

RESEARCH ARTICLE

Validity, reliability, and readability of single-item and short physical activity questionnaires for use in surveillance: A systematic review

Antonina Tcymbal^{1*}, Sven Messing¹, Rachel Mait¹, Roberto Galindo Perez¹, Taiyeba Akter¹, Ivo Rakovac², Peter Gelius³, Karim Abu-Omar¹

1 Department of Sport Science and Sport, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, **2** World Health Organization Regional Office for Europe, Copenhagen, Denmark, **3** Institute of Sport Sciences, Université de Lausanne, Lausanne, Switzerland

* antonina.tcymbal@fau.de



OPEN ACCESS

Citation: Tcymbal A, Messing S, Mait R, Perez RG, Akter T, Rakovac I, et al. (2024) Validity, reliability, and readability of single-item and short physical activity questionnaires for use in surveillance: A systematic review. PLoS ONE 19(3): e0300003. <https://doi.org/10.1371/journal.pone.0300003>

Editor: Heather Macdonald, UBC: The University of British Columbia, CANADA

Received: August 25, 2023

Accepted: February 19, 2024

Published: March 12, 2024

Copyright: © 2024 Tcymbal et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are included in the paper, and lists of included and excluded records are available on the Open Science Framework (<https://osf.io/dypur/>; doi: [10.17605/OSF.IO/DYPUR](https://doi.org/10.17605/OSF.IO/DYPUR)).

Funding: This research was conducted as part of projects funded by the German Federal Ministry of Health (ZMI5-2522WH0001, ZMI5-2523WH0001). The ministry was neither involved in writing this manuscript nor in the decision to submit the article for publication.

Abstract

Background

Accurate and fast measurement of physical activity is important for surveillance. Even though many physical activity questionnaires (PAQ) are currently used in research, it is unclear which of them is the most reliable, valid, and easy to use. This systematic review aimed to identify existing brief PAQs, describe and compare their measurement properties, and assess their level of readability.

Methods

We performed a systematic review based on the PRISMA statement. Literature searches were conducted in six scientific databases. Articles were included if they evaluated validity and/or reliability of brief (i.e., with a maximum of three questions) physical activity or exercise questionnaires intended for healthy adults. Due to the heterogeneity of studies, data were summarized narratively. The level of readability was calculated according to the Flesch-Kincaid formula.

Results

In total, 35 articles published in English or Spanish were included, evaluating 32 distinct brief PAQs. The studies indicated moderate to good levels of reliability for the PAQs. However, the majority of results showed weak validity when validated against device-based measurements and demonstrated weak to moderate validity when validated against other PAQs. Most of the assessed PAQs met the criterion of being "short," allowing respondents to complete them in less than one minute either by themselves or with an interviewer. However, only 17 questionnaires had a readability level that indicates that the PAQ is easy to understand for the majority of the population.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: BPAAT, Brief Physical Activity Assessment Tool; EHIS-PAQ, European Health Interview Survey Physical Activity Questionnaire; GPAQ, Global Physical Activity Questionnaire; IPAQ, International Physical Activity Questionnaire; JACC, Japan Collaborative Cohort Questionnaire; L-CAT, The Stanford Leisure-Time Activity Categorical Item; NPAQ-short, The Nordic Physical Activity Questionnaire-short; PA, physical activity; PAQ, physical activity questionnaire; SIPAM, Single Item Physical Activity Measure; SR-PA L7, The Seven-level Single-Question Scale for Self-Reported Leisure Time Physical Activity; TAM, Total Activity Measure.

Conclusions

This review identified a variety of brief PAQs, but most of them were evaluated in only a single study. Validity and reliability of short and long questionnaires are found to be at a comparable level, short PAQs can be recommended for use in surveillance systems. However, the methods used to assess measurement properties varied widely across studies, limiting the comparability between different PAQs and making it challenging to identify a single tool as the most suitable. None of the evaluated brief PAQs allowed for the measurement of whether a person fulfills current WHO physical activity guidelines. Future development or adaptation of PAQs should prioritize readability as an important factor to enhance their usability.

Background

It has been demonstrated that regular physical activity (PA) can help improve physical and mental functions as well as reverse some effects of chronic diseases [1]. Regularly engaging in 150 minutes of PA per week [2], alongside following a healthy diet and abstaining from smoking and alcohol consumption, is seen as being key to prevent non-communicable diseases.

However, accurate measurement of PA levels is important to determine the amount of activity needed to improve health and identify links with other health outcomes and behaviors [3]. To this day, self-report questionnaires are the most common measures to collect PA data, often as part of surveillance systems such as the Behavioral Risk Factor Surveillance System [4], the WHO STEPwise approach to noncommunicable disease risk factor surveillance [5], and the European Health Interview Survey [6]. While such self-report questionnaires have seen widespread use because of their efficiency, they have also been shown to have limitations related to bias and data accuracy [7].

Many different self-report questionnaires have been developed over the years with the Global Physical Activity Questionnaire (GPAQ), the International Physical Activity Questionnaire (IPAQ; also available as a short form, IPAQ-SF), and the European Health Interview Survey Physical Activity Questionnaire (EHIS-PAQ) being the most widely utilized in global PA surveillance [8]. All three questionnaires assess PA across different domains, asking respondents to report their PA in a typical week (GPAQ, EHIS-PAQ) or during the last 7 days (IPAQ-SF). All three are comparatively complex and range from seven items (IPAQ-SF) to 16 (GPAQ). Nevertheless, the measurement properties of these questionnaires are modest. For GPAQ [9], IPAQ-SF [10, 11], and the EHIS-PAQ [12], different studies have demonstrated reasonable reliability but comparatively low validity. In a recent review of IPAQ-SF, GPAQ, and EHIS-PAQ, the questionnaires showed low-to-moderate validity against device-based measures of PA such as accelerometers, and moderate-to-high validity against subjectively measured PA such as other questionnaires [8].

It is well known that questionnaires with many items can increase the response burden on respondents [13, 14], which has resulted in the development of several short self-report instruments. In contrast to the detailed questionnaires mentioned above, the purpose of short PA questionnaires is to simplify and speed up the procedure for assessing PA levels.

In addition, language-related difficulties are currently an important topic in public health. Patient education materials can increase patient compliance, but only if they are written in a language that is easy for the patient to understand [15]. Regarding PA questionnaires, Altschuler et al. [16] found significant gaps between respondents' interpretations of some PA

questions and researchers' original assumptions about what those questions were intended to measure. One of the characteristics of language difficulty is the level of readability, which indicates how easily readers can understand the text. Research of texts used in healthcare consistently shows that materials intended for patients often require a high level of education and are too complicated for the average person [15, 17]. In relation to physical activity questionnaires (PAQs), readability can influence the amount of time a person needs to understand the question and, if the text is too complicated, may potentially decrease the response rate and accuracy of the answer.

A number of existing reviews have investigated the measurement properties of PA questionnaires. Van Poppel et al. [18] reviewed the validity and reliability methodology of 85 versions of PAQs with no clear consensus regarding the best questionnaire for PA measurement. Helmerhost et al. [19] studied reliability and objective criterion-related validity of 34 newly-developed and 96 existing PAQs. Both reviews included PAQs regardless of their length. To our knowledge, a dedicated review of short PA questionnaires measurement properties has not yet been performed. Therefore, this review was conducted upon request of representatives of European surveillance systems to provide an overview of short PAQs, their measurement properties, and their level of readability. The intention of this initiative was to identify PAQs that are brief, valid, reliable, and easy-to-understand for the general population. From a more general perspective, this review aims to inform the further development and harmonization of surveillance systems.

Methods

This review follows the guidelines of the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) statement [20] and the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) guideline for systematic reviews of patient-reported outcome measures [21]. As the study did not involve the collection and analysis of participant data, ethical approval and informed consent were not required.

Information sources and search strategy

A systematic search was performed in six databases (PubMed, Web of Science, Scopus, CINAHL, SportDiscus, PsycInfo) in March 2022, and it was repeated on July 21, 2023, to check for newly published articles. Additionally, reference lists from previous reviews of PA questionnaires and other relevant publications were screened to identify additional studies. A comprehensive search strategy was developed with a combination of keywords in the categories of measured construct (physical activity) and type of instruments (brief/short questionnaire).

The resulting search string was as follows:

("physical activit*" OR "physical inactivit*")

AND

(questionnaire OR measure OR evaluat* OR assess* OR surveillance OR monitor* OR screening)

AND

("single item" OR single-item OR "single question" OR "one item" OR one-item OR "one question" OR "brief questionnaire" OR "short questionnaire" OR "brief assessment" OR "short assessment" OR "two-item*" OR "two-item*" OR "two questions" OR "brief physical activity assessment" OR "Single response" OR "Single-response")

No restrictions were made regarding language or publication date.

Eligibility criteria

Articles were included in the review if they fulfilled the following eligibility criteria:

1. The article described a self-report PA or exercise questionnaire intended for healthy adults.
2. The evaluated questionnaire was brief and included a maximum of three questions.
3. The article investigated one or more measurement properties of the questionnaire.
4. The article was published in a peer-reviewed journal.

Articles were excluded based on the following exclusion criteria:

1. The article focused on a questionnaire that measured only physical inactivity, screen time, or sedentary behavior.
2. The evaluated questionnaire was intended solely for children and adolescents or people with specific conditions.
3. The article did not investigate the measurement properties of the questionnaire.

Study selection

Two reviewers independently screened and selected the relevant articles. First, all articles were screened based on titles and abstracts. If the title and/or abstract indicated that the study fulfilled the inclusion criteria, both reviewers screened the full text for eligibility. When necessary, supplementary files were also reviewed for additional information. Disagreements between the reviewers were discussed within the research team until a consensus was reached.

Records were managed using the Covidence systematic review software (Veritas Health Innovation, Melbourne, Australia; www.covidence.org) and EndNote X9 (Clarivate Analytics, Philadelphia, PA, USA).

Data extraction

Data of included studies were extracted and summarized by one reviewer and verified by a second reviewer to reduce bias and error. Discrepancies were discussed between reviewers to achieve consensus. Extracted information included: publication details (first author, year of publication, country), sample characteristics (number of participants, age category, special health conditions), the measurement tool(s) explored, who assessed PA levels of participants, assessed measurement properties, other measurement tool(s) used as a comparison, reliability test-retest interval, and the results of the study.

Risk of bias assessment

The methodological quality of the individual studies included in this review was assessed with 18 questions based on the Appraisal tool for Cross-Sectional Studies (AXIS) [22]. In order to specifically assess the quality of the tools' validity and reliability measurement properties, some questions were modified based on the COSMIN risk of bias checklist [21] and on suggestions from previous systematic reviews examining PA assessment measures [23]. This quality assessment evaluated articles based on study design, sample size, participant selection process, appropriate blinding, examiner experience, method of measurement, adequate data reporting,

internal consistency, and six other categories (see [S1 File](#)). Risk of bias assessment was conducted independently by two reviewers, with any discrepancies resolved through discussion in the research team. Studies were scored 1 if they satisfied the quality element, and 0 if they did not. The summary score (range: 0–18) indicates the risk of bias, with a higher score indicating higher quality and therefore a lower risk of bias.

Data synthesis and analysis of measurement properties

The primary objective of this systematic review was to identify existing short questionnaires suitable for assessing PA levels in surveillance and primary care settings, and to compare their measurement properties. To accomplish this goal, relevant information from the included studies was summarized separately for each questionnaire. In order to evaluate the reliability and validity of the identified questionnaires, a range of tests were employed in the included studies. Some studies reported results for a total questionnaire summary score, while others assessed reliability and validity for specific aspects, intensities, or domains of the questionnaire. Additionally, certain studies examined these measurement properties within subgroups of the test population. Due to the heterogeneity in the methods used and the lack of standardized reporting across studies, a quantitative meta-analysis was not feasible. Consequently, the information from the included studies was summarized narratively, highlighting the key findings for each questionnaire. This narrative synthesis allows for a comprehensive overview of the reliability and validity findings, highlighting strengths and limitations of each questionnaire.

Analysis of length and readability of questionnaires

To determine how quickly questionnaires could be answered and how easy it is to understand the questions, identified PAQs were analyzed for word count and readability level. The expected time the tool would take for self-administration (silent reading speed) and interviewer administration (spoken word speed) was calculated based on the respective questionnaire's word count and English reading speeds established by Brysbaert [24]. The level of readability was calculated according to the Flesch-Kincaid formula, which was chosen because it is the most commonly used tool to calculate the readability level of written health information [25]. The formula has two forms: the Flesch Reading-Ease-Score, and the Flesch-Kincaid Grade Level. The Flesch Reading-Ease-Score test produces a score from 0 to 100, and higher scores indicate material that is easier to read; lower scores mark passages that are more difficult to read. The Flesch-Kincaid Grade Level formula matches the text to the grade level achievement (number of years of education) required to understand the text. We applied both formulas for each questionnaire using an online calculator [26].

Results

Study selection process

The search across six databases resulted in 2,693 publications. After removing 1,394 duplicates, 1,299 articles were screened based on title and abstract. 69 studies were found eligible for full text assessment. Two recently published articles [27, 28] and one article identified while screening the reference lists of relevant publications [29] satisfied all eligibility criteria and were therefore included in the analysis. Subsequently, 34 full text articles were excluded due to the length of the questionnaire ($n = 15$), the publication type ($n = 14$), a lack of self-reporting on PA and exercise ($n = 2$), and a study design that did not consider validity and reliability ($n = 3$). In total, 35 studies were included for data extraction and methodological quality assessment. A summary of the search results is presented in [Fig 1](#).

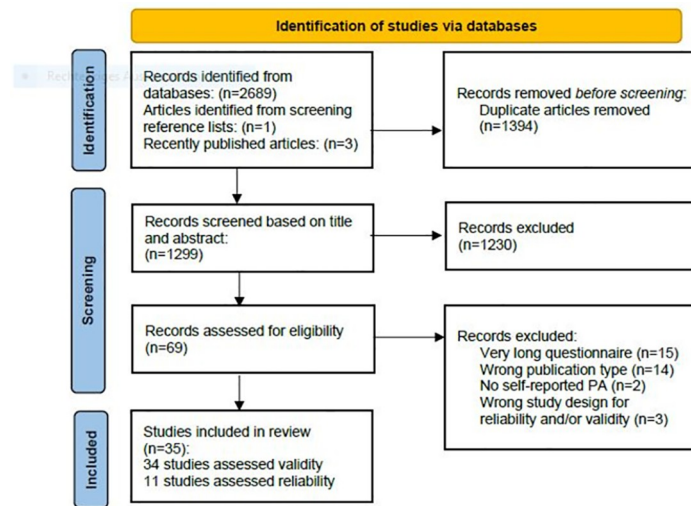


Fig 1. PRISMA flowchart of study eligibility. PA = physical activity.

<https://doi.org/10.1371/journal.pone.0300003.g001>

Characteristics of included studies

A summary of the characteristics of the 35 included studies is presented in [S1 Table](#). The included studies were conducted in Western European countries ($n = 14$), the USA ($n = 11$), Australia ($n = 5$), Canada ($n = 2$), New Zealand ($n = 2$), and Japan ($n = 1$). Thirty-four of the included studies were published in English, and one was published in Spanish language.

A total of 114,199 adults were assessed using the 32 unique brief PAQs (with sample sizes ranging from 9 to 39,379). In 26 studies, the sample consisted only of healthy adults, and nine studies also included specific populations such as older adults [30] [31, 32], overweight/obese adults [33, 34], as well as patients with rheumatoid arthritis [35], coronary heart disease [36, 37], and chronic obstructive pulmonary disease [38].

Thirty-two studies documented who completed the PAQ. In 17 studies, respondents filled out the questionnaire by themselves, and in 15 studies the PA level was collected from an interviewer reading a questionnaire.

Questionnaires asked for the amount of PA ($n = 6$), the number of days per week with a sufficient amount of PA ($n = 4$), general exercise participation ($n = 4$), self-reported activity compared with peers ($n = 2$), or for respondents to choose a categorical descriptor of PA levels ranging from “inactive” to “very active” ($n = 16$).

Seven of the included studies investigated the measurement properties of the Single Item Physical Activity Measure (SIPAM) [27, 31, 39–43], five were related to the Brief Physical Activity Assessment Tool (BPAAT) [38, 44–47], and two to the Stanford Leisure-Time Activity Categorical Item (L-Cat) [33, 34]. The other 29 PAQs were assessed in a single study.

Risk of bias assessment

The average quality score of included studies was 12, with a range between 7 and 17 (out of 18). Thirteen studies received 14 or more points, which can be considered a high methodological quality. In most of the studies, the aims and objectives, the process of the measurement properties investigation, the statistical analysis and the results were sufficiently described, and the study design was appropriate. However, the included papers reported poorly on whether

the examiners had enough experience with the tool and whether they were blinded to participant characteristics, previous findings, or other observed findings. Only four studies justified their choice of sample size. Additionally, many studies failed to consistently report reasons for dropout and characteristics of non-responders. Quality scores of the individual studies are reported in [S1 Table](#).

Questionnaires' measurement properties

A summary of the reliability, validity, and diagnostic test accuracy data found in the identified studies is presented in [S2 Table](#). Studies commonly use a number of different statistical analyses to define absolute (agreement between the two measurement tools) or relative (the degree to which the two measurement tools rank individuals in the same order) validity and reliability. These types of statistical analysis include correlations (Pearson's; Spearman's; interclass), regression, kappa statistics, and area under the receiver operating curve (AUC-ROC).

Eleven studies measured the reliability of brief PAQs. All of them used a test-retest procedure to measure the consistency of the PAQs. Statistical methods and test-retest intervals varied widely between studies. Overall, studies showed moderate to good reliability levels of the PAQs.

The validity of PAQs was assessed in 34 studies. As a "gold standard" for validation, 13 studies used other PAQs, 15 studies validated brief PAQs against accelerometers or pedometers, and 12 studies compared results of the brief PAQs to other objective measurements, such as BMI [29, 33, 48, 49], VO₂ max [48, 50], or doubly labeled water [35]. Validity coefficients were in general considerably lower than reliability coefficients. The majority of results showed weak validity of brief PAQs against device-based and objective measurements and weak to moderate validity against other PAQs.

Length and complexity of the questionnaires

The texts of all questionnaires and information about their length and readability level are presented in [S3 Table](#). It should be noted that, while some PAQs were created and conducted in other languages, length and readability were calculated for the English versions.

Twenty-seven out of thirty-two assessed PAQs can be read silently by the respondent or aloud by the interviewer in less than one minute. However, some questionnaires presented as "brief" by the authors of the identified studies contain a large amount of text and require more than one minute for reading: the Nordic Physical Activity Questionnaire (NPAQ) short, the Total Activity Measure (TAM), the Self-report Scale to Assess Habitual Physical Activity, the Stanford Leisure-Time Activity Categorical Item, and the Stanford Brief Activity Survey.

The calculation of the readability levels with the Flesch-Kincaid formula showed that only 17 out of 32 brief PAQs had readability levels of "easy to read" ($n = 9$) or "plain English," ($n = 8$) and can be easily understood by the majority of the population. Other questionnaires have long sentences and/or many complicated words (three syllables and more), making them difficult to read. Seven questionnaires are fairly difficult to read. Eight questionnaires were rated as "difficult to read" or "very difficult to read" and require college-level education; it is likely that people with lower education levels will find them hard to understand. The original version of the SIPAM was rated as "fairly difficult to read", but adjusting it to special population and adding additional terms to the same sentences made the questionnaire more complicated. For example, parents' version was appraised as "very difficult to read".

Table 1. Comparison of the brief physical activity questionnaires by reliability, validity, length and readability.

Tool	Reliability	Validity		Word count	Readability*
		Against device-based and objective measures	Against other PAQ		
Single Item Physical Activity Measure (SIPAM) [27, 31, 39, 40, 41, 42, 43]	Good (0,63–0,82)	Poor to Good (0,22–0,81)	Poor to Good (0,33–0,81)	63 –original 111 – parents version	Fairly difficult to read Parents version—Very difficult to read
Brief Physical Activity Assessment Tool (BPAAT) [44, 45, 46, 47]	Moderate to good (0,53–0,72)	Poor to moderate (0,18–0,43)	Moderate to good (0,45–0,64)	106	Difficult to read
Three-Question Assessment [47]	Moderate to good (0,56–0,63)	Poor (0,24–0,31)	Moderate (0,39–0,43)	140	Fairly difficult to read
Nordic physical activity questionnaire (NPAQ) -short (open ended) [51]	Good (0,82)	Poor (0,33)	NM	237	Difficult to read
Total Activity Measure (Version 2) [37]	Good (0,82)	Poor (0,36–0,38)	NM	241	Fairly difficult to read
Japan Collaborative Cohort (JACC) Questionnaire [52]	Moderate (0,39–0,56)	NM	Moderate (0,43–0,60)	90	Plain English
Relative PA Question (Compared to peers) [30]	Moderate (0,56)	Poor to moderate (0,28–0,57)	NM	25	Easy to read
The Seven-level Single-Question Scale [53]	Good (0,7)	Poor (0,33)	NM	144	Fairly difficult to read
Absolute PA Question [30]	Good (0,7)	Poor (0,1–0,33)	NM	34	Fairly difficult to read
A single-item 5-point rating of usual PA (Usual PA Scale) [49]	Good (0,68–0,88)	NM	Good (0,66)	147	Plain English
The Stanford Leisure-Time Activity Categorical Item (L-CAT) [33, 34]	Good (0,64–0,8)	Poor (0,36–0,38)	NM	331	Plain English

PAQ = physical activity questionnaire; NM = not measured; green = good level of reliability, validity or readability, low amount of words; yellow = moderate level of reliability, validity or readability, moderate amount of words; red = poor level of reliability, validity or readability, high amount of words;

* = the readability level was calculated using the Flesch Reading-Ease-Score based on the English versions of the questionnaires.

<https://doi.org/10.1371/journal.pone.0300003.t001>

Comparison of brief PAQs

Data on both validity and reliability were only available for nine of the 32 assessed PAQs. These brief PAQs were chosen for an in-depth comparison. Their summarized measurement properties and linguistic characteristics are presented in [Table 1](#).

The Single Item Physical Activity Measure (SIPAM) uses a single question to assess the number of days per week on which 30 minutes or more of PA are performed (excluding housework and work-related PA). For this PAQ, the largest number of studies on validity and reliability were found. The SIPAM showed good reliability levels in all studies; however, results on validity varied considerably between studies—from poor to good—against both device-based measures and other self-reported PAQs. Zwolinsky et al. (2015) also measured SIPAM's ability to identify people that meet or do not meet WHO PA guidelines. The tool showed a low diagnostic capacity compared to the IPAQ. The agreement between the SIPAM and the IPAQ was low for identification of participants who meet WHO PA guidelines ($\kappa = 0.13$, 95% CI 0.12 to 0.14) and moderate for the classification of inactive participants ($\kappa = 0.45$, 95% CI 0.43 to 0.47) [43].

The Brief Physical Activity Assessment Tool (BPAAT) consists of two questions, one regarding the frequency and duration of vigorous PA and the other regarding moderate PA and walking performed in an individual's usual week. By combining the results of both questions (scores can range from 0 to 8), the subject can be classified as insufficiently (0–3 score) or sufficiently active (≥ 4 score). Results of reviewed studies showed that the BPAAT has moderate to good levels of reliability and validity in comparison with other PAQs; however, comparison with accelerometers identified only poor to moderate validity.

Smith et al. (2005) evaluated the Three-Question Assessment variant of the BPAAT which has separate questions about moderate PA and walking. The results of the study did not find a considerable difference in validity and reliability between the two- and three-question versions. One study also reported that more physicians preferred the two-question version (BPAAT) as it is shorter and therefore easier to use (Smith, 2005).

The Nordic Physical Activity Questionnaire-short (NPAQ-short) includes one question about moderate to vigorous PA and a second question about vigorous PA. Danquah et al. (2018) compared open-ended and closed-ended questions with the open-ended version achieving better results. The open-ended version showed good reliability but performed similarly to other questionnaires, showing poor validity when compared against device-based PA measures. The analyses showed that the questionnaire was one of the longest and was rated “difficult to read”. The agreement with accelerometer data in identification of people that meet or do not meet the WHO PA guidelines was low ($\kappa = 0.42$).

The Total Activity Measures (TAM) includes three open-ended questions about strenuous, moderate, and mild PA. The revised second version, TAM2, asks about the total time spent at each activity level over a 7-day period. The TAM2 showed good reliability when validated against device-based measured PA. The word count was high at 241, and readability was rated as “fairly difficult”.

The Japan Collaborative Cohort (JACC) Questionnaire has three items. Two questions focus on leisure-time PA, i.e., time per week engaging in sport or PA (with options ranging from “little” to “at least 5 hours”), and frequency of engagement in sport over the past year (options from “seldom” to “at least twice a week”). The third question asks about daily walking patterns (options from “little” to “more than 1 hour”). The questionnaire showed moderate reliability after one year and moderate validity when compared against a more in-depth interview by a trained researcher. The questions are 90 words in total and were rated as “plain English.”

The Relative PA Question is the shortest of the compared PAQs and has an easy readability level. It allows respondents to compare their level of PA with other people of the same age and to choose from five categories ranging from “much more active” to “much less active.” The questionnaire has a moderate reliability level and showed poor validity against device-based and objective measures.

The Absolute PA Question asked participants to choose what best describes their activity level from three options: (1) vigorously active for at least 30 minutes, three times per week; (2) moderately active at least three times per week; or (3) seldom active, preferring sedentary activities. The questionnaire is relatively short, but as most of the other brief PAQs, it showed high reliability and poor validity against device-based and objective measures.

The Seven-Level Single-Question Scale (SR-PA L7) requires respondents to assign themselves to a level ranging from “I do not move more than is necessary in my daily routines/chores” and “I participate in competitive sports and maintain my fitness through regular training.” These seven items are supposed to categorize respondents as maintaining low, medium, or high PA. The SR-PA L7 showed poor validity when compared to device-based measured PA but good reliability. The scale is moderately long at 144 words and was rated as “fairly difficult to read.”

A single-item 5-point rating of usual PA (Usual PA Scale) includes descriptions of three PA levels: highly active, moderately active, and inactive. Respondents need to read a description of each category and identify their usual PA level from the list. The questionnaire’s readability level was ranked as “plain English,” and it showed good reliability and validity levels. However, validation was done only against another PAQ which is not well-known and requires further research.

The Stanford Leisure-Time Activity Categorical Item (L-CAT) is a single-item questionnaire that consists of six descriptive PA categories ranging from inactive to very active. Each category consists of one or two statements describing common activity patterns over the past month, differing in frequency, intensity, duration, and types of activity. The categories are described in “plain English,” but this made the questionnaire longer than the other questionnaires. The L-CAT showed high reliability but poor validity against pedometer and accelerometer validation.

Discussion

This review assessed the validity, reliability, and readability of brief PAQs for adults. To our knowledge, it is the first review focused specifically on brief PAQs and also the first to assess their readability. The sheer number of brief PAQs we identified ($n = 32$) indicates a high research interest in such instruments. However, it also indicates a lack of harmonization when it comes to PA assessment using brief questionnaires [54].

Overall, most PAQs were reported to have poor or moderate validity. They showed higher validity levels against other self-reported tools than against device-based and other objective PA measures. Although reliability is an important measurement property, it was assessed only in 11 studies. About half of all PAQs showed a moderate to good level of reliability. These results are in line with other reviews of PAQs [18, 19, 55, 56]. The validity and reliability of short questionnaires is similar to the respective measurement properties of longer questionnaires, such as GPAQ, IPAQ-SF and EHIS-PAQ [8–12].

A significant difficulty in conducting this review was that the studies used different methods for validation, varying time-intervals between repeated measurements, and different statistical methods to analyze data. Complete data on validity and reliability were only available for nine PAQs, and only those could be compared in greater detail. However, it has not been possible to identify a specific questionnaire that is most accurate. For the SIPAM, the largest number of studies on validity and reliability were found. However, there was a high variation from poor to good levels of validity. In addition, the questionnaire has a poor level of readability. Another example is the Usual PA Scale. Even though it has good levels of validity and reliability, the PAQ was investigated only in one study and was not validated against device-based measures.

In general, the methodological quality of the included studies was modest. The most common flaws were comparably small sample sizes, a lack of sample size justification, and a poor description of the validity and reliability assessment process. Additionally, most studies utilized convenience samples, making it impossible to assess if measurement properties would differ between adults with different levels of socioeconomic status and/or educational attainment.

Also, the included PAQs used different concepts of measuring PA, further complicating a direct comparison between instruments. For example, the SIPAM measures on how many days per week respondents perform 30 minutes or more of PA; the Absolute PAQ requires respondents to choose from several descriptions of different PA levels; and the Relative PA question asks them to compare their level of PA with peers. Some PAQs, such as the TAM, aim to assess the total volume of moderate-to-vigorous PA. Others focus on particular PA domains. And yet others assess only leisure-time PA. This variety can be partly explained by efforts to keep PAQs short at the price of excluding some PA dimensions (type, duration, intensity, volume). However, it could also be interpreted as a lack of common understanding about which dimensions of PA should be assessed with brief PAQs. It should also be taken into account that none of the reviewed brief PAQs allowed for the measurement of whether a person fulfills the WHO physical activity guidelines [2]. This raises the important question of

whether specialists should collaborate to improve current brief questionnaires to meet the needs of surveillance systems.

All of the included studies were conducted in highly developed nations, and 24 of them took place in English-speaking countries. This most certainly biased the results, since terminology related to PA differs between languages, as do levels of literacy [57]. Ultimately, to assess reliability and validity of PAQs, many more studies should be conducted in developing nations to obtain a more accurate assessment of their suitability for international surveillance systems. Certainly, in order to stimulate such research, funding opportunities need to be made available to research teams from such nations.

The results of this study point out that many PAQs have low readability levels. This is particularly disturbing when considering that most of these PAQs were tested in developed nations with comparatively high levels of literacy. Potentially, poor readability is related to the modest measurement properties that PAQs commonly come with. This highlights the need to revise current PAQs considering their readability levels and linguistic features. It is also crucial to keep this in mind when developing questionnaires in the future. Questionnaires that are more readable or might even have been co-developed [58] with the intended population groups are likely to yield better measurement properties. This would also strengthen the case of integrating them into PA surveillance systems. However, at this point in time, there is a dearth of research relating PAQs and their measurement properties to readability. Advancing knowledge in this field would also benefit longer PA questionnaires (such as IPAQ and GPAQ) that are widely utilized in surveillance and potentially score low on readability as well.

This review comes with certain limitations. The search was limited to scientific databases, and only studies published in peer reviewed scientific journals were included. This can lead to a publication bias, as all other types of publications and gray literature were excluded. The varying measurement methods and conditions complicated the comparison of findings from different studies and limited data analysis to a narrative description of differences between tools. This led to more subjective results and greater difficulty in identifying the tools with the best measurement properties. It should be also taken into account that the Flesch-Kincaid readability formula was created for longer texts and not adapted for short questionnaires. Consequently, the readability levels of PAQs presented here should be interpreted with caution.

Recommendations for physical activity surveillance:

Based on the results of this systematic review, we recommend experts and decision-makers to take the following aspects into account when further developing or harmonizing surveillance systems:

1. Shorter PAQs could be considered for use instead of long PAQs for some use cases, as they demonstrate comparable validity and reliability levels. The primary advantage of using shorter questionnaires is the reduction of response time, which is important for both surveillance systems and primary care. Nevertheless, it should be taken into account that short questionnaires allow to collect a fairly small amount of information and are not suitable for all PA measurement purposes.
2. The majority of PAQs have the goal to identify individuals with an insufficient level of PA. However, none of the reviewed questionnaires captures all necessary components of the complete WHO guidelines on PA (both duration of moderate- and vigorous-intensity PA and frequency of strength training per week). This highlights the necessity of discussing the content that should be included in questionnaires and the potential development of new tools based on the WHO guidelines.

3. It is necessary to consider linguistic characteristics when developing, testing and translating questionnaires. Potential strategies for improving readability and understandability of questionnaires can be the collaboration with linguists and the involvement of various population groups in the development process.

Conclusion

This systematic review sheds light on the validity, reliability, and readability of short PAQs. The diversity of research methods and insufficient information about measurement properties for some PAQs made it impossible to compare questionnaires in detail and to identify the most accurate tool. Results indicate that additional research on PAQs is needed, notably regarding their reliability and validity, readability, and applicability to non-English speaking and/or developing countries.

Recent years have seen a shift from self-report questionnaires towards device-based measures in PA surveillance. This shift has been partially motivated by the persistently modest measurement properties of PAQs, as well as technological advancements in the quality and affordability of accelerometers and other PA measurement tools. However, device-based PA assessment comes with its own set of limitations [59], and self-report and device-based measures have been described as measuring entirely different parameters. Right now, it seems unclear what the future for PA surveillance might hold and if brief PAQs will continue to play a role.

In this regard, the key merit of short PAQs is the significantly shorter response time compared to established PA surveillance questionnaires (such as GPAQ, IPAQ, and EHIS-PAQ), while validity and reliability of short and long questionnaires are found to be at a comparable level. This makes them highly appealing for use in surveillance systems. However, short PAQs need to be improved to allow for comparison with WHO's guidelines on PA.

Supporting information

S1 File. Blank for studies quality assessment.

(DOCX)

S1 Table. Studies overview.

(DOCX)

S2 Table. Summary of measurement properties.

(DOCX)

S3 Table. Length and readability levels of short physical activity questionnaires.

(DOCX)

Author Contributions

Conceptualization: Antonina Tcymbal, Sven Messing, Ivo Rakovac, Peter Gelius, Karim Abu-Omar.

Data curation: Antonina Tcymbal, Rachel Mait, Roberto Galindo Perez, Taiyeba Akter.

Formal analysis: Antonina Tcymbal, Karim Abu-Omar.

Funding acquisition: Karim Abu-Omar.

Investigation: Antonina Tcymbal.

Methodology: Antonina Tcymbal, Sven Messing, Ivo Rakovac, Peter Gelius, Karim Abu-Omar.

Project administration: Antonina Tcymbal.

Software: Antonina Tcymbal, Rachel Mait, Roberto Galindo Perez, Taiyeba Akter.

Supervision: Ivo Rakovac, Karim Abu-Omar.

Visualization: Antonina Tcymbal.

Writing – original draft: Antonina Tcymbal, Rachel Mait, Roberto Galindo Perez, Taiyeba Akter, Karim Abu-Omar.

Writing – review & editing: Antonina Tcymbal, Sven Messing, Rachel Mait, Roberto Galindo Perez, Taiyeba Akter, Ivo Rakovac, Peter Gelius, Karim Abu-Omar.

References

1. McPhee JS, French DP, Jackson D, Nazroo J, Pendleton N, Degens H. Physical activity in older age: perspectives for healthy ageing and frailty. *Bio gerontology*. 2016; 17(3):567–80. <https://doi.org/10.1007/s10522-016-9641-0> PMID: 26936444
2. World Health Organization. Guidelines on physical activity and sedentary behaviour. Geneva: World Health Organization; 2020.
3. Welk G. Physical Activity Assessments for Health-related Research. USA, Ames: Human Kinetics; 2002.
4. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System Survey Questionnaire. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; 2021.
5. World Health Organization. WHO STEPS Surveillance Manual: The WHO STEPwise approach to non-communicable disease risk factor surveillance. Geneva: World Health Organization; 2020.
6. Finger JD, Tafforeau J, Gisle L, Oja L, Ziese T, Thelen J, et al. Development of the European Health Interview Survey—Physical Activity Questionnaire (EHIS-PAQ) to monitor physical activity in the European Union. *Arch Public Health*. 2015; 73:59. <https://doi.org/10.1186/s13690-015-0110-z> PMID: 26634120
7. Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. *Research quarterly for exercise and sport*. 2000; 71:1–14.
8. Sember V, Meh K, Sorić M, Starc G, Rocha P, Jurak G. Validity and Reliability of International Physical Activity Questionnaires for Adults across EU Countries: Systematic Review and Meta Analysis. *International journal of environmental research and public health*. 2020; 17(19):7161. <https://doi.org/10.3390/ijerph17197161> PMID: 33007880
9. Bull FC, Maslin TS, Armstrong T. Global physical activity questionnaire (GPAQ): nine country reliability and validity study. *J Phys Act Health*. 2009; 6(6):790–804. <https://doi.org/10.1123/jpah.6.6.790> PMID: 20101923
10. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. *Medicine and science in sports and exercise*. 2003; 35(8):1381–95. <https://doi.org/10.1249/01.MSS.0000078924.61453.FB> PMID: 12900694
11. Lee PH, MD J., Lam TH, Stewart SM. Validity of the International Physical Activity Questionnaire Short Form (IPAQ-SF): a systematic review. *The international journal of behavioral nutrition and physical activity*. 2011; 8:115. <https://doi.org/10.1186/1479-5868-8-115> PMID: 22018588
12. Baumeister SE, Ricci C, Kohler S, Fischer B, Topfer C, Finger JD, et al. Physical activity surveillance in the European Union: reliability and validity of the European Health Interview Survey-Physical Activity Questionnaire (EHIS-PAQ). *The international journal of behavioral nutrition and physical activity*. 2016; 13:61. <https://doi.org/10.1186/s12966-016-0386-6> PMID: 27215626
13. Ware J Jr., M. K, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Medical Care*. 1996; 34(3):220–33. <https://doi.org/10.1097/00005650-199603000-00003> PMID: 8628042
14. Rolstad S, Adler J, Rydén A. Response burden and questionnaire length: is shorter better? A review and meta-analysis. *Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2011; 14(8):1101–8. <https://doi.org/10.1016/j.jval.2011.06.003> PMID: 22152180

15. Grabeel KL, Russomanno J, Oelschlegel S, Tester E, Heidel RE. Computerized versus hand-scored health literacy tools: a comparison of Simple Measure of Gobbledygook (SMOG) and Flesch-Kincaid in printed patient education materials. *J Med Libr Assoc*. 2018; 106(1):38–45. <https://doi.org/10.5195/jmla.2018.262> PMID: 29339932
16. Altschuler A, Picchi T, Nelson M, Rogers JD, Hart J, Sternfeld B. Physical activity questionnaire comprehension: lessons from cognitive interviews. *Med Sci Sports Exerc*. 2009; 41(2):336–43. <https://doi.org/10.1249/MSS.0b013e318186b1b1> PMID: 19127192
17. Wallace LS, Lennon ES. American Academy of Family Physicians patient education materials: Can patients read them? *Family Medicine*. 2004; 36(8):571–4. PMID: 15343418
18. van Poppel MN, Chinapaw MJ, Mokka LB, van Mechelen W, Terwee CB. Physical activity questionnaires for adults: a systematic review of measurement properties. *Sports Medicine (Auckland, N Z)*. 2010; 40(7):565–600. <https://doi.org/10.2165/11531930-000000000-00000> PMID: 20545381
19. Helmerhorst HJH, S. B, Warren J, Besson H, Ekelund U. A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. *The international journal of behavioral nutrition and physical activity*. 2012; 9:103. <https://doi.org/10.1186/1479-5868-9-103> PMID: 22938557
20. 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. <https://doi.org/10.1136/bmj.n71> PMID: 33782057
21. Prinsen CAC, Mokka LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care, and Rehabilitation*. 2018; 27(5):1147–57. <https://doi.org/10.1007/s11136-018-1798-3> PMID: 29435801
22. Downes MJ, Brennan ML, Williams HC, Dean RS. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ Open*. 2016; 6(12):e011458. <https://doi.org/10.1136/bmjopen-2016-011458> PMID: 27932337
23. Williams MA, McCarthy CJ, Chorti A, Cooke MW, Gates S. A systematic review of reliability and validity studies of methods for measuring active and passive cervical range of motion. *Journal of Manipulative and Physiological Therapeutics*. 2010; 33(2):138–55. <https://doi.org/10.1016/j.jmpt.2009.12.009> PMID: 20170780
24. Brysbaert M. How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*. 2019; 109.
25. Wang LW, Miller MJ, Schmitt MR, Wen FK. Assessing readability formula differences with written health information materials: application, results, and recommendations. *Res Social Adm Pharm*. 2013; 9(5):503–16. <https://doi.org/10.1016/j.sapharm.2012.05.009> PMID: 22835706
26. charactercalculator.com. Flesch Reading Ease 2023 [<https://charactercalculator.com/flesch-reading-ease/>].
27. Bauman AE, Richards JA. Understanding of the Single-Item Physical Activity Question for Population Surveillance. *Journal of Physical Activity and Health*. 2022; 19(10):681–6. <https://doi.org/10.1123/jpah.2022-0369> PMID: 36126942
28. Hart PD. Initial Assessment of a Brief Health, Fitness, and Spirituality Survey for Epidemiological Research: A Pilot Study. *J Lifestyle Med*. 2022; 12(3):119–26. <https://doi.org/10.15280/jlm.2022.12.3.119> PMID: 36628178
29. Schechtman KB, Barzilai B, Rost K, B. FJE. Measuring physical activity with a single question. *American journal of public health*. 1991; 81(6):771–3. <https://doi.org/10.2105/ajph.81.6.771> PMID: 2029054
30. Gill DP, Jones GR, Zou G, Speechley M. Using a single question to assess physical activity in older adults: A reliability and validity study. *BMC Medical Research Methodology*. 2012; 12(20). <https://doi.org/10.1186/1471-2288-12-20> PMID: 22373159
31. Macdonald HM, Nettlefold L, Bauman A, Sims-Gould J, McKay HA. Pragmatic Evaluation of Older Adults' Physical Activity in Scale-Up Studies: Is the Single-Item Measure a Reasonable Option? *Journal of Aging and Physical Activity*. 2022; 30(1):25–32. <https://doi.org/10.1123/japa.2020-0412> PMID: 34348228
32. Portegijs E, Sipilä S, Viljanen A, Rantakokko M, Rantanen T. Validity of a single question to assess habitual physical activity of community-dwelling older people. *Scandinavian Journal of Medicine & Science in Sports*. 2017; 27(11):1423–30. <https://doi.org/10.1111/sms.12782> PMID: 27747944
33. Kiernan M, Schoffman DE, Lee K, Brown SD, Fair JM, Perri MG, et al. The stanford leisure-time activity categorical item (L-Cat): A single categorical item sensitive to physical activity changes in overweight/obese women. *International Journal of Obesity*. 2013; 37(12):1597–602. <https://doi.org/10.1038/ijo.2013.36> PMID: 23588625

34. Ross KM, Leahey TM, Kiernan M. Validation of the Stanford Leisure-Time Activity Categorical Item (L-Cat) using armband activity monitor data. *Obesity science & practice*. 2018; 4(3):276–82. <https://doi.org/10.1002/osp4.155> PMID: 29951218
35. Johansson G, Westerterp KR. Assessment of the physical activity level with two questions: Validation with doubly labeled water. *International Journal of Obesity*. 2008; 32(6):1031–3. <https://doi.org/10.1038/ijo.2008.42> PMID: 18392036
36. Blomqvist A, Bäck M, Klompstra L, Strömberg A, Jaarsma T. Utility of single-item questions to assess physical inactivity in patients with chronic heart failure. *ESC Heart Fail*. 2020; 7(4):1467–76. <https://doi.org/10.1002/ehf2.12709> PMID: 32372549
37. Orrell A, Doherty P, Miles J, Lewin R. Development and validation of a very brief questionnaire measure of physical activity in adults with coronary heart disease. *European Journal of Cardiovascular Prevention and Rehabilitation*. 2007; 14(5):615–23. <https://doi.org/10.1097/HJR.0b013e3280ecfd56> PMID: 17925619
38. Cruz J, Jácome C, Oliveira A, Paixão C, Rebelo P, Flora S, et al. Construct validity of the brief physical activity assessment tool for clinical use in COPD. *The Clinical Respiratory Journal*. 2021; 15(5):530–9. <https://doi.org/10.1111/crj.13333> PMID: 33484059
39. Hamilton K, White KM, Cuddihy T. Using a single-item physical activity measure to describe and validate parents' physical activity patterns. *Research Quarterly for Exercise and Sport*. 2012; 83(2):340–5. <https://doi.org/10.1080/02701367.2012.10599865> PMID: 22808720
40. Milton K, Bull FC, Bauman A. Reliability and validity testing of a single-item physical activity measure. *British Journal of Sports Medicine*. 2011; 45(3):203–8. <https://doi.org/10.1136/bjsm.2009.068395> PMID: 20484314
41. Milton K, Clemes S, Bull F. Can a single question provide an accurate measure of physical activity? *British Journal of Sports Medicine*. 2013; 47(1):44–8. <https://doi.org/10.1136/bjsports-2011-090899> PMID: 22522584
42. O'Halloran P, Kingsley M, Nicholson M, Staley K, Randle E, Wright A, et al. Responsiveness of the single item measure to detect change in physical activity. *PLoS ONE*. 2020; 15(6). <https://doi.org/10.1371/journal.pone.0234420> PMID: 32584830
43. Zvolinsky S, McKenna J, Pringle A, Widdop P, Griffiths C. Physical activity assessment for public health: Efficacious use of the single-item measure. *Public Health*. 2015; 129(12):1630–6. <https://doi.org/10.1016/j.puhe.2015.07.015> PMID: 26296848
44. Marshall AL, Smith BJ, Bauman AE, Kaur S. Reliability and validity of a brief physical activity assessment for use by family doctors. *British Journal of Sports Medicine*. 2005; 39(5):294–7. <https://doi.org/10.1136/bjsm.2004.013771> PMID: 15849294
45. Puig Ribera A, Peña Chimenis Ò, Romaguera Bosch M, Duran Bellido E, Heras Tebar A, Solà Gonfau M, et al. How to identify physical inactivity in Primary Care: Validation of the Catalan and Spanish versions of 2 short questionnaires. *Atencion Primaria*. 2012; 44(8):485–93.
46. Puig-Ribera A, Martín-Cantera C, Puigdomenech E, Real J, Romaguera M, Magdalena-Belio JF, et al. Screening physical activity in family practice: Validity of the Spanish version of a brief physical activity questionnaire. *PLoS ONE*. 2015; 10(9). <https://doi.org/10.1371/journal.pone.0136870> PMID: 26379036
47. Smith BJ, Marshall AL, Huang N. Screening for physical activity in family practice: Evaluation of two brief assessment tools. *American journal of preventive medicine*. 2005; 29(4):256–64. <https://doi.org/10.1016/j.amepre.2005.07.005> PMID: 16242587
48. Gionet NJ, Godin G. Self-Reported Exercise Behavior of Employees—A Validity Study. *Journal of Occupational and Environmental Medicine*. 1989; 31(12):969–73.
49. Li S, Carlson E, Holm K. Validation of a single-item measure of usual physical activity. *Perceptual and Motor Skills*. 2000; 91(2):593–602. <https://doi.org/10.2466/pms.2000.91.2.593> PMID: 11065323
50. Graff-Iversen S, Anderssen SA, Holme IM, Jennum AK, Raastad T. Two short questionnaires on leisure-time physical activity compared with serum lipids, anthropometric measurements and aerobic power in a suburban population from Oslo, Norway. *European journal of epidemiology*. 2008; 23(3):167–74. <https://doi.org/10.1007/s10654-007-9214-2> PMID: 18095171
51. Danquah IH, Petersen CB, Skov SS, Tolstrup JS. Validation of the NPAQ-short—A brief questionnaire to monitor physical activity and compliance with the WHO recommendations. *BMC Public Health*. 2018; 18(1). <https://doi.org/10.1186/s12889-018-5538-y> PMID: 29739383
52. Iwai N, Hisamichi S, Hayakawa N, Inaba Y, Nagaoka T, Sugimori H, et al. Validity and reliability of single-item questions about physical activity. *Journal of Epidemiology*. 2001; 11(5):211–8. <https://doi.org/10.2188/jea.11.211> PMID: 11579928

53. Hyvärinen M, Sipilä S, Kulmala J, Hakonen H, Tammelin TH, Kujala UM, et al. Validity and reliability of a single question for leisure-time physical activity assessment in middle-aged women. *Journal of Aging and Physical Activity*. 2020; 28(2):231–41. <https://doi.org/10.1123/japa.2019-0093> PMID: 31585436
54. Ernährung und Bewegung [Nutrition and physical activity]: Ministère de la Santé (Ministerium für Gesundheit); 2016.
55. Forsén L, Loland NW, Vuillemin A, Chinapaw MJ, van Poppel MN, Mokkink LB, et al. Self-administered physical activity questionnaires for the elderly: a systematic review of measurement properties. *Sports Medicine*. 2010; 40(7):601–23. <https://doi.org/10.2165/11531350-000000000-00000> PMID: 20545382
56. Smith TO, McKenna MC, Salter C, Hardeman W, Richardson K, Hillsdon M, et al. A systematic review of the physical activity assessment tools used in primary care. *Family Practice*. 2017; 34(4):384–91. <https://doi.org/10.1093/fampra/cmz011> PMID: 28334801
57. Levin-Zamir D, Leung AYM, Dodson S, Rowlands G. Health Literacy in Selected Populations: Individuals, Families, and Communities from the International and Cultural Perspective. *Studies in Health Technology and Informatics*. 2017; 240:392–414. PMID: 28972530
58. Marsilio M, Fusco F, Gheduzzi E, Guglielmetti C. Co-Production Performance Evaluation in Healthcare. A Systematic Review of Methods, Tools and Metrics. *International Journal of Environmental Research and Public Health*. 2021; 18(7). <https://doi.org/10.3390/ijerph18073336> PMID: 33804862
59. Hankinson A. Association of activity and chronic disease risk factors: utility and limitations of objectively measured physical activity data. *Journal of the American Dietetic Association*. 2008; 108(6):945–7. <https://doi.org/10.1016/j.jada.2008.03.018> PMID: 18502224