REVIEW
Open Access

Comparing national device-based physical activity surveillance systems: a systematic review

Inge de Wolf[1,2]*, Anne Elevelt[2], Femke van Nassau[1], Vera Toepoel[2], Ellen de Hollander[4], Maaike E. Kompier[2], Annemieke Luiten[2], Barry Schouten[3], G. C. Wanda Wendel-Vos[4] and Hidde P. van der Ploeg[1]

Abstract

Background Physical activity surveillance systems are important for public health monitoring but rely mostly on self-report measurement of physical activity. Integration of device-based measurements in such systems can improve population estimates, however this is still relatively uncommon in existing surveillance systems. This systematic review aims to create an overview of the methodology used in existing device-based national PA surveillance systems.

Methods Four literature databases (PubMed, Embase.com, SPORTDiscus and Web of Science) were searched, supplemented with backward tracking. Articles were included if they reported on population-based (inter)national surveillance systems measuring PA, sedentary time and/or adherence to PA guidelines. When available and in English, the methodological reports of the identified surveillance studies were also included for data extraction.

Results This systematic literature search followed the PRISMA guidelines and yielded 34 articles and an additional 18 methodological reports, reporting on 28 studies, which in turn reported on one or multiple waves of 15 different national and 1 international surveillance system. The included studies showed substantial variation between (waves of) systems in number of participants, response rates, population representativeness and recruitment. In contrast, the methods were similar on data reduction definitions (e.g. minimal number of valid days, non-wear time and necessary wear time for a valid day).

Conclusions The results of this review indicate that few countries use device-based PA measurement in their surveillance system. The employed methodology is diverse, which hampers comparability between countries and calls for more standardized methods as well as standardized reporting on these methods. The results from this review can help inform the integration of device-based PA measurement in (inter)national surveillance systems.

Keywords Accelerometer, Pedometer, Physical activity, Sedentary behaviour, Surveillance

*Correspondence:
Inge de Wolf
i.dewolf@amsterdamumc.nl
[1]Department of Public and Occupational Health, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Public Health Research Institute, van der Boechorststraat 7, Amsterdam 1081BT, the Netherlands
[2]Statistics Netherlands, CBS-weg 11, Heerlen 6412EX, the Netherlands
[3]Statistics Netherlands, Henri Faasdreef 312, 2492JP, The Hague, the Netherlands
[4]National Institute for Public Health and the Environment, Antonie van Leeuwenhoeklaan 9, 3721MA, Bilthoven, the Netherlands

Background

The importance of physical activity (PA) for our health is well-known [1]. In 2016, it was estimated that 27.5% of the global adult population was insufficiently physically active [2]. Estimates suggest that physical inactivity is responsible for 10% of deaths worldwide [3] and increases the risk for developing chronic diseases such as cardiovascular disease, diabetes and several forms of cancer [4]. The WHO (World Health Organization) 2020 guidelines [5] on PA and sedentary behaviour (SB) recommend adults to undertake 150–300 min of moderate-to-vigorous intensity PA (MVPA) per week and to do muscle-strengthening activities at least two days a week. Furthermore, daily SB should be limited.

To assess and report on these PA and SB levels and guidelines adherence, national population-based surveillance systems are in place to support public health policies and strategies [6]. Since 2004 the WHO recommends to systematically monitor PA in national surveillance systems [7]. Between 2001 and 2016, population-based surveillance has been done in 168 countries worldwide [2]. National surveillance systems differ in what aspects they assess. Both PA, SB and sport participation are measured in different ways and the degree of inquiry of these three varies widely [8].

There are a variety of different methods to assess PA and SB: such as questionnaires, pedometers, accelerometers and heart rate monitoring [9]. The most common method to monitor PA (and SB) in national surveillance systems is by sending out questionnaires [8, 10]. Questionnaires can offer comprehensive information on various dimensions, including domain, type, intensity, frequency, duration and context [11, 12]. However, questionnaires are also prone to social desirability, recall bias and over reporting [11–14]. In addition, the different questionnaires to assess PA and SB vary on a variety of variables: period of interest (e.g., past week, past month, regular week), categories of activity that are included (e.g., leisure, occupation, transport), input (e.g., duration, frequency, intensity) and output (e.g., hours, minutes, total energy expenditure) [15]. As a result, the percentage of participants meeting the PA recommendations can vary between 26% and 92% depending on the questionnaire and guideline that is used [16]. Geographical and cultural differences are currently also an issue with questionnaires when comparing estimates of PA and SB levels across countries [17]. The methodological variations and the resulting variation in reported outcomes make it difficult to give a reliable estimate of PA levels of populations based on questionnaires [16].

Next to that, correlations of PA questionnaires with device-based measures are generally poor to modest [12, 17]. Questionnaires often tend to overestimate PA for both amount of PA and intensity levels when compared to objective measures [18, 19]. Hence, adherence to PA guidelines derived from self-report should be interpreted cautiously [17].

Although questionnaires are the most common method to assess PA, technological advances have led to the growing popularity of device-based PA measurement methods, such as accelerometers. Wearable, accelerometer-based activity monitors have become more commonly used in relatively large national samples to measure PA and SB levels [20–22]. Accelerometer-based activity monitors can measure frequency, intensity and duration of activities, and give information on activity bouts (periods of continuous PA) [23, 24]. These device-based PA measurements have shown to be more reliable and valid compared to questionnaires for assessing PA levels over time [19, 25]. Therefore, the use of activity monitors to assess PA and SB might help to overcome limitations associated with self-reported PA. Implementation of device-based measurement methods in surveillance systems is therefore recommended [17].

When conducting PA research with device-based measurements on a population level numerous methodological decisions have to be made. This includes the type of device, number of wear days, attachment place but also sample selection, data collection and data processing aspects [22, 26]. Furthermore, algorithms are used to separate wake, sleep and non-wear times and estimate intensity levels [24, 27]. With this review, we aim to create an overview of the methodology used in existing device-based PA surveillance systems, which can help inform the development of such systems in other countries and help harmonise data collection in the future.

Methods

In order to produce such a systematic overview, we conducted a systematic literature review. In preparation of the literature search a review protocol was written, specifying inclusion and exclusion criteria for the studies and data analysis. We registered this review (CRD42022329755) in the PROSPERO database [28]. The reporting of this systematic review adheres to the reporting items of the PRISMA 2020 checklist [29].

Search strategy

To identify all relevant publications we conducted systematic searches in the bibliographic databases of PubMed, Embase.com, SPORTDiscus (Ebsco) and Web of Science (Core Collection) from inception to March 15th 2024, in collaboration with a medical information specialist at the University Library. The following terms were used (including synonyms and closely related words) as index terms or free-text words: "Accelerometry", "Actigraphy", "Actimetry", "PA", "Physical inactivity", "Population surveillance".

Duplicate papers were removed. The full search strategies for all databases can be found in the Supplementary Material (Additional file 1). In addition, after the full-text review phase, the reference lists of the included papers were also scanned for relevant publications. When references to methodological reports or -papers of the studied surveillance system were given, these were also recovered in order to gain the most complete insight in the used methods of that system.

Article selection

Only peer-reviewed, English journal papers, dissertations and reports of relevant surveillance systems were included. Laboratory, intervention, cohort and methodological studies were excluded. The papers described studies using national population-based samples of people aged > 16 years that used device-based measurements to measure PA. Papers on studies done in a particular setting (e.g., hospital, school) were excluded. To be included in the study the papers had to report on PA or SB variables, for example, minutes of MVPA, sedentary time, or adherence to PA guidelines.

All retrieved papers were imported into reference manager software (Rayyan and EndNote). Two researchers (IdW, AE) were involved in the article selection, data extraction and quality assessment. For title and abstract selection the two researcher first checked agreement by screening a random sample of 1% of the results together. This was followed by a random sample of 200 that was screened by both researchers, but independently. The remaining results were divided and screened independently. In case of uncertainty, the paper was included for full text review. For the full-text selection and quality assessment both researchers screened all full text papers independently. The few differences in judgement were resolved through a consensus meeting between the two reviewers.

Data extraction

A standardized, pre-piloted form was used to extract data from the included studies, this was done by one researcher (IdW). In case of uncertainty it was discussed with another researcher. The extracted data included study characteristics, population characteristics, data collection methods, data processing, data analyses and outcome measures.

The quality of the included surveillance systems was linked to the Total Survey Error Framework (TSEF) [30]. This framework can be seen in Fig. 1. This framework consists of two dimensions in which possible sources of error can occur. Within these dimensions the framework describes a set of quality concepts. The first four sources of errors (coverage-, sampling, nonresponse- and adjustment error) are the quality concepts about the

representativeness of the study and start with the target population and end at the post survey adjustments. The other sources of error are related to measurement and consist of three quality concepts (validity, measurement- and processing error). The concepts of this framework were linked to PA surveillance systems to identify specific areas of potential improvement in reporting and methodology.

Results

Study selection

The flow chart of the search and selection process is presented in Fig. 2. After screening 8151 titles and abstracts, 98 full-text papers were assessed for eligibility. Sixty-six papers were excluded based on the exclusion criteria, leaving 32 papers for final inclusion. Most papers were excluded because they assessed PA in local cohorts. Two additional papers were found in the reference lists of the eligible papers resulting in a total of 34 included papers [31–64]. The 34 included papers described 28 studies that used device-based measurement to monitor PA and or SB on an (inter)national level, covering 15 different national surveillance systems and 1 international surveillance system. In some papers one or more methodological reports/papers on the studied surveillance system were referenced. We identified and included 18 additional methodological reports/papers [65–82]. Finally, we obtained additional information through correspondence with the authors of the included studies by email.

Study information and sample characteristics

The study information and sample characteristics of the included papers are shown the Supplementary Material (Additional file 2 Tables 1 and 2). Surveillance systems that use accelerometers were separated from the systems using pedometers. This was done because not all aspects from the data extraction are applicable to both.

Most of the surveillance systems were from European countries (*n*=8) [31–33, 38, 42–53, 56–59]and North-America (*n*=3) [34–37, 39–41, 50, 60–64]. The other systems were from Asia (*n*=3) [38, 54, 59] and South-America (*n*=2) [43–45]. Only two systems were from low- and middle-income countries (LMIC) [43–45]. The oldest surveillance system originated from 1995 (Japan) [54], while the most recent identified paper is on the Finnish second wave from 2017/19 [39, 53]. Half of the systems used device-based measurement during one wave only (*n*=7) [31, 32, 38, 42–45, 50, 57, 59]. The other half of the systems (Finland, Norway, Sweden, Canada, Denmark, Portugal, United States and Japan) [33–37, 39–41, 46–49, 51–58, 60–64] collected more than one wave of surveillance with device-based measurements. In seven systems device-based measurement was only applied in a sub-sample [31, 32, 38, 43–45, 50–52, 56, 59].
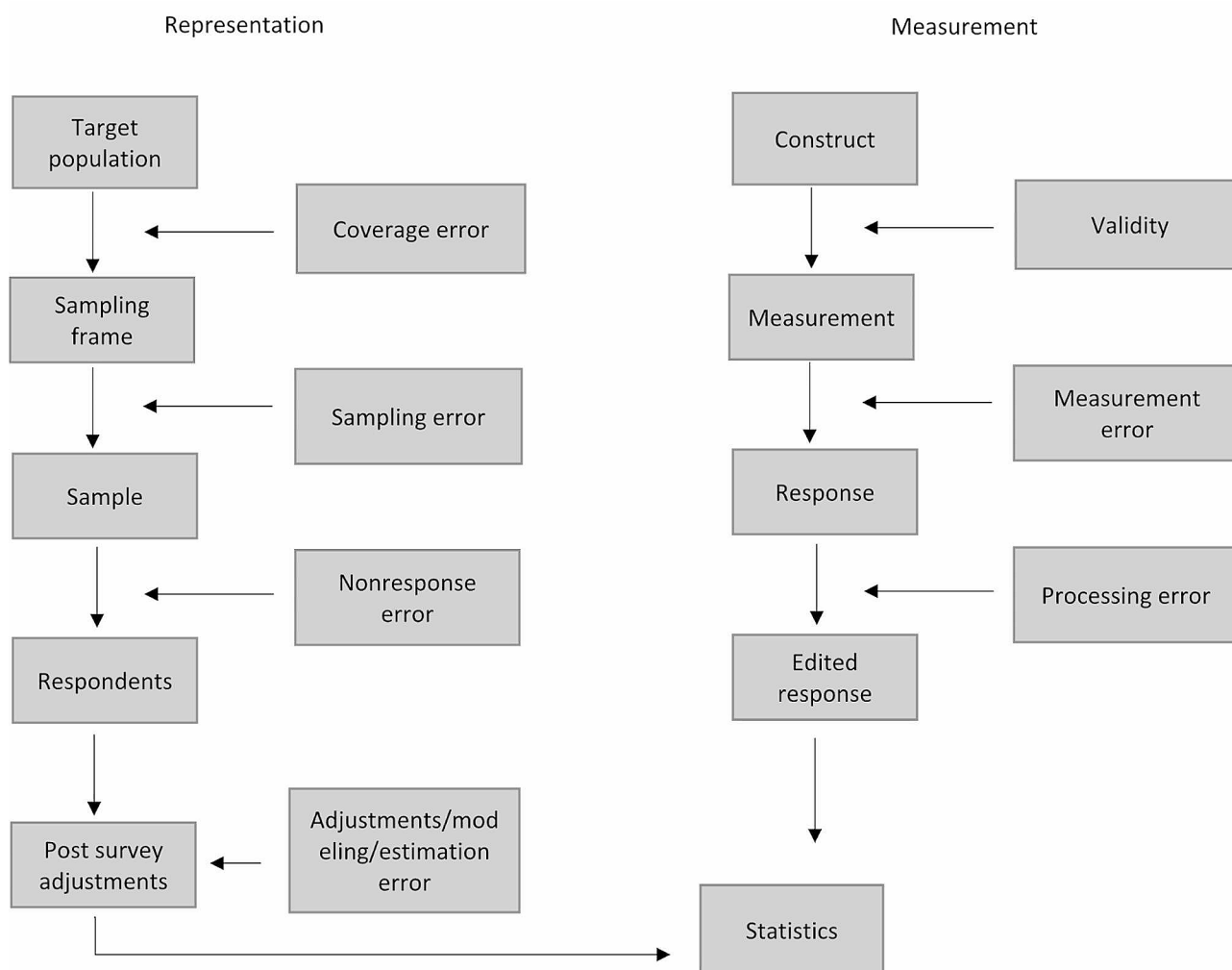
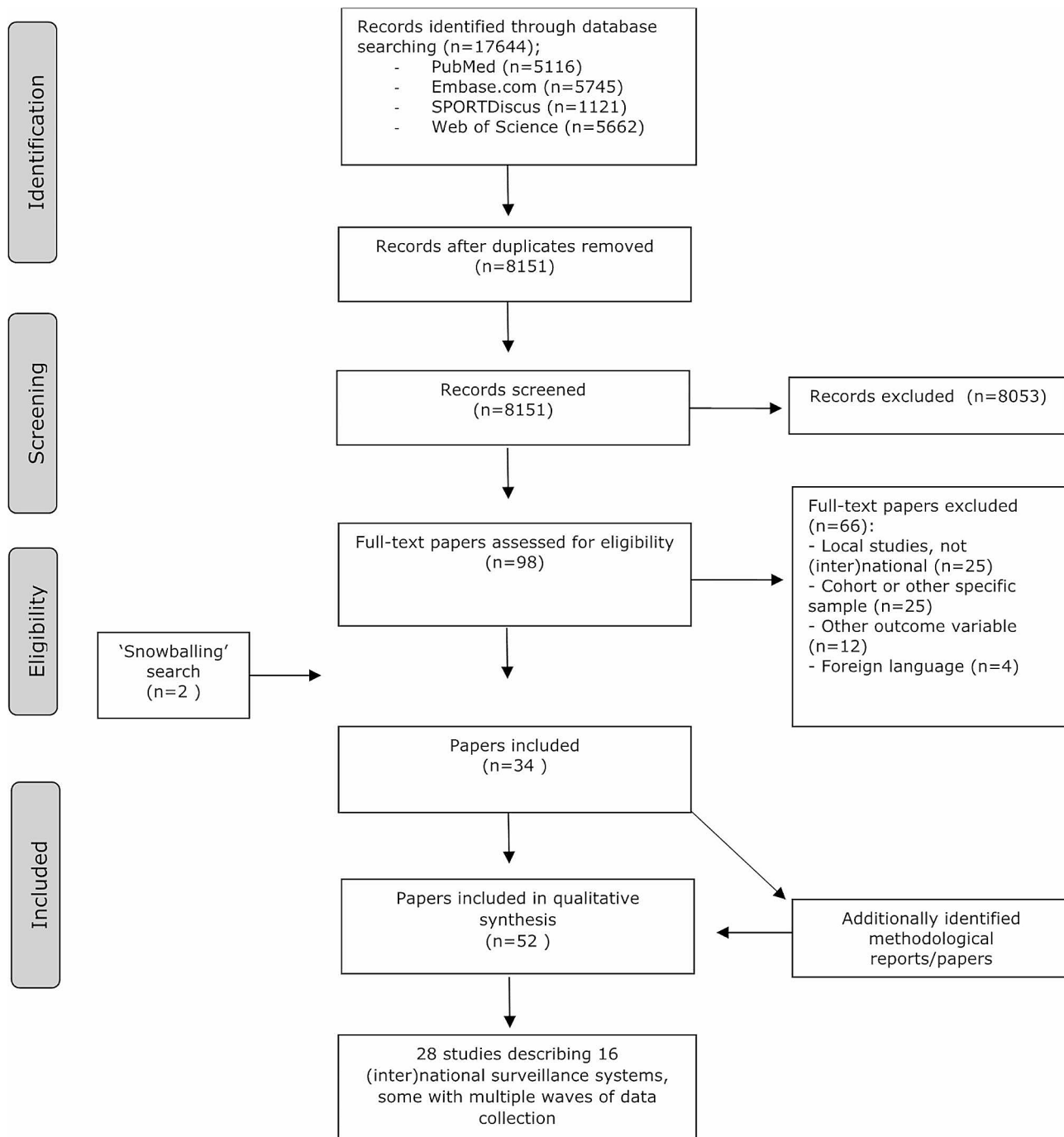**Fig. 1** Components of the Total Survey Error Framework

The number of participants ranged from 224 for the first wave in Denmark in 2007 [56] to more than 6000 participants in a single wave for Japan, Portugal and the United States [33, 54, 55, 58, 60, 62–64].

Response rates ranged from 17.6% in Finland for a completely new sample when calculating it from the total potential participants belonging to the sample [53], to 98% in Brazil for a sub-sample when participants were selected from the participants of an original study [45]. Besides, response rates were calculated in different ways (i.e., response rates were sometimes reported as people that complied with the wear time protocol, or calculated from different original samples) and some studies did not report response rates at all or lacked sufficient information to calculate them [34–37, 39–41, 43, 44, 47, 55, 59–64]. Almost all studies reported having nationally representative samples and having selected participants with a variety of probability sampling methods. Nonresponse analyses were not always performed or reported. When this was done it was most often for gender and age

[31, 32, 43, 44, 46–48, 50] and less frequent for educational level and body mass index (BMI) [42, 45, 51–53]. Differences between respondents and non-respondents were not frequently reported. The adjustments for nonresponse bias and other biases were also not always performed or reported. The first two quality concepts of the TSEF (coverage- and sampling error) are mostly well described and executed in the identified surveillance systems. More room for improvement can be found for the other two quality concepts (nonresponse error and adjustment error). Recruitment was most often done via introductory letters followed by a call or visit from an interviewer. Canada and Norway used media coverage in the first waves [41, 48] to facilitate recruitment and two studies recruited from schools, work sites and community settings [33, 55, 58].

Device information and data collection methods

Tables 3 and 4 in the Supplementary Material (Additional file 2 Tables 3 and 4) give an overview of device

Identification

Records identified through database searching (n=17644);
- PubMed (n=5116)
- Embase.com (n=5745)
- SPORTDiscus (n=1121)
- Web of Science (n=5662)

Records after duplicates removed (n=8151)

Screening

Records screened (n=8151)

Records excluded (n=8053)

Eligibility

Full-text papers assessed for eligibility (n=98)

Full-text papers excluded (n=66):
- Local studies, not (inter)national (n=25)
- Cohort or other specific sample (n=25)
- Other outcome variable (n=12)
- Foreign language (n=4)

'Snowballing' search (n=2)

Papers included (n=34)

Included

Papers included in qualitative synthesis (n=52)

Additionally identified methodological reports/papers

28 studies describing 16 (inter)national surveillance systems, some with multiple waves of data collection

Fig. 2 Flowchart of the search and selection procedure of studies

information and data collection methods of the included papers. A variety of different monitors were used in the surveillance systems. The most used devices are different models of the ActiGraph accelerometer. Finland used the Hookie in wave 1 [51, 52] and the UKKRM42 in wave 2 [53], while Barbados [50] used the Actiheart and Canada the Actical in all their five waves [37, 39–41]. All countries using a pedometer (Czech republic, Denmark and Japan) used a Yamax [54, 56, 57]. The monitors were worn on the right hip/waist/lower back, except for Barbados were participants had to wear the device on the chest [50], and for the third and fourth wave in the United States and in Luxembourg where it was worn on the wrist [35, 42, 61]. For the majority of the studies, participants were instructed to wear the monitor for all waking hours and only to take it off during water-based activities [31–34, 36–41, 43–49, 51, 52, 55, 58–60, 62–64]. In Barbados [50], the second wave in Finland [53], the third and

fourth wave in the United States [35, 61] and in Luxembourg [42] participants were asked to wear the monitor continuously, so also when sleeping. In Finland, participants were asked to wear the device on the right hip during waking hours and on the non-dominant wrist when sleeping. In most surveillance systems the distribution of the device was done in person (*n*=10) [31–33, 36–45, 51–53, 58–60, 62–64]. Only Norway [48] and Sweden [47] specified for wave 1 that they distributed the accelerometers via mail. Epoch length used for analyses was usually set at 60 s but ranged from 10 [48] to 60 s. The monitoring period ranged from one to seven days, with Japan [54] measuring during one "typical" day, Barbados [50] during four days and for all other systems during seven consecutive days. Wear-time compliance ranged from 59% in England [31, 32] to 97% in Barbados [50] and was most often around 80–90%. For some systems wear time compliance was not reported (waves 2 and 3 Canada [39, 40], wave 2 Sweden [46], Portugal [55], and Japan [54]). Regarding financial incentives, only England [31, 32] and the United States [34–36, 60–64] reported giving a £20,- gift voucher and $40 respectively. Some systems gave feedback to the participants on their own PA levels as an incentive [33, 48, 49, 51–53, 58].

Data processing and outcome measures
Information on data processing and outcome measures for the included studies are shown in Tables 5 and 6 of the Supplementary Material (Additional file 2 Tables 5 and 6). In the majority of studies four days with valid data were required, with the exception of four systems: for the second wave in Norway [49] and Japan [54] one day with valid data was sufficient, while Brazil [45] and the Latin America study [43, 44] required at least five days, including one weekend day. Regarding required wear-time, a valid day was defined by almost all studies as at least 10 h of wear-time per day. For Barbados and the second wave in Finland a wear time of 24 h/d was required [41, 51, 72], for the third and fourth wave of the United states 23 h/d was required [35, 61]. For non-wear time various definitions were used: the most common definition for non-wear time was any period of continuous zero counts for 60 or more consecutive minutes. In the second wave in Finland [53] 120 min of continuous zero counts was considered non-wear time, while for the first wave in Sweden [47] this was set at 20 min.

PA intensities were estimated based on different cut-points linking activity counts to metabolic equivalents (METs). For the ActiGraph the cut-points from Troiano [62] and Freedson (113) were most often used. Studies that reported on outliers used a variety of cleaning rules. The most common outcome measures were adherence to the PA guidelines and minutes per day at different PA intensities. Meeting the PA guidelines in most studies

meant accumulating at least 150 min of MVPA per week in bouts of 10 min or more. For England and Luxembourg the bouts of at least 10 min were not required [31, 32, 42] and Norway did not specify their used PA guidelines [48, 49]. For systems using pedometers, steps per day as a continuous variable was the most commonly reported outcome.

Discussion
In this systematic review we aimed to map the methodology used in existing national PA surveillance systems using device-based measurement. This was done to support the development of such systems in other countries and to harmonise data collection in the future. A total of 34 eligible papers were identified, reporting on fifteen national and one international surveillance system. Substantial variation between waves of systems was found in number of participants, response rates, population representativeness and recruitment. Data reduction definitions were more often comparable between systems (e.g., minimal number of valid days, non-wear time and necessary wear time for a valid day). This was not the case within the systems using pedometers. Pedometers were the forerunners of accelerometers, but are now less often used for population surveillance [83], and more as behavioral chance tools in interventions. For this reason, we focus more on the systems using accelerometers in the discussion. Because data was collected before the launch of the 2020 WHO guidelines on PA and SB, they refer to the WHO 2010 guidelines that required MVPA to be accumulated in bouts of at least 10 min. In the 2020 guidelines all MVPA counts, regardless of the duration of the bouts in which it was accumulated [5]. Only the paper using the data from Luxembourg refers to both WHO guidelines [42].

Representativeness in PA surveillance with device-based measurement
A national surveillance system aims to assess the PA levels of the whole population. Representativeness of the study sample is therefore an important aspect to consider when conducting PA surveillance with device-based measurement. Most systems identified in our review claim to have "a nationally representative sample" or "nationally representative data". However, response rates in the included studies were as low as 17.6% and regularly around 50%. In many PA intervention trials participants tend to be women, Caucasian and higher educated compared to nonparticipants [84]. The people that participated in the included studies might therefore differ from the whole original sample, resulting in potential nonresponse bias. It is difficult to determine if the participating sample is truly representative of characteristics of the whole population. Using for example census data,

it would be possible to weigh the participating sample towards the population distributions for factors such as age, gender, income and area of residence. However, representativeness of other relevant factors, such as health status, is not always possible. Another issue is the representativeness of samples with regard to PA itself. Population PA levels are often not known (which is why the surveillance system is in place) so there is no data to calibrate the surveillance system sample against. Hence, it is not possible to determine if, for example, more active people are more likely to participate in the surveillance system which would result in an overestimation of population PA levels. Often the only way to attempt to adjust for such potential confounding is by weighing the population towards the known population distribution for the factors mentioned above (e.g., age, gender, income). However, it remains to be seen if selection bias towards PA level for example can be adjusted for in this way and hence if the participating sample is truly representative of the population PA levels.

Not all identified surveillance systems performed non-response analyses and the surveillance systems that did a non-response bias analysis only looked at basic demographic variables such as gender, age, education level and sometimes body mass index. To avoid biased PA estimates in survey research it is preferable to adjust for non-response error [85]. We believe that non-response bias analysis should also be done on PA levels. This was not done in the included studies in this review. As PA data are usually unavailable for non-respondents it is hard to know if the weights and other adjustment procedures in the surveillance systems are sufficient to keep the results representative of the whole population.

Non-response bias and the related adjustment procedures are important to consider when conducting PA surveillance with devices. Claims of population representativeness of PA levels need to be substantiated and potential limitations in such representativeness should be discussed.

Measurement in PA surveillance with device-based measurement

Besides the issues around representativeness mentioned in the previous paragraph, we also want to discuss measurement error and processing error of the measurement side of the TSEF, that show room for improvement.

The placement of the device has implications for the PA outcome measures. Accelerometers cannot detect all types of movements depending on the attachment place. The most common accelerometer placement used in the systems included in this review was worn on the right hip/waist. Accelerometers placed on the hip or waist can underestimate certain movements (e.g., upper body movements, cycling) and can be uncomfortable to wear during sleep [86]. The USA (NHANES) changed the attachment location of the accelerometer from the hip to the wrist after two waves to improve compliance and collect sleep measures [87]. In the second wave in Finland, they used a 24 h approach by changing the accelerometer from the hip during the day to the wrist during the night. This two wear-site approach did not impact compliance and it was found feasible among working-aged adults [53]. Determining the best possible attachment place can be a challenge. Thigh-worn devices might be more suitable to wear 24 h a day as they do not need to change position and lower participant burden [88]. Thigh-worn monitors might therefore be more accurate in predicting time spent in different PA intensity categories and SB breaks compared to hip- and wrist-worn accelerometers [89].

The next quality aspect is the processing error which is connected to the measurement as well. The identified studies used different criteria for valid wear time, which has major implications on the outcome measures of PA. Studies have shown that the amount of wear time influences the amount of PA measured by the device during the day [90, 91]. It also influences the number of individuals having enough valid days to be included [92]. These possible processing errors can influence the survey statistics. Standardization of the data processing would help increase comparability of the results.

Strengths and limitations

This study provides an overview of all PA surveillance systems worldwide that use device-based measurements. The main strength of this study is the thorough and systematic review process. The search was performed in four databases complimented with additional search strategies (i.e., looking at the references of included papers, looking at the websites of the systems and contacting the authors). In addition, the article selection and quality assessment were conducted independently by two researchers.

Even though the search was extensive the possibility exists that papers or background reports are missed because we only looked for English results. Especially methodological reports and other background reports might have been written in other languages and therefore be missed by the search strategy.

Although some methodological reports might have been missed because of the language restrictions, a strength of this study is the inclusion of those background reports, which have provided extra information. Several of the 34 papers that were eligible for inclusion reported on the same study. We decided to combine all identified papers and reports to describe the methodology used for that particular system in the best possible way.

A limitation of this study is the lack of a quality assessment using a quality assessment tool. Currently, to our knowledge, no quality assessment tool for surveillance systems exists. To overcome this we used the TSEF [30] as an alternative to look at the quality and identify aspects with room for improvement.

The identified papers presented on data collected from 1995 to 2019. More recent (ongoing) studies might not have been found in the search as it can take time before the data becomes available. A check of the websites of the included surveillance systems revealed more recent or ongoing waves on which no publications were available yet. A new wave of CHMS (Canada, 2019) and Fin-Fit (Finland, 2022) have recently been finished. CHMS 8 (Canada) and Kan 3 (Norway) are ongoing at the time of writing this paper. These cycles have not been included in this paper since results were not yet presented at the time of writing this paper.

Another strength of this review is the extensive data extraction. This gives an overview of the aspects that the included surveillance systems have described, but also which methodological aspects have limitations. However, with limited available space in papers it might be hard to report on all aspects. Our paper might therefore not describe the quality aspects of the surveillance system but mainly the quality of reporting in the identified papers and reports.

Recommendations

The results of this review highlight the need for standardisation of measurement methods used in device-based surveillance systems worldwide. The different methodological aspects reported and the incompleteness of information on certain system aspects made it difficult to compare systems and create a complete overview. More standardized reporting would enable better comparison of the surveillance systems and create a complete overview of all important aspects to consider when implementing device-based PA monitoring.

For surveillance systems representativeness is especially important. According to the used quality framework [30] both nonresponse and adjustment error are important quality aspects to consider. Information to assess these aspects was not always reported or investigated. We advocate for more standardized reporting on these aspects. Standardized reporting on the difference between respondents and non-respondents on as many variables as possible should be the goal. To be able to do this, clear reporting on the sample, number of participants and the associated response rates are necessary. If necessary, the adjustments in the face of, for example, nonresponse error to improve the sample estimate should be reported. The framework might be a helpful tool to improve reporting on these steps in a more standardized manner.

Additionally, the framework might also be helpful to further standardize the measurement dimension [30]. Considering the validity, measurement error and processing error can result in more standardized methods. Even though there is some standardization in certain methodological aspects, there is generally insufficient standardization to compare different national device-based PA surveillance systems [25].

Furthermore, this review showed that 15 countries currently employ a device-based surveillance system which may be implemented by more countries. Using device-based measurement in PA surveillance systems can provide more accurate and comparable estimates of PA and SB levels across countries and over time [17, 19, 25]. An important issue here is the affordability as device-based data are more expensive to collect than questionnaire data. Therefore, device-based measurements are harder to implement in surveillance systems, especially in LMIC. Nonetheless, two surveillance systems in South American LMIC have already used device-based measurements [43–45]. Furthermore, the WHO has also piloted wrist-worn accelerometry in Africa (Malawi) to determine its feasibility for implementation into the global WHO STEPS surveillance system (STEPwise approach to non-communicable disease risk factor surveillance) [93]. They concluded that it is feasible to implement devices within the STEPS surveillance system in countries such as Malawi. Hence, the standardisation of methodologies across countries that STEPS facilities could work for device-based methods in addition to the more commonly used questionnaire based methods that are likely to still employed in many countries, at least in the short term.

Ideally a combination of device-based measurements and questionnaires in surveillance systems should be implemented. The differences in the dimensions studied by questionnaires and devices make these two methods complementary to each other. Combining both methods and fine tuning their complementary outcomes in a surveillance system will result in the most complete picture of national physical activity levels [25]. Device-based measurements will provide more valid data about intensity and duration while questionnaires are useful for information about context and activity types and domains [17, 94]. However, more research on how to best utilize device-based data in synergy with questionnaire data for national surveillance would be helpful.

Conclusion

In this systematic review, we mapped the methodology used in existing national PA surveillance systems using device-based measurement. Currently, only fifteen countries have monitored population level PA with

device-based measurements. Although they do not vary much in the device used, differences are found in many other methodological aspects and harmonization of methodologies would improve comparability. Representativeness of surveillance systems with regard to PA levels is an important aspect that needs attention in the future. This review and the associated recommendations can be used to incorporate device-based measurement of PA in surveillance systems of other countries.

Abbreviations

PA	Physical activity
WHO	World Health Organization
SB	Sedentary behaviour
MVPA	Moderate-to-vigorous intensity physical activity
TSEF	Total Survey Error Framework
BMI	Body mass index
LMIC	Low- and middle-income countries
MET	Metabolic equivalent of task

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12966-024-01612-8.

Supplementary Material 1

Supplementary Material 2

Declarations

References
1. WHO. Global action plan on physical activity 2018–2030: more active people for a healthier world. Geneva, Switzerland: World Health Organization; 2018.
2. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health. 2018;6(10):e1077–86.
3. Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet. 2012;380(9838):219–29.
4. WHO. Global status report on noncommunicable diseases 2014. Switzerland: World Health Organization;: Geneva; 2014. Report No.: 9241564857 Contract No.: WHO/NMH/NVI/15.1.
5. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med. 2020;54(24):1451–62.
6. Troiano RP, Stamatakis E, Bull FC. How can global physical activity surveillance adapt to evolving physical activity guidelines? Needs, challenges and future directions. Br J Sports Med. 2020;54(24):1468–73.
7. WHO. Global strategy on diet, physical activity and health. Geneva, Switzerzerland: World Health Organization; 2004.
8. Sorić M, Meh K, Rocha P, Wendel-Vos W, de Hollander E, Jurak G. An inventory of national surveillance systems assessing physical activity, sedentary behaviour and sport participation of adults in the European Union. BMC Public Health. 2021;21(1):1797.
9. Aparicio-Ugarriza R, Mielgo-Ayuso J, Benito PJ, Pedrero-Chamizo R, Ara I, González-Gross M. Physical activity assessment in the general population; instrumental methods and new technologies. Nutr Hosp. 2015;31(Suppl 3):219–26.
10. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet. 2012;380(9838):247–57.
11. Bauman A, Ainsworth BE, Bull F, Craig CL, Hagströmer M, Sallis JF, et al. Progress and pitfalls in the use of the International Physical Activity Questionnaire (IPAQ) for adult physical activity surveillance. J Phys Act Health. 2009;6(Suppl 1):S5–8.
12. Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol. 2003;56(12):1163–9.
13. Kapteyn A, Banks J, Hamer M, Smith JP, Steptoe A, van Soest A, et al. What they say and what they do: comparing physical activity across the USA, England and the Netherlands. J Epidemiol Community Health. 2018;72(6):471–6.
14. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. Br J Sports Med. 2003;37(3):197–206. discussion.
15. Sylvia LG, Bernstein EE, Hubbard JL, Keating L, Anderson EJ. Practical guide to measuring physical activity. J Acad Nutr Diet. 2014;114(2):199–208.
16. Loyen A, Van Hecke L, Verloigne M, Hendriksen I, Lakerveld J, Steene-Johannessen J, et al. Variation in population levels of physical activity in European adults according to cross-european studies: a systematic literature review within DEDIPAC. Int J Behav Nutr Phys Act. 2016;13:72.
17. Steene-Johannessen J, Anderssen SA, Van der Ploeg HP, Hendriksen IJ, Donnelly AE, Brage S, Ekelund U. Are self-report measures able to define individuals as physically active or inactive? Med Sci Sports Exerc. 2016;48(2):235.
18. Lee PH, Macfarlane DJ, Lam TH, Stewart SM. Validity of the International Physical Activity Questionnaire Short Form (IPAQ-SF): a systematic review. Int J Behav Nutr Phys Act. 2011;8:115.
19. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act. 2008;5:56.
20. Bauman A, Pedisic Z, Bragg K. Objective measurement in physical activity surveillance: Present Role and future potential. In: Shephard R, Tudor-Locke C, editors. The objective monitoring of physical activity: contributions of Accelerometry to Epidemiology, Exercise Science and Rehabilitation. Switzerland: Springer Interntional; 2016. pp. 347–67.
21. Loyen A, Clarke-Cornwell AM, Anderssen SA, Hagströmer M, Sardinha LB, Sundquist K, et al. Sedentary time and physical activity Surveillance through Accelerometer Pooling in four European countries. Sports Med. 2017;47(7):1421–35.
22. Wijndaele K, Westgate K, Stephens SK, Blair SN, Bull FC, Chastin SF, et al. Utilization and harmonization of adult Accelerometry data: Review and Expert Consensus. Med Sci Sports Exerc. 2015;47(10):2129–39.
23. Butte NF, Ekelund U, Westerterp KR. Assessing physical activity using wearable monitors: measures of physical activity. Med Sci Sports Exerc. 2012;44(1 Suppl 1):S5–12.

24. Quante M, Kaplan ER, Rueschman M, Cailler M, Buxton OM, Redline S. Practical considerations in using accelerometers to assess physical activity, sedentary behavior, and sleep. Sleep Health. 2015;1(4):275–84.

25. Skender S, Ose J, Chang-Claude J, Paskow M, Brühmann B, Siegel EM, et al. Accelerometry and physical activity questionnaires - a systematic review. BMC Public Health. 2016;16:515.

26. Trost SG, McIver KL, Pate RR. Conducting accelerometer-based activity assessments in field-based research. Med Sci Sports Exerc. 2005;37(11 Suppl):S531–43.

27. Pedišić Ž, Bauman A. Accelerometer-based measures in physical activity surveillance: current practices and issues. Br J Sports Med. 2015;49(4):219–23.

28. Centre for Reviews and Dissemination. PROSPERO database [http://www.crd. york.ac.uk/PROSPERO.

29. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.

30. Groves RM, Lyberg L. Total Survey Error: past, Present, and Future. Public Opin Quart. 2010;74(5):849–79.

31. Anokye NK, Trueman P, Green C, Pavey TG, Taylor RS. Physical activity and health related quality of life. BMC Public Health. 2012;12:624.

32. Bakrania K, Edwardson CL, Bodicoat DH, Esliger DW, Gill JM, Kazi A, et al. Associations of mutually exclusive categories of physical activity and sedentary time with markers of cardiometabolic health in English adults: a cross-sectional analysis of the Health Survey for England. BMC Public Health. 2016;16:25.

33. Baptista F, Santos DA, Silva AM, Mota J, Santos R, Vale S, et al. Prevalence of the Portuguese population attaining sufficient physical activity. Med Sci Sports Exerc. 2012;44(3):466–73.

34. Barreira TV, Harrington DM, Katzmarzyk PT. Cardiovascular health metrics and accelerometer-measured physical activity levels: National Health and Nutrition Examination Survey, 2003–2006. Mayo Clin Proc. 2014;89(1):81 – 6.

35. Belcher BR, Wolff-Hughes DL, Dooley EE, Staudenmayer J, Berrigan D, Eberhardt MS, Troiano RP. US Population-referenced percentiles for wrist-worn accelerometer-derived activity. Med Sci Sports Exerc. 2021;53(11):2455–64.

36. Camhi SM, Sisson SB, Johnson WD, Katzmarzyk PT, Tudor-Locke C. Accelerometer-determined lifestyle activities in US adults. J Phys Act Health. 2011;8(3):382–9.

37. Carson V, Wong SL, Winkler E, Healy GN, Colley RC, Tremblay MS. Patterns of sedentary time and cardiometabolic risk among Canadian adults. Prev Med. 2014;65:23–7.

38. Chu AHY, Bernard JY, Koh D, Muller-Riemenschneider F. Accelerometer Profile of Physical Activity and Sedentary Behavior in a multi-ethnic urban Asian Population. Res Q Exerc Sport. 2021;92(3):361–8.

39. Clarke J, Colley R, Janssen I, Tremblay MS. Accelerometer-measured moderate-to-vigorous physical activity of Canadian adults, 2007 to 2017. Health Rep. 2019;30(8):3–10.

40. Colley RC, Butler G, Garriguet D, Prince SA, Roberts KC. Comparison of self-reported and accelerometer-measured physical activity among Canadian youth. Health Rep. 2019;30(7):3–12.

41. Colley RC, Garriguet D, Janssen I, Craig CL, Clarke J, Tremblay MS. Physical activity of Canadian adults: accelerometer results from the 2007 to 2009 Canadian Health measures Survey. Health Rep. 2011;22(1):7–14.

42. Collings PJ, Backes A, Aguayo GA, Malisoux L. Group O-Ls. Device-measured physical activity and sedentary time in a national sample of Luxembourg residents: the ORISCAV-LUX 2 study. Int J Behav Nutr Phys Act. 2022;19(1):161.

43. Ferrari GLM, Kovalskys I, Fisberg M, Gomez G, Rigotti A, Sanabria LYC, et al. Association of moderate-to-vigorous physical activity with neck circumference in eight latin American countries. BMC Public Health. 2019;19(1):809.

44. Ferrari GLM, Kovalskys I, Fisberg M, Gómez G, Rigotti A, Sanabria LYC, et al. Comparison of self-report versus accelerometer - measured physical activity and sedentary behaviors and their association with body composition in latin American countries. PLoS ONE. 2020;15(4):e0232420.

45. Ferrari GLM, Victo ER, Kovalskys I, Mello AV, Previdelli AN, Solé D, Fisberg M. Sedentary behavior, physical activity and body composition in adults. Rev Assoc Med Bras (1992). 2020;66(3):314–20.

46. Hagströmer M, Kwak L, Oja P, Sjöström M. A 6 year longitudinal study of accelerometer-measured physical activity and sedentary time in Swedish adults. J Sci Med Sport. 2015;18(5):553–7.

47. Hagströmer M, Oja P, Sjöström M. Physical activity and inactivity in an adult population assessed by accelerometry. Med Sci Sports Exerc. 2007;39(9):1502–8.

48. Hansen BH, Kolle E, Dyrstad SM, Holme I, Anderssen SA. Accelerometer-determined physical activity in adults and older people. Med Sci Sports Exerc. 2012;44(2):266–72.

49. Hansen BH, Kolle E, Steene-Johannessen J, Dalene KE, Ekelund U, Anderssen SA. Monitoring population levels of physical activity and sedentary time in Norway across the lifespan. Scand J Med Sci Sports. 2019;29(1):105–12.

50. Howitt C, Brage S, Hambleton IR, Westgate K, Samuels TA, Rose AM, Unwin N. A cross-sectional study of physical activity and sedentary behaviours in a Caribbean population: combining objective and questionnaire data to guide future interventions. BMC Public Health. 2016;16(1):1036.

51. Hukkanen H, Husu P, Sievänen H, Tokola K, Vähä-Ypyä H, Valkeinen H, et al. Aerobic physical activity assessed with accelerometer, diary, questionnaire, and interview in a Finnish population sample. Scand J Med Sci Sports. 2018;28(10):2196–206.

52. Husu P, Suni J, Vähä-Ypyä H, Sievänen H, Tokola K, Valkeinen H, et al. Objectively measured sedentary behavior and physical activity in a sample of Finnish adults: a cross-sectional study. BMC Public Health. 2016;16(1):920.

53. Husu P, Tokola K, Vähä-Ypyä H, Sievänen H, Suni J, Heinonen OJ, et al. Physical activity, sedentary behavior, and Time in Bed among Finnish adults measured 24/7 by Triaxial Accelerometry. J Meas Phys Behav. 2021;4(2):163–73.

54. Inoue S, Ohya Y, Tudor-Locke C, Tanaka S, Yoshiike N, Shimomitsu T. Time trends for step-determined physical activity among Japanese adults. Med Sci Sports Exerc. 2011;43(10):1913–9.

55. Magalhaes JP, Hetherington-Rauth M, Rosa GB, Correia IR, Pinto GM, Ferreira JP, et al. Physical activity and sedentary behavior in the Portuguese Population: what has changed from 2008 to 2018? Med Sci Sports Exerc. 2023;55(8):1416–22.

56. Matthiessen J, Andersen EW, Raustorp A, Knudsen VK, Sørensen MR. Reduction in pedometer-determined physical activity in the adult Danish population from 2007 to 2012. Scand J Public Health. 2015;43(5):525–33.

57. Pelclová J, Frömel K, Řepka E, Bláha L, Suchomel A, Fojtík I et al. Is pedometer-determined physical activity decreasing in Czech adults? Findings from 2008 to 2013. Int J Environ Res Public Health. 2016;13(10).

58. Santos DA, Júdice PB, Magalhães JP, Correia IR, Silva AM, Baptista F, Sardinha LB. Patterns of accelerometer-derived sedentary time across the lifespan. J Sports Sci. 2018;36(24):2809–17.

59. Shin SW, Kim H, Kang YH, Kim J. Accelerometer-derived physical activity and sedentary behavior patterns among Korean adults. Phys Act Nutr. 2023;27(2):25–33.

60. Sisson SB, Camhi SM, Tudor-Locke C, Johnson WD, Katzmarzyk PT. Characteristics of step-defined physical activity categories in U.S. adults. Am J Health Promot. 2012;26(3):152–9.

61. To QG, Stanton R, Schoeppe S, Doering T, Vandelanotte C. Differences in physical activity between weekdays and weekend days among U.S. children and adults: cross-sectional analysis of NHANES 2011–2014 data. Prev Med Rep. 2022;28:101892.

62. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc. 2008;40(1):181–8.

63. Tucker JM, Welk GJ, Beyler NK. Physical activity in U.S.: adults compliance with the physical activity guidelines for americans. Am J Prev Med. 2011;40(4):454–61.

64. Tudor-Locke C, Johnson WD, Katzmarzyk PT. Accelerometer-determined steps per day in US adults. Med Sci Sports Exerc. 2009;41(7):1384–91.

65. Alkerwi A, Pastore J, Sauvageot N, Coroller GL, Bocquet V, d'Incau M, et al. Challenges and benefits of integrating diverse sampling strategies in the observation of cardiovascular risk factors (ORISCAV-LUX 2) study. BMC Med Res Methodol. 2019;19(1):27.

66. Chu AHY, van Dam RM, Biddle SJH, Tan CS, Koh D, Muller-Riemenschneider F. Self-reported domain-specific and accelerometer-based physical activity and sedentary behaviour in relation to psychological distress among an urban Asian population. Int J Behav Nutr Phys Act. 2018;15(1):36.

67. Craig R, Mindell J, Hirani V. Volume 1: physical activity and fitness. Health Survey for England 2008. The Health and Social Care Information Centre; 2009.

68. Craig R, Mindell J, Hirani V. Health Survey for England 2008. Volume 2: methods and documetation. The Health and Social Care Information Centre; 2009.

69. Curtin LR, Mohadjer LK, Dohrmann SM, Montaquila JM, Kruszan-Moran D, Mirel LB, et al. The National Health and Nutrition Examination Survey: Sample Design, 1999–2006. Vital Health Stat. 2012;2(155):1–39.

70. Ferrari GLM, Kovalskys I, Fisberg M, Gomez G, Rigotti A, Sanabria LYC, et al. Methodological design for the assessment of physical activity and sedentary

time in eight latin American countries - the ELANS study. MethodsX. 2020;7:100843.

71. Fisberg M, Kovalskys I, Gomez G, Rigotti A, Cortes LY, Herrera-Cuenca M, et al. Latin American Study of Nutrition and Health (ELANS): rationale and study design. BMC Public Health. 2016;16:93.

72. Fisberg M, Kovalskys I, Previdelli AN, Zimberg IZ, Del'Arco A, Ferrari GLM. Brazilian Study of Nutrition and Health (EBANS) - Brazilian data of ELANS: methodological opportunities and challenges. Rev Assoc Med Bras (1992). 2019;65(5):669–77.

73. Howitt C, Hambleton IR, Rose AM, Hennis A, Samuels TA, George KS, Unwin N. Social distribution of diabetes, hypertension and related risk factors in Barbados: a cross-sectional study. BMJ Open. 2015;5(12):e008869.

74. Lundqvist A, Mäki-Opas T. Health 2011 survey – methods. Tampere, Finland: National Institute for Health and Welfare; 2016.

75. Ma J, Kim H, Kim J. Isotemporal Substitution analysis of accelerometer-derived sedentary behavior and physical activity on Cardiometabolic Health in Korean adults: a Population-based cross-sectional study. Int J Environ Res Public Health. 2021;18:21.

76. NHANES. Laboratory Procedures Manual. National Health and Nutrition Examination Survey. 2004.

77. NHANES. Laboratory Procedures Manual. National Health and Nutrition Examination Survey. 2005.

78. Canadian Health Reports. Canadian Health Measures Survey (CHMS) data user guide: cycle 1. 2011.

79. Canadian Health Reports. Canadian Health measures Survey (CHMS) data user guide: cycle 2. Statistics Canada; 2012.

80. Canadian Health Reports. Canadian Health measures Survey (CHMS) data user guide: cycle 3. Statistics Canada; 2014.

81. Canadian Health Reports. Canadian Health measures Survey (CHMS) data user guide: cycle 4. Statistics Canada; 2017.

82. Canadian Health Reports. Canadian Health measures Survey (CHMS) data user guide: cycle 5. Statistics Canada; 2019.

83. Shephard RJ. A history of physical activity measurement in epidemiology. The objective monitoring of physical activity: contributions of accelerometry to epidemiology, exercise science and rehabilitation. Springer; 2016. pp. 39–83.

84. Waters LA, Galichet B, Owen N, Eakin E. Who participates in physical activity intervention trials? J Phys Act Health. 2011;8(1):85–103.

85. Beyler N, Beyler A. Adjusting for measurement error and nonresponse in physical activity surveys: a Simulation Study. J off Stat. 2017;33(2):533–50.

86. Bammann K, Thomson NK, Albrecht BM, Buchan DS, Easton C. Generation and validation of ActiGraph GT3X + accelerometer cut-points for assessing physical activity intensity in older adults. The OUTDOOR ACTIVE validation study. PLoS ONE. 2021;16(6):e0252615.

87. Freedson PS, John D. Comment on estimating activity and sedentary behavior from an accelerometer on the hip and wrist. Med Sci Sports Exerc. 2013;45(5):962–3.

88. Stevens ML, Gupta N, Eroglu EI, Crowley PJ, Eroglu B, Bauman A, et al. Thigh-worn accelerometry for measuring movement and posture across the 24-hour cycle: a scoping review and expert statement. BMJ open Sport Exerc Med. 2020;6(1):e000874.

89. Montoye AHK, Pivarnik JM, Mudd LM, Biswas S, Pfeiffer KA. Validation and comparison of Accelerometers worn on the hip, thigh, and wrists for measuring physical activity and sedentary behavior. AIMS Public Health. 2016;3(2):298–312.

90. Herrmann SD, Barreira TV, Kang M, Ainsworth BE. How many hours are enough? Accelerometer wear time may provide bias in daily activity estimates. J Phys Act Health. 2013;10(5):742–9.

91. Herrmann SD, Barreira TV, Kang M, Ainsworth BE. Impact of accelerometer wear time on physical activity data: a NHANES semisimulation data approach. Brit J Sport Med. 2014;48(3):278–82.

92. Colley R, Connor Gorber S, Tremblay MS. Quality control and data reduction procedures for accelerometry-derived measures of physical activity. Health Rep. 2010;21(1):63–9.

93. Westgate K, Ridgway C, Rennie K, Strain T, Wijndaele K, Brage S. Feasibility of incorporating objective measures of physical activity in the STEPS program. A pilot study in Malawi; 2020.

94. Haskell WL. Physical activity by self-report: a brief history and future issues. J Phys Act Health. 2012;9(Suppl 1):S5–10.

Publisher's Note