$) was associated with a lower likelihood of hypertension. Each PA measure demonstrated expected negative associations with hemoglobin A1C that did not reach statistical significance (ρ range = -0.01 to $-0.17, ps > 0.05$).

3.4. Brain MRI outcomes

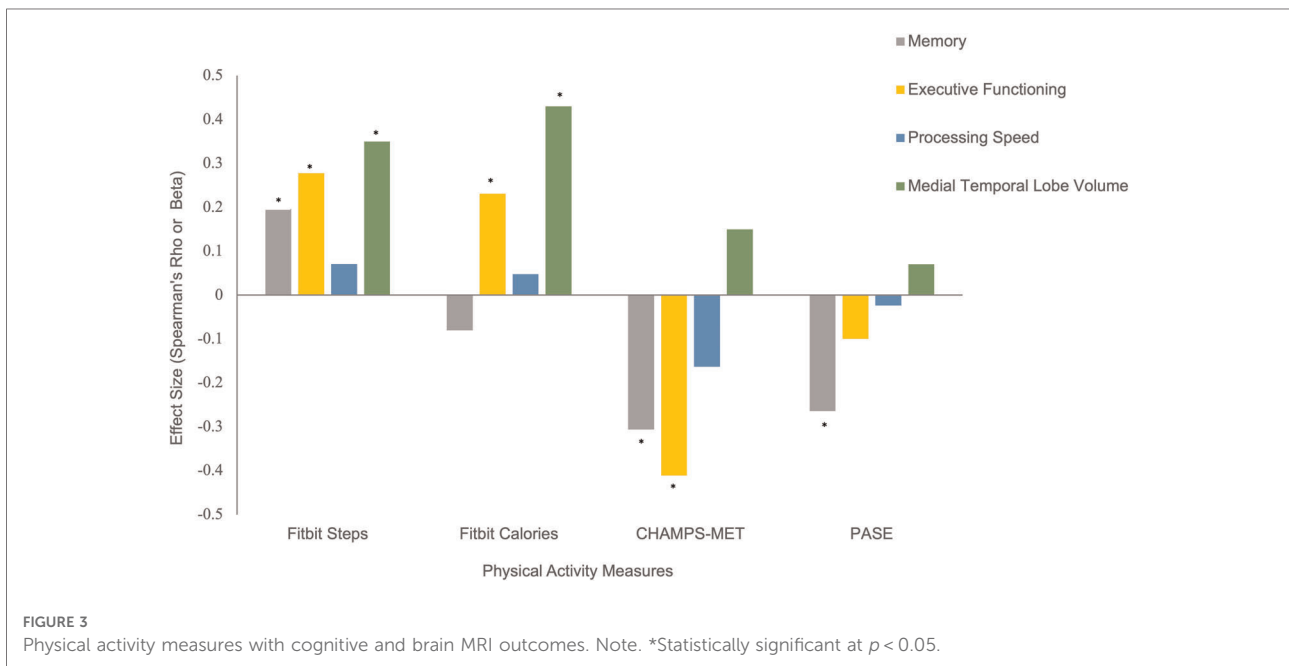
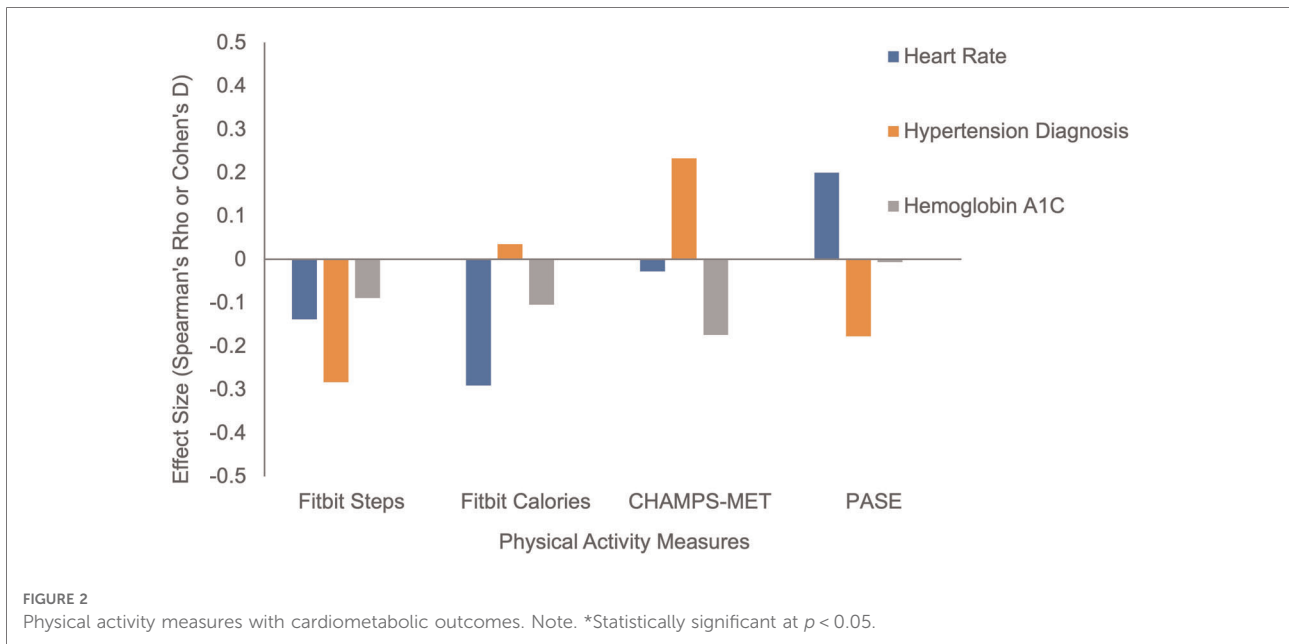
Lastly, greater Fitbit step count ($\beta = 0.35, p < 0.001$) and calories burned ($\beta = 0.43, p < 0.001$), but not PASE ($\beta = 0.15, p = 0.17$) or CHAMPS-MET ($\beta = 0.07, p = 0.56$), were associated with larger medial temporal lobe volumes (see Figure 3).

Given there were sex and age-related differences, we ran models adjusting for these factors and the patterns remained the same, with the exception of the relationship between Fitbit calories burned and memory, which showed a positive association ($\beta = 0.31, p = 0.07$)

TABLE 2 Correlations between physical activity measures, age, and education.

	1. Age	2. Education	3. Fitbit steps	4. Fitbit calories	5. PASE
1. Age					
2. Education					
3. Fitbit steps	-0.36*	0.02			
4. Fitbit calories	-0.38*	0.13	0.50*		
5. PASE	-0.13	-0.11	0.35*	0.20*	
6. CHAMPS-MET	-0.03	0.04	0.20	0.31*	0.44*

Note. *Statistically significant at $p < 0.05$.



3.5. Discrepancy scores

Using Fitbit step count as the standard, discrepancy score analysis showed that females tended to overreport to a greater degree on both self-report PA scales, though this effect reached significance only for PASE (PASE Cohen's $d = 0.61$, $p = 0.02$; CHAMPS Cohen's $d = 0.22$, $p = 0.45$), see **Figure 4**. Older age and overreporting showed small associations that

did not reach significance (ρ range = 0.09–0.17, $ps > 0.14$). Additionally, those with a diagnosis of hypertension overreported on CHAMPS-MET (PASE Cohen's $d = 0.20$, $p = 0.45$; CHAMPS Cohen's $d = 1.02$, $p < 0.01$), see **Figure 4**. Generally, worse cognition was associated with greater degree of overreporting particularly for measures of memory (ρ range = -0.23 to -0.26 , $ps = 0.05$ – 0.09) and executive functioning (PASE $\rho = -0.15$, $p = 0.28$; CHAMPS $\rho = -0.32$, $p < 0.01$), see **Figure 5**.

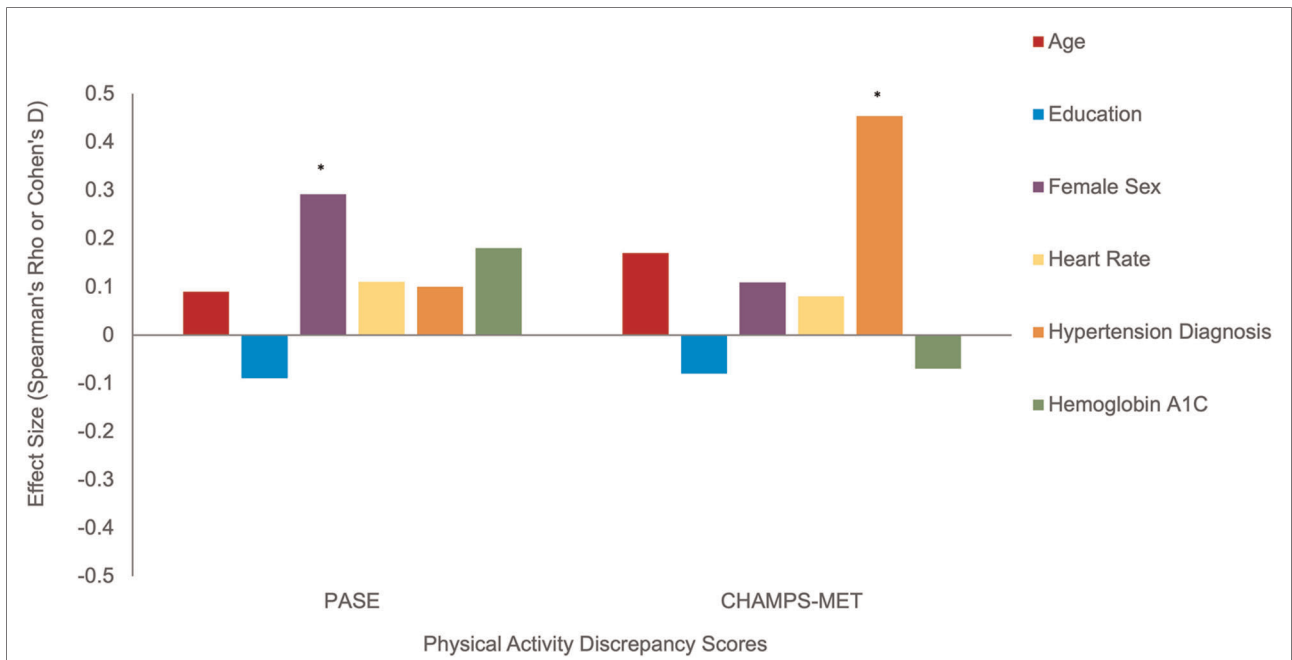


FIGURE 4 Demographic and cardiometabolic correlates of physical activity overreporting (positive discrepancy score indicates greater overreporting). Note. *Statistically significant at $p < 0.05$.

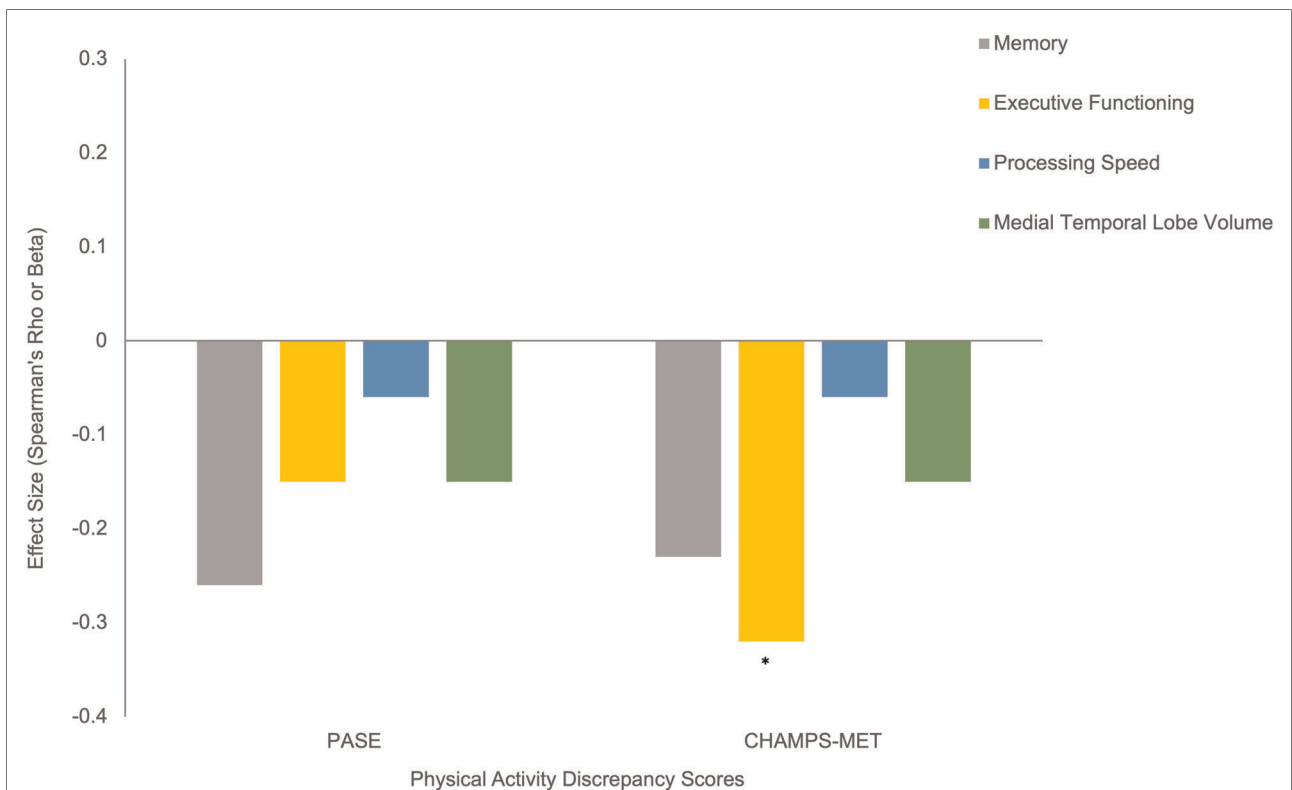


FIGURE 5 Cognitive correlates of physical activity overreporting (positive discrepancy score indicates greater overreporting). Note. *Statistically significant at $p < 0.05$.

Decreased medial temporal lobe volume showed small associations with increased overreporting, though they did not reach statistical significance (β range = -0.14 to -0.15 $ps > 0.19$), see [Figure 5](#).

4. Discussion

We build on previous findings that wearable, objective indicators of PA demonstrate greater overall construct validity compared to self-report measures now extending to older adults, and further detail the characteristics that may impact reporting styles on PA measures. We also highlight several novel findings on the limited interchangeability of PA metrics. Both Fitbit steps and calories burned demonstrated more consistent convergence with relevant outcomes of interest when compared to subjective measures. These results are consistent with prior studies in younger adults showing closer approximation of measuring physical activity with objective monitors in comparison to self-report ([38](#), [39](#)). Within Fitbit metrics, step count particularly demonstrated stronger and more consistent expected associations with each cardiometabolic, cognitive, and brain volume outcome of interest compared to calories burned, consistent with previous studies ([11](#)). These findings converge with previous literature that has demonstrated distinct associations between physical activity and executive functioning, memory, and MTL volume ([36](#), [40](#), [41](#)). While we again identify that objective digital health measures show preferable construct validity over subjective measures of physical activity, objective measures still demonstrated substantial variability with expected outcomes of interest (e.g., hypertension diagnosis, memory) and more work is warranted to elucidate factors that contribute to variability in physical activity metrics for older adults.

Notably, the associations found among the four physical activity metrics used in this study suggest a lack of cohesion within the field's current standard measurement approaches. For instance, our results demonstrated only modest relationships across the four physical activity measures (i.e., $r = 0.2$ – 0.30) indicating small variance explained between each measure (i.e., $R^2 = 4\%$ – 9% variance explained). These data suggest only minimal overlap among metrics that were created to quantify the same construct. There was also variation between Fitbit metrics and their associations with the self-report measures. Fitbit step count demonstrated a greater association with subjective reports of physical activity measured by the PASE, whereas Fitbit calories was more strongly associated with physical activity measured by CHAMPS-MET. These findings illustrate an important implication. While these subjective and objective measures were intended to capture the same construct, there is clearly variability in how each measure assesses physical activity. However, it is worth noting that CHAMPS-MET appears to

be more closely related to the construct of calories burned, as expected. It is important to continue to clarify what each measure is assessing (with regard to physical activity) in order to pinpoint what factors of physical activity are beneficial in the context of brain and age-related health.

In addition to variation across the physical activity measures, our results also indicated considerable heterogeneity with associations between each activity measure and outcomes of interest. For example, Fitbit step count showed a stronger relationship to decreased likelihood of hypertension, while Fitbit calories demonstrated a stronger association with lower resting heart rate. Medial temporal lobe volume, and hemoglobin A1C demonstrated consistent, expected associations with each measure of PA; however, effects were small and often did not reach statistical significance (with the exception of Fitbit metrics and brain volume). Overall, there was notable variability in the strength and direction of examined relationships, particularly between self-reported physical activity measures and each outcome. This again suggests that physical activity may be comprised of several constructs that are differentially tapped into by self-report questionnaires and wearable devices, and/or there is imprecision when capturing physical activity across measurement tools. These findings are particularly relevant in scientific research when comparing across studies, as these objective and subjective measures do not appear to be equivalent to one another. Our results revealed that metrics capturing physical activity intensity *via* calories burned (e.g., Fitbit calories, CHAMPS-MET) may not be interchangeable with metrics of overall movement (i.e., Fitbit steps, PASE). Our data highlight how challenging measurement of physical activity can be, and that there is still room for improvement, even within “gold-standard” objective measures.

Fitbit steps demonstrated the greatest construct validity. However, there are still mixed findings in current literature with regard to its validity and reliability ([42](#)). For example, one study found that Fitbit total steps underestimated activity in healthy adults walking at faster treadmill speeds, but overestimated total steps at slower speeds ([43](#)). Similarly, another study found that Fitbit underestimated caloric expenditure in comparison to CHAMPS ([39](#)). The results of the current study, in conjunction with the variability in results from previous studies examining physical activity measures, highlight the importance of improving the current standard measurements of physical activity. In addition, these inconsistencies point to a need for refinement of the operational definitions associated with each of these physical activity measures to best understand what they are evaluating.

Notably, more precise measurement and specification of the broad range of physical activity constructs currently

being utilized would allow for greater understanding of the specific movement patterns that are most critical for brain health. To date, there is not strong evidence for a particular movement (e.g., walking) or intensity (calories burned, heart rate during exercise) that is most impactful for brain health trajectories. For example, in a meta-analysis of exercise randomized control trials (RCTs) in older adults, type of exercise was not a significant predictor of cognitive benefit (44). Indeed, activities ranging from tai chi to jogging demonstrated comparable benefit. Some of the earliest evidence linking physical activity with cognitive aging demonstrated beneficial effects even with low impact activities, such as walking (45). However, several epidemiologic and RCTs indicate that cardiorespiratory activities aimed at increasing VO_2 max may be particularly beneficial for cognitive outcomes and future dementia risk (2, 46, 47). Nonetheless, more work needs to be done in this area to understand what particular aspects of physical activity are most important for brain health.

Because of the degree of variability found within associations between measures of physical activity and important outcomes, we elected to more closely examine which factors could be driving these relationships. More specifically, we leveraged a discrepancy score analysis to identify factors that characterize participants who may overreport on physical activity questionnaires. The results demonstrated a greater degree of overreporting in females, particularly for the PASE. This finding may impact the ability of studies to determine sex-related differences in physical activity for brain and age-related health. In addition, participants who were most discrepant generally had higher levels of physical activity across all measures, which may suggest that self-report measures are less accurate in detecting physical activity levels for those who are very active.

We also found that individuals who engaged in the greatest degree of overreporting were older adults who performed worse on cognitive assessments, suggesting subjective measures may be systematically confounded by cognitive ability when assessing activity level in the aging population. This finding converges with prior studies demonstrating that utilization of self-report measures in older adults for measuring physical activity is less accurate (48). In addition, cognitive functioning declines with age as a group, which in turn may increase risk of inaccurate responses on self-report measures (49). In addition to decline in cognition, a greater degree of overreporting was also associated with smaller medial temporal lobe volume. Generally, our findings demonstrated that individuals with poorer vascular and cognitive health tended to overreport to a greater degree. These novel findings increase our understanding of possible factors that could be contributing

to reporting bias in older adults, namely their vascular and cognitive health. Future studies should examine whether these discrepancy scores could be used to predict individuals at risk for adverse brain aging.

Our study is not without limitations. Simply wearing a Fitbit may have increased participants' motivation to move, which could lead to possible a possible confound in our data. However, we noted that all feedback from the Fitbit was removed from the device so participants were otherwise blinded to real time activity levels. Our sample was not representative of the general population, with 85 percent of participants identifying as White, limiting generalizability. Furthermore, our sample was relatively small and had some variability across measures, which may have biased the outcomes. There are also inherent biases that can occur when utilizing digital technology with older adults; notably, data collection is contingent on successful device use, which may be impacted by technology familiarity. Therefore, study findings have limited generalizability to older adults with low levels of technology literacy. In addition, the high educational attainment of this cohort also impacts our generalizability and is particularly important considering higher education has been associated with lower risk of developing neurodegenerative disease. It is possible that our results are limited by a slight gap in time between obtaining self-report data and gathering metrics from Fitbit monitors. At the visit, participants reported physical activity from the last seven days *via* PASE, and from the prior 30 days *via* CHAMPS. Fitbit monitoring took place in the following 30 days post visit. However, physical activity levels are a generally stable trait (50), so it is unlikely this time discrepancy significantly affected our findings.

Our study is also limited by the lack of other "gold-standard" objective measures of physical activity and fitness, such as the 6-min walk or a research-grade accelerometer (e.g., Actigraph GT3X+). With these other objective measures, we could have more comprehensively examined Fitbit as an actigraphy metric. In addition, a longitudinal study design would have provided the opportunity to understand the reliability of subjective and objective PA measures over time and meaningfully track changes in cognitive and vascular health related to physical activity.

Without precision and specificity, it is difficult to pinpoint which aspects of physical activity contribute to brain health. Our results suggest that objective quantification of physical activity demonstrates the best validity and high clinical relevance. Moreover, issues regarding reporting bias may be especially important in older adults with lower vascular, cognitive and brain structural statuses. These findings also begin to broaden our understanding of what physical activity metrics represent to facilitate better-informed recommendations for healthy aging.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by UCSF Human Research Protection Program (HRPP). The patients/participants provided their written informed consent to participate in this study.

Author contributions

AV and KC contributed to the conception and design of the study and performed the statistical analysis. ND and MY organized the database. AV wrote the manuscript. LG, EG, and CF wrote sections of the manuscript. JK provided critical edits and study funding. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the NIH-NIA (K23 AG058752, R01AG032289, and R01AG048234), and the UCSF ADRC P30AG062422. Our work was also supported by the

References

- Chen F-T, Hopman RJ, Huang C-J, Chu C-H, Hillman CH, Hung T-M, et al. The effect of exercise training on brain structure and function in older adults: a systematic review based on evidence from randomized control trials. *J Clin Med.* (2020) 9(4):914. doi: 10.3390/JCM9040914
- Hörder H, Johansson L, Guo X, Grimby G, Kern S, Östling S, et al. Midlife cardiovascular fitness and dementia: a 44-year longitudinal population study in women. *Neurology.* (2018) 90(15):e1298–305. doi: 10.1212/WNL.00000000000005290
- Nystoriak M, Bhatnagar A. Cardiovascular effects and benefits of exercise. *Front Cardiovasc Med.* (2018) 5:135. doi: 10.3389/FCVM.2018.00135
- Buchman AS, Boyle PA, Yu L, Shah RC, Wilson RS, Bennett DA. Total daily physical activity and the risk of AD and cognitive decline in older adults. *Neurology.* (2012) 78(17):1323–9. doi: 10.1212/WNL.0B013E3182535D35
- Casaleto KB, Staffaroni AM, Wolf A, Appleby B, Brushaber D, Coppola G, et al. Active lifestyles moderate clinical outcomes in autosomal dominant frontotemporal degeneration. *Alzheimer's Dement.* (2020) 16(1):91–105. doi: 10.1002/alz.12001
- Noah JA, Spierer DK, Gu J, Bronner S. Comparison of steps and energy expenditure assessment in adults of fitbit tracker and ultra to the actual and indirect calorimetry. *J Med Eng Technol.* (2013) 37(7):456–62. doi: 10.3109/03091902.2013.831135
- Dontje ML, de Groot M, Lengton RR, van der Schans CP, Krijnen WP. Measuring steps with the fitbit activity tracker: an inter-device reliability study. *J Med Eng Technol.* (2015) 39(5):286–90. doi: 10.3109/03091902.2015.1050125
- Diaz KM, Krupka DJ, Chang MJ, Peacock J, Ma Y, Goldsmith J, et al. Fitbit™: an accurate and reliable device for wireless physical activity tracking. *Int J Cardiol.* (2015) 185:138–40. doi: 10.1016/J.IJCARD.2015.03.038
- Jones D, Crossley K, Dascombe B, Hart HF, Kemp J. Validity and reliability of the fitbit FLEX™ and actigraph gt3x+ at jogging and running speeds. *Int J Sports Phys Ther.* (2018) 13(5):860. doi: 10.26603/ijsp20180860
- Reid RER, Insogna JA, Carver TE, Comptour AM, Bewski NA, Sciortino C, et al. Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. *J Sci Med Sport.* (2017) 20(6):578–82. doi: 10.1016/J.JSAMS.2016.10.015
- Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act.* (2015) 12(1):1–22. doi: 10.1186/S12966-015-0314-1
- Hagiwara A, Ito N, Sawai K, Kazuma K. Validity and reliability of the physical activity scale for the elderly (PASE) in Japanese elderly people. *Geriatr Gerontol Int.* (2008) 8(3):143–51. doi: 10.1111/J.1447-0594.2008.00463.X
- Washburn RA, Ficker JL. Physical activity scale for the elderly (PASE): the relationship with activity measured by a portable accelerometer. *J Sports Med Phys Fitness.* (1999) 39(4):336–40. PMID: 10726435
- Chad KE, Reeder BA, Harrison EL, Ashworth NL, Sheppard SM, Schultz SL, et al. Profile of physical activity levels in community-dwelling older adults. *Med Sci Sports Exerc.* (2005) 37(10):1774–84. doi: 10.1249/01.MSS.0000181303.51937.9C
- Logan SL, Gottlieb BH, Maitland SB, Meegan D, Spriet LL. The physical activity scale for the elderly (PASE) questionnaire: does it predict physical

Alzheimer's Association (AARG-20-683875, PI: KC) and the Larry L. Hillblom Foundation Network (2014-A-004-NET).

Acknowledgments

The authors would like to thank the National Institute on Aging, the Alzheimer's Association, and the Larry L. Hillblom Foundation Network for funding this study. We would also like to thank the participants and family members involved.

Conflict of interest

Joel H. Kramer receives royalties from Pearson's Inc. The authors report no other conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- health? *Int J Environ Res Public Heal.* (2013) 10(9):3967–86. doi: 10.3390/IJERPH10093967
16. Washburn RA, Smith KW, Jette AM, Janney CA. The physical activity scale for the elderly (PASE): development and evaluation. *J Clin Epidemiol.* (1993) 46(2):153–62. doi: 10.1016/0895-4356(93)90053-4
17. Stewart L, Mills KM, King AC, Haskell WL, Gillis D, et al. CHAMPS physical activity questionnaire for older adults: outcomes for interventions. *Med Sci Sports Exerc.* (2001) 33(7):1126–41. doi: 10.1097/00005768-200107000-00010
18. Hekler EB, Buman MP, Haskell WL, Conway TL, Cain KL, Sallis JF, et al. Reliability and validity of CHAMPS self-reported sedentary-to-vigorous intensity physical activity in older adults. *J Phys Act Heal.* (2012) 9(2):225–36. doi: 10.1123/JPAH.9.2.225
19. Feldman LS, Kaneva P, Demyttenaere S, Carli F, Fried GM, Mayo NE. Validation of a physical activity questionnaire (CHAMPS) as an indicator of postoperative recovery after laparoscopic cholecystectomy. *Surgery.* (2009) 146(1):31–9. doi: 10.1016/J.SURG.2009.02.019
20. Prince SA, Cardilli L, Reed JL, Saunders TJ, Kite C, Douillette K, et al. A comparison of self-reported and device measured sedentary behaviour in adults: a systematic review and meta-analysis. *Int J Behav Nutr Phys Act.* (2020) 17(1):1–17. doi: 10.1186/S12966-020-00938-3
21. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. *Br J Sports Med.* (2003) 37(3):197–206. doi: 10.1136/BJSM.37.3.197
22. Sims J, Smith F, Duffy A, Hilton S. The vagaries of self-reports of physical activity: a problem revisited and addressed in a study of exercise promotion in the over 65s in general practice. *Fam Pract.* (1999) 16(2):152–7. doi: 10.1093/FAMPRA/16.2.152
23. Mandolesi L, Polverino A, Montuori S, Foti F, Ferraioli G, Sorrentino P, et al. Effects of physical exercise on cognitive functioning and wellbeing: biological and psychological benefits. *Front Psychol.* (2018) 9:509. doi: 10.3389/FPSYG.2018.00509
24. Rikli RE, Edwards DJ. Effects of a three-year exercise program on motor function and cognitive processing speed in older women. *Res Q Exerc Sport.* (2013) 62(1):61–7. doi: 10.1080/02701367.1991.10607519
25. Kramer JH, Jurik J, Sha SJ, Rankin KP, Rosen HJ, Johnson JK, et al. Distinctive neuropsychological patterns in frontotemporal dementia, semantic dementia, and Alzheimer disease. *Cogn Behav Neurol.* (2003) 16(4):211–8. doi: 10.1097/00146965-200312000-00002
26. Possin KL, Laluz VR, Alcantar OZ, Miller BL, Kramer JH. Distinct neuroanatomical substrates and cognitive mechanisms of figure copy performance in Alzheimer's disease and behavioral variant frontotemporal dementia. *Neuropsychologia.* (2011) 49(1):43–8. doi: 10.1016/J.NEUROPSYCHOLOGIA.2010.10.026
27. Kerchner GA, Racine CA, Hale S, Wilhelm R, Laluz V, Miller BL, et al. Cognitive processing speed in older adults: relationship with white matter integrity. *PLoS One.* (2012) 7(11). doi: 10.1371/journal.pone.0050425
28. Kramer JH, Mungas D, Possin KL, Rankin KP, Boxer AL, Rosen HJ, et al. NIH EXAMINER: conceptualization and development of an executive function battery. *J Int Neuropsychol Soc.* (2014) 20(1):11. doi: 10.1017/S1355617713001094
29. American Heart Association. All about heart rate (pulse). Available at: <https://www.heart.org/en/health-topics/high-blood-pressure/the-facts-about-high-blood-pressure/all-about-heart-rate-pulse> (Accessed October 25, 2021).
30. Fox K, Borer JS, Camm AJ, Danchin N, Ferrari R, Lopez Sendon JL, et al. Resting heart rate in cardiovascular disease. *J Am Coll Cardiol.* (2007) 50(9):823–30. doi: 10.1016/J.JACC.2007.04.079
31. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE Trans Med Imaging.* (1998) 17(1):87–97. doi: 10.1109/42.668698
32. Ashburner J, Friston KJ. Unified segmentation. *Neuroimage.* (2005) 26(3):839–51. doi: 10.1016/J.NEUROIMAGE.2005.02.018
33. Ashburner J. A fast diffeomorphic image registration algorithm. *Neuroimage.* (2007) 38(1):95–113. doi: 10.1016/J.NEUROIMAGE.2007.07.007
34. Mazziotta JC, Toga AW, Evans A, Fox P, Lancaster J. A probabilistic atlas of the human brain: theory and rationale for its development: the international consortium for brain mapping (ICBM). *Neuroimage.* (1995) 2(2):89–101. doi: 10.1006/NIMG.1995.1012
35. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage.* (2006) 31(3):968–80. doi: 10.1016/J.NEUROIMAGE.2006.01.021
36. Erickson KI, Voss MW, Prakash RS, Basak C, Szabo A, Chaddock L, et al. Exercise training increases size of hippocampus and improves memory. *Proc Natl Acad Sci U S A.* (2011) 108(7):3017–22. doi: 10.1073/PNAS.1015950108
37. Cohen J. *Statistical power analysis for the behavioral sciences.* New York City: Routledge (1988).
38. Colbert LH, Matthews CE, Havighurst TC, Kim K, Schoeller DA. Comparative validity of physical activity measures in older adults. *Med Sci Sports Exerc.* (2011) 43(5):867–76. doi: 10.1249/MSS.0B013E3181FC7162
39. Stahl ST, Insana SP. Caloric expenditure assessment among older adults: criterion validity of a novel accelerometry device. *J Health Psychol.* (2014) 19(11):1382–7. doi: 10.1177/1359105313490771
40. Snowden M, Steinman L, Mochan K, Grodstein F, Prohaska TR, Thurman DJ, et al. Effect of exercise on cognitive performance in community-dwelling older adults: review of intervention trials and recommendations for public health practice and research. *J Am Geriatr Soc.* (2011) 59(4):704–16. doi: 10.1111/J.1532-5415.2011.03323.X
41. Erickson KI, Raji CA, Lopez OL, Becker JT, Rosano C, Newman AB, et al. Physical activity predicts gray matter volume in late adulthood: the Cardiovascular Health Study. *Neurology.* (2010) 75(16):1415–22. doi: 10.1212/WNL.0B013E3181F88359
42. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Young Yoo J, et al. Accuracy of fitbit devices: systematic review and narrative syntheses of quantitative data. *JMIR mHealth uHealth.* (2018) 6(8):e10527. doi: 10.2196/10527
43. Takacs J, Pollock CL, Guenther JR, Bahar M, Napier C, Hunt MA. Validation of the Fitbit One activity monitor device during treadmill walking. *J Sci Med Sport.* (2014) 17(5):496–500. doi: 10.1016/J.JSAMS.2013.10.241
44. Northey JM, Cherbuin N, Pampa KL, Smee DJ, Rattray B. Exercise interventions for cognitive function in adults older than 50: a systematic review with meta-analysis. *Br J Sports Med.* (2018) 52(3):154–60. doi: 10.1136/BJSPORTS-2016-096587
45. Yaffe K, Barnes D, Nevitt M, Lui LY, Covinsky K. A prospective study of physical activity and cognitive decline in elderly women: women who walk. *Arch Intern Med.* (2001) 161(14):1703–8. doi: 10.1001/ARCHINTE.161.14.1703
46. Pentikainen H, Savonen K, Ngandu T, Solomon A, Komulainen P, Paajanen T, et al. Cardiorespiratory fitness and cognition: longitudinal associations in the FINGER study. *J Alzheimers Dis.* (2019) 68(3):961–8. doi: 10.3233/JAD-180897
47. Vidoni ED, Johnson DK, Morris JK, Van Sciver A, Greer CS, Billinger SA, et al. Dose-response of aerobic exercise on cognition: a community-based, pilot randomized controlled trial. *PLoS One.* (2015) 10(7):e0131647. doi: 10.1371/JOURNAL.PONE.0131647
48. Harada ND, Chiu V, King AC, Stewart AL. An evaluation of three self-report physical activity instruments for older adults. *Med Sci Sports Exerc.* (2001) 33(6):962–70. doi: 10.1097/00005768-200106000-00016
49. Baranowski T. Validity and reliability of self report measures of physical activity: an information-processing perspective. *Res Q Exerc Sport.* (2013) 59(4):314–27. doi: 10.1080/02701367.1988.10609379
50. Friedman HS, Martin LR, Tucker JS, Criqui MH, Kern ML, Reynolds CA. Stability of physical activity across the lifespan. *J Health Psychol.* (2008) 13(8):1092–104. doi: 10.1177/1359105308095963