Survey Research Methods (2023)
Vol. 17, No. 4, pp. 429-445
https://doi.org/10.18148/srm/2023.v17i4.7919
European Survey Research Association

The feasibility of using consumer-level activity trackers in population monitoring of physical activity

Rianne Kraakman[1] · Maaike Kompier[2] · Annemieke Luiten[2] · Vera Toepoel[2]
[1]Utrecht University
[2]Statistics Netherlands

Consumer-level activity trackers can potentially be used for population monitoring of physical activity, without suffering from the recall and social desirability bias that occur in self-report and at lower costs and effort compared to research-grade devices. However, other drawbacks may be at play when using personal activity trackers. The current study compares response, representativeness and measurement quality of consumer-level activity trackers to a research-grade accelerometer (ActivPAL) and self-report (the SQUASH survey). The study existed of a questionnaire on physical activity, possession and usage of a personal activity tracker and willingness to share data or wear a research-grade device. Subsequently, a smaller follow-up study was conducted in which respondents wore an ActivPAL and their own personal tracker to allow comparison of the different methods. The results showed a loss of respondents in each step of the process. Additionally, the representativeness of the response was influenced by both demographics and physical activity level, introducing nonresponse bias. The measurements of personal trackers had a decent agreement with the ActivPAL for number of minutes MVPA and steps, while both objective measures differed substantially from self-report on all indicators. It is concluded that consumer-level trackers are not a full replacement for self-report in large-scale studies for estimation of the exact physical activity level of a population due to representation and measurement issues, but could be used, possibly together with research-grade devices, to supplement or calibrate survey results. More research to identify and lower the barriers for respondents to participate in research with activity trackers is warranted.

Keywords: physical activity; nonresponse; consumer-level activity trackers; self-report; accelerometry

1 Introduction

In the last decades, physical inactivity and sedentary behavior have become more prevalent, with adverse consequences for health (Lee et al., 2012; López-Valenciano et al., 2020; World Health Organization, 2011). This phenomenon has led to the development of global physical activity (PA) recommendations (World Health Organization, 2020) and national guidelines (Weggemans et al., 2018 for the Dutch guidelines). Most countries use self-report questionnaires to monitor adherence of the population to these guidelines, but subjective measures of PA are prone to recall and social desirability bias (Adams et al., 2005; Sallis and Saelens, 2000), and correlate poorly with objective measures (Prince et al., 2008; Skender et al., 2016; Steene-Johannessen et al., 2016).

Using objective monitoring devices to replace or supplement self-report can reduce measurement error and augment survey estimates. This is increasingly done in health research and already implemented in some countries for national surveillance (de Wolf et al., 2023). Objective monitoring provides opportunities to study additional variables such as sleep, number of steps and sedentary behaviour, as well as patterns in activity. This may provide a more complete and comprehensive understanding of one's physical activity rhythm, which can create additional health-related insights (DiPietro et al., 2020). However, the costs of professional, research-grade monitoring devices are high and the logistics involved in distributing them to participants are important (e.g., de Wolf et al., 2023). Since an increasing amount of people possess a consumer-level (personal) activity tracker (e.g., fitness tracker, smartwatch), collecting data from those devices might be a convenient alternative. Especially if people would donate historical data, the data would give an unbiased understanding of the participant's regular physical activity, instead of behaviour they

Corresponding author: Maaike Kompier, Statistics Netherlands, Utrecht, Netherlands (Email: me.kompier@cbs.nl)

show when they know they are monitored and that might be adapted to the demanded task (e.g., McCambridge, de Bruin & Witton, 2012). To our knowledge, personal activity trackers are not used in population monitoring, although they are increasingly used in health studies (see Albert et al., 2014; Meyer and Hein, 2013; Radin et al., 2020 for examples). Possibly this is due to some major disadvantages: differences in performance between devices, limited access to raw data, and changing algorithms, unknown to researchers. For population monitoring, time series are crucial, and this kind of fluctuation is extremely undesirable. The potential gain in efficiency and reduction in cost need to be balanced carefully against these drawbacks, and additionally against the pros and cons of other methods which are further discussed in section 2. This is part of a larger debate in survey research over whether sensors can replace surveys.

Statistics Netherlands monitors PA of the Dutch population for their Health Survey currently by means of self-report data, gathered with a questionnaire. Statistics Netherlands considers supplementing or replacing the questionnaire data with objective measurement and is presently researching various alternatives towards that end, together with the National Institute for Public Health and the Environment (RIVM). The current study is a small scale feasibility study into the possibilities of using personal activity trackers, by comparing them to a research-grade accelerometer (the ActivPAL™) and self-report measures. The study focusses on the two pillars of the Total Survey Error framework: representation and measurement (Groves et al., 2004). Representation is studied by determining what part of the population has (and uses) their own activity tracker, what part of activity-tracker-owners is willing to share their data and what part is willing to wear the research-grade device. Furthermore, a comparison of the demographic characteristics and PA profile of the willing and unwilling will give insight in the representativeness of the respondents. The measurement aspects are studied by comparing three kinds of measurements (by questionnaire, research-grade device and consumer-level activity tracker) for a small subset of respondents.

2 Measuring Physical Activity

2.1 Physical activity measurement and guidelines

Physical activity has been defined as "any bodily movement produced by skeletal muscles that results in energy expenditure" (Caspersen et al., 1985, p. 126). It is usually characterized by its type, duration and intensity, and can be expressed in terms of energy expenditure or metabolic equivalents (METs). PA is often classified into light, moderate and vigorous-intensity (Beyler, 2010; Schutz et al., 2001), as shown in Fig. 1. Sedentary behaviour is defined as "waking behavior characterized by an energy expenditure ≤1.5 METs, while in a sitting, reclining or lying posture" (Tremblay et al., 2017, p. 9).

The amount of PA can also be expressed by the number of steps per day. Step count is an easy-to-understand metric. In the general population, steps per day have been used to classify people as more or less active, see Table 1 (Tudor-Locke et al., 2013).

Most countries base their national PA guidelines on the global recommendations of the WHO (2020). Following the

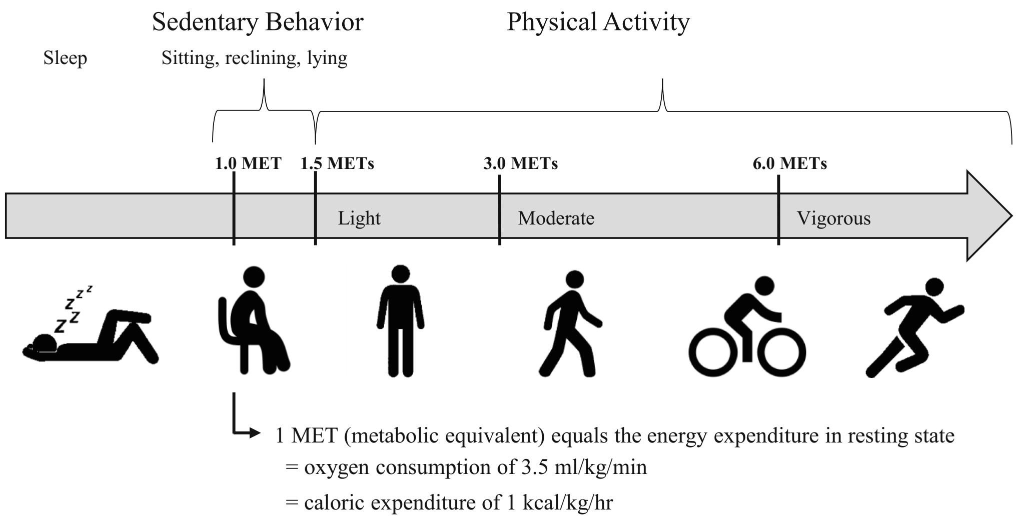

Fig. 1

Physical activity spectrum, showing energy expenditure in METs and related intensity

WHO, the most recent Dutch PA guidelines advise adults to perform moderate-to-vigorous PA (MVPA) at least 150 minutes per week, to engage in muscle- and bone-strengthening activities twice per week, and to avoid long periods of sedentary behavior (Gezondheidsraad, 2017). National (self-report) data show that only 53% of the Dutch population meets these guidelines (Statistics Netherlands, 2021). Globally, more than a quarter of the adult population are insufficiently active, with levels of inactivity being twice as high in high-income countries compared to low-income countries (Guthold et al., 2018). Statistics Netherlands uses the Short Questionnaire to Assess Health-enhancing PA (SQUASH), developed by the RIVM for national surveillance. The SQUASH is reasonably valid compared to other PA questionnaires, but lacks strong correlation with accelerometer measurement in the Dutch population (Kwak et al., 2007; Nicolaou et al., 2016; Wendel-Vos et al., 2003).

2.2 Research-grade monitoring devices

Over the last decades, several monitoring devices for research purposes have been developed and the technology is advancing quickly (Nweke et al., 2018; Troiano et al., 2014). These devices are often small and can be worn on the body, typically around the wrist, the waist, on the ankle, or on the upper leg. They track bodily movements (e.g., motion, acceleration, speed, postural allocation) and/or physiological reactions (e.g., heart rate, temperature), which can be used to obtain energy expenditure (Ainsworth et al., 2015; Beyler, 2010; Doherty, 2009). Accelerometry-based monitors, often enhanced with other sensors such as pedometers and heart rate monitors, are most commonly used. For more on the technology, characteristics and overviews of commonly used devices in research (e.g., Actigraph, Actical, RT3), we refer to Chen et al., (2012) and Reilly et al., (2008).

2.3 Personal activity trackers

Personal activity trackers (e.g., fitness trackers, smartwatches) have become widely available. Pew Research (2020) estimates that 21% of Americans used a smartwatch or fitness tracker to monitor their activity in 2019. Kantar (2021) shows that the sale of smartwatches in the US, the UK and Germany increased with 20% (Germany) to 25% (US) in one years' time. These devices allow people to track their own PA and provide immediate feedback to the user. These trackers could provide a cheap and unobtrusive alternative for PA measurement (Evenson et al., 2015; Lee et al., 2014a; 2014b; Loh et al., 2017; Nweke et al., 2018). To date, numerous personal activity trackers, such as the Fitbit, Garmin Vivofit and Apple Watch, have undergone validation studies in laboratory and free-living settings (e.g. Case et al., 2015; Ferguson et al., 2015; Guo et al., 2013; Lee et al., 2014b; Woodman et al., 2017). These studies show that the accuracy of commercial devices in estimating PA is mixed, and depend strongly on the device and the activity indicator (e.g., step count, energy expenditure, sleep) that is used (Evenson et al., 2015; Ferguson et al., 2015; Lee et al., 2014a). Step count seems to be estimated more accurately than energy expenditure (e.g., Evenson et al., 2015; Ferguson et al., 2015). A study employing indirect calorimetry (an often used 'golden standard' laboratory instrument) found that consumer monitors can provide similar, and sometimes even better, accuracy compared to an established research-grade device (Lee et al., 2014b). As a result researchers have started to use these devices in health studies (e.g., Albert et al., 2014; Meyer and Hein, 2013, Radin et al., 2020).

Table 1

Step-based classification of physical activity

Cut-off point	Classification
<5000 steps per day	Sedentary
5000-7500 steps per day	Low active
7500-10000 steps per day	Somewhat active
10000-12500 steps per day	Active
≥12500 steps per day	Very active

Even though many people may possess an activity tracker, they also would need to share the data with the researcher, or, in this case, the National Statistics Office. Hyde et al., (2020) describe that between 40 and 76% of American owners of activity trackers is willing to share the data of their trackers. Willingness varies with the (potential) receiver of the data: it is least with public health agencies, and most with respondents' health-care provider. To what extent (Dutch) people are willing to share their PA data for statistical purposes is one of the questions in the present research.

2.4 Challenges of (objective) measurement: nonresponse and bias

Although objective measurement of PA may decrease measurement errors, it could increase problems related to representation. Nonresponse is a growing concern for survey researchers since response rates in national surveillance studies show a decreasing trend (Luiten et al., 2020; Williams and Brick, 2017). Objective measurement of PA is only feasible when response rates are acceptable and nonresponse bias is minimal. Ideally, a substantial and representative part of the population should participate in PA monitoring.

A prior study among members of a Dutch research panel found that, while 77% of invited members completed the invitation questionnaire, only 57% of the respondents was willing to participate with accelerometers. Of those subsequently invited to wear the device, only 90% actually adhered (Scherpenzeel, 2017). Another study in the Netherlands, among an independent cross-section of the population, showed even lower response to monitoring with a research-grade accelerometer, with a response rate of 10%. It was also found that the participants were healthier and more active than the average Dutch person (Loyen et al., 2020). In a Norwegian national representative sample, response to an invitation to wear an accelerometer was related to sex, age, level of education, and ethnic background with higher participation rates for women, people with higher education, middle aged people and people with a Western European background (Hansen et al., 2019). Additionally, in a Canadian study it was found that people willing to participate were older, more often women and also less obese (Colley et al., 2011). In contrast, in a study in the United Kingdom no demographic, health and lifestyle differences were found between people declining to wear an accelerometer and people with valid data (Roth and Mindell, 2013).

Ownership of commercial activity trackers is likely to be biased too: a Dutch study in 2018 showed that 31% of the population used at least one health internet-of-things (IoT) device, with activity trackers and heart rate monitors most commonly used. The use of health-related devices was highest among young, high income and highly educated people (van Deursen et al., 2018).

Overall, a lot is still unknown about within- and between-respondent differences in willingness to share sensor data, and research is still accumulating (Revilla et al., 2019; Struminskaya et al., 2020; 2021). The current study contributes to this on-going effort by examining: 1) willingness to copy or donate data from personal activity trackers,

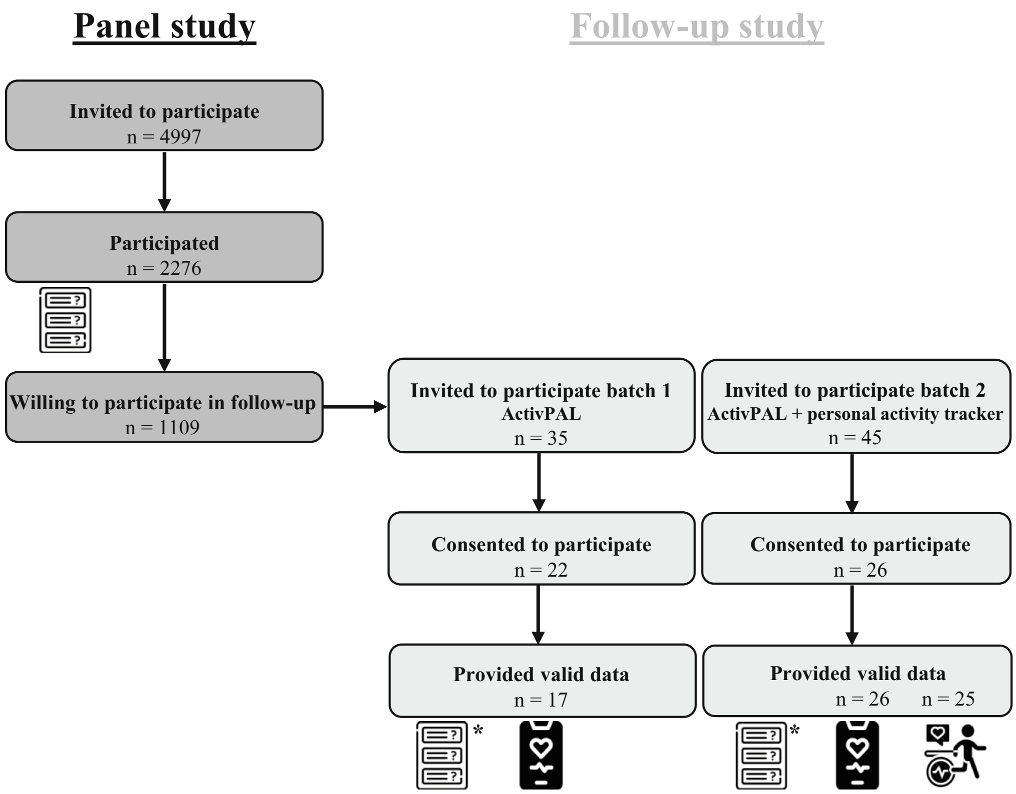

Fig. 2

Overview of respondents and available data in the panel and follow-up study
** Only self-report data of participants who completed the SQUASH was used in the follow-up study, meaning that the number of respondents for the analyses with self-report is lower.*
Note: valid ActivPAL data: at least four days with 24 hours of recording time, and not more than 4 hours of continuous stillness per day. Valid Personal activity tracker data: data for at least one of the PA indicators on at least four measurement days.

2) willingness to wear a research-grade device, 3) predictors of willingness that could introduce nonresponse bias, and 4) conditions that could change unwillingness to participation. Moreover, 5) the study adds to the literature by investigating how personal activity trackers compare to the ActivPAL and subjective SQUASH self-report in measuring PA.

3 Data and methods

3.1 Design

Data was gathered using a panel study and a subsequent, small follow-up study (see Fig. 2), both conducted by I&O Research on behalf of Statistics Netherlands and in collaboration with Utrecht University. In the panel study, data were collected with a survey in the Dutch, probability-based I&O Research Panel in June 2020. A total of 4997 people of age 18 and older were invited by email to participate in a survey on physical activity in exchange for I&O credits. The PA questionnaire consisted of a questionnaire to assess the level of PA, questions related to the possession and usage of personal activity trackers and willingness to participate in follow-up research. In total, 2276 respondents participated in the survey. In the follow-up study, 80 of the willing respondents were asked to wear an ActivPAL for a week in June or July 2020. A selection of these willing respondents (45) also owned a personal activity tracker and were asked to share that data as well. Only 43 participants provided valid data for the ActivPAL, of which 25 also shared the data from their personal tracker.

Participants gave informed consent and the data were pseudonomised to guarantee anonymity and confidentiality. The research was approved by the Ethical Review Board of Utrecht University's Faculty of Social and Behavioural Sciences.

3.2 PA Questionnaire

All respondents were randomly assigned to either the SQUASH PA questionnaire or a PA questionnaire of I&O Research[1]. The other part of the questionnaire was the same for all respondents. These questions were on possession and usage of any device or application that measures PA. Those possessing a personal activity tracker were asked to copy data from the day before into the questionnaire. In addition, willingness to copy a weeks' worth of data into a questionnaire, willingness to download and share data from the device (either a months' or a weeks' data) and willingness to wear a research-grade device were asked. Additionally, some questions were included on objections against sharing data or wearing a research-grade device as well as conditions in which respondents would be willing to participate. The entire questionnaire can be found in Supplementary Materials A.

3.3 Follow-up study

In the follow-up study, 80 respondents, selected from the people who indicated to be willing to wear a research-grade device and share data from their personal activity tracker in the questionnaire, were asked to wear an ActivPAL for a week. The ActivPAL is a triaxial thigh-worn accelerometer, which was selected based on prior research comparing several physical activity monitors in a lab setting (van Hoek et al., 2022). Several studies show the ActivPAL's good performance on multiple aspects, for example estimating posture (Grant et al., 2006) and time spent in activity intensity categories (Lyden et al., 2017). The device distinguishes all behaviour into upright and non-upright. Subsequently upright activity is classified as standing or stepping (including cycling and other active behaviours), and non-upright activity is classified as primary lying (sleeping), secondary lying (such as lying on the couch watching tv), or sitting.

The respondents were divided into two batches: the first batch consisted of 35 respondents without a personal activity tracker, whereas in the second batch 45 respondents with a personal tracker were invited. The latter group was also asked to report the data of their personal device of the week they wore the ActivPAL by copying it into a questionnaire (see Supplementary Materials B). Data donation in terms of uploading the data of the personal devices was only asked hypothetically in the prior survey, but not actually performed. The samples were randomly drawn, but then slightly adjusted to ensure representative distribution of age, sex, education and PA. All selected respondents received an information letter and were asked for consent to participate. An ActivPAL and wear instructions were sent to the respondents who gave consent, and they received a reminder email two days before they had to start wearing the ActivPAL. Participants received 20 euros as compensation and feedback on their PA after sending the ActivPAL back to the researchers.

[1] The two versions allowed I&O Research to validate their own physical activity questionnaire with the SQUASH. Results of this validation study are not reported in this article.

Table 2

Operationalization of the physical activity variables

	PA Questionnaire[a]	ActivPAL	Personal activity tracker
MVPA (minutes)	MVPA	Stepping (cadence >75)	Number of active minutes
150 min. MVPA (yes/no)	MVPA dichotomized	–	–
LPA (minutes)	LPA	Standing + stepping (cadence <75)	–
Sleep (minutes)	–	Primary lying	Sleeping time
Step count	–	Total steps/cycling steps	Step count

Note: MVPA moderate-to-vigorous physical activity, *LPA* light physical activity
[a] For MVPA and LPA, only the data of respondents who filled in the SQUASH was used. The dichotomous MVPA-variable is also available for respondents who filled in the I&O questionnaire

3.4 Measures

Adherence to the PA guidelines was calculated for all respondents of the PA questionnaire. Furthermore, the number of minutes of moderate-to-vigorous physical activity (MVPA), light physical activity (LPA) and sleep were determined as well as step count for the respondents that completed the SQUASH and wore the ActivPAL and/or the personal activity trackers to compare the three measurement methods. An overview of the operationalization is provided in Table 2.

For the respondents who completed the SQUASH in the PA questionnaire, only the self-reported minutes of LPA and MVPA and adherence to the guidelines were computed. The standardized methods to compute these parameters (Wendel-Vos et al., 2020; syntax version 2020 04 01) were used with some small adjustments (see Supplementary Materials C).

The self-reported data from the I&O questionnaire was only used to determine the dichotomous variable for adherence to the MVPA guidelines. This was calculated by dichotomizing the sum of the average time of MVPA (without sports) and the average time performing muscle- or bone-strengthening activities (including sports; see Supplementary Materials C).

The ActivPAL data was processed using the PAL software. Only data of respondents with at least four valid measurement days, with 24 hours of recording time and not more than 4 hours of continuous stillness per day, was used in the analyses. This software computes the number of minutes lying, sitting, standing and stepping at a cadence < 75, stepping at a cadence > 75 for each valid day. For the number of minutes MVPA, the number of minutes stepping at a cadence > 75 was taken. LPA was the sum of standing and stepping with a cadence < 75. For step count, cycling steps were extracted from the total steps, to be consistent with step count in personal activity trackers. Lastly, sleep equalled primary lying. For a maximum of three invalid days per respondent, the mean values of the valid days were imputed. Subsequently, the number of weekly minutes for each parameter was calculated by summing the minutes per day for each parameter.

The data from the personal activity trackers was cleaned by removing erroneous values (7% of the data for active minutes and 8% for sleep). The number of active minutes was regarded as minutes MVPA following earlier studies (Ferguson et al., 2015). Step count and sleep could directly be obtained from the given responses. Subsequently, per parameter missing values for a maximum of three measurement days were imputed with the mean value on the other days. When more days were missing, that parameter was deemed invalid. For each valid parameter, the weekly minutes was calculated by summing the daily minutes for that parameter.

3.5 Statistical analysis

First, the response percentages were calculated for the response on the questionnaire and the indicated willingness to participate in several elements of PA measurements. Chi-square tests were used to conduct a non-response analysis in which bivariate relationships between the questionnaire response and age, sex and education were tested. Analyses were performed to investigate potential predictors for the possession and usage of personal activity trackers, willingness to copy or upload data, and actually copying data. For these analyses, logistic regression models were fitted with sex, age, education and adherence to PA guidelines as predictors and the response on the various willingness questions as dependent variable. For the model for willingness to wear a research-grade device the same predictors were used in addition to the type of device. Average Marginal Effects (AME) for these models[2] were obtained

[2] Generalized Linear Models had to be fitted to obtain Average Marginal Effects (AME). In the current paper only AME are reported, but odds ratios (OR) are also available.

Table 3

Demographics of respondents and nonrespondents for the PA questionnaire

Characteristic	Respondents		Nonrespondents		χ^2 test
	N = 2276	%	N = 2721	%	
Sex					χ^2 (1) = 5.90, p < 0.05
Male	1129	50	1256	46	
Female	1147	50	1465	54	
Age					χ^2 (3) = 67.13, p < 0.001
18–34	412	18	622	23	
35–49	391	17	620	23	
50–64	818	36	913	34	
65+	655	29	566	21	
Education level					χ^2 (2) = 1.32, p = 0.52
Low	540	24	649	24	
Medium	870	38	1077	40	
High	866	38	995	37	

with the R-package “mfx” (Fernihough and Henning, 2019) to examine what person characteristics were the strongest predictors of response (or possession, usage, etc). To control for multiple testing, a Bonferroni-corrected significance level was applied for all analyses.

To compare the output of the personal activity trackers to the self-report and the ActivPAL, total minutes of MVPA, LPA, and sleep and step count were compared using scatterplots, comparisons of the mean for the different methods and calculation of the correlation. Due to the small sample, the Pearson’s r or Spearman’s ρ (when the scores were not normally distributed) should be interpreted with extreme care.

All analyses were performed in R version 4.0.3.

4 Results

4.1 Non-response analysis PA questionnaire.

The response rate on the PA questionnaire in the panel study was 46%. Table 3 provides the composition of the sample on demographic characteristics. Chi-square tests comparing the groups of respondents and nonrespondents on sex, age and education show that respondents differ from nonrespondents on sex and age, but not on education. Males were more likely to respond than females, and people from the oldest two age groups (50–64 and 65+) had a higher response rate than people from the youngest two age groups (18–34 and 35–49).

4.2 Willingness in PA questionnaire

Table 4 shows the response and willingness to participate in the several elements of PA measurements. Of the 2276 respondents, 49% would be willing to wear a research-grade device (ActivPAL) for a week when it is given to them on loan.

About half (49%) of the 2276 respondents were in possession of a personal activity tracker, of which 492 (44% of 1119) possessed a smartwatch or activity tracker and the remaining 627 (56% of 1119) possessed (only) a phone that can track PA. Of all respondents possessing a personal activity tracker, 878 (78%) used it daily or sometimes. Of all those respondents using a personal activity tracker, 469

Table 4

Response and hypothetical future response in PA measurement

Response step	N	%
Invited to participate	4997	
Responded to questionnaire	2276	100
≥ 150 minutes MVPA per week (self-report)	2000	88
< 150 minutes MVPA per week (self-report)	276	12
Willing to wear ActivPAL	1109	49
Possession of tracking device[a]	1119	49
Usage of tracking device[a]	878	39
Daily/sometimes, activity tracker or smartwatch	411	18
Daily/sometimes, phone	467	21
Copied at least one indicator	754	33
Willing to copy data (a week)	469	21
Willing to upload data	332	15

[a] A smartwatch, activity (fitness) tracker or phone that measures PA

Table 5

PA indicators that respondents using a tracking device[a] could and want to report

Indicator	Uses tracker[a] that is able to measure this indicator		Copied yesterday's data		Did not know how to copy		Did not want to copy	
	n	%	n	%	n	%	n	%
Number of steps	732	83	637	87	29	4	66	9
Distance (traveled)	724	83	565	78	89	12	70	10
Active minutes	503	57	350	70	102	20	51	10
Calories burned	463	53	326	70	73	16	64	14
Heart rate	353	40	236	67	75	21	42	12
Sleep duration	244	28	209	86	16	7	19	8
Speed (current/mean)	278	32	128	46	119	43	31	11
Stairs climbed	229	26	176	77	31	14	22	10

[a] A smartwatch, activity (fitness) tracker or phone that measures PA

(53%) were willing to copy a weeks' data, 332 (38%) were willing to upload tracker data, and 754 (86%) copied at least one PA indicator from the day before. Table 5 gives an overview of the PA indicators that were asked for and how many respondents copied, did not know how to copy and did not want to copy these data in the survey.

4.3 Likelihood to possess, use and share data of personal tracker measures

Multivariate logistic regression models were used to investigate which respondents possess and use a personal activity tracker, would be willing to copy or upload their data, and actually copied a PA indicator from the day before. The results are presented in Table 6. The models are conditional on the previous step, meaning that the benchmark of every model is the amount of respondents that 'passed' the previous step (e.g., possess a tracker for the model of using a tracker).

Model 1 shows that age, education level and PA were significantly related to personal activity tracker possession among respondents. The probability of possession was lower for people aged 50–64 and 65+ (9% and 19%, respectively) compared to people aged 35–49. Moreover, higher educated people were more likely (11%) to possess a tracking device compared to medium level educated people. Physically active people, who meet the requirement of 150 minutes of MVPA per week (based on their self-report), were 15% more likely to possess an activity tracker than people not meeting this PA requirement. Model 2 shows that of those respondents possessing an activity tracker, physically active people were also more likely (13%) to actually use their tracking device.

Among respondents using a personal activity tracker, sex, age, education and PA level were not significantly related to willingness to copy data and actually copying data from the day before (Model 3 and 5). However, willingness to upload a data file (Model 4) was significantly related to sex and age: the probability of being willing to upload a data file from a personal activity tracker was lower for people aged 50–64 and 65+ (16% and 14%, respectively) compared to people aged 35–49. Furthermore, the probability of willingness to upload a data file was reduced by 14% for females. The model fit is low for all 5 models, with Model 1 having the best fit with a Nagelkerke R^2 of 0.09.

4.4 Likelihood to wear a research-grade device

Table 7 presents the results of the multivariate logistic regression model investigating which respondents were willing to wear a research-grade device for a week. Sex, age, PA and personal activity tracker usage were all significantly related to willingness to wear a research-grade device. Females were 6% more willing to wear a research-grade device than males. People aged 50-64 and 65+ were less willing (9% and 18%, respectively) compared to people aged 35-49. Physically active people were 11% more likely to be willing to wear a research-grade device. Lastly, for people using a personal smartwatch or tracker, the probability to be willing was 18% higher compared to people not using a tracking device.

Table 6

Average Marginal Effects[a] for the likelihood to possess, use and share data of a personal activity tracker measure of PA

	Model 1		Model 2		Model 3		Model 4		Model 5	
	Possession of tracking device		Usage of tracking device (conditional on possession)		Willingness to copy data from tracking device (conditional on usage)		Willingness to upload data file from tracking device (conditional on usage)		Actually copied at least one indicator (conditional on usage)	
	AME	Std. Err.	AME	Std. Err.	AME	Std. Err.	AME	Std. Err.	AME	Std. Err.
Sex (female, ref.: male)	0.75	2.20	-4.75	2.49	0.90	3.47	-13.50*	3.35	1.36	2.36
Age (ref.: 35-49)										
18-34	4.04	3.68	-4.04	3.82	6.24	5.19	-0.22	4.91	2.08	3.20
50-64	-9.23*	3.16	1.89	3.42	-10.30	4.74	-15.58*	4.29	1.47	3.02
65+	-19.23*	3.24	1.26	3.86	-12.22	5.37	-14.12*	4.62	6.70	2.93
Education (ref.: medium)										
Lower educated	-6.26	2.94	-2.16	3.87	5.66	5.04	7.43	5.28	-6.52	4.24
Higher educated	11.18*	2.47	-0.42	2.75	4.23	3.79	2.59	3.72	-1.74	2.67
PA (≥ 150 min. MVPA)	14.71*	3.20	12.83*	4.91	4.47	6.41	9.97	5.70	5.23	4.88
Nagelkerke R^2	0.09		0.02		0.03		0.05		0.02	
N	2276		1119		878		878		878	

Note: The models are conditional on the previous step, meaning that the benchmark of every model is the amount of respondents that 'passed' the previous step

[a] The Average Marginal Effects are transformed into percentages (x100)

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table 7

Average Marginal Effects[a] for the likelihood to be willing to wear a research-grade device

	Willingness to wear research-grade device	
	AME	Std. Err.
Sex (female)	5.60*	2.19
Age (ref.: 35-49)		
18-34	-1.30	3.66
50-64	-9.22**	3.17
65+	-18.36***	3.27
Education (ref.: medium)		
Lower educated	-5.14	2.93
Higher educated	2.84	2.51
PA (≥ 150 min. MVPA)	11.35***	3.25
Tracking device usage (ref.: no usage)		
Tracker/smartwatch	17.60***	2.84
Only phone	4.83	2.74
Nagelkerke R^2	0.08	
N	2276	

[a] The Average Marginal Effects are transformed into percentages (x100)
* $p<0.05$, ** $p<0.01$, *** $p<0.001$

4.5 Conditions for willingness to participate in objective monitor-based measurement

Respondents not willing to share data from their own tracking device or wear a research-grade device were asked under which conditions they would change their mind. Frequently mentioned conditions were classified in nine categories. The condition most mentioned was 'burden'. One third of the unwilling respondents (n = 546) indicated they would be willing if it would be less of a burden, take less time or be less complicated. Other conditions mentioned often were 'information' (25% needed more information about the research and its purpose) and 'privacy' (24% mentioned privacy concerns or desire to control their data or which info is shared). 'Compensation' (wants financial compensation) and 'feedback' (wants feedback or advice on the PA measured) were mentioned by respectively 16% and 13% of respondents. 29% of respondents indicated that they would not share data with I&O under any circumstance. For wearing an ActivPAL 'never' was reported most often (55% of 1125 unwilling respondents), followed by 'information' (15%), 'privacy' (12%), 'feedback' (8%) and 'compensation' (8%).

4.6 Comparison of self-report, ActivPAL and personal device

The dataset for the comparison between the three measurement methods was fairly small as not all people invited for the follow-up study actually participated and provided valid data. In batch 1 (no personal activity tracker, but willing to wear the ActivPAL), 22 of 35 invitees consented to participate in the follow-up study, and 17 of those 22 provided valid data. In batch 2 (owners of personal activity tracker and willing to wear the ActivPAL), 26 of 45 invitees consented to participate and all of them provided valid ActivPAL data. Of those 26, 25 people also provided data for at least one of the indicators from their personal activity tracker. The kind of tracking devices used by participants were phone (9), Fitbit (5), Garmin (4), Apple (2), Polar (2), Samsung (1), Huawei (1), TomTom (1) and an unknown tracker (1). The number of steps was provided by 25 out of 26 participants, but only 14 people provided sleep duration and 15 provided active minutes, used to infer moderate-to-vigorous PA (MVPA).

Table 8

Means (sd) for all comparisons on physical activity indicators

Indicator	Self-report[a]		ActivPAL		Personal tracker			
	Mean	sd	Mean	sd	Mean	sd	N	Spearman's ρ/Pearson's r
MVPA min/week	926.9	865.9	453.9	225.6			21	ρ = 0.11
			580.4	248.0	698.4	257.5	15	ρ = 0.60*
	653.3	385.5			640.1	311.3	7	r = –0.01
LPA min/week	786.3	436.3	2152.8	908.2			20[b]	ρ = 0.24
Sleep min/week			3330.0	574.4	2904.6	413.8	14	r = 0.28
Steps/week			75148.6	9851.5	75478.4	33998.6	25	r = 0.83***

Note. MVPA moderate-to-vigorous PA, *LPA* light PA
[a] Only self-report of follow-up study participants who filled in the SQUASH was used
[b] 1 participant with an extreme value on self-reported LPA was removed. The self-report values for MVPA were not extreme
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

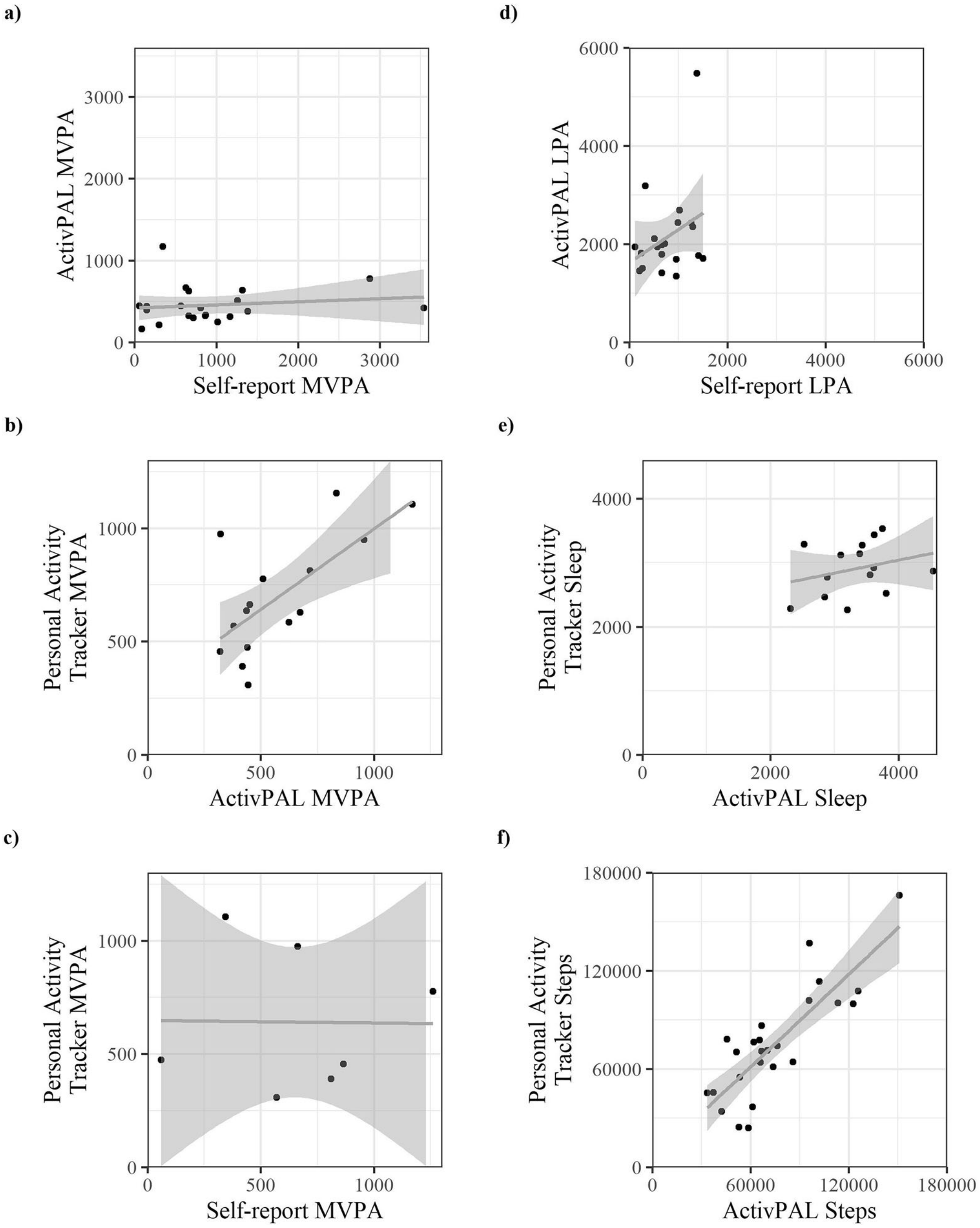

Fig. 3

*Scatterplots with trend line comparing **a** ActivPAL measured and self-reported MVPA), **b** personal activity tracker and ActivPAL measured MVPA, **c** personal activity tracker measured and self-reported MVPA, **d** personal activity tracker measured and self-reported LPA, **e** personal activity tracker and ActivPAL measured sleep, and f personal activity tracker and ActivPAL measured steps. The grey areas indicate 95% Confidence Intervals*

In Table 8, the means and standard deviation for the PA indicators are given pairwise for each comparison. As not all indicators could be determined for each method as a result of unavailable data, a different N was available for every comparison, resulting in different means and standard deviations for each comparison. For minutes MVPA, the relative agreement ($\rho = 0.60$, $p < 0.05$) between the ActivPAL and personal trackers was moderate to good, while it was poor between self-report and the ActivPAL, and self-report and personal trackers, as can be seen in the scatterplots in Fig. 3a–c. Table 8 also shows the large difference in means between the self-reported and ActivPAL measured MVPA, with the self-reported MVPA being much higher.

LPA is compared for self-report and the ActivPAL only. The agreement is weak and participants score much higher on the ActivPAL (Table 8 and Fig. 3d). Sleep and steps were only compared between the ActivPAL and the personal activity trackers, as those measures were not part of the SQUASH questionnaire (Fig. 3e–f). For sleep the agreement is weak: the correlation is not significant, and the ActivPAL scores are higher on average. In contrast, the strong correlation for steps ($r = 0.83$, $p < 0.001$) and the similar means show good agreement between the ActivPAL and the personal activity trackers.

5 Discussion

Measuring physical activity (PA) and adherence to PA guidelines is important to monitor the health of the population. The currently used self-report measures often suffer from response bias and measurement error. Hence, objective measurement with monitoring devices is increasingly applied. For monitoring the physical activity of the general population, research-grade devices are used (de Wolf et al., 2023). Additionally, a growing number of people have their own personal activity trackers, which may provide another opportunity for the monitoring of physical activity.

In this study, we investigated whether using personal tracking devices would be feasible for large scale population monitoring, in terms of (potential) response, representativeness, and measurement qualities. Response and representativeness were studied by tallying possession of personal trackers amongst respondents to a recruitment questionnaire in a probability-based web panel, gauging respondents' willingness to wear a research-grade accelerometer for a week, gauging their willingness to copy a weeks' data from their personal device into a questionnaire, and gauging willingness to download and share an overview file of their personal data. Response and representativeness of persons who indicated willingness were compared with the response and representativeness of the PA questionnaire.

In a small scale follow-up study, 80 willing respondents were invited to wear the research-grade accelerometer. About half of the invitees owned and used a personal tracker and wore that as well in the measurement week. The follow-up study allowed comparison of physical activity between the SQUASH questions posed in the PA questionnaire, the research-grade ActivPAL and the personal trackers, by means of the data that respondents copied into a follow-up questionnaire. Data donation by uploading tracker data was not part of this study.

The nonresponse analyses showed that (hypothetical) response rates for objective monitoring measures were substantially lower than the response rates of the PA questionnaire. In each step of the response process, sample persons were lost: in the PA questionnaire, in the recruitment for objective measurement, in the actual participation, and in wearing the device for the required minimal number of days. Extrapolating the final response of the small subsample to the initial recruitment sample, only 12% of the invited persons wore the research-grade device for a valid number of days, similar to one Dutch study (Loyen et al., 2020), but lower than another (Scherpenzeel, 2017). We encountered a large drop-out after respondents had indicated to be willing to participate. This may be due to (limited) information in the initial questionnaire and the (duration of) the process to receive the informed consent and ActivPAL. Response from personal trackers would be even lower, as only about half of the respondents indicated having a personal tracker in some form, and among those, again about half was willing to copy their data. Willingness to upload data would be lower still, with 38% of owners willing to upload data. This willingness is lower than that found by Hyde et al., (2020), suggesting that donating for research may not be sufficiently relevant for respondents. Interestingly, a substantial number (71%) of people who indicated that they would not copy their data for a week did copy one day in the PA questionnaire. Alternatively, 37% of people who copied the data for one day were not willing to do it again. The explanation of these findings is not clear though; did the exertion of copying the data of the previous day discourage people when the recruitment question for one weeks' data donation was posed? The findings show that several people who did not copy data indicated that they did not know how to do this. We may have underestimated the task involved. Giving clear instructions tailored to commonly used commercial activity trackers could increase response rates. This is a general lesson from the entire study: in order to be successful, nonresponse in each step of the process needs to be addressed; even when respondents indicate that they are willing to participate, a substantial number still drop out. Respondents indicated that they needed more information, both about the procedure and about privacy, and clearer instructions. Copying data or uploading data may be a needlessly complicated and burdensome way of accessing the data from the personal trackers. Some trackers allow third party access, where the only thing the participant has to do is give consent. This would potentially increase both response rates and data accuracy. However, extensive instructions are not necessarily beneficial for response either, in spite of the expressed wishes. One way of meeting the respondents' needs may be by involving interviewers in the recruitment, in contrast to the web-only recruitment in this study. Interviewers can be instrumental in addressing the respondents' needs, precluding the necessity to elucidate all information in a recruitment question, or supplementary material.

Low response rates need not be problematic if there is no relation between response and the target variable of interest. Yet, in this study, responding to the PA questionnaire, usage of an activity tracker and willingness to wear a research-grade device were all strongly related to adhering to the PA guidelines. Whereas in the general population 53% of people adhere to the guidelines for PA, and 57% move (moderately) vigorous at least 150 minutes in a general week (see Supplementary Materials D), the latter percentage was 88% among the respondents of the PA survey. In addition, the mean number of minutes MVPA in the measurement week was far above the threshold of 150 minutes, both when measured with the personal trackers and with the ActivPAL.

In addition to a strong bias related to physical activity, there were also strong correlates with demographic characteristics influencing the representativeness of the results. Older persons were more likely to participate in the PA questionnaire, but younger persons were more likely to possess a personal tracker, and more likely to donate their data. Thus, the age bias found in the questionnaire would be partly corrected by the use of personal trackers. This correction does not happen for the bias that was found for sex: men were more likely to fill in the questionnaire, and were also more willing to donate their data. On the other hand, women were more willing to wear the ActivPAL. A final demographic response bias was found for educational attainment: highly educated persons were more likely to possess a personal tracker. Demographical biases can to some extent be calibrated, but the bias as a result of differential physical activity can not, and is thus more detrimental.

Some respondents indicated that they would never participate in this kind of research, but others indicated conditions in which they could be persuaded. Their decision is more situational or condition-dependent, which corresponds to the findings of a recent study on the mechanisms of willingness (Struminskaya et al., 2021). However, whether the indicated conditions will actually result in response is yet to be determined. A large knowledge gap to be filled is how to persuade people to participate with this kind of objective measurement, or sensor measurements in general. Potentially, response rates can be increased by better procedures, better recruitment questions and better instructions. Furthermore, we foresee a large role for interviewers in gaining trust and response, and in diminishing bias. As a further step in the area of PA, a larger scale study has already been performed where we tried to implement the lessons learned from this feasibility study. The results are eagerly awaited.

The second main research question concerned the measurement quality of people's personal trackers, compared to the research-grade trackers and the self-report findings. This question was studied in a small follow-up study. Regretfully, of the 80 people that were invited, an unforeseen large number did not actually participate. For an additional few persons, data was lost as result of device malfunctioning. As a result, the basis of comparison is rather slim. Nevertheless, we found good convergent validity between the ActivPAL and personal activity trackers on the number of steps and minutes MVPA. This is in line with earlier studies showing agreement between commercial trackers and research-grade devices on these outcomes (e.g., Ferguson et al., 2015; Kooiman et al., 2015). Minutes MVPA on the monitoring devices differed substantially from self-report, with the self-report scores often being higher, which is also in line with earlier research (Nicolaou et al., 2016). Other measures, like LPA and sleep did not match between the two objective measurement methods, and can also not be compared to self-reported values as there is no direct equivalent measurement for these topics in the subjective instruments.

Overall, the convergent validity between the personal activity trackers and the ActivPAL was better than between the monitoring devices and self-report. Yet, nine different type of personal trackers were used and systematic variation in performance may exist between the different types of trackers. It would be worthwhile to study the performance of different trackers in more detail, and limit data donation to those trackers with the highest correlation with the ActivPAL or another research-grade device of choice. Some measurement difference between the ActivPAL and the personal devices will always exist, as a result of the different part of the body where the devices are worn: the leg for the ActivPAL, the wrist for most personal trackers, and perhaps other parts of the body for smartphones (waist or upper arm). Accelerometers placed on the hip, waist or wrist can underestimate movements like cycling, whereas thigh-worn monitors may be more accurate in determining different PA intensity categories than hip- and wrist-worn accelerometers (Montoye et al., 2016).

In spite of the relatively high convergent validity, the quality of the copied data from the personal trackers left to be desired. Not all trackers could measure all physical activity indicators needed, and the indicators measured might be measured differently. Furthermore, not all participants copied the data from their trackers, especially if they needed to look back a number of days. Better instruction could help, but data donation could be another venue, either through downloading and uploading data, as proposed here, or by asking respondents permission to access their data for them, by means of an API. Data donation would also preclude another phenomenon that occurred: doctoring of the values. Some of the values gave ground for suspicion, for example when only perfectly round values were provided.

This indicates that copying data from a personal tracker may not solve the problem of social desirability bias.

We started this study in an endeavor to find out whether personal activity trackers could be used as an alternative to self-report questionnaires and/or objective measurement using research-grade activity trackers. We found that response rates for data sharing are very low due to the conditional structure; people had to complete the questionnaire, possess a tracker, use a tracker and also be willing to share data. Additionally, whereas the bias in questionnaire response may be partially counteracted for age by the bias in the data sharing of the personal trackers, it accumulated for gender, education and level of physical activity. Furthermore, we found that convergent validity of personal trackers with research-grade trackers is high for some indicators.

Even though when using personal trackers to assess PA, response rates will definitely be lower and response bias will occur, they still might be a relevant alternative or supplement to self-report questionnaires, as this study shows (again) that the agreement between the personal trackers and research-grade devices is better than with the self-reported data. Although no ground truth measure of respondents' PA was done in this study, it is highly likely that the objective measures provided a more accurate estimate than the self-report. This small scale feasibility study indicates that personal tracking devices can play a role for some purposes, like estimating number of steps or minutes MVPA, or calibrating subjective PA measures. More advanced indicators of PA can only be measured by the most advanced personal trackers, further limiting the availability in the population. Research-grade accelerometers are better suited for that purpose.

Future research should investigate whether the data of both types objective measurements is of such quality that we can accept the lower response rates and additional bias of self-report measures. An important consideration for the eventual answer will depend on the goal: one would probably not use a host of different devices with varying measurement quality to make point estimates of the physical activity level of the population. On the other hand: the personal devices could probably give a fair estimate of the chance that a person adheres to (part of) the physical activity guidelines: the number of minutes of MVPA, and especially whether this exceeds the threshold of 150 minutes or not.

To gather information from the general population, it is essential to increase the response and reduce the bias that occurs when implementing objective measurement methods. Therefore, future research should focus on identifying and lowering the barriers for people to participate in PA research with these methods. For example, people who do not have a personal tracker should be loaned a research-grade device. Yet, even with these barriers addressed, there will still be a substantial part of the population that does not want to participate with objective measurements. For these people, we might have to contend ourselves with subjective data of lower quality if we want to know anything about them at all. The resulting multi-mode, multi-device design will be complicated to analyze but is inevitable if we strive for both population representation and high quality data.

Acknowledgements We thank I&O research for their generous support of this study, and two anonymous reviewers for their helpful comments.

The views expressed in this report are those of the authors and do not necessarily correspond to the policies of Statistics Netherlands.

References

Adams, S. A., Matthews, C. E., Ebbeling, C. B., Moore, C. G., Cunningham, J. E., Fulton, J., & Hebert, J. R. (2005). The effect of social desirability and social approval on self-reports of physical activity. *American Journal of Epidemiology*, *161*(4), 389–398. https://doi.org/10.1093/aje/kwi054.

Ainsworth, B. E., Cahalin, L., Buman, M., & Ross, R. (2015). The current state of physical activity assessment tools. *Progress in Cardiovascular Diseases*, *57*(4), 387–395. https://doi.org/10.1016/j.pcad.2014.10.005.

Albert, M. V., Deeny, S., McCarthy, C., Valentin, J., & Jayaraman, A. (2014). Monitoring daily function in persons with transfemoral amputations using a commercial activity monitor: a feasibility study. *PM&R*, *6*(12), 1120–1127. https://doi.org/10.1016/j.pmrj.2014.06.006.

Beyler, N. K. (2010). *Statistical methods for analyzing physical activity data*. Ames: Iowa State University. https://doi.org/10.31274/etd-180810-2230. Graduate Theses and Dissertations 11223.

Case, M. A., Burwick, H. A., Volpp, K. G., & Patel, M. S. (2015). Accuracy of Smartphone applications and wearable devices for tracking physical activity data. *JAMA*, *313*(6), 625–626. https://doi.org/10.1001/jama.2014.17841.

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise and physical fitness: definitions and distinctions for health-related research. *Public Health Reports*, *100*(2), 126–131.

Chen, K. Y., Janz, K. F., Zhu, W., & Brychta, R. J. (2012). Re-defining the roles of sensors in objective physical activity monitoring. *Medicine & Science in Sports & Exercise*, *44*(s1), 13–23. https://doi.org/10.1249/MSS.0b013e3182399bc8.

Colley, R. C., Garriguet, D., Janssen, I., Craig, C. L., Clarke, J., & Tremblay, M. S. (2011). Physical activity of

Canadian adults: accelerometer results from the 2007 to 2009 Canadian health measures survey. *Health Reports*, *22*(1), 7–14.

de Wolf, I., Elevelt, A., van Nassau, F., Toepoel, V., de Hollander, E., Kompier, M., Luiten, A., Schouten, B., Wendel-Vos, W., & van der Ploeg, H. (2023). Comparing national device-based physical activity surveillance systems: a systematic review. Submitted.

DiPietro, L., Al-Ansari, S.S., Biddle, S.J.H., et al. (2020). Advancing the global physical activity agenda: recommendations for future research by the 2020 WHO physical activity and sedentary behavior guidelines development group. *Int J Behav Nutr Phys Act*, *17*, 143–145. https://doi.org/10.1186/s12966-020-01042-2.

Doherty, S.T. (2009). Emerging methods and technologies for tracking physical activity in the built environment. In P. Bonnel, M. Lee-Gosselin, J. Zmud & J.-L. Madre (Eds.), *Transport survey methods: keeping up with a changing world* (pp. 153–190). Bingley: Emerald Group Publishing Limited. https://doi.org/10.1108/9781848558458-009.

Evenson, K.R., Goto, M.M., & Furberg, R.D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, *12*(1), 159–181. https://doi.org/10.1186/s12966-015-0314-1.

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*(1), 42–57. https://doi.org/10.1186/s12966-015-0201-9.

Fernihough, A., & Henningsen, A. (2019). ”MFX: Marginal Effects, Odds Ratios and Incidendence Rate Ratios for GLMs”. R package version 1.2-2. https://cran.r-project.org/web/packages/mfx/index.html .

Gezondheidsraad (2017). “Beweegrichtlijnen 2017: Samenvatting (No. Publication: 2017/08)”. Den Haag: Gezondheidsraad. https://www.gezondheidsraad.nl/documenten/adviezen/2017/08/22/beweegrichtlijnen-2017 .

Grant, P.M., Ryan, C.G., Tigbe, W.W., & Granat, M.H. (2006). The validation of a novel activity monitor in the measurement of posture and motion during everyday activities. *British Journal of Sports Medicine*, *40*(12), 992–997. https://doi.org/10.1136/bjsm.2006.030262.

Groves, R., Fowler, F., Couper, M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. New York: Wiley.

Guo, F., Li, Y., Kankanhalli, M.S., & Brown, M.S. (2013). An Evaluation of Wearable Activity Monitoring Devices. In *Proceedings of the 1st ACM International Workshop on Personal Data Meets Distributed Multimedia*. PDM ’13. (pp. 31–34). New York: Association for Computing Machinery. https://doi.org/10.1145/2509352.2512882.

Guthold, R., Stevens, G.A., Riley, L.M., & Bull, F.C. (2018). Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1.9 million participants. *The Lancet Global Health*, *6*(10), 1077–1086. https://doi.org/10.1016/S2214-109X(18)30357-7.

Hansen, B.H., Kolle, E., Steene-Johannessen, J., Dalene, K.E., Ekelund, U., & Anderssen, S.A. (2019). Monitoring population levels of physical activity and sedentary time in Norway across the lifespan. *Scandinavian journal of medicine & science in sports*, *29*(1), 105–112.

Hyde, E.T., Omura, J.D., Fulton, J.E., Weldy, A., & Carlson, S.A. (2020). Physical activity surveillance using wearable activity monitors: are US adults willing to share their data? *American Journal of Health Promotion*, *34*, 672–676. https://doi.org/10.1177/0890117119900587.

Kantar (2021). Smartwatch populatity continues as we enter the festive season. https://www.kantar.com/inspiration/technology/smartwatch-popularity-continues-as-we-enter-the-festive-season .

Kooiman, T.J., Dontje, M.L., Sprenger, S.R., Krijnen, W.P., van der Schans, C.P., & de Groot, M. (2015). Reliability and validity of ten consumer activity trackers. *BMC Sports Science, Medicine and Rehabilitation*, *7*(1), 1–11. https://doi.org/10.1186/s13102-015-0018-5.

Kwak, L., Kremers, S.P., Brug, J., & Baak, M.A. (2007). Measuring physical activity in field studies: comparison of a questionnaire, 24-hour recall and an accelerometer. *European Journal of Sport Science*, *7*(4), 193–201. https://doi.org/10.1080/17461390701674088.

Lee, I.M., Shiroma, E.J., Lobelo, F., Puska, P., Blair, S.N., & Katzmarzyk, P.T. (2012). Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *The Lancet*, *380*(9838), 219–229. https://doi.org/10.1016/S0140-6736(12)61031-9.

Lee, J.-M., Kim, Y.-W., & Welk, G.J. (2014a). TRACK IT: validity and utility of consumer-based physical activity monitors. *ACSM’s Health & Fitness Journal*, *18*(4), 16–21. https://doi.org/10.1249/FIT.0000000000000051.

Lee, J.-M., Kim, Y.-W., & Welk, G.J. (2014b). Validity of consumer-based physical activity monitors. *Medicine & Science in Sports & Exercise*, *46*(9), 1840–1848. https://doi.org/10.1249/MSS.0000000000000287.

Loh, M., Sarigiannis, D., Gotti, A., Karakitsios, S., Pronk, A., Kuijpers, E., & Annesi-Maesano, I. (2017). How sensors might help define the external exposome. *International Journal of Environmental Research and Public Health*, *14*(4), 434–445. https://doi.org/10.3390/ijerph14040434.

López-Valenciano, A., Mayo, X., Liguori, G., Copeland, R.J., Lamb, M., & Jimenez, A. (2020). Changes in sedentary behaviour in European Union adults between 2002 and 2017. *BMC Public Health*, *20*(1), 1–10. https://doi.org/10.1186/s12889-020-09293-1.

Loyen, A., Duijvestijn, M., & Hollander, E.L. (2020). *Monitoren van het Beweeg-, Sport-, en Zitgedrag van de Nederlandse Bevolking met Objectieve Meetmethoden*. Bilthoven: RIVM. https://doi.org/10.21945/RIVM-2020-0053.

Luiten, A., Hox, J., & Leeuw, E. (2020). Survey Nonresponse trends and fieldwork effort in the 21st century: results of an international study across countries and surveys. *Journal of Official Statistics*, *36*(3), 469–487. https://doi.org/10.2478/jos-2020-0025.

Lyden, K., Kozey-Keadle, S., Staudenmayer, J.W., & Freedson, P.S. (2017). The ActivPAL accurately classifies activity intensity categories in healthy adults. *Medicine & Science in Sports & Exercise*, *49*(5), 1022–1028. https://doi.org/10.1249/MSS.0000000000001177.

McCambridge, J., de Bruin, M., & Witton, J. (2012). The effects of demand characteristics on research participant behaviours in non-laboratory settings: a systematic review. *PloS one*, *7*(6), e39116. https://doi.org/10.1371/journal.pone.0039116.

Meyer, J., & Hein, A. (2013). Live long and prosper: potentials of low-cost consumer devices for the prevention of cardiovascular diseases. *Medicine*, *2*(2), 7. https://doi.org/10.2196/med20.2667.

Montoye, A.H.K., Pivarnik, J.M., Mudd, L.M., Biswas, S., & Pfeiffer, K.A. (2016). Validation and comparison of accelerometers worn on the hip, thigh, and wrists for measuring physical activity and sedentary behavior. *AIMS Public Health*, *3*, 298–312.

National Institute for Public Health and the Environment (RIVM) (2020). *Leefstijlmonitor [Lifestyle Monitor]*. Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu. https://www.rivm.nl/leefstijlmonitor.

Nicolaou, M., Gademan, M.G., Snijder, M.B., Engelbert, R.H., Dijkshoorn, H., Terwee, C.B., & Stronks, K. (2016). Validation of the SQUASH physical activity questionnaire in a multi-ethnic population: the HELIUS study. *PLoS ONE*, *11*(8), e161066. https://doi.org/10.1371/journal.pone.0161066.

Nweke, H.F., Teh, Y.W., Al-garadi, M.A., & Alo, U.R. (2018). Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the Art and research challenges. *Expert Systems with Applications*, *105*, 233–261. https://doi.org/10.1016/j.eswa.2018.03.056.

Pew Research Center (2020). About one-in-five Americans use a smart watch or fitness tracker. https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/.

Prince, S.A., Adamo, K.B., Hamel, M.E., Hardt, J., Gorber, S.C., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, *5*(1), 56. https://doi.org/10.1186/1479-5868-5-56.

Radin, J.M., Wineinger, N.E., Topol, E.J., & Steinhubl, S.R. (2020). Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: a population-based study. *The Lancent Digital health*, *2*(2), 85–93. https://doi.org/10.1016/S2589-7500(19)30222-5.

Reilly, J.J., Penpraze, V., Hislop, J., Davies, G., Grant, S., & Paton, J.Y. (2008). Objective measurement of physical activity and sedentary behaviour: review with new data. *Archives of Disease in Childhood*, *93*(7), 614–619. https://doi.org/10.1136/adc.2007.133272.

Revilla, M., Couper, M.P., & Ochoa, C. (2019). Willingness of Online panelists to perform additional tasks. *Methods, Data, Analyses*, *13*(2), 223–252. https://doi.org/10.12758/MDA.2018.01.

Roth, M.A., & Mindell, J.S. (2013). Who provides accelerometry data? Correlates of adherence to wearing an accelerometry motion sensor: the 2008 health survey for england. *Journal of Physical Activity and Health*, *10*(1), 70–78. https://doi.org/10.1123/jpah.10.1.70.

Sallis, J.F., & Saelens, B.E. (2000). Assessment of physical activity by self-report: status, limitations, and future directions. *Research Quarterly for Exercise and Sport*, *71*(s2), 1–14. https://doi.org/10.1080/02701367.2000.11082780.

Scherpenzeel, A. (2017). Mixing Online Panel Data Collection with Innovative Methods. In S. Eifler & F. Faulbaum (Eds.), *Methodische Probleme von Mixed-Mode-Ansätzen in Der Umfrageforschung* (pp. 27–49). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-15834-7_2.

Schutz, Y., Weinsier, R.L., & Hunter, G.R. (2001). Assessment of free-living physical activity in humans: an overview of currently available and proposed new measures. *Obesity Research*, *9*(6), 368–379. https://doi.org/10.1038/oby.2001.48.

Skender, S., Ose, J., Chang-Claude, J., Paskow, M., Brühmann, B., Siegel, E.M., Steindsorf, K., & Ulrich, C.M. (2016). Accelerometry and physical activity questionnaires - a systematic review. *BMC Public Health*, *16*(1), 515. https://doi.org/10.1186/s12889-016-3172-0.

Statistics Netherlands (CBS) (2021). Leefstijl en (Preventief) Gezondheidsonderzoek; Persoonskenmerken". Den Haag: Centraal Bureau voor de Statistiek. https://opendata.cbs.nl/#/CBS/nl/dataset/83021NED/table?dl=36FC6%2F .

Steene-Johannessen, J., Anderssen, S.A., Ploeg, H.P., Hendriksen, I.J., Donnelly, A.E., Brage, S., & Ekelund, U. (2016). Are self-report measures able to define individuals as physically active or inactive? *Medicine & Science in Sports & Exercise*, *48*(2), 235–244. https://doi.org/10.1249/MSS.0000000000000760.

Struminskaya, B., Lugtig, P., Keusch, F., & Höhne, J.K. (2020). Augmenting surveys with data from sensors and Apps: opportunities and challenges. *Social Science Computer Review*, *20*(10), 1–13. https://doi.org/10.1177/0894439320979951.

Struminskaya, B., Toepoel, V., Lugtig, P., Haan, M., Luiten, A., & Schouten, B. (2021). Understanding willingness to share Smartphone-sensor data. *Public Opinion Quarterly*. https://doi.org/10.1177/0894439320979951.

Tremblay, M.S., Aubert, S., Barnes, J.D., Saunders, T.J., Carson, V., Latimer-Cheung, A.E., Chastin, S.F.M., & Chinapaw, M.J.M. (2017). Sedentary Behavior Research Network (SBRN) – Terminology Consensus Project Process and Outcome. *International Journal of Behavioral Nutrition and Physical Activity*, *14*(1), 75. https://doi.org/10.1186/s12966-017-0525-8.

Tudor-Locke, C., Craig, C., Thyfault, J.P., & Spence, J.C. (2013). A step-defined sedentary lifestyle index: ⟨5000 steps/day. *Appl Physiol Nutr Metab*, *38*(2), 100–114. https://doi.org/10.1139/APNM-2012-0235.

Troiano, R.P., McClain, J.J., Brychta, R.J., & Chen, K.Y. (2014). Evolution of accelerometer methods for physical activity research. *British Journal of Sports Medicine*, *48*(13), 1019–1023. https://doi.org/10.1136/bjsports-2014-093546.

van Deursen, A.J., Zeeuw, A., Boer, P., Jansen, G., & Rompay, T. (2018). *Gebruik van Internet of Things in Nederland Anno 2018*. Enschede: Universiteit Twente.

van Hoek, S., Windmeijer, D., Luiten, A., Bolte, J., & Schouten, B. (2022). "Comparing activity trackers to investigate physical activity: balancing quality, costs, and usability". CBS Discussion paper. https://www.cbs.nl/nl-nl/achtergrond/2022/08/comparing-activity-trackers-to-investigate-physical-activity .

Weggemans, R.M., Backx, F.J., Borghouts, L., Chinapaw, M., Hopman, M.T., Koster, A., Kremers, S., van Loon, L.J.C., May, A., Mosterd, A., van der Ploeg, H.P., Takken, T., Visser, M., Wendel-Vos, G.W., & de Geus, E.J. (2018). The 2017 Dutch physical activity guidelines. *International Journal of Behavioral Nutrition and Physical Activity*, *15*, 58. https://doi.org/10.1186/s12966-018-0661-9.

Wendel-Vos, G.W., Berg, S., Duijvestijn, M., & Hollander, E. (2020). *Beweegrichtlijnen en Wekelijks Sporter: Van Vragenlijst tot Cijfer*. Bilthoven: RIVM. https://www.rivm.nl/publicaties/beweegrichtlijnen-en-wekelijks-sporter-van-vragenlijst-tot-cijfer .

Wendel-Vos, G.W., Schuit, A.J., Saris, W.H., & Kromhout, D. (2003). Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. *Journal of Clinical Epidemiology*, *56*(12), 1163–1169. https://doi.org/10.1016/S0895-4356(03)00220-8.

Williams, D., & Brick, M. (2017). Trends in US face-to-face household survey nonresponse and level of effort. *Journal of Survey Statistics and Methodology*, *0*, 1–26.

Woodman, J.A., Crouter, S.E., Bassett, D.R., Fitzhugh, E.C., & Boyer, W.R. (2017). Accuracy of consumer monitors for estimating energy expenditure and activity type. *Medicine & Science in Sports & Exercise*, *49*(2), 371–377. https://doi.org/10.1249/MSS.0000000000001090.

World Health Organization (2011). *Global status report on noncommunicable diseases 2010: description of the global burden of NCDs, their risk factors and determinants*. Geneva: WHO. https://apps.who.int/iris/handle/10665/44579.

World Health Organization (2020). *Global recommendations on phyisical activity and sedentary behaviour*. Geneva: WHO. https://www.who.int/publications/i/item/9789240015128 .