Current Evidence of Measurement Properties of Physical Activity Questionnaires for Older Adults: An Updated Systematic Review

Matteo C. Sattler1, Johannes Jaunig1, Christoph Tösch1, Estelle D. Watson2, Lidwine B. Mokkink3, Pavel Dietz4, Mireille N. M. van Poppel1,5

Abstract

Background Questionnaires provide valuable information about physical activity (PA) behaviors in older adults. Until now, no firm recommendations for the most qualified questionnaires for older adults have been provided.

Objectives This review is an update of a previous systematic review, published in 2010, and aims to summarize, appraise and compare the measurement properties of all available self-administered questionnaires assessing PA in older adults.

Methods We included the articles evaluated in the previous review and conducted a new search in PubMed, Embase, and SPORTDiscus from September 2008 to December 2019, using the following inclusion criteria (1) the purpose of the study was to evaluate at least one measurement property (reliability, measurement error, hypothesis testing for construct validity, responsiveness) of a self-administered questionnaire; (2) the questionnaire intended to measure PA; (3) the questionnaire covered at least one domain of PA; (4) the study was performed in the general, healthy population of older adults; (5) the mean age of the study population was > 55 years; and (6) the article was published in English. Based on the Quality Assessment of Physical Activity Questionnaires (QAPAQ) checklist, we evaluated the quality and results of the studies. The content validity of all included questionnaires was also evaluated using the reviewers’ rating. The quality of the body of evidence was evaluated for the overall construct of each questionnaire (e.g., total PA), moderate-to-vigorous physical activity (MVPA) and walking using a modified Grading of Recommendation, Assessment, Development, and Evaluation (GRADE) approach.

Results In total, 56 articles on 40 different questionnaires (14 from the previous review and 26 from the update) were included. Reliability was assessed for 22, measurement error for four and hypotheses testing for construct validity for 38 different questionnaires. Evidence for responsiveness was available for one questionnaire. For many questionnaires, only one measurement property was assessed in only a single study. Sufficient content validity was considered for 22 questionnaires. All questionnaires displayed large measurement errors. Only versions of two questionnaires showed both sufficient reliability and hypotheses testing for construct validity, namely the Physical Activity Scale for the Elderly (PASE; English version, Turkish version) for the assessment of total PA, and the Physical Activity and Sedentary Behavior Questionnaire (PASB-Q; English version) for the assessment of MVPA. The quality of evidence for these results ranged from very low to high.

Conclusions Until more high-quality evidence is available, we recommend the PASE for measuring total PA and the PASB-Q for measuring MVPA in older adults. However, they are not equally qualified among different languages. Future studies on the most promising questionnaires should cover all relevant measurement properties. We recommend using and improving existing PA questionnaires—instead of developing new ones—and considering the strengths and weaknesses of each PA measurement instrument for a particular purpose.

1 Introduction

The aging of the world’s population represents one of the key challenges over the next decades. Both life expectancy and the proportion of older adults are increasing [1] and, therefore, promoting and maintaining quality of life at an older age is essential. Current evidence shows that physical activity (PA) can increase health in later life [2] through...
Key Points

Based on low-to-moderate-quality evidence of both sufficient reliability and hypotheses testing for construct validity, we recommend using the Physical Activity Scale for the Elderly (PASE—English version) for the assessment of total PA and the Physical Activity and Sedentary Behavior Questionnaire (PASB-Q—English version) for the assessment of MVPA.

To ensure high quality of and comparability across studies, we recommend using and improving existing questionnaires, rather than developing new versions, as well as evaluating strengths and weaknesses of each PA measurement instrument with respect to the study purpose.

We recommend performing high-quality studies on the most promising questionnaires, including an assessment of content validity and responsiveness, and the use of standards for study design and evaluation (e.g., Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklists).

Increasing quality of life [3, 4], cognitive and physical functioning [5, 6] and decreasing the risks for neurodegenerative diseases (e.g., Alzheimer’s disease, vascular dementia) [7], depressive symptoms [8, 9] and all-cause mortality [10].

Several instruments are available to measure PA in older adults such as questionnaires, diaries, accelerometers and pedometers. Although several aspects (e.g., strengths, weaknesses and practical considerations) have to be considered when selecting an instrument [11], questionnaires appear to be popular for the measurement of PA in older adults [12]. In contrast to accelerometers, they are usually feasible in large epidemiological studies and well accepted by participants. For example, questionnaires are used in large national surveys to determine and compare PA levels among different countries [13]. The use of the same measurement method in these surveys facilitates comparability among PA estimates [14]. Furthermore, in addition to the total volume of PA, questionnaires can provide valuable information about different domains (e.g., home, leisure time) and types (e.g., walking, resistance training) of activities [15]. Finally, questionnaires can be used as a screening tool to determine PA levels of individuals in healthcare settings. The assessment can be integrated into the clinical workflow and linked to electronic record systems, whereas the obtained results can be used for counseling and PA promotion [16, 17].

Both researchers and healthcare professionals should use instruments with high measurement quality. The quality of an instrument is determined by evaluating its measurement properties such as reliability, validity and responsiveness. Sufficient measurement properties are indispensable to trust the results of studies on the efficacy of PA interventions, health benefits of PA, dose–response relationships as well as trends of PA over time. However, many PA questionnaires and modified versions of these have been developed. The great number of available questionnaires makes it difficult to choose the instrument with the best measurement properties. Moreover, the use of different questionnaires decreases the comparability of PA estimates and its relationship with health outcomes across studies and countries. To limit methodological biases and to draw study conclusions with the highest quality, it is important to select the questionnaire with the best measurement properties for a particular purpose.

Already in 2000, Sallis and Saelens [15] recognized a profusion of PA questionnaires and suggested to select only a few, most qualified ones for future studies. Existing reviews on measurement properties of PA self-reports [18–28] usually focused on the adult population or a specific population of older adults (e.g., older adults with dementia). However, although research on PA in older adults has grown continuously [2], no firm recommendations for the most-qualified self-administered PA questionnaires for older adults have been provided.

In 2010, a series of systematic reviews on measurement properties of PA questionnaires in youth [29], adults [30] and older adults [28] were published. Regarding older adults, we concluded that the evidence for measurement properties of PA questionnaires is scarce and future high-quality validation studies are needed. Specifically, the reliability of the Physical Activity Scale for the Elderly (PASE) was rated as sufficient but the results for validity were inconsistent. Recently, the review for youth was updated [19] and a new one for pregnancy was published [18]. The present review is an update for older adults and aims to summarize, compare and appraise the measurement properties (i.e., reliability, measurement error, hypotheses testing for construct validity, responsiveness) of all available self-administered PA questionnaires in older adults aged > 55 years. In addition, we evaluated the content validity of all included questionnaires and aimed to provide recommendations for choosing the best available PA questionnaires in older adults.

2 Methods

For reporting, we followed the Preferred Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [31]. A definition of all quoted measurement properties is provided in Table 1.
2.1 Literature Search

We performed systematic literature searches in the databases PubMed, SPORTDiscus and Embase (using the filter ‘Embase only’). The search strategy involved (variations of) the terms ‘physical activity’, ‘questionnaire’ and ‘measurement properties’ [32] (see Electronic Supplementary Material Appendix S1). We excluded publication types such as case reports, interviews or biographies and adapted our search for Embase and SPORTDiscus following their guidelines. In 2010 [28], we included all publications until May 2009 in the initial title/abstract search. For this update, to avoid any losses of publications, we considered all results from September 2008 to 17 December 2018 (day of search) as potentially relevant. The search was updated on 3 December 2019.

2.2 Eligibility Criteria

The following eligibility criteria were defined [18, 28, 33]:

1. The purpose of the study was to evaluate at least one of the following measurement properties of a self-administered questionnaire: reliability, measurement error, hypotheses testing for construct validity or responsiveness. Because no gold standard exists to measure PA [25, 34], results from studies referring to the criterion validity of a questionnaire were considered as evidence for hypotheses testing for construct validity.
2. The purpose of the questionnaire was to assess PA, which was defined as any bodily movement produced by skeletal muscles which results in energy expenditure (EE; p. 126) [35].
3. The questionnaire should cover at least one domain of PA (household, occupation, recreation, sports or transport [cycling and/or walking]).
4. The study was performed in the general population of older adults (i.e., healthy older adults), regardless of the population for which the questionnaire was developed (e.g., general population, patients with cardiovascular disease).
5. The mean or median age of the study population was > 55 years.
6. The article was published in English.

Consistent with our previous review [18], we did not evaluate measurement properties regarding the internal structure of the questionnaire (structural validity, internal consistency)

| Table 1 | Definition of measurement properties for PA questionnaires, adapted from the COSMIN methodology [135] (p. 743) |
| --- | --- |
| Domain | Measurement property | Aspect | Definition |
| Reliability | Internal consistency | The degree to which the measurement is free from measurement error |
| Reliability | Reliability | The degree of the interrelatedness among the items |
| Measurement error | Measurement error | The proportion of the total variance in the measurements which is because of true differences among participants |
| Validity | Content validity | The degree to which the content of an instrument is an adequate reflection of the construct |
| Face validity | Face validity | The degree to which the items of an instrument indeed look as though they are an adequate reflection of the construct |
| Construct validity | Construct validity | The degree to which the scores of an instrument are consistent with hypotheses (for example with respect to internal relationships, relationships to scores of other instruments) based on the assumption that the instrument validly measures the construct |
| Structural validity | Structural validity | The degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct |
| Hypotheses testing | Hypotheses testing | Idem construct validity |
| Cross-cultural validity | Cross-cultural validity | The degree to which the performance of the items on a translated or culturally adapted instrument are an adequate reflection of the performance of the items of the original version of the instrument |
| Criterion validity | Criterion validity | The degree to which the scores of an instrument are an adequate reflection of a gold standard |
| Responsiveness | Responsiveness | The ability of an instrument to detect change over time in the construct |

COSMIN COnsensus-based Standards for the selection of health Measurement INstruments, PA physical activity
(e.g., using Cronbach’s alpha), cross-cultural validity). Internal structure is only relevant for questionnaires based on a reflective model assuming items to be correlated [33]. This is not the case for PA questionnaires (e.g., time spent in walking does not necessarily have to correlate with time spent in other behaviors) [36]. In addition, we did not perform an exhaustive evaluation of content validity but rather applied a subjective rating to assess the content validity of all included questionnaires [33]. A detailed evaluation of content validity may be performed in future reviews and would require the inclusion of all studies focusing on any aspect of content validity (e.g., studies on the development of the questionnaire, pilot tests among older adults, expert opinions).

Finally, the following exclusion criteria were applied:

1. Questionnaires measuring physical functioning or sweating, diaries, interviews (face-to-face, telephone), and interviewer-administered questionnaires. However, we did include self-administered PA questionnaires where some participants had received help with the completion.
2. Questionnaires assessing specific behaviors within one domain of PA (e.g., commuting to work).
3. Studies performed solely in patients or in a priori defined subpopulations (e.g., stroke patients, obese older adults).
4. Studies assessing the agreement between a PA questionnaire and a non-PA measure such as body mass index (BMI), health functioning, performance, fitness, well-being or cardiovascular risk factors. This was done because we found it difficult to define specific cut points for sufficient measurement properties.

### 2.3 Selection of Articles and Data Extraction

Two researchers independently screened titles and abstracts for eligible studies. MCS and either CT or JJ inspected full-text articles, performed data extraction, result rating and quality assessment. Disagreements were discussed during consensus meetings. If no agreement could be reached, a third researcher (LBM, MVP) was consulted. Consistent with our previous reviews [18, 28], we extracted all relevant information using a standardized form. This form was based on the Quality Assessment of Physical Activity Questionnaire (QAPAQ) checklist [36]. We included the results for the overall construct of PA [i.e., total PA, total physical activity energy expenditure (PAEE)] and for any subdimension (e.g., leisure time physical activity (LTPA), moderate-to-vigorous physical activity (MVPA), walking) in our tables for which information about at least one measurement property was available. It is important note that, depending on the purpose of the questionnaire (overall construct), the total score of the questionnaire can either represent total PA, total PAEE or a specific subdimension of PA. For example, a questionnaire may aim in assessing LTPA and, hence, the total score of the questionnaire does not necessarily represent total PA.

### 2.4 Assessment of Measurement Properties

Each result on a measurement property was either rated as sufficient (+) or insufficient (−). Our criteria for sufficient measurement properties were based on the QAPAQ checklist [36] and have been described previously [18, 28, 30]. However, a short description will be provided herein. The content validity of all included questionnaires was assessed following the reviewers’ ratings on three principal criteria [18, 30]: (1) If the questionnaire measures total PA (or MVPA), it should at least include the domains of household, recreation, sports and transport. Regarding transport, at least walking should be included since it represents one of the most common activities in older adults [37]. Occupational PA was considered as optional for older adults; (2) the questionnaire should assess at least the parameters frequency and duration of PA (e.g., to further define dose–response patterns between PA and health [38]); and (3) the recall period should be at least one week (if not assessing daily PA).

We included results for reliability [intraclass correlation coefficient (ICC), concordance, kappa, Pearson/Spearman correlation] and measurement error [coefficient of variation (CV), standard error of measurement (SEM), smallest detectable change (SDC), change in the mean or mean difference (d; systematic error), limits of agreement (LOA; random error)]. Previous research has shown that already low doses of PA (e.g., < 150 min of MVPA, 1–2 times running per week) were associated with substantial health benefits in older adults such as reductions in all-cause mortality [10, 39]. Therefore, we defined a change in the frequency of two times per week and a change in MVPA of 30 min [≥ 90 metabolic equivalent (MET) minutes] per week as clinically important [18]. These values represent a minimal important change (MIC) and were used to evaluate measurement error. If the LOA or SDC are smaller than the MIC, changes as large as the MIC represent true changes beyond measurement error. In other words, a PA questionnaire should be able to measure changes of ± 20% of current PA guidelines [2].

A result for reliability was sufficient if ICC/kappa/concordance was ≥ 0.70 or Pearson/Spearman ≥ 0.80 and a result for measurement error if MIC (e.g., 30 min of MVPA per week) > LOA/SDC or CV ≤ 15%. Otherwise, the result was insufficient. Cut points for sufficient hypotheses testing for construct validity are shown in Table 2 [18, 36]. We used the same set of hypotheses to appraise responsiveness which, in this case, concern a change score of PA [40, 41].
2.5 Quality of Individual Studies

The standards for the assessment of the quality of each study were based on the QAPAQ checklist [36] and were described in our previous reviews [18, 28–30]. Briefly, if the study did not show any substantial flaws in the design or analysis (4: inadequate quality), we assigned one of the three different levels of quality (1: very good, 2: adequate, 3: doubtful) for each construct/subdimension of the questionnaire (e.g., total PA or MVPA) and measurement property (i.e., reliability, measurement error, hypotheses testing for construct validity, and responsiveness).

Reliability and measurement error are usually assessed by repeated measurements in stable participants. To guarantee that the behavior was sufficiently stable over this period [42], we defined an adequate time interval between test and retest as follows: > 1 day and ≤ 3 months for questionnaires recalling a usual week/month; > 1 day and ≤ 2 weeks for questionnaires recalling the previous week; > 1 day and ≤ 1 week for questionnaires recalling the previous day; > 1 day and ≤ 1 year for questionnaires recalling the previous year or assessing lifetime PA. Thus, the following levels of quality for studies on reliability and measurement error were applied:

1. Very good (1): reporting of ICC, LOA, SDC, SEM, CV, kappa or concordance and an adequate time interval between test and retest.
2. Adequate (2): reporting of ICC, LOA, SDC, SEM, CV, kappa or concordance and an inadequate time interval between test and retest; or reporting of Pearson/Spearman correlation and an adequate time interval between test and retest.
3. Doubtful (3): reporting of Pearson/Spearman correlation and an inadequate time interval between test and retest.

Regarding hypotheses testing for construct validity and responsiveness, higher quality was considered with increasing degree of comparability between the measured construct/subdimension and other PA measures (Table 2). For example, the quality was higher for comparisons with accelerometers compared to diaries or other questionnaires.

2.6 Inclusion of the Evidence from the Previous Review

All studies from the previous review [28] were included in this update. Compared to the previous review, the following changes were made within this update: (1) all results were rated irrespective of the sample size. The sample size was considered in the assessment of the quality of the body of evidence; (2) results for measurement error were rated; (3) results based on comparisons with non-PA measures such as health or performance associations were not included; (4) we did not evaluate group differences based on significance levels and instead, only evaluated the magnitude of the effect (e.g., correlation coefficients) [36]; and (5) we used updated levels of quality, as described earlier [18] [e.g., sports/exercise was included in the list, PAEE was distinguished from PA (e.g., as behavior typically measured using raw units such as minutes)]. Due to these differences, two researchers independently (MCS, JJ) reassessed all studies included in the previous review.

2.7 Quality of the Body of Evidence

Based on all studies included from the new and previous review, the quality of evidence was evaluated for the overall construct of each questionnaire (e.g., total PA, total PAEE, total LTPA), also called the ‘total’ score, as well as for the subdimensions MVPA and walking. This was done using the Grading of Recommendation, Assessment, Development, and Evaluation (GRADE) approach [43]. Specifically, we applied a modified approach, as recommended (and described) in the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guideline [33], and assessed the evidence for each measurement property (reliability, measurement error, hypotheses testing for construct validity, and responsiveness) and questionnaire separately. Where applicable, the results from multiple studies on the same questionnaire were summarized. Although different language versions should be treated separately, one may consider summarizing the results if the results have been consistent [33]. Thus, we also assessed the quality of evidence based on the summarized results across multiple studies on different language versions of the same questionnaire.

The grading procedure was described previously [18, 33]. Briefly, the quality of evidence could be high, moderate, low or very low depending on the assessment of four factors (risk of bias (methodological quality of the study), inconsistency in results, indirectness, imprecision). Due to serious flaws in one or more of these factors, the quality of evidence could be downgraded by up to three levels (serious, very serious, extremely serious). For example, serious risk of bias and serious indirectness would result in low-quality evidence (downgraded by two levels).

The assessment of risk of bias was based on the quality ratings of each study (see Sect. 2.5). We considered risk of bias as serious when there were multiple studies of doubtful quality or only one study of adequate quality available, and as very serious when there were multiple studies of inadequate quality or only one study of doubtful quality. We considered downgrading by three levels (extremely serious), if there was only one study of inadequate quality available. Due to inconsistency in results among multiple studies (e.g.,
Table 2: Cut points for sufficient correlations per construct and dimension of PA measured by the questionnaire, and level of quality

| Construct/dimension | 1: Very good | 2: Adequate | 3: Doubtful |
|---------------------|--------------|-------------|-------------|
| Total PAEE (MET/kcal) | Doubly labeled water \( \geq 0.70 \) | Accelerometer total counts or average counts \( \geq 0.50 \) | Diary, logbook, other questionnaire, interview \( \geq 0.70 \); pedometer steps \( \geq 0.40 \); accelerometer time in moderate, moderate-to-vigorous or vigorous intensity \( \geq 0.40 \) |
| Total PA (min/score) | Accelerometer total counts or average counts \( \geq 0.50 \) | Accelerometer time in moderate-to-vigorous intensity \( \geq 0.40 \) | Diary, logbook, other questionnaire, interview \( \geq 0.70 \); pedometer steps \( \geq 0.40 \) |
| By intensity | | | |
| Vigorous | Accelerometer time in vigorous intensity \( \geq 0.50 \) | Accelerometer total counts or average counts \( \geq 0.40 \) | Diary, logbook, other questionnaire, interview \( \geq 0.70 \); accelerometer time in light, moderate or moderate-to-vigorous intensity \( \geq 0.40 \); pedometer steps \( \geq 0.40 \) |
| Moderate-to-vigorous | Accelerometer time in moderate-to-vigorous intensity \( \geq 0.50 \) | Accelerometer total counts or average counts \( \geq 0.40 \) | Diary, logbook, other questionnaire, interview \( \geq 0.70 \); accelerometer time in light, moderate or vigorous intensity \( \geq 0.40 \); pedometer steps \( \geq 0.40 \) |
| Moderate | Accelerometer time in moderate intensity \( \geq 0.50 \) | Accelerometer total counts or average counts \( \geq 0.40 \) | Diary, logbook, other questionnaire, interview \( \geq 0.70 \); accelerometer time in light, moderate or vigorous intensity \( \geq 0.40 \); pedometer steps \( \geq 0.40 \) |
| Light | Accelerometer time in light intensity \( \geq 0.50 \) | Accelerometer total counts or average counts \( \geq 0.40 \) | Diary, logbook, other questionnaire, interview \( \geq 0.70 \); accelerometer time in light, moderate or vigorous intensity \( \geq 0.40 \); pedometer steps \( \geq 0.40 \) |
| By type | | | |
| Walking | Pedometer or accelerometer walking total counts \( \geq 0.70 \) | Accelerometer total counts or average counts \( \geq 0.40 \) | Diary, logbook, other questionnaire, interview \( \geq 0.70 \); accelerometer time in moderate, moderate-to-vigorous or vigorous intensity \( \geq 0.40 \) |
| Leisure time | Accelerometer total counts or average counts in leisure time \( \geq 0.50 \) | Accelerometer total counts or average counts \( \geq 0.40 \) | Diary, logbook, other questionnaire, interview \( \geq 0.70 \); pedometer steps \( \geq 0.40 \); accelerometer time in moderate, moderate-to-vigorous or vigorous intensity \( \geq 0.40 \) |
| Occupational | Direct observational method \( \geq 0.60 \); accelerometer total counts or average counts during working hours \( \geq 0.50 \) | Accelerometer total counts or average counts \( \geq 0.40 \) | Diary, logbook, other questionnaire, interview \( \geq 0.70 \); accelerometer time in light, moderate, moderate-to-vigorous or vigorous intensity \( \geq 0.40 \); pedometer steps \( \geq 0.40 \) |
| Household/caregiving | Accelerometer time in light, light-to-moderate or moderate intensity \( \geq 0.50 \) | Accelerometer total counts or average counts \( \geq 0.40 \) | Diary, logbook, other questionnaire, interview \( \geq 0.70 \); accelerometer time in moderate-to-vigorous or vigorous intensity \( \geq 0.40 \); pedometer steps \( \geq 0.40 \) |
| Sports/exercise | Accelerometer time in moderate-to-vigorous or vigorous intensity \( \geq 0.50 \) | Accelerometer total counts or average counts \( \geq 0.40 \) | Diary, logbook, other questionnaire, interview \( \geq 0.70 \); accelerometer time in light or moderate intensity \( \geq 0.40 \); pedometer steps \( \geq 0.40 \) |

Kcal kilocalories, MET metabolic equivalent, min minutes, PA physical activity, PAEE physical activity energy expenditure
some have been sufficient but others insufficient), downgrading by one or two levels was considered. If this inconsistency could be explained, for instance by differences in the study methods (e.g., different subpopulations) or handling of questionnaire data (e.g., score calculation), the results from these studies were not summarized, and the evidence was provided separately. With respect to the purpose of this review (e.g., eligibility criteria), differences in populations and questionnaire scores were evaluated and if applicable, downgrading by one or two levels because of serious or very serious indirectness was considered. For example, one may consider serious indirectness if a study included only male older adults. Finally, imprecision was assessed using the previously determined optimal information sizes for reliability and hypotheses testing for construct validity [18]. If the total sample size did not meet the criteria, we downgraded the evidence by one (serious imprecision, reliability and measurement error: \( n < 45 \); hypotheses testing for construct validity and responsiveness: \( n < 123 \)) or two (very serious imprecision, reliability and measurement error: \( n < 12 \); hypotheses testing for construct validity and responsiveness: \( n < 32 \)) levels. Based on the quality of evidence (high, moderate, low, very low) and overall result of the measurement properties (sufficient, insufficient), recommendations for the most-qualified questionnaires were given.

### 3 Results

#### 3.1 Literature Search

The update resulted in 29,831 hits (Fig. 1). Based on titles and abstracts, 61 articles were selected, of which 23 were excluded after reading the full texts. Consequently, 38 articles [44–81] were included in the update. A summary of all included studies, questionnaires and evaluated measurement properties of this update is provided in Table 3.

In the previous review from 2010 [28], 18 articles [82–99] on versions of 13 different questionnaires were included. However, during the reference check of our update, we found two articles [75, 76] which were not included in the previous review. These articles fulfilled all our inclusion criteria, have been published before September 2008, and, thus, were now included. Results from studies reported in these two articles were shown together with those from previously included studies in order to allow comparisons. An overview of all previously included studies (including the latter two articles) is provided in Electronic Supplementary Material Table S1. In contrast to 2010, we considered the Cambridge Index as a stand-alone instrument which means that we reassessed 14 (instead of 13) different questionnaires. Six questionnaires [Cambridge Index, Community...
Health Activities Model Program for Senior (CHAMPS), International Physical Activity Questionnaire—short-form (IPAQ-SF), PASE, Stanford Brief Activity Survey (SBAS), Women’s Health Initiative Physical Activity Questionnaire (WHI-PAQ)] were assessed in studies included both in the update and previous review.

Previous review and update combined, we included studies on measurement properties of versions of 40 different questionnaires (14 from the previous review and 26 from the update) derived from 56 articles. Information about reliability was available for versions of 22, measurement error for four, and hypotheses testing for construct validity for 38 different questionnaires. Results for responsiveness were available for one questionnaire. Regarding the latter measurement property, one study [100] from the update was excluded after reading the full text because the reported results for responsiveness could not be evaluated with respect to our set of hypotheses. Likewise, another study [82] from the previous review evaluated the sensitivity to change of the CHAMPS but did not use a PA comparison measure or test hypotheses about expected effect sizes.

Three studies [49, 65, 83] considered doubly labeled water (DLW) as a comparison method, whereas most often accelerometers, pedometers and other PA questionnaires were used. Both original and modified versions were assessed. For example, two studies modified the CHAMPS by replacing questions and adjusting MET values [59] or changing the recall period to the past 7 days (instead of past 4 weeks) and using modified response categories [84]. Some studies evaluated measurement properties of new indices [e.g., Cambridge Index derived from the questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC)].

Finally, although all studies evaluated a ‘PA questionnaire’, two studies questionednaires intending to measure the construct total EE (i.e., Questionnaire d’Activité Physique Saint-Etienne (QAPSE) [85], Questionnaire preceding EPIC (Pre-EPIC) [86]) and one study presented multiple results concerning both total EE and PA (i.e., Flemish Physical Activity Computerized Questionnaire (FPACQ) [87]). The construct total EE is different from PA, since it also includes a detailed assessment of all activities summing up to 24 h (e.g., rest, sleep, eating). Whenever reported, results for total EE were not evaluated but included in the tables to allow the reader to interpret the results.

### 3.2 Description of Questionnaires

A detailed description of all questionnaires included in the update is provided in the Electronic Supplementary Material Table S2 whereas a description of previously included questionnaires was provided in 2010 [28]. The populations for which the questionnaires were developed varied (e.g., older adults, female adults). Most questionnaires intend to measure total PA, total PAEE, MVPA or domain-specific PA such as LTPA. Some questionnaires [e.g., Web-based Physical Activity Questionnaire Active-Q (Active-Q)] measure frequency and duration of activities but not the relative intensity in which these activities were performed (i.e., subjective rating of the participants). Although intensity may not be measured in this way, usually absolute MET values were assigned to activities to obtain time spent in different intensity levels (e.g., light, moderate, vigorous). Finally, sometimes information about parameters of PA (frequency, duration, intensity) is only obtained for some but not all listed activities [e.g., Arizona Activity Frequency Questionnaire (AAFQ)].

### 3.3 Assessment of Measurement Properties

#### 3.3.1 Content Validity

Based on our three criteria, the content validity was sufficient for 22 questionnaires [AAFQ, Active Australia Survey (AAS), Aerobic Center Longitudinal Study—Physical Activity Long Survey (ACLS-PALS), Active-Q, CHAMPS, EPIC-Norfolk Physical Activity Questionnaire (EPAQ2), FPACQ, International Physical Activity Questionnaire for the Elderly (IPAQ-E), International Physical Activity Questionnaire—long form (IPAQ-LF), IPAQ-SF, Modified Leisure Time Physical Activity Questionnaire (mLTPA-Q), Modified version of the Minnesota Leisure Time Physical Activity Questionnaire (Modified Minnesota LTPA-Q), Older Adult Exercise Status Inventory (OA-ESI), PASE, Physical Activity and Sedentary Behavior Questionnaire (PASB-Q), Physical Activity Questionnaire for Elderly Japanese (PAQ-EJ), Physical Activity Vital Sign Questionnaire (PAVS), Physical Activity Questionnaire for the Elderly (QAPPA), Pre-EPIC, Two questions asking about time spent in Moderate-to-vigorous Physical Activities (MVPA questions), Walking question, Zutphen Physical Activity Questionnaire (ZPAQ)].

It should be noted that the content validity of the original version of the ZPAQ was insufficient due to the lack of household-related activities [101]. However, the content validity of the modified version of the ZPAQ was sufficient because the authors included the missing domain [57].

#### 3.3.2 Reliability and Measurement Error

Table 4 summarizes the results for reliability and measurement error of studies included in the update. The results of the reassessment of all studies included in the previous review are shown in Electronic Supplementary Material Table S3. The quality of studies was usually very good or adequate. Versions of the CHAMPS (English version,
| Abbreviation | Full name of questionnaire | Studies on measurement properties | Assessed measurement properties | Comparison measures | Sample N (of consented), n (women), age (years), BMI (kg/m²), specific characteristics, nationality |
|--------------|---------------------------|----------------------------------|-------------------------------|---------------------|---------------------------------------------------------------------------------|
| AAFQ         | Arizona Activity Frequency Questionnaire | Neuhouser et al. [65] English version | Reliability, Measurement error, Hypotheses testing for construct validity | DLW                 | 450 (of 450), all ♂, age ≥ 60, 31.8% BMI (18.5–24.9), USA                           |
| AAS          | Active Australia Survey   | Vandelanotte et al. [77] English version |                                | Acc                 | 342 (N/A); 207 ♂, age n (%): 50–64 years = 200 (58.8%), age n (%): ≥ 65 years = 142 (41.5%), Australia |
|              |                           | Freene et al. [55] English version |                                |                     | First group: 39 (of 56), 29 ♂, mean age = 56.7 (SD= 4.7), mean BMI = 26.9 (SD = 5.1), Australia |
|              |                           | Heesch et al. [58] English version |                                | Ped                 | 53 (N/A), 26 ♂, mean age = 72.6 (SD = 5.9), 37.7% BMI (18.5–24.9), 49.1% BMI (25–29.9), 13.2% BMI (≥ 30), Australia |
| ACLS-PALS    | Aerobic Center Longitudinal Study—Physical Activity Long Survey | Banda et al. [46] English version |                                | Acc                 | 71 (of 80), 49 ♂, mean age = 57.4 (SD = 9.9), mean BMI = 27.9 (SD = 4.9), 70.5% overweight/obese, 74.6% Caucasian, USA |
| ACLS-PASS    | Aerobic Center Longitudinal Study—Physical Activity Short Survey | Banda et al. [46] English version |                                | Acc                 | 71 (of 80), 49 ♂, mean age = 57.4 (SD = 9.9), mean BMI = 27.9 (SD = 4.9), 70.5% overweight/obese, USA |
| Active-Q     | Web-based Physical Activity Questionnaire Active-Q | Bonn et al. [48] Swedish version |                                | Acc                 | 148 (of 167), all ♂, mean age = 65.4 (SD = 8.7), mean BMI = 25.7 (SD = 2.9), Sweden |
| BRHS         | British Regional Heart Study Physical Activity Questionnaire | Jefferis et al. [62] English version |                                | Acc                 | 1377 (of 1655), all ♂, mean age = 78.5 (SD = 4.6), mean BMI = 27.1 (SD = 3.8), UK |
| Cambridge Index | Simple Physical Activity Index of the European Prospective Investigation into Cancer (EPIC) study | España-Romero et al. [53] English version |                                | Acc+HR              | 1689 (of 1829), 876 ♂, age (range) = 60–64, mean BMI ♂ = 27.9 (SD = 5.3), mean BMI ♂ = 27.8 (SD = 4.2), 32.5% normal-weight ♂, 27.3% normal-weight ♂, UK |
| Abbreviation | Full name of questionnaire | Studies on measurement properties | Assessed measurement properties | Comparison measures | Sample |
|--------------|---------------------------|-----------------------------------|--------------------------------|-------------------|--------|
| CHAMPS       | Community Health Activities Model Program for Seniors | Colbert et al. [49] English version | ●  ●  ●  ● | Acc, DLW | 56 (of 70), 79% ♀, mean age = 74.7 (SD = 6.5), mean BMI = 25.8 (SD = 4.2), n (arthritis) = 50, USA |
|              |                           | Hekler et al. [59] Modified English version | ●  ●  ●  ● | Acc | 870 (25% of 3911), 493 ♀, age ≥ 66, 29.4–56.0% BMI (25–29.9), 13.0–25.2% BMI (≥ 30), USA |
| EPAQ2        | EPIC-Norfolk Physical Activity Questionnaire (based on the EPAQ) | España-Romero et al. [53] Modified English version | ●  ●  ●  ● | Acc+HR | 1689 (of 1829), 876 ♀, age (range) = 60–64, mean BMI ♀ = 27.9 (SD = 5.3), mean BMI ♂ = 27.8 (SD = 4.2), 32.5% normal-weight ♀, 27.3% normal-weight ♂, UK |
| GPPAQ        | General Practice Physical Activity Questionnaire (based on the Cambridge Index) | Ahmad et al. [44] English version | ●  ●  ●  ● | Acc | 298 (N/A), 160 ♀, age (range) = 60–74, 67% overweight or obese, adults within primary health care, UK |
| IPAQ-E       | International Physical Activity Questionnaire for the Elderly (based on the IPAQ-SF) | Hurtig-Wennlöf et al. [60] Swedish version | ●  ●  ●  ● | Acc | 54 (of 70), 31 ♀, median age ♀ = 74 (IQR = 69–77), median age ♂ = 71 (IQR = 68–76), Sweden |
| IPAQ-LF      | International Physical Activity Questionnaire—long-form | Cleland et al. [78] English version | ●  ●  ●  ● | Acc | 226 (of 253), 97 ♀, mean age = 71.8 (SD = 6.6), 81.9% retired, Northern Ireland |
|              |                           | Winkers et al. [74] Modified Dutch version | ●  ●  ●  ● | Acc | 196 (of 202), 111 ♀, mean age = 57.1 (SD = 15.4), BMI = 24.8 (SD = 4.2), The Netherlands |
|              |                           | Milanović et al. [64] Serbian version | ●  ●  ●  ● | Acc | 660 (of 700), 308 ♀, mean age = 67.7 (SD = 5.8, mean BMI = 25.9 (SD = 3.7, Serbia |
| IPAQ-SF      | International Physical Activity Questionnaire—short-form | Grimm et al. [56] English version | ●  ●  ●  ● | Acc | 127 (N/A), 96 ♀, mean age = 63.9 (SD = 7.7, mean BMI = 28.3 (SD = 5.8, USA |
| Abbreviation | Full name of questionnaire | Studies on measurement properties | Assessed measurement properties | Comparison measures | Sample |
|--------------|---------------------------|----------------------------------|---------------------------------|---------------------|--------|
| Tomioka et al. [72] Japanese version | | | Reliability | Measurement error | Hypotheses testing for construct validity | Responsiveness | Acc | 325 (of 349), 161 ♀, median age ♀ (young old) = 70, median age ♂ (young old) = 65–74, age range, young old = 65–74, median age ♀ (old old) = 77, median age ♂ (old old) = 78, age range, old old = 75–89, 4.8–18.2% BMI (> 25), Japan |
| Colpani et al. [50] Portuguese version | | | | | | | Ped | 292 (of 301), all ♀, mean age = 57.1 (SD = 5.4), mean BMI = 28.3 (SD = 7.0), Brazil |
| IPEQ | Incidental and Planned Exercise Questionnaire | Delbaere et al. [51] English version | | | | | | 500 (N/A), 279 ♀, mean age = 77.4 (SD = 6.08), Australia |
| LAPAQ | Longitudinal Aging Study Amsterdam Physical Activity Questionnaire | Koolhaas et al. [63] Dutch version | | | | | | 1410 (of 3156), 742 ♀, mean age = 73.8 (SD = 7.6), 70.4% BMI (overweight/obese), The Netherlands |
| Siebeling et al. [69] Dutch version | | | | | | | | 89 (of 92), 46 ♀, median age = 72.4, age range = 65.4–87.6, median BMI = 25.0 BMI (range) = 17.0–35.7, The Netherlands |
| mLTPA-Q | Modified Leisure Time Physical Activity Questionnaire | Fowles et al. [54] English version | | | | | | 32 (of 35), 26 ♀, mean age ♀ = 55 (SD = 10), mean age ♂ = 63 (SD = 9), mean BMI ♀ = 31 (SD = 6), mean BMI ♂ = 26 (SD = 3), Canada |
| Modified Minnesota LTPA-Q | Modified version of the Minnesota Leisure Time Physical Activity Questionnaire | Sabia et al. [67] English version | | | | | | 3975 (of 4492), 26% ♀, age ≥ 60, UK |
| MVPA questions | Two questions asking about time spent in Moderate-to-vigorous Physical Activities | Ekblom et al. [35] Swedish version | | | | | | 948 (of 1111), 486 ♀, median age ♀ = 57.5 (IQR = 53.7–61.4), median age ♂ = 57.7 (IQR = 53.8–62.0), Sweden |
| NC85+PAQ | Newcastle 85 + Study Physical Activity Questionnaire | Innerd et al. [61] English version | | | | | | 484 (N/A), 308 ♀, age range = 87–89, 43% BMI (18.5–24.9), UK |
| Abbreviation | Full name of questionnaire | Studies on measurement properties | Assessed measurement properties | Comparison measures | Sample | N (of consented), n (women), age (years), BMI (kg/m²), specific characteristics, nationality |
|--------------|-----------------------------|-----------------------------------|---------------------------------|---------------------|--------|--------------------------------------------------------------------------------------------|
| NPAQ         | Neighborhood Physical Activity Questionnaire | Bödeker et al. [47] German version | ● | Ped | 58 (of 132), 70.7%♀, age ≥ 60, Germany |
| PASB-Q       | Physical Activity and Sedentary Behavior Questionnaire (of the Canadian Society for Exercise Physiology) | Fowles et al. [54] English version | ● | Acc | 32 (of 35), 26♀, mean age♀ = 55 (SD = 10), mean age♂ = 63 (SD = 9), mean BMI♀ = 31 (SD = 6), mean BMI♂ = 26 (SD = 3), Canada |
| PASE         | Physical Activity Scale for the Elderly | Ngai et al. [66] Chinese version | ● | Acc | 90 (N/A), 54♀, mean age = 77.7 (SD = 7.7), mean BMI = 24.4 (SD = 3.8), China |
|              |                             | Vaughan et al. [73] Chinese version | ● | Acc | 73 (N/A), 71%♀, mean age = 79.0 (SD = 8.5), Chinese immigrants living in Vancouver for at least 5 years, Canada |
|              |                             | Covotta et al. [79] Italian version | ● | Acc | 94 (of 100), 49.5%♀, mean age = 62.9 (SD = 7.2), Italy |
|              |                             | Keikavoosi-Arani et al. [80] Persian version | ● | Acc | 278 (N/A), 65%♀, mean age = 74.2 (SD = 14.8), mean BMI = 28.2 (SD = 9.9), Iran |
|              |                             | Ayvat et al. [81] Turkish version | ● | Acc | 80 (N/A), 29♀, mean age = 69.7 (SD = 4.6), mean BMI = 27.7 (SD = 4.9), Turkey |
| PAVS         | Physical Activity Vital Sign Questionnaire | Ball et al. [45] English version | ● | Q | 298 (of 305), 115♀, age n (%): ≥ 55 years = 202 (67.8%), adults within primary health care, USA |
| PHAS question | Public Health Agency of Sweden Physical Activity Question | Ekblohm et al. [52] Swedish version | ● | Acc | 948 (of 1111), 486♀, median age♀ = 57.5 (IQR = 53.7–61.4), median age♂ = 57.7 (IQR = 53.8–62.0), Sweden |
| QAPPA        | Questionnaire d’Activité Physique pour les Personnes Âgées (Physical Activity Questionnaire for the Elderly) | de Souto Barreto [70] French version | ● | Q | 265 (N/A), 62.9%♀, mean age = 70.7 (SD = 7.3), France |
| Abbreviation | Full name of questionnaire | Studies on measurement properties | Assessed measurement properties | Comparison measures | Sample |
|--------------|---------------------------|----------------------------------|-------------------------------|-------------------|--------|
| SBAS         | Stanford Brief Activity Survey | Taylor-Piliae et al.\(^a\) \[71\] English version | ● Reliability, Measurement error, Hypotheses testing for construct validity |  | 1017 (of 1023), 382 ♀, mean age = 65.8 (SD = 2.8), mean BMI = 28.4 (SD = 5.2), USA |
| SGPALS (LT question) | Saltin-Grimby Physical Activity Level Scale (single question about LTPA) | Ekblom et al. \[52\] Swedish version | ● | Acc 948 (of 1111), 486 ♀, median age ♀ = 57.5 (IQR = 53.7–61.4), median age ♂ = 57.7 (IQR = 53.8–62.0), Sweden |
| Single item on Recreational and Domestic Activity | Single item on Recreational and Domestic Activity (from the British Regional Heart Study) | Jefferis et al. \[62\] English version | ● | Acc 1377 (of 1655), all ♂, mean age = 78.5 (SD = 4.6), mean BMI = 27.1 (SD = 3.8), UK |
| Walking question | Single question asking about time spent Walking | Ekblom et al. \[52\] Swedish version | ● | Acc 948 (of 1111), 486 ♀, median age ♀ = 57.5 (IQR = 53.7–61.4), median age ♂ = 57.7 (IQR = 53.8–62.0), Sweden |
| WHI-PAQ | Women’s Health Initiative Physical Activity Questionnaire | Neuhouser et al. \[65\] English version | ● | DLW 450 (of 450), all ♀, age ≥ 60, 31.8% BMI (18.5–24.9), USA |
| WHS-AASPA | Women’s Health Study: Accelerometer Ancillary Study Physical Activity Form (based on the NHS II Activity Questionnaire) | Shiroma et al. \[68\] English version | ● | Acc 10,115 (of 16,689), all ♀, mean age = 71.6 (SD = 5.7), mean BMI = 26.1, USA |
| ZPAQ | Zutphen Physical Activity Questionnaire | Harris et al. \[57\] English version | ● | Acc 234 (of 240), 110 ♀, mean age = 73.6 (SD = 6.1), mean BMI = 27.0 (SD = 4.0), UK |
|            |                           | Harris et al. \[57\] Modified English version | ● | Acc, Ped 234 (of 240), 110 ♀, mean age = 73.6 (SD = 6.1), mean BMI = 27.0 (SD = 4.0), UK |

\(^a\)Results for hypotheses testing for construct validity were not included since comparisons were performed with non-PA measures.
in Table 5 shows the results for different hypotheses for construct validity and responsiveness of studies included in this update. The results of the reassessment of all studies included in the previous review are shown in Electronic Supplementary Material Table S4. The level of quality varied but most studies were of very good or adequate quality. Versions of the AAS (English version), Cambridge Index (English version), CHAMPS (English version, Modified English version), IPAQ-LF (English version, Modified Dutch version), IPAQ-SF (Chinese version, English version, Japanese version, Portuguese version), LAPAQ (Dutch version), PASE (Dutch version, English version, Japanese version, Turkish version) and the Self-administered PAQ (Swedish version) were evaluated in multiple studies.

In at least one study, versions of 10 questionnaires [CHAMPS, FPACQ, IPAQ-LF, IPAQ-SF, Incidental and Planned Exercise Questionnaire (IPEQ), Modified Baecke, PASB-Q, PASE, QAPSE, WHI-PAQ] showed sufficient reliability in assessing the overall construct (e.g., total PA, total LTPA) and/or subdimensions (i.e., MVPA, walking) of PA. Measurement error was assessed for versions of four questionnaires [CHAMPS, Longitudinal Aging Study Amsterdam Physical Activity Questionnaire (LAPAQ), PASE, Questionnaire used in the EPIC (EPIC)]. The measurement errors of these versions were insufficient for all scores.

### 3.3.3 Construct Validity and Responsiveness

Table 5 shows the results for different hypotheses for construct validity and responsiveness of studies included in this update. The results of the reassessment of all studies included in the previous review are shown in Electronic Supplementary Material Table S4. The level of quality varied but most studies were of very good or adequate quality. Versions of the AAS (English version), Cambridge Index (English version), CHAMPS (English version, Modified English version), IPAQ-LF (English version, Modified Dutch version), IPAQ-SF (Chinese version, English version, Japanese version, Portuguese version), LAPAQ (Dutch version), PASE (Dutch version, English version, Japanese version, Turkish version) and the Self-administered PAQ (Swedish version) were evaluated in multiple studies.

In at least one study, versions of 13 questionnaires (AAS, ACLS-PALS, ACLS-PASS, BRHS, CHAMPS, IPAQ-LF, mLTPA-Q, Neighborhood Physical Activity Questionnaire (NPAQ), PAQ-EJ, PASB-Q, PASE, PAVS, Single item on Recreational and Domestic Activity) showed sufficient hypotheses testing for construct validity in assessing the overall construct (e.g., total PA, total LTPA) and/or subdimensions (i.e., MVPA, walking) of PA. The results for the SBAS [99] and QAPPA [70] were not rated because the authors reported p-values rather than effect sizes. Similarly, the results for the General Practice Physical Activity Questionnaire (GPPAQ) [44] were not rated since no combined effect size for sensitivity and specificity was reported [e.g., area under the curve (AUC)]. The responsiveness of the AAS for the assessment of MVPA and other subdimensions of PA was insufficient.

### 3.4 Quality of the Body of Evidence

The quality of the body of evidence (i.e., all studies from the previous review and update combined) together with the rating of measurement properties for all available self-administered questionnaires assessing PA in older adults is shown in Table 6. None of the included questionnaires provided evidence for all relevant measurement properties (reliability, measurement error, hypotheses testing for construct validity, responsiveness). Overall, the quality of evidence for both sufficient and insufficient measurement properties was often low to moderate. The CHAMPS, IPAQ-SF and PASE were the most frequently assessed.

In addition to the evidence provided for each questionnaire version, we considered summarizing the results from multiple studies on eight questionnaires (AAS, Cambridge Index, CHAMPS, IPAQ-LF, IPAQ-SF, LAPAQ, OA-ESI, PASE). Regarding reliability and measurement error, results from studies on versions of the IPAQ-SF and PASE (i.e., for the assessment of walking only) were not summarized due to the observed inconsistency in results. Likewise, we did not summarize the results on hypotheses testing for construct validity on versions of the IPAQ-LF and PASE. It is likely that these inconsistent results can be explained by cultural adaptations and modifications of the questionnaire. Results of versions of the ZPAQ were not summarized because they were assessed in the same sample. Two studies [59, 84] assessed modified English versions of the CHAMPS. Because of moderate-to-strong modifications of the original questionnaire (e.g., replacing items; see Sect. 3.1), we considered these versions as different instruments and provided the quality of evidence separately.

Several limitations associated with the quality of evidence were observed. First, for some questionnaires, serious indirectness was considered when the evidence was based on a single study including only women or men (e.g., BRHS [62]). Second, sometimes, a positive result was only reported in a subsample of participants such as in men at older age [e.g., reliability of the IPAQ-SF (Japanese version) for the assessment of walking [72]]. Furthermore, some studies reported results based on different levels of quality (e.g., very good and doubtful). If this was the case, we considered results based on higher quality for the grading. For example, one study [49] aimed to investigate the agreement between PAEE estimated by the CHAMPS and DLW and also presented results compared to the accelerometer. Although the comparison to the accelerometer was sufficient, we used the results based on DLW for the evaluation of the quality of evidence. The use of modified versions and selective reporting of results across different measurement properties resulted in the disadvantage that the evidence could not be considered for the same questionnaire. For instance, two studies [65, 88] evaluated the measurement properties...
| Questionnaire                  | Study population (n) for analysis | Interval       | Results                                                                 | Study quality and result rating<sup>a</sup> |
|-------------------------------|-----------------------------------|----------------|--------------------------------------------------------------------------|----------------------------------------------|
| Active-Q                      |                                  |                | Light: ICC = 0.66 [0.57–0.75]                                             | 1–                                            |
| Swedish version               | 148                               | 3 weeks        | Moderate: ICC = 0.69 [0.60–0.77]                                           | 1–                                            |
| Bonn et al. [48]              |                                   |                | Vigorous: ICC = 0.51 [0.39–0.63]                                           | 1–                                            |
|                               |                                   |                | Moderate-to-vigorous: ICC = 0.67 [0.58–0.76]                                | 1–                                            |
|                               |                                   |                | Sedentary-to-light: ICC = 0.67 [0.58–0.76]                                  | 1–                                            |
|                               |                                   |                | Sedentary: ICC = 0.80 [0.74–0.86]                                          | 1–                                            |
| CHAMPS                        | 56                                | 10 days        | Total (PAEE): ICC = 0.64                                                   | 1–                                            |
| English version               |                                   |                | Measurement error:                                                         |                                               |
| Colbert et al. [49]           |                                   |                | Total (PAEE): \( d = -11, \text{LOA} \) = -11 ± 1.96 \times 181 (kcal/day) |                                               |
| CHAMPS                        | 748                               | 6 months       | Total (duration): ICC = 0.69                                                | 2–                                            |
| Modified English version      |                                   |                | Total (PAEE): ICC = 0.64                                                   | 2–                                            |
| Hekler et al. [59]            |                                   |                | Low-light (duration): ICC = 0.70                                            | 2+                                            |
|                               |                                   |                | High-light (duration): ICC = 0.68                                           | 2–                                            |
|                               |                                   |                | Moderate-to-vigorous (duration): ICC = 0.66                                 | 2–                                            |
|                               |                                   |                | Moderate-to-vigorous (PAEE): ICC = 0.61                                    | 2–                                            |
|                               |                                   |                | Sedentary (duration): ICC = 0.56                                           | 1–                                            |
| GPPAQ                         | 126                               | 3 months       | Total: \( \kappa = 0.57 \)                                                 | 1–                                            |
| English version               |                                   |                | Total: \( \kappa = 0.63 \)                                                 | 2–                                            |
| Ahmad et al. [44]             |                                   |                |                                                                         |                                               |
| IPAQ-LF                       | 660 (\( n_{\text{men}} = 352, \)
| Serbian version               |                                   | 2 weeks         | Total (PAEE): ICC<sub>\text{men}</sub> = 0.71 [0.58–0.82]; ICC<sub>\text{women}</sub> = 0.74 [0.59–0.83] | 1+ 1+                                         |
| Milanović et al. [64]         |                                   |                | Moderate: ICC<sub>\text{men}</sub> = 0.77 [0.71–0.87]; ICC<sub>\text{women}</sub> = 0.64 [0.53–0.69] | 1+ 1–                                         |
|                               |                                   |                | Vigorous: ICC<sub>\text{men}</sub> = 0.88 [0.79–0.94]; ICC<sub>\text{women}</sub> = 0.82 [0.75–0.89] | 1+ 1+                                         |
|                               |                                   |                | Walking: ICC<sub>\text{men}</sub> = 0.69 [0.55–0.81]; ICC<sub>\text{women}</sub> = 0.61 [0.58–0.72] | 1– 1–                                         |
|                               |                                   |                | Work: ICC<sub>\text{men}</sub> = 0.64 [0.51–0.71]; ICC<sub>\text{women}</sub> = 0.85 [0.79–0.93] | 1– 1+                                         |
|                               |                                   |                | Transport: ICC<sub>\text{men}</sub> = 0.71 [0.62–0.79]; ICC<sub>\text{women}</sub> = 0.91 [0.81–0.96] | 1+ 1+                                         |
|                               |                                   |                | Housework/gardening: ICC<sub>\text{men}</sub> = 0.68 [0.56–0.75]; ICC<sub>\text{women}</sub> = 0.90 [0.80–0.95] | 1– 1+                                         |
|                               |                                   |                | Leisure: ICC<sub>\text{men}</sub> = 0.53 [0.42–0.64]; ICC<sub>\text{women}</sub> = 0.74 [0.68–0.81] | 1– 1+                                         |
| IPAQ-SF                       | 325                               | 2 weeks        | Total (PAEE; age group: 65–74): ICC<sub>\text{men}</sub> = 0.65 [0.46–0.78]; ICC<sub>\text{women}</sub> = 0.57 [0.34–0.72] | 1– 1–                                         |
| Japanese version              |                                   |                | Total (PAEE; age group: 75–89): ICC<sub>\text{men}</sub> = 0.50 [0.22–0.68]; ICC<sub>\text{women}</sub> = 0.56 [0.30–0.72] | 1– 1–                                         |
| Tomioka et al. [72]           | (\( n_{\text{women+aged 65–74}} = 88; \)
|                               |                                   |                | Moderate (age group: 65–74): ICC<sub>\text{men}</sub> = 0.52 [0.25–0.69]; ICC<sub>\text{women}</sub> = 0.47 [0.18–0.65] | 1– 1–                                         |
|                               |                                   |                | Moderate (age group: 75–89): ICC<sub>\text{men}</sub> = 0.63 [0.43–0.76]; ICC<sub>\text{women}</sub> = 0.60 [0.36–0.75] | 1– 1–                                         |
|                               |                                   |                | Vigorous (age group: 65–74): ICC<sub>\text{men}</sub> = 0.55 [0.31–0.71]; ICC<sub>\text{women}</sub> = 0.58 [0.36–0.73] | 1– 1–                                         |
|                               |                                   |                | Vigorous (age group: 75–89): ICC<sub>\text{men}</sub> = 0.39 [0.06–0.61]; ICC<sub>\text{women}</sub> = 0.30 [−0.11–0.56] | 1– 1–                                         |
### Table 4 (continued)

| Questionnaire | Study population (n) for analysis | Interval | Results | Study quality and result ratinga |
|---------------|----------------------------------|----------|---------|----------------------------------|
| **Walking** | | | | |
| (age group: 65–74): | 1 week | | | |
| ICCmen = 0.73 [0.59–0.83]; ICCwomen = 0.55 [0.32–0.71] | | 1+ | 1– | |
| (age group: 75–89): | | | | |
| ICCmen = 0.65 [0.46–0.77]; ICCwomen = 0.60 [0.36–0.75] | | 1– | 1– | |
| **Sitting** | | | | |
| (age group: 65–74): | 2 weeks | | | |
| ICCmen = 0.82 [0.71–0.88]; ICCwomen = 0.70 [0.54–0.80] | | 2– | 2– | |
| (age group: 75–89): | | | | |
| ICCmen = 0.66 [0.48–0.78]; ICCwomen = 0.67 [0.48–0.80] | | 2– | 2– | |

**IPEQ**

- **English version**
- Delbaere et al. [51]
- past week version = 30; past 3 months version = 50

| | 1 week | | | |
|––––––|––––––|––––––|––––––|––––––|
| Total (last week version): | ICC = 0.77 | | 1+ | |
| Total (last 3 months version): | ICC = 0.84 | | 1+ | |

**LAPAQ**

- **Dutch version**
- Siebeling et al. [69]
- (nrepresentative sample = 50)

| | 2 weeks | | | |
|––––––|––––––|––––––|––––––|––––––|
| Total (overall sample): | r = 0.68 [0.55–0.80] | | 2– | |
| Total (representative sample): | r = 0.73 [0.59–0.88] | | 2– | |
| Mild (overall sample): | r = 0.58 [0.42–0.72] | | 2– | |
| Mild (representative sample): | r = 0.69 [0.54–0.84] | | 2– | |
| Moderate (overall sample): | r = 0.79 [0.69–0.88] | | 2– | |
| Moderate (representative sample): | r = 0.81 [0.69–0.93] | | 2+ | |
| Vigorous (overall sample): | r = 0.75 [0.47–0.87] | | 2– | |
| Vigorous (representative sample): | r = 0.81 [0.49–0.93] | | 2+ | |

**mLTPA-Q**

- **English version**
- Fowles et al. [54]
- 35

| | 1 week | | | |
|––––––|––––––|––––––|––––––|––––––|
| Total (LTPA): | r = 0.04 | | 2– | |
| Moderate (LTPA): | r = 0.49 | | 2– | |
| Strenuous (LTPA): | r = 0.45 | | 2– | |
| Moderate-to-vigorous (LTPA): | r = 0.66 | | 2+ | |
| Muscle-strengthening (frequency): | r = 0.92 | | 2+ | |

**PASB-Q**

- **English version**
- Fowles et al. [54]
- 35

| | 1 week | | | |
|––––––|––––––|––––––|––––––|––––––|
| Moderate-to-vigorous (PAVS): | r = 0.83 | | 2+ | |

**PASE**

- **Chinese version**
- Ngai et al. [66]
- 32

| | N/A | | | |
|––––––|––––––|––––––|––––––|––––––|
| Total: | ICC = 0.81 | | ? + | |

**PASE**

- **Chinese version**
- Vaughan et al. [73]
- 66

| | 2 weeks | | | |
|––––––|––––––|––––––|––––––|––––––|
| Walking outside home: | \( \kappa = 0.45 \) | | 1– | |
| Light sports/recreational activities: | \( \kappa = 0.33 \) | | 1– | |
| Moderate sports/recreational activities: | \( \kappa = 0.51 \) | | 1– | |
| Strenuous sports/recreational activities: | \( \kappa = 0.65 \) | | 1– | |
| Muscle strength/endurance exercise: | \( \kappa = 0.43 \) | | 1– | |
| Light housework: | \( \kappa = 0.78 \) | | 1+ | |
| Heavy housework or chores: | \( \kappa = 0.64 \) | | 1– | |
| Home repairs: | \( \kappa = 0.39 \) | | 1– | |
| Lawn work or yard care: | \( \kappa = 0.17 \) | | 1– | |
| Outdoor gardening: | \( \kappa = 0.85 \) | | 1+ | |
| Caring for another person: | \( \kappa = 0.62 \) | | 1– | |
| Work for pay or as a volunteer: | \( \kappa = 0.92 \) | | 1+ | |

**Measurement error:**

- Total: \( \text{MDD}_{95} = 63.1, \text{SEM} = 22.8 \) (weighted total score) | 1– | |
- Total: \( \overline{d} = 2.4, \text{LOA} = 2.4 \pm 68.5 \) (weighted total score) | 1– | |
| Questionnaire                  | Study population (n) for analysis | Interval       | Results                                                                 | Study quality and result rating¹ |
|-------------------------------|----------------------------------|----------------|--------------------------------------------------------------------------|----------------------------------|
| **PASE**                      |                                  |                |                                                                          |                                  |
| Italian version               | 48                               | 1 week         | Total: ICC = 0.98 (0.96–0.99)                                           | 1+                               |
| Covotta et al. [79]           |                                  |                | Leisure time activity: ICC = 0.99 (0.99–0.99)                            | 1+                               |
|                               |                                  |                | Household activity: ICC = 0.99 (0.98–0.99)                               | 1+                               |
|                               |                                  |                | Work-related activity: ICC = 0.97 (0.94–0.98)                            | 1+                               |
|                               |                                  |                | Walking outside home: ICC = 0.90 (0.92–0.94)                             | 1+                               |
|                               |                                  |                | Light sports/recreational activities: ICC = 0.89 (0.87–0.91)             | 1+                               |
|                               |                                  |                | Moderate sports/recreational activities: ICC = 0.93 (0.90–0.95)          | 1+                               |
|                               |                                  |                | Strenuous sports/recreational activities: ICC = 0.91 (0.89–0.92)         | 1+                               |
|                               |                                  |                | Household activity: ICC = 0.86 (0.82–0.87)                               | 1+                               |
|                               |                                  |                | Light housework: ICC = 0.86 (0.82–0.86)                                  | 1+                               |
|                               |                                  |                | Heavy housework or chores: ICC = 0.81 (0.80–0.84)                        | 1+                               |
|                               |                                  |                | Home repairs: ICC = 0.76 (0.72–0.77)                                     | 1+                               |
|                               |                                  |                | Lawn work or yard care: ICC = 0.80 (0.79–0.81)                           | 1+                               |
|                               |                                  |                | Caring for another person: ICC = 0.95 (0.92–0.97)                        | 1+                               |
|                               |                                  |                | Job—standing or walking: ICC = 0.91 (0.90–0.94)                          | 1+                               |
| **PASE**                      | 278                              | 2 weeks        |                                                                          |                                  |
| Persian version               |                                  |                |                                                                          |                                  |
| Keikavoosi-Arani et al. [80]  |                                  |                |                                                                          |                                  |
|                               |                                  |                |                                                                          |                                  |
| **PASE**                      | 80                               | 1 week         | Total: ICC = 0.99 (0.99–0.99)                                           | 1+                               |
| Turkish version               |                                  |                | Leisure time activity: ICC = 0.99 (0.99–0.99)                            | 1+                               |
| Ayvat et al. [81]             |                                  |                | Household activity: ICC = 0.99 (0.99–0.99)                               | 1+                               |
|                               |                                  |                | Work-related activity: ICC = 1.00 (1.00–1.00)                            | 1+                               |
| **QAPPA**                     | 225                              | 1 year         | Moderate (PAEE): ICC = 0.46                                             | 2−                               |
| French version                |                                  |                | Vigorous (PAEE): ICC = 0.63                                              | 2−                               |
| de Souto Barreto [70]         |                                  |                | Moderate-to-vigorous (PAEE): ICC = 0.64                                  | 2−                               |
|                               |                                  |                | Classification (active/inactive): κ = 0.44                               |                                  |
| **SBAS**                      | 996                              | 2 years        | Total: ρ = 0.62                                                           | 3−                               |
| Taylor-Piliae et al. [71]     |                                  |                |                                                                          |                                  |
| English version               |                                  |                |                                                                          |                                  |

©As described in Sect. 2.5, the quality of the individual study was evaluated per questionnaire and construct/dimension of PA and can be either very good (1), adequate (2), doubtful (3) or inadequate (4). Additionally, the reported results were rated [i.e., sufficient (+), insufficient (−)] as described in Sect. 2.4

²Based on the reported results, we calculated the LOA using the formula LOA = $\bar{d} \pm 1.96 \times s \times \sqrt{2}$, where $s$ = within-subject standard deviation (typical error) [146]

Active-Q Web-based Physical Activity Questionnaire, Active-Q, CHAMPS Community Health Activities Model Program for Seniors, $\bar{d}$ change in the mean, GPPAQ General Practice Physical Activity Questionnaire, ICC intraclass correlation coefficient, κ Kappa coefficient; IPAQ-LF International Physical Activity Questionnaire—long-form, IPAQ-SF International Physical Activity Questionnaire—short-form, IPEQ Incidental and Planned Exercise Questionnaire, kcal kilocalories, LAPAQ Longitudinal Aging Study Amsterdam Physical Activity Questionnaire, LOA limits of agreement, LTPA leisure time physical activity; $MDD_{95}$ minimal detectable difference based on the 95% confidence interval, $min$ minutes, mLTPA-Q Modified Leisure Time Physical Activity Questionnaire, N/A not applicable, PA physical activity, PAEE physical activity energy expenditure, PASB-Q Physical Activity and Sedentary Behavior Questionnaire, PASE Physical Activity Scale for the Elderly, PAVS physical activity vital sign, QAPPA Questionnaire d’Activité Physique pour les Personnes Âgées (Physical Activity Questionnaire for the Elderly), $r$ Pearson correlation coefficient, $ρ$ Spearman correlation coefficient, SBAS Stanford Brief Activity Survey, $SEM$ standard error of measurement, $?/\text{unclear}$
Table 5 Hypotheses testing for construct validity and responsiveness of PA questionnaires for older adults

| Questionnaire | Study population \((n)\) for analysis | Comparison measure (type, placement, registration period [valid week], epoch length, cut points) | Results | Study quality and result rating\(^a\) |
|---------------|----------------------------------------|-------------------------------------------------|---------------------------------|----------------------|
| AAFQ          | 450                                    | DLW                                             | Total (PAEE): \(R^2 = 7.6\%\) (24.0\% when corrected for measurement error) | 1−                   |
| English version |                                       | Neuhouser et al. [65]                           | 50–64 years of age               |                      |
| AAS           | \(n_{\text{50–64 years of age}} = 186,\) | Accelerometer (ActiGraph GT3X, hip, waking hours of 7 days [5 days], 1 s, Freedson et al. [136]) | 50–64 years of age               |                      |
| English version | \(n_{\text{over 65 years of age}} = 132\) |                                                   | Moderate: \(\rho = 0.24\ [0.10–0.38]\) | 1−                   |
| Vandelanotte et al. [77] | |                                                   | Vigorous: \(\rho = 0.41\ [0.27–0.54]\) | 1−                   |
|               | \(n_{\text{50–64 years of age}} = 134,\) |                                                   | Moderate-to-vigorous: \(\rho = 0.28\ [0.15–0.43]\) | 1−                   |
|               | \(n_{\text{over 65 years of age}} = 104\) |                                                   | Over 65 years of age             |                      |
|               |                                       |                                                   | Moderate: \(\rho = 0.20\ [0.02–0.37]\) | 1−                   |
|               |                                       |                                                   | Vigorous: \(\rho = 0.20\ [0.02–0.38]\) | 1−                   |
|               |                                       |                                                   | Moderate-to-vigorous: \(\rho = 0.21\ [0.05–0.38]\) | 1−                   |
| AAS           | \(n_{\text{first group}} = 39, n_{\text{second group}} = 37\) | Accelerometer (ActiGraph GT1M, hip, waking hours of 7 days [4 days], 5 s, Freedson et al. [136]) | No minimum bout length |                      |
| English version |                                       |                                                   | Total (LTPA): \(\rho = 0.56; \rho = 0.49\) | 3+ 3+                |
| Freene et al. [55] | |                                                   | Moderate (LTPA, including walking): \(\rho = 0.56; \rho = 0.55\) | 3+ 3+                |
|               |                                       |                                                   | Vigorous (LTPA): \(\rho = 0.33; \rho = -0.08\) | 3− 3−                |
|               |                                       |                                                   | Classification of LTPA (active/inactive): \(\phi = 0.41; \phi = 0.16\) |                      |
|               |                                       |                                                   | 10-min minimum bout length        |                      |
|               |                                       |                                                   | Total (LTPA): \(\rho = 0.56; \rho = 0.64\) | 3+ 3+                |
|               |                                       |                                                   | Moderate (LTPA, including walking): \(\rho = 0.63; \rho = 0.64\) | 3+ 3+                |
|               |                                       |                                                   | Vigorous (LTPA): \(\rho = 0.17\) | 3−                   |
|               |                                       |                                                   | Classification of LTPA (active/inactive): \(\phi = 0.47; \phi = 0.21\) |                      |
| AAS           | 50                                     | Pedometer (Yamax SW-200, waking hours of 7 days [4 days]) | Total (LTPA): \(\rho = 0.42\) | 3+                   |
| English version |                                       |                                                   | Moderate-to-vigorous (LTPA, excluding walking): \(\rho = 0.31\) | 3−                   |
| Heesch et al. [58] | |                                                   | Walking (LTPA): \(\rho = 0.42\) | 1−                   |
| Questionnaire   | Study population (n) for analysis | Comparison measure (type, placement, registration period [valid week], epoch length, cut points)                                                                 | Results                                                                                                      | Study quality and result rating |
|----------------|----------------------------------|-----------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|---------------------------------|
| ACLS-PALS      | 71                               | Accelerometer (Actical, waist, 7 days [4 days], 60 s, Hooker et al. [137])                                       | 1-min minimum bout length                                                                                     | 1+                             |
| English version |                                  | Total (Exercise): \( r = 0.55 \) [0.35–0.75]                                                                   | Classification (active/inactive): \( \kappa = 0.38 \) [0.16–0.60]                                             |                                 |
| Banda et al. [46] |                                 | 10-min minimum bout length                                                                                       | Total (Exercise): \( r = 0.49 \) [0.29–0.70]                                                                 | 1–                             |
|                |                                  | Classification (active/inactive): \( \kappa = 0.15 \) [0.03–0.28]                                                |                                                                                                               |                                 |
| ACLS-PASS      | 71                               | Accelerometer (Actical, waist, 7 days [4 days], 60 s, Hooker et al. [137])                                       | 1-min minimum bout length                                                                                     | 1+                             |
| English version |                                  | Moderate-to-vigorous: \( r = 0.53 \) [0.32–0.73]                                                                 | Classification (active/inactive): \( \kappa = 0.26 \) [0.04–0.48]                                             |                                 |
| Banda et al. [46] |                                 | 10-min minimum bout length                                                                                       | Total (Exercise): \( r = 0.37 \) [0.15–0.60]                                                                 | 1–                             |
|                |                                  | Classification (active/inactive): \( \kappa = 0.04 \) [− 0.04 to 0.11]                                            |                                                                                                               |                                 |
| Active-Q       | 148                              | Accelerometer (GENEA, wrist, two times for 7 days [6 days/week], 60 s, Bonn et al. [48])                         | Light: \( \rho = 0.15 \) [0.00–0.31]                                                                         | 1−                             |
| Swedish version |                                  | Moderate: \( \rho = 0.27 \) [0.12–0.42]                                                                         |                                                                                                               |                                 |
| Bonn et al. [48] |                                 | Vigorous: \( \rho = 0.54 \) [0.42–0.67]                                                                          |                                                                                                               |                                 |
|                |                                  | Moderate-to-vigorous: \( \rho = 0.35 \) [0.21–0.48]                                                              |                                                                                                               |                                 |
|                |                                  | Sedentary-to-light: \( \rho = 0.35 \) [0.19–0.51]                                                               |                                                                                                               |                                 |
|                |                                  | Sedentary: \( \rho = 0.19 \) [0.04–0.34]                                                                          |                                                                                                               |                                 |
|                |                                  | Moderate (classification based on quartiles): \( \kappa = 0.16 \)                                                |                                                                                                               |                                 |
|                |                                  | Vigorous (classification based on quartiles): \( \kappa = 0.39 \)                                               |                                                                                                               |                                 |
|                |                                  | Moderate-to-vigorous (classification based on quartiles): \( \kappa = 0.22 \)                                  |                                                                                                               |                                 |
|                |                                  | Light: \( \bar{d} = 87 \), LOA = – 398 to 571 (min/day)                                                        |                                                                                                               |                                 |
|                |                                  | Moderate: \( \bar{d} = 76 \), LOA = – 157 to 309 (min/day)                                                      |                                                                                                               |                                 |
|                |                                  | Vigorous: \( \bar{d} = 15 \), LOA = – 33 to 64 (min/day)                                                        |                                                                                                               |                                 |
|                |                                  | Moderate-to-vigorous: \( \bar{d} = 91 \), LOA = – 147 to 329 (min/day)                                        |                                                                                                               |                                 |
|                |                                  | Sedentary-to-light: \( \bar{d} = 91 \), LOA = – 329 to 146 (min/day)                                          |                                                                                                               |                                 |
|                |                                  | Sedentary: \( \bar{d} = 178 \), LOA = – 606 to 250 (min/day)                                                   |                                                                                                               |                                 |
| BRHS           | 1377                             | Accelerometer (Actigraph, GT3X, hip, waking hours of 7 days [3 days], Copeland et al. [138])                     | Total (compared to cpm; steps; MVPA): \( \rho = 0.49; \rho = 0.49; \rho = 0.49 \)                            | 1− 3+ 2+                       |
| English version |                                  |                                                                                                                 |                                                                                                               |                                 |
| Jefferis et al. [62] |                                 |                                                                                                                 |                                                                                                               |                                 |
| Questionnaire                  | Study population (n) for analysis | Comparison measure (type, placement, registration period [valid week], epoch length, cut points)                                                                 | Results                                                                                                                                          | Study quality and result rating^a |
|-------------------------------|----------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------|
| Cambridge Index               | 1689                             | Accelerometer + heart rate (Actiheart/Red Dot 2570: 3M, 5 days [48 h], 30 s, individual/group calibration)                                                   | Women                                                                                                                                              | 2− 1−                             |
| English version               |                                  |                                                                                                                                                             | Total (compared to MVPA; PAEE): $\rho = 0.21; \rho = 0.17$                                                                                       |                                   |
| España-Romero et al. [53]     |                                  |                                                                                                                                                             | Men                                                                                                                                               |                                   |
|                               |                                  |                                                                                                                                                             | Total (compared to MVPA; PAEE): $\rho = 0.24; \rho = 0.25$                                                                                       |                                   |
|                               |                                  |                                                                                                                                                             | Total (PAEE): $\rho = 0.52$                                                                                                                     | 2+                               |
| CHAMPS                        | 56                               | Accelerometer (ActiGraph GT1M, waist, waking hours of 10 days [$\geq 10$ h for any day], 10 s, Crouter et al. [139])                                          | Total (PAEE): $\rho = 0.28$                                                                                                                     | 1−                               |
| English version               |                                  |                                                                                                                                                             | Total (PAI): $\rho = 0.23$                                                                                                                     | 1−                               |
| Colbert et al. [49]           |                                  |                                                                                                                                                             | Total (PAEE): LOA$^d$ = − 968 to 130 (− 419 ± 1.96*280 kcal/day)                                                                               |                                   |
| CHAMPS                        | 56                               | DLW                                                                                                                                                         | Total (PAE): $\rho = 0.38$                                                                                                                     | 1−                               |
| Modified English version      |                                  |                                                                                                                                                             | Total (PAE): $\rho = 0.39$                                                                                                                     | 2−                               |
| Hekler et al. [59]            | 850                              | Accelerometer (Actigraph models 7164 and 71256, waist, waking hours of 7 days [5 days], 60 s, Freedson et al. [136], Cope-land et al. [138])                 | Low-light (duration): $\rho = 0.06$                                                                                                               | 1−                               |
|                               |                                  |                                                                                                                                                             | High-light (duration): $\rho = 0.27$                                                                                                              | 1−                               |
|                               |                                  |                                                                                                                                                             | Moderate-to-vigorous (duration): $\rho = 0.37$                                                                                                    | 1−                               |
|                               |                                  |                                                                                                                                                             | Moderate-to-vigorous (PAEE): $\rho = 0.38$                                                                                                       | 1−                               |
|                               |                                  |                                                                                                                                                             | Sedentary (duration): $\rho = 0.12$                                                                                                              |                                   |
|                               |                                  |                                                                                                                                                             | Total (duration): $\bar{d} = 618$, LOA = − 504 to 1740 (min/week)                                                                            |                                   |
|                               |                                  |                                                                                                                                                             | Low-light (duration): $\bar{d} = − 473$, LOA = − 1937 to 992 (min/week)                                                                         |                                   |
|                               |                                  |                                                                                                                                                             | High-light (duration): $\bar{d} = 396$, LOA = − 346 to 1137 (min/week)                                                                          |                                   |
|                               |                                  |                                                                                                                                                             | Moderate-to-vigorous (duration): $\bar{d} = 222$, LOA = − 403 to 848 (min/week)                                                                |                                   |
| Questionnaire                  | Study population (n) for analysis | Comparison measure (type, placement, registration period [valid week], epoch length, cut points)                                                                 | Results                                                                                           | Study quality and result rating¹ |
|-------------------------------|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|----------------------------------|
| EPAQ2                        | 1689                              | Accelerometer + heart rate (Actiheart/Red Dot 2570: 3M, 5 days [48 h], 30 s, individual/group calibration)                       | Women                                                                                             | 1–                               |
| España-Romero et al. [53]     |                                   |                                                                                                                                | Total (PAEE): \( \rho = 0.26 \)                                                                  |                                  |
|                               |                                   |                                                                                                                                | Light: \( \rho = 0.12 \)                                                                          |                                  |
|                               |                                   |                                                                                                                                | Moderate-to-vigorous: \( \rho = 0.36 \)                                                          |                                  |
|                               |                                   |                                                                                                                                | Sedentary: \( \rho = 0.18 \)                                                                     |                                  |
|                               |                                   |                                                                                                                                | Total (PAEE): \( \bar{d} = 29, \text{LOA} = -39 \text{ to } 95 \text{ (kJ/kg/day)} \)          |                                  |
|                               |                                   |                                                                                                                                | Light: \( \bar{d} = -60, \text{LOA} = -368 \text{ to } 247 \text{ (min/day)} \)               |                                  |
|                               |                                   |                                                                                                                                | Moderate-to-vigorous: \( \bar{d} = 55, \text{LOA} = -117 \text{ to } 228 \text{ (min/day)} \) |                                  |
|                               |                                   |                                                                                                                                | Sedentary: \( \bar{d} = -6.0, \text{LOA} = -10.9 \text{ to } 1.0 \text{ (h/day)} \)           |                                  |
| Men                          |                                   |                                                                                                                                |                                                                                                  |                                  |
|                               |                                   |                                                                                                                                |                                                                                                  |                                  |
| GPPAQ                        | 289                               | Accelerometer (Actigraph, GT3X+, waist, waking hours of 7 days [5 days], 5 s, Freedson et al. [136])                         | Women                                                                                             | 1–                               |
| English version              |                                   |                                                                                                                                | Classification (active/inactive): Sensitivity = 19%                                             |                                  |
| Ahmad et al. [44]            |                                   |                                                                                                                                | Classification (active/inactive): Specificity = 85%                                             |                                  |
|                               |                                   |                                                                                                                                | Classification (active/inactive; including walking): Sensitivity = 39%                          |                                  |
|                               |                                   |                                                                                                                                | Classification (active/inactive; including walking): Specificity = 70%                          | Not rated                        |
|                               |                                   |                                                                                                                                |                                                                                                  |                                  |
Table 5 (continued)

| Questionnaire                     | Study population (n) for analysis | Comparison measure (type, placement, registration period [valid week], epoch length, cut points) | Results                                                                 | Study quality and result ratinga |
|-----------------------------------|----------------------------------|-----------------------------------------------------------------------------------------------|--------------------------------------------------------------------------|----------------------------------|
| IPAQ-E Swedish version Hurtig-Wennlöf et al. [60] | 54                               | Accelerometer (Actigraph GT1M, waist, waking hours of 7 days [4 days], 15 s, Matthews et al. [140], Swartz et al. [141], Troiano et al. [142]) | Moderate: \( \rho = 0.47; \rho = 0.17; \rho = 0.40 \) Moderate plus walking: \( \rho = 0.42, \rho = 0.34, \rho = 0.41 \) Vigorous: \( \rho = 0.37 \) Walking: \( \rho = 0.30 \) Sitting: \( \rho = 0.28 \) Classification (active/inactive): \( \kappa = 0.45 \ [0.18–0.72] \), Sensitivity = 81%, Specificity = 85% | 1− 1− 1− |
| IPAQ-LF English version Cleland et al. [78] | 226                              | Accelerometer (ActiGraph GT3X+, waist, waking hours of 7 days [5 days], 15 s, Copeland et al. [138]) | Moderate-to-vigorous: \( \rho = 0.52 \) Moderate-to-vigorous: \( \bar{\Delta} = -100, \text{LOA} = 1766 \text{ to } -1965 \ (\text{min/week}) \) | 1+ |
| IPAQ-LF Modified Dutch version Winckers et al. [74] | 196                              | Accelerometer (ActiGraph GT3X+, hip, waking hours of 7 days [5 days]) | Total (PAEE): \( \rho = 0.16 \) Moderate-to-vigorous (compared with Actigraph total counts): \( \rho = 0.27 \) Moderate-to-vigorous (compared with Actigraph MVPA): \( \rho = 0.16 \) | 3− |
| IPAQ-SF English version Grimm et al. [56] | 127                              | Accelerometer (ActiGraph 7164, hip, waking hours of 7 days [5 days], 60 s, Crouter et al. [139], Freedson et al. [136], Matthews et al. [140]) | No minimum bout length: Total (PAEE): \( \rho = 0.24 \) Moderate: \( \rho = 0.16 \) Vigorous: \( \rho = 0.18 \) Walking: \( \rho = 0.29 \) Sitting: \( \rho = 0.36 \) Classification (active/passive): \( \kappa = 0.02 \) 10-min minimum bout length: Total (PAEE): \( \rho = 0.17 \) Moderate: \( \rho = 0.04 \) Vigorous: \( \rho = 0.17 \) Walking: \( \rho = 0.28 \) Classification (active/passive): \( \kappa = 0.10 \) Moderate lifestyle: \( \bar{\Delta} = -15, \text{LOA} = -214 \text{ to } 184 \ (\text{min/day}) \) Moderate walking: \( \bar{\Delta} = -27, \text{LOA} = -143 \text{ to } 90 \ (\text{min/day}) \) Vigorous: \( \bar{\Delta} = -43, \text{LOA} = -176 \text{ to } 90 \ (\text{min/day}) \) Sitting: \( \bar{\Delta} = 262, \text{LOA} = -84 \text{ to } 608 \ (\text{min/day}) \) | 3− |
| Questionnaire | Study population (n) for analysis | Comparison measure (type, placement, registration period [valid week], epoch length, cut points) | Results | Study quality and result rating |
|---------------|----------------------------------|-------------------------------------------------|---------|-------------------------------|
| IPAQ-SF Japanese version Tomioka et al. [72] | 306 (n<sub>women+aged 65–74</sub> = 84; n<sub>men+aged 65–74</sub> = 76; n<sub>women+aged 75–89</sub> = 69; n<sub>men+aged 75–89</sub> = 77) | Accelerometer (Kenz Lifecorder PLUS, waist, waking hours of at least 28 days [14 days], Tomioka et al. [72]) | Total (PAEE; age group: 65–74): ρ<sub>men</sub> = 0.42, ρ<sub>women</sub> = 0.49 | 2– 2− |
| | | | Total (PAEE; age group: 75–89): ρ<sub>men</sub> = 0.53, ρ<sub>women</sub> = 0.49 | 2+ 2− |
| | | | Moderate (age group: 65–74): ρ<sub>men</sub> = 0.26, ρ<sub>women</sub> = 0.13 | 2– 2− |
| | | | Moderate (age group: 75–89): ρ<sub>men</sub> = 0.05, ρ<sub>women</sub> = 0.03 | 2– 2− |
| | | | Vigorous (age group: 65–74): ρ<sub>men</sub> = 0.25, ρ<sub>women</sub> = 0.12 | 2– 2− |
| | | | Vigorous (age group: 75–89): ρ<sub>men</sub> = 0.17, ρ<sub>women</sub> = 0.17 | 2– 2− |
| | | | Walking (age group: 65–74): ρ<sub>men</sub> = 0.30, ρ<sub>women</sub> = 0.48 | 1– 1– |
| | | | Walking (age group: 75–89): ρ<sub>men</sub> = 0.59, ρ<sub>women</sub> = 0.55 | 1– 1– |
| | | | Classification (tertiles; age group: 65–74): κ<sub>men</sub> = 0.50 [0.36–0.64], κ<sub>women</sub> = 0.39 [0.22–0.56] |  |
| | | | Classification (tertiles; age group: 75–89): κ<sub>men</sub> = 0.47 [0.31–0.63], κ<sub>women</sub> = 0.47 [0.28–0.66] |  |
| IPAQ-SF Portuguese version Colpani et al. [50] | 292 | Pedometer (BP 148 TECHLINE, waist, waking hours of 7 days) | Total (PAEE): ρ = 0.11 | 3– |
| | | | Classification (active/inactive): κ = 0.11 |  |
| | | | Classification (active/moderate/inactive): κ = 0.08 |  |
| LAPAQ Dutch version Koolhaas et al. [63] | 1410 | Accelerometer (GeneActiv, wrist, 7 days [4 days], White et al. [143]) | Total: $\ddot{d} = -0.17$, LOA = -2.36 to 2.03 (Z scores of steps and MET<sub>log</sub>) | 1− |
| | | | Total: $\ddot{d} = -529$, LOA<sub>d</sub> = -529 ± 1.96*1080 (min/week) |  |
| | | | Light: $\ddot{d} = -708$, LOA<sub>d</sub> = -708 ± 1.96*484 (min/week) |  |
| | | | Moderate: $\ddot{d} = 205$, LOA<sub>d</sub> = 205±1.96*781 (min/week) |  |
| | | | Vigorous: $\ddot{d} = -26$, LOA<sub>d</sub> = -26 ± 1.96*338 (min/week) |  |
| LAPAQ Dutch version Siebeling et al. [69] | 88 | Accelerometer (Sensewear Pro, upper arm, 14 days) | Total: $r = 0.25$ [0.07–0.44] | 1− |
| | | | Mild: $r = 0.05$ [−0.16 to 0.24] |  |
| | | | Moderate: $r = 0.27$ [0.07–0.48] |  |
| | | | Vigorous: $r = 0.01$ [−0.07 to 0.25] |  |
| | | | Classification (active/inactive): AUC: 0.73 [0.59–0.86] |  |
| | | | Total: $\ddot{d} = -354$, LOA<sub>d</sub> = -354 ± 1.96*1830 (min/2 weeks) |  |
| | | | Mild: $\ddot{d} = -267$, LOA<sub>d</sub> = -267 ± 1.96*1423 (min/2 weeks) |  |
| | | | Moderate: $\ddot{d} = -234$, LOA<sub>d</sub> = -234 ± 1.96*852 (min/2 weeks) |  |
| | | | Vigorous: $\ddot{d} = 148$, LOA<sub>d</sub> = 148 ± 1.96*403 (min/2 weeks) |  |
| Questionnaire                        | Study population (n) for analysis | Comparison measure (type, placement, registration period [valid week], epoch length, cut points)                                                                 | Results                                                                                                                                                                                                 | Study quality and result rating |
|-------------------------------------|----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------|
| mLTPA-Q English version Fowles et al. [54] | 32                               | Accelerometer (ActiGraph GT3X, hip, waking hours of 7 days [4 days], Freedson et al. [136])                                                                                                       | Moderate (LTPA): $r = 0.53$  
Strenuous (LTPA): $r = 0.18$  
Moderate-to-vigorous (LTPA): $r = 0.56$  
Classification (active/inactive): Sensitivity = 73%, Specificity = 82%  
Moderate-to-vigorous (LTPA): LOA = $-223$ to $262$ (min/week)                                                                                                                                 | 3+                            |
| Modified Minnesota LTPA-Q English version Sabia et al. [67] | 3975                             | Accelerometer (GeneActiv. Wrist, 9 days [4 days])                                                                                                                                               | Total (PAEE): $\rho = 0.33$ [0.30–0.36]  
Mild: $\rho = 0.21$ [0.18–0.24]  
Moderate: $\rho = 0.25$ [0.22–0.28]  
Vigorous: $\rho = 0.24$ [0.21–0.26]  
Walking: $\rho = 0.21$ [0.18–0.24]  
Cycling: $\rho = 0.15$ [0.12–0.18]  
Sports: $\rho = 0.22$ [0.19–0.25]  
Gardening: $\rho = 0.16$ [0.13–0.19]  
Do-it-yourself activities: $\rho = 0.15$ [0.12–0.18]  
Housework: $\rho = 0.09$ [0.05–0.12]  
Other: $\rho = 0.07$ [0.04–0.10]  
Classification (tertiles): $\kappa = 0.16$  
Moderate-to-vigorous: $\rho = 0.14$  
Classification (active/inactive): AUC = 0.57 [0.54–0.63], Sensitivity = 62%, Specificity = 56%  
Moderate-to-vigorous: Median difference = $-21$, 5th to 95th percentile: $-81$ to $111$ (min/day) | 2–                            |
| MVPA questions Swedish version Ekblom et al. [52] | 948                              | Accelerometer (ActiGraph GT3X and GT3X+, hip, waking hours of 7 days [4 days], 60s, Sasaki et al. [144])                                                                                           | Classification (tertiles): $\kappa = 0.16$  
Moderate-to-vigorous: $\rho = 0.14$  
Classification (active/inactive): AUC = 0.57 [0.54–0.63], Sensitivity = 62%, Specificity = 56%  
Moderate-to-vigorous: Median difference = $-21$, 5th to 95th percentile: $-81$ to $111$ (min/day)                                                                                                                                 | 1–                            |
| NC85+PAQ English version Innerd et al. [61] | 337                              | Accelerometer (GENEA, wrist, 7 days [5 days], Esliger et al. [145])                                                                                                                             | Total (low active group): $\rho = 0.10$  
Total (moderate active group): $\rho = 0.38$  
Total (high active group): $\rho = 0.34$ | 1–                            |
| Questionnaire          | Study population (n) for analysis | Comparison measure (type, placement, registration period [valid week], epoch length, cut points)                                                                 | Results                                                                 | Study quality and result rating \(^a\) |
|------------------------|----------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|----------------------------------------|
| NPAQ German version    | 58                               | Pedometer (HJ-720 IT-E2, waking hours of 7 days [3 days])                                                                                                                                   | Total: \( \rho = 0.45 [0.21–0.76] \)
Moderate: \( \rho = 0.22 [-0.10 \text{ to } 0.56] \)
Vigorous: \( \rho = -0.05 [-0.40 \text{ to } 0.50] \)
Moderate-to-vigorous: \( \rho = 0.15 [-0.18 \text{ to } 0.48] \)
Walking (total): \( \rho = 0.49 [0.24–0.82] \)
Walking (within neighborhood): \( \rho = 0.43 [0.16-0.76] \)
Walking (within neighborhood for transport): \( \rho = 0.20 [-0.11 \text{ to } 0.51] \)
Walking (within neighborhood for recreation): \( \rho = 0.20 [0.12–0.76] \)
Walking (outside neighborhood): \( \rho = 0.29 [-0.001 \text{ to } 0.60] \)
Walking (outside neighborhood for transport): \( \rho = 0.21 [-0.10 \text{ to } 0.53] \)
Walking (outside neighborhood for recreation): \( \rho = 0.52 [0.23–0.91] \)
Walking for transport (in and outside): \( \rho = 0.25 [-0.35 \text{ to } 0.55] \)
Walking for recreation (in and outside): \( \rho = 0.44 [0.17–0.77] \)
Total biking: \( \rho = 0.33 [0.05–0.64] \) | 1−                                                                 |
| PASB-Q English version | 32                               | Accelerometer (ActiGraph GT3X, hip, waking hours of 7 days [4 days], Freedson et al. [136])                                                                                                   | Moderate-to-vigorous (PAVS): \( r = 0.50 \)
Classification (active/inactive): Sensitivity = 60%, Specificity = 83% | 1+                                                                 |
| PASE                   | 80                               | Questionnaire (IPAQ-LF)                                                                                                                                                                        | Total: \( r = 0.74 \)
Leisure time activity: \( r = 0.68 \)
Household activity: \( r = 0.65 \)
Work-related activity: \( r = 0.57 \) | 3+                                                                 |
| PAVS                   | 269                              | Questionnaire (Modifiable Activity Questionnaire)                                                                                                                                             | Moderate-to-vigorous: \( r = 0.71 \)
Classification (active/inactive): \( \kappa = 0.55 [0.45–0.64] \)
Moderate-to-vigorous: \( \hat{d} = -86.3, \text{LOA} = -371 \text{ to } 199 \text{ (min/week)} \) | 3+                                                                 |
### Table 5 (continued)

| Questionnaire                      | Study population (n) for analysis | Comparison measure (type, placement, registration period [valid week], epoch length, cut points) | Results | Study quality and result rating |
|------------------------------------|----------------------------------|-----------------------------------------------------------------------------------------------------------------|---------|-------------------------------|
| PHAS question Swedish version      | 948                              | Accelerometer (ActiGraph GT3X and GT3X+, hip, waking hours of 7 days [4 days], 60 s, Sasaki et al. [145])       | Total (LTPA): $\rho = 0.26$ | 2−   |
| QAPPA French version de Souto Barreto [70] | 265                              | Questionnaire (exercise behavior [yes/no] in the last 2 months)                                                | Moderate (PAEE): Significant difference between exercisers and non-exercisers (Wilcoxon rank sum test) | Not rated |
|                                    |                                  |                                                                    | Vigorous (PAEE): Significant difference between exercisers and non-exercisers (Wilcoxon rank sum test) | Not rated |
|                                    |                                  |                                                                    | Moderate-to-vigorous (PAEE): Significant difference between exercisers and non-exercisers (Wilcoxon rank sum test) | Not rated |
|                                    |                                  |                                                                    | Classification (active/inactive): Significant difference between exercisers and non-exercisers (Chi-squared test) |  |
|                                    |                                  |                                                                    |  |
| SGPALS (LT question) Swedish version | 948                              | Accelerometer (ActiGraph GT3X and GT3X+, hip, waking hours of 7 days [4 days], 60 s, Sasaki et al. [144])       | Total (LTPA): $\rho = 0.21$ | 2−   |
| Single item on recreational and domestic activity English version Jefferis et al. [62] | 1377 | Accelerometer (Actigraph, GT3X, hip, waking hours of 7 days [3 days], Copeland et al. [138]) | Total (domestic/recreational PA, compared to cpm; steps; MVPA): $\rho = 0.46; \rho = 0.45; \rho = 0.43$ | 2+ 3+ 3+ |
| Walking question Swedish version Ekblom et al. [52] | 948 | Accelerometer (Actigraph GT3X and GT3X+, hip, waking hours of 7 days [4 days], 60 s, Sasaki et al. [144]) | Walking: $\rho = 0.26$ | 2−   |
|                                    |                                  |                                                                    | Classification (active/inactive): AUC = 0.61 [0.55–0.66], Sensitivity = 70%, Specificity = 48% |  |
| WHI-PAQ English version Neuhouser et al. [65] | 450 | DLW                                                              | Total (PAEE): $R^2 = 3.4\%$ (10.7% when corrected for biomarker measurement error) | 1−   |
| WHS-AASPA English version Shinma et al. [68] | 10,115 | Accelerometer (Actigraph GT3X+, hip, waking hours of 7 days, Copeland et al. [138], Freedson et al. [136], Matthews et al. [140], Sasaki et al. [144]) | Moderate-to-vigorous: $\rho = 0.35 [0.33–0.37]; \rho = 0.36 [0.35–0.38]; \rho = 0.39 [0.37–0.40]; \rho = 0.37 [0.36–0.39]$ | 1− 1− | 1− |
|                                    |                                  |                                                                    | No minimum bout length |  |
|                                    |                                  |                                                                    | Classification (active/inactive): $\kappa = 0.09; \kappa = 0.21; \kappa = 0.18; \kappa = 0.25$ |  |
|                                    |                                  |                                                                    | 10-min minimum bout length |  |
|                                    |                                  |                                                                    | Classification (active/inactive): $\kappa = 0.25; \kappa = 0.22; \kappa = 0.11; \kappa = 0.15$ |  |
| ZPAQ English version Harris et al. [57] | 234 | Accelerometer (Actigraph GT1M, hip, 7 days [5 days], 5 s) | Total (excluding household): $r = 0.35$ | 2−   |
| Questionnaire | Study population (n) for analysis | Comparison measure (type, placement, registration period [valid week], epoch length, cut points) | Results | Study quality and result rating |
|---------------|---------------------------------|------------------------------------------------|--------|-----------------------------|
| ZPAQ Modified English version Harris et al. \[57\] | 234 | Accelerometer (Actigraph GT1M, hip, 7 days [5 days], 5 s) | Total (including household): \( r = 0.34 \) | 2– |
|               | 121 | Pedometer (Yamax Digi-walker SW-200, hip, 7 days [5 days]) | Total (including household): \( r = 0.36 \) | 3– |

AAFQ Arizona Activity Frequency Questionnaire, AAS Active Australia Survey, ACLS-PALS Aerobic Center Longitudinal Study—Physical Activity Long Survey, ACLS-PASS Aerobic Center Longitudinal Study—Physical Activity Short Survey, Active-Q Web-based Physical Activity Questionnaire Active-Q, AUC area under the curve, BRHS British Regional Heart Study Physical Activity Questionnaire, CHAMPS Community Health Activities Model Program for Seniors, cpm counts per minute, \( \bar{d} \) change in the mean, DLW doubly labeled water, IPAQ2 Norfolk cohort of the European Prospective Investigation into Cancer (EPIC-Norfolk) Physical Activity Questionnaire, EPIC European Prospective Investigation into Cancer, GAPPAQ General Practice Physical Activity Questionnaire, \( \eta \) hours, IPAQ-E International Physical Activity Questionnaire for the Elderly, IPAQ-LF International Physical Activity Questionnaire—long-form, IPAQ-SF International Physical Activity Questionnaire—short-form, \( \kappa \) Kappa coefficient, \( kcal \) kilocalories, \( kg \) kilogram, \( kj \) kilojoules, LAPAQ Longitudinal Aging Study Amsterdam Physical Activity Questionnaire, LOA limits of agreement, \( log \) logarithm, \( LT \) leisure time, LTPA Modified Leisure Time Physical Activity Questionnaire, Modified Minnesota LTPA-Q Modified version of the Minnesota Leisure Time Physical Activity Questionnaire, MVPA moderate-to-vigorous physical activity, NC85+PAQ Newcastle 85+ Study Physical Activity Questionnaire, NPAQ Neighborhood Physical Activity Questionnaire, PA physical activity, PAEE physical activity energy expenditure, PAI physical activity index, PASB-Q Physical Activity and Sedentary Behavior Questionnaire, PASE Physical Activity Vital Sign Questionnaire, PHAS question Public Health Agency of Sweden physical activity question, \( \phi \) phi correlation coefficient, QAPPA Questionnaire d’Activité Physique pour les Personnes Âgées (Physical Activity Questionnaire for the Elderly), \( r \) Pearson correlation coefficient, \( R^2 \) R-squared, \( \rho \) Spearman correlation coefficient, \( s \) seconds, SGPALS Saltin-Grimby Physical Activity Level Scale, WHI-PAQ Women’s Health Initiative Physical Activity Questionnaire, WHS-AAPSA Women’s Health Study Physical Activity Questionnaire, ZPAQ Zutphen Physical Activity Questionnaire

*As described in Sect. 2.5, the quality of the individual study was evaluated per questionnaire and construct/dimension of PA and can be either very good (1), adequate (2), doubtful (3) or inadequate (4). Additionally, the reported results were rated (i.e., sufficient [+], insufficient [–]) as described in Sect. 2.4

1Results based on the first group (home based); second group (group exercise)
2Results based on second group (group exercise)
3Based on the reported results, we calculated the LOA using the formula \( \text{LOA} = \bar{d} \pm 1.96 \times s / \sqrt{2} \), where \( s \) is within-subject standard deviation (typical error) \[146\]
4The comparison was considered of high quality due to combined sensing and individual calibration
5Results based on different lower and upper accelerometer cut points: 760–2019 counts/min; 2020–4944 counts/min; 760–4944 counts/min
6Results based on both recreational and household-related PA. However, information about household-related PA was obtained from a previous data collection wave
7Results based on different accelerometer cut points: 760 cpm (vertical axis) \[140\]; 1041 cpm (vertical axis) \[138\]; 1952 cpm (vertical axis) \[136\]; 2690 cpm (triaxial) \[144\]
| Measurement property | Construct/dimension per questionnaire | Results | No. of studies (n) | GRADE Risk of bias | Inconsistency | Indirectness | Imprecision | Quality of evidence |
|----------------------|--------------------------------------|---------|-------------------|-------------------|---------------|--------------|-------------|-------------------|
| **Reliability**      |                                      |         |                   |                   |               |              |             |                   |
| Active-Q             | Swedish version                      | −       | 1 (148) [48]      | None –            | Seriousb       | None         | Moderate    |                   |
| MVPA                 |                                      |         |                   |                   |               |              |             |                   |
| Cambridge Index      | English version                      | −       | 1 (182) [93]      | None –            | None          | None         | High        |                   |
| Total                |                                      |         |                   |                   |               |              |             |                   |
| CHAMPS               | English versionc                     | −       | 4 (326) [49, 82, 91, 94] | None | None| None | None | High |
| Total                |                                      |         |                   |                   |               |              |             |                   |
| MVPA                 | +                                    | 3 (270) [82, 91, 94] | None | Serious | None | None | Moderate |
| CHAMPS               | Modified English version by Giles et al. | + | 1 (39) [84] | None – | None | None | Serious | Moderate |
| Total                |                                      |         |                   |                   |               |              |             |                   |
| MVPA                 | +                                    | 1 (42) [84] | None – | None | None | Serious | Moderate |
| Walking              |                                      |         |                   |                   |               |              |             |                   |
| CHAMPS               | Modified English version by Hekler et al. | + | 1 (748) [59] | Serious – | None | None | Moderate |
| Total                |                                      |         |                   |                   |               |              |             |                   |
| MVPA                 | −                                    | 1 (748) [59] | Serious – | None | None | Moderate |
| EPIC                 | English version                      | −       | 1 (182) [93]      | None –            | None          | None         | Moderate    |                   |
| Total                |                                      |         |                   |                   |               |              |             |                   |
| FPACQ                | Flemish version                      | +       | 1 (36) [87]       | None –            | None          | Serious      | Moderate    |                   |
| GPPAQ                | English version                      | +       | 1 (126) [44]      | None –            | None          | None         | High        |                   |
| Total                |                                      |         |                   |                   |               |              |             |                   |
| IPAQ-LF              | Serbian version                      | +       | 1 (660) [64]      | None –            | None          | None         | High        |                   |
| Walking              |                                      | −       | 1 (660) [64]      | None –            | None          | None         | High        |                   |
| IPAQ-SF              | Chinese version                      | +       | 1 (224) [89]      | None –            | None          | None         | High        |                   |
| Walking              |                                      | +       | 1 (224) [89]      | None –            | None          | None         | High        |                   |
| IPAQ-SF              | Japanese version                     | −       | 1 (325) [72]      | None –            | None          | None         | High        |                   |
| Walking              |                                      | −       | 1 (325) [72]      | None –            | None          | None         | High        |                   |
| IPEQ                 | English version                      | +       | 1 (50) [51]       | None –            | None          | None         | High        |                   |
| LAPAQ                | Dutch version                        | −       | 1 (86) [69]       | Serious –         | None          | None         | Moderate    |                   |
### Table 6 (continued)

| Measurement property | Construct/dimension per questionnaire | Results | No. of studies (n<sup>r</sup>) | GRADE | Risk of bias | Inconsistency | Indirectness | Imprecision | Quality of evidence |
|----------------------|--------------------------------------|---------|-------------------------------|-------|--------------|---------------|--------------|-------------|---------------------|
|                      |                                      |         |                               |       |              |               |              |             |                     |
| mLTPA-Q              | English version                      |         | MVPA                          | 1 (35) [54] | Serious – None | None | Serious | Low       |                     |
| Modified Baecke      | Dutch version                        |         | Total +                         | 1 (30) [86] | Serious – Serious | Serious | Serious | Very low |                     |
| OA-ESI               | English version                      |         | Total −                         | 2 (46) [95] | Serious None | None | None | Moderate |                     |
| PAQ-EJ               | Japanese version                     |         | Total                          | 1 (147) [96] | Serious – None | None | None | Moderate |                     |
| PAQ-B-Q              | English version                      |         | MVPA                          | 1 (147) [96] | Serious – None | None | None | Moderate |                     |
| PASE                 | All versions                          |         | Total +                        | 7 (1064) [66, 73, 76, 92, 79, 81, 97] | None | None | None | None | High |
| PASE                 | Chinese version                       |         | Total +                        | 2 (98) [66, 73] | None | None | None | None | High |
|                      | Walking                              |         | Total −                        | 1 (66) [73] | None – Serious | None | None | Moderate |
| PASE                 | English version                      |         | Total                          | 1 (254) [92] | Very serious – None | None | None | Low      |
| PASE                 | Italian version                      |         | Total                          | 1 (48) [79] | None – None | None | None | High      |
| PASE                 | Japanese version                     |         | Total                          | 1 (257) [97] | Serious – None | None | None | Moderate |
| PASE                 | Norwegian version                    |         | Total                          | 1 (327) [76] | None – None | None | None | High      |
| PASE                 | Persian version                      |         | Walking                        | 1 (278) [80] | None – None | None | None | High      |
| PASE                 | Turkish version                      |         | Total                          | 1 (80) [81] | None – None | None | None | High      |
| QAPPA                | French version                       |         | MVPA                          | 1 (225) [70] | Serious – None | None | None | Moderate |
| QAPSE                | French version                       |         | MVPA                          | 1 (44) [85] | Serious – None | None | None | Low      |
| SBAS                 | English version                      |         |                               |       |              |               |              |             |                     |
| Measurement property | Construct/dimension per questionnaire | Results | No. of studies (n) | GRADE |  |
|----------------------|--------------------------------------|---------|--------------------|--------|---|
|                      |                                      |         |                    | Risk of bias | Inconsistency | Indirectness | Imprecision | Quality of evidence |
|                      |                                      |         |                    | Very serious | – | None | None | Low |
|                      |                                      |         |                    | None | None | None | None | High |
|                      |                                      |         |                    | None | – | Seriousb | None | Moderate |
|                      |                                      |         |                    | None | – | Seriousb | None | Moderate |
|                      |                                      |         |                    | None | – | Seriousb | None | Moderate |
|                      |                                      |         |                    | None | – | None | None | High |
|                      |                                      |         |                    | None | – | None | None | Moderate |
|                      |                                      |         |                    | None | – | Serious | None | Moderate |
|                      |                                      |         |                    | Serious | None | None | Serious | Low |
|                      |                                      |         |                    | None | None | None | None | High |
|                      |                                      |         |                    | None | None | None | None | Moderate |
|                      |                                      |         |                    | None | – | None | Serious | Moderate |
|                      |                                      |         |                    | None | – | None | None | Moderate |
|                      |                                      |         |                    | None | – | None | Serious | Moderate |
|                      |                                      |         |                    | None | – | None | None | Moderate |
|                      |                                      |         |                    | None | – | None | None | High |
|                      |                                      |         |                    | None | – | None | None | High |
|                      |                                      |         |                    | None | – | None | None | High |

**Measurement error**

| CHAMPS                      | English versionc |
|-----------------------------|------------------|
| Total                       | –                | 1 (56) [49]    | None | – | None | None | High |
| EPIC                        | Totalb           | 1 (182) [93]   | None | – | None | None | High |
| LAPAQ                       | Total            | 1 (86) [69]    | None | – | None | None | High |
| PASE                        | Total            | 1 (66) [73]    | None | – | Seriousi | None | Moderate |

**Hypotheses testing for construct validity**

| AAFQ                        | English version |
|-----------------------------|------------------|
| Total                       | –                | 1 (450) [65]   | None | – | Seriousb | None | Moderate |
| AAS                         | English version |
| Total                       | +                | 2 (89) [55, 58]| Serious | None | None | Serious | Low |
| MVPA                        | –                | 2 (368) [58, 77]| None | None | None | None | High |
| Walking                     | –                | 1 (50) [58]    | None | – | None | Serious | Moderate |
| ACLS-PALS                   | English version |
| MVPA                        | +1               | 1 (71) [46]    | None | – | None | Serious | Moderate |
| ACLS-PASS                   | English version |
| MVPA                        | +1               | 1 (71) [46]    | None | – | None | Serious | Moderate |
| Active-Q                    | Swedish version |
| MVPA                        | –                | 1 (148) [48]   | None | – | Seriousb | None | Moderate |
| BRHS                        | English version |
| Total                       | +                | 1 (1377) [62]  | None | – | Serious | None | Moderate |
| Cambridge Index             | English version |
| Total                       | –                | 2 (1871) [53, 93]| None | None | None | None | High |
| CHAMPS                      | English versionc |
| Total                       | –                | 2 (134) [49, 91]| None | None | None | None | High |
| MVPA                        | –                | 1 (78) [91]    | Serious | – | None | Serious | Low |
| Measurement property | Construct/dimension per questionnaire | Results | No. of studies (n) | GRADE | Risk of bias | Inconsistency | Indirectness | Imprecision | Quality of evidence |
|----------------------|--------------------------------------|---------|--------------------|--------|-------------|--------------|-------------|-------------|-------------------|
|                      |                                      |         |                    |        |             |              |              |             |                   |
| CHAMPS               | Modified English version by Giles et al. |         |                    |        |             |              |              |             |                   |
| MVPA                 | − 1 (38) [84]                       | Very serious – None | Serious | Very low |
| Walking              | − 1 (44) [84]                       | None – None | None | Serious | Moderate |
| CHAMPS               | Modified English version by Hekler et al. |         |                    |        |             |              |              |             |                   |
| Total                | − 1 (850) [59]                      | None – None | None | None | High |
| MVPA                 | − 1 (850) [59]                      | None – None | None | None | High |
| EPAQ2                | Modified English version            |         |                    |        |             |              |              |             |                   |
| Total                | − 1 (1689) [53]                     | None – None | None | None | High |
| MVPA                 | − 1 (1689) [53]                     | None – None | None | None | High |
| EPIC                 | English version                     |         |                    |        |             |              |              |             |                   |
| Total                | − 1 (182) [93]                      | None – None | None | None | High |
| IPAQ-LF              | Modified Dutch version              |         |                    |        |             |              |              |             |                   |
| Total                | − 1 (196) [74]                      | Very serious – None | None | None | Low |
| MVPA                 | − 1 (196) [74]                      | None – None | None | None | High |
| IPAQ-SF              | All versions                        |         |                    |        |             |              |              |             |                   |
| Total                | − 4 (949) [50, 56, 72, 89]           | None – None | None | None | Moderate |
| Walking              | 3 (657) [56, 72, 89]                | None – None | None | None | High |
| IPAQ-SF              | Chinese version                     |         |                    |        |             |              |              |             |                   |
| Total                | − 1 (224) [89]                      | Very serious – None | None | None | Low |
| Walking              | − 1 (224) [89]                      | None – None | None | None | High |
| IPAQ-SF              | English version                     |         |                    |        |             |              |              |             |                   |
| Total                | − 1 (127) [56]                      | Very serious – None | None | None | Low |
| Walking              | − 1 (127) [56]                      | Very serious – None | None | None | Low |
| IPAQ-SF              | Japanese version                    |         |                    |        |             |              |              |             |                   |
| Total                | − 5 (306) [72]                      | Serious – None | None | None | Moderate |
| Walking              | − 1 (306) [72]                      | None – None | None | None | High |
| IPAQ-SF              | Portuguese version                  |         |                    |        |             |              |              |             |                   |
| Total                | − 1 (292) [50]                      | Very serious – | Serious | None | Very low |
Table 6 (continued)

| Measurement property | Construct/dimension per questionnaire | Results | No. of studies (n\(^a\)) | GRADE |
|----------------------|---------------------------------------|---------|--------------------------|-------|
|                      |                                       |         |                          | Risk of bias | Inconsistency | Indirectness | Imprecision | Quality of evidence |
| LAPAQ                | Dutch version                         | −       | 2 (1498) [63, 69]       | None         | None          | None         | None         | High               |
| mLTPA-Q              | English version                       | +       | 1 (32) [54]             | Very serious | –             | None         | Serious       | Very low           |
| Modified Baecke      | Dutch version                         | −       | 1 (28) [86]             | None         | –             | Very serious | Very serious  | Very low           |
| Modified Minnesota LTPA-Q | English version                    | −       | 1 (3975) [67]          | Serious      | –             | None         | None         | Moderate           |
| Walking              |                                       | −       | 1 (3975) [67]          | Serious      | –             | None         | None         | Moderate           |
| MVPA questions       | Swedish version                       | −       | 1 (948) [52]           | None         | –             | None         | None         | High               |
| NC85+PAQ             | English version                       | −       | 1 (337) [61]           | None         | –             | None         | None         | High               |
| NPAQ                 | German version                        | +       | 1 (58) [47]            | Very serious | –             | None         | Serious       | Very low           |
| Walking              |                                       | +       | 1 (58) [47]            | Very serious | –             | None         | None         | Very low           |
| OA-ESI               | English version                       | −       | 1 (327)                | Very serious | –             | Serious\(^b\) | None         | Very low           |
| PAQ-EJ               | Japanese version                      | +       | 1 (147) [96]           | Very serious | –             | None         | None         | Low                |
| MVPA                 |                                       | +       | 1 (147) [96]           | None         | –             | None         | None         | High               |
| PASB-Q               | English version                       | +       | 1 (32) [54]            | None         | –             | None         | Serious       | Moderate           |
| PASE                 | Dutch version                         | −       | 1 (21) [83]            | None         | –             | None         | Very serious  | Low                |
| PASE                 | English version                       | +       | 1 (78) [91]            | None         | –             | None         | Serious       | Moderate           |
| PASE                 | Japanese version                      | −       | 1 (200) [97]           | None         | –             | None         | None         | High               |
| PASE                 | Turkish version                       | +       | 1 (80) [81]            | Very serious | –             | None         | Serious       | Very low           |
| PAVS                 | English version                       | +       | 1 (269) [45]           | Very serious | –             | Very serious | None         | Very low           |
### Table 6 (continued)

| Measurement property | Construct/dimension per questionnaire | Results | No. of studies (n) | GRADE | Risk of bias | Inconsistency | Indirectness | Imprecision | Quality of evidence |
|----------------------|--------------------------------------|---------|-------------------|--------|--------------|---------------|--------------|-------------|---------------------|
| PHAS question        | Swedish version                      | Total   | − 1 (948) [52]    | Serious | –            | None          | None         | Moderate    |                     |
| Self-administered PAQ| Swedish version                      | Total   | − 2 (227) [75, 98]| Serious | None         | None          | None         | Moderate    |                     |
| SGPALs (LT question)| Swedish version                      | Total   | − 1 (948) [52]    | Serious | –            | None          | None         | Moderate    |                     |
| Single item on Recreational and Domestic Activity | English version | Total   | + 1 (1377) [62]   | Serious | –            | Serious<sup>b</sup> | None        | Low        |                     |
| Walking question     | Swedish version                      | Walking | − 1 (948) [52]    | Serious | –            | None          | None         | Moderate    |                     |
| WHI-PAQ              | English version                      | Total   | − 1 (450) [65]    | None    | –            | Very serious<sup>g</sup> | None        | Low        |                     |
| WHS-AASPA            | English version                      | MVPA    | − 1 (10115) [68]  | None    | –            | Serious<sup>b</sup> | None        | Moderate    |                     |
| ZPAQ English version | Total                               | − 1 (234) [57] | Serious | –            | None          | None         | Moderate    |           |                     |
| ZPAQ Modified English version | Total    | − 1 (234) [57]   | Serious | –            | None          | None         | Moderate    |           |                     |

### Responsiveness

| AAS | Total | − 1 (238) [77] | None | – | None | None | High |

<sup>AAPQ</sup> Arizona Activity Frequency Questionnaire, <sup>AAS</sup> Active Australia Survey, <sup>ACLS-PALS</sup> Aerobic Center Longitudinal Study—Physical Activity Long Survey, <sup>ACLS-PASS</sup> Aerobic Center Longitudinal Study—Physical Activity Short Survey, <sup>Active-Q</sup> Web-based Physical Activity Questionnaire Active-Q, <sup>BRHS</sup> British Regional Heart Study Physical Activity Questionnaire, <sup>CHAMPS</sup> Community Health Activities Model Program for Seniors, <sup>EPACQ</sup> Norfolk cohort of the European Prospective Investigation into Cancer (EPIC-Norfolk) Physical Activity Questionnaire, <sup>EPIC</sup> European Prospective Investigation into Cancer, <sup>FPACQ</sup> Flemish Physical Activity Computerized Questionnaire, <sup>GPPAQ</sup> General Practice Physical Activity Questionnaire, <sup>GRADE</sup> Grading of Recommendation, Assessment, Development and Evaluation, <sup>HEPA</sup> health enhancing physical activity, <sup>IPAQ-E</sup> International Physical Activity Questionnaire for the Elderly, <sup>IPAQ-LF</sup> International Physical Activity Questionnaire—long-form, <sup>IPAQ-SF</sup> International Physical Activity Questionnaire—short-form, <sup>IPEQ</sup> Incidental and Planned Exercise Questionnaire, <sup>LAPAQ</sup> Longitudinal Aging Study Amsterdam Physical Activity Questionnaire, <sup>LT</sup> leisure time, <sup>LTPA</sup> leisure time physical activity, <sup>min</sup> minutes, <sup>mLTPA-Q</sup> Modified Leisure Time Physical Activity Questionnaire, <sup>Modified Minnesota LTPA-Q</sup> Modified version of the Minnesota Leisure Time Physical Activity Questionnaire, <sup>MVPA</sup> moderate-to-vigorous physical activity, <sup>NC85+PAQ</sup> Newcastle 85+ Study Physical Activity Questionnaire, <sup>NPAQ</sup> Neighborhood Physical Activity Questionnaire, <sup>OA-ESI</sup> Older Adult Exercise Status Inventory, <sup>PA</sup> physical activity, <sup>PAQ</sup> Physical Activity Questionnaire, <sup>PAQ-EJ</sup> Physical Activity Questionnaire for Elderly Japanese, <sup>PASB-Q</sup> Physical Activity and Sedentary Behavior Questionnaire, <sup>PASE</sup> Physical Activity Scale for the Elderly, <sup>PASF</sup> Physical Activity Vital Sign Questionnaire, <sup>PHAS question</sup> Public Health Agency of Sweden physical activity question, <sup>QAPPA</sup> Questionnaire d’Activity Physique pour les Personnes Âgées (Physical Activity Questionnaire for the Elderly), <sup>QAPSE</sup> Questionnaire d’Activité Physique Saint-Etienne, <sup>SBAS</sup> Stanford Brief Activity Survey, <sup>SGPALS</sup> Saltin-Grimby Physical Activity Level Scale, <sup>WHI-PAQ</sup> Women’s Health Initiative Physical Activity Questionnaire, <sup>WHS-AASPA</sup> Women’s Health Study: Accelerometer Ancillary Study Physical Activity Form, <sup>ZPAQ</sup> Zutphen Physical Activity Questionnaire

Results are shown as sufficient (+) or insufficient (−) measurement properties depending on scores and rating obtained from Tables 4 and 5, as well as from Electronic Supplementary Material Table S3 and Electronic Supplementary Material Table S4. Results are shown for the overall construct of the questionnaire (e.g., total PA, total PAEE, total LTPA), also called ‘total’ score, and for the subdimensions MVPA and walking.

<sup>a</sup>Total number of participants across all studies.
of the WHI-PAQ. However, the evidence cannot be considered together because the results for hypotheses testing for construct validity were based on both recreational and household-related PA [65], but results for reliability were reported separately for these domains [88]. Finally, the different measurement properties were assessed across a variety of language versions (e.g., reliability of the IPAQ-LF was assessed for the Serbian version but information about hypotheses testing for construct validity was available only for other languages).

Regarding the overall construct, there was at least low-quality evidence that versions of six questionnaires (FPACQ, IPAQ-LF, IPAQ-SF, IPEQ, PASE, WHI-PAQ) showed sufficient reliability and versions of five questionnaires (AAS, BRHS, PAQ-EJ, PASE, Single item on Recreational and Domestic Activity) showed sufficient hypotheses testing for construct validity. Versions of two questionnaires (mLTPA-Q, PAVS) showed also sufficient hypotheses testing for construct validity. Overall, corresponding versions of two questionnaires showed both sufficient reliability and hypotheses testing for construct validity, namely the PASE (i.e., English version, Turkish version) concerning the assessment of total PA, and the PASB-Q (English version) concerning the assessment of MVPA. The quality of evidence for these results ranged from very low to high.

4 Discussion

The present review is an update of a previous review published in 2010 [28] and aimed to evaluate the measurement properties of all available self-administered PA questionnaires for older adults and to provide recommendations for
the most-qualified questionnaires based on the quality of the body of evidence.

The overall evidence of measurement properties for questionnaires assessing PA in older adults is often of low to moderate quality. None of the included questionnaires provided evidence for all relevant measurement properties (reliability, measurement error, hypotheses testing for construct validity, responsiveness). For versions of 14 questionnaires (Active-Q, Cambridge Index, CHAMPS, EPIC, FPACQ, IPAQ-SF, LAPAQ, mLTPA-Q, Modified Baecke, OA-ESI, PAQ-EJ, PASB-Q, PAHE, Self-administered PAQ) combined evidence (i.e., on the same version) for reliability and hypotheses testing for construct validity was available. Of these, there was very-low-to-high-quality evidence of both sufficient reliability and hypotheses testing for construct validity for one questionnaire [PASE (English version, Turkish version)] regarding the measurement of total PA, and for another questionnaire [PASB-Q (English version)] regarding the measurement of MVPA. These two questionnaires also met our criteria for sufficient content validity.

The quality of individual studies was often very good or adequate. Only few studies used inadequate statistical approaches such as Pearson or Spearman correlation coefficients for reliability analyses [36, 102]. Although the ICC is the preferred method [36], a low coefficient does not necessarily indicate low reliability. Correlation coefficients are susceptible to several influences such as the variability of PA behaviors (heterogeneity), differences in the shape of the distribution and non-linearity [103, 104]. For example, any serious lack of variability in the sample (e.g., one may consider PA levels of the very old or other subgroups) could have reduced the observed coefficient. Therefore, we recommend considering the limitations of correlation coefficients when interpreting results concerning both reliability and hypotheses testing for construct validity.

The choice of the comparison measure and use of different intensity levels of PA often reduced the quality of the individual study. For example, both accelerometers and pedometers were often used to test hypotheses for construct validity. Although pedometers can be considered as the reference to measure daily steps, they are unable to capture frequency, duration and intensity of PA [105]. Thus, they can be considered as the best choice to evaluate walking but not MVPA or total PA measured by a questionnaire [e.g., IPAQ-SF (Portuguese version) [50]]. In other studies (e.g., on the Modified Minnesota LTPA-Q [67]), moderate PA measured by the questionnaire was compared to total PA from the accelerometer (including also light and vigorous PA). In this case, the best comparison measure would also be moderate PA due to highest similarity to the construct [106]. The need to choose comparison measures as similar as possible was also demonstrated by studies using novel statistical approaches to handle accelerometer data [107]. Specifically, it was shown that the correlation was much lower for distal (light and vigorous PA), compared to proximal PA intensity levels. However, calculating the time spent in different intensity levels using accelerometer data is clearly challenging because of the dependency on intensity-specific cut points [106].

We observed considerable heterogeneity in the collection, processing and reporting of accelerometer data among individual studies. Although most studies considered a 7-day registration period, a broad range of different cut points, epoch lengths (e.g., 5–60 s) and criteria for a valid week (e.g., 1–14 days) were used. These decision rules will impact the obtained PA estimates [108]. Several studies (e.g., on the AAS [55], mLTPA-Q and PASB-Q [54]) did not use population-specific intensity cut points which may result in an under- or overestimation of time spent in different intensity levels [109]. Another shortcoming was that not all studies reported all decisions such as sampling frequency, non-wear definition and use of filters [110]. Therefore, the use of standards for the design of studies on measurement properties of PA questionnaires (e.g., COSMIN study design checklist) [111, 112] is highly recommended. Likewise, experts in the field emphasized the need for standards for using and reporting accelerometer data [106, 113, 114]. However, despite some attempts [110, 115, 116], it seems that there is currently no consensus on the most appropriate use of accelerometers in older adults [117].

Not only the comparison measure, but also PA questionnaires themselves have important limitations which must be considered. Reporting errors can result from problems in recalling the duration of activities, differences in the interpretation of their intensity [38], social desirability [118] or telescoping of events [119]. Moreover, the accuracy of the recall is influenced by factors such as age, weight status, education and mental health [120, 121]. This is problematic when using questionnaires to define dose–response patterns with health outcomes and strongly reduces the comparability of results among studies with different populations. Hence, it is important to consider advantages and disadvantages of each measurement instrument (e.g., questionnaire, accelerometer, pedometer) when selecting a tool for a particular purpose [11].

Many studies used MET values to estimate the energy costs of activities [i.e., to obtain (rates of) PAEE]. These values are multiples of an adult’s average resting metabolic rate (energy expenditure at rest) [122] and are usually obtained from a compendium of physical activities [123, 124]. However, as emphasized by the authors [124], the compendium does not provide specific energy costs of activities for older adults. So far, there exists no comparable list for older adults although recent studies demonstrated that MET values obtained from daily activities of older adults differed considerably to those listed in the compendium [125], including...
a strong inter-subject variability and a decrease in resting metabolic rate with age [126]. Therefore, the error associated with the universal application of MET values will likely increase when values from a different population will be applied to older adults [127]. It follows that experts in the field have called for studies of subgroup-specific MET values (e.g., regarding age, sex, body mass, disease status) and questioned the accuracy of conventional MET values to describe the energy costs of activities in older adults [128].

After combining the studies from the previous review and our update, we observed serious shortcomings associated with the quality of the body of evidence. First, only one study assessed the responsiveness of a PA questionnaire. Questionnaires are commonly applied in intervention studies in older adults [12] and sufficient responsiveness is indispensable to accurately measure changes of PA over time [36]. Secondly, only three studies [49, 65, 83] used DLW as a comparison method although (rates of) PAEE was often estimated. Furthermore, for most questionnaire versions, there was only a single study available. This often decreased the overall quality of evidence, especially when this study was of lower quality, the sample size was small or the sample was too restricted (e.g., only women). Finally, we also observed inconsistency in the results when trying to summarize the results from multiple studies on different language versions (e.g., reliability of the Chinese and Japanese version of the IPAQ-SF [72, 89]). The varying results (sufficient, insufficient) of different language versions can partly be explained by cultural adaptations and differences in the conceptualisation and interpretation of PA [129]. If inconsistency in the results is observed and/or studies on the cross-cultural validity revealed important differences between the versions, these language versions should be treated separately. Despite careful cross-cultural adaptation, sufficient measurement quality in one version does not guarantee the same quality for other languages and populations [18, 33].

More than half (i.e., 22 of 40) of all questionnaires met our principal criteria for sufficient content validity. Older adults engage in less exercise-related behaviors; whereas low-to-moderate-intensity activities such as walking and gardening become more prevalent [130]. Nevertheless, these light activities are under-represented in available PA questionnaires for older adults and there is a lack of consensus on the conceptualisation of PA in this population [131, 132]. Light activities are less reliably reported than higher intensity activities which outlines a challenge for the measurement of PA in older adults using self-reports [38]. We recommend that the included questionnaires are further appraised with respect to these considerations, as suggested earlier [131, 132].

Whenever assessed, absolute measurement errors were large (e.g., > 2000 min for total PA of the LAPAQ [69]). Although researchers may define a different MIC, it seems that the ability of questionnaires to detect important changes of PA beyond measurement error is limited [36]. Moreover, we observed a substantial lack of absolute agreement between the questionnaire and the comparison measure (usually the accelerometer), such as for the mLTPA-Q (LOA = −223 to 262 min per week) [54]. This means that the two instruments do not assess the same absolute dose of PA. However, because of a missing gold standard for the measurement of PA [25, 34], the interpretation of these absolute agreements for construct validity is flawed. We simply do not know what the true dose of PA was. Absolute agreements can only be interpreted when a reference method is available, for instance, when total EE estimated by the questionnaire or accelerometer is compared to the accepted standard of DLW [11].

Of the overall body of evidence, versions of the CHAMPS, IPAQ-SF and PASE were assessed the most often. A great number of results were based on low- or very-low-quality evidence which means that we cannot be confident in the observed measurement properties. Lower quality of the evidence was often related to the reliance on single studies with serious shortcomings in quality, sample size or indirectness. Some results (e.g., for total PA, MVPA) were slightly below [e.g., reliability of the Self-Administered PAQ (Swedish version) [90], hypotheses testing for construct validity of the CHAMPS (English version) [91] and PASE (Dutch version) [83]] or above [e.g., reliability of the IPAQ-LF (Serbian version) [64], hypotheses testing for construct validity of the PAVS [45]] our acceptance levels. These results, if based on high-quality evidence, should not be entirely disregarded when selecting a questionnaire to measure PA in older adults.

### 4.1 Recommendations for Choosing a Questionnaire

The purpose of the study guides the choice of the questionnaire. In addition to earlier recommendations [36], we suggest the following for the selection of a questionnaire to measure PA in older adults:

- Choose a questionnaire which provides sufficient content validity for a particular purpose and evaluate the content of the questionnaire before using it. For instance, we observed noticeable differences not only in format but also in the obtained information (e.g., frequency, duration or intensity may not be obtained for all included activities). Some attempts regarding the evaluation of content validity have been made previously [131, 132]. If the content validity is insufficient, evaluation of further measurement properties is irrelevant [18].

- When measuring total PA, the questionnaire should include all relevant domains of PA (household, recrea-
Physical Activity Questionnaires for Older Adults

- The questionnaire should include at least parameters of frequency and duration of PA and a representative list of light-to-moderate activities which are more frequently performed by older adults [130].
- The choice of the recall period depends on several factors such as cognitive demands, intended construct (e.g., usual PA, lifetime PA) and the intensity of activities [38]. For example, experts in the field have called for improvements in PA self-reports by reducing the recall period (e.g., multiple 24 h recalls) [38]. However, until high-quality evidence for superior recall periods is available, we recommend that the recall period should capture at least an entire week when using a single administration.
- Due to serious differences in PAEE in older adults and the lack of age-specific energy costs of activities [128], we recommend not using MET values. Instead, raw units such as total time or time spent in different intensity levels can be used.
- It is important to choose a questionnaire with both sufficient reliability and hypotheses testing for construct validity in the target population (e.g., older adults). Unfortunately, this was not often the case in the past [12]. If the questionnaire is used to measure change in PA, sufficient responsiveness is required.
- We recommend considering modified versions of questionnaires as separate instruments, especially when inconsistent results were observed and/or studies on cross-cultural validity showed critical differences [33]. This may also be the case for different language versions when questions are replaced and/or the wording is changed during cultural adaptations. The same questionnaire may not be equally qualified in different settings and populations of older adults.
- If evidence for the measurement properties of a particular modified version is missing, we recommend performing pilot tests.

Not only researchers but also healthcare professionals (e.g., practitioners) are interested in the measurement of PA using questionnaires. In this setting, our recommendations can be followed because they represent general recommendations for the use of questionnaires in order to improve the quality of the measurement. However, further aspects such as clinical feasibility, mode of administration and linkage to electronic record systems should be considered [16]. For instance, clinical feasibility was not part of this review, although included in another review evaluating PA questionnaires in healthcare settings [17]. We propose the following additional recommendations for the use of PA questionnaires in healthcare settings:

- Because the administration should be integrated into the daily workflow, we recommend considering the length of the questionnaire (i.e., time to completion). For this, the PASB-Q may serve as a suitable tool with sufficient measurement properties.
- Healthcare professionals should be aware that the mode of administration likely impacts the obtained results (e.g., interviewer- versus self-administered) [133].
- PA questionnaires show inevitable limitations (e.g., reporting errors due to social desirability or difficulties in recalling the duration of activities) [38, 118] and in this review, only limited high-quality evidence for sufficient measurement properties and usually large measurement errors were observed. Therefore, we recommend bearing in mind that the assessment of PA on the individual level (e.g., determining the PA level of a single patient) is likely associated with large measurement errors.

In general, we recommend using questionnaires with sufficient content validity and at least low-quality evidence for sufficient measurement properties (for at least reliability and hypotheses testing for construct validity) [33]. This was the case for the English versions of the PASE, concerning the assessment of total PA, and PASB-Q, concerning the assessment of MVPA. Also, the Turkish version of the PASE revealed sufficient measurement properties, but the results of hypotheses testing for construct validity were based on very low-quality evidence. The PASE measures PA over the past 7 days and provides an overall weighted score but does not intend to measure EE [92]. The PASB-Q obtains time spent in MVPA in a typical week [54]. It is a brief measure and does not provide separate information for different domains of PA.

We recommend not using questionnaires with insufficient content validity and/or high-quality evidence for insufficient measurement properties (for at least reliability and hypotheses testing for construct validity) [33]. Hence, we recommend not using the Cambridge Index (English version) for total PA, CHAMPS (English version) for total PA, EPIC (English version) for total PA and the IPAQ-SF (Japanese version) for walking. Several more questionnaires showed insufficient content validity (see Sect. 3.3.1) and would not be recommended. However, future studies performing a comprehensive evaluation of the content validity of these questionnaires are needed in order to be able to give solid recommendations based on only content validity.
4.2 Limitations and Strengths of this Review

We used standardized criteria [36] for the rating of measurement properties which are in accordance with our previous reviews [18, 19, 28–30]. However, the common problem when using cut points like this is dichotomization and loss of information. This can be seen in the results when questionnaires showed results just below or above the cut point. Although one may consider both types of results as acceptable, our cut points represent minimal important criteria for sufficient measurement properties.

The quality of evidence for the measurement properties of many (versions of) questionnaires was limited. Moreover, we observed considerable heterogeneity in the use, analysis and reporting of accelerometer data. We did not use standardized criteria to include these methodological aspects into our quality ratings. Although attempts have been made for certain devices [110], a consensus on the most appropriate use of accelerometers in older adults is lacking [114, 117].

Future reviews may be able to include different decision rules such as epoch length, filter and valid wear time into their assessment. Furthermore, different researchers were involved in the previous review and this update which could have influenced the quality (e.g., level of agreement).

The lack of a gold standard to measure PA resulted in the use of various proxy measures (e.g., accelerometers, pedometers, diaries) to test hypotheses for construct validity. The measurement quality of these instruments varies [25], which means that the construct validity of a PA questionnaire is assessed by comparisons to instruments also showing shortcomings in construct validity. This is a serious problem for any study addressing measurement properties of PA measurement instruments. However, we tried to include differences in the measurement quality of the comparison measure in our quality assessment.

The strengths of this review are that it expands the former evidence [28] and provides the latest recommendations for the use of PA questionnaires in older adults. Data extraction and all assessments were performed independently by at least two researchers. We applied transparent methodological guidelines [33, 36, 43] to assess each result with the same set of criteria as well as to evaluate the quality of individual studies and the overall body of evidence. Finally, we presented all results of the included studies in our tables and, therefore, researchers in the field are invited to discuss the results with regards to their own expertise, probably assigning different criteria.

4.3 Recommendations for Future Research

In 2010 [28], it was recommended that a study should provide a detailed description of the sample and should include at least 50 participants. Such a sample size was considered acceptable to address reliability and hypotheses testing for construct validity [103]. We found that newer studies followed these recommendations. Future studies evaluating the quality of PA questionnaires in older adults should consider the following:

- Because the remaining measurement properties (e.g., reliability, hypotheses testing for construct validity) should only be addressed when the questionnaire has sufficient content validity, we recommend evaluating the content validity of the most promising questionnaires.
- Because many results were based on low-quality evidence and, hence, confidence in these is limited, we recommend evaluating questionnaires for which there is currently only low- or very-low-quality evidence available.
- Because for the majority of questionnaires (>60%) no combined evidence for reliability and hypotheses testing for construct validity was available, we recommend evaluating questionnaires for which there is currently at least low-quality evidence for sufficiency in one measurement property but information on others is missing.
- We found that many questionnaires were available in only one language (usually English, e.g., PASB-Q). Therefore, we recommend evaluating different language versions of the most promising questionnaires (including correct translation and cultural adaptation).
- Because there was a clear lack of studies assessing responsiveness, we recommend assessing the responsiveness of the most promising questionnaires.
- Because many different (versions of) questionnaires exist, we recommend improving the most promising questionnaires rather than developing new ones [19].
- Because the way we handle accelerometer data influences derived PA estimates [108], we recommend not only working on consensus-based standards but also providing a transparent description of accelerometer data collection and processing rules.
- Due to the observed heterogeneity in the design of studies, we recommend using standards [e.g., COSMIN (http://www.cosmin.nl)] for the study design and evaluation of measurement properties of PA measurement instruments.

5 Conclusions

Since our review in 2010 [28], many new PA questionnaires for older adults have been developed. All evidence combined, no questionnaire showed sufficient content validity, reliability, hypotheses testing for construct validity and responsiveness, due to the lack of studies. For most questionnaires, only one study was available, and responsiveness
was usually not included in the assessment. The quality of the body of evidence was often reduced. However, two questionnaires (PASB-Q, PASE) can be recommended although the quality of different language versions varied. Because an accepted gold standard to measure PA is missing [34], it is difficult to select the best comparison measure to test hypotheses for construct validity of a questionnaire. We concurred with experts in the field that researchers should consider strengths and weaknesses of each instrument, and select the best available comparison measure for a particular construct measured by the questionnaire [11, 134]. For the future, we recommend using existing questionnaires without performing minor modifications to the questionnaire. Rather than developing new questionnaires, we should work on improving existing ones.

Acknowledgements Open access funding was provided by the University of Graz.

Data Availability Statement Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Compliance with Ethical Standards

Funding No sources of funding were used to assist in the preparation of this article.

Conflict of interest Matteo Sattler, Johannes Jaunig, Christoph Tösch, Estelle Watson, Mireille van Poppel, and Pavel Dietz declare that they have no conflicts of interest relevant to the content of this review. Lidwine Mokkink is one of the developers of the Quality Assessment of Physical Activity Questionnaire (QAPAQ), and the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist and methodology, and the institute of which Lidwine Mokkink is a part receives royalties for one of the references [103] cited in this review.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, and indicate if changes were made.

References

1. World Health Organization. World report on ageing and health. Geneva: WHO; 2015.
2. Physical Activity Guidelines Advisory Committee. 2018 Physical Activity Guidelines Advisory Committee Scientific Report. Washington, DC: Department of Health and Human Services; 2018.
3. Vagetti GC, Barbosa Filho VC, Moreira NB, Oliveira BD, Mazzardo O, Campos WD. Association between physical activity and quality of life in the elderly: a systematic review, 2000–2012. Braz J Psychiatry. 2014;36:76–88. https://doi.org/10.1590/1516-4446-2012-0895.
4. Park S-H, Han KS, Kang C-B. Effects of exercise programs on depressive symptoms, quality of life, and self-esteem in older people: a systematic review of randomized controlled trials. Appl Nurs Res. 2014;27:219–26. https://doi.org/10.1016/j.apnr.2014.01.004.
5. Prakash RS, Voss MW, Erickson KI, Kramer AF. Physical activity and cognitive vitality. Annu Rev Psychol. 2015;66:769–97. https://doi.org/10.1146/annurev-psych-010814-015249.
6. Chase J-AD, Philips LJ, Brown M. Physical activity intervention effects on physical function among community-dwelling older adults: a systematic review and meta-analysis. J Aging Phys Act. 2017;25:149–70. https://doi.org/10.1123/japa.2016-0040.
7. Gallaway PJ, Miyake H, Buchowski MS, Shimada M, Yoshitake Y, Kim AS, Hongu N. Physical activity: a viable way to reduce the risks of mild cognitive impairment, Alzheimer’s disease, and vascular dementia in older adults. Brain Sci. 2017. https://doi.org/10.3390/brainsci7020022.
8. Catalan-Matamoros D, Gomez-Conesa A, Stubbs B, Vancampfort D. Exercise improves depressive symptoms in older adults: an umbrella review of systematic reviews and meta-analyses. Psychiatry Res. 2016;244:202–9. https://doi.org/10.1016/j.psychres.2016.07.028.
9. Rhyner KT, Watts A. Exercise and depressive symptoms in older adults: a systematic meta-analytic review. J Aging Phys Act. 2016;24:234–46. https://doi.org/10.1123/japa.2015-0146.
10. Hopin D, Roche F, Gremeaux V, Chatteraj C-J, Oriol M, Gaspoz J-M, et al. Even a low-dose of moderate-to-vigorous physical activity reduces mortality by 22% in adults aged ≥ 60 years: a systematic review and meta-analysis. Br J Sports Med. 2015;49:1262–7. https://doi.org/10.1136/bjsports-2014-094306.
11. Strath SJ, Kaminsky LA, Ainsworth BE, Ekelund U, Freedson PS, Gary RA, et al. Guide to the assessment of physical activity: clinical and research applications: a scientific statement from the American Heart Association. Circulation. 2013;128:2259–79. https://doi.org/10.1161/01.cir.0000435708.67487.da.
12. Falck RS, McDonald SM, Beets MW, Brazendale K, Liu-Ambrose T. Measurement of physical activity in older adult interventions: a systematic review. Br J Sports Med. 2016;50:464–70. https://doi.org/10.1136/bjsports-2014-094413.
13. 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:e1077–86. https://doi.org/10.1016/S2214-109X(18)30357-7.
14. Loyen A, van Heck L, Verloigne M, Hendrikse 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. https://doi.org/10.1186/s12966-016-0398-2.
15. Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. Res Q Exerc Sport. 2000;71(Suppl 2):1–14. https://doi.org/10.1080/0270131994473780.
16. Heron N, Tully MA, McKinley MC, Cupples ME. Physical activity assessment in practice: a mixed methods study of GPQPAQ use in primary care. BMC Fam Pract. 2014;15:11. https://doi.org/10.1186/1471-2296-15-11.
17. Lobelo F, Rohm Young D, Sallis R, Garber MD, Billinger SA, Duperly J, et al. Routine assessment and promotion of physical activity in healthcare settings: a scientific statement from the
Activity Questionnaire in German older adults. Meas Phys Educ Exerc Sci. 2018;22:61–9. https://doi.org/10.1080/109167X.2017.1383256.

48. Bonn SE, Bergman P, Trolle Lagerros Y, Sjölander A, Bälter K. A validation study of the web-based physical activity questionnaire active-Q against the GENEa accelerometer. JMIR Res Protoc. 2015;4:e86. https://doi.org/10.1101/resprot.3896.

49. Colbert LH, Matthews CE, Havighurst TC, Kim K, Schoeller DA. Comparative validity of physical activity measures in older adults. Med Sci Sports Exerc. 2011;43:867–76. https://doi.org/10.1249/MSS.0b013e3181f7162.

50. Colpani V, Spritzer PM, Lodì AP, Dorigo GG, de Miranda IAS, Hahn LB, et al. Physical Activity in climacteric women: comparison between self-reporting and pedometer. Rev Saude Publica. 2014;48:258–65. https://doi.org/10.1590/S0034-8910.2014048004765.

51. Delbaere K, Hauer K, Lord SR. Evaluation of the incidental and planned activity questionnaire (IPEQ) for older people. Br J Sports Med. 2010;44:1029–34. https://doi.org/10.1136/bjsm.2009.060350.

52. Ekblom Ö, Ekblom-Bak E, Martin KR, Hardy R, Ekelund U, Kuh D, et al. Comparison of the EPIC Physical Activity Questionnaire with combined heart rate and movement sensing in a nationally representative sample of older British adults. PLoS One. 2014;9:e87085. https://doi.org/10.1371/journal.pone.00870 85.

53. España-Romero V, Golubic R, Martin KR, Hardy R, Ekelund U, Kuh D, et al. Concurrent and predictive validity of physical activity measurement items commonly used in clinical settings—data from SCAPIS pilot study. BMC Public Health. 2015;15:978. https://doi.org/10.1186/s12889-015-2316-y.

54. Freene N, Waddington G, Chesworth W, Davey R, Cochrane T. Are Active Questionnaires for the Elderly in Chinese population. J Rehabil Med. 2012;44:462–5. https://doi.org/10.2340/16501977-0953.

55. Fowles JR, O'Brien MW, Wojcik WR, d'Entremont L, Shields CA. A pilot study: validity and reliability of the CSEP-PATH PASB-Q and a new leisure time physical activity questionnaire to assess physical activity and sedentary behaviours. Appl Physiol Nutr Metab. 2017;42:677–80. https://doi.org/10.1113/apnn-2016-0412.

56. Grimm EL, Swartz AM, Hart T, Miller NE, Strath SJ. Comparisons of physical activity in older adults. J Aging Phys Act. 2012;20:64–79.

57. Harris TJ, Owen CG, Victor CR, Adams R, Ekelund U, Cook DG. A comparison of questionnaire, accelerometer, and pedometer: measures in older people. Med Sci Sports Exerc. 2009;41:1392–402. https://doi.org/10.1249/MSS.0b013e31819b3533.

58. Heesch KC, Hill RL, van Uffelen JGZ, Brown WJ. Are Active Australia physical activity questions valid for older adults? J Sci Med Sport. 2011;14:233–7. https://doi.org/10.1016/j.jsams .2010.11.004.

59. Hekler EB, Buman MP, Haskell WL, Conway TL, Cain KL, Saliris JF, et al. Reliability and validity of CHAMPS self-reported sedentary-to-vigorous intensity physical activity in older adults. J Phys Act Health. 2012;9:225–36.

60. Hurtig-Wennlöf A, Hagströmer M, Olsson LA. The International Physical Activity Questionnaire modified for the elderly: aspects of validity and feasibility. Public Health Nutr. 2010;13:1847–54. https://doi.org/10.1017/S136894650000157.

61. Innerd P, Catt M, Collerton J, Davies K, Trenell M, Kirkwood TBL, Jagger C. A comparison of subjective and objective measures of physical activity from the Newcastle 85+ study. Age Ageing. 2015;44:691–4. https://doi.org/10.1093/ageing/afv062.

62. Jefferis BJ, Sartini C, Ash S, Lennon LT, Wannamethee SG, Whincup PH. Validity of questionnaire-based assessment of sedentary behaviour and physical activity in a population-based cohort of older men; comparisons with objectively measured physical activity data. Int J Behav Nutr Phys Act. 2016;13:14. https://doi.org/10.1186/s12966-016-0338-1.

63. Koohsari SM, van Rooy JF, Cepeda M, Tiemeier H, Franco OH, Schoufour JD. Physical activity derived from questionnaires and wrist-worn accelerometers: comparability and the role of demographic, lifestyle, and health factors among a population-based sample of older adults. Clin Epidemiol. 2018;10:1–16. https:// doi.org/10.2147/CLEP.S147613.

64. Koolhaas CM, van Rooij FJ, Cepeda M, Tiemeier H, Franco OH, Schoufour JD. Physical activity derived from questionnaires and wrist-worn accelerometers: comparability and the role of demographic, lifestyle, and health factors among a population-based sample of older adults. Clin Epidemiol. 2018;10:1–16. https://doi.org/10.2147/CLEP.S147613.

65. Maltais F, Pantelić S, Trajković N, Jorgić B, Sporiš G, Bratić M. Reliability of the Serbian version of the International Physical Activity Questionnaire for older adults. Clin Interv Aging. 2014;9:581–7. https://doi.org/10.2147/CIA.S57379.

66. Neuhausser ML, Di C, Tinker LF, Thomson C, Sternfeld B, Mosavari-Rahmani Y, et al. Physical activity assessment: biomarkers and self-report of activity-related energy expenditure in the WHI. Am J Epidemiol. 2013;177:576–85. https://doi.org/10.1093/aje/ kws269.

67. Ngai SPC, Cheung RTH, Lam PL, Chiu JKW, Fung EYH. Validation and reliability of the Physical Activity Scale for the Elderly in Chinese population. J Rehabil Med. 2012;44:462–5. https://doi.org/10.2340/16501977-0953.

68. Sabia S, van Hees VT, Shipley MJ, Trenell MI, Hagger-Johnson G, Elbaz A, et al. Association between questionnaire- and accelerometer-assessed physical activity: the role of sociodemographic factors. Am J Epidemiol. 2014;179:781–90. https://doi.org/10.1093/aje/kwt330.

69. Shimroa EJ, Cook NR, Manson JE, Buring JE, Rimm EB, Lee I-M. Comparison of self-reported and accelerometer-assessed physical activity in older women. PLoS One. 2015;10:e0145950. https://doi.org/10.1371/journal.pone.0145950.

70. Siebeling L, Wiebers S, Beem L, Pahan MA, Ter Riet G. Validity and reproducibility of a physical activity questionnaire for older adults: questionnaire versus accelerometer for assessing physical activity in older adults. Clin Epidemiol. 2012;4:171–80. https://doi.org/10.2147/CLEP.S30848.

71. de Souto Barreto P. Construct and convergent validity and repeatability of the Questionnaire d’Activité Physique pour les Personnes Âgées (QAPPA), a physical activity questionnaire for the elderly. Public Health. 2013;127:844–53. https://doi.org/10.1016/j.puhe.2012.10.018.

72. Taylor-Piliae RE, Fair JM, Haskell WL, Varady AN, Iribarren C, Hlatky MA, et al. Validation of the Stanford Brief Activity Survey: examining psychological factors and physical activity levels in older adults. J Phys Act Health. 2010;7:87–94.

73. Tomioaka K, Iwamoto J, Saeki K, Okamoto N. Reliability and validity of the International Physical Activity Questionnaire (IPAQ) in elderly adults: the Fujiwara-kyo study. J Epidemiol. 2011;21:459–65. https://doi.org/10.2188/jeaJE20110003.

74. Vaughan K, Miller WC. Validity and reliability of the Chinese translation of the Physical Activity Scale for the Elderly (PASE). Disabil Rehabil. 2013;35:191–7. https://doi.org/10.3109/09638 288.2012.690498.

75. Winckers ANE, Mackenbach JD, Compernelle S, Nicolaou M, van der Ploeg HP, de Bourdeaudhuij I, et al. Educational differences in the validity of self-reported physical activity. BMC Public Health. 2015;15:1299. https://doi.org/10.1186/s1288 9-015-2656-7.

76. Norman A, Bellocco R, Bergström A, Wolk A. Validity and reproducibility of self-reported total physical activity—differences by...
relative weight. Int J Obes Relat Metab Disord. 2001;25:682–8. 
https://doi.org/10.1038/sj.ijo.0801597.

76. Lolland N. Reliability of the physical activity scale for the elderly (PASE). Eur J Sport Sci. 2002;2:1–12. 
https://doi.org/10.1016/S1475-191X(02)00072-5.

77. Vanderanotte C, Duncan MJ, Stanton R, Rosenkranz RR, Capernichio CM, Rebar AL, et al. Validity and responsiveness to change of the Active Australia Survey according to gender, age, BMI, education, and physical activity level and awareness. BMC Public Health. 2019;19:407. 
https://doi.org/10.1186/s12889-019-6717-1.

78. Cleland C, Ferguson S, Ellis G, Hunter RF. Validity of the International Physical Activity Questionnaire (IPAQ) for assessing moderate-to-vigorous physical activity and sedentary behaviour of older adults in the United Kingdom. BMC Med Res Methodol. 2018;18:176. 
https://doi.org/10.1186/s12874-018-0642-3.

79. Covotta A, Gagliardi M, Berardi A, Maggi G, Pierelli F, Mollica R, et al. Physical Activity Scale for the Elderly: translation, cultural adaptation, and validation of the italian version. Curr Gerontol Geriatr Res. 2018;2018:8294568. 
https://doi.org/10.1155/2018/8294568.

80. Keikavossi-Arani L, Salehi L. Cultural adaptation and psychometric adequacy of the Persian version of the physical activity scale for the elderly (P-PASE). BMC Res Notes. 2019;12:555. 
https://doi.org/10.1186/s13104-019-4591-7.

81. Ayvat E, Kiliç M, Kirdi N. The Turkish version of the Physical Activity Scale for the Elderly (PASE): its cultural adaptation, validation, and reliability. Turk J Med Sci. 2017;47:908–15. 
https://doi.org/10.3906/sag-1605-7.

82. Stewart AL, Mills KM, King AC, Haskell WL, Gillis D, Ritter PL. CHAMPS physical activity questionnaire for older adults: outcomes for interventions. Med Sci Sports Exerc. 2001;33:1126–41.

83. Schuit AJ, Schouten EG, Westerterp KR, Saris WH. Validity of the Physical Activity Scale for the Elderly (PASE): according to energy expenditure assessed by the doubly labeled water method. J Clin Epidemiol. 1997;50:541–6.

84. Giles K, Marshall AL. Repeatability and accuracy of CHAMPS as a measure of physical activity in a community sample of older Australian adults. J Phys Act Health. 2009;6:221–9. 
https://doi.org/10.1123/jpah.6.1.221.

85. Bonnefoy M, Kostka T, Berthouze SE, Lacour JR. Validation of a computerized questionnaire for elderly Japanese: the Nakanojo study. J Aging Health. 2005;17:263–82. 
https://doi.org/10.1177/0898264304274121.

86. Pols MA, Peeters PH, Kemper HC, Collette HJ. Repeatability and accuracy of the Physical Activity Scale for the Elderly (PASE): development and evaluation of the physical activity questionnaire for elderly Japanese: the Nakanojo study. J Aging Phys Act. 2007;15:398–411.

87. Matton L, Wijndaele K, Duvigneaud N, Duquet W, Philippaerts RM. C-Sattler et al. The prevalence of selected physical activities and their relation with coronary heart disease risk factors in elderly men: the Zutphen Study, 1985. Am J Epidemiol. 1991;133:1078–92. 
https://doi.org/10.1093/aje/kwj248.

88. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19:231–40. 
https://doi.org/10.1519/00004009-00004009-00000001.

89. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: a practical guide. Cambridge: Cambridge University Press; 2011.

90. Meyer A-M, Evenson KR, Moriimoto L, Siscovich D, White E. Test-retest reliability of the Women's Health Initiative physical activity questionnaire. Med Sci Sports Exerc. 2009;41:530–8. 
https://doi.org/10.1249/MSS.0b013e31818ace55.

91. Harada ND, Chiu V, King AC, Stewart AL. An evaluation of three self-report physical activity instruments for older adults. Med Sci Sports Exerc. 2001;33:962–70. 
https://doi.org/10.1093/medsci/33.5.962.

92. Washburn RA, Smith KW, Jette AM, Janney CA. The Physical Activity Scale for the Elderly (PASE): development and evaluation. J Clin Epidemiol. 1993;46:153–62.

93. Cusick AE, Smith BJ, Chau J, van der Ploeg HP, Friedenreich CM, Armstrong BK, Bauman A. Validity and repeatability of the EPIC physical activity questionnaire: a validation study using accelerometers as an objective measure. Int J Behav Nutr Phys Act. 2008;5:33. 
https://doi.org/10.1186/1479-5868-5-33.

94. Cyarto EV, Marshall AL, Dickinson RK, Brown WJ. Measurement properties of the CHAMPS physical activity questionnaire in a sample of older Australians. J Sci Med Sport. 2006;9:319–26. 
https://doi.org/10.1016/j.jsams.2006.03.001.

95. O'Brien-Cousins S. An older adult exercise status inventory: reliability and validity. J Sport Behav. 1996;19:288–306.

96. Yasunaga A, Park H, Watanabe E, Togo F, Park S, Shephard R, Aoyagi Y. Development and evaluation of the physical activity questionnaire for elderly Japanese: the Sapporo study. J Aging Phys Act. 2015;107:48–59.
109. Barnett A, van den Hoek D, Barnett D, Cerin E. Measuring moderate-intensity walking in older adults using the ActiGraph accelerometer. BMC Geriatr. 2016;16:211. https://doi.org/10.1186/s12877-016-0380-5.

110. Migueles JH, Cadenas-Sanchez C, Ekelund U, Delisle Nystroem C, Mora-Gonzalez J, Löf M, et al. Accelerometer Data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. Sports Med. 2017;47:1821–45. https://doi.org/10.1007/s40279-017-0716-0.

111. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, Terwee CB. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018;27:1171–9. https://doi.org/10.1007/s11136-017-1765-4.

112. Mokkink LB, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, Vet HCW de, Terwee CB. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. 2019. https://www.cosmin.nl. Accessed 8 Sep 2019.

113. Wijndaele K, Westgate K, Stephens SK, Blair SN, Bull FC, Chas-...
140. Matthew CE. Calibration of accelerometer output for adults. Med Sci Sports Exerc. 2005;37:SS12–22. https://doi.org/10.1249/01.mss.0000185659.11982.3d.

141. Swartz AM, Strath SJ, Bassett DR, O’Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. Med Sci Sports Exerc. 2000;32:S450–6.

142. 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;40:181–8. https://doi.org/10.1249/MSS.0b013e31815a51b3.

143. White T, Westgate K, Wareham NJ, Brage S. Estimation of physical activity energy expenditure during free-living from wrist accelerometer in UK adults. PLoS One. 2016;11:e0167472. https://doi.org/10.1371/journal.pone.0167472.

144. Hopkins WG. Measures of reliability in sports medicine and science. Sports Med. 2000;30:1–15. https://doi.org/10.2165/00001856-200030010-00001.

145. Esliger DW, Rowlands AV, Hurst TL, Catt M, Murray P, Eston RG. Validation of the GENEA accelerometer. Med Sci Sports Exerc. 2011;43:1085–93. https://doi.org/10.1249/MSS.0b013e31820513be.

146. Hopkins WG. Measures of reliability in sports medicine and science. Med Sci Sports Exerc. 1991;23:974–9.

147. Sluiter JW, van der Beek AJ, Frings-Dresen MH. Validation of a new questionnaire for the assessment of physical activity at work. Occup Environ Med. 1996;53:377–81. https://doi.org/10.1136/oem.53.4.377.

148. Staten LK, Taren DL, Howell WH, Tobar M, Poehlman ET, Hill MJ. Validation of the Arizona Activity Frequency Questionnaire for Activity Questionnaire. Int J Epidemiol. 2002;31:168–74. https://doi.org/10.1093/ije/31.1.168.

149. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc. 2003;35:1381–95. https://doi.org/10.1249/01.MSS.0000078924.61453.FB.

150. 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;40:181–8. https://doi.org/10.1249/MSS.0b013e31815a51b3.

151. Shaper AG, Wannamethee G, Weatherall R. Physical activity and ischaemic heart disease in middle-aged British men. Br Heart J. 1991;66:384–94. https://doi.org/10.1136/hrt.66.5.384.

152. Souto Barreto P de, Ferrandez A-M, Saliba-Serre B. Questionnaire d’activité physique pour les personnes âgées (QAPPA): validation d’un nouvel instrument de mesure en langue française. Sci Sports. 2011;26:11–8. https://doi.org/10.1016/j.scisp.2010.09.006.

153. Wareham NJ, Nunn RM, Gao Y, Bingham SA, Cummings RJ, Khaw KT. Validation of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. Public Health Nutr. 2003;6:407–13. https://doi.org/10.1079/PHN2002439.

154. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc. 2003;35:1381–95. https://doi.org/10.1249/01.MSS.0000078924.61453.FB.

155. The IPAQ group. International Physical Activity Questionnaires (IPAQ); 2002. https://sites.google.com/site/theipaq/. Accessed 10 Aug 2017.

156. Stel VS, Smit JH, Pluijm SMF, Visser M, Deeg DJH, Lips P. Comparison of the LASSA Physical Activity Questionnaire with a 7-day diary and pedometer. J Clin Epidemiol. 2004;57:252–8. https://doi.org/10.1016/j.jclinepi.2003.07.008.

157. Godin G, Shephard RJ. A simple method to assess exercise behavior in the community. Can J Appl Sport Sci. 1985;10:141–6.

158. Taylor HL, Jacobs DR, Schucker B, Knudsen J, Leon AS, Debacker G. A questionnaire for the assessment of leisure time physical activities. J Chronic Dis. 1978;31:741–55.

159. Giles-Corti B, Timperio A, Cutf H, Pikora TJ, Bull CFL, Knuiman M, et al. Development of a reliable measure of walking within and outside the local neighborhood: RESIDE’s Neighborhood Physical Activity Questionnaire. Prev Med. 2006;42:455–9. https://doi.org/10.1016/j.ypmed.2006.01.019.

160. Voorrips LE, Ravelli AC, Dongelmans PC, Deurenberg P, van Staveren WA. A physical activity questionnaire for the elderly. Med Sci Sports Exerc. 1991;23:974–9.

161. CSEP. Canadian Society for Exercise Physiology-Physical Activity Training for Health (CSEP-PATH), Ottawa: Canadian Society for Exercise Physiology; 2013.

162. Sweden PHao. The Swedish national public health survey; 2012. https://www.folkhalsomyndigheten.se/contentassets/840c39c076eb48bc8a1cbfdff0da22/formular-nationella-folkhaltsoenkate-n-2012.pdf.

163. Souto Barreto P de, Ferrandez A-M, Saliba-Serre B. Questionnaire d’activité physique pour les personnes âgées (QAPPA): validation d’un nouvel instrument de mesure en langue française. Sci Sports. 2011;26:11–8. https://doi.org/10.1016/j.scisp.2010.09.006.

164. Saltin B, Grimby G. Physiological analysis of middle-aged and old former athletes. Comparison with still active athletes of the same ages. Circulation. 1968;38:1104–15.

165. McTiernan A, Kooperberg C, White E, Wilcox S, Coates R, Adams-Campbell LL, et al. Recreational physical activity and the risk of breast cancer in postmenopausal women: the Women’s Health Initiative Cohort Study. JAMA. 2003;290:1331–6. https://doi.org/10.1001/jama.290.10.1331.

166. Wolf AM, Hunter DJ, Colditz GA, Manson JE, Stampfer MJ, Corsano KA, et al. Reproducibility and validity of a self-administered physical activity questionnaire. Int J Epidemiol. 1994;23:991–9. https://doi.org/10.1093/ije/23.5.991.
Affiliations

Matteo C. Sattler\(^1\) · Johannes Jaunig\(^1\) · Christoph Tösch\(^1\) · Estelle D. Watson\(^2\) · Lidwine B. Mokkink\(^3\) · Pavel Dietz\(^4\) · Mireille N. M. van Poppel\(^{1,5}\)

\(^1\) Institute of Sport Science, University of Graz, Graz, Austria
\(^2\) School of Therapeutic Sciences, Faculty of Health Sciences, University of Witwatersrand, Johannesburg, South Africa
\(^3\) Department of Epidemiology and Biostatistics, Amsterdam University Medical Centers, Vrije Universiteit Amsterdam, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands
\(^4\) Institute of Occupational, Social and Environmental Medicine, University Medical Centre, University of Mainz, Mainz, Germany
\(^5\) Department of Public and Occupational Health, Amsterdam University Medical Centers, Vrije Universiteit Amsterdam, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands