REVIEW ARTICLE

Sports Med 2010; 40 (7): 565-600
0112-1642/10/0007-0565/$49.95/0

Physical Activity Questionnaires for Adults

A Systematic Review of Measurement Properties

Mireille N.M. van Poppel,[1] *Mai J.M. Chinapaw,*[1] *Lidwine B. Mokkink,*[2] *Willem van Mechelen*[1] and *Caroline B. Terwee*[2]

1 Department of Public and Occupational Health, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands
2 Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands

Contents

Abstract

Many questionnaires have been developed to measure physical activity (PA), but an overview of the measurement properties of PA questionnaires is lacking. A summary of this information is useful for choosing the best questionnaire available. Therefore, the objective of this study was to evaluate and compare measurement properties of self-administered questionnaires assessing PA in adults. We searched MEDLINE, EMBASE and SportDiscus®, using 'exercise', 'physical activity', 'motor activity' and 'questionnaire' as keywords. We included studies that evaluated the measurement properties of self-report questionnaires assessing PA. Article

selection, data extraction and quality assessment were performed by two independent reviewers. The quality and results of the studies were evaluated using the Quality Assessment of Physical Activity Questionnaires (QAPAQ) checklist. Construct validity, reliability and responsiveness were rated as positive, negative or indeterminate, depending on the methods and results. We included 85 (versions of) questionnaires. Overall, the quality of the studies assessing measurement properties of PA questionnaires was rather poor. Information on content validity was mostly lacking. Construct validity was assessed in 76 of the questionnaires, mostly by correlations with accelerometer data, maximal oxygen uptake or activity diaries. Fifty-one questionnaires were tested for reliability. Only a few questionnaires had sufficient construct validity and reliability, but these need to be further validated. Responsiveness was studied for only two questionnaires and was poor. There is a clear lack of standardization of PA questionnaires, resulting in many variations of questionnaires. No questionnaire or type of questionnaire for assessing PA was superior and therefore could not be strongly recommended above others. In the future, more attention should be paid to the methodology of studies assessing measurement properties of PA questionnaires and the quality of reporting.

Adequately measuring physical activity (PA) is important for determining trends in PA levels over time, for evaluation of the effect of PA interventions and for determining health benefits of PA. Poor measurement of PA may hinder detection of important associations or effects.[1] Many questionnaires have been developed to measure PA. Some questionnaires were developed specifically for a certain subgroup or setting, others because researchers were not aware of existing questionnaires or because they were not satisfied with available questionnaires. Often researchers needed to translate and/or adapt existing questionnaires to other target groups. This has led to a large number of (versions of) questionnaires available, which makes it difficult to choose the most suitable instrument. Furthermore, the use of different instruments in different studies and surveys makes comparison of PA levels across countries or studies difficult.

To our knowledge, an overview of the measurement properties of PA questionnaires is lacking. A summary of these findings might be helpful for choosing the best questionnaire available for a specific purpose. Furthermore, a critical assessment of the methodological quality of the studies assessing the measurement properties of PA questionnaires is lacking, while the methodological quality of these studies might be variable. If the methodological quality of a study is poor, the results and conclusions can be seriously biased. For example, wrong conclusions can be drawn from a validation study if no adequate comparison instrument was used. It is therefore important to assess the methodological quality of a study to be confident that the design, conduct, analysis and interpretation of the study is adequate, and to inform about possible bias that might have influenced the results.

In this article, we aim to evaluate and compare the measurement properties of all available self-administered questionnaires measuring PA in adults, using a systematic approach for the literature search, data extraction and assessment of the quality of the studies. This article is one of a series of four articles on measurement properties of PA questionnaires published in *Sports Medicine*.

1. Methods

1.1 Literature Search

Literature searches were performed in PubMed, EMBASE using 'EMBASE only', and in

SportDiscus® (complete databases until May 2009) on the topic of self-report questionnaires of PA. Additional papers were identified by manually searching references of the retrieved papers and the authors' own literature databases.

The full search strategy in PubMed was as follows: (exercise[MeSH] OR 'physical activity' [tiab] OR motor activity[MeSH]) AND (questionnaire[MeSH] OR questionnaire*[tiab]), and limited to humans. In EMBASE and SportDiscus®, 'physical activity' and 'questionnaire' were used as free text words and in EMBASE this was complemented with the EMTREE term 'exercise'.

1.2 Eligibility Criteria

We used the following inclusion criteria:

1. The aim of the study should be to develop or evaluate the measurement properties – i.e. content validity, construct validity, reliability or responsiveness – of a self-report questionnaire.
2. The aim of the questionnaire should be to measure PA, which was defined as any bodily movement produced by skeletal muscles that results in energy expenditure above resting level.[2] PA in daily life can be categorized into occupational, sports, conditioning, household or other activities. Questionnaires were included regardless of the time frame; thus, questionnaires measuring lifetime PA or historical activity were also included.
3. The questionnaire could be used to measure PA in adults in the general population, and was not developed or evaluated in a specific population, such as patients or pregnant or obese participants.
4. The study sample should have a mean age between 18 and 55 years.
5. The article should have been published in the English language.
6. Information on (at least one of) the measurement properties of the self-report questionnaire should be provided. We included information on measurement properties only if it was intentionally collected or calculated to assess the measurement properties of the particular self-report questionnaire. If, for example, correlations between a self-report questionnaire and an accelerometer were presented to assess the validity of the accelerometer (while the self-report questionnaire was used as a gold standard) or if correlations between different PA questionnaires were calculated without one questionnaire considered as the standard, these data were not included in this review.

We excluded PA interviews or diaries. We also excluded studies that evaluated the measurement properties of a self-report questionnaire administered in an interview form. Finally, questionnaires measuring physical functioning (e.g. the degree to which one is limited in carrying out activities) and questionnaires asking about sweating in a single question were excluded.

1.3 Selection of Papers

Abstract selection, selection of full-text articles, data extraction and quality assessment were performed by two independent reviewers. Disagreements were discussed and resolved. We retrieved the full-text paper of all abstracts that fulfilled the inclusion criteria and of abstracts that did not contain measurement properties, but in which indications were found that these properties were presented in the full-text paper.

1.4 Data Extraction

We extracted a description of the self-report questionnaires from the included papers, using a standardized data extraction form. Data extracted included (i) the target population for which the questionnaire was developed; (ii) the dimension(s) of PA that the questionnaire intends to measure (e.g. habitual PA); (iii) the parameters of PA that the questionnaire is measuring (i.e. frequency, duration and intensity or activities); (iv) the setting in which PA is being measured (i.e. sport, recreational, transport, occupational/school activities, household activities [including gardening], other); (v) the number of questions; (vi) the recall period that the questions refer to; and (vii) the type and number of scores that were calculated (e.g. total energy expenditure or minutes of activity per day).

1.5 Quality Assessment of the Studies on Measurement Properties

To assess the methodological quality and results of the studies on measurement properties, we used the QAPAQ checklist (see table I for acronym definitions). We developed this checklist specifically for PA questionnaires, based on two recently developed checklists to evaluate the measurement properties of patient-reported outcomes COSMIN[8] and self-report health status questionnaires.[33] The QAPAQ is described elsewhere.[29] We extracted and rated the methods and results of all evaluated measurement properties (see sections 1.7–1.9).

1.6 Content Validity

No criterion exists to rate whether the content of a questionnaire is relevant and comprehensive for measuring PA. Therefore, we formed our own opinion on content validity. Questionnaires should measure at least duration and frequency of PA, and if the intention was to measure total PA, the questionnaire should cover activities in all settings (work, home, transport, recreation, sport).

1.7 Construct Validity

The more similar the constructs that are being compared, the more evidence is provided for validity. Comparison with objective measures of PA (doubly labelled water, accelerometers, pedometers) was considered the best level of evidence (Level 1 or 2, depending on the use of the objective data). We considered constructs not really measuring current PA (maximal oxygen uptake [$\dot{V}O_{2max}$], body mass index [BMI], etc.) or another questionnaire, a diary or interview as less adequate comparison measures (Level 3). Depending on the strength of the hypothesized association with the comparison measure, different correlations were considered to be adequate (table II).

A positive score was given if the study population consisted of ≥50 participants and the correlation was above the specified cut-off point. If the correlation was below the specified cut-off point, a negative score was given. If the

Table I Explanation of acronyms or abbreviated names of questionnaires

Abbreviation	Full name of questionnaire
ARIC/Baecke[3,4]	Atherosclerosis Risk in Communities (ARIC)/Baecke Questionnaire
CARDIA Q[3,5]	Coronary Artery Risk Development in Young Adults Questionnaire
CHAMPS[6]	Cardiovascular Health after Maternal Placental Syndromes
CMH[7]	California Men's Health Study
COSMIN[8]	COnsensus-based Standards for the selection of health Measurement INstruments
EPAQ2[9]	EPIC-Norfolk Physical Activity Questionnaire
EPIC original Q[10]	European Prospective Investigation into Cancer and Nutrition original Questionnaire
HLAQ[11]	Historical Leisure Activity Questionnaire
HUNT 1 and 2[12,13]	The Nord-Trøndelag Health Study 1 and 2
IPAQ[14]	International Physical Activity Questionnaire
JACC Q[15]	Japan Collaborative Cohort Study for Evaluation of Cancer Risk Questionnaire
LACE PA Q[7]	Life After Cancer Epidemiology Study Physical Activity Questionnaire
Minnesota LTPA Q[16]	Minnesota Leisure Time Physical Activity Questionnaire
MOSPA[17]	Monica Optional Study of Physical Activity
NASA Q[18]	National Aeronautics and Space Administration Questionnaire
NHS II Activity Q[19]	Nurses' Health Study II Activity Questionnaire
NPAQ[20]	Neighbourhood Physical Activity Questionnaire
NZPAQ-SF[21]	New Zealand Physical Activity Questionnaire – Short Form
PAFQ[22]	Physical Activity Frequency Questionnaire
PAQ-AD[23]	Physical Activity Questionnaire – Adults
PAS[24,25]	Physical Activity Survey
PYTPAQ[26,27]	Past Year Total Physical Activity Questionnaire
QAPSE[28]	Questionnaire d'Activité Physique Saint-Etienne
QAPAQ[29]	Quality Assessment of Physical Activity Questionnaire Checklist
RWJ[30]	Historical Walking, Running and Jogging Questionnaire
SDR[31]	7-day recall
SQUASH[32]	short questionnaire to assess health-enhancing physical activity
TOQ[31]	Tecumseh Occupational Questionnaire
YPAS[6]	Yale Physical Activity Survey

Table II. Cut-off points for sufficient correlations per dimension of physical activity (PA) measured by the questionnaire, and level of evidence

Dimension of PA measured	Level 1	Level 2	Level 3
Total energy expenditure	Doubly labelled water ≥0.70	Accelerometer total counts ≥0.50	$\dot{V}O_{2max}$ ≥0.40 Diary, other questionnaire, interview ≥0.70 caloric intake, BMI, % BF ≥0.50
Vigorous activity	Accelerometer vigorous counts ≥0.50	Accelerometer total counts ≥0.40	$\dot{V}O_{2max}$ ≥0.60 Diary, other questionnaire, interview ≥0.70 caloric intake, BMI, % BF ≥0.50
Moderate plus vigorous activity	Accelerometer moderate and vigorous counts ≥0.50	Accelerometer total counts ≥0.40	$\dot{V}O_{2max}$ ≥0.50 Diary, other questionnaire, interview ≥0.70 caloric intake, BMI, % BF ≥0.50
Moderate activity	Accelerometer moderate counts ≥0.50	Accelerometer total counts ≥0.40	Diary, other questionnaire, interview ≥0.70 caloric intake, BMI, % BF ≥0.50
Walking	Pedometer or accelerometer walking counts ≥0.70		Diary, other questionnaire, interview ≥0.70 caloric intake, BMI, % BF ≥0.50
Leisure time PA	Accelerometer total counts in leisure time ≥0.50	Accelerometer total counts ≥0.40	$\dot{V}O_{2max}$ ≥0.40 Diary, other questionnaire, interview ≥0.70 caloric intake, BMI, % BF ≥0.50
Occupational PA	Direct observational method ≥0.60	Accelerometer during working hours ≥0.40	$\dot{V}O_{2max}$ ≥0.40 Diary, other questionnaire, interview ≥0.70 caloric intake, BMI, % BF ≥0.50

BF = body fat; **BMI** = body mass index; $\dot{V}O_{2max}$ = maximal oxygen uptake.

sample size was <50 participants, the score was indeterminate (?).

1.8 Reliability

The time interval between the test and retest must have been described and short enough to ensure that subjects had not changed their PA levels, but long enough to prevent recall. The most optimal time interval depends on the construct to be measured and the recall period of the questionnaire. For measuring PA during the past or usual week or in the past year, a time interval of 1 day to 3 months was considered appropriate. For measuring lifetime PA, a time interval from 1 day to 12 months was considered appropriate.

For reliability, three levels of evidence were formulated:

- Level 1: an adequate time interval between test and retest and an intraclass correlation coefficient (ICC), Kappa or Concordance.
- Level 2: an inadequate time interval between test and retest and an ICC, Kappa or Concordance; or an adequate time interval between test and retest and a Pearson/Spearman correlation.
- Level 3: an inadequate time interval between test and retest and Pearson/Spearman correlation.

An ICC >0.70 was considered acceptable.[34] The use of Pearson or Spearman correlation coefficients was considered inadequate, because it neglects systematic errors.[35] However, Pearson/Spearman correlations >0.80 would probably result in ICCs >0.70 and were therefore also rated positively, but on a second level of evidence. Pearson or Spearman correlations <0.80 were rated negatively.

A positive score was given if the study population consisted of ≥50 participants and the ICC, Kappa, Concordance or Pearson/Spearman correlation was above the specified cut-off point. If the correlation was below the specified cut-off point, a negative score was given. If the sample size was <50 participants, the score was rated as indeterminate (?).

1.9 Responsiveness

Responsiveness is the ability of an instrument to detect change over time in the construct to be measured.[36] It should be considered an aspect of

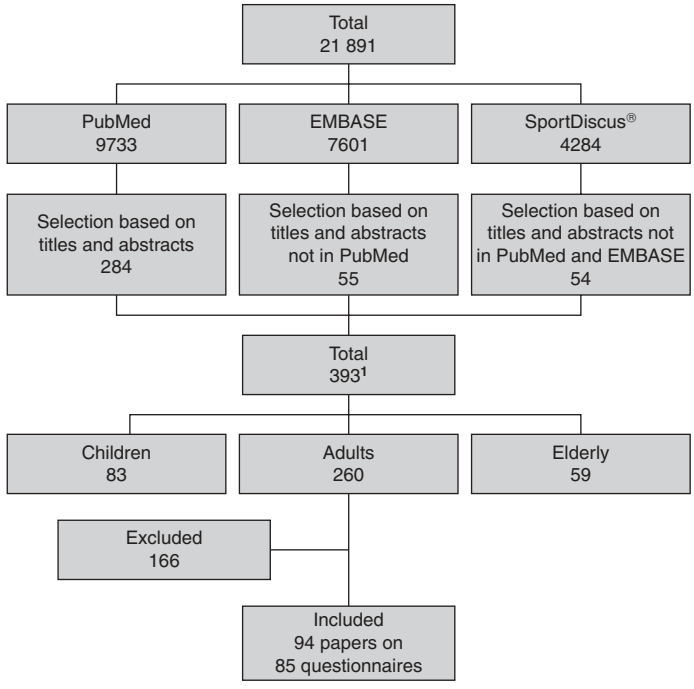

Fig. 1. Flowchart of literature search and paper selection. **1** One paper appears in both the review for adults and for elderly.

validity in a longitudinal setting. Responsiveness was assessed by comparing changes in the PA questionnaire with changes in other instruments that measure closely related constructs. The same approach as for assessing validity was applied, except that change scores were being compared instead of absolute scores. Depending on the strength of the hypothesized association, different correlations were considered to be adequate.

2. Results

The search resulted in 21 891 hits, of which 260 abstracts were selected. Of the full-text articles with relevant titles and/or abstracts, 166 were excluded. Most of the papers were excluded because the questionnaire was administered in an interview or because no measurement properties of the questionnaire were assessed. Finally, 94 papers on 85 (versions of) questionnaires were included in the review (figure 1). Descriptive information on the questionnaires included in the review is provided in table III.

2.1 Quality of the Studies

Construct validity was assessed for 77 questionnaires in 85 studies. Of these 77 questionnaires, 16 were validated at Level 1 and an additional 22 questionnaires at Level 2. Objective comparison measures were often $\dot{V}O_{2max}$ (n = 40), accelerometers (n = 41), heart rate monitor (n = 5), doubly labelled water (n = 7) or pedometer (n = 6) [table IV]. Two of the three questionnaires specifically designed to measure walking were validated against pedometers (Level 1). Surprisingly, appropriate cut-off points for analysing accelerometer data were often not used when assessing time spent in moderate to vigorous PA, but instead total counts were used, which does not discriminate between light, moderate or vigorous PA.

Reliability was assessed for 51 (versions of) questionnaires in 49 studies. Only 15 questionnaires were reliability-tested at Level 1 and an additional 36 questionnaires at Level 2 (table V). The most frequently occurring methodological

Table III. Description of physical activity (PA) questionnaires (Q)

Questionnaire	Construct				Format		
	dimension	setting	recall period	no. of questions/activities	parameters	scores	unit of measurement
Modified Active Australia Survey[37]	PA	Leisure, walking	Past wk	24	F, D	TEE	MET • min/wk
Activity History Q[38]	Physical training	?	Past y	?	F, D	TEE Vigorous EE	kcal/kg/wk kcal/kg/wk
Aires[39]	LTPA	LTPA	Past 12 mo	1	I	Total leisure	Activity score (1–4)
Arizona Activity Frequency Q[40]	TEE	Sport, recr, occup, home, sleeping, personal care	Past 28 d	68	F, D	TEE; daily PA EE	kJ/day
Baecke[18,41,42]	Habitual PA	Sport, recr, occup, sleeping	Not defined	16	F, D	Work; sport; leisure	Activity score (1–5)
Modified Baecke 1[43]	?	?	Past y	19	F, D	Work; sport; leisure; total	Activity score (1–5, total 3–15)
Modified Baecke (ARIC/Baecke)[3,4]	LTPA	Sport, recr, trans, occup, watching TV	?	15	F, D	Sport- and exercise-related leisure index; non-sport- and exercise-related leisure index; total leisure activity	Activity score (1–5)
Modified Baecke 2[44]	?	Sport, recr, trans, watching TV, sweating	?	5	F, D	Sport activity index Leisure activity index	Activity score (?) Activity score (1–5)
Extended Baecke (QAPSE)[28]	DEE	Sport, recr, trans, occup, home, sleeping, eating, washing	Usual wk	35	F, D	TEE	MET/day
Bharathi Q[45]	?	Sport, recr, trans, occup, home, sleeping, sedentary activities	Past mo	13	F, D	TEE PAL	kJ/day 24h EE/BMR
Black Women's Health Study[46]	?	Sport, recr, home, walking	Previous y	?	D	Weekly PA EE	MET • h/wk
CARDIA[3,5]	?	?	Past y	3	F	Total; moderate; heavy	Weighted F
Modified CHAMPS[6]	?	?	Past 2 wk	31	F, D	TEE; moderate/vigorous; vigorous; sports;	kcal/kg/wk
CMH Q[7]	?	Sport, recr, occup, sedentary act	Past 3 mo	24	F, D, I	Total; moderate; vigorous	MET • h/wk

Continued next page

Table III. Contd

Questionnaire	Construct				Format		
	dimension	setting	recall period	no. of questions/activities	parameters	scores	unit of measurement
EPIC original Q[10]	Daily EE	Sport, recr, trans, occup, home, rest	Past y	28	F, D, I	Total; occup; leisure; rest	kJ/24 h
Modified EPIC Q (short PA Index)[47]	?	Sport, recr, trans, occup, home	Past y	4	D	PA index	Activity score (4 categories)
EPAQ2[9]	TEE	Sport, recr, trans, occup, home, sleeping	Past y	85	F, D	TV time Activity at home Activity at work Recreational activity Vigorous activity PA index	h/wk MET • h/wk MET • h/wk MET • h/wk h/wk MET • h/wk
Flemish PA computerized Q[48]	PA and sedentary behaviour	Sport, recr, trans, occup, home, sedentary behaviour, sleeping	Usual wk	57–90	D	15 different activity scores	kcal/wk h/wk
Framingham Q[42]	?	?	Usually	?	?	TEE	kcal/day
Single PA Q Gionet and Godin[49]	LTPA	LTPA	Past 6 mo	1	F	Total	Activity score (1–6)
SDR Q Gionet and Godin[49] (based on Godin Q[50])	LTPA	Sport, recr	Past 7 d	29	F	Total; strenuous; moderate; mild	MET/wk
Godin Q[3,18,50]	LTPA	Sport, recr	Usual wk	4	F	Total; strenuous; moderate; light	Times/wk
Harvard/College Alumnus Q[42,51-53]	LTPA	Sport, recr, trans, stair climbing	Past 7 d	3	F, D	Leisure EE; light; moderate; vigorous; TEE	MET • min/wk, kcal/wk
Harvard/College Alumnus Q[3]	?	Sport, recr, trans, stair climbing	Currently	3	F, D, I	TEE; sports	MET • min/day
HUNT 1[12]	LTPA	sport	Usually	3	F, D, I		
HUNT 2[13]	LTPA in past y	Sport, recr, occup	Past y	3	D I	Light PA Hard PA Work PA	Activity score (0–3) Activity score (0–3) Activity score (1–4)
IPAQ[14]	?	Sport, recr, trans, occup, home, sitting	Past 7 d/usual wk	9 (S7S, SUS) 31 (L7S, LUS)	F, D, I	TEE Meeting ACSM norm	MET • min/wk Yes/no

Continued next page

Table III. Contd

Questionnaire	Construct				Format		
	dimension	setting	recall period	no. of questions/activities	parameters	scores	unit of measurement
Adapted IPAQ[54,55]	?	Sport, recr, trans, occup, home, sitting	Usual wk, summer and winter	59	F, D, I	TEE	MET • min/wk
JACC Q[15]	?	Sport, recr, trans	Usually usually past 2 y	3	F, D	PA time; walking time; PA F	Activity score (1–4)
Kaiser PA Survey[56]	?	Sport, recr, trans, occup, home, watching TV	Past y	75	F, D	Caregiving; housework; housework/caregiving; sports/exercise; active living habits; occup; 3-point summary;	Activity score (1–4 caregiving, 1–5 other)
Kuopio Q[57]	Habitual activity	Sport, recr, trans, occup	Currently	39	F, I	Total	F of conditioning exercise
LACE PA Q[7]	?	Sport, recr, trans, occup, home	Past 12 mo	56	F, D, I	Domain and intensity specific summaries	MET • h/wk
Life in NZ National Survey[58]	?	Recr, occup, home	Past 4 wk	100	F, D, I	$Activity_{hi}$ $Activity_{lo}$	min/wk min/wk MET/wk
Lipid Research Clinics Q[3,42,59]	PA level relative to peers; regular engagement in strenuous activities	Unspecified	Currently	4	Comparative rating	Active/inactive/ highly active/ moderately active/ low active/very low active	Activity score (1–2) Activity score (1–4)
Löf Q[60]	?	?	Past 2 wk	6	?	TEE	kcal/24 h
Leisure Time PA Q[61]	EE during LTPA	Sport, recr, trans, home	Past 3 mo Past y	47	D, I	LTPA	kcal/wk/kg
Mail Survey of PA habits[62]	Exercise habits and participation	Sport, recr, trans	Past 3 mo	6	F, D	TEE RWJ Index Sweat F	MET/wk Activity score Times/wk
Minnesota LTPA Q[16]	?	?	Past y	63	F, D	Leisure EE	MET/h
Minnesota LTPA Q[3,42,63]	?	?	Past y	74	D, I	Leisure EE	MET • min/wk

Continued next page

Table III. Contd

Questionnaire	Construct				Format		
	dimension	setting	recall period	no. of questions/activities	parameters	scores	unit of measurement
Modified Minnesota LTPA Q (Canada Fitness Survey)[64]	?	Sport, recr, home	Weekly, past mo, past y	?	F, D, I	Total; leisure; non-leisure	Time, kJ/kg/wk
Modified Minnesota LTPA Q (Year 11 Q)[65]	?	Sport, recr, home	Past y, recalled 1–10 y later	27	F, D	Leisure EE; light EE; moderate EE; vigorous EE	kcal/wk
Modified Minnesota LTPA Q + TOQ Q + new household activity measure[66]	EE	Sport, recr, occup, home	Past y	98	F, D	Occup EE; leisure EE; household EE	MET/h
Modified Minnesota LTPA Q + TOQ Q + general Q + sleeping[67]	EE	Sport, recr, trans, occup, home, sleeping, watching TV, reading, parenting	Past 4 wk	107	F, D,	TEE	MJ/day
MOSPA[17]	?	Sport, recr, trans, occup, home	?	?	?	TEE Work Transport Household LTPA	kcal/day min/wk min/wk min/wk min/wk
Mundal Q[68]	Habitual LTPA	Sport, recr, home	?	1	PA level	LTPA	Activity score (6 categories)
NASA Q[18]	?	?	?	?	?	?	?
NHS II Activity Q[19]	?	Sport, recr, trans, home, sedentary activities	Past y	14	F, D, I	Activity score Inactivity score	MET • h/wk
Modified NHS II Activity Q[69]	?	Sport, recr, trans, home, sedentary activities, stairs climbed	Past y	15	D	Vigorous activity; non-vigorous activity; sum of activities; inactivity at home; inactivity at work; overall inactivity	MET/wk
Norman Q[70]	Total PA	Sport, occup, home, walking/cycling, watching TV/reading, sleeping	Past y	6	D, I	Crude total PA; total PA	MET

Continued next page

Table III. Contd

Questionnaire	Construct				Format		
	dimension	setting	recall period	no. of questions/activities	parameters	scores	unit of measurement
NZPAQ-SF[21]	?	Sport, recr, trans, occup, home	Past 7 d	7 and 1 optional	F, D, I	EE	MET • min/wk
One-week recall Q[71]	Current PA guidelines	Sport, recr, trans, occup, home	Past wk	6	F, D	Walking Moderate Vigorous Total duration Meeting fitnorm[71]	times/wk times/wk times/wk min/wk Yes/no
PAFQ[22]	Total and act-specific EE	Sport, recr, trans, occup, home, sleeping	Past 7 d	71	F, D	TEE	kcal/day
PA History Q[72]	Usual activity	Sport, recr, trans, occup, home	Past y	13	F, D	Moderate intensity Heavy intensity Total	Activity score
PAS[24,25]	Total PA in 24 h	Sport, recr, trans, occup, home, sleeping, sitting	Average 24 h wk day	9	D	TEE	24 h MET • time
PAQ-AD[23]	Moderate to vigorous PA	Sport, vigorous act	Past 7 d	7	F	Total PA	PA score (1–5)
PYTPAQ[26,27]	Total PA	Sport, recr, trans, occup, home	Past y	Open table format	F, D, I	Total PA; occup PA; household PA; recr PA	h/wk MET • h/wk
Pennsylvania Alumni[42]	?	Recr, occup, ?	Past 7 d Usually Past y	?	?	TEE	kcal/day
Scottish PA Q[73,74]	PA of at least moderate intensity	?	Past 7 d	?	?	Total; leisure; occup	Min/wk
Modified Scottish PA Q for students[75]	PA of at least moderate intensity	?	Past 7 d	?	?	Total; leisure; occup	Min/wk
Saltin and Grimby Q[76]	Lifetime PA	?	?	?	?	Lifetime occup PA; lifetime LTPA	Activity score (1–4)
Singh Q[77,78]	?	Sport, recr, trans, occup, home, sleeping	Past 3 mo	26	F, D	PA index; RWJ index; total activity index; vigorous activity; sport/recr index	MET • min/wk
Single Q[79]	PA for maintaining or improving physical fitness	PA to improve fitness	Currently	1	Yes/no	Meeting fitnorm[71]	Yes/no

Continued next page

Table III. Contd

Questionnaire	Construct				Format		
	dimension	setting	recall period	no. of questions/activities	parameters	scores	unit of measurement
Stanford SDR[3,38]	?	Moderate, vigorous activities	Past 7 d	2	D	Moderate; vigorous	h/wk
Modified Stanford SDR Auckland Heart Study PA Q[80]	?	Moderate, vigorous, activities, resting, sleeping	Past 3 mo	?	F, D	Moderate; vigorous; TEE	kcal/day
Modified Stanford SDR[72]	?	Moderate, hard, very hard activities, sleeping	Past 7 d	?	D	TEE; occup EE; leisure EE	kcal/kg/day
Stanford Usual Act Q[3]	?	Moderate act, vigorous act	Past 3 mo	11		Moderate Vigorous	Activity score (1–6) Activity score (1–5)
Suzuki Q[81]	Energy expenditure	Sport, recr, trans, occup, home, sleeping, sitting	Past 7 d	9	F	TEE TEE	kcal/day kcal/wk
SQUASH[32]	Habitual PA	Sport, recr, trans, occup, home	Usual wk	11	F, D, I	Total; commuting; activities at work; household; leisure time	min/wk Activity score
Total PA[82]	TEE	Sport, recr, trans, occup, home, sitting, sleeping	Usual day	9	D, I	TEE	kcal/day
Usual PA measure[83]	Usual PA		Usually	1	PA level	Total	Activity score (1–5)
YPAS[6]	Current PA	Vigorous, trans, standing, sitting	Typical wk last mo	5	F, D	Total	Activity score (0–98)
Walking activities							
Historical RWJ[30]	Historical RWJ	Sport, recr, trans	Past 10 y	3	F, D, I	TEE Sufficiently active or not	MET • h/wk Activity score (1–2)
NPAQ[20]	Walking, overall index of PA	Recr, trans	Usual wk	11?	F, D, destination	Overall PA index; walking	MET • min/wk F of walking inside and outside neighbourhood; duration of walking inside and outside neighbourhood

Continued next page

Table III. Contd

Questionnaire	Construct				Format		
	dimension	setting	recall period	no. of questions/activities	parameters	scores	unit of measurement
Walking Q[84]	Walking	Walking time	Usually	1	D	Walking	Activity score (1–3)
Walking Q (one question from College Alumni Q)[85]	Walking	Walking distance	?	1	F	Walking	km
Historical/lifetime PA							
Modified HLAQ[11]	Lifetime PA	Sport, recr, occup, home, childcare	Past y, 14–21 y, 22–34 y, 35–50 y, 51–65 y	32	F, D	TEE	MET • h/wk
Occupational PA							
Modified Baecke ARIC/Baecke Work Index[31]	?	Occup	?	?	F	Work index	Activity score (1–5)
CARDIA occup Q[31]	?	Occup	Past y	1	F	Total occup	Activity score
Health Insurance Plan of NY Q[3,42]	?	Trans, occup	Usually	6	F, D	Total occup	Activity score (1–28)
Lipid Research Clinics occup Q[31]	?	Occup	?	1	Comparative rating	Total occup	Activity score (1–5)
Minnesota Heart Health Program Q[3]	?	Occup	Currently	6	?, D, I	Work index Leisure index	MET • min/day
Minnesota Heart Health Program occup Q[31]	?	Occup	Usually	2	% vigorous act	Total occup	Activity score (1–4)
Modified Stanford SDR (SDR)[31]	?	Occup	Past 7 days	5	D	Total occup score	Activity score/wk h/wk MET • min/wk
TOQ[31]	Occup-related PA	Trans, occup	Past y	29	F, D	Total occup score	Activity score/wk, h/wk, MET•min/wk
Bone loading PA							
Bone Loading History Q[86]	Bone loading PA	Sport, recr, occup	Life time (4–45 y)	36	F, D	Total hip loading score Total spine loading score	Hip and spine bone loading score Hip and spine bone loading exposure

Continued next page

Table III. Contd

Questionnaire	Construct			Format			
	dimension	setting	recall period	no. of questions/activities	parameters	scores	unit of measurement
Historical Activity Q[87]	Lifetime PA related to bone	Sport, recr, occup, home	5–11 y, 12–13 y, 14–17 y, >18 y	89–140 activities listed for each time period	F, D, I	Total Occup Athletics Leisure Exercise Lifting/carrying Impact level	(MET) h/day (MET) h/day (MET) h/day (MET) h/day (MET) h/day (MET) h/day Score (1–3)
London PA Q[88]	PA related to bone health	Sport, recr, trans, occup, home, standing, sitting	Currently	5	F, D	Total	h/wk

ACSM = meeting PA guidelines of the American College of Sports Medicine; **$Activity_{hi}$** = activity of high intensity; **$Activity_{lo}$** = activity of low intensity; **BMR** = basal metabolic rate; **D** = duration; **EE** = energy expenditure; **F** = frequency; **home** = home activities (household and gardening); **I** = intensity; **L7S** = long form, last 7 d; **LTPA** = leisure time physical activity; **LUS** = long form, usual wk; **MET** = metabolic equivalent; **occup** = occupational; **PAL** = PA level; **Recr** = recreational; **RWJ** = run-walk-jog; **S7S** = short form, last 7 d; **SUS** = short form, usual week; **TEE** = total energy expenditure; **trans** = transport; ? indicates not specified or unclear.

shortcoming was that Pearson correlations instead of ICCs or Kappas were calculated. Another frequently occurring methodological shortcoming was an inadequate time interval between the test and retest.

Responsiveness was assessed for only two (versions of) questionnaires, and the quality of these studies was rated as Level 3.

2.2 Qualitative Attributes of the Questionnaires

In the study by Altschuler et al.,[7] it was tested whether respondents interpreted the LACE PA questionnaire and the CMH questionnaire as intended. In cognitive interviews, respondents described their thought processes while completing these two questionnaires. It was demonstrated that the term ‘intensity’ was frequently interpreted as emotional or psychological intensity rather than physical effort. In addition, it was found that respondents often counted the same activity more than once, overestimated occupational PA and mistook a list of examples for a definitive list.

We did not find studies in which the *content validity* of a PA questionnaire was assessed. However, we formed our own opinion on the content of the questionnaires.

Of the 85 (versions of) questionnaires included in this review, 23 had sufficient content validity: i.e. they covered all relevant settings of PA (e.g. for total PA all five settings; and for occupational PA only transport and work) and measured duration and frequency (Bharati,[45] EPIC original Questionnaire (Q),[10] EPAQ2,[9] Harvard/College Alumnus Q,[3,51] the long version of the IPAQ,[14] the adapted IPAQ,[54] Kaiser PA Survey,[56] LACE PA Q,[7] Minnesota LTPA Q,[61] Mail Survey of PA,[62] Norman Q,[70] NZPAQ-SF,[21] One-week recall Q,[71] PAFQ,[22] PA History Q,[72] PYTPAQ,[26] Singh Q,[77,78] SQUASH,[32] Historical RWJ questionnaire,[30] NPAQ,[20] Health Insurance Plan of NY,[3] TOQ[31,89] and London PA Q[88]).

2.3 Validation Results

Only the 48 studies that assessed construct validity at Level 1 or 2 are discussed below.

Table IV. Construct validity of physical activity (PA) questionnaires (Q)

Questionnaire	Study population (n; mean age; nationality)	Comparison measure	Results	Levels of evidence
Modified Active Australia Survey	44♀; 55 y; AUS[37]	Accelerometer	Walk mod r = 0.39 Vig r = 0.54 Total activity (≥3 MET) r = 0.52	1 ? 1 ? 1 ?
	159♀; 55 y; AUS	Pedometer	Walk mod r = 0.40 Vig r = 0.55 Total r = 0.48	3 – 3 + 3 –
Activity History Q	24♂; 18–31 y; US[38]	$\dot{V}O_{2max}$	TEE r = 0.76 Heavy PA r = 0.64	3 ? 3 ?
Aires	160 105♂ 172 032♀; 40–42 y; NOR[39]	BMI Cholesterol	Lower BMI with higher levels of total leisure Lower cholesterol with higher levels of total leisure	3 ? 3 ?
Arizona Activity Frequency Q	35♀; 44 y; US[40]	Doubly labelled water	TEE r = 0.58	1 ?
Baecke	64–73; 37 y; US[3]	Accelerometer $\dot{V}O_{2max}$ 4 wk history % BF	r = 0.19 r = 0.54 r = 0.37 r = –0.49	2 – 3 – 3 – 3 –
	7♂; 30 y: 26♀; 28 y; US[18]	Accelerometer	Total r = 0.40	2 ?
	21♂; 36 y; US[42]	Resting EE Caloric intake	Total r = 0.21 Total r = 0.38	3 ? 3 ?
	139♂ 167♀; 20–32 y; NL[41]	Lean body mass	Work b = 1.36♂, b = 0.48♀ Sport b = 1.23♂, b = 0.23♀ Leisure b = 0.15♂, b = –0.27♀	3 – 3 – 3 –
Modified Baecke (ARIC/Baecke)	28♂ 49♀; 37 y; US[4]	Accelerometer $\dot{V}O_{2max}$ % BF 48 h activity diary	Total leisure activity r = 0.24♂, r = 0.19♀ Total leisure activity r = 0.57♂, r = 0.46♀ Total leisure activity r = –0.30♂, r = –0.51♀ Total leisure activity r = 0.59♂, r = 0.33♀	2 ? 3 ? 3 ? 3 ?
Modified Baecke 1	60♂; 20–60 y: 54♀; 20–70 y; NL[43]	3 d activity diary	Total r = 0.66♂, r = 0.42♀	3 –
Modified Baecke 2	195♂; 41 y; FR[44]	$\dot{V}O_{2max}$ % BF Quetelet index	SAI r = 0.31, LAI r = 0.09 SAI r = –0.20, LAI r = –0.14 SAI r = 0.03, LAI r = –0.21	3 – 3 – 3 –
Extended Baecke (QAPSE)	20♂; 56–72 y; FR[28]	Caloric intake	DEE r = 0.58	3 ?
Bharathi Q	14♂; 34♀; 18–60 y; IN[45]	Energy intake Age (young [n = 57] vs elderly [n = 49])	TEE r = 0.33 PAL young 1.52, elderly 1.22, $p < 0.01$	3 ? 3 ?

Continued next page

Table IV. Contd

Questionnaire	Study population (n; mean age; nationality)	Comparison measure	Results	Levels of evidence
Black Women's Health Study	101♀; 48 y; US[46]	Accelerometer	Total r = 0.28, vig r = 0.40, mod r = –0.04	2 –
		7 d PA diary	Total r = –0.32, vig r = 0.41, mod r = 0.26	3 –
CARDIA	64–73; 37 y; US[3]	Accelerometer	Mod r = 0.11, vig r = 0.31	2 – 2 –
		$\dot{V}O_{2max}$	Mod r = 0.08, vig r = 0.63	3 – 3 +
		4 wk history	Mod r = 0.08, vig r = 0.54	3 – 3 –
		% BF	Mod r = –0.09, vig r = –0.35	3 – 3 –
Modified CHAMPS	109♀ 29♂; 41 y; US[6]	$\dot{V}O_{2max}$	TEE r = 0.42, mod/vig, r = 0.43 vig ?	3 ?
			sports r = 0.50 ♂	3 ?
			TEE r = 0.07, mod/vig r = 0.05, vig r = –0.01	3 –
			sports r = 0.19 ♀	3 –
		BMI	TEE r = 0.15, mod/vig r = 0.19, vig r = ?,	3 –
			sports r = 0.07 ♂	3 –
			TEE r = –0.01, mod/vig r = 0.02, vig r = –0.03,	3 –
			sports r = –0.01 ♀	3 –
EPIC original Q	59♂; 41 y: 52♀; 49 y; NL[10]	Activity diary	Total r = 0.66♂, r = 0.43♀	3 –
Modified EPIC Q (short PA Index)	84♂; 59 y: 89♀; 55 y; UK[47]	HR-EE	PA index p = 0.003	3 –
		$\dot{V}O_{2max}$	PA index p = 0.01	3 –
		7-d food diaries	PA index p < 0.05	3 –
EPAQ2	84♂; 59 y: 89♀; 55 y; UK[9]	HR-EE	PA index r = 0.28, work r = 0.17	3 – 3 –
		$\dot{V}O_{2max}$	PA index r = 0.15, work r = 0.01	3 – 3 –
Flemish PA computerized Questionnaire	31♂; 39 y: 35♀; 42 y; BE[48]	Accelerometer + 7 d activity record	PAL r = 0.56♂, r = 0.44♀	2 ?
Framingham Q	21♂; 36 y; US[42]	Resting EE	TEE r = 0.24	3 ?
		Caloric intake	TEE r = 0.43	3 ?
Godin Q	163♂; 31 y: 143♀; 30 y; CA[50]	% BF	Total r = 0.13; strenuous r = 0.21	3 –
		$\dot{V}O_{2max}$	Total r = 0.24; strenuous r = 0.38	3 – 3 –
	7♂; 30 y: 26♀; 28 y; US[18]	Accelerometer	Total r = 0.45	2 ?
	64–73; 37 y; US[3]	Accelerometer	Total r = 0.32	2 –
		$\dot{V}O_{2max}$	Total r = 0.56	3 +
		4 wk history	Total r = 0.36	3 –
		% BF	Total r = –0.43	3 –
Single PA Q Gionet and Godin	456♂; 36 y: 95♀; 33 y; CA[49]	$\dot{V}O_{2max}$	Total (pattern) r = 0.22♂, r = 0.40♀	3 –
		BMI	Total (pattern) r = –0.10♂, r = –0.05♀	3 –
		Musculoskeletal endurance	Total (pattern) r = 0.25♂, r = 0.32♀	3 –
SDR Q Gionet and Godin	456♂; 36 y: 95♀; 33 y; CA[49]	$\dot{V}O_{2max}$	TEE r = 0.11♂, r = 0.05♀	3 –
			strenuous r = 0.25♂, r = 0.28♀	3 –
		BMI	TEE r = 0.01♂, r = 0.02♀	3 –
			strenuous r = –0.04♂, r = –0.04♀	3 –

Continued next page

Table IV. Contd

Questionnaire	Study population (n; mean age; nationality)	Comparison measure	Results	Levels of evidence
		Musculoskeletal endurance	TEE r = 0.18♂, r = 0.10♀	3 –
			strenuous r = 0.36♂, r = 0.36♀	3 –
Harvard/College Alumnus Q	21♂; 36 y; US[42]	Resting EE	TEE r = 0.32	3 ?
		Caloric intake	TEE r = 0.49	3 ?
	64–73; 37 y; US[3]	Accelerometer	TEE r = 0.30	2 –
		$\dot{V}O_{2max}$	TEE r = 0.52	3 +
		4 wk history	TEE r = 0.31	3 –
		% BF	TEE r = –0.30	3 –
	12♂; 31 y: 13♀; 30 y; US[52]	HR monitor combined with two accelerometers	Total r = 0.35	2 ?
			Light r = 0.20	2 ?
			Mod r = 0.27	2 ?
			Vig r = 0.47	1 ?
	28♂ 50♀; 38 y; US[51]	Accelerometer	Leisure EE r = 0.19♂, r = 0.19♀	2 ? 2 –
		$\dot{V}O_{2max}$	Leisure EE r = 0.58♂, r = 0.53♀	3 ? 3 +
		3 × 48 h activity diaries	Leisure EE r = 0.60–0.65♂, r = 0.34–0.54♀	3 ? 3 –
		% BF	Leisure EE r = –0.36♂, r = –0.36♀	3 ? 3 –
	138; 41 y; US[6]	$\dot{V}O_{2max}$	City blocks r = –0.06, stairs r = 0.11	3 –
			walking min/day r = 0.32♂, r = 0.02♀	3 –
		BMI	City blocks r = 0.14, stairs r = –0.02	3 –
			walking min/day r = –0.21♂, r = 0.12♀	3 –
	36♂; 41 y: 32♀; 42 y; US[53]	$\dot{V}O_{2max}$	Leisure EE r = 0.29	3 –
		Sweat Q	Leisure EE r = 0.57	3 –
	21♂; 36 y; US[42]	Resting EE	Total r = 0.05	3 ?
		Caloric intake	Total r = 0.19	3 ?
	105♂; 40 y: 87♀; 38 y; US[90]	BMI	PA index: no significant regression coefficient;	3 ?
			total wkly activity: significant regression coefficient	3 ?
HUNT 1	S7S: 108♂; 32 y; NOR[12]	Accelerometer (PAL)	Frequency r = 0.03	2 –
			Intensity r = 0.06	2 –
			Duration r = 0.12	2 –
			Index r = 0.07	2 –
		$\dot{V}O_{2max}$	Frequency r = 0.43	3 +
			Intensity r = 0.40	3 +
			Duration r = 0.31	3 –
			Index r = 0.48	3 +
HUNT 2	108♂; 32 y; NOR[13]	Accelerometer	Light r = –0.10, hard r = 0.31, work r = 0.39	1 –
		$\dot{V}O_{2max}$	Light r = –0.03, hard r = 0.46, work r = –0.06	3 –
		IPAQ	Light r = 0.19, hard r = 0.48, work r = 0.34	3 –

Continued next page

Table IV. Contd

Questionnaire	Study population (n; mean age; nationality)	Comparison measure	Results	Levels of evidence
IPAQ	S7S: 26–151; 18–65 y[14]	Accelerometer	S7S total r = 0.02–0.47, ACSM r = 0.46–0.93	2 – 1 +
	SUS: 26–127; 18–65 y[14]		SUS total r = –0.12–0.32, ACSM r = 0.50–0.75	2 – 1 +
	L7S: 26–151; 18–65 y[14]		L7S total r = 0.05–0.52, ASCM r = 0.31–1.0	2 – 1 +
	LUS: 26–127; 18–65 y[14]		LUS total r = –0.02–0.36, ASCM r = 0.35–0.72	2 – 1 +
	S7S: 847–928; 29 y; FIN[91]	$\dot{V}O_{2max}$	Tot METs: some dose-response relation with $\dot{V}O_{2max}$; Vig: dose-response relation with $\dot{V}O_{2max}$	3 ?
	L7S: 22♂ 24♀; 41 y; SW[92]	Accelerometer	Total r = 0.55, vig r = 0.63, mod r = 0.12	1 +
		$\dot{V}O_{2max}$	Total r = 0.21, vig r = 0.14, mod r = 0.21	3 –
		BMI	Total r = 0.25, vig r = 0.27, mod r = 0.17	3 –
		PA log	Leisure r = 0.58, trans r = 0.18, work r = 0.64, home r = 0.47	3 –
	S7S: 32♂ 91♀; 21 y; US[93]	Accelerometer	Total r = 0.23, vig r = 0.47, mod r = 0.23	1 –
		Pedometer	Tot r = 0.25, vig r = 0.38, mod r = 0.17, walking r = 0.12	1 –
	S7S: 74♂ 76♀; 31 y; JAP[94]	Doubly labelled water	Significant diff between insufficiently and highly active categories, but non-significant diff between insufficiently active and sufficiently active group or sufficiently active and highly active group	1 –
	S7S: 108♂; 32 y; NOR[12]	Accelerometer	Vig r = 0.07	1 –
			Mod r = 0.17	1 –
			Total r = 0.26	2 –
		$\dot{V}O_{2max}$	Vig r = 0.41	3 –
	S7L: 16♂ 20♀; 39 y; NZ[21]	Doubly labelled water	AEE r = 0.31	1 ?
			METmin r = 0.33	
	S7S: 51♂ 91♀; 44 y; US[95]	Accelerometer	EE 1 min bout: r = 0.58♂, r = 0.21♀	2 –
			EE 10 min bout r = 0.48♂, r = 0.07♀	2 –
			Meeting guidelines 1 min bout κ = 0.21	2 –
			Meeting guidelines 10 min bout κ = 0.04	2 –
	S7S: 30♂; 26 y: 19♀; 34 y; CH[96]	Accelerometer	Total r = 0.09	2 –
		PA log	Total r = 0.29	3 –
computerized IPAQ	23♂ 30♀; 31 y; BE[97]	Accelerometer	r = 0.38 (total PA) r = 0.37 (MVPA)	2 –
		PA diary	r = 0.39 (min)	3 –
Adapted IPAQ	LUS: 1068♂ 1372♀; 47 y; NOR[54]	BMI	r = –0.02♂, r = –0.04♀	3 –
IPAQ sitting Q	L7S and S7S: 65♂ 79♀; 35 y; UK[98]	Accelerometer	L7S r = 0.22♂, r = 0.35♀	2 –
	L7S and S7S: 30; 33 y; NL[98]		S7S r = 0.24♂, r = 0.29♀	2 –
	L7S and S7S: 26; 49 y; US[98]		L7S r = 0.35, S7S r = 0.22	2 –
	L7S and S7S: 26; 36 y; US[98]		L7S r = 0.26, S7S r = 0.45	2 –
			L7S r = 0.49, S7S r = 0.49	2 –

Continued next page

Table IV. Contd

Questionnaire	Study population (n; mean age; nationality)	Comparison measure	Results	Levels of evidence
JACC Q	739♂ 991♀; 22–80; JAP[15]	Interview	PA time r = 0.53♂, r = 0.58♀	3 –
			PA freq r = 0.53♂, r = 0.59♀	3 –
Kaiser PA Survey	50♀; 39 y; US[56]	Accelerometer,	4-point summary r = 0.49, occup r = 0.16	2 – 3 –
		$\dot{V}O_{2max}$	4-point summary r = 0.59, occup r = 0.04	3 + 3 –
		2 7 d activity diaries,	4-point summary r = 0.35, occup r = 0.35	3 – 3 –
		% BF	4-point summary r = –0.53, occup r = –0.06	3 – 3 –
Kuopio Q	1162♂; 54 y; FIN[57]	$\dot{V}O_{2max}$	Total b = 0.15	3 –
Life in NZ National Survey	140; 37 y; NZ[58]	$\dot{V}O_{2max}$	$Activity_{hi}$ r = 0.40, $activity_{lo}$ r = –0.10, $metab_{2.5}$ r = 0.03	3 – 3 – 3 –
		BMI	$Activity_{hi}$ r = –0.05, $activity_{lo}$ r = 0.14, $metab_{2.5}$ r = 0.16	3 – 3 – 3 –
		Stanford SDR Q	$Activity_{hi}$ r = 0.39, $activity_{lo}$ r = 0.31, $metab_{2.5}$ r = 0.30	3 – 3 – 3 –
Lipid Research Clinics Q	21♂; 36 y; US[42]	Resting EE	Total r = 0.24	3 ?
		Caloric intake	Total r = 0.40	3 ?
	28♂ 50♀; 40 y; US[59]	Accelerometer	2-point score $r^2 = 0.04$, 4-point score $r^2 = 0.04$	2 –
		$\dot{V}O_{2max}$	2-point score $r^2 = 0.29$, 4-point score $r^2 = 0.29$	3 –
		% BF	2-point score $r^2 = 0.10$, 4-point score $r^2 = 0.17$	3 –
		BMI	2-point score $r^2 = 0.15$, 4-point score $r^2 = 0.22$	3 –
Löf Q	24♀; 30 y; SW[60]	Doubly labelled water	TEE r = 0.56	1 ?
			LOA = –800–1200	
London PA Q	26♀; 43–54 y; UK[88]	4 d activity diary	Total r = 0.45	3 ?
		$\dot{V}O_{2max}$	Total NS	3 ?
Leisure Time PA Q	166♂; 43 y; US[61]	$\dot{V}O_{2max}$	LTPA r = 0.43	3 +
		% BF	LTPA r = –0.35	3 –
		HDL cholesterol	LTPA r = 0.17	3 –
		Sys blood pressure	LTPA r = 0.02	3 –
Mail survey of PA habits	375♂; 47 y; US[62]	Treadmill time	TEE r = 0.05	3 –
			RWJ index r = 0.51	3 +
			Sweat freq r = 0.51	3 +
Minnesota LTPA Q	21♂; 36 y; US[42]	Resting EE	Leisure EE r = 0.17	3 ?
		Caloric intake	Leisure EE r = 0.13	3 ?
version 22 items (correspond closely to Canada Fitness Survey)	43♀; 47 y; US[76]	3 d beeper-cued diary	Leisure EE r = 0.14	3 ?
		3 occup groups	No significant diff between 3 occup groups	

Continued next page

Table IV. Contd

Questionnaire	Study population (n; mean age; nationality)	Comparison measure	Results	Levels of evidence
	64–73; 37 y; US[3]	Accelerometer	Leisure EE r = 0.18	2 –
		$\dot{V}O_{2max}$	Leisure EE r = 0.43	3 +
		4 wk history	Leisure EE r = 0.74	3 +
		% BF	Leisure EE r = –0.24	3 –
	2356 ♂; 49 y; US[99]	Caloric intake	Total LTPA r = 0.11	3 –
			Light LTPA r = 0.08	3 –
			Mod LTPA r = 0.04	3 –
			Intense LTPA r = 0.08	3 –
Minnesota Heart Health Program Q	64–73; 37 y; US[3]	Accelerometer	Work index r = 0.04, leisure index r = 0.28	3 – 2 –
		$\dot{V}O_{2max}$	Work index r = 0.0, leisure index r = 0.56	3 – 3 +
		4 wk history	Work index r = –0.09, leisure index r = 0.39	3 – 3 –
		% BF	Work index r = 0.07, leisure index r = –0.37	3 – 3 –
Modified Minnesota LTPA Q + TOQ + new household activity measure	59 ♀; 47 y; US[66]	$\dot{V}O_{2max}$ (score 1–5)	Occupational EE $\dot{V}O_2$ r = 0.03, BMI r = 0.02	3 – 3 –
		BMI	Leisure EE $\dot{V}O_2$ r = 0.21, BMI r = –0.11	3 – 3 –
			Household EE $\dot{V}O_2$ r = 0.14, BMI r = 0.00	3 – 3 –
MOSPA	108 ♂ 59 ♀; 36 y; BE[17]	$\dot{V}O_{2max}$	TEE $\dot{V}O_2$ r = 0.39, BMI r = 0.30, %BF r = –0.26	3 – 3 – 3 –
		BMI	Work $\dot{V}O_2$ r = 0.21, BMI r = 0.08, %BF r = –0.17	3 – 3 – 3 –
		% BF	Transport $\dot{V}O_2$ r = 0.16, BMI r = –0.17, %BF r = –0.13	3 – 3 – 3 –
			Household $\dot{V}O_2$ r = –0.01, BMI r = –0.15, %BF r = –0.01	3 – 3 – 3 –
			LTPA $\dot{V}O_2$ r = 0.30, BMI r = 0.04, %BF r = –0.25	3 – 3 – 3 –
Mundal Q	1769 ♂; 40–59 y; NOR[68]	Interview	LTPA κ = 0.62	3 –
NASA Q	7 ♂ 30 y; 26 ♀ 28 y; US[18]	Accelerometer	Total r = 0.32	2 ?
NHS II Activity Q	147 ♀; 39 y; US[19]	Past wk recall	Activity score r = 0.79	3 +
		7 d activity diary	Activity score r = 0.56; Inactivity score r = 0.41	3 – 3 –
Modified NHS II Activity Q	238 ♂; 40–75 y; US[69]	4 7 d activity diaries	Vig activity r = 0.58	3 –
			Non-vig activity r = 0.28	3 –
			Sum of activities r = 0.65	3 –
			Inactivity at home r = 0.30	3 –
			Inactivity at work r = 0.40	3 –
			Overall inactivity r = 0.41	3 –
Norman Q	111 ♂; 63 y; SW[70]	7 d activity diary	Crude total PA r = 0.23	3 –
			Total PA r = 0.56	3 –
			r = 0.65 age 44–64 y	3 –
			r = 0.50 age 65–78 y	3 –
			r = 0.73 BMI ≤26	3 +
			r = 0.39 BMI >26	3 –
			Occup r = 0.40	3 –

Continued next page

Table IV. Contd

Questionnaire	Study population (n; mean age; nationality)	Comparison measure	Results	Levels of evidence
			Home r = 0.62	3 –
			Leisure r = 0.40	3 –
			TV/reading r = 0.52	3 –
			Sleeping r = 0.61	3 –
NZPAQ	16♂ 20♀; 39 y; NZ[21]	Doubly labelled water	AEE r = 0.38	1 ?
			METmin r = 0.39	
One-week recall Q	55♂; 38 y: 63♀; 40 y; AUS[71]	Accelerometer	≥3 MET r = 0.29♂, r = 0.25♀	1 –
			3.0–5.9 MET r = 0.40♂, r = 0.19♀	1 –
			6.0+ MET r = 0.19♂, r = 0.10♀	1 –
PAQ-AD	61♂ 122♀; 31 y; CAN[23]	Accelerometer CAL	r = 0.43	2 –
		Accelerometer MTI (n = 41)	r = 0.26	2 ?
		Several other PA Q	r = 0.54–0.63	3 –
PAFQ	18♂ 23♀; 35–69 y; SWZ[22]	Heart rate EE	TEE r = 0.76	3 ?
		24 h recall	TEE r = 0.80	3 ?
PA History Q	4956; 18–30 y; US[72]	Treadmill time	Significant regression coefficients	3 ?
		Caloric intake	Significant regression coefficients	3 ?
		BMI	Significant regression coefficients only in ♀	3 ?
PAS	19♂ 20♀; 20–60 y; DK[24]	Accelerometer	TEE r = 0.05♂, r = 0.31♀	1 ?
		4 d activity diary	TEE r = 0.86♂, r = 0.49♀	3 ?
	53♂ 47♀; 35–65 y; DK[25]	$\dot{V}O_{2max}$	TEE non-significant association	3 ?
			Vig PA significant association	
	45♂ 62♀; 21 y; AUS[100]	Pedometer	TEE r = 0.48	3 +
			TEE r = 0.56♂, r = 0.38♀	3 ? 3 –
PYTPAQ	75♂ 79♀; 49 y; CAN[26]	Accelerometer	r = 0.26	2 –
		$\dot{V}O_{2max}$	r = 0.32/0.37	3 –
		BMI	r = –0.07/0.22	3 –
		PA log	r = 0.41	3 –
Pennsylvania Alumni Q	21♂; 36 y; US[42]	Resting EE	TEE r = 0.30	3 ?
		Caloric intake	TEE r = 0.47	3 ?
Saltin and Grimby Q	43♀; 47 y; US[76]	3 d beeper-cued diary	Lifetime occup PA r = 0.45; significant diff	3 ?
		3 occup groups	between 3 occup groups	
			Lifetime LTPA r = 0.55; no significant diff between 3 occup groups	
Scottish PA Q	30; 37 y; SC[73]	Accelerometer	Total r = 0.13	2 ?
			Total without occup walking + outliers r = 0.52	
Modified Scottish PA Q	23; 18–48 y; UK[75]	HR monitor	Total r = 0.0003 (0.34 without 3 outliers)	3 –
	21; 18–48 y; UK[75]	HR monitor	Total r = 0.59	3 +

Continued next page

Table IV. Contd

Questionnaire	Study population (n; mean age; nationality)	Comparison measure	Results	Levels of evidence
Singh Q	115♂ 90♀; 52 y; US[77]	Treadmill time	PA index r = 0.27–0.38♂, r = 0.07–0.15♀	3 –
			RWJ index r = 0.28–0.48♂, r = 0.10–0.34♀	3 –
			Total act index r = 0.24♂, (n = 24) r = 0.03♀ (n = 28)	3 –
	44♂ 94♀; 49 y; US[78]	PAR	Total activity r = 0.51♂, r = 0.65♀	3 – 3 –
			Vig activity r = 0.13♂, r = 0.85♀	3 – 3 +
			Mod activity r = 0.53♂, r = 0.44♀	3 – 3 –
			Inactivity activity r = 0.69♂, r = 0.59♀	3 – 3 –
			Sleep r = 0.39♂, r = 0.52♀	3 – 3 –
		Pedometer	Total activity r = 0.14♂, r = 0.24♀	3 – 3 –
		Treadmill time	Total activity r = 0.23♂, r = –0.09♀	3 – 3 –
Single Q	371♂ 733♀; 37 y; US[79]	$\dot{V}O_{2max}$ (n = 304)	p = 0.0007♂, p = 0.002♀	3 ?
		BMI	p = 0.0001♂, p = 0.001♀	3 ?
		HDL cholesterol	p = 0.0001♂, p = 0.46♀	3 ?
Stanford SDR	375♂; 47 y; US[62]	Treadmill time	Mod r = –0.08	3 –
			Vig r = 0.18	3 –
			TEE r = 0.14	3 –
	7♂; 30 y: 26♀; 28 y; US[18]	Accelerometer	TEE r = 0.79 κ = 0.61	2 ?
	24♂; 18–31 y; US[38]	$\dot{V}O_{2max}$	Vig r = 0.46	3 ?
			TEE r = 0.61	3 ?
	158; 22 y; US[38]	7-day activity diary	TEE r = 0.82	3 +
	74; 22 y; US[38]	12 min run	TEE r = 0.32	3 –
		Skinfolds	TEE r = –0.04	3 –
version unclear	21♂; 36 y; US[42]	Resting EE	TEE r = 0.10	3 ?
		Caloric intake	TEE r = 0.35	3 ?
Stanford SDR	64–73; 37 y; US[3]	Accelerometer	TEE r = 0.33	2 –
		$\dot{V}O_{2max}$	TEE r = 0.30	3 –
		4 wk history	TEE r = 0.36	3 –
		% BF	TEE r = –0.12	3 –
Modified Stanford SDR: Auckland Heart Study PA Q	77♂; 53 y: 75♀; 56 y; NZ[80]	7 d activity diary	Mod r = 0.60 κ = 0.36	3 –
			Vig r = 0.48 κ = 0.23	3 –
			TEE r = 0.91 κ = 0.62	3 –
Modified Stanford SDR	4956; 18–30 y; US[72]	Treadmill time	Significant regression coefficients	3 ?
		Caloric intake	Significant regression coefficients	3 ?
		BMI	Non-significant regression coefficients	3 –
Modified Stanford SDR	46♀; 39 y; US[89]	Accelerometer	NS	2 ?
		$\dot{V}O_{2max}$	NS	3 ?
		2 7 d occup activity diaries	Total occup score h/wk r = 0.78	3 ?
			Total occup score MET·min/wk r = 0.45	3 ?

Continued next page

Table IV. Contd

Questionnaire	Study population (n; mean age; nationality)	Comparison measure	Results	Levels of evidence
	27♂ 48♀; 37 y; US[31]	6×48 h occup activity diaries	Total occup score h/wk r = 0.16 Total occup score MET min/wk r = 0.30	3 – 3 –
Stanford Usual Act Q	64–73; 37 y; US[3]	Accelerometer $\dot{V}O_{2max}$ 4 wk history % BF	Mod r = 0.23, vig r = 0.22 Mod r = 0.27, vig r = 0.38 Mod r = 0.05, vig r = 0.28 Mod r = –0.33, vig r = –0.16	2 – 3 – 3 – 3 –
Suzuki Q	49♂; 27 y: 32♀; 32 y; JAP[81]	Accelerometer	Daily EE r = 0.57♂, r = 0.68♀ Weekly PA r = 0.69♂, r = 0.69♀	2 + 2 +
SQUASH	36♂ 14♀; 44 y; NL[32]	Accelerometer	r = 0.45; κ for comparing tertiles: 0.30	2 –
	24♂ 16♀; 37 y; NL[101]	Accelerometer	TEE r = 0.62♂, r = –0.49♀ κ_w for tertiles: r = 0.29♂ r = –0.15♀	1 ? 1 ?
TOQ + Minnesota LTPA Q	34♀; 37 y; US[16]	Doubly labelled water	TEE r = 0.40	1 ?
Modified Minnesota LTPA Q + TOQ + general Q + sleeping	24♂; 42 y; US[67]	Doubly labelled water	TEE r = 0.39 LOA 1.32 ± 0.73 (EE was more overestimated with higher EE values)	1 ?
Total PA	39 + 94; 41 y; SW[82]	24 h recall	Total PA r = 0.73; concordance = 0.57	3 +
Usual PA measure	188♀; 47 y; US[83]	BMI	Total F = 16.38, p < 0.01	3 ?
YPAS	138; 41 y; US[6]	$\dot{V}O_{2max}$ BMI	Total r = 0.36♂, r = 0.01♀ Total r = 0.16♂, r = 0.08♀	3 – 3 –
Lifetime PA				
Modified HLAQ	131♀; 50 y; US[102]	4 7 d activity diaries	TEE r = 0.29 Mod EE r = 0.16 Vig EE r = 0.63	3 – 3 – 3 –
Walking activities				
Walking Q	51♂ 55♀; 62 y; JAP[84]	Pedometer	Walking p < 0.001–0.006	1 ?
Walking Q (one question from CAQ)	48♂; 41 y: 48♀; 39 y; US[85]	Pedometer	Walking r = 0.35♂, r = 0.48♀	1 –
Historical RWJ Q	4100♂; 48 y: 963♀; 45 y; US[30]	Treadmill time	TEE r = 0.53♂, r = 0.47♀ Significant diff between sufficient/insufficient (effect size 0.68♂, 0.81♀)	3 –
Bone loading PA				
Bone Loading History Q	80♀; 31 y; US[86]	Femoral neck BMI Spine BMI	Total hip loading exposure r = 0.32 Total spine loading exposure r = 0.34 No correlations with spine BMD	1 – 1 –

Continued next page

Table IV. Contd

Questionnaire	Study population (n; mean age; nationality)	Comparison measure	Results	Levels of evidence
Occup PA				
Modified Baecke, ARIC/Baecke Work Index	27♂ 48♀; 37 y; US[31]	6×48 h occup activity diaries	Work index r = 0.04	3 –
CARDIA Occup Q	27♂ 48♀; 37 y; US[31]	6×48 h occup activity diaries	Total occup r = –0.05	3 –
Health Insurance Plan occup Q	27♂ 48♀; 37 y; US[31]	6×48 h occup activity diaries	Total occup r = 0.10	3 –
Health Insurance Plan of NY Q	n = 64–73; 37 y; US[3]	Accelerometer	Total occup r = 0.14	2 –
		$\dot{V}O_{2max}$	Total occup r = 0.07	3 –
		4 wk history	Total occup r = 0.00	3 –
		% BF	Total occup r = –0.03	3 –
Lipid Research Clinics Q	n = 64–73; 37 y; US[3]	Accelerometer	Total occup r = 0.21	2 –
		$\dot{V}O_{2max}$	Total occup r = 0.49	3 +
		4 wk history	Total occup r = 0.24	3 –
		% BF	Total occup r = –0.43	3 –
Lipid Research Clinics occup Q	27♂ 48♀; 37 y; US[31]	6×48 h occup activity diaries	Total occup r = 0.09	3 –
Minnesota Heart Health Program occup Q	27♂ 48♀; 37 y; US[31]	6×48 h occup activity diaries	Total occup r = 0.33	3 –
TOQ	46♀; 39 y; US[89]	Accelerometer	Total occup score MET • min/wk r < 0.25	3 ?
		$\dot{V}O_{2max}$	Total occup score MET • min/wk r < 0.25	3 ?
		2×7 d occup activity diaries	Total occup score h/wk r = 0.18	3 ?
			Total occup score MET • min/wk r = 0.46	3 ?
	27♂ 48♀; 37 y; US[31]	6×48 h occup activity diaries	Total occup score h/wk r = 0.11	3 –
			Total occup score MET • min/wk r = 0.52	3 –
version 17 items	43♀; 47 y; US[76]	3d beeper-cued diary	Total occup score MET/h = 0.29	3 ?
		Three occup groups	Significant diff between three occup groups	

ACSM = meeting PA guidelines of the American College of Sports Medicine; **AUS** = Australia; **b** = regression coefficient; **BE** = Belgium; **BF** = body fat; **BMD** = bone mineral density; **BMI** = body mass index; **CA** = Canada; **CH** = China; **DEE** = dietary energy expenditure; **diff** = differences; **DK** = Denmark; **EE** = energy expenditure; **F** = F-test for mean differences in PA between different levels of the comparison measure; **FIN** = Finland; **FR** = France; **HDL** = high density lipoprotein; **HR** = heart rate; **IN** = India; **IPAQ** = International Physical Activity Questionnaire; **JAP** = Japan; **κ** = Kappa; **κ_w** = weighted Kappa; **L7S** = long form, last 7d; **LAI** = leisure activity index; **LOA** = limits of agreement; **LTPA** = leisure-time physical activity; **LUS** = long form, usual wk; **MET** = metabolic equivalent; **mod** = moderate; **NOR** = Norway; **NL** = Netherlands; **NS** = not significant; **NZ** = New Zealand; **occup** = occupational; **PAL** = physical activity level; **PAR** = physical activity recall(s); **r** = correlation coefficient; **S7S** = short form, last 7d; **SAI** = sport activity index; **SC** = Scotland; **SUS** = short form, usual wk; **SW** = Sweden; **Sweat Q** = number of times/wk vigorous activity sufficient to 'work up a sweat'; **SWZ** = Switzerland; **Sys** = systolic; **TEE** = total energy expenditure; **TV** = television; **UK** = United Kingdom; **US** = United States; **vig** = vigorous; **$\dot{V}O_2$** = oxygen uptake; **$\dot{V}O_{2max}$** = maximal $\dot{V}O_2$; ? indicates indeterminate; ♀ indicates female; ♂ indicates male.

Table V. Reliability of physical activity (PA) questionnaires (Q)

Questionnaire	Study population (n; mean age; nationality)	Interval	Results	Rating
Modified Active Australian Survey	169♀ 55 y; AUS[37]	13 d	Total frequency r = 0.58 Total min/wk r = 0.64	2 – 2 –
Baecke	277; 20–32 y; NL[41]	3 mo	Work r = 0.88 Sport r = 0.81 Leisure r = 0.74	2 + 2 + 2 –
Modified Baecke 1	63♂; 20–60 y: 56♀; 20–70 y; NL[43]	5 mo	Work r = 0.89♂, r = 0.80♀ Sport r = 0.88♂, r = 0.71♀ Leisure r = 0.76♂, r = 0.83♀ Total r = 0.85♂, r = 0.83♀	3 + 3 + 3 + 3 – 3 – 3 + 3 + 3 +
Modified Baecke (ARIC/Baecke)	28♂ 49♀; 37 y; US[4]	26 d	Sport and exercise-related leisure index r = 0.92♂, r = 0.87♀ Non-sport and exercise-related leisure index r = 0.88♂, r = 0.86♀ Total leisure activity r = 0.92♂, r = 0.90♀	2 + 2 + 2 + 2 + 2 + 2 +
	28♂ 50♀; 37 y; US[3]	1 mo	Total r = 0.93 Work r = 0.78 Sport r = 0.90 Leisure r = 0.86	2 + 2 – 2 + 2 +
Extended Baecke (QAPSE)	7♂ 13♀; 23–54 y; FR[28]	6 wk	TEE r = 0.997	2 ?
Bharathi Q	45♂ 67♀; 18–60 y; IN[45]	2–4 wk	TEE r = 0.86 PAL r = 0.54	2 + 2 –
CARDIA	28♂ 50♀; 37 y; US[3]	1 mo	Total r = 0.88 Mod r = 0.66 Heavy r = 0.91	2 + 2 – 2 +
EPIC original Q	62♂; 41 y: 50♀; 49 y; NL[10]	5 mo	Total r = 0.76♂, r = 0.58♀ Occup r = 0.90♂, r = 0.79♀ Leisure r = 0.85♂, r = 0.68♀ Rest r = 0.67♂, r = 0.65♀	3 – 3 – 3 + 3 – 3 + 3 – 3 – 3 –
Modified EPIC Q (short PA Index)	2271; UK[47]	18–21 mo	PA index κ = 0.60	2 –
EPAQ2	187♂; 65 y: 212♀; 64 y; UK[9]	3 mo	TV time κ = 0.71♂, κ = 0.74♀ Activity at home κ = 0.61♂, κ = 0.62♀ Activity at work κ = 0.79♂, κ = 0.82♀ Recreational activity κ = 0.54♂, κ = 0.55♀ Vig activity κ = 0.58♂, κ = 0.67♀ PA index κ = 0.66♂, κ = 0.70♀	1 + 1 + 1 – 1 – 1 + 1 + 1 – 1 – 1 – 1 – 1 – 1 +
Flemish PA computerized Q	31♂; 39 y: 35♀; 42 y; BE[48]	2 wk	PAL ICC = 0.92♂, ICC = 0.78♀	1 ?
Godin Q	53; 18–65 y; CA[50]	2 wk	Total ICC = 0.74 Strenuous ICC = 0.94 Mod ICC = 0.46 Light ICC = 0.48	1 + 1 + 1 – 1 –
	28♂ 50♀; 37 y; US[3]	1 mo	Leisure r = 0.62 Mod r = 0.36 Vig r = 0.84	2 – 2 – 2 +
Harvard/College Alumnus Q	28♂ 50♀; 37 y; US[3]	1 mo	TEE r = 0.72 Sports r = 0.75	2 – 2 –
	21♂ 38♀; 39 y; US[51]	28 d	Leisure EE r = 0.61♂, r = 0.75♀	2 ? 2 ?

Continued next page

Table V. Contd

Questionnaire	Study population (n; mean age; nationality)	Interval	Results	Rating
HUNT 1	S7S: 108♂; 32 y; NOR[12]	1 wk	Frequency $\kappa_w = 0.80$	1 +
			Intensity $\kappa_w = 0.82$	1 +
			Duration $\kappa_w = 0.69$	1 –
HUNT 2	108♂; 32 y; NOR[13]	1 wk	Light $\kappa = 0.20$	1 –
			Hard $\kappa = 0.41$	1 –
			Work $\kappa = 0.80$	1 +
IPAQ	S7S: 111; 21 y; US[93]	4–6 d	Total ICC = 0.86	1 +
			Vig ICC = 0.89	1 +
			Mod ICC = 0.71	1 +
			Walking ICC = 0.89	1 +
	S7S: 292[a]; 18–65 y[14]	3–7 d	S7S TEE r = 0.75, ACSM r = 0.93–1.0	2 – 2 +
	SUS: 906; 18–65 y[14]		SUS TEE r = 0.79, ACSM r = 0.77–0.99	2 – 2 +
	L7S: 294; 18–65 y[14]		L7S TEE r = 0.77, ASCM r = 0.92–1.0	2 – 2 +
	LUS: 904; 18–65 y[14]		LUS TEE r = 0.83, ASCM r = 0.90–1.0	2 + 2 +
	S7S: 108♂; 32 y; NOR[12]	1 wk	Vig ICC = 0.61–0.62	1 –
			Mod ICC = 0.30–0.34	1 –
			Walking ICC = 0.42–0.56	1 –
			Sitting ICC = 0.80	1 +
	S7S; 30♂; 26 y: 19♀; 34 y; CH[96]	3 d	Total ICC = 0.79	1 +
			Vig ICC = 0.75	1 +
			Mod ICC = 0.31	1 –
			Walking ICC = 0.93	1 +
			Sitting ICC = 0.97	1 +
computerized IPAQ	LUS; 23♂ 30♀; 31 y; BE[97]	7 + 3–6 d	Total ICC = 0.69 (ICC over three meas)	1 –
			Vig ICC = 0.82 (ICC over three meas)	1 +
			Mod ICC = 0.63 (ICC over three meas)	1 –
IPAQ Sitting Q	L7S and S7S: 65♂ 78♀; 35 y; UK[97]	3–7 d	L7S r = 0.82♂, r = 0.65♀	2 + 2 –
	L7S and S7S: 66; 33 y; NL		S7S r = 0.81♂, r = 0.63♀	2 + 2 –
	L7S and S7S: 25; 49 y; US		L7S r = 0.87, S7S r = 0.95	2 + 2 +
	L7S and S7S: 29; 36 y; US		L7S r = 0.95, S7S r = 0.92	2 ? 2 ?
			L7S r = 0.85, S7S r = 0.85	2 ? 2 ?
JACC Q	425♂ 650♀; 40–79 y; JAP[15]	1 y	PA time $\kappa = 0.45$♂, $\kappa = 0.40$♀	2 – 2 –
			Walking time $\kappa = 0.32$♂, $\kappa = 0.31$♀	2 – 2 –
			PA freq $\kappa = 0.50$♂, $\kappa = 0.51$♀	2 – 2 –
Kaiser PA Survey	50♀; 39 y; US[56]	1 mo	3-point summary ICC = 0.82	1 +
			4-point summary ICC = 0.83	1 +
			Caregiving ICC = 0.01	1 –
			Housework ICC = 0.79	1 +
			Housework/caregiving ICC = 0.81	1 +
			Sports/exercise ICC = 0.84	1 +
			Active living habits ICC = 0.82	1 +
			Occup ICC = 0.85	1 +
Life in NZ National Survey	36–48; 43 y; NZ[103]	?	$Activity_{hi}$ ICC = 0.70–0.88	2 ?
			$Activity_{lo}$ ICC = 0.50–0.71	2 ?
Lipid Research Clinics Q	28♂ 50♀; 37 y; US[3]	1 mo	4-point scorer = 0.93	2 +
	28♂ 50♀; 40 y; US[59]	4 wk	2-point scorer = 0.85	2 +
			4-point score r = 0.88	2 +
Minnesota LTPA Q	28♂ 50♀; 37 y; US[3]	1 mo	Leisure EE r = 0.92	2 +
			Mod r = 0.80	2 +
			Heavy r = 0.95	2 +

Continued next page

Table V. Contd

Questionnaire	Study population (n; mean age; nationality)	Interval	Results	Rating
Modified Minnesota LTPA Q (Canada Fitness Survey)	64♂; 49 y: 63♀; 46 y; CA[64]	3–4 wk	Total ICC = 0.53 (time) ICC = 0.48 (TEE) Leisure ICC = 0.52 (time) ICC = 0.58 (TEE) Non-leisure ICC = 0.62 (time) ICC = 0.26 (TEE) Strenuous ICC = 0.86♂, ICC = 0.31♀	1 – 1 – 1 – 1 – 1 – 1 – 1 + 1 –
Minnesota Heart Health Program Q	28♂ 50♀; 37 y; US[3]	1 mo	Work index r = 0.91 Leisure index r = 0.86	2 + 2 +
Modified Minnesota LTPA Q (y11 Q)	129♂ 322♀; 41 y; US[65]	1–10 y	Leisure EE r = 0.20♂, r = 0.29♀ Leisure EE κ = 0.49♂, κ = 0.40♀ *(high v low)* Light EE r = 0.17♂, r = 0.25♀ Mod EE r = 0.17♂, r = 0.25♀ Vig EE r = 0.47♂, r = 0.41♀ Vig EE κ = 0.67♂, κ = 0.32♀ *(high v low)*	3 – 3 – 2 – 2 – 3 – 3 – 3 – 3 – 3 – 3 – 2 – 2 –
Modified Minnesota LTPA Q + TOQ Q + new household activity measure	59♀; 47 y; US[66]	2 wk	Occup EE r = 0.75; LOA = –0.009 ± 0.90 Leisure EE r = 0.46; LOA = –0.05 ± 2.25 Household EE r = 0.64; LOA = –0.25 ± 1.80	2 – 2 – 2 –
MOSPA	65; 36 y; BE[17]	<3 mo	TEE ICC = 0.68 Work ICC = 0.85 Transport ICC = 0.62 Household ICC = 0.91 LTPA ICC = 0.87	2 – 2 + 2 – 2 + 2 +
NHS II Activity Q	147♀; 39 y; US[19]	2 y	Activity score r = 0.59 Inactivity score r = 0.52	3 – 3 –
Modified NHS II Activity Q	238♂ 40–75 y; US[69]	2 y	Vig activity ICC = 0.52 Non-vig activity ICC = 0.42 Sum of activities ICC = 0.41 Inactivity at home ICC = 0.39 Inactivity at work ICC = 0.50 Overall inactivity ICC = 0.39	2 – 2 – 2 – 2 – 2 – 2 –
Norman Q	222♂ 63 y; SW[70]	7 mo	Crude total PA C = 0.66 Total PAC = 0.67 C = 0.78 age 44–64 y C = 0.51 age 65–78 y C = 0.70 BMI ≤26 C = 0.64 BMI >26 Occup C = 0.70 Home C = 0.66 Leisure C = 0.61 TV/reading C = 0.67 Sleeping C = 0.75	2 – 2 – 2 + 2 – 2 + 2 – 2 + 2 – 2 – 2 – 2 +
One-wk recall Q	55♂ 38 y; 63♀; 40 y; AUS[71]	3 d	Walking ICC = 0.67♂, ICC = 0.86♀ Mod ICC = 0.71♂, ICC = 0.53♀ Vig ICC = 0.38♂, ICC = 0.89♀ Total duration ICC = 0.45♂, ICC = 0.80♀ Meeting fitnorm[71] κ = 0.64♂, κ = 0.55♀	1 – 1 + 1 + 1 – 1 – 1 + 1 – 1 + 1 – 1 –
PYTPAQ	75♂ 79♀; 49 y; CA[26,27]	9 wk	Total ICC = 0.66 Vig ICC = 0.72 Low/Mod ICC = 0.55 Occup ICC = 0.58	1 – 1 + 1 – 1 –
Scottish PA Q	9♂ 25♀; 33 y; SC[73]	2 d	Total r = 0.998, COR = 53 min Leisure COR = 29 min, occup COR = 55 min	2 ?

Continued next page

Table V. Contd

Questionnaire	Study population (n; mean age; nationality)	Interval	Results	Rating
web-based vs *paper version*	16; UK[74]	1 wk	Total r = 0.67	2 ?
Stanford SDR	28♂ 50♀; 37 y; US[3]	1 mo	Total r = 0.34 Mod r = 0.12 Vig r = 0.37	2 – 2 – 2 –
	90♂ 73♀; 22 y; US[38]	3 wk 4 wk 7 wk	TEE r = 0.58 TEE r = 0.63 TEE r = 0.42	2 – 2 – 2 –
Stanford Usual Act Q	28♂ 50♀; 37 y; US[3]	1 mo	Mod r = 0.77 Vig r = 0.67	2 – 2 –
Usual PA measure	37♀; 40–55 y; US[83]	14 d	Total r = 0.88	2 ?
SQUASH	36♂ 14♀; 44 y; NL[32]	5 wk	Total r = 0.58 Sports r = 0.90	2 – 2 +
Suzuki Q	95♂; 37–72 y: 119♀; 35–73 y; JAP[81]	1 y	TEE (day) r = 0.59♂, r = 0.62♀ TEE (wk) r = 0.37♂, r = 0.43♀	3 – 3 – 3 – 3 –
Singh Q	59♂ 53♀; 52 y; US[77]	6 wk	PA index r = 0.56–0.80♂, r = 0.76♀ RWJ index r = 0.77–0.78♂, r = 0.70–0.85♀ Total activity index r = 0.51♂	2 – 2 – 2 – 2 – 2 –
	29♂ 70♀; 49 y; US[78]	6 wk	RWJ index r = 0.65♂, r = 0.64♀ Vig activity r = 0.82♂, r = 0.78♀ Sport/recreational index r = 0.91♂, r = 0.65♀ Total activity[b] r = 0.78 ♂, r = 0.64♀	2 ? 2 – 2 ? 2 – 2 ? 2 – 2 ? 2 –
Total PA	39 + 94; 41 y; SW[82]	3 wk	Total r = 0.73	2 –
Lifetime PA				
Modified HLAQ	134♀; 50 y; US[11,102]	1 y	TEE ICC = 0.82 Mod EE ICC = 0.80 Vig EE ICC = 0.86 Recreational ICC = 0.87 Household ICC = 0.78	1 + 1 + 1 + 1 + 1 +
Occup PA				
Modified Baecke ARIC/Baecke Work Index	27♂ 48♀; 37 y; US[31]	1 mo	Work index r = 0.74	2 –
Health Insurance Plan of NY Q	27♂ 48♀; 37 y; US[31]	1 mo	Total occup r = 0.83	2 +
	28♂ 50♀; 37 y; US[3]	1 mo	Total occup r = 0.86	2 +
Minnesota Heart Health Program Occup Q	27♂ 48♀; 37 y; US[31]	1 mo	Total occup r = 0.84	2 +
Modified Stanford SDR	27♂ 48♀; 37 y; US[31]	1 mo	Total occup score activity score/wk r = 0.58 Total occup score h/wk r = 0.56 Total occup score MET • min/wk r = 0.20	2 – 2 – 2 –
TOQ	27♂ 48♀; 37 y; US[31]	1 mo	Total occup score activity score/wk r = 0.83 Total occup score h/wk r = 0.63 Total occup score MET • min/wk r = 0.37	2 + 2 – 2 –
Lipid Research Clinics Occup Q	27♂ 48♀; 37 y; US[31]	1 mo	Total occup r = 0.73	2 –
CARDIA Occup	27♂ 48♀; 37 y; US[31]	1 mo	Total occup r = 0.37	2 –

Continued next page

Table V. Contd

Questionnaire	Study population (n; mean age; nationality)	Interval	Results	Rating
Walking activities				
NPAQ	82; 20–71 y; AUS[20]	1 wk	Total walking ICC = 0.91	1 +
Walking Q	51 ♂ 55 ♀; 62 y; JAP[84]	3 mo	Walking 59–74% agreement	1 –
Bone loading PA				
Bone Loading History Q	78 ♀; 31 y; US[86]	4–6 wk	Total PA hip ICC = 0.89 Total PA spine ICC = 0.92	1 + 1 +
Historical Activity Q	31 ♀; 21 y, US[87]	6.5 mo	Total r = 0.76 Athletics r = 0.82 Exercise r = 0.55 Leisure r = 0.70 Occup r = 0.48 Lifting/carrying r = 0.51	2 –

a Pooled data from 12 countries.

b Calculated slightly differently from the Total Activity Index in Singh et al.[77]

ACSM = meeting PA guidelines of the American College of Sports Medicine; **Activity$_{hi}$** = activity of high intensity; **Activity$_{lo}$** = activity of low intensity; **AUS** = Australia; **BE** = Belgium; **BMI** = body mass index; **C** = concordance; **CA** = Canada; **CH** = China; **COR** = coefficient of repeatability; **EE** = energy expenditure; **FR** = France; **ICC** = intraclass correlation coefficient; **IN** = India; **JAP** = Japan; **κ** = Kappa; **κ$_w$** = weighted Kappa; **LOA** = limits of agreement; **LTPA** = leisure time physical activity; **LUS** = long form, usual week; **meas** = measurements; **mod** = moderate; **NL** = Netherlands; **occup** = occupational; **NOR** = Norway; **PAL** = PA level; **r** = correlation coefficient; **RWJ** = run-walk-jog; **S7S** = short form, last 7 d; **SC** = Scotland; **SUS** = short form, usual week; **SW** = Sweden; **TEE** = total energy expenditure; **TV** = television; **UK** = United Kingdom; **US** = United States; **vig** = vigorous; ♀ indicates female; ♂ indicates male.

Construct validity was assessed by validation against doubly labelled water for seven questionnaires.[16,21,40,81,104,105] In all these studies, the correlation of total energy expenditure assessed with the questionnaire and with doubly labelled water was lower than our criterion of 0.70, with Pearson correlations ranging between 0.31 and 0.58 (table IV).

In 41 studies, construct validity was assessed by validation against accelerometers (table IV). For only one questionnaire, validated in a study with >50 participants, the correlation between accelerometer data and total PA was >0.50 (Suzuki Q[81]).

In an attempt to find out which type of questionnaire performed best, we averaged the correlations found in the 41 studies using accelerometers as the comparison measure. It was clear that correlations differed slightly between vigorous and moderate activity, with higher correlations for vigorous activity (r = 0.32 vs 0.22). Also, a higher correlation was found for questionnaires asking about the past week, instead of a usual week/usual PA/current PA or about the past year (r = 0.41 vs 0.26 and 0.30, respectively).

Two questionnaires designed for measuring walking were validated against pedometers (Level 1). One scored negative[85] and the other was rated as indeterminate because of a statistical analysis that could not be interpreted.[84] The *reliability* of 15 versions of PA questionnaires was assessed at Level 1 (table V), and only five showed positive results: the self-administered, short version of the IPAQ on PA in the past 7 days (S7S),[93] the Modified HLAQ,[11,102] the NPAQ[20] and the Bone Loading History Q[86] scored positive on all aspects, and the Kaiser PA Survey[56] scored positive on all aspects, except 'care giving'. The other questionnaires showed mixed results or scored negative on most aspects, or scored indeterminate because of a small sample size.

In addition to the 15 questionnaires for which evidence on Level 1 was available, Level 2 evidence was found for another 36 (versions of) questionnaires. For only six questionnaires, a positive score on Level 2 was given (Modified Baecke [(ARIC) Baecke],[4] Health Insurance Plan of NY Q,[3,31] Lipid Res Clin Q,[3,59] Minnesota LTPA Q,[3] the Minnesota Heart Health Program Q,[3] and the Minnesota Heart Health

Program Occupational Q[31]). The other questionnaires showed mixed results or scored negative on most aspects, or scored indeterminate because of a small sample size.

When averaging the results of the reliability studies, no clear differences were found between questionnaires with different recall periods, between different time intervals between test and retest or between sexes. The only difference found was that, on average, the reliability for vigorous activity was higher than for moderate activity.

The *responsiveness* of a questionnaire was assessed in only two studies,[38,54] and seemed to be poor. The correlation between changes in self-reported PA and changes in supervised activity in a training programme was −0.07 for total energy expenditure and 0.01 for vigorous activity.[38] The correlation of change in PA assessed with an adapted version of the long form of the IPAQ with change in $\dot{V}O_{2max}$ was 0.20 for men and 0.12 for women.[54]

3. Discussion

Although more than 90 papers have been published on the validity or reliability of PA questionnaires, this is the first systematic review of studies assessing the measurement properties of PA questionnaires, in which the results as well as the methodological quality of the individual studies have been taken into account. Our results indicate that the overall methodological quality of the studies could be much improved. Most common flaws were small sample size and inadequate analyses, and for construct validity, comparison measures that were not measuring the same construct.

An important finding of our review was the poor reporting of methods and results of the studies. It was often unclear what dimension of PA the questionnaire was supposed to measure. This made assessing content validity sometimes impossible. Furthermore, it was extremely difficult, if not impossible, to assess whether the same or slightly modified versions of questionnaires were used in some studies, and it was not always clear whether the data were derived from a self-report questionnaire or whether the questionnaire was part of an interview.

For assessing *construct validity*, it is important to formulate specific hypotheses in advance about expected correlations between the questionnaire under study and other measures. However, almost none of the studies had formulated such hypotheses. To be able to provide levels of evidence we formulated hypotheses regarding the strength of the association between comparison instruments. This methodology is not new, and the idea behind it is that, in retrospect, it is always easy and tempting to come up with explanations for the findings and conclude that the questionnaire is valid. In fact, most studies in our review concluded that the questionnaire under study was valid. However, when we applied our criteria we found that these conclusions were overly optimistic in almost all cases.

Reliability was also often poorly assessed. Many studies used large time intervals between the test and retest, and in most studies Pearson or Spearman correlation coefficients were calculated instead of ICCs or Kappas. This is partly because we included studies performed many years ago, when Pearson correlation was still an accepted method, but nowadays there is a consensus that calculating ICCs or Kappas is the preferred method for assessing reliability.

Only two studies evaluated *responsiveness*, i.e. the ability of a questionnaire to detect change in PA over time. This is amazing, given the importance of responsiveness of a questionnaire when used in PA intervention studies. If a questionnaire has poor responsiveness, treatment effects cannot be detected, or only with large sample sizes. For some questionnaires, the majority of the population scored the highest or lowest possible score (e.g. with the modified CHAMPS[6]). When this happens, there is little opportunity for change, leading to low responsiveness. Although the methodology of assessing responsiveness tends to be less well understood, there is a consensus that responsiveness should be considered an aspect of validity, in a longitudinal context.[106] While construct validity is about the validity of a single score, responsiveness is about the validity of a change score. This means that

similar methods can be applied as for assessing validity to assess the validity of changes in PA scores over time, i.e. stating *a priori* hypotheses.

We found that correlations between PA questionnaire data and accelerometer data were slightly higher in questionnaires asking about the previous week compared with those asking about a usual week. Often, accelerometers were worn in the week that was captured by the questionnaire. It might be that this explains why higher correlations were found for these questionnaires compared with those that asked about a usual week or usual PA. So, whether questionnaires asking about the previous week are really better in assessing PA, or that this is a consequence of the testing procedures, needs to be determined.

3.1 Limitations of this Review

As with any other systematic review, it is possible we missed some relevant papers with our literature search. We only used the search terms 'questionnaire', 'physical activity', 'exercise' and 'motor activity' and did not include alternative wordings, such as 'survey'. However, after checking all references of relevant papers retrieved in our search, it proved that very few papers were missed.

Because of an overwhelming amount of data available, we had to be selective in what to present in this review. First of all, we chose to limit the review to self-administered questionnaires, realizing that some questionnaires have been used in other forms as well, such as interview-administered. We realized that with this restriction we have ignored some studies on questionnaires that can be either self-administered or used as an interview. The measurement properties of these questionnaires may be different in these two applications. Therefore, by restricting the review to one form of administration, the studies were more homogeneous and we felt better comparisons across questionnaires could be made, without allowing for the type of administration as well. Further, when assessing validity, only correlations with accelerometer data, $\dot{V}O_{2max}$, BMI and percentage body fat were extracted from the papers, because we felt that, although these are different constructs, these comparison measures were most closely related to the construct being measured in the questionnaires. We have ignored correlations with, for example, cholesterol or blood pressure in these comparisons because only a limited correlation with PA can be expected. Lastly, not all scores resulting from the questionnaires could be presented. We often restricted the information to the overall or total PA scores. Data were presented for men and women separately when relevant (i.e. in case of sex differences).

Interpretation of the results was difficult for some studies, mostly due to poor reporting. Although two reviewers independently extracted data from the papers, interpretation may have been incorrect in some cases. Given the number of studies included in the review, and the number of studies conducted a long time ago, we chose not to contact the authors of the original studies.

Many of the choices for scoring the quality of the studies have been made without a very strong basis on theory or evidence, simply because there is not much available to base these choices on. Others might have chosen different cut-off points for scoring negative or positive on validity or reliability. The same is true for the decision on what is a sufficient sample size and what is the appropriate time interval between test-retest. However, readers can decide according to their own insights and draw their own conclusions from the data provided in the tables.

3.2 Recommendations for Choosing a Questionnaire

Current US recommendations state that every adult should participate 2.5 hours a week in moderate intensity or 75 minutes a week in vigorous intensity aerobic PA or in an equivalent combination of moderate and vigorous intensity activity. Aerobic activity should be performed in episodes of at least 10 minutes, preferably spread throughout the week. Based on these recommendations, questionnaires for measuring total PA should at least measure duration and frequency, and measure PA in all settings (work, home, transport, recreation, sport) to have sufficient content validity. Especially older questionnaires, such as the Baecke questionnaire,[41] do not fulfil this criterion,

because insight into what PA for health should entail has changed over time.

Of course, some researchers will need a PA questionnaire not only for measuring total PA but also for different purposes, and different aspects of PA might be relevant for their study. For instance, when looking at bone health, energy expended in cycling or swimming might be less important, but carrying loads would be of interest. So there will not be one questionnaire suitable for all purposes or target groups. The choice for a certain questionnaire should therefore always start with defining the purpose of the study and the PA measurement, after which the content validity of a possible questionnaire should be judged. Only then do construct validity and reliability need to be considered.

In this review, the content of 23 questionnaires was deemed appropriate for the dimension of PA it was intended to measure (Bharati,[45] EPIC original Q,[10] EPAQ2,[9] Harvard/College Alumnus Q,[3,51] the long version of the IPAQ,[14] the adapted IPAQ,[54] Kaiser PA Survey,[56] LACE PA Q,[7] LTPA Q,[61] Mail Survey of PA,[62] Norman Q,[70] NZPAQ-SF,[21] One-week recall Q,[71] PAFQ,[22] PA History Q,[72] PYTPAQ,[26] Singh Q,[77,78] SQUASH,[32] Historical walking, running and jogging questionnaire,[30] NPAQ,[20] Health Insurance Plan of NY,[3] TOQ[31,89] London PA Q[88]). Unfortunately, for only 13 of these 23 questionnaires was both reliability and construct validity studied (Bharati,[45] EPIC original Q,[10] EPAQ2,[9] Harvard/College Alumnus Q,[3,51] Kaiser PA Survey,[56] the long version of the IPAQ,[14] Norman Q,[70] One-week recall Q,[71] PYTPAQ,[26] Singh Q,[77,78] SQUASH,[32] Health Insurance Plan of NY,[3] TOQ[31,89]).

Of the 23 questionnaires with sufficient content validity, the Kaiser PA Survey,[56] the Godin Q,[50] the NPAQ,[20] Bharati Q,[45] the LUS version of the IPAQ,[14] One-week recall Q,[71] and the Health Insurance Plan of NY[3] scored good for reliability at Level 1 or 2. Construct validity was sufficient according to our criteria only for the L7S version of the IPAQ in one study,[92] although validity for the Kaiser PA Survey[56] was 0.49, which is only just below the (arbitrarily chosen) cut-off point of 0.50.

In recent studies, the IPAQ seems to be used most often and it is by far the most widely validated questionnaire at present.[14,91-95,97,107] Reliability of the IPAQ was not shown consistently within or between studies, although the short version for the past 7 days (S7S) and the long version for a usual week (LUS) seemed to perform best. We therefore recommend additional reliability studies of the IPAQ. Validity of the IPAQ seems questionable. First, content validity of the short forms seems limited because it does not discriminate between different settings. The long form, which does discriminate between five settings therefore has a better content validity, but it was reported to be "too boring and repetitive" and too long for routine surveillance.[14] The construct validity of both the short and the long forms varied widely, but were mostly below our criteria. Of the self-administered IPAQ forms, only for the L7S was a correlation found with an accelerometer – of 0.52 found in Finland[14] and 0.55 in Sweden[92] – and for the S7S in the US in men only.[95] Discrimination of the IPAQ between groups of people with different activity levels as measured with DLW[94] was questionable, although differentiation between groups with different fitness levels was adequate.[91] Therefore, we feel that additional well designed studies on the measurement properties, with specific attention to responsiveness, of the IPAQ are required.

3.3 Recommendations for Further Research

For future studies, we recommend choosing from the abovementioned 23 questionnaires that we identified as having sufficient content validity, and validating those further for reliability, construct validity and especially responsiveness.

The results of this review indicate that one study on validity and reliability of a questionnaire is not enough. A number of other questionnaires were validated in more than one study, and without exception the results were conflicting: the questionnaires showed sufficient validity in one study and not in another. Also, in the large international study on validity and reliability of the IPAQ, huge differences were found between countries. This indicates that it is important for

researchers to assess the measurement properties of a questionnaire in their own language and in their own target population. As the majority of the studies on measurement properties of PA questionnaires have been conducted in the US, it remains to be seen whether the results can be generalized to other countries. We therefore strongly recommend researchers to assess measurement properties of a questionnaire carefully in their own target group.

Although PA questionnaires are frequently used for the evaluation of the effects of intervention, surprisingly little attention has been paid to the responsiveness of these questionnaires. A prerequisite for detecting differences in PA after an intervention would be that the questionnaire is responsive to change. The two studies assessing responsiveness did not show positive results in that regard.

Finally, more attention should be paid to reporting on studies assessing measurement properties of PA questionnaires, since, for instance, it was often unclear what questionnaire was used and for what purpose the questionnaire was intended. The QAPAQ might be a useful tool when reporting on measurement properties.

4. Conclusions

Based on our review of the literature concerning measurement properties of questionnaires measuring PA, no conclusion can be drawn regarding the best questionnaire at the moment. Researchers should determine which questionnaire would fit their purposes best regarding the content of the questionnaire. Questionnaires with good content validity need to be validated in well designed studies and in different countries. Data on the responsiveness of PA questionnaires are urgently needed for the use of questionnaires in intervention studies.

Acknowledgements

No sources of funding were used to assist in the preparation of this review. The authors have no conflicts of interest that are directly relevant to the content of this review.

References

1. Powell KE, Thompson PD, Caspersen CJ, et al. Physical activity and the incidence of coronary heart disease. Annu Rev Public Health 1987; 8: 253-87
2. 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): 126-31
3. Jacobs Jr DR, Ainsworth BE, Hartman TJ, et al. A simultaneous evaluation of 10 commonly used physical activity questionnaires. Med Sci Sports Exerc 1993; 25 (1): 81-91
4. Richardson MT, Ainsworth BE, Wu HC, et al. Ability of the Atherosclerosis Risk in Communities (ARIC)/Baecke Questionnaire to assess leisure-time physical activity. Int J Epidemiol 1995; 24 (4): 685-93
5. Jacobs J, Hahn LP, Haskell WL, et al. Validity and reliability of short physical activity history: Cardia and the Minnesota Heart Health Program. J Cardiopulm Rehabil 1989; 9 (11): 448-59
6. Resnicow K, McCarty F, Blissett D, et al. Validity of a modified CHAMPS physical activity questionnaire among African-Americans. Med Sci Sports Exerc 2003; 35 (9): 1537-45
7. Altschuler A, Picchi T, Nelson M, et al. Physical activity questionnaire comprehension: lessons from cognitive interviews. Med Sci Sports Exerc 2009; 41 (2): 336-43
8. Mokkink LB, Terwee CB, Knol DL, et al. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. BMC Med Res Methodol 2006 Jan 24; 6: 2
9. Wareham NJ, Jakes RW, Rennie KL, et al. Validity and repeatability of the EPIC-Norfolk Physical Activity Questionnaire. Int J Epidemiol 2002; 31 (1): 168-74
10. Pols MA, Peeters PH, Ocke MC, et al. Relative validity and repeatability of a new questionnaire on physical activity. Prev Med 1997; 26 (1): 37-43
11. Chasan-Taber L, Erickson JB, McBride JW, et al. Reproducibility of a self-administered lifetime physical activity questionnaire among female college alumnae. Am J Epidemiol 2002; 155 (3): 282-9
12. Kurtze N, Rangul V, Hustvedt BE, et al. Reliability and validity of self-reported physical activity in the Nord-Trondelag Health Study: HUNT 1. Scand J Public Health 2008; 36 (1): 52-61
13. Kurtze N, Rangul V, Hustvedt BE, et al. Reliability and validity of self-reported physical activity in the Nord-Trondelag Health Study (HUNT 2). Eur J Epidemiol 2007; 22 (6): 379-87
14. Craig CL, Marshall AL, Sjostrom M, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003; 35 (8): 1381-95
15. Iwai N, Hisamichi S, Hayakawa N, et al. Validity and reliability of single-item questions about physical activity. J Epidemiol 2001; 11 (5): 211-8
16. Walsh MC, Hunter GR, Sirikul B, et al. Comparison of self-reported with objectively assessed energy expenditure in black and white women before and after weight loss. Am J Clin Nutr 2004; 79 (6): 1013-9
17. Roeykens J, Rogers R, Meeusen R, et al. Validity and reliability in a Flemish population of the WHO-MONICA

Optional Study of Physical Activity Questionnaire. Med Sci Sports Exerc 1998; 30 (7): 1071-5

18. Miller DJ, Freedson PS, Kline GM. Comparison of activity levels using the Caltrac accelerometer and five questionnaires. Med Sci Sports Exerc 1994; 26 (3): 376-82
19. Wolf AM, Hunter DJ, Colditz GA, et al. Reproducibility and validity of a self-administered physical activity questionnaire. Int J Epidemiol 1994; 23 (5): 991-9
20. Giles-Corti B, Timperio A, Cutt H, et al. Development of a reliable measure of walking within and outside the local neighborhood: RESIDE's Neighborhood Physical Activity Questionnaire. Prev Med 2006; 42 (6): 455-9
21. Maddison R, Ni MC, Jiang Y, et al. International Physical Activity Questionnaire (IPAQ) and New Zealand Physical Activity Questionnaire (NZPAQ): a doubly labelled water validation. Int J Behav Nutr Phys Act 2007 Dec 3; 4: 62
22. Bernstein M, Sloutskis D, Kumanyika S, et al. Data-based approach for developing a physical activity frequency questionnaire. Am J Epidemiol 1998; 147 (2): 147-54
23. Copeland JL, Kowalski KC, Donen RM, et al. Convergent Validity of the Physical Activity Questionnaire for Adults: the new member of the PAQ family. J Phys Act Health 2005; 2 (2): 216
24. Aadahl M, Jorgensen T. Validation of a new self-report instrument for measuring physical activity. Med Sci Sports Exerc 2003; 35 (7): 1196-202
25. Aadahl M, Kjaer M, Kristensen JH, et al. Self-reported physical activity compared with maximal oxygen uptake in adults. Eur J Cardiovasc Prev Rehabil 2007; 14 (3): 422-8
26. Friedenreich CM, Courneya KS, Neilson HK, et al. Reliability and validity of the Past Year Total Physical Activity Questionnaire. Am J Epidemiol 2006; 163 (10): 959-70
27. Ferrari P, Friedenreich C, Matthews CE. The role of measurement error in estimating levels of physical activity. Am J Epidemiol 2007; 166 (7): 832-40
28. Berthouze SE, Minaire PM, Chatard JC, et al. A new tool for evaluating energy expenditure: the "QAPSE" development and validation. Med Sci Sports Exerc 1993; 25 (12): 1405-14
29. Terwee CB, Mokkink LB, van Poppel MNM, et al. Qualitative attributes and measurement properties of physical activity questionnaires: the QAPAQ checklist. Sports Med 2010; 40 (7): 525-37
30. Bowles HR, FitzGerald SJ, Morrow Jr JR, et al. Construct validity of self-reported historical physical activity. Am J Epidemiol 2004; 160 (3): 279-86
31. Ainsworth BE, Jacobs Jr DR, Leon AS, et al. Assessment of the accuracy of physical activity questionnaire occupational data. J Occup Med 1993; 35 (10): 1017-27
32. Wendel-Vos GC, Schuit AJ, Saris WH, et al. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol 2003; 56 (12): 1163-9
33. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007; 60 (1): 34-42
34. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. New York: Oxford University Press, 2003
35. de Vet HCW. Observer reliability and agreement. In: Armitage P, Colton T, editors. Encyclopedia of biostatistics. Boston (MA): John Wiley & Sons Ltd, 1998: 3123-8
36. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. J Chronic Dis 1986; 39: 897-906
37. Brown WJ, Burton NW, Marshall AL, et al. Reliability and validity of a modified self-administered version of the Active Australia physical activity survey in a sample of mid-age women. Aust NZ J Public Health 2008; 32 (6): 535-41
38. Dishman RK, Steinhardt M. Reliability and concurrent validity for a 7-d re-call of physical activity in college students. Med Sci Sports Exerc 1988; 20 (1): 14-25
39. Aires N, Selmer R, Thelle D. The validity of self-reported leisure time physical activity, and its relationship to serum cholesterol, blood pressure and body mass index: a population based study of 332,182 men and women aged 40-42 years. Eur J Epidemiol 2003; 18 (6): 479-85
40. Staten LK, Taren DL, Howell WH, et al. Validation of the Arizona Activity Frequency Questionnaire using doubly labeled water. Med Sci Sports Exerc 2001; 33 (11): 1959-67
41. Baecke JA, Burema J, Frijters JE. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. Am J Clin Nutr 1982; 36 (5): 936-42
42. Albanes D, Conway JM, Taylor PR, et al. Validation and comparison of eight physical activity questionnaires. Epidemiology 1990; 1 (1): 65-71
43. Pols MA, Peeters PH, Bueno-de-Mesquita HB, et al. Validity and repeatability of a modified Baecke questionnaire on physical activity. Int J Epidemiol 1995; 24 (2): 381-8
44. Canon F, Levol B, Duforez F. Assessment of physical activity in daily life. J Cardiovasc Pharmacol 1995; 25 Suppl. 1: S28-34
45. Bharathi AV, Sandhya N, Vaz M. The development & characteristics of a physical activity questionnaire for epidemiological studies in urban middle class Indians. Indian J Med Res 2000; 111: 95-102
46. Carter-Nolan PL, Adams-Campbell LL, Makambi K, et al. Validation of physical activity instruments: Black Women's Health Study. Ethn Dis 2006; 16 (4): 943-7
47. Wareham NJ, Jakes RW, Rennie KL, et al. Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. Public Health Nutr 2003; 6 (4): 407-13
48. Matton L, Wijndaele K, Duvigneaud N, et al. Reliability and validity of the Flemish Physical Activity Computerized Questionnaire in adults. Res Q Exerc Sport 2007; 78 (4): 293-306
49. Gionet NJ, Godin G. Self-reported exercise behavior of employees: a validity study. J Occup Med 1989; 31 (12): 969-73
50. Godin G, Shephard RJ. A simple method to assess exercise behavior in the community. Can J Appl Sport Sci 1985; 10 (3): 141-6
51. Ainsworth BE, Leon AS, Richardson MT, et al. Accuracy of the College Alumnus Physical Activity Questionnaire. J Clin Epidemiol 1993; 46 (12): 1403-11
52. Strath SJ, Bassett Jr DR, Swartz AM. Comparison of the college alumnus questionnaire physical activity index with objective monitoring. Ann Epidemiol 2004; 14 (6): 409-15

53. Siconolfi SF, Lasater TM, Snow RC, et al. Self-reported physical activity compared with maximal oxygen uptake. Am J Epidemiol 1985; 122 (1): 101-5
54. Graff-Iversen S, Anderssen SA, Holme IM, et al. An adapted version of the long International Physical Activity Questionnaire (IPAQ-L): construct validity in a low-income, multiethnic population study from Oslo, Norway. Int J Behav Nutr Phys Act 2007 April 20; 4: 13
55. Graff-Iversen S, Anderssen SA, Holme IM, et al. Two short questionnaires on leisure-time physical activity compared with serum lipids, anthropometric measurements and aerobic power in a suburban population from Oslo, Norway. Eur J Epidemiol 2008; 23 (3): 167-74
56. Ainsworth BE, Sternfeld B, Richardson MT, et al. Evaluation of the kaiser physical activity survey in women. Med Sci Sports Exerc 2000; 32 (7): 1327-38
57. Salonen JT, Lakka T. Assessment of physical activity in population studies: validity and consistency of the methods in the Kuopio ischemic heart disease risk factor study. Scand J Sports Sci 1987; 9 (3): 89-95
58. Hopkins WG, Wilson NC, Russell DG. Validation of the physical activity instrument for the Life in New Zealand national survey. Am J Epidemiol 1991; 133 (1): 73-82
59. Ainsworth BE, Jacobs Jr DR, Leon AS. Validity and reliability of self-reported physical activity status: the Lipid Research Clinics questionnaire. Med Sci Sports Exerc 1993; 25 (1): 92-8
60. Lof M, Hannestad U, Forsum E. Assessing physical activity of women of childbearing age: ongoing work to develop and evaluate simple methods. Food Nutr Bull 2002; 23 (3 Suppl.): 30-3
61. Parker DL, Leaf DA, McAfee SR. Validation of a new questionnaire for the assessment of leisure time physical activity. Ann Sports Med 1988; 4 (2): 72-81
62. Kohl HW, Blair SN, Paffenbarger Jr RS, et al. A mail survey of physical activity habits as related to measured physical fitness. Am J Epidemiol 1988; 127 (6): 1228-39
63. Taylor HL, Jacobs Jr DR, Schucker B, et al. A questionnaire for the assessment of leisure time physical activities. J Chronic Dis 1978; 31 (12): 741-55
64. Weller IM, Corey PN. A study of the reliability of the Canada Fitness Survey questionnaire. Med Sci Sports Exerc 1998; 30 (10): 1530-6
65. Blair SN, Dowda M, Pate RR, et al. Reliability of long-term recall of participation in physical activity by middle-aged men and women. Am J Epidemiol 1991; 133 (3): 266-75
66. Wilbur J, Holm K, Dan A. A quantitative survey to measure energy expenditure in midlife women. J Nurs Meas 1993; 1 (1): 29-40
67. Conway JM, Irwin ML, Ainsworth BE. Estimating energy expenditure from the Minnesota Leisure Time Physical Activity and Tecumseh Occupational Activity questionnaires: a doubly labeled water validation. J Clin Epidemiol 2002; 55 (4): 392-9
68. Mundal R, Erikssen J, Rodahl K. Assessment of physical activity by questionnaire and personal interview with particular reference to fitness and coronary mortality. Eur J Appl Physiol Occup Physiol 1987; 56 (3): 245-52
69. Chasan-Taber S, Rimm EB, Stampfer MJ, et al. Reproducibility and validity of a self-administered physical activity questionnaire for male health professionals. Epidemiology 1996; 7 (1): 81-6
70. Norman A, Bellocco R, Bergstrom A, et al. Validity and reproducibility of self-reported total physical activity: differences by relative weight. Int J Obes Relat Metab Disord 2001; 25 (5): 682-8
71. Timperio A, Salmon J, Crawford D. Validity and reliability of a physical activity recall instrument among overweight and non-overweight men and women. J Sci Med Sport 2003; 6 (4): 477-91
72. Sidney S, Jacobs Jr DR, Haskell WL, et al. Comparison of two methods of assessing physical activity in the Coronary Artery Risk Development in Young Adults (CARDIA) Study. Am J Epidemiol 1991; 133 (12): 1231-45
73. Lowther M, Mutrie N, Loughlan C, et al. Development of a Scottish physical activity questionnaire: a tool for use in physical activity interventions. Br J Sports Med 1999; 33 (4): 244-9
74. Marsden J, Jones RB. Validation of Web-based questionnaires regarding osteoporosis prevention in young British women. Health Bull (Edinb) 2001; 59 (4): 254-62
75. Bulley C, Donaghy M, Payne A, et al. Validation and modification of the Scottish Physical Activity Questionnaire for use in a female student population. Int J Health Promot Edu 2005; 43 (4): 117-24
76. Wilbur J, Miller A, Dan AJ, et al. Measuring physical activity in midlife women. Public Health Nurs 1989; 6 (3): 120-8
77. Singh PN, Tonstad S, Abbey DE, et al. Validity of selected physical activity questions in white Seventh-day Adventists and non-Adventists. Med Sci Sports Exerc 1996; 28 (8): 1026-37
78. Singh PN, Fraser GE, Knutsen SF, et al. Validity of a physical activity questionnaire among African-American Seventh-day Adventists. Med Sci Sports Exerc 2001; 33 (3): 468-75
79. Schechtman KB, Barzilai B, Rost K, et al. Measuring physical activity with a single question. Am J Public Health 1991; 81 (6): 771-3
80. Arroll B, Jackson R, Beaglehole R. Validation of a three-month physical activity recall questionnaire with a seven-day food intake and physical activity diary. Epidemiology 1991; 2 (4): 296-9
81. Suzuki I, Kawakami N, Shimizu H. Reliability and validity of a questionnaire for assessment of energy expenditure and physical activity in epidemiological studies. J Epidemiol 1998; 8 (3): 152-9
82. Lagerros YT, Mucci LA, Bellocco R, et al. Validity and reliability of self-reported total energy expenditure using a novel instrument. Eur J Epidemiol 2006; 21 (3): 227-36
83. Li S, Carlson E, Holm K. Validation of a single-item measure of usual physical activity. Percept Mot Skills 2000; 91 (2): 593-602
84. Tsubono Y, Tsuji I, Fujita K, et al. Validation of walking questionnaire for population-based prospective studies in Japan: comparison with pedometer. J Epidemiol 2002; 12 (4): 305-9
85. Bassett Jr DR, Cureton AL, Ainsworth BE. Measurement of daily walking distance-questionnaire versus pedometer. Med Sci Sports Exerc 2000; 32 (5): 1018-23

86. Dolan SH, Williams DP, Ainsworth BE, et al. Development and reproducibility of the bone loading history questionnaire. Med Sci Sports Exerc 2006; 38 (6): 1121-31
87. Eagan MS, Lyle RM, George PM, et al. A new self-reported comprehensive historical activity questionnaire for young women. J Phys Act Health 2005; 2 (1): 35
88. Suleiman S, Nelson M. Validation in London of a physical activity questionnaire for use in a study of postmenopausal osteopaenia. J Epidemiol Community Health 1997; 51 (4): 365-72
89. Ainsworth BE, Richardson MT, Jacobs Jr DR, et al. Accuracy of recall of occupational physical activity by questionnaire. J Clin Epidemiol 1999; 52 (3): 219-27
90. Rundle A, Hagins M, Orjuela M, et al. Traditional physical activity indexes derived from the Harvard Alumni Activity Survey have low construct validity in a lower income, urban population. Urban Health 2008; 84 (5): 722-32
91. Fogelholm M, Malmberg J, Suni J, et al. International Physical Activity Questionnaire: validity against fitness. Med Sci Sports Exerc 2006; 38 (4): 753-60
92. Hagstromer M, Oja P, Sjostrom M. The International Physical Activity Questionnaire (IPAQ): a study of concurrent and construct validity. Public Health Nutr 2006; 9 (6): 755-62
93. Dinger MK, Behrens TK, Han JL. Validity and reliability of the International Physical Activity Questionnaire in college students. Am J Health Edu 2006; 37 (6): 337-43
94. Ishikawa-Takata K, Tabata I, Sasaki S, et al. Physical activity level in healthy free-living Japanese estimated by doubly labelled water method and International Physical Activity Questionnaire. Eur J Clin Nutr 2008 Jul; 62 (7): 885-91
95. Wolin KY, Heil DP, Askew S, et al. Validation of the international physical activity questionnaire-short among blacks. J Phys Act Health 2008; 5 (5): 746-60
96. MacFarlane DJ, Lee CCY, Ho EYK, et al. Reliability and validity of the Chinese version of IPAQ (short, last 7 days). J Sci Med Sport 2007; 10 (1): 45-51
97. Vandelanotte C, de Bourdeaudhuij I, Philippaerts R, et al. Reliability and validity of a computerized and Dutch version of the International Physical Activity Questionnaire (IPAQ). J Phys Act Health 2005; 2 (1): 63
98. Rosenberg DE, Bull FC, Marshall AL, et al. Assessment of sedentary behavior with the International Physical Activity Questionnaire. J Phys Act Health 2008; 5 Suppl. 1: S30-44
99. Slattery ML, Jacobs Jr DR. The inter-relationships of physical activity, physical fitness, and body measurements. Med Sci Sports Exerc 1987; 19 (6): 564-9
100. Leicht A. Validation of a one-day self-report questionnaire for physical activity assessment in healthy adults. Eur J Sport Sci 2008; 8 (6): 389-97
101. Kwak L, Kremers SPJ, van Baak MA, et al. Measuring physical activity in field studies: comparison of a questionnaire, 24-hour recall and an accelerometer. Eur J Sport Sciences 2007; 7 (4): 193-201
102. Chasan-Taber L, Erickson JB, Nasca PC, et al. Validity and reproducibility of a physical activity questionnaire in women. Med Sci Sports Exerc 2002; 34 (6): 987-92
103. Hopkins WG, Wilson NC, Worsley FA, et al. Reliability of the core questionnaire in the life in New Zealand Survey. NZ J Health Phys Edu Rec 1991; 24 (3): 21-2
104. Lof M, Hannestad U, Forsum E. Comparison of commonly used procedures, including the doubly-labelled water technique, in the estimation of total energy expenditure of women with special reference to the significance of body fatness. Br J Nutr 2003; 90 (5): 961-8
105. Conway JM, Seale JL, Jacobs Jr DR, et al. Comparison of energy expenditure estimates from doubly labeled water, a physical activity questionnaire, and physical activity records. Am J Clin Nutr 2002; 75 (3): 519-25
106. Terwee CB, Dekker FW, Wiersinga WM, et al. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Qual Life Res 2003; 12 (4): 349-62
107. Kurtze N, Rangul V, Hustvedt BE. Reliability and validity of the international physical activity questionnaire in the Nord-Trondelag health study (HUNT) population of men. BMC Med Res Methodol 2008 Oct 9; 8: 63

Correspondence: Dr *Mireille N.M. van Poppel*, Department of Public and Occupational Health, EMGO Institute for Health and Care Research, VU University Medical Center, Van der Boechorststraat 7, 1081 BT Amsterdam, the Netherlands.
E-mail: mnm.vanpoppel@vumc.nl