Using consumer-friendly wearables to correlate patient and technology-reported physical activity in healthy seniors

Manea, Vlad; Berrocal, Allan; Wac, Katarzyna

Published in:
Procedia Computer Science

DOI:
10.1016/j.procs.2020.07.036

Publication date:
2020

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Manea, V., Berrocal, A., & Wac, K. (2020). Using consumer-friendly wearables to correlate patient and technology-reported physical activity in healthy seniors. *Procedia Computer Science*, *175*, 245-252. https://doi.org/10.1016/j.procs.2020.07.036

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 175 (2020) 245–252

Procedia Computer Science

www.elsevier.com/locate/procedia

The 17th International Conference on Mobile Systems and Pervasive Computing (MobiSPC)
August 9-12, 2020, Leuven, Belgium

Using consumer-friendly wearables to correlate patient and technology-reported physical activity in healthy seniors

Vlad Manea[a], Allan Barrocal[b], Katarzyna Wac[a,b,*]

[a] *Quality of Life Technologies Lab, Department of Computer Science, University of Copenhagen, Sigurdsgade 41, Copenhagen 2200, Denmark*
[b] *Quality of Life Technologies Lab, Institute of Service Science, University of Geneva, 7 Route de Drize, Geneva 1227, Switzerland*

Abstract

A leading risk factor for chronic disease is physical inactivity. In efforts to assess physical activity and inform designs for prevention, health professionals currently use inexpensive, but subjective validated scales, or objective, but expensive research-grade wearables. In the meanwhile, individuals increasingly use affordable consumer-friendly wearable devices that can objectively monitor behaviours while daily life unfolds. However, the relationships between their outcomes and the validated scales are yet to be calibrated. We report our results from a study on 31 seniors from Hungary and Spain (mean age 70.6 +/- 3.2). Our study quantified the relations between physical activity outcomes, as patient-reported through 53 answers (1.71 +/- 0.96 / person) on the International Physical Activity Questionnaire (IPAQ) with a 7-day recall period, and 5615 days (mean 181.1 +/- 179.2 days collected / person) technology-reported by Fitbit Charge 2. The wearables monitored daily life behaviours of physical activity and sleep for long durations (7 to 120 days). We found strong Spearman correlations between light and moderate IPAQ physical activity in the domestic activity domain, and light-fair intensity Fitbit physical activity (e.g., $r_S = 0.88$, $p < 0.005$). We also found negative moderate-strong correlations between Fitbit sedentary duration and all IPAQ physical activity domains and intensities (e.g., $r_S = 0.64$, $p < 0.005$). We obtained increasingly stronger relationships across all IPAQ domains and Fitbit intensities by monitoring physical activity beyond the scale recall period, quantifying physical activity relative to all activities of the day, and including sleep. Our findings inform the design of longitudinal observations and personalized, focused, and potentially effective interventions for physical activity in seniors.

Keywords: observational study; healthy senior; physical activity; statistical correlation; questionnaire; IPAQ; consumer-friendly wearable; Fitbit

* Corresponding author. Tel.: +41-22-379-024
E-mail address: katarzyna.wac@unige.ch

10.1016/j.procs.2020.07.036

1. Introduction

Chronic diseases represent a significant share of the burden of disease globally [21]. They are responsible for 86% of all deaths [36]. In Europe, chronic diseases affect over 80% of adults over 65 and incur 70% of the increasing healthcare costs [3]. The most common chronic diseases are cardiovascular, pancreatic, pulmonary, and neoplastic.

Unhealthy lifestyle and behaviours, such as physical inactivity, poor nutrition, and tobacco intake, explain up to 50% of the risk of chronic disease [12]. A leading behaviour of risk is physical inactivity. There is *"overwhelming evidence that proves the notion that reductions in daily physical activity are primary causes of chronic diseases"* [5]. However, *"the evidence is currently insufficiently precise to warrant separate guidelines for each specific disease, but it is strong enough to cover all health outcomes"* [35].

The gold standard for the measurement of physical activity is the subjective patient-reported outcome (PRO, [19]) administered as a questionnaire based on a qualitative scale, statistically validated on a population of interest. However, validated scales for physical activity have the inherent shortcomings of PROs: they are inconvenient, infrequent, memory-biased, socially conditioned, and qualitative.

Research-grade wearables measure physical activity objectively. They provide quantitative, technology-reported physical activity outcomes (TechRO, [19]), such as acceleration and heart rate, and have been clinically validated. However, they are uncomfortable and expensive [30]. Several studies used PROs from validated scales and TechROs from research-grade wearables to quantify the relationships between subjective and objective physical activity [11, 34]. However, their participants wore the devices for a *short-term* period, and without owning the devices.

Consumer-friendly wearables measure continuously, accurately, and objectively quantitative TechROs of physical activity during daily life for long, or *longitudinal* periods [9]. Also, more individuals opt for consumer-friendly wearable devices; the market size will likely double by 2022 [16]. However, the few studies using exclusively consumer-friendly wearables to measure longitudinal physical activity, e.g., [6], focused on younger populations.

To our knowledge, there are no studies aimed at quantifying the relationships between physical activity PROs (obtained from validated scales) and TechROs (collected from consumer-friendly wearables) at different intensities in longitudinal, daily, and free-living conditions for seniors. Our study observed N = 31 healthy seniors along 2017-2019. They provided 53 PRO answers and collected 5615 TechRO days of physical activity and sleep. We included sleep in the TechROs to model the interdependence of active, sedentary, and sleep duration during the day [26].

From over 80 scales that provide physical activity PROs [22], we chose the International Physical Activity Questionnaire (IPAQ). The IPAQ is *"developed to measure health-related physical activity in populations"* [14] and has been validated on seniors. In our study, it was feasible to administer the long (and more detailed) variant of the IPAQ.

From over 200 models of consumer-friendly wearables that provide physical activity TechROs [15], we chose Fitbit (Fitbit, Inc.). Fitbit aims at motivating consumers to *"reach health and fitness goals by tracking activity, exercise, sleep, weight, and more"* [10]. It has been selected for Digital Health Software Precertification by the US FDA [32]. Fitbit monitors daily life behaviours accurately and continuously, operationalizes the critical human factors for prolonged wear by senior end-users, and facilitates reliable behavioural data collection. We selected Fitbit Charge 2, a watch model with a user-friendly display, previously validated with seniors [31].

Our paper is structured as follows. Section 2 reviews related work. Section 3 describes our method. Section 4 foregrounds our results. Section 5 discusses our findings. The last section concludes the paper.

2. Related Work

Validated scales of physical activity and sleep have only moderate validity and reliability [24]. Sims et al. have shown that seniors reporting physical activity overestimate the amount undertaken (N = 20, mean age 72.2, [28]). Also, two studies by Anderson (N = 421, ages 87-89, [4]) and Van Der Berg (N = 69, ages 57-97, [33]) have shown that subjective sleep is less reliable than objective sleep.

Several studies have compared PROs to TechROs by using a validated scale in tandem with a research-grade wearable. For example, Garriguet et al. have found a Spearman $r_S = 0.23$ correlation between PRO and TechRO moderate and vigorous activity. The latter was reported by Actical accelerometers worn for seven days (N = 112, ages 18-79, [11]). Wanner et al. have obtained a Spearman $r_S = 0.41$ correlation between vigorous physical activity from

PROs and TechROs. They used ActiGraph GTX3+ accelerometers worn for eight days (N = 346, mean age 54.6, [34]). These studies have focused on younger populations for a *short-term* period (typically one week).

Other studies have compared TechROs from both research-grade and consumer-friendly wearables by requiring the participants to wear them simultaneously. Gomersall et al. have reported Spearman $r_S = 0.80$ correlations between ActiGraph GTX3+ and Fitbit One for moderate and vigorous physical activity (N = 29, mean age 39.6, [13]). Participants wore both devices on the hips for two sessions of seven days each. Brewer et al. have found a Pearson $r = 0.69$ correlation between moderate and vigorous physical activity measured by ActiGraph GT3X+ and Fitbit devices (N = 50, university students, [6]). Participants wore the devices on the hip and wrist for seven days. These studies have focused on younger populations and covered a short-term period as well due to the research-grade wearable and the discomfort from wearing both devices.

To our knowledge, no studies quantified the relationships between physical activity PROs and TechROs at different intensities in longitudinal, daily, and free-living conditions for seniors. In our study, we used PROs from the IPAQ validated scale and TechROs from the Fitbit Charge 2 consumer wearable.

Studies have previously assessed physical activity by using the IPAQ. Silsbury et al. have reported that IPAQ has very good reliability and good agreement with accelerometer measures [27]. Van Poppel et al. included IPAQ in the list of scales appropriate for measuring their intended dimension of physical activity [24]. Prior studies in Hungary and Spain, the countries of our research, have also used the IPAQ. In Hungary, Makai et al. used the IPAQ and found differences in physical activity levels by sociodemographic parameters in N = 910 adults [18]. In Spain, Roman-Viñas et al. have found that the IPAQ has good reliability for all its intensities and domains in N = 110 adults [25].

Previous studies measured the accuracy of Fitbit consumer-friendly devices in reporting daily life behaviours of physical activity and sleep. Ferguson et al. have found the Fitbit One, Fitbit Zip, and Withings Pulse to perform the most reliably across physical activity and sleep constructs (N = 21 adults, [9]). Brewer et al. have reported that Fitbit (Charge HR, Charge, Flex, Surge, Zip, Alta) agrees with the ActiWatch GT3X+ accelerometer in assessing active minutes of physical activity in a study run for 7 days (N = 50 university students, [6]). For seniors, Tedesco et al. compared the Fitbit Charge 2 and the Garmin Vivosmart HR+ in free-living environments in a senior cohort (N = 20, age over 65, [31]). They have found that Fitbit had better overall results in step count, energy expenditure, and sleep duration. Paul et al. have also found that Fitbit One and Flex monitor and provide feedback on steps accurately for seniors (N = 32, age over 60, [23]).

The user experience with the wearable of the target group is an essential factor affecting the duration of wear. McMahon et al. have reported that Fitbit One (N = 95, 70+ years old, [20]) is easy to use on the System Usability Scale [7]. Steinert et al. have found that seniors (N = 20, age over 60) graded the Fitbit the highest in usability.

Fitbit provides a well-documented application programming interface (API). The Fitbit API exposes behavioural data for the entire day, including physical activity and sleep.

3. Methods

Our study had three objectives. First, we aimed at quantifying the relationships between PRO intensities and domains of physical activity and TechRO intensities of physical activity. Second, we aimed at identifying stronger relationships beyond the typical questionnaire recall period of several days. Third, we aimed at reporting the quantified relations from data collected while daily life unfolds.

3.1. Study Design

We conducted the study as part of the EU AAL Caregiver and Me (CoME, No. 14-7) research project and software application (2015-2020) for self-management aimed at healthy seniors [2]. The goal of CoME was to reduce the risk of developing dementia [17] by monitoring its risk factors, including physical activity. The institutional review board at the University of Geneva had approved the study in 2016.

Seniors who owned a smartphone, or were willing to receive one for ownership, were invited to a care centre in their city (Lleida, Spain and Budapest, Hungary). A total of 42 individuals (age: 68.78 +/- 6.30; gender: 26 female and 16 male; location: 28 in Spain and 14 in Hungary) agreed to participate from January 2017 to December 2019. They signed written consent. Their identities were pseudonymized.

Upon arrival at the care centre, participants attended an informational workshop about the project. They optionally received a smartphone and a Fitbit Charge 2 consumer wearable as their own (for the study duration and beyond). At the beginning and in subsequent visits throughout 2018 and 2019, the participants answered several questionnaires, including the IPAQ. They were not explicitly informed about when they would fill the questionnaires to avoid any activity pattern change before the visit. Caregivers assisted them throughout the process. Three distinct periods of answers, or *waves*, have resulted: wave 1: mid-2018, wave 2: end-2018 and start-2019, and wave 3: mid-2019.

3.2. Patient-Reported Physical Activity (IPAQ)

The IPAQ long contains 25 questions about the typical duration and frequency of physical activity at *walking*, *moderate*, and *vigorous* intensities, in several domains: *work*, *transport*, *domestic and garden*, and *leisure*. The questions refer to a recall period of one week. The scale provides separate scores of physical activity for each intensity, domain, and overall, derived from the cumulative weekly duration and the energy expenditure. We calculated the IPAQ weekly durations for the intensities and domains (11 variables corresponding to the non-score rows in Table 1). We separately included the scores for the intensities and domains (7 variables). We then added the overall score (1 variable). We obtained a total of 19 PRO variables represented as the rows in Table 1.

We processed the individuals' answers by adhering to the data cleaning, maximum values for excluding outliers, and minimum values for the duration of activity from the IPAQ scoring guideline [14]. The guide does not provide a threshold for converting the *"duration reported as weekly (not daily) to daily into an average daily time"*. For example, if a senior individual reported seven hours of vigorous physical activity per day, the duration would likely reflect one hour per day. We allowed at most 7 hours of physical activity per day in any intensity by dividing all excessive durations by 7 days.

3.3. Technology-Reported Physical Activity (Fitbit)

We assessed the TechRO behaviours of physical activity and sleep. We derived behaviour variables in two amounts, *absolute* and *relative*, with separate semantics. Absolute variables refer to each behaviour separately. Relative variables reflect the difference between a behaviour and the (geometric) mean of all behaviours during the 24 hours of the day. The relative amount was motivated by the interdependence of behaviours during the day [26].

In the absolute amount, we derived the variables directly. For physical activity, we calculated the daily distance (denoted *distance*), energy expenditure (*energy*), step count (*steps*), sedentary duration (*sedentary*), and the duration at three intensities (*light*, *moderate*, and *vigorous*) as reported by Fitbit (7 variables). As Fitbit had not published intensity thresholds, we also derived the cumulative durations in sedentary and light (*sedentary+light*), light and fair (*light+fair*), and fair and vigorous (*fair+vigorous*) intensities (3 variables). We also calculated the total daily active duration (*active*) cumulating the light, fair, and vigorous durations (1 variable). For sleep, we included the entire sleep duration of the day (1 variable). We derived 12 TechRO variables in the absolute amount.

In the relative amount, we derived variables denoting compositional components of physical activity intensities and sleep throughout the day. We derived variables for each component of the centred log-ratio (*CLR*, [1]) transformation. The CLR is a symmetric transformation that does not require a reference component behaviour. We computed the CLRs of two separate compositions: (1) from all physical activity durations (*PAC*) (4 variables) and (2) from all physical activity durations and the sleep duration (*PASC*) (5 variables). We included both relative amounts as the CLRs of a composition are not preserved in sub-compositions [1], but some studies may not be able to monitor sleep. We derived 9 TechRO variables in the absolute amount.

We considered *valid* only those days where the duration covered by wearable monitoring was at least 21 hours. We allowed at most three hours of missing data for device battery charging and handling (15-20 minutes to 2 hours).

Then we derived intervals with fixed durations of 7, 14, 21, 28, 60, 90, and 120 days to balance the number of included days in the analysis with the available intraday monitoring quality. The choice of 7 days for the lower bound was motivated by the need to acquire enough representative data for daily life, the IPAQ recall period of 7 days, and the significant improvements in Fitbit accuracy for active minutes from 7 days onwards [6]. The choice of increasing intervals to the upper bound of 120 days reflected the duration of a wave, a large number of valid days per person (mean 181.1 days), but also the high variance (σ = 179.9 days).

We only included intervals with at least 70% of their days valid, such that both weekdays and weekends were expectedly present in a week; the limit is compatible with previously reported consumer wearable use in seniors [8]. For each interval and variable, we aggregated the mean and geometric mean for the daily absolute and relative amounts, respectively. We included 21 TechRO variables in total, represented as columns in Table 1.

3.4. Statistical Analysis

We aligned in time PROs with TechROs by using the administration date of the PRO answer and the end date of the TechRO measurement interval, with a leeway of at most the interval duration due to scarce exact matches. For each participant, we included only the last alignment in a wave, to discard repeated answers within a few minutes and reduce bias towards overly diligent responders. When we aligned PROs with TechROs of increasing durations, the number of paired observations decreased; we thus required a minimum of 10 observations.

We applied the Spearman rank test measuring the direction and strength of a correlation [29]. We chose this test because the PRO and TechRO assessments were interdependent (they referred to the same participant), not all variables were normally distributed (Shapiro Wilk normality test $p < 0.05$), and the variables had distinct units of measurement (making rank correlation appropriate). The $19 \times 21 = 399$ correlations are depicted in Table 1.

The correlation strength and direction are denoted by a signed real value in the interval $[-1, 1]$, denoted r_S. We considered correlations (r_S) to be *weak* ($r_S \in [0, 0.25]$), *moderate-weak* ($r_S \in (0.25, 0.45]$), *moderate* ($r_S \in (0.45, 0.55]$), *moderate-strong* ($r_S \in (0.55, 0.75]$), or *strong* ($r_S \in (0.75, 1]$) in absence of consensus in the literature. For brevity, we reported only stronger or adjusted correlations. Table 1 depicts the correlations in its cells.

We reported both statistically significant and non-significant correlations. We adjusted the significance following a partial Bonferroni correction by dividing the significance threshold α by 12, the maximum number of interval-agnostic TechRO variables in a given amount, and then by 7, the number of interval durations. The small sample size motivated our choice to balance between no adjustment ($\alpha = 0.05$) and full Bonferroni adjustment ($\alpha = 0.05$ / 19 PRO variables / 21 TechRO variables / 7 interval durations). We placed the significant correlations in three levels of increasing significance: $*$ for $p < 0.05$ (unadjusted), 1 for $p < 0.05 \div 12$ (adjusted), and 2 for $p < 0.05 \div 12 \div 7$ (adjusted). We reported the strength and direction of all correlations regardless of significance. For brevity, we only reported one correlation per TechRO interval duration in the cells of Table 1. If two correlations differed by the TechRO interval duration, we chose the one with a higher level of significance and indicated its duration.

4. Results

We included in further analysis 31 out of 42 initial participants who had filled IPAQ PROs (mean age 70.6 +/- 3.2). The included participants had contributed 53 IPAQ answers (1.71 +/- 0.96 / person) and provided 9836 Fitbit days (317.3 +/- 256.9 / person) from which 5615 were valid days (181.1 +/- 179.2 / person).

As observed in Table 1, the energy expenditure had adjusted moderate-strong correlations with the work moderate and vigorous physical activity. The distance had adjusted moderate-strong correlations with the leisure moderate. The steps had an adjusted moderate-strong correlation with the walking score and the leisure moderate physical activity.

For the absolute physical activity intensities, sedentary duration had negative moderate-strong correlations with the leisure and work walking. Light physical activity duration had adjusted moderate-strong correlations with domestic and leisure moderate physical activity, as well as the score for moderate physical activity. Cumulative light+fair duration had a strong adjusted correlation with the score for moderate physical activity ($r_S = 0.79^1$). Fair duration correlated negatively and moderate-strongly with the moderate physical activity score and the domestic moderate physical activity. Cumulative fair+vigorous duration correlated negatively and moderate-strongly with garden moderate physical activity. The vigorous duration had moderate-strong negative correlations with the domestic and garden score, and garden moderate physical activity. The total active duration correlated moderately-strongly with leisure and moderate scores.

For the PAC relative amount, we report an adjusted negative moderate-strong correlation between the sedentary log-ratio and the work vigorous physical activity as well as unadjusted negative moderate-strong correlations for the transport and overall scores. Light log-ratio had a strong adjusted correlation with the domestic moderate physical activity ($r_S = 0.88^1$) and other moderate-strong correlations in the domestic and garden domain, and the score for moderate physical activity. Fair log-ratio had a strong correlation with garden vigorous physical activity ($r_S = 0.75^*$) and moderate-strong correlations with moderate and vigorous physical activity. Vigorous log-ratio had a positive correlation with the domestic moderate physical activity and the leisure vigorous physical activity.

For the PASC relative amount (including sleep), sedentary log-ratio had negative and moderate-strong correlations in the work and domestic+garden domains. Light log-ratio had strong correlations with domestic moderate physical activity ($r = 0.84^1$) and the score for moderate physical activity ($r_S = 0.75^*$). Fair log-ratio had strong positive correlations with the work moderate physical activity ($r_S = 0.79^*$) and the vigorous score ($r_S = 0.78^*$). Vigorous log-ratio had a strong positive correlation with the domestic moderate physical activity ($r_S = 0.77^1$), but a negative moderate-strong correlation with the leisure vigorous physical activity.

We report higher correlations in the following objective-subjective pairs as compared to the other pairs where only the TechRO intensity changed. In the absolute amount, we found the strongest correlations between (1) cumulative light+fair duration and moderate physical activity at $r_S = 0.79^1$, (2) light duration and walking at $r_S = 0.64^1$ and moderate physical activity at $r_S = 0.71^1$, (3) fair duration and vigorous physical activity at $r_S = 0.44^*$, (4) cumulative fair+vigorous and vigorous physical activity at $r_S = 0.43^*$, (5) vigorous duration and moderate physical activity at $r_S = 0.46$ (non-significant). In the PAC and PASC relative amounts (1) light CLR correlated the highest with the moderate physical activity at $r_S = 0.68^*$ and $r_S = 0.75^*$, (2) fair CLR correlated the strongest with the vigorous physical activity at $r_S = 0.6^*$ and $r_S = 0.78^*$, and (3) vigorous CLR correlated the strongest again with the moderate physical activity at $r_S = 0.66^*$ and $r_S = 0.63^*$.

Table 1. Rank correlations (cells) between aligned PROs (rows) and TechROs (columns) of physical activity (Spearman r_S)

	Energy	Distance	Steps	Sedentary	Sedentary+Light	Light	Light+Fair	Fair	Fair+Vigorous	Vigorous	Active	Sleep	PAC sedentary	PAC light	PAC fair	PAC vigorous	PASC sedentary	PASC light	PASC fair	PASC vigorous	PASC sleep
Work walking	$0.46^{120,1}$	$0.48^{120,1}$	$0.52^{120,1}$	$0.56^{21,*}_{-}$	0.14_{-}	0.33	0.04_{-}	0.15_{-}	0.06_{-}	0.03	$0.45^{14,*}$	0.1_{-}	$0.51^{14,*}_{-}$	0.21	0.21_{-}	0.02	$0.56^{14,*}_{-}$	0.19	0.34_{-}	0.04	0.05
Work moderate	$0.55^{120,2}$	0.32	$0.4^{120,*}$	$0.49^{21,*}_{-}$	0.18_{-}	0.22	0.01	0.14_{-}	0.14_{-}	0.01_{-}	$0.43^{28,*}$	0.12	$0.5^{14,*}_{-}$	0.16	$0.74^{120,*}$	0.18	$0.55^{14,*}_{-}$	0.26	$0.79^{90,*}$	0.17	0.23
Work vigorous	$0.57^{120,2}$	$0.34^{120,*}$	$0.34^{120,*}$	$0.39^{90,*}_{-}$	$0.54^{60,*}_{-}$	0.1	0.15_{-}	$0.51^{14,*}$	$0.53^{120,*}$	$0.53^{120,*}$	$0.56^{90,*}$	0.35	$0.61^{14,1}_{-}$	0.16_{-}	$0.62^{120,*}$	0.08_{-}	$0.67^{14,1}_{-}$	0.01_{-}	$0.55^{7,*}_{-}$	0.13_{-}	0.21_{-}
Work score	$0.49^{120,1}$	$0.4^{120,*}$	$0.55^{90,1}$	$0.59^{21,*}_{-}$	$0.52^{60,*}_{-}$	0.26	0.03_{-}	0.06	0.0	0.11	$0.47^{14,*}$	0.05	$0.48^{28,*}_{-}$	0.0	0.38_{-}	0.03	$0.59^{14,*}_{-}$	0.12	$0.57^{7,*}_{-}$	0.03	0.15
Transport walking	0.26	$0.38^{120,*}$	$0.48^{120,1}$	$0.64^{60,1}_{-}$	0.25_{-}	$0.49^{60,*}$	0.32	0.08_{-}	0.44_{-}	0.48_{-}	$0.46^{14,*}$	0.0	$0.5^{14,*}_{-}$	$0.56^{60,*}$	0.42_{-}	0.18	$0.46^{14,*}_{-}$	$0.63^{90,*}$	0.3_{-}	0.22	0.0_{-}
Transport cycling	$0.4^{21,*}$	$0.44^{60,*}$	$0.47^{28,*}$	$0.49^{21,*}_{-}$	$0.38^{14,*}_{-}$	0.1	0.09	0.07	0.17	0.18	$0.49^{28,*}$	0.02_{-}	$0.59^{14,*}_{-}$	0.34	0.11_{-}	0.05	0.14_{-}	0.46	0.36_{-}	0.03	0.31_{-}
Transport score	0.31	$0.37^{14,*}$	$0.44^{120,*}$	$0.58^{60,*}_{-}$	$0.53^{21,*}_{-}$	0.34	0.38	0.04_{-}	0.3_{-}	0.37_{-}	$0.56^{90,*}$	$0.55^{21,*}$	$0.62^{14,*}_{-}$	$0.63^{120,*}$	0.42_{-}	0.22	0.35_{-}	$0.64^{90,*}$	0.46_{-}	0.25	0.04_{-}
Domestic moderate	$0.49^{120,1}_{-}$	0.22	$0.46^{28,*}$	0.25_{-}	0.06	$0.72^{60,1}$	$0.53^{90,*}$	$0.55^{21,*}_{-}$	$0.58^{21,*}_{-}$	$0.53^{14,*}_{-}$	0.06	0.07_{-}	0.52	$0.88^{120,1}$	0.06_{-}	$0.58^{120,*}$	$0.66^{120,*}$	$0.84^{120,1}$	0.05_{-}	$0.77^{28,1}$	$0.62^{28,*}$
Garden moderate	0.04_{-}	0.15	0.19	0.16_{-}	0.11	$0.53^{120,1}$	$0.58^{120,*}$	$0.6^{120,*}_{-}$	$0.69^{90,*}_{-}$	$0.64^{90,*}_{-}$	0.11	0.06_{-}	$0.5^{21,*}$	$0.63^{120,*}$	0.06_{-}	$0.54^{120,*}$	$0.63^{120,*}$	$0.69^{120,*}$	0.15_{-}	$0.58^{120,*}$	0.34
Garden vigorous	$0.44^{120,*}$	$0.44^{120,*}$	$0.46^{120,*}$	$0.42^{90,*}_{-}$	$0.45^{90,*}_{-}$	0.24	0.11_{-}	0.02	0.1_{-}	0.04	0.0_{-}	$0.43^{120,*}$	0.06	0.06_{-}	$0.75^{120,*}$	0.18	$0.47^{21,*}_{-}$	0.14	$0.73^{90,*}$	0.11	0.53
Domestic+garden score	$0.38^{21,*}_{-}$	0.17	0.22	0.23_{-}	0.03	$0.43^{120,*}$	$0.53^{90,*}$	0.4_{-}	$0.58^{90,*}_{-}$	$0.61^{90,*}_{-}$	0.23	0.03	0.44	$0.62^{90,*}$	0.23_{-}	$0.53^{21,*}$	$0.63^{120,*}$	$0.67^{21,1}$	0.31_{-}	$0.66^{120,*}$	0.51
Leisure walking	0.06	$0.48^{28,1}$	$0.51^{60,1}$	$0.66^{60,1}_{-}$	0.37_{-}	$0.5^{60,*}$	0.34	0.09_{-}	0.17_{-}	0.23_{-}	$0.54^{120,*}$	$0.45^{28,*}$	0.02_{-}	0.21	0.46_{-}	0.04	0.32_{-}	0.09	0.13_{-}	0.18	0.0_{-}
Leisure moderate	0.09	$0.59^{120,2}$	$0.62^{120,2}$	$0.55^{120,1}_{-}$	0.36_{-}	$0.65^{14,1}$	0.52	0.02	0.05_{-}	0.12_{-}	$0.58^{120,*}$	0.0	$0.52^{14,*}_{-}$	0.53	0.07_{-}	$0.47^{14,*}_{-}$	0.05_{-}	$0.72^{120,*}$	0.07	0.24	0.25_{-}
Leisure vigorous	0.15	0.09	0.04	$0.4^{14,*}_{-}$	$0.37^{14,*}_{-}$	0.15_{-}	0.01	0.18_{-}	$0.47^{7,*}$	0.42	$0.43^{7,*}$	0.1_{-}	$0.5^{7,*}_{-}$	0.27_{-}	$0.59^{60,*}$	$0.7^{28,*}_{-}$	0.05_{-}	0.21_{-}	$0.66^{90,*}$	$0.73^{14,*}_{-}$	0.17_{-}
Leisure score	0.06	$0.52^{14,1}$	$0.48^{120,1}$	$0.65^{90,1}_{-}$	$0.43^{7,*}_{-}$	$0.51^{14,*}$	0.4	0.04_{-}	0.03_{-}	0.03_{-}	$0.68^{120,1}$	0.04_{-}	$0.48^{120,*}_{-}$	0.3	0.17_{-}	0.16_{-}	$0.46^{7,*}_{-}$	0.04_{-}	0.06	0.03_{-}	0.22_{-}
Walking score	0.18	$0.54^{120,1}$	$0.62^{120,2}$	$0.65^{60,1}_{-}$	0.25_{-}	$0.64^{60,1}$	0.31	0.13_{-}	0.34_{-}	0.37_{-}	$0.59^{90,*}$	0.07_{-}	$0.51^{14,*}_{-}$	0.43	0.46_{-}	0.21	$0.51^{14,*}_{-}$	0.23	0.4_{-}	0.25	0.01_{-}
Moderate score	0.09	0.26	$0.39^{120,*}$	$0.41^{14,*}_{-}$	0.12	$0.71^{14,1}$	$0.79^{90,1}$	$0.58^{21,*}_{-}$	0.52_{-}	$0.56^{21,*}_{-}$	$0.62^{14,1}$	0.35_{-}	0.11	$0.68^{90,*}$	0.35_{-}	$0.66^{120,*}$	0.43	$0.75^{120,*}$	0.48_{-}	$0.63^{120,*}$	0.22
Vigorous score	$0.37^{120,*}$	$0.38^{7,*}$	$0.43^{28,*}$	$0.5^{14,1}_{-}$	$0.36^{28,*}_{-}$	0.0	0.04_{-}	$0.44^{7,*}$	$0.43^{7,*}$	0.46	$0.52^{14,*}$	0.15	$0.5^{14,*}_{-}$	0.15_{-}	$0.6^{90,*}$	$0.64^{28,*}_{-}$	$0.52^{14,*}_{-}$	0.08_{-}	$0.78^{90,*}$	$0.59^{14,*}_{-}$	0.04_{-}
Overall score	0.09	$0.39^{120,*}$	$0.48^{120,1}$	$0.41^{14,*}_{-}$	0.04	0.35	0.16	0.0_{-}	0.03	0.09_{-}	$0.55^{14,*}$	0.24_{-}	$0.58^{14,*}_{-}$	0.52	0.03_{-}	0.1_{-}	$0.49^{14,*}_{-}$	0.03	0.17_{-}	0.05_{-}	0.3_{-}

Cells: correlation strength (script), duration and significance (superscript), and direction (subscript), e.g., $0.66^{28,1}_{-}$ depicts a (negative) Spearman correlation with $r_S = -0.66$ and $p < 0.05 \div 12$ for an interval with 28 days. Significance: * for $p < 0.05$; 1 for $p < 0.05 \div 12$; 2 for $p < 0.05 \div 12 \div 7$. Colors: orange (weaker correlation) to green (stronger correlation), only for significant correlations.

5. Discussion

The correlations consistently reflected the negative relationship between the objective sedentary duration and the subjective physical activity across all intensities and domains. The sedentary duration had the strongest negative correlations with walking physical activity across all domains and scores, e.g., $r_S = -0.66^1$. The light duration had stronger correlations in the domestic and leisure domains when compared to other domains, e.g., $r_S = 0.72^1$ vs $r_S = 0.49$. Energy expenditure, distance, and steps had mostly moderate correlations in the work and leisure domains, and weaker correlations in the domestic domain, e.g., $r_S = 0.62^2$ vs $r_S = 0.49^1$. The correlations indicate that the seniors engage in physical activity while they are at home or in the garden, but their absolute sedentary time may be unrelated to such activity. Instead, sedentary duration correlated with decreased physical activity in leisure, transport, and work settings, e.g., $r_S = -0.66^1$. Furthermore, energy, distance, and steps did not appear to measure physical activity at home accurately; energy correlated more with the work domain, e.g., $r_S = 0.57^2$ vs $r_S = 0.44^*$, while steps and distance correlated more with the leisure moderate physical activity, e.g., $r_S = 0.62^2$ vs $r_S = 0.48^1$. This observed difference is consistent with the placement of IPAQ domestic and garden moderate activities in different scoring intensities.

There were stronger correlations across all objective intensities of physical activity for the domestic and garden moderate physical activity, also reflected in the score for moderate physical activity, as compared to other domains, e.g., $r_S = 0.69^*$ vs $r_S = 0.5^*$. This observation indicates that seniors perceive most of the moderate activity to take place around their homes. Objective light duration correlated more with domestic moderate activity, e.g., $r_S = 0.72^1$, while objective fair+vigorous duration correlates more with garden moderate, e.g., $r_S = -0.69^*$. However, the range of objective intensities, e.g., negative correlations, indicates high variability in seniors' descriptions of domestic and garden activities at moderate intensity.

The longitudinal analysis of relative intensities of physical activity leads to stronger correlations between objective and subjective physical activity. In the fair-vigorous intensity spectrum, correlations of absolute intensities are short-term negative, e.g., $r_S = -0.69^*$. However, correlations of relative intensities indicate positive and generally stronger correlations, e.g., $r_S = 0.75^*$. In the PAC relative amount, there are stronger correlations than in the absolute amount for sedentary, light, and fair durations, e.g., $r_S = 0.88^1$ vs $r_S = 0.72^1$. Objectively monitoring seniors longitudinally (up to 120 days) increases the strength of the PRO and TechRO relationships, despite the IPAQ recall period of 7 days.

Measuring sleep in PASC relative amount further strengthened the relations overall from the PAC relative amount in the sedentary to moderate spectrum across all domains, e.g., $r_S = 0.75^*$ vs $r_S = 0.68^*$. Objectively monitoring sleep, in conjunction with physical activity, increased the strength of the physical activity correlations.

Within a small sample size, we report an initial calibration between the definitions of physical activity intensities in TechRO and PRO. In the absolute amount, cumulative light+fair duration and light duration correspond to the moderate physical activity, fair duration corresponds to the vigorous physical activity, and the cumulative fair+vigorous corresponds to the vigorous physical activity. In the relative amounts, the light ratio corresponds to the moderate physical activity, and the fair ratio corresponds to the vigorous physical activity.

Several limitations characterize the study. A first limitation is the presence of multiple answers per individual, but with high variability, for which we only included one answer per wave. A second limitation is a significant decrease in alignments from the original 53 answers; we allowed for a leeway proportional to the interval duration to allow PRO and TechRO alignments that are both (1) short-term, but strict, and (2) longitudinal, but permissive. The study highlights the challenge of retaining individuals (shared by many health studies) that can provide physical activity outcomes through both questionnaire and wearable. A third limitation refers to the simplicity of the chosen variables and the analysis method (correlations with partial adjustment), driven by the reduced sample size.

We expect to employ more advanced techniques and obtain more results within statistical significance as we increase the sample size in further studies aimed at calibrating PROs and TechROs for health outcomes and longitudinal behaviours such as physical activity and sleep in seniors.

Conclusion

We quantified the relationships between physical activity durations reported by the IPAQ questionnaire and the Fitbit wearable in a sample of seniors. Several methodological approaches yielded increasingly stronger relationships across all IPAQ domains and Fitbit intensities, facilitating the calibration of physical activity PROs and TechROs. First, monitoring physical activity longitudinally (beyond the questionnaire recall period). Second, deriving quantifications of physical activity relative to all behaviours throughout the day (compositional). Third, including sleep even in studies targeting physical activity. Our results can inform the design of observational studies that monitor and assess daily life behaviours continuously and longitudinally, and personalized, focused, and effective interventions for senior individuals' targeting physical activity to reduce the risk of chronic disease and improve health and Quality of Life.

Acknowledgements

We thank the projects H2020 WellCO (769765), AAL CoME (14-7), and AAL GUARDIAN (6-120-CP).

References

[1] Aitchison, J., 1982. The statistical analysis of compositional data. J. of the Royal Statistical Society: Series B (Methodological) 44, 139–160.
[2] Ambient Assisted Living, 2020. Caregiver and me. URL: http://come-aal.eu.
[3] Anderson, G., 2011. Responding to the growing cost and prevalence of people with multiple chronic conditions. Johns Hopkins University .
[4] Anderson, K.N., Catt, M., Collerton, J., Davies, K., von Zglinicki, T., Kirkwood, T.B., Jagger, C., 2014. Assessment of sleep and circadian rhythm disorders in the very old: the newcastle 85+ cohort study. Age and Ageing 43, 57–63.
[5] Booth, F.W., Roberts, C.K., Laye, M.J., 2011. Lack of exercise is a major cause of chronic diseases. Comprehensive Physiology 2, 1143–1211.
[6] Brewer, W., Swanson, B.T., Ortiz, A., 2017. Validity of fitbit's active minutes as compared with a research-grade accelerometer and self-reported measures. BMJ Open Sport & Exercise Medicine 3, e000254.
[7] Brooke, J., et al., 1996. Sus-a quick and dirty usability scale. Usability Evaluation in Industry 189, 4–7.
[8] Cooper, C., Gross, A., Brinkman, C., Pope, R., Allen, K., Hastings, S., Bogen, B.E., Goode, A.P., 2018. The impact of wearable motion sensing technology on physical activity in older adults. Experimental Gerontology 112, 9–19.
[9] Ferguson, T., Rowlands, A.V., Olds, T., Maher, C., 2015. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. International Journal of Behavioral Nutrition and Physical Activity 12, 42.
[10] Fitbit, Inc., 2020. Fitbit official website. URL: https://www.fitbit.com.
[11] Garriguet, D., et al., 2015. Comparison of physical activity adult questionnaire results with accelerometer data. Health Reports 26, 11.
[12] General, S., 1979. The surgeon general's report on health promotion and disease prevention. DHEW (PHS) Publication .
[13] Gomersall, S.R., et al., 2016. Estimating physical activity and sedentary behavior in a free-living context: a pragmatic comparison of consumer-based activity trackers and actigraph accelerometry. Journal of Medical Internet Research 18, e239.
[14] Hallal, P.C., Victora, C.G., 2004. Reliability and validity of the international physical activity questionnaire (ipaq). Medicine & Science in Sports & Exercise 36, 556.
[15] Inkin, 2020. Wearables: the comprehensive list of wearables on the market. URL: https://www.inkin.com/wearables/.
[16] International Data Corporation, 2020. IDC forecasts sustained double-digit growth for wearable devices led by steady adoption of smartwatches. URL: https://www.idc.com/getdoc.jsp?containerId=prUS44553518.
[17] Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S.G., Huntley, J., Ames, D., Ballard, C., Banerjee, S., Burns, A., Cohen-Mansfield, J., et al., 2017. Dementia prevention, intervention, and care. The Lancet 390, 2673–2734.
[18] Makai, A., Füge, K., Breitenbach, Z., Figler, M., 2016. Physical activity levels of adults in two counties of hungary in view of sociodemographic inequalities. Health Problems of Civilization 10, 38–46.
[19] Mayo, N.E., Figueiredo, S., Ahmed, S., Bartlett, S.J., 2017. Montreal accord on patient-reported outcomes (pros) use series–paper 2: terminology proposed to measure what matters in health. Journal of Clinical Epidemiology 89, 119–124.
[20] McMahon, S.K., Lewis, B., Oakes, M., Guan, W., Wyman, J.F., Rothman, A.J., 2016. Older adults' experiences using a commercially available monitor to self-track their physical activity. JMIR mHealth and uHealth 4, e35.
[21] Naghavi, M., Abajobir, A.A., Abbafati, C., et al., 2017. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the global burden of disease study 2016. The Lancet 390, 1151–1210.
[22] National Cancer Institute, 2020. Validation of Physical Activity. URL: https://epi.grants.cancer.gov/paq/validation.html.
[23] Paul, S.S., Tiedemann, A., Hassett, L.M., Ramsay, E., Kirkham, C., Chagpar, S., Sherrington, C., 2015. Validity of the fitbit activity tracker for measuring steps in community-dwelling older adults. BMJ Open Sport & Exercise Medicine 1.
[24] van Poppel, M.N., et al., 2010. Physical activity questionnaires for adults. Sports Medicine 40, 565–600.
[25] Roman-Viñas, B., Serra-Majem, L., Hagströmer, M., Ribas-Barba, L., Sjöström, M., Segura-Cardona, R., 2010. International physical activity questionnaire: reliability and validity in a spanish population. European Journal of Sport Science 10, 297–304.
[26] Rosenberger, M.E., Fulton, J.E., Buman, M.P., Troiano, R.P., Grandner, M.A., Buchner, D.M., Haskell, W.L., 2019. The 24-hour activity cycle: a new paradigm for physical activity. Medicine & Science in Sports & Exercise 51, 454.
[27] Silsbury, Z., Goldsmith, R., Rushton, A., 2015. Systematic review of the measurement properties of self-report physical activity questionnaires in healthy adult populations. BMJ Open 5, e008430.
[28] Sims, J., Smith, F., Duffy, A., Hilton, S., 1999. 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. Family Practice 16, 152–157.
[29] Spearman, C., 1961. The proof and measurement of association between two things. .
[30] Sylvia, L.G., et al., 2014. Practical guide to measuring physical activity. Journal of the Academy of Nutrition and Dietetics 114, 199–208.
[31] Tedesco, S., Sica, M., Ancillao, A., Timmons, S., Barton, J., O'Flynn, B., 2019. Validity evaluation of the fitbit charge2 and the garmin vivosmart hr+ in free-living environments in an older adult cohort. JMIR mHealth and uHealth 7, e13084.
[32] US FDA, 2020. Digital Health Software Precertification (Pre-Cert) Program. URL: https://www.fda.gov/medical-devices/digital-health/digital-health-software-precertification-pre-cert-program.
[33] Van Den Berg, J.F., Van Rooij, F.J., Vos, H., Tulen, J.H., Hofman, A., Miedema, H.M., Neven, A.K., Tiemeier, H., 2008. Disagreement between subjective and actigraphic measures of sleep duration in a population-based study of elderly persons. Journal of Sleep Research 17, 295–302.
[34] Wanner, M., Probst-Hensch, N., Kriemler, S., Meier, F., Autenrieth, C., Martin, B.W., 2016. Validation of the long international physical activity questionnaire: influence of age and language region. Preventive Medicine Reports 3, 250–256.
[35] WHO, 2020. Global recommendations on physical activity for health. URL: https://www.who.int/dietphysicalactivity/global-PA-recs-2010.pdf.
[36] WHO regional office for Europe, 2020. Noncommunicable diseases. URL: http://www.euro.who.int/en/health-topics/noncommunicable-diseases.