Brazilian Journal of Physical Therapy 2020;24(3):187–218

abrapg·ft
Associação Brasileira de Pesquisa e Pós-Graduação em Fisioterapia

Brazilian Journal of Physical Therapy

https://www.journals.elsevier.com/brazilian-journal-of-physical-therapy

SYSTEMATIC REVIEW

Critical evaluation of physical activity questionnaires translated to Brazilian-Portuguese: a systematic review on cross-cultural adaptation and measurements properties

Fernanda Gonçalves Silva[a], Crystian Bitencourt Oliveira[a], Thalysi Mayumi Hisamatsu[a], Ruben Faria Negrão Filho[a], Caio Russo Dutra Rodrigues[a], Marcia Rodrigues Franco[b,c], Rafael Zambelli Pinto[d,*]

[a] *Department of Physical Therapy, Faculdade de Ciências e Tecnologia, Universidade Estadual Paulista (UNESP), Presidente Prudente, SP, Brazil*
[b] *Centro Universitário UNA, Contagem, MG, Brazil*
[c] *Regional Public Hospital of Betim, Betim, MG, Brazil*
[d] *Department of Physical Therapy, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brazil*

Received 3 August 2018; received in revised form 13 April 2019; accepted 18 April 2019
Available online 5 May 2019

KEYWORDS
Psychometric properties;
Validation;
Reliability;
Questionnaire;
Physical activity;
Physical therapy

Abstract
Background: The number of questionnaires that measures physical activity levels has increased considerably. For Brazilian population it becomes a challenge, due to the need of a rigorous translation, adaptation and testing of measurement properties.
Objective: Evaluate the methodological quality and criteria of physical activity questionnaires translated to Brazilian-Portuguese.
Methods: Methodological quality and quality criteria was assessed using the COnsensus-based Standards for the selection of health Measurement INstruments checklist.
Results: Sixty-nine studies were included, the most frequent questionnaires investigated were the International Physical Activity Questionnaire ($n = 16$) and the Baecke Physical Activity Questionnaire ($n = 12$). Translation ($n = 13$), reliability ($n = 37$) and construct validity ($n = 44$) were the measurement properties commonly investigated. For reliability, most studies were rated as 'adequate' for methodological quality. The Intraclass Correlation Coefficient of the questionnaires ranged from 0.20 to 1.0. For construct validity, 31 analyses showed 'inadequate' methodological quality, due to poor description of the comparator instrument. High level

* Corresponding author at: Avenida Pres. Antônio Carlos, 6627, Universidade Federal de Minas Gerais (UFMG), CEP 31270-901 Belo Horizonte, MG, Brazil.
E-mail: rafaelzambelli@gmail.com (R.Z. Pinto).
URLs: https://www.ufmg.br (R.Z. Pinto), https://www.eeffto.ufmg.br (R.Z. Pinto).

https://doi.org/10.1016/j.bjpt.2019.04.002

of evidence on reliability were found for Baecke Physical Activity Questionnaire, Self-administered Physical Activity Checklist and Physical Activity Questionnaire of the Surveillance System of Risk Factors and Protection for Chronic Diseases; on construct validity for Self-administered Physical Activity Checklist, Physical activity Questionnaire for Adolescents, Physical activity Questionnaire for Older Children and *Saúde naBoa* Questionnaire.
Conclusion: Most questionnaires showed poor methodological quality and measurement properties. The Baecke Physical Activity Questionnaire and Self-administred Physical Activity Checklist showed better scorings for methodological quality and quality criteria. Further high methodological quality studies are still warranted.

Introduction

Evidence shows that regular physical activity is associated with low level of mortality in adults and elderly.[1] Sedentary lifestyle and physical inactivity are estimated to be responsible for between 6% and 10% of the major non-communicable diseases.[2] Taken together, the available evidence suggest that physical inactivity is the biggest public health problem of the 21st century worldwide.[3]

Physical activity is defined as any activity involving bodily movement that produces energy expenditure greater than at rest.[4] The term can be interpreted to include activities ranging from structured exercise programs to incidental daily activities.[5] Currently, there are several methods described in the literature for measuring physical activity levels.[6] Choosing the ideal method may depend on several factors, such as the physical activity domains of interest, number of individuals to be analyzed, population of interest and feasibility of the instrument.[7] Physical activity levels can be measured by self-reported and objective assessment methods. The difference is that the self-reported methods rely on information provided by individuals, whereas the objective methods utilize technology to measure and record in real time the biomechanical and/or physiologic consequences of performing physical activity.[8] The self-reported assessment methods have the advantages of being, quick, cheap and easy to administer in comparison to the objective methods.[9]

While the development of one instrument generates various costs, a commonly used and highly effective method is the translation and cultural adaptation of valid questionnaires.[10] Another important step is to assess the measurement properties of the questionnaires to check whether the translated questionnaire behaviors the same way as the original one.[11] More recently, the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) was proposed to evaluate the methodological quality of studies on the measurement properties of health instruments.[12–14] A methodological quality criteria are helpful to legitimize what is the best instrument and whether a measuring instrument has adequate measurement properties.[15,16]

Given that the number of self-reported physical activity questionnaires available has increased considerably over the past decades, the choice of which questionnaire to use has become a challenge for clinicians and research. In Brazil, this challenge is even greater due to the need of a rigorous translation and cultural adaptation process. Therefore, the purpose of this systematic review was to evaluate the procedures of translation and cultural adaptation as well as the measurement properties of physical activity questionnaires translated and adapted into Brazilian-Portuguese.

Methods

Study selection

We included studies that: (1) presented a self-reported questionnaire; (2) included a questionnaire measuring aspects related to physical activity; (3) tested in the Brazilian population; (5) were published as full-text in peer-reviewed journals; and (6) tested its measurement properties (i.e.: assessing reliability, construct validity, responsiveness, content validity, measurement error or internal consistency). In addition, in person or online self-administered questionnaires and questionnaires administered by trained assessors were considered eligible. Questionnaires fully developed and tested in the Brazilian population were also considered eligible for this review. Studies conducted with healthy individuals of different age as well as populations with specific clinical diagnosis (i.e.: cancer, pregnancy, chronic low back pain, and cardiovascular disease) were also included in this review. We excluded studies that: (1) presented an instrument translated and/or adapted in another language.

Search strategy

The literature search was conducted in five electronic databases (MEDLINE, EMBASE, CINAHL, SCIELO and LILACS) from their inception until September 2018. Three groups of search terms were used: Terms of physical activity: exercise, physical inactivity, motor activity, physical fitness, sedentary, life style, leisure activities, walking, sports, aerobic and cycle; questionnaire terms: Questionnaire, index, scale, score, outcome assessment, self-assessment, self-report and inventory; and terms related to language: Portuguese, Brazil, Brasil and Brazilian. There were no restrictions to any specific language and date of publication, but only full texts publications in scientific journals were considered

eligible. Appendix 1 shows the search strategy performed in MEDLINE.

Two independent reviewers (F.G.S. and C.B.O.) screened title and abstracts. Then, full-texts of the potentially eligible papers were evaluated according to the inclusion criteria. If there were any disagreement between the two reviewers a third reviewer (R.Z.P) was consulted to arbitrate. All reviewers are physical therapists with expertise in conducting systematic reviews and studies assessing measurement properties of health instruments.

Data extraction

Two independent reviewers (F.G.S. and C.B.O.) using a standardized form performed data extraction. The following information from self-reported questionnaires were extracted for each included study: (i) domains of physical activity (e.g. leisure time, household, transportation and occupational activity); (ii) recall period (e.g. activities performed in the last day or seven days, last month or last year); (iii) number of items; (iv) unit of measure and (v) type of population. Data regarding the measurement properties were also extracted.

Methodological quality assessment

We assessed the methodological quality of the included studies using the COSMIN Checklist.[12–14] Two reviewers (F.G.S. and C.B.O.) rated independently each study, and, in case of disagreement, a third reviewer (R.Z.P.) was consulted to arbitrate. The checklist is composed by nine measurement properties: cross-cultural validity, measurement error, internal consistency, content validity, structural validity, reliability, construct validity, hypothesis-testing, criterion validity and responsiveness. Definition of each measurement property is provided in Table 1. Each measurement property consists of a number of items evaluated using a 4-point scale (i.e. very good, adequate, doubtful and inadequate). The final methodological quality score for each measurement property was determined considering the worst score among all items. For reliability, the time intervals considered appropriate were: (i) for a recall period of an usual week a time interval between 1 day and 3 months; (ii) for a recall period of the previous week a time interval between 1 day and 2 weeks; and (iii) a recall period of the previous day, a time interval between 1 day and 1 week.[17]

Quality criteria assessment

We also assessed whether the measurement properties reported in the included studies were adequate using the quality criteria proposed in the COSMIN checklist. For each measurement property a criterion was defined for sufficient (+), insufficient (−) or indeterminate (?) rating. Details of the quality criteria are described in Table 1. For reliability assessment, it was considered the 'Total physical activity' score from each questionnaire if available. For studies reporting only separate data by Physical activity (PA) domains, we assessed each domain separately and the final quality criteria was determined considering the consistency for most (>50%) physical activity domains (e.g. a study to be rated as + more than 50% of the domains had to show an Intraclass Correlation Coefficient greater than 0.70). For construct validity, in addition to the quality criteria assessment we also assessed the degree of similarity between the physical activity domain measured with the questionnaire and the comparator instrument using the level of evidence classification for the comparator instrument (i.e. Levels 1, 2 and 3) described elsewhere.[18] In this classification, the level of evidence varies depending on how the physical activity dimension of interest was measured. In brief, comparator instrument close to level 1 indicates that the comparator instrument chosen has the highest degree of similarity to the physical activity domain of the questionnaire.

Assessment of overall level of evidence

Each measurement property analyzed from all questionnaires was assessed according to the overall level of evidence. Data from all studies investigating the same questionnaire were combine and levels of evidence provided for each measurement property. The quality of the evidence refers to the confidence that the summarized result is trustworthy. We assesed the quality of evidence using a modified version of the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach and downgraded the evidence level considering the following four domains: (1) methodological quality (2) inconsistency of results across studies, (3) imprecision (i.e. total sample size of the available studies) and (4) indirectness (i.e. evidence from different populations than the population of interest in the review). The quality of the evidence was classified as graded as high, moderate, low, or very low evidence.

Results

The search strategy identified a total of 11022 records. After title and abstracts screening, 69 records were considered potentially eligible and the full-text retrieved. Details of the selection process are described in Fig. 1.

Physical activity questionnaires

Sixty-nine included studies investigated 30 different self-report physical activity questionnaires. Table 2 describes the included questionnaires in terms of target population, recall period, domain of activity investigated, number of items, and unit of measurement. The target population varied across studies including broad population of adolescents, young adults, adults and older people as well as specific clinical population such as adults with claudication, pregnant women, individuals with heart disease, low back pain and juvenile dermatomyositis population. The recall period varied across questionnaires including past 24 h, last week, last month, last 12 months and present way of life.

Measurement properties

A total of 110 analyses on measurement properties were investigated in 69 included studies. Of these, 44 (40%)

Table 1 Description of measurement properties definition and the criteria adopted for methodological and results assessments.

Measurement property	Definition[12]	Methodological quality assessment[12]	Quality criteria assessment[12]
Content validity	The degree to which the content of an instrument is an adequate reflection of the construct to be measured.	Assessment of general requirements (e.g. relevance of items, comprehensiveness of the instrument and any important flaws in the design or methods of the study)	(+) A clear description is provided of the measurement aim, the target population, the concepts that are being measured, and the item selection AND target population and (investigators OR experts) were involved in item selection; (?) A clear description of above-mentioned aspects is lacking OR only target population involved OR doubtful design or method; (−) No target population involvement[15]
Structural validity	The degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured	Assessment of design requirements and statistical methods (e.g. adequate sample size, information on exploratory factor analysis or IRT tests and any important flaws in the design or methods of the study)	(+) CTT CFA: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.082 IRT/Rasch No violation of unidimensionality: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.08 AND no violation of local independence: residual correlations among the items after controlling for the dominant factor <0.20 OR Q3's <0.37 AND no violation of monotonicity: adequate looking graphs OR item scalability >0.30 AND adequate model fit: IRT: $\chi^2 > 0.01$ Rasch: infit and outfit mean squares $\geq$0.5 and $\leq$1.5 OR *Z*-standardized values >−2 and <2 (?) Not all information for '+' reported IRT/Rasch: Model fit not reported (−) Criteria for '+' not met
Internal consistency	The degree of the interrelatedness among the items.	Assessment of design requirements and statistical methods (e.g. information on Cronbach's alpha analysis and any important flaws in the design or methods of the study)	(+) At least low evidence for sufficient structural validity AND Cronbach's alpha(s) $\geq$0.70 for each unidimensional scale or subscale; (?) Criteria for ''At least low evidence for sufficient structural validity'' not met; (−) At least low evidence for sufficient structural validity AND Cronbach's alpha(s) <0.70 for each unidimensional scale or subscale
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.	Assessment of design requirements and statistical methods (e.g. adequate sample size, characteristics similarity on sample and if the regression analysis or IRT was assessed)	(+) No important differences found between group factors (such as age, gender, language) in multiple group factor analysis OR no important DIF for group factors (McFadden's $R^2 < 0.02$) (?) No multiple group factor analysis OR DIF analysis performed (−) Important differences between group factors OR DIF was found

Table 1 (*Continued*)

Measurement property	Definition[12]	Methodological quality assessment[12]	Quality criteria assessment[12]
Reliability	The proportion of the total variance in the measurements which is due to true differences between Individuals. The extent to which scores for individuals who have not changed are the same for repeated measurement under several conditions.	Assessment of design requirements and statistical methods (e.g. test conditions, information on time interval, ICC or Kappa analysis assessment)	(+) ICC or weighted Kappa >0.70; (?) ICC or weighted Kappa not reported; (−) ICC or weighted Kappa <0.70
Measurement error	The systematic and random error of an individual's score that is not attributed to true changes in the construct to be measured.	Assessment of design requirements (e.g. information on time interval, test conditions, SEM, SDC or LoA analysis assessment and any important flaws in the design or methods of the study)	(+) SDC or LoA < MIC; (?) MIC not defined; (−) SDC or LoA > MIC5
Criterion validity	The degree to which the scores of an instrument are an adequate reflection of a 'gold standard'.	Assessment of design requirements and statistical methods (e.g. AUC analysis, sensitivity and specificity determined and any important flaws in the design or methods of the study)	(+) Correlation with gold standard ≥0.70 OR AUC ≥0.70; (?) Not all information for '+' reported; (−) Correlation with gold standard <0.70 OR AUC <0.70
Construct validity	The degree to which the scores of an instrument are consistent with hypotheses (*for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups*) based on the assumption that the instrument validly measures the construct to be measured.	Assessment of design requirements and statistical methods (e.g. measurement properties of comparator instrument, comparison between subgroups and any important flaws in the design or methods of the study)	(+) The result is in accordance with the hypothesis; (?) No hypothesis defined (by the review team); (−) The result is not in accordance with the hypothesis
Responsiveness	The ability of an instrument to detect change over time in the construct to be measured.	Assessment of design requirements and statistical methods (e.g. gold standard use, ROC curve calculated, sensitivity and specificity determined, measurement properties of comparator instrument and any important flaws in the design or methods of the study)	(+) The result is in accordance with the hypothesis 7 OR AUC ≥ 0.70; (?) No hypothesis defined (by the review team); (−) The result is not in accordance with the hypothesis 7 OR AUC < 0.70

AUC, area under the curve; CFA, confirmatory factor analysis; CFI, comparative fit index; CTT, classical test theory; CTV, content validity; DIF, differential item functioning; ICC, intraclass correlation coefficient; IRT, item response theory; LoA, limits of agreement; MIC, minimal important change; RMSEA, root mean square error of approximation; ROC, receiver operator curve; SDC, smallest detectable change; SEM, standard error of measurement; SRMR, standardized root mean residuals; TLI, Tucker–Lewis index.
(+) = sufficient rating, (?) = indeterminate rating, (−) = insufficient rating.

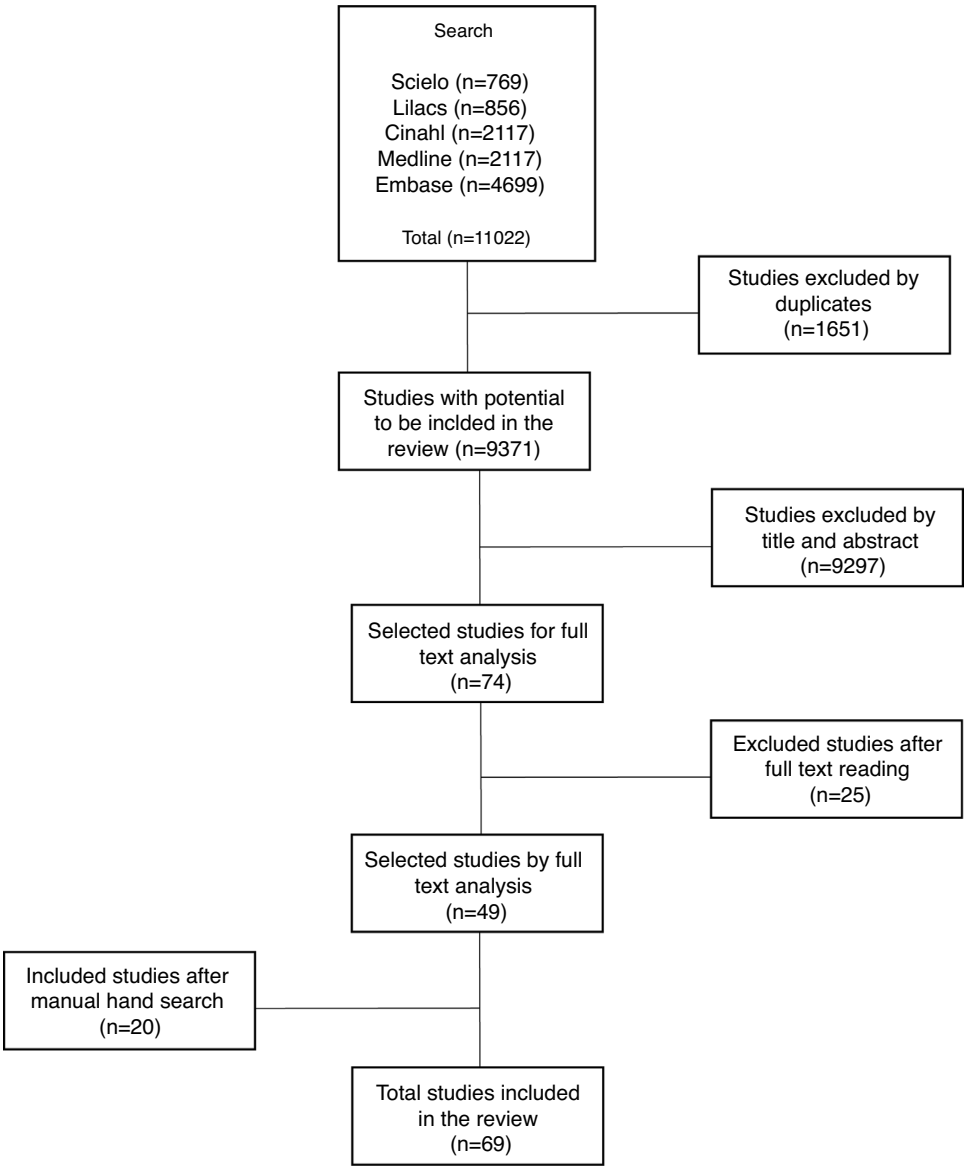

Figure 1 Flowchart for the included studies after systematic review.

analyses were on construct validity, 37 (33.6%) on reliability, 13 (11.8%) on translation and cross-cultural adaptation, nine (8%) on measurement error, four (3.6%) on internal consistency, two (2%) on content validity and one (1%) on responsiveness. According to COSMIN checklist, 18 (16.4%) measurement properties were rated as ''very good'', 28 (25.5%) as ''adequate'', 25 (22.7%) as ''doubtful'' and 39 (35.4%) as ''inadequate''. Table 3 shows the methodological quality and quality criteria assessment for all measurement analyses investigated in each study.

Translation and cross-cultural adaptation

Thirteen[19–30] translation and cross-cultural adaptation analyses of physical activity questionnaires were reported. The methodological quality assessment was rated as 'Doubtful', due to unclear information regarding whether the study samples were similar for relevant characteristics. Additionally, all studies were rated as 'Inadequate' because the sample size and the statistical methods used to analyze the data were inappropriate.

Measurement error

Nine[31–39] studies performed the analysis of measurement error. Of these, eight[31,32,34–39] studies were rated as 'Very Good' and one as 'Inadequate' in the methodological quality assessment. The study rated as 'Inadequate' used an inappropriate time interval between assessments. For the quality criteria assessment, all studies received an indeterminate (?) rating because minimal important change was not calculated. The Baecke Physical Activity Questionnaire for adolescents showed the highest percent of agreement (95.5%). The International Physical Activity Questionnaire – long version and the Baltimore Activity Scale for Intermittent Claudication showed the highest limits of

Table 2 Characteristics of physical activity questionnaires.

Physical activity questionnaire (abbreviation)	Domains:	Recall period:	No. of items	Unit of measure:	Target population:
Active Australia Questionnaire (AAQ)	Walk, Yard-work PA, Sports, Household PA, Leisure-time PA	Last week	8	Min/week	Elderly
Baltimore Activity Scale for Intermittent Claudication (BASIC)	Leisure-time PA and Transportation PA	Lately, weekly	7	Dimensionless score	Individuals with intermittent claudication
Baecke Physical Activity Questionnaire (BPAQ)	Occupational PA; Sports; Leisure-time PA	Last 12 months	16	Dimensionless score	Adults, elderly, people living with HIV, youth
Questionnaire of a typical physical activity and food intake day (DAFA)	Transportation PA; Sports; Leisure-time PA	Typical day	11	Dimensionless score	Youth
Internet Version of the Questionnaire of a typical physical activity and food intake day (DAFA)	Transportation PA; Household PA, Leisure-time PA	Typical day	11	Dimensionless score	Youth
Questionnaire of previous physical activity and food intake day (DAFA)	Transportation PA; Sports, Leisure-time PA	Previous day	11	Dimensionless score	Youth
Internet Version of the Questionnaire of previous physical activity and food intake day (DAFA)	Transportation PA; Sports, Leisure-time PA	Previous day	11	Dimensionless score	Youth
Godin Shephard Leisure-Time Physical Activity Questionnaire (GSLTPAQ)	Leisure-time PA	Typical week	11	METs and dimensionless score	Adults, people with heart disease
Human Activity Profile Questionnaire (HAP)	Transportation PA; Sports; Leisure-time PA; Sedentary activities	In the present moment, lately.	94	Sedentary activity/day Moderate activity/day – min/day	Elderly
Health-Promoting Lifestyle Profile-II (HPLP-II)	Leisure-time PA; Sports; Transportation PA	Present way of life or personal habits	52	Dimensionless score	Adults
International Physical Activity Questionnaire (IPAQ) Long Version	Occupational PA; Transportation PA; Household PA; Leisure-time PA; Sedentary activities	Last week	27	METs.min/week	Adults, youth, adult with high blood pressure, elderly with Alzheimer's disease
International Physical Activity Questionnaire (IPAQ) Short Version	Occupational PA; Transportation PA; Household PA; Leisure-time PA; Sedentary activities	Last week	8	METs.min/week	Adults, climateric women
Minnesota Leisure Time Activities Questionnaire in elderly (MLTAQ)	Household PA, Sports; Leisure-time PA	Last year	63	METs/min/week/ month/year	Elderly
Netherlands Physical Activity Questionnaire (NPAQ)	Leisure-time PA and Sports	Usual preferences	7	Dimensionless score	Youth
Physical Activity Checklist Interview (PACI)	Regular PA; Leisure PA; Transportation PA; Sedentary activities	Past 24 h	21	min/min × MET/ min × MET × intensity rate	Youth

Table 2 (*Continued*)

Physical activity questionnaire (abbreviation)	Domains:	Recall period:	No. of items	Unit of measure:	Target population:
Physical Activity Questionnaires for Adolescents (PAQ-A)	Sports and Leisure-time PA	Last week	8	Dimensionless score	Adolescents from 14 to 18 years old
Physical Activity Questionnaires for Older Children (PAQ-C)	Sports and Leisure-time PA	Last week	9	Dimensionless score	Childrens from 8 to 13 years old
Physical Activity Questionnaire for Pregnant Women (PAQPW)	Leisure-time PA; Sports; Transportation PA; Sedentary activities; Household PA	Present way of life	Not provided	Dimensionless score	Pregnant women
Physical Activity Rating (PAR)	Overall level of PA	Last month	0–7 (Scale)	Dimensionless score	Elderly
Three day physical activity questionnaire (3DPAR)	Transportation PA; Sports; Leisure-time PA	Habitual PA	Recordatory	Min/day – Hour/day MET	Adolescents
24 h physical activity recordatory (24PAR)	Transportation PA; Sports; Leisure-time PA	24 h	Recordatory	Min/day – Hour/day MET	Adults
Physical activity level and sedentary behavior evaluation questionnaire for school students (PASBEQ)	Sports, Leisure-time PA, Transportation PA, School-time PA, Sedentary activities	Typical week		Hour/week and METs/week	Adolescents from 10 to 13 years old
Questionnaire to measure physical activity and sedentary behavior (PASBQ)	Leisure-time PA and Sedentary activity	Typical weekday, weekend	12	Min/day score (0–24)	Youth
Brazilian National School-Based Health Survey (PeNSE)	Sports; Leisure-time PA; Sedentary activities	Last week	11	Minutes/week.	Adolescents
Pregnancy Physical Activity Questionnaire (PPAQ)	Household/caregiving PA; Occupational PA; Sports; Transportation PA and Sedentary activities	Daily routine activity	33	Minutes or hours per day – MET – MET-hour/week.	Pregnant women
Self-Administered Physical Activity Checklist (SAPAC)	Regular PA; Leisure PA; Transportation PA; Sedentary activities	Last week	24	min/min × MET/ min × MET × intensity rate	Youth
Saúdes Vitória Study's physical activity assessment questionnaire for children (Saúdes)	Sedentary activities; Transportation PA; Sports; Leisure-time PA	Typical day	13	Hours and minutes	Youth
Saúde na Boa Questionnaire (SBQ)	Not provided	Typical week and last seven days	Not provided	Not provided	Adolescents
Short version Physical Activity Questionnaire (SVPAQ)	Transportation PA, Sports; Leisure-time PA	Last week	8	Min/week	Adolescents
Questionário de atividade física do sistema de vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico (VIGITEL)	Transportation PA; Occupational PA; Leisure-time PA; Household PA	Last three months, last week, lately	20	Min/day – Min/week	Adults

Description of the characteristics of each questionnaire, such as the domains evaluated through the items, number of questions (items), the period considered when answering the questions, the unit of measure generated by the questionnaire and the population to which the instrument was submitted.
MET, metabolic equivalent of task; PA, physical activity.

Table 3 Characteristics of included studies and detailed information on measurement properties investigated.

Reference	Physical activity questionnaire[a]	Analysis performed	Study characteristics	Sample size (gender); Mean age (SD); Target population	Results	COSMIN rank	Quality criteria assessment
Rocha et al., 2017[30]	AAQ	Translation and cross-cultural adaptation Reliability	Translation process Test re-test: 4 h	22 (F: 22); 72.5 (5.3); Elderly women	**Reliability:** ICC: 0.97	Cross-cultural validity: Doubtful Reliability: Inadequate	Reliability: (+)
Souza Barbosa et al., 2012[38]	BASIC	Reliability; Measurement error	Test re-test: 7 days apart	38 (F: 20/M: 18); 64 (11.4); Individuals with intermittent claudication	**Reliability:** ICC: 0.87 (95% CI: 0.74,0.93) **Measurement error**: LoA: −117 to 250 kcal	Reliability: Adequate Measurement error: Very good	Reliability: (+) Measurement error: (?)
Lopes et al., 2013[76]	BASIC	Construct validity	Comparator: Pedometer	150 (F: 56/M: 94); 64 (9); Individuals with intermittent claudication	**Construct validity**: $r = 0.34$	Construct validity: Doubtful	Construct validity: (?), level of evidence = 3−
Florindo et al., 2004[40]	BPAQ	Translation and cross-cultural adaptation; Internal consistency	Translation process	326 (M: 326); 62.5 (7.9); Men aged 50 or more	**Internal consistency**: Cronbach alpha OPA − 0.52 SPA − 0.52 LPA + TPA − 0.62	Cross-cultural validity: Doubtful Internal consistency: Very good	Internal consistency: (?)
Sardinha et al., 2010[21]	BPAQ	Translation and cross-cultural adaptation	Translation process	30 (M: 11/F: 19); 48.13 (15.99); Adults	n/a	Cross-cultural validity: Doubtful	n/a
Garcia et al., 2013[66]	BPAQ	Construct validity	Comparator: accelerometer	58 (F: 40/M: 18); 39.9 (11.5); Adults	**Construct validity**: SPA + LPA + TPA: $r = 0.36$ Total score: $r = 0.54$	Construct validity: Inadequate	Construct validity: (?), level of evidence = 2−/2 +
Florindo et al., 2003[48]	BPAQ	Reliability; Construct validity	Test re-test: 45 days Comparator: VO_2max	27 (M: 21); 32.6 (3.1); Adult men	**Reliability**: Total score − ICC: 0.77 LPA + TPA score − ICC: 0.80 SPA − ICC: 0.69 **Construct validity**: Total score: $r = 0.17$ LPA + TPA score: $r = 0.24$ SPA: $r = 0.04$	Reliability: Adequate Construct validity: Inadequate	Reliability: (+) Construct validity: (?), level of evidence = 3?/3?/3?

Table 3 (*Continued*)

Reference	Physical activity questionnaire[a]	Analysis performed	Study characteristics	Sample size (gender); Mean age (SD); Target population	Results	COSMIN rank	Quality criteria assessment
Florindo et al., 2006[49]	BPAQ	Reliability; Construct validity	Test re-test: 15–30 days Comparator: VO_2max	29; 37.2 (range: 26.0–49.5); HIV population	**Reliability**: Total score – ICC: 0.72 LPA + TPA score – ICC: 0.44 SPA – ICC: 0.70 OPA – ICC: 0.85 **Construct validity**: Total score: $r = 0.27$ LPA + TPA score: $r = 0.19$ SPA: $r = 0.41$ OPA: $r = -0.14$	Reliability: Adequate Construct validity: Inadequate	Reliability: (+) Construct validity: (?), level of evidence= 3?/3?/3?/3?
Glaner et al., 2007[70]	BPAQ	Construct validity	Comparator: VO_2max	105 (F: 28/M: 77); 24.8 (5.3); Adults	**Construct validity**: % concordance = 64.1%	Construct validity: Inadequate	Construct validity: (?), level of evidence = 3+
Mazo et al., 2001[50] (Modified for elderly women)	BPAQ	Reliability; Construct validity	Test re-test: 15 days apart Comparator: Pedometer	30 (F: 30); 71.2 (4.6); Elderly women	**Reliability**: SPA: ICC = 0.84 LPA: ICC = 0.85 Home activities: ICC = 0.82 Total PA: ICC = 0.83 **Construct validity**: Total score: 0.27	Reliability: Adequate Construct validity: Inadequate	Reliability: (+) Construct validity: (?), level of evidence = 3?
Carvalho et al., 2017[62]	BPAQ	Reliability; Construct validity	Test re-test: 7 days Comparator: Accelerometer	73 (F: 23/M: 50); 37.2 (12.2); Adults with chronic low back pain	**Reliability**: Total PA $ICC_{2,1}$: 0.77 OPA $ICC_{2,1}$: 0.84 SPA $ICC_{2,1}$: 0.83 LPA $ICC_{2,1}$: 0.61 **Construct validity**: Total PA (counts/min) $r = 0.18$ Total PA (VM counts/min) $r = 0.26$ Total PA (MVPAmin/day) $r = 0.17$ Total PA (Steps/day) $r = 0.27$	Reliability: Adequate Construct validity: Inadequate	Reliability: (+) Construct validity: (?), level of evidence: 2−/3−/3−/3−

Table 3 (*Continued*)

Reference	Physical activity questionnaire[a]	Analysis performed	Study characteristics	Sample size (gender); Mean age (SD); Target population	Results	COSMIN rank	Quality criteria assessment
Guedes et al., 2005B[63]	BPAQ	Reliability; Construct validity	Test re-test: 2 weeks apart Comparator: PA recordatory	≤14 years girls 59; 12.92 (0.86); >14 years girls 33; 15.8 (0.93); ≤14 years boys 38; 13 (0.81); >14 years boys 31; 15.81 (0.70); Adolescents	**Reliability**: ≤14 years/≥ 14 years ♀ OPA ICC: 0.55/ICC: 0.61 SPA ICC: 0.79/ICC: 0.85 TPA ICC: 0.61/ICC: 0.70 Total PA ICC: 0.66/ICC: 0.76 ≤14 years/≥14 years ♂ OPA ICC: 0.68/ICC: 0.69 SPA ICC: 0.73/ICC: 0.82 TPA ICC: 0.71/ICC: 0.76 Total PA ICC: 0.75/ICC: 0.80 **Construct validity**: ≤14 years/≥14 years ♀ Total PA *r*: 0.36/*r*: 0.46 ≤14 years/≥14 years ♂ Total PA *r*: 0.41/*r*: 0.59	Reliability: Adequate Construct validity: Doubtful	Reliability: (+) for ≥14 years old girls and boys and ≤14 years old boys. (−) for ≤14 years old girls. Construct validity: (?), level of evidence= 3−/3?/3?/3?
Romero et al., 2011[36] (Internet version)	BPAQ	Reliability; Measurement error	Test re-test: 14 days apart	135 (F: 74/M: 61); Youth	**Reliability**: $k = 0.47$ **Measurement error**: % of agreement: 95.5%	Reliability: Adequate Measurement error: Very good	Reliability: (−) Measurement error: (?)
Florindo et al., 2006b[51]	BPAQ	Development reliability; Construct validity	Test re-test: 15 days apart Comparator: 20-m shuttle run test Frequency meter VO_2max Waist circumference	94 (F: 64/M: 30); 13 (1.1); Youth	**Reliability**: ICC > 0.60 **Construct validity**: Yearly/weekly VO_2max: $r = 0.18/r = 0.28$ Total speed: $r = 0.15/r = 0.24$ Total time: $r = 0.19/r = 0.30$ Maximum heart rate: $r = 0.05/r = 0.08$ Waist circumference: $r = -0.12/r = -0.06$	Reliability: Adequate Construct validity: Doubtful	Reliability: (−) Construct validity: (?), level of evidence = 3− (all analysis)

Table 3 (*Continued*)

Reference	Physical activity questionnaire[a]	Analysis performed	Study characteristics	Sample size (gender); Mean age (SD); Target population	Results	COSMIN rank	Quality criteria assessment
Morelhão et al., 2018[82]	BPAQ	Responsiveness	Follow up period: 2 months	106 (F: 56/M: 50); 40 (11.6); Adults with chronic low back pain	**Responsiveness**: Mean difference: 0.18 (2.25) Effect size (84% CI): 0.12 (−0.08 to 0.34)	Responsiveness: Inadequate	Responsiveness: (−)
Costa et al., 2010[53]	DAFA	Reliability	Test re-test: 15 days apart	101 (F: 44/M: 57); 9.4 (1.0); Youth	Dance – ICC: 0.50 Walk/run – ICC: 0.51 Play with the dog – ICC: 0.75 Household – ICC: 0.68 Cycle – ICC: 0.79 Rope jump – ICC: 0.51 Climb stairs – ICC: 0.62 Play soccer – ICC: 0.86 Swim – ICC: 0.79 Skateboard – ICC: 0.83 Gymnastics – ICC: 0.77	Reliability: Adequate	Reliability: (+)
Barros et al., 2007[54]	DAFA	Reliability; Construct validity	Test re-test: days apart Comparator: Questionnaire answered by parents/teachers	69 (F: 35/M: 35); 7–10 years old	**Reliability**: Dance ICC: 0.62 Walk/run ICC: 0.55 Play with the dog ICC: 0.77 Household ICC: 0.75 Cycle ICC: 0.63 Jump rope ICC: 0.65 Climb stairs ICC: 0.75 Play soccer ICC: 0.79 Swim ICC: 0.33 Skate ICC: 0.63 Gymnastics ICC: 0.75 General PA ICC: 0.85 **Construct validity**: $k = 0.28$	Reliability: Adequate Construct validity: Inadequate	Reliability: (+/−) Unclear Construct validity: (?), level of evidence = 3−
Legnani et al., 2013[37]	DAFA (Internet version)	Reliability; Measurement error	Test re-test: a day apart	127 (F: 58/M: 69); 8.4 (1.1); Youth	**Reliability**: General PA ICC: 0.94 **Measurement error**: Mean error: 1.7 (95% CI: −25.6,29.1)	Reliability: Adequate Measurement error: Very good	Reliability: (+) Measurement error: (?)

Table 3 (*Continued*)

Reference	Physical activity questionnaire[a]	Analysis performed	Study characteristics	Sample size (gender); Mean age (SD); Target population	Results	COSMIN rank	Quality criteria assessment
Cabral et al., 2011[79]	DAFA (previous day version)	Construct validity	Comparator: Pedometer	50 (F: 25/M: 25); 10.2 (1.49); Youth	**Construct validity**: $r = 0.45$	Construct validity: Very good	Construct validity: (?), level of evidence = 3−
Jesus et al., 2016[64]	DAFA (Internet version of the previous day)	Reliability; Construct validity	Test re-test: 3 h apart Comparator: Direct observation	**Reliability**: 94 (F: 25/M: 25) **Validity**: 390 (F: 194/M: 196) 9.53 (1.53); Youth	**Reliability**: Incidence ratio ranged from 0.63 to 7.52 **Construct validity**: Childs play incidence ratio ranged from 0.52 to 18.1	Reliability: Inadequate Construct validity: Inadequate	Reliability: (?) Construct validity: (?), level of evidence = 1?
Sao Joao et al., 2013[25]	GSLTPAQ	Translation and cross-cultural adaptation Reliability; Content validity	Translation process Test re-test: 15 days apart	80 (F: 48/M: 32); 53.2 (10.4); Healthy individuals and individuals with cardiovascular disease	**Reliability**: Strenuous PA ICC: 0.79 Moderate PA ICC: 0.80 Mild PA ICC: 0.82 Total PA ICC: 0.84	Cross-cultural validity: Doubtful Reliability: Adequate Content validity: Inadequate	Reliability: (+) Content validity: (−)
São João et al., 2015[80]	GSLTPAQ	Construct validity	Comparator: VO_2peak, VO_2pred, PA Questionnaires	236 (F: 138/M: 98); 52.8 (11.1); Healthy individuals and with cardiovascular disease	VO_2peak: Total PA $r = 0.09$/MVPA $r = 0.03$ VO_2pred: Total PA $r = 0.15$: MVPA $r = 0.19$ PA Questionnaire (VSAQ): Total PA $r = 0.23$/MVPA $r = 0.34$ PA Questionnaire (Baecke): Total PA $r = 0.36$/MVPA $r = 0.25$ PA in leisure: $r = 0.62$	Construct validity: Very good	Construct validity: (+), level of evidence = 3− (all analysis)
Souza et al., 2006[24]	HAP	Translation and cross-cultural adaptation; Internal consistency	Translation process	230 (F: 198/M: 32); 66.32 (8.50); Elderly	**Internal consistency**: Rash analysis: 0.91	Cross-cultural validity: Doubtful Internal consistency: Inadequate	Internal consistency: (?)

Table 3 (*Continued*)

Reference	Physical activity questionnaire[a]	Analysis performed	Study characteristics	Sample size (gender); Mean age (SD); Target population	Results	COSMIN rank	Quality criteria assessment
Bastone et al., 2014[73]	HAP	Construct validity	Comparator: Accelerometer	120 (F: 120); 71.8 (6.6); Elderly women	**Construct validity**: Counts/day: $r = 0.61$ Moderate activity/day: $r = 0.63$ Steps/day: $r = 0.69$ Energy expenditure/day: $r = 0.55$	Construct validity: Inadequate	Construct validity: (?), level of evidence = 2+/1+/1−/2+
Tajik et al., 2010[22]	HPLP-II	Translation and cross-cultural adaptation; Internal consistency	Translation process	30 (F: 18/M: 12); 37.4; Adults	**Internal consistency**: Cronbach alpha Total – 0.93 PA subscale – 0.85	Cross-cultural validity: Doubtful Internal consistency: Very good	Internal consistency: (?)
Barros et al., 2000[42]	IPAQ – long version	Reliability	Test re-test: 7 days apart	42 (F: 20/M: 22); 34.7 (8.8); Adult	OPA ICC: $0.88/k = 0.33$ Household PA ICC: $0.67/k = 0.25$ TPA ICC: $0.68/k = 0.41$ LPA ICC: $0.71/k = 0.32$ Total ICC: $0.86/k = 0.39$	Reliability: Adequate	Reliability: (+)
Benedetti et al., 2004[33]	IPAQ – long version	Reliability; Construct validity	Test re-test: 15 days apart Comparator: Pedometer and PA diary	41 (F: 41); 67 (4.8); Elderly women	**Reliability**: OPA ICC: $0.97/r = 1.00$ Household PA ICC: $0.89/r = 0.77$ TPA ICC: $0.73/r = 0.67$ LPA ICC: $0.86/r = 0.95$ Sitting time ICC: $0.76/r = 0.60$ Total ICC: $0.88/r = 0.77$ **Construct validity**: $r = 0.12$–0.27 Pedometer $r = 0.16$–0.37 PA diary	Reliability: Adequate Construct validity: Inadequate	Reliability: (+) Construct validity: (?), level of evidence = 3?/3?
Benedetti et al., 2007[33]	IPAQ – long version	Reliability; Construct validity; Measurement error	Test re-test: 21 days apart Comparator: Pedometer and PA diary	29 (M: 29); 66.6 (4.3); Elderly men	**Reliability**: $r_s = 0.95$ **Construct validity**: $r = 0.24/k = 0.03$ Pedometer $r = 0.38/k = 0.35$ PA diary **Measurement error**: LoA: 7.29 kcal/min to −14.0 kcal/min	Reliability: Doubtful Construct validity: Inadequate Measurement error: Very good	Reliability: (?) Construct validity: (?), level of evidence = 3?/3? Measurement error: (?)

Table 3 (*Continued*)

Reference	Physical activity questionnaire[a]	Analysis performed	Study characteristics	Sample size (gender); Mean age (SD); Target population	Results	COSMIN rank	Quality criteria assessment
Lima et al., 2010[44]	IPAQ – long version	Reliability; Construct validity	Test re-test: 7 days apart Comparator: Pedometer	26 (F: 22/M: 4); 74.4 (6.5); Elderly with Alzheimer's disease	**Reliability**: ICC: 0.56 **Construct validity**: $r = 0.57$	Reliability: Adequate Construct validity: Inadequate	Reliability: (−) Construct validity: (?), level of evidence = 3?
Guedes et al., 2005A[34]	IPAQ – long version	Reliability; Construct validity; Measurement error	Test re-test: 4 days apart Comparator: 24 h recordatory	≤14 years 97 (F: 59/M: 33); 12.9 (0.84) ≥14 years 64 (F: 33/M: 31); 15.8 (0.84); Youth	**Reliability**: ≤14 years/≥14 years ♀ Walk $r_s = 0.52/r_s = 0.55$ Moderate PA $r_s = 0.49/r_s = 0.63$ Intense PA $r_s = 0.70/r_s = 0.55$ Sitting $r_s = 0.58/r_s = 0.61$ ≤14 years/≥14 years ♂ Walk $r_s = 0.56/r_s = 0.61$ Moderate PA $r_s = 0.59/r_s = 0.66$ Intense PA $r_s = 0.67/r_s = 0.83$ Sitting $r_s = 0.62/r_s = 0.82$ **Construct validity**: ≤14 years/≥14 years ♀ Walk $r_s = 0.17/r_s = 0.11$ Moderate PA $r_s = 0.24/r_s = 0.35$ Intense PA $r_s = 0.26/r_s = 0.43$ Sitting $r_s = 0.16/0.24$ ≤14 years/≥14 years ♂ Walk $r_s = 0.09/r_s = 0.12$ Moderate PA $r_s = 0.29/r_s = 0.34$ Intense PA $r_s = 0.35/r_s = 0.51$ Sitting $r_s = 0.29/r_s = 0.39$ **Measurement error**: LoA: ♂ ≥ 14 years: 16 ± 92 min LoA: ♀ ≤ 14 years: 131 ± 429 min	Reliability: Doubtful Construct validity: Doubtful Measurement error: Very good	Reliability: (?) Construct validity: (?), level of evidence = 3+ (for ≤14 years girls at Int. efforts and ≥14 years boys at Int. efforts and sitting) 3− (for all other analyses) Measurement error: (?)
Garcia et al., 2013[66]	IPAQ – long version	Construct validity	Comparator: Accelerometer	58 (F: 40/M: 18); 39.9 (11.5); Adults	**Construct validity**: Moderate vigorous intensity: $r = 0.34$	Construct validity: Inadequate	Construct validity: (?), level of evidence = 1−

Table 3 (*Continued*)

Reference	Physical activity questionnaire[a]	Analysis performed	Study characteristics	Sample size (gender); Mean age (SD); Target population	Results	COSMIN rank	Quality criteria assessment
Hallal et al., 2010[35]	IPAQ – long version	Reliability; Construct validity; Measurement error	Test re-test: 5 days Comparator: Accelerometer	156 (F: 81/M: 75); 40.3 (15.1); Adults	**Reliability**: TPA $r_s = 0.87$ LPA $r_s = 0.92$ TPA + LPA $r_s = 0.90$ **Construct validity**: Moderate intensity: $r = 0.23$ Vigorous intensity: $r = 0.30$ Total score: $r = 0.22$ **Measurement error**: Mean difference: 3 min % of agreement: 89.8%	Reliability: Doubtful Construct validity: Inadequate Measurement error: Very good	Reliability: (?) Construct validity: (?), level of evidence = 1−/1−/2− Measurement error: (?)
Carvalho et al., 2017[62]	IPAQ –long version	Reliability; Construct validity	Test re-test: 7 days Comparator: Accelerometer	73 (F: 23/M: 50); 37.2 (12.2); Adults with low back pain	**Reliability**: Total PA $ICC_{2,1}$: 0.37 OPA $ICC_{2,1}$: 0.32 TPA $ICC_{2,1}$: 0.20 Household PA $ICC_{2,1}$: 0.40 LPA $ICC_{2,1}$: 0.38 Walking $ICC_{2,1}$: 0.72 MVPA $ICC_{2,1}$: 0.25 **Construct validity**: Total PA (counts/min) $r = 0.33$ Total PA (VM counts/min) $r = 0.33$ Total PA (MVPAmin/day) $r = 0.22$ Total PA (steps/day) $r = 0.37$ MVPA (counts/min) $r = 0.18$ MVPA (VM counts/min) $r = 0.21$ MVPA (MVPAmin/day) $r = 0.22$ MVPA (steps/day) $r = 0.25$	Reliability: Adequate Construct validity: Inadequate	Reliability: (−) Construct validity: (?), level of evidence = 2−/3−/3−/3−/2−/1−/1−/3−
Pardini et al., 2001[46]	IPAQ –long version	Reliability; Construct validity	Test re-test: one day apart Comparator: PA recordatory and accelerometer	43 (F: 21/M: 22); 24 (4.5); Young adults	**Reliability**: Total PA r_s: 0.71 **Construct validity**: PA Recordatory $r = 0.49$ Accelerometer $r = 0.24$	Reliability: Doubtful Construct validity: Inadequate	Reliability: (?) Construct validity: (?), level of evidence = 3?/2?

Table 3 (*Continued*)

Reference	Physical activity questionnaire[a]	Analysis performed	Study characteristics	Sample size (gender); Mean age (SD); Target population	Results	COSMIN rank	Quality criteria assessment
Lopes et al., 2015[77]	IPAQ – long version	Construct validity	Comparator: Questionnaire for physical activity and sedentary lifestyle	240 (F: 157/M: 83); 54.6 (range: 18–69 years); Adults with high blood pressure	**Construct validity**: Accuracy – ROC curve: 0.70 (95% CI: 0.64–0.75)	Construct validity: Doubtful	Construct validity: (?), level of evidence = 3+
Alves et al., 2010[67]	IPAQ – Short Version	Construct validity	Comparator: Celafisc criteria	173 (F: 98/M: 75); 40 (13); Adults	**Construct validity**: $k = 0.85$	Construct validity: Inadequate	Construct validity: (?), level of evidence = 3+
Matsudo et al., 2001[47]	IPAQ – Short Version	Reliability; Construct validity	Test re-test: 3–10 days apart Comparator: Accelerometer	**Reliability**: 257 (F: 149/M: 108); 36.8 (13.8) **Validity**: 28 (F: 16/M: 12) 42.9 (14.2); Adults	**Reliability**: Total PA ICC: 0.77 Total PA: $r_s = 0.74$ **Construct validity**: Total PA $r = 0.75$	Reliability: Adequate Construct validity: Inadequate	Reliability: (+) Construct validity: (?), level of evidence = 2?
Colpani et al., 2014[69]	IPAQ – Short Version	Construct validity	Comparator: Pedometer	292 (F: 292); 57.1 (5.3); Climateric women	**Construct validity**: $r = 0.13$	Construct validity: Inadequate	Construct validity: (?), level of evidence = 3–
Glaner et al., 2007[70]	IPAQ – Short Version	Construct validity	Comparator: VO_2max	105 (F: 28/M: 77); 24.8 (5.3); Young adults	**Construct validity**: % concordance = 47%	Construct validity: Inadequate	Construct validity: (?), level of evidence = 3–
Pinto et al., 2016[71]	IPAQ – Short Version	Construct validity	Comparator: Accelerometer	19; 14.6 (3.9); Juvenile dermatomyositis 20; 14.5 (2.4); Juvenile systemic lupus erythematosus	**Construct validity**: Total PA $r = 0.51$ JSLE Total PA $r = 0.29$ JDM Light-intensity PA and MVPA ranged from $r = 0.05$ to $r = 0.32$	Construct validity: Inadequate	Construct validity: (?), level of evidence = 2?/2?/2/1

Table 3 (*Continued*)

Reference	Physical activity questionnaire[a]	Analysis performed	Study characteristics	Sample size (gender); Mean age (SD); Target population	Results	COSMIN rank	Quality criteria assessment
Moraes et al., 2013[68]	IPAQ – Short Version	Construct validity	Comparator: American College of Sport Medicine criteria	2197 Adults	**Construct validity**: Male: $k = 0.95$ Female: $k = 0.93$	Construct validity: Inadequate	Construct validity: (?), level of evidence= 3+/3+
Lustosa et al., 2011[29]	MLTAQ	Translation and cross-cultural adaptation	Translation process	39 (F: 32/M: 7); 71.2 (6.8); Elderly	n/a	Cross-cultural validity: Doubtful	n/a
Bielemann et al., 2011[74]	NPAQ	Construct validity	Comparator: Accelerometer	239 (F: 123/M: 116); Youth	Daily counts $r = 0.24$ Mean counts per min $r = 0.21$ Sedentary activity $r = -0.08$ Moderate activity $r = 0.27$ Vigorous activity $r = 0.21$ Moderate to vigorous activity $r = 0.27$	Construct validity: Inadequate	Construct validity: (?), level of evidence = 1−/1−/3−/1−/1−/1−
Cruciani et al., 2011[19]	PACI	Translation and cross-cultural adaptation; Content validity	Translation process	24; 8.5 (1.5); Youth	n/a	Cross-cultural validity: Doubtful Content validity: Inadequate	Content validity: (−)
Adami et al., 2011[31]	PACI	Reliability; Measurement error	Test re-test: 3 h apart	83 (F: 42/M: 41); 9.3 (1.0); Youth	**Reliability**: PA time ICC: 0.89/$r = 0.83$ Total MET ICC: 0.91/$r = 0.87$ Total weighted MET ICC: 0.89/$r = 0.86$ Sedentary time ICC: 0.97/$r = 0.97$ **Measurement error**: PA time MD: 4.48 (LoA = 50 min) Total MET MD: 17.8 (LoA = 294.6 min) Total weighed MET MD: 25.8 (LoA = 298.4 min) Sedentary time MD: 1.35 (LoA = 47.4 min)	Reliability: Inadequate Measurement error: Inadequate	Reliability: (+) Measurement error: (?)

Table 3 (*Continued*)

Reference	Physical activity questionnaire[a]	Analysis performed	Study characteristics	Sample size (gender); Mean age (SD); Target population	Results	COSMIN rank	Quality criteria assessment
Adami et al., 2013[65]	PACI	Construct validity	Comparator: Accelerometer	83 (F: 42/M: 41); 9.3 (1.0); Youth	**Construct validity**: PA time – $r = 0.34$ (counts/min) MET – $r = 0.38$ (counts/min) Weighted MET – $r = 0.34$ (counts/min)	Construct validity: Inadequate	Construct validity: (?), level of evidence = 2−/2−/2−
Guedes et al., 2015[28]	PAQ-A	Translation and cross-cultural adaptation; Reliability; Construct validity	Translation process Test re-test: 14 days apart Comparator: Accelerometer	296 (F: 161/M: 135); F: 15.96 (1.25) M: 15.41 (1.09); Adolescents	**Reliability**: Total PA ICC: 0.77 **Construct validity**: Total PA: $r = 0.56$ MVPA: $r = 0.54$	Cross-cultural validity: Doubtful Reliability: Adequate Construct validity: Very good	Reliability: (+) Construct validity: (?), level of evidence = 2+/1+
Guedes et al., 2015[28]	PAQ-C	Translation and cross-cultural adaptation; Reliability; Construct validity	Translation process Test re-test: 14 days apart Comparator: Accelerometer	232 (F: 124/M: 108); F: 11.12 (1.38) M: 11.48 (1.15); Childrens	**Reliability**: Total PA ICC: 0.74 **Construct validity**: Total PA: $r = 0.40$ MVPA: $r = 0.48$	Cross-cultural validity: Doubtful Reliability: Adequate Construct validity: Very good	Reliability: (+) Construct validity: (?), level of evidence = 2−/1−
Takito et al., 2008[60]	PAQPW	Reliability; Construct validity	Test re-test: 7 days apart Comparator: Heart rate monitor	68 (F: 68); 26.9 (6.1); Pregnant women	**Reliability**: Sport $k = 0.41$ Vigorous PA $k = 0.32$ Moderate PA $k = 0.29$ Sedentary activity ICC: 0.81 Light PA ICC: 0.85 Moderate ICC: 0.75 Walking ICC: 0.80 **Construct validity**: LoA: 7–11 h	Reliability: Adequate Construct validity: Inadequate	Reliability: (+) for ICC (−) for Kappa Construct validity: (?), level of evidence = 3?
Neto et al., 2008[26]	PAR	Translation and cross-cultural adaptation; Reliability	Translation process Test re-test: 14 days apart	12 (F: 11/M: 1); 75 (4); Elderly population	Reliability: ICC: 0.92	Cross-cultural validity: Doubtful Reliability: Adequate	Reliability: (+)
Neto et al., 2011[75]	PAR	Construct validity	Comparator: VO_2max	98 (F: 43/M: 55); 67 (7); Elderly population	Construct validity: $r = 0.61$	Construct validity: Inadequate	Construct validity: (?), level of evidence = 3+

Table 3 (*Continued*)

Reference	Physical activity questionnaire[a]	Analysis performed	Study characteristics	Sample size (gender); Mean age (SD); Target population	Results	COSMIN rank	Quality criteria assessment
Damasceno et al., 2017[27]	PAR-3D	Translation and cross-cultural adaptation	Translation process	n/a	n/a	Cross-cultural validity: Doubtful	n/a
Farias Júnior et al., 2002[58]	PAR-3D	Reliability	Test re-test: 24 h apart	45 (F: 20/M: 25); 16 (1.28); Adolescents	**Reliability**: Habitual PA: ICC: 0.84 Light PA: ICC: 0.51 Moderate PA: ICC: 0.80 Vigorous PA: ICC: 0.78	Reliability: Adequate	Reliability: (+)
Ribeiro et al., 2011[81]	PAR-24	Development; Construct validity	Comparator: Accelerometer	98 (F: 65/M: 33); 39.4 (11); Adults	**Construct validity**: Counts – $r = 0.38$ kcal – $r = 0.31$	Construct validity: Very good	Construct validity: (?), level of evidence = 2—/2—
Militão et al., 2013[41] (Developed)	PASBEQ	Reliability; Construct validity; Internal consistency	Test re-test: 72 h apart Comparator: Shuttle run test (VO_2max)	**Reliability**: 47 **Validity**: 46 (F: 23/M: 23); 10–13 years old	**Reliability**: SPA ICC: 0.63–0.85 LPA (week days) ICC: 0.42–0.74 LPA (weekend) ICC: 0.44–0.75 Total LPA ICC: 1.00 TPA ICC: 0.60–0.86 PA in school ICC: 0.63–0.85 Total PA ICC: 0.61–0.84 **Construct validity**: SPA $r = 0.04$ LPA (week days) $r = 0.27$ LPA (weekend) $r = 0.28$ Total LPA $r = 0.35$ TPA $r = 0.07$ PA in school $r = 0.19$ Total PA $r = 0.37$ **Internal consistency**: Cronbach Alfa SPA – 0.86 LPA (week days) – 0.75 LPA (weekend) – 0.77 Total LPA – 1.0 TPA – 0.87 PA in school – 0.86 Total PA – 0.86	Reliability: Adequate Construct validity: Inadequate Internal consistency: Very good	Reliability: bv Unclear Construct validity: (?), level of evidence = 3? (all analysis) Internal consistency: (?)

Table 3 (*Continued*)

Reference	Physical activity questionnaire[a]	Analysis performed	Study characteristics	Sample size (gender); Mean age (SD); Target population	Results	COSMIN rank	Quality criteria assessment
Oliveira et al., 2011[56] (Developed)	PASBQ	Reliability	Test re-test: 7 days apart	65 (F: 27/M: 38); 4.2 (1.2); Youth	**Reliability**: Outdoor playtime $r_s = 0.92$ Sedentary behavior $r_s = 0.75$	Reliability: Doubtful	Reliability: (?)
Tavares et al., 2014[72]	PeNSE	Construct validity	Comparator: 24 h Recordatory	174 (F: 94/M: 80); 14.7; Adolescents	**Construct validity**: Accuracy: ≥300 min – 73.1% ≥150 min – 78.4% Inactive – 92.4%	Construct validity: Inadequate	Construct validity: (?), level of evidence = 3+/3+/3+
Silva et al., 2015[23]	PPAQ	Translation and cross-cultural adaptation	Translation process	305 (F: 305) Pregnant women	n/a	Cross-cultural validity: Doubtful	n/a
Farias Junior et al., 2012[32]	SAPAC	Reliability; Construct validity; Measurement error	Test re-test: 7 days apart Comparator: 24 h PA recalls	**Test re-test**: 239 (F: 133/M: 106); 16 (1.2) **Validity**: 70 (F: 39/M: 31); 15.7 (1.2); Youth	**Reliability**: ICC: 0.88/$k = 0.52$ **Construct validity**: All: rho = 0.62/$k = 0.59$ Male: rho = 0.52/0.41 Female: rho = 0.51/$k = 0.69$ 14–15 years: rho = 0.52/$k = 0.58$ 16–19 years: rho = 0.60/0.61 **Measurement error**: % of agreement: 75.7% LoA: 871.1 to −639.4	Reliability: Adequate Construct validity: Very good Measurement error: Very good	Reliability: (+) Construct validity: (?), level of evidence = 3−/3−/3−/3−/3− Measurement error: (?)
Prazeres Filho et al., 2017[61]	SAPAC	Reliability; Construct validity	Test re-test: 2 days apart Comparator: Accelerometer	**Test re-test**: 171 (F: 102/M: 69); 12.3 (1.1) **Validity**: 341 (F: 172/M: 169); 11.9 (1.0); Youth	**Reliability**: ICC: 0.73/$k = 0.58$ **Construct validity**: All: rho = 0.37 Male: rho = 0.38 Female: rho = 0.37 10–11 years: rho = 0.36 12–14 years: rho = 0.39	Reliability: Adequate Construct validity: Inadequate	Reliability: (+) Construct validity: (?), level of evidence = 2−
Checon et al., 2011[52] (Developed)	Saúdes	Reliability	Test re-test: 15 days apart	91 (F: 49/M: 42); Youth	**Reliability**: k or r_s: from −0.01 to 1.00	Reliability: Doubtful	Reliability: (?)

Table 3 (*Continued*)

Reference	Physical activity questionnaire[a]	Analysis performed	Study characteristics	Sample size (gender); Mean age (SD); Target population	Results	COSMIN rank	Quality criteria assessment
Nahas et al., 2007[57] (Developed)	SBQ	Reliability; Construct validity	Test re-test: 7 days apart Comparator: Pedometer	122 (F: 78/M: 44); 15.8 (1.6); Adolescents	**Reliability**: ICC from 0.76 to 0.93 **Construct validity**: $r = 0.23$	Reliability: Adequate Construct validity: Very good	Reliability: (+) Construct validity: (?), level of evidence = 3−
Hallal et al. 2013[78]	SVPAQ	Construct validity	Comparator: Doubly labeled water	25 (F: 16/M: 9); 13 (0.3); Adolescents	**Construct validity**: Total energy expenditure: $r = 0.41$ Physical activity energy expenditure: $r = 0.30$	Construct validity: Doubtful	Construct validity: (?), level of evidence = 1?
Monteiro et al., 2008[59] (Developed)	VIGITEL	Reliability; Construct validity	Test re-test: 7–15 days apart Comparator: 24 h Recordatory	**Reliability**: 110 (F: 63/M: 47); 45 **Construct validity**: 111 (F: 61/M: 50); 44; Adults	**Reliability**: Sufficient active in LPA: $k = 0.80$ Inactive in four domains of PA: $k = 0.78$ Television for long periods: $k = 0.53$ **Construct validity**: Specificity greater than 80% Sensibility: long period on TV: 69.7% Sedentary activity: 59.1% Sufficiently active in leisure: 50%	Reliability: Adequate Construct validity: Inadequate	Reliability: (+) Construct validity: (?), level of evidence = 3−/3−/3−

Table 3 (*Continued*)

Reference	Physical activity questionnaire[a]	Analysis performed	Study characteristics	Sample size (gender); Mean age (SD); Target population	Results	COSMIN rank	Quality criteria assessment
Moreira et al., 2017[39]	VIGITEL	Reliability; Construct-validity; Measurement error	Test re-test: 7–15 days apart Comparator: 24 h Recordatory	305 (F: 177/M: 128); 49.7 (18.2); Adults	**Reliability**: Active in LPA $k = 0.70$ Active in TPA $k = 0.35$ Inactive – $k = 0.64$ Watch Television – $k = 0.56$ **Construct validity**: PA Leisure Sensibility = 67.7% Specificity = 82.8% PA Locomotion Sensibility = 11.9% Specificity = 91.2% Sedentary Level Sensibility = 54.8% Specificity = 87.8% **Measurement error**: % of Agreement: 65%	Reliability: Adequate Construct validity: Inadequate Measurement error: Very good	Reliability: (−) Construct validity: (?), level of evidence = 3−/3−/3− Measurement error: (?)

Table presents the description of the instruments evaluated in alphabetical order, also the psychometric properties assessed in each study, relevant characteristics of each property, such as the time between each test–retest (reproducibility) analysis, comparator instrument (construct validity), sample characteristics, relevant results to the statistical analyzes carried out, classification regarding the methodological quality, classification of the results of the analyzes and level of evidence of the comparator instrument (construct validity). The type of ICC used was specified in subscript when informed by the article.

ICC, intraclass correlation coefficient; LoA, limits of agreement; LPA, leisure physical activity; MD, mean difference; MET, metabolic equivalent of task; MVPA, moderate and vigorous physical activity; n/a, not applicable; OPA, occupational physical activity; PA, physical activity; ROC, receiver operating characteristic; SD, standard deviation; SPA, sports physical activity; SPPB, short physical performance battery; TPA, transportation physical activity; VO_2, oxygen volume; WHODAS, Word Health Organization Disability Assessment Schedule.

(+) = sufficient rating, (?) = indeterminate rating, (−) = insufficient rating.

[a] For the abbreviation listed in the physical activity questionnaire refer to Table 2.

agreement (IPAQ-LV = −14.0 to 7.29 kcal/min; BASIC = −117 to 250 kcal).

Internal consistency

Four[22,24,40,41] studies performed the internal consistency analysis for the following questionnaires Habitual Physical Activity Baecke Questionnaire, Health Promoting Lifestyle Profile-II (HPLP-II), Physical Activity Level and Sedentary Behavior Evaluation Questionnaire (PASBEQ) and the Human Activity Profile Questionnaire (HAP). The methodological quality assessment revealed that three[22,40,41] studies were rated as 'Very Good' with Cronbach's alpha ranging from 0.52 to 1.0 (including analysis of sub-dimensions), whereas one[24] study was rated as 'Inadequate' due to inadequate statistical analysis. In the quality criteria assessment, all four studies received an indeterminate (?) rating because they failed to meet the criterion for low evidence for sufficient structural validity.

Reliability

Thirty-seven[25,26,30–39,41–59,28,60,61] reliability analyses were reported in the included studies. Of these, twenty-eight[25,26,32,36–39,41–44,47,49–51,53,54,57–59,28,60,61,48,62,63] analyses were rated as 'Adequate', six[33–35,46,52,56] rated as 'Doubtful', and three[30,31,64] rated as 'Inadequate' in the methodological quality assessment. The item most rated as 'Inadequate' referred to the time interval between test–retest and the item most rated as 'doubtful' referred to statistical method used. In addition, most studies were rated as 'Adequate' because they failed to include a detailed description for the test conditions and whether the individuals were stable in the interim period on the construct to be measured. The studies assessed reliability using ICC, correlation coefficient or Kappa coefficient and the interval for test-retest ranged from 3 h to 45 days. The ICC of the questionnaires varied from 0.20 to 1.00, the coefficient correlation from 0.49 to 1.00 correlation and the Kappa coefficients from −0.01 to 1.00. Overall, the most reliable questionnaire was the internet version of Questionnaire of a Typical Physical Activity and Food Intake to youth population, which was rated as 'Adequate' in the methodological quality assessment and showed an ICC of 0.94 for total physical activity score which indicates a positive rating in the quality criteria assessment. For healthy adults, the Baecke Physical Activity Questionnaire, International Physical Activity Questionnaire – long version and short version questionnaires were rated as 'Adequate' in the methodological quality assessment and achieved a sufficient (+) rating in the quality criteria assessment (ICC > 0.70). For the elderly population, the Baecke Physical Activity Questionnaire, International Physical Activity Questionnaire – long version and Physical Activity Rating was rated as 'Adequate' in the methodological quality assessment but only the International Physical Activity Questionnaire – long version (ICC = 0.88) and Physical Activity Rating (ICC 0.92) achieved a sufficient (+) rating in the quality criteria assessment. The reliability analyses for people with specific conditions (i.e. individuals with intermittent claudication and cardiovascular disease), the Baltimore Activity Scale for Intermittent Claudication and Godin Shepard Leisure-time Physical Activity Questionnaire were rated as 'Adequate' in the methodological quality assessment and both questionnaires achieved a sufficient (+) rating in the quality criteria assessment (ICC ranging from 0.84 to 0.87).

Content validity

Two[19,25] studies assessed the content validity of the Physical activity checklist interview and Godin shepard leisure-time physical activity questionnaire. In the methodological assessment, both studies were rated as 'Inadequate' because they failed to consult the target population about relevance and comprehensiveness of the questionnaire items. In the quality criteria assessment, both studies received an insufficient (−) rating because did not involve the target population in the process of content validation. Overall, the most studies focused predominantly on the terminologies and language expressions.

Construct validity

Among included studies, fifty-seven construct validity analyses were identified. Forty-four analyses[33,35,39,41,43,44,46–48,50,51,54,55,59–62,65–75] were rated as 'Inadequate' and six[34,51,63,76–78] rated as 'Doubtful' in the methodological quality assessment due to the lack or insufficient information on the measurement properties of the comparator instrument. Also, seven[32,57,28,79–81] analyses, reporting the results for Questionnaire of a Typical Physical Activity and Food Intake (previous day), Godin Shepard Leisure-time Physical Activity Questionnaire, Physical Activity Questionnaire for Adolescents, Physical Activity Questionnaire for Older Children, 24-Physical Activity Rating, Self-administered Physical Activity Checklist and *Saúde na Boa* Questionnaire, were rated as 'Very Good'. Regarding the level of evidence for the comparator instrument, ten[35,55,28,62,66,71,73,74,78] analyses (23%) used a comparator instrument classified as level 1, eight[46,47,61,62,65–67,81] analyses (18%) as level 2 and twenty-six[32–34,39,41,43,44,48–51,54,57,59,60,63,67–70,72,75–77,79,80] analyses (59%) as level 3. Thirty-five analyses reported the construct validity using objective measures as comparator, such as accelerometer[35,45–47,28,61,65–67,71,73,74,81] ($n = 16$), pedometer[33,43,44,50,57,69,70,76,79] ($n = 9$), frequency meter[51,60] ($n = 2$), doubly labeled water[78] ($n = 1$) and VO_2max[41,48,49,51,70,75,80] ($n = 7$). For the criterion validity, all studies received an indeterminate (?) rating because they failed to test prior-formulated hypotheses. The correlation coefficients between the questionnaire and the comparator instrument ranged from −0.08 to 0.75 with accelerometer as the comparator, from 0.12 to 0.57 with pedometer, from 0.05 to 0.08 with frequency meter, from 0.30 to 0.41 with doubly labeled water and from −0.14 to 0.61 with VO_2max. In addition, the physical activity questionnaires were compared with American College of Sports Medicine criteria, other questionnaires, recordatory and diary. The correlation coefficients of physical activity questionnaires with other self-report measures[32–34,39,43,46,54,59,63,68,72,77,80] ($n = 13$) ranged from 0.09 to 0.88 and the kappa coefficient ranged from 0.28 to 0.95.

Responsiveness

Only one[82] study assessed the responsiveness of a physical activity questionnaire. The methodological quality was

rated as 'Inadequate', due to the lack or insufficient information on measurement properties of the comparator instrument. In the quality criteria assessment, this study received an insufficient (−) rating, indicating that the result was not in accordance with the hypothesis.

Overall findings and level of evidence

Our findings showed that none of the questionnaire had their measurement properties fully tested. The Baecke Physical Activity Questionnaire and Physical Activity Checklist Interview were the most frequent investigated questionnaires, with five measurement properties evaluated. Only eight questionnaires had at least one measurement property classified as high for level of evidence, of those, the Baecke Physical Activity Questionnaire demonstrated high level of evidence for reliability, measurement error and internal consistency; and the Self-administered Physical Activity Checklist Questionnaire presented high level of evidence for reliability and construct validity. Table 4 presents the levels of evidence for each measurement property from each questionnaire.

Discussion

To our knowledge this is the first systematic review assessing the methodological quality of physical activity questionnaires translated to Brazilian-Portuguese. Systematic reviews on physical activity questionnaire translated to a specific language is important because valid and reliable questionnaires are needed in the context of clinical practice to evaluate and monitor physical activity outcomes as well as in the context of research to allow physical activity data to be generalized to a specific population and to be compared across countries.[18] Our findings identified a wide variety of questionnaires, designed for different target populations and assessing different constructs and dimensions of physical activity. In summary, our review shows that the International Physical Activity Questionnaire, the Baecke Physical Activity Questionnaire and the Physical Activity Checklist Interview were the most frequent investigated questionnaires. The measurement properties most frequently investigated among included studies were construct validity, reliability and translation. In addition, the methodological quality of the included studies revealed that majority of the included studies were rated as 'Inadequate', 'Doubtful' or 'Adequate'. Importantly, most of the questionnaires identified were found to have their measurement properties partially tested or even not tested. Previous systematic reviews investigating physical activity and sedentary behavior questionnaires also showed similar results.[17,18] The common methodological flaws found in this review were poor reporting methods, the lack of prior-formulated hypothesis, inadequate statistical analyses and sample size. Another possible explanation for these findings is that the COSMIN criteria is a considerably recent tool.

Strengths and limitations

A strength of our review was the use of two independent reviewers to perform the study selection, data extraction and quality rating. Our review also had some limitations that should be considered in the interpretation of the results. Although we conducted an extensively search in five electronic database aided by hand searching of reference list of included studies, we could not exclude the possibility of missing studies.

Physical activity questionnaires recommendation

The most recent recommendation for physical activity[5] states that adults should do at least 150 min of moderate-intensity throughout the week. Therefore, physical activity questionnaires should include the five domains (leisure-time, occupational, transportation, sports and household) to comprise total physical activity levels, as well as, duration and frequency of the practices. Of the included questionnaires ($n = 30$), only two questionnaires (i.e. International Physical Activity Questionnaire – long and short versions) included the five domains of physical activity. However, high level of evidence was demonstrated only for measurement error on International Physical Activity Questionnaire – long version. In contrast, questionnaires specifically designed to assess physical activity levels of children or adolescents included relevant domains for these populations (e.g. School-time PA). In addition, the choice of the questionnaire should involve the physical activity domain of interest, which does not necessarily characterize the individual's total physical activity level. Our findings showed that 'high' level of evidence were found for Baecke Physical Activity Questionnaire, Self-administered Physical Activity Checklist and Physical Activity Questionnaire of the Surveillance System of Risk Factors and Protection for Chronic Diseases on reliability; for Self-administered Physical Activity Checklist, Physical Activity Questionnaire for Adolescents, Physical Activity Questionnaire for Older Children and *Saúde na boa* Questionnaire on construct validity; for Baecke Physical Activity Questionnaire, Questionnaire of a Typical Physical Activity and Food Intake (Internet version) and International Physical Activity Questionnaire – long version on measurement error; and finally only for Baecke Physical Activity Questionnaire on internal consistency. For all other measurement properties, the evidence remains untested or with moderate, low or very low level of evidence due to poor methodological quality of the studies, insufficient quality criteria or even lack of evidence.

Recommendations for future research

The results of this review should be used to guide future high methodological quality studies investigating measurement properties of physical activity questionnaires. For instance, studies investigating construct validity of physical activity questionnaires should use reference instruments that measures the construct to be investigated[83] as well as test prior-formulated hypotheses. Additional studies are warranted in other measurement properties, such as content

Table 4 Summary of included studies and classification of measurement properties investigated.

Physical activity questionnaire[d]	Analysis performed	Total sample size (no. of studies)	Results	COSMIN rank (no. of studies)	Quality criteria assessment (no. of studies)	Grade
AAQ	Translation	–	–	Doubtful (1)	–	⊖(2)⊕⊕⊕ Low[a]
	Reliability	22 (1)	ICC: 0.97	Inadequate (1)	(+) (1)	⊖(3)⊕⊖(2)⊕ Very low[a,c]
BASIC	Reliability	38 (1)	ICC: 0.87	Adequate (1)	(+) (1)	⊖⊕⊖(2)⊕ Very low[a,c]
	Measurement error	38 (1)	LoA: −117 kcal to 250 kcal	Very Good (1)	(?) (1)	⊕⊕⊖(2)⊕ Low[3]
	Construct validity	150 (1)	$r = 0.34$	Doubtful (1)	(?) (1)	⊖(2)⊕⊕⊕ Low[a]
BPAQ	Translation	–	–	Doubtful (2)	–	⊖⊖⊕⊕ Low[a,b]
	Reliability	594 (7)	ICC: 0.44–0.85	Adequate (7)	(−) (3) (+) (7)	⊕⊕⊕⊕ High
	Measurement error	135 (1)	% of agreement: 95.5%	Adequate (1)	(?) (1)	⊕⊕⊕⊕ High
	Construct validity	577 (8)	$r = -0.14$ to 0.59	Inadequate (6) Doubtful (2)	(?) (8)	⊖⊖⊕⊕ Low[a,b]
	Internal consistency	326 (1)	Cronbach alpha: 0.52–0.76	Very Good (1)	(?) (1)	⊕⊕⊕⊕ High
DAFA	Reliability	170 (2)	ICC: 0.33–0.86	Adequate (2)	(+) (1) Unclear (1)	⊕⊖⊕⊕ Moderate[b]
	Construct validity	69 (1)	$k = 0.28$	Inadequate (1)	(?)	⊖(2)⊕⊖⊕ Very low[a,c]
DAFA-IV	Reliability	127 (1)	ICC: 0.94	Adequate (1)	(+) (1)	⊖⊕⊕⊕ Moderate[a]
	Measurement error	127 (1)	Mean error: 1.7	Very Good (1)	(?) (1)	⊕⊕⊕⊕ High
DAFA-PD	Construct validity	50 (1)	$r = 0.45$	Very Good (1)	(?) (1)	⊕⊕⊖⊕ Moderate[3]
DAFA-IVPD	Reliability	94 (1)	Incidence ratio: 0.63–7.52	Inadequate (1)	(?) (1)	⊖(3)⊕⊖⊕ Very low[a,c]
	Construct validity	390 (1)	Incidence ration: 0.52–18.1	Inadequate (1)	(?) (1)	⊖(3)⊕⊕⊕ Very low[a]
GSLTPAQ	Translation	–	–	Doubtful (1)	–	⊖(2)⊕⊕⊕ Low[a]
	Reliability	80 (1)	ICC: 0.79–0.84	Adequate (1)	(+) (1)	⊖⊕⊖⊕ Low[a,c]
	Content validity	80 (1)	–	Inadequate (1)	(−) (1)	⊖(3)⊕⊖⊕ Very low[a,c]
	Construct validity	236 (1)	$r = 0.03$–0.62	Very Good (1)	(+) (1)	⊕⊖⊕⊕ Moderate[b]

Table 4 (*Continued*)

Physical activity questionnaire[d]	Analysis performed	Total sample size (no. of studies)	Results	COSMIN rank (no. of studies)	Quality criteria assessment (no. of studies)	Grade
HAP	Translation	–	–	Doubtful (1)	–	⊖(2)⊕⊕⊕ Low[a]
	Internal Consistency	230 (1)	Rash analysis: 0.91	Inadequate (1)	(?) (1)	⊖(3)⊕⊕⊕ Very low[a]
	Construct validity	120 (1)	r = 0.55–0.69	Inadequate (1)	(?) (1)	⊖(3)⊕⊕⊕ Very low[a]
HPLP-II	Translation	–	–	Doubtful (1)	–	⊖(2)⊕⊕⊕ Low[a]
	Internal consistency	30 (1)	Cronbach alpha: 0.85–0.93	Very Good (1)	(?) (1)	⊕⊕⊖(2)⊕ Low[3]
IPAQ-LV	Reliability	571 (8)	ICC: 0.20–0.97 k = 0.25–1.00 r_s = 0.49–0.95	Adequate (4) Doubtful (4)	(?) 4 (−) 2 (+) 2	⊕⊖⊕⊕ Moderate[b]
	Measurement error	346 (3)	LoA: −14.0 to 131 min	Adequate (3)	(?) 3	⊕⊕⊕⊕ High
	Construct validity	827 (9)	r = 0.12–0.57 r_s = 0.09–0.51 k = 0.03–0.35 ROC curve: 0.70	Inadequate (7) Doubtful (2)	(?) (9)	⊖⊕⊕⊕ Moderate[a]
IPAQ-SV	Reliability	257 (1)	ICC: 0.77 r = 0.74	Adequate (1)	(+) (1)	⊖⊕⊕⊕ Moderate[a]
	Measurement error	39 (1)	% of agreement: 95% LoA: 236.6–278.8	Adequate (1)	(?) (1)	⊖(2)⊕⊖⊕ Very low[a,c]
	Construct validity	2834 (6)	k = 0.85–0.95 r = 0.05–0.75 % Concordance = 47%	Inadequate (6)	(?) (6)	⊖⊕⊕⊕ Low[a]
MLTAQ	Translation	30 (1)	–	Doubtful (1)	–	⊖(2)⊕⊕⊕ Low[a]
NPAQ	Construct validity	239 (1)	r = −0.08 to 0.27	Inadequate (1)	(?) (1)	⊖(3)⊕⊕⊕ Very low[a]
PACI	Translation	24 (1)	–	Doubtful (1)	–	⊖(2)⊕⊖(2)⊕ Very low[a,c]
	Content validity	24 (1)	–	Inadequate (1)	(−) (1)	⊖(3)⊕⊖(2)⊕ Very low[a,c]
	Reliability	83 (1)	ICC: 0.89–0.97 r = 0.83–0.97	Inadequate (1)	(+) (1)	⊖(3)⊕⊖⊕ Very low[a,c]
	Measurement error	83 (1)	Mean difference: 1.35–25.8 LoA: 50–294.6 min	Inadequate (1)	(?) (1)	⊖(3)⊕⊖⊕ Very low[a,c]
	Construct validity	83 (1)	r = 0.34–0.38	Inadequate (1)	(?) (1)	⊖(3)⊕⊖⊕ Very low[a,c]

Table 4 (*Continued*)

Physical activity questionnaire[d]	Analysis performed	Total sample size (no. of studies)	Results	COSMIN rank (no. of studies)	Quality criteria assessment (no. of studies)	Grade
PAQ-A	Translation	–	–	Doubtful (1)	–	⊖(2)⊕⊕⊕ Low[a]
	Reliability	296 (1)	ICC: 0.77	Adequate (1)	(+) (1)	⊖⊕⊕⊕ Moderate[a]
	Construct validity	296 (1)	$r = 0.54–0.56$	Very Good (1)	(?) (1)	⊕⊕⊕⊕ High
PAQ-C	Translation	–	–	Doubtful (1)	–	⊖(2)⊕⊕⊕ Low[a]
	Reliability	232 (1)	ICC: 0.74	Adequate (1)	(+) (1)	⊖⊕⊕⊕ Moderate[a]
	Construct validity	232 (1)	$r = 0.40–0.48$	Very Good (1)	(?) (1)	⊕⊕⊕⊕ High
PAQPW	Reliability	68 (1)	ICC: 0.75–0.85 $k = 0.29–0.41$	Adequate (1)	(+) (1)	⊖⊕⊖⊕ Low[a,c]
	Construct validity	68 (1)	LoA: 7–11 h	Inadequate (1)	(?) (1)	⊖(3)⊕⊖⊕ Very Low[a,c]
PAR	Translation	–	–	Doubtful (1)	–	⊖(2)⊕⊕⊕ Low[a]
	Reliability	12 (1)	ICC: 0.92	Adequate (1)	(+) (1)	⊖⊕⊖(2)⊕ Very Low[a,c]
	Construct validity	98 (1)	$r = 0.61$	Inadequate (1)	(?) (1)	⊖(3)⊕⊖⊕ Very Low[a,c]
PAR-3D	Translation	–	–	Doubtful (1)	–	⊖(2)⊕⊕⊕ Low[a]
	Reliability	45 (1)	ICC: 0.51–0.84	Adequate (1)	(+) (1)	⊖⊕⊖(2)⊕ Very Low[a,c]
PAR-24	Construct validity	98 (1)	$r = 0.31–0.38$	Very Good (1)	(?) (1)	⊕⊕⊖⊕ Moderate[3]
PASBEQ	Reliability	47 (1)	ICC: 0.42–1.00	Adequate (1)	Unclear (1)	⊖⊖⊖(2)⊕ Very Low[a,b,c]
	Construct validity	46 (1)	$r = 0.04–0.37$	Inadequate (1)	(?) (1)	⊖(3)⊕⊖(2)⊕ Very Low[a,c]
	Internal consistency	46 (1)	Cronbach alfa: 0.75–1.00	Very Good (1)	(?) (1)	⊕⊕⊖(2)⊕ Low[a]
PASBQ	Reliability	65 (1)	$r_s = 0.75–0.92$	Doubtful (1)	(?) (1)	⊖(2)⊕⊖⊕ Very Low[a,c]
PeNSE	Construct validity	174 (1)	Accuracy: 73.1%–92.4%	Inadequate (1)	(?) (1)	⊖(3)⊕⊕⊕ Very Low[a]
PPAQ	Translation	305 (1)	–	Doubtful (1)	–	⊖(2)⊕⊕⊕ Low[a]

Table 4 (*Continued*)

Physical activity questionnaire[d]	Analysis performed	Total sample size (no. of studies)	Results	COSMIN rank (no. of studies)	Quality criteria assessment (no. of studies)	Grade
SAPAC	Reliability	410 (2)	ICC: 0.73–0.88 k = 0.52–0.58	Adequate (2)	(+) (2)	⊕⊕⊕⊕ High
	Measurement error	239 (1)	% of agreement: 75.7% LoA: 871.1–-639.4	Adequate (1)	(?) (1)	⊖⊕⊕⊕ Moderate[a]
	Construct validity	411 (2)	rho = 0.36–0.62 k = 0.41–0.69	Very Good (1) Inadequate (1)	(?) (2)	⊕⊕⊕⊕ High
Saudes	Reliability	91 (1)	k or r_s: from −0.01 to 1.00	Doubtful (1)	(?) (1)	⊖(2)⊖⊖⊕ Very low[a,b,c]
SBQ	Reliability	122 (1)	ICC: 0.76–0.93	Adequate (1)	(+) (1)	⊖⊕⊕⊕ Moderate[a]
	Construct validity	122 (1)	r = 0.23	Very Good (1)	(?) (1)	⊕⊕⊕⊕ High
SVPAQ	Construct validity	25 (1)	r = 0.30–0.41	Doubtful (1)	(?) (1)	⊖(2)⊕⊖(2)⊕ Very low[a,c]
VIGITEL	Reliability	415 (2)	k = 0.35–0.80	Adequate (2)	(+) (1) (−) (1)	⊕⊕⊕⊕ High
	Measurement error	305 (1)	PoA: 65%	Adequate (1)	(?) (1)	⊖⊕⊕⊕ Moderate[a]
	Construct validity	416 (2)	Specificity: >80% Sensibility: 11%–69.7%	Inadequate (2)	(?) (2)	⊖(2)⊕⊖⊕ Low[1]

[a] Downgraded for risk of bias by one level if there is serious risk of bias (i.e. multiple studies of doubtful quality available, or one study of adequate quality), two levels (e.g. from high to low) if there is very serious risk of bias (i.e. multiple studies of inadequate quality, or one study of doubtful quality available), or three levels (i.e. from high to very low) of there is extremely risk of bias (i.e. only one study of inadequate quality available).
[b] Downgraded for inconsistency if the results are inconsistent (i.e. interpretation of the summary results generates different interpretations considering the range).
[c] Downgraded for imprecision by one level if the total sample size of the summary results was below 100 and two level if the total sample size is below 50.
[d] For the abbreviation listed in the physical activity questionnaire refer to Table 2.

validity and responsiveness. In addition, the COSMIN checklist should be used in future studies to ensure the high methodological quality.

Conclusion

Given the results obtained with this review, few conclusions could be made about the best physical activity questionnaire, since many of them did not have their measurement properties fully tested and the studies showed in general poor methodological quality. Nevertheless, the questionnaires Baecke Physical Activity Questionnaire for adults and the Self-administered Physical Activity Checklist for youth demonstrated better scores considering methodological quality, quality criteria and also high level of evidence for some measurement properties tested. Therefore, further high methodological quality studies investigating the measurement properties of physical activity questionnaires are still needed in this area.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

F.G.S was supported with a scholarship, grant number 2014/09560-1, from Sao Paulo Research Foundation (FAPESP).

Appendix 1.

Search strategy for Medline database:

1 exp exercise/
2 physical inactivity.mp.
3 physical activity.mp.
4 exp motor activity/
5 Physical Fitness/
6 sedentary.ab. or sedentary.ti.
7 exp life style/
8 exp leisure activities/
9 exp walking/
10 exp sports/
11 (exercise$ adj aerobic$).tw.
12 (physical$ adj5 (fit$ or train$ or activ$ or endur$)).tw.
13 (exercis$ adj5 (train$ or physical$ or activ$)).tw.
14 sport$.tw.
15 walk$.tw.
16 cycle$.tw.
17 ((''lifestyle'' or life-style) adj5 activ$).tw.
18 ((''lifestyle'' or life-style) adj5 physical$).tw.
19 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18
20 Questionnaires/
21 index.mp.
22 scale.mp.
23 score.mp.
24 Patient Outcome Assessment/ or Self-Assessment/
25 Evaluation Studies as Topic/
26 Psychometrics/ or Self Report/
27 inventory.mp.
28 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27
29 Brazil/
30 brasil.mp.
31 Brazilian.mp.
32 Brazilian Portuguese.mp.
33 29 or 30 or 31 or 32
34 19 and 28 and 33

References

1. Haskell WL, Lee IM, Pate RR, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Med Sci Sports Exerc*. 2007;39(8):1423–1434.
2. Lee IM, Shiroma EJ, Lobelo F, et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet*. 2012;380(9838):219–229.
3. Blair SN. Physical inactivity: the biggest public health problem of the 21st century. *Br J Sports Med*. 2009;43(1):1–2.
4. Caspersen CJ, Powell KE, Christenson GM. Physical-activity, exercise, and physical-fitness: definitions and distinctions for health-related research. *Public Health Rep*. 1985;100(2):126–131.
5. WHO. *Global recommendations on physical activity for health*. Geneva: World Health Organization; 2010. Available from: https://www.ncbi.nlm.nih.gov/books/NBK305057/.
6. Prince SA, Adamo KB, Hamel MB, Hardt J, Gerber SC, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int J Behav Nutr Phys Act*. 2008;5(56):24.
7. Nahas MV. Revisão de métodos para determinação dos níveis de atividade física habitual em diversos grupos populacionais. *Rev Bras Ativ Fis Saúde*. 1996;1(4):27–37.
8. Trost SG, O'Neil M. Clinical use of objective measures of physical activity. *Br J Sports Med*. 2014;48(3):178–181.
9. Dishman RK, Washburn RA, Schoeller DA. Measurement of physical activity. *Quest*. 2001;53(3):295–309.
10. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*. 2000;25(24):3186–3191.
11. Warren JM, Ekelund U, Besson H, et al. Assessment of physical activity – a review of methodologies with reference to epidemiological research: a report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. *Eur J Cardiovasc Prev Rehabil*. 2010;17(2):127–139.
12. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147–1157.
13. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1171–1179.
14. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018;27(5):1159–1170.
15. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34–42.
16. Martins JC, Aguiar LT, Nadeau S, et al. Measurement properties of self-report physical activity assessment tools for patients with stroke: a systematic review. *Braz J Phys Ther*. 2019;23(6):476–490.
17. Hidding LM, Altenburg TM, Mokkink LB, Terwee CB, Chinapaw MJM. Systematic review of childhood sedentary behavior questionnaires: what do we know and what is next? *Sports Med*. 2017;47(4):677–699.
18. van Poppel MN, Chinapaw MJ, Mokkink LB, van Mechelen W, Terwee CB. Physical activity questionnaires for adults: a systematic review of measurement properties. *Sports Med*. 2010;40(7):565–600.
19. Cruciani F, Adami F, Assuncao NA, Bergamaschi DP. Conceptual, item, and semantic equivalence of a Brazilian version of the Physical Activity Checklist Interview (PACI). *Cad Saude Publica*. 2011;27(1):19–34.
20. Florindo AA, Latorre MR, Jaime PC, Tanaka T, Zerbini CA. Methodology to evaluation the habitual physical activity in men aged 50 years or more. *Rev Saude Publica*. 2004;38(2):307–314.
21. Sardinha A, Levitan MN, Lopes FL, et al. Translation and cross-cultural adaptation of the Habitual Physical Activity Questionnaire. *Rev Psiquiatr Clin*. 2010;37(1):16–22.
22. Tajik M, Galvao HM, Siqueira EC. Health survey instrument development through a community-based participatory research approach: Health Promoting Lifestyle Profile (HPLP-II) and Brazilian immigrants in greater Boston. *J Immigr Minor Health*. 2010;12(3):390–397.
23. Silva FT, Araujo Junior E, Santana EF, Lima JW, Cecchino GN, Silva Costa FD. Translation and cross-cultural adaptation of the Pregnancy Physical Activity Questionnaire (PPAQ) to the Brazilian population. *Ceska Gynekol*. 2015;80(4):290–298.
24. Souza AC, Magalhães LC, Teixeira-Salmela LF. Cross-cultural adaptation and analysis of the psychometric properties in the Brazilian version of the human activity profile. *Cad Saude Publica*. 2006;22(12):2623–2636.
25. Sao-Joao TM, Rodrigues RCM, Gallani MCBJ, Miura CTP, Domingues GBL, Godin G. Cultural adaptation of the Brazilian

version of the Godin-Shephard Leisure-Time Physical Activity Questionnaire. *Rev Saude Publica*. 2013;47(3):479–487.
26. Maranhão Neto GA, Leon ACMP, Farinatti PTV. Cross-cultural equivalence of three scales used to estimate cardiorespiratory fitness in the elderly. *Cad Saude Publica*. 2008;24(11):2499–2510.
27. Damasceno VO, Silva AC, Amorim PR, et al. Adaptação transcultural do questionário three day physical activity recall. *Rev Bras Med Esporte*. 2017;23(2):93–97.
28. Guedes DP, Guedes JERP. Measuring physical activity in Brazilian youth: reproducibility and validity of the paq-c and paq-a. *Rev Bras Med Esporte*. 2015;21(6):425–432.
29. Lustosa LP, Pereira DS, Dias RC, Britto RR, Parentoni AN, Pereira LSM. Translation and cultural adaptation of the Minnesota Leisure Time Activities Questionnaire in community-dwelling older people. *Rev Bras Geriatr Gerontol*. 2011;5(2):57–65.
30. Rocha VTM, Soares TM, Leopoldino AAO, et al. Cross-cultural adaptation and reliability of the active Australia questionnaire for the elderly. *Rev Bras Med Esporte*. 2017;23(1):46–49.
31. Adami F, Cruciani F, Douek M, et al. Reliability of the Brazilian version of the physical activity checklist interview in children. *Rev Saude Publica*. 2011;45(2):321–333.
32. Farias Júnior JC, Lopes AS, Mota J, Santos MP, Ribeiro JC, Hallal PC. Validade e reprodutibilidade de um questionário para medida de atividade física em adolescentes: uma adaptação do Self-Administered Physical Activity Checklist. *Rev Bras Epidemiol*. 2012;15(1):198–210.
33. Benedetti TRB, Antunes PC, Rodriguez-Añez CR, Mazo GZ, Petroski EL. Reprodutibilidade e validade do Questionário Internacional de Atividade Física (IPAQ) em homens idosos. *Rev Bras Med Esporte*. 2007;13(1):11–16.
34. Guedes DP, Lopes CC, Guedes JERP. Reproducibility and validity of the International Physical Activity Questionnaire in adolescents. *Rev Bras Med Esporte*. 2005;11(2):151–158.
35. Hallal PC, Simoes E, Reichert FF, et al. Validity and reliability of the telephone administered international physical activity questionnaire in Brazil. *J Phys Act Health*. 2010;7(3):402–409.
36. Romero A, Florindo AA, Voci SM, Slater B. Reliability of computerized questionnaire of physical activity in adolescents. *Rev Bras Ativ Fis Saúde*. 2011;16(3):234–239.
37. Legnani E, Legnani RFS, Rech CR, Barros MVG, Campos W, Assis MAA. Agreement and reliability of an electronic questionnaire for children (WEBDAFA). *Rev Bras Cineantropom Desempenho Hum*. 2013;15(1):38–48.
38. Souza Barbosa JP, Lima RA, Gardner AW, de Barros MV, Wolosker N, Ritti-Dias RM. Reliability of the Baltimore activity scale questionnaire for intermittent claudication. *Angiology*. 2012;63(4):254–258.
39. Moreira AD, Claro RM, Felisbino-Mendes MS, Velasquez-Melendez G. Validity and reliability of a telephone survey of physical activity in Brazil. *Rev Bras Epidemiol*. 2017;20(1):136–146.
40. Florindo AA, Latorre MRDO, Jaime PC, Tanaka T, Zerbini CAF. Methodology to evaluation the habitual physical activity in men aged 50 years or more. *Rev Saude Publica*. 2004;38(2):307–314.
41. Militão AG, Silva FR, Peçanha LM, Souza JWS, Militão ESG, Campbell CSG. Reproducibility and validity of a physical activity level and sedentary behavior evaluation questionnaire for school students aged 10 to 13 years, Federal District, Brazil, 2012. *Epidemiol Serv Saúde*. 2013;22(1):111–120.
42. Barros MVG, Nahas MV. Reprodutibilidade (teste–reteste) do questionário internacional de atividade física (QIAF-Versão 6): um estudo piloto com adultos no Brasil. *Rev Bras Cienc Mov*. 2000;8(1):23–26.
43. Benedetti TB, Mazo GZ, de Barros MVG. Application of the International Physical Activity Questionnaire (IPAQ) for evaluation of elderly women: concurrent validity and test–retest reproducibility. *Rev Bras Cienc Mov*. 2004;12(1):25–33.
44. Lima RA, Freitas CMSM, Smethurst WS, Santos CM, Barros MVG. Physical activity level among elderly with Alzheimer disease by using the IPAQ questionnaire and pedometers. *Rev Bras Ativ Fis Saúde*. 2010;15(3):180–185.
45. Carvalho FA, Maher CG, Franco MR, et al. Fear of movement is not associated with objective and subjective physical activity levels in chronic nonspecific low back pain. *Arch Phys Med Rehabil*. 2017;98(1):96–104.
46. Pardini R, Matsudo S, Araújo T, et al. Validation of the International Physical Activity Questionaire (IPAQ version 6): pilot study in Brazilian young adults. *Rev Bras Cienc Mov*. 2011;9(3):45–51.
47. Matsudo S, Araujo T, Matsudo V, et al. Questionário Internacional de Atividade Física (IPAQ): Estudo de validade e reprodutibilidade no Brasil. *Rev Bras Ativ Fís Saúde*. 2001;6(2):1–14.
48. Florindo AA, Latorre MRDO. Validation and reliability of the Baecke questionnaire for the evaluation of habitual physical activity in adult men. *Rev Bras Med Esporte*. 2003;9(3):129–135.
49. Florindo AA, Latorre MRDO, Santos EC, Negrão CE, Azevedo LF, Segurado AA. Validity and reliability of the Baecke questionnaire for the evaluation of habitual physical activity among people living with HIV/AIDS. *Cad Saude Publica*. 2006;22(3):535–541.
50. Mazo GZ, Mota J, Benedetti TB, Barros MVG. Validade concorrente e reprodutibilidade teste-reteste do questionário de Baecke modificado para idosos. *Rev Bras Ativ Fís Saúde*. 2001;6(1):5–11.
51. Florindo AA, Romero A, Peres SV, Silva MV, Slater B. Development and validation of a physical activity assessment questionnaire for adolescents. *Rev Saude Publica*. 2006;40(5):802–809.
52. Checon K, Fonseca VM, Faria CP, Carletti L, Molina MCB. The reproducibility of the Saúdes – Vitória Study's physical activity assessment questionnaire for children. *Rev Bras Saúde Mater Infant*. 2011;11(2):173–180.
53. Costa FF, Liparotti JR. Reliability of a new questionnaire for the evaluation of habitual physical activity and food consumption in children. *Rev Bras Cineantropom Desempenho Hum*. 2010;12(1):21–28.
54. Barros MVG, Assis MAA, Pires MC, et al. Validity of physical activity and food consumption questionnaire for children aged seven to ten years old. *Rev Bras Saúde Mater Infant*. 2007;7(4):437–448.
55. Jesus GM, Assis MAA, Kupek E, Dias LA. Assessment of physical activity in schoolchildren using a web-based questionnaire. *Rev Bras Med Esporte*. 2016;22(4):261–266.
56. Oliveira NKR, Lima RA, Mélo EN, Santos CM, Barros SSH, Barros MVG. Reprodutibilidade de questionário para medida da atividade física e comportamento sedentário em crianças pré-escolares. *Rev Bras Ativ Fís Saúde*. 2011;16(3):228–233.
57. Nahas MV, Barros MVG, Florindo AA, Farias Junior JCD. Reprodutibilidade e validade do questionário saúde na boa para avaliar atividade física e hábitos alimentares em escolares do ensino médio. *Rev Bras Ativ Fís Saúde*. 2007;(1):12–20.
58. Farias Junior JC, Pires MC, Lopes AS. Questionnaire's reliability to assess health-related behaviors in adolescents. *Rev Bras Cienc Mov*. 2002;10(3):43–48.
59. Monteiro CA, Florindo AA, Claro RM, Moura EC. Validity of indicators of physical activity and sedentariness obtained by telephone survey. *Rev Saude Publica*. 2008;42(4):575–581.
60. Takito MY, Neri LCL, Benicio MHAD. Evaluation of the reproducibility and validity of a physical activity questionnaire for pregnant women. *Rev Bras Med Esporte*. 2008;14(2):132–138.

61. Prazeres Filho A, Barbosa AO, Mendonça G, Farias Junior JC. Reproducibility and concurrent validity of the Physical Activity Questionnaire for Adolescents (QAFA) aged 10–14 years. *Rev Bras Cineantropom Hum*. 2017;17(3):270–282.
62. Carvalho FA, Morelhao PK, Franco MR, et al. Reliability and validity of two multidimensional self-reported physical activity questionnaires in people with chronic low back pain. *Musculoskelet Sci Pract*. 2017;27:65–70.
63. Guedes DP, Lopes CC, Guedes JERP, Stanganelli LC. Reprodutibilidade e validade do questionário Baecke para avaliação da atividade física habitual em adolescentes. *Rev Port Cien Desp*. 2006;6(3):265–274.
64. Jesus GM, Assis MAA, Kupek E, Dias LA. Assessment of physical activity in school children using a web-based questionnaire. *Rev Bras Med Esporte*. 2016;22(4):261–266.
65. Adami F, Bergamaschi DP, Hinnig PF, Oliveira NS. Validity study of the ''Physical Activity Checklist'' in children. *Rev Saude Publica*. 2013;47(3):488–496.
66. Garcia LMT, Osti RFI, Ribeiro EHC, Florindo AA. Validation of two questionnaires to assess physical activity in adults. *Rev Bras Ativ Fis Saúde*. 2013;18(3):317–331.
67. Alves VV, Borges KCS, Ribeiro LFP, Gadelha SR, Santos SC. Agreement between criteria for physical activity level categorization from the international physical activity questionnaire. *Rev Bras Ativ Fis Saúde*. 2010;15(2):111–114.
68. Moraes SA, Suzuki CS, Freitas ICM. Comparison between the International Physical Activity Questionnaire and the American College of Sports Medicine/American Heart Association criteria to classify the physical activity profile in adults. *Rev Lat Am Enfermagem*. 2013;21(4):835–840.
69. Colpani V, Spritzer PM, Lodi AP, et al. Physical activity in climacteric women: comparison between self-reporting and pedometer. *Rev Saude Publica*. 2014;48(2):258–265.
70. Glaner MF. Agreement of physical activity questionnaire with cardiorespiratory fitness. *Rev Bras Cineantropom Desempenho Hum*. 2007;9(1):61–66.
71. Pinto AJ, Roschel H, Benatti FB, et al. Poor agreement of objectively measured and self-reported physical activity in juvenile dermatomyositis and juvenile systemic lupus erythematosus. *Clin Rheumatol*. 2016;35(6):14–1507.
72. Tavares LF, Castro IR, Cardoso LO, Levy RB, Claro RM, Oliveira AF. Validity of indicators on physical activity and sedentary behavior from the Brazilian National School-Based Health Survey among adolescents in Rio de Janeiro, Brazil. *Cad Saude Publica*. 2014;30(9):1861–1874.
73. Bastone AC, Moreira BS, Vieira RA, Kirkwood RN, Dias JM, Dias RC. Validation of the human activity profile questionnaire as a measure of physical activity levels in older community-dwelling women. *J Aging Phys Act*. 2014;22(3):348–356.
74. Bielemann RM, Reichert FF, Paniz VMV, Gigante DP. Validation of the Netherlands physical activity questionnaire in Brazilian children. *Int J Behav Nutr Phys Act*. 2011;8(45):1–8.
75. Maranhão Neto GA, Leon ACP, Farinatti PTV. Validade do critério e equivalência de mensutação da escala Physical Activity Rating em idosos. *Rev Bras Ativ Fis Saúde*. 2011;16(4):295–299.
76. Lopes PR, Barbosa JPAS, Farah BQ, et al. Correlation between physical activity levels of patients with intermittent claudication estimated using the Baltimore Activity Scale for Intermittent Claudication and a pedometer. *J Vasc Bras*. 2013;12(3):187–192.
77. Lopes MV, Silva VM, Araujo TL, Guedes NG, Martins LC, Teixeira IX. Instrument for evaluation of sedentary lifestyle in patients with high blood pressure. *Rev Bras Enferm*. 2015;68(3):443–449.
78. Hallal PC, Reichert FF, Clark VL, et al. Energy expenditure compared to physical activity measured by accelerometry and self-report in adolescents: a validation study. *PLoS ONE*. 2013;(11):e67703.
79. Cabral LGA, Costa FF, Liparotti JR. Preliminary validation of the physical activity section of the Previous Day Physical Activity and Food Consumption Questionnaire. *Rev Bras Ativ Fis Saúde*. 2011;16(2):100–106.
80. Joao TMS, Rodrigues RCM, Gallani MCBJ, et al. Validity of the Brazilian version of the Godin-Shephard Leisure-Time Physical Activity Questionnaire. *Cad Saude Publica*. 2015;31(9):1825–1838.
81. Ribeiro EH, Costa EF, Sobral GM, Florindo AA. Development and validation of a 24-hour recall for the evaluation of physical activity. *Rev Bras Ativ Fis Saúde*. 2011;16(2):132–137.
82. Morelhão PK, Franco MR, Oliveira CB, et al. Physical activity and disability measures in chronic non-specific low back pain: a study of responsiveness. *Clin Rehabil*. 2018;32(12):1684–1695.
83. Terwee CB, Mokkink LB, van Poppel MN, Chinapaw MJ, van Mechelen W, de Vet HC. Qualitative attributes and measurement properties of physical activity questionnaires: a checklist. *Sports Med*. 2010;40(7):525–537.