



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/137046/>

Version: Accepted Version

Article:

Smith, RD, Dziedzic, KS, Quicke, JG et al. (2019) Identification and evaluation of self-reported physical activity instruments in adults with osteoarthritis: A systematic review. *Arthritis Care and Research*, 71 (2). pp. 237-251. ISSN: 2151-464X

<https://doi.org/10.1002/acr.23787>

This article is protected by copyright. This is the peer reviewed version of the following article: Identification and evaluation of self-reported physical activity instruments in adults with osteoarthritis: A systematic review, Smith, R. D., Dziedzic, K. S., Quicke, J. G., Holden, M. A., McHugh, G. A. and Healey, E. L. (2018), Identification and evaluation of self-reported physical activity instruments in adults with osteoarthritis: A systematic review. *Arthritis Care Res.* Accepted Author Manuscript, which has been published in final form at <https://doi.org/10.1002/acr.23787>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Title: Identification and evaluation of self-reported physical activity instruments in adults with osteoarthritis: A systematic review.

Running head: Measuring physical activity in osteoarthritis

Authors: Robert D. Smith, PhD¹;

Krysia S. Dziedzic, PhD²;

Jonathan G. Quicke, PhD²;

Melanie A. Holden, PhD²;

Gretl A. McHugh, PhD, MSc Public Health, B.Nurs (Hons) ³;

Emma L. Healey, PhD².

Affiliations:

¹ School of Nursing, The University of Hong Kong, Hong Kong, China. ²Research Institute for Primary Care and Health Sciences, Keele University, Keele, Staffordshire, United Kingdom, ST5 5BG. ³ School of Healthcare, University of Leeds, Leeds, LS2 9JT.

Funding: This work was funded by a Keele University ACORN studentship and supported by a National Institute for Health Research (NIHR) Programme Grant (RP-PG-0407-10386). This research was also funded by the Arthritis Research UK Centre in Primary Care grant (Grant Number 18139).

Corresponding Author: Robert Smith

Email: robsmith@hku.hk

Fax: +852 2872 6079

Abstract word count: 250

Manuscript word Count: 3776/3800

Tables/Figures: 6/6

1 **Abstract**

2 *Objective:* To identify and evaluate the measurement properties of self-report physical
3 activity (PA) instruments suitable for those with osteoarthritis (OA).

4 *Methods:* A comprehensive two-stage systematic review using multiple electronic
5 databases from inception until July 2018. Stage One sought to identify all self-reported
6 PA instruments used in populations with joint pain attributable to OA in the foot, knee,
7 hip or hand. Stage Two searched for and appraised studies investigating the
8 measurement properties of the instruments identified. For both stages all articles were
9 screened for study eligibility criteria, completed data extraction using the Qualitative
10 Attributes and Measurement Properties of Physical Activity Questionnaires (QAPAQ)
11 checklist, and conducted methodology quality assessments using a modified
12 Consensus-based Standards for the selection of health Measurement Instruments
13 (COSMIN) checklist. Measurement properties for each physical activity instrument
14 were evaluated and combined using narrative synthesis.

15 *Results:* Stage One identified 23 unique self-report PA instruments. Stage Two
16 identified 53 studies that evaluated the measurement properties of 13 of the 23
17 instruments identified. Instrument reliability varied from inadequate to adequate
18 ($ICC \geq 0.7$). Instrument construct and criterion validity assessment demonstrated
19 small to moderate correlations with direct measures of PA. Responsiveness was
20 assessed in only 1 instrument and was unable to detect changes in comparison to
21 accelerometers.

22 *Conclusion:* While many instruments were identified as potentially suitable for use in
23 individuals with OA, none demonstrated adequate measurement properties across all
24 domains of reliability, validity and responsiveness. Further high-quality assessment of

1 self-reported PA instruments is required before such measures can be recommended
2 for use in OA research.

3

4 **Significance and innovation:**

- 5 • Physical activity (PA) is a recommended core treatment for osteoarthritis (OA)
6 and is a commonly used outcome in clinical trials, therefore accurately
7 measuring current PA levels and changes in PA in individuals with OA is vital.
- 8 • This systematic review updates and builds on a previous systematic review
9 examining the measurement properties of PA instruments suitable for adults
10 with OA, collecting evidence from 53 studies.
- 11 • This study highlights the need for high-quality assessment (following COSMIN
12 guidelines) across all measurement properties of self-reported PA instruments
13 before such measures can be recommended for use in OA research.

14

15

16

17

18

19

20

21

1 **Introduction**

2 Osteoarthritis (OA) is a clinical syndrome of joint pain with varying degrees of limitation
3 in physical function and reduced quality of life and most commonly affects the knee,
4 hip, hand and foot (1). Physical activity (PA), such as therapeutic strengthening
5 exercises or aerobic exercise, can reduce joint pain symptoms and improve physical
6 function. PA is recommended as a core treatment for people with OA in the foot, knee,
7 hip or hand (2, 3). However, pain is an important predictor of physical inactivity (4) and
8 less than half of adults with OA are meeting the current guideline of 150 minutes of
9 moderate intensity PA per week (5, 6). Accurately measuring current PA levels and
10 changes of PA in individuals with OA is important in research.

11 PA can be measured using direct methods such as accelerometry or indirect methods
12 such as self-reported PA instruments (7). Self-reported PA instruments are a popular
13 approach for measuring levels of PA in larger population studies (8). This is due to
14 their ease of use, their ability to allow immediate access to information about an
15 individual's PA, and the low cost involved in their administration to a large number of
16 study participants (9). To accurately measure PA using self-report instruments, the
17 appropriate instrument must be selected according to the demographics of the
18 participants (10). An example are instruments developed specifically to measure PA
19 for adults age 65 years and over (11).

20 Multi domain instruments such as the Western Ontario and McMaster Universities
21 Osteoarthritis Index (WOMAC) and the Knee injury and Osteoarthritis Outcome Score
22 (KOOS), have been designed specifically for use in populations with OA. While these
23 multi domain instruments do measure PA as a component or sub-scale score, they
24 have been excluded from this review as their purpose is not to assess PA levels

1 explicitly in terms of frequency, duration and intensity, which are required to make
2 comparisons to current PA guidelines.

3 To date there is still no consensus on which self-reported PA instrument is the most
4 suitable for OA research. In 2011, Terwee et al evaluated the measurement properties
5 of PA instruments in OA populations but focused solely on those with a diagnosis of
6 knee or hip OA (12). This previous systematic review identified 9 studies, however
7 none of these included the Physical Activity Scale for the Elderly (PASE) (13), an
8 instrument that has more recently been used in OA research (14-16). Other systematic
9 reviews that have evaluated the measurement properties of PA instruments for adults
10 in non-joint pain populations restricted to adults aged between 18-65 years or adults
11 aged 65 years or over (7, 8, 11). Therefore, there is a gap in the literature for a
12 comprehensive, broader and updated systematic review that captures relevant
13 information regarding the measurement of PA in those with OA, a group that are most
14 commonly aged 45 years and over. Rather than just focusing on those with a diagnosis
15 of OA, by including studies that have evaluated the measurement properties of
16 relevant instruments in other populations (i.e. 1. those with joint pain attributable to OA
17 in the foot, knee, hip or hand and 2. community dwelling adults in the same age bracket
18 as those with OA), it will be possible to identify and evaluate the measurement
19 properties of a range of instruments suitable for those with OA. To our knowledge, no
20 instrument measuring PA levels has been specifically developed for populations with
21 OA. Instruments developed for other populations, such as general adult or elderly adult
22 populations, have been used in OA research. It is, therefore, important to understand
23 how well these instruments reflect the construct of PA levels in OA populations by
24 assessing the instruments' measurement properties as defined in the COSMIN
25 taxonomy (17).

1 A two-stage systematic review was conducted and aimed to identify and evaluate the
2 measurement properties of self-report physical activity (PA) instruments suitable for
3 those with OA.

4

5 **Patients and Methods**

6 Stage One identified all self-report PA instruments used in published research
7 involving populations aged 45 years and over with joint pain attributable to OA in the
8 feet, knee, hips or hands. The age range and joint sites were selected following the
9 National Institute for Health and Care Excellence guideline on the management of OA
10 and the most commons peripheral joints affected by OA (1). Stage Two subsequently
11 identified all the published evidence on the measurement properties of the instruments
12 identified in Stage One. Both stages of the systematic review involved electronic
13 database searching of MEDLINE, EMBASE and Web of Science from inception until
14 19th July 2018 combined with hand searching of reference lists from included articles.
15 The primary reviewer (RS) screened all titles and the abstracts, full articles were
16 independently double reviewed by the primary reviewer and at least one of the
17 secondary reviewing team (MH, JQ, EH, GM, KD), with any disagreements resolved
18 via consensus discussion between reviewers. Titles and abstracts were reviewed by
19 the primary reviewer only due to time limitations of the secondary reviewers, to
20 minimise risk of reviewer error, 10% of all titles and abstracts were independently
21 reviewed with at least one of the secondary review team.

22

23

1 **Stage One**

2 **Selection Criteria**

3 The selection criteria for Stage One were quantitative research studies that focused
4 on populations with joint pain attributable to OA in the foot, knee, hip or hand and
5 measured self-reported PA (Table 1). Populations were included if other sites of pain
6 were present alongside pain in the foot, knee, hip or hand. Due to cases where study
7 sample include both OA and inflammatory arthritis populations, we only include those
8 with more than 50% of the sample having OA or joint pain attributable to OA. Search
9 terms for articles in Stage One were synthesised from previous joint pain and PA
10 systematic review search strategies (18, 19). The full search strategy for Stage One
11 is shown in appendix 1.

12 ***add Table 1 here***

13 **Data extraction**

14 Data extraction for Stage One involved extracting the citation of the included studies
15 and identifying the self-reported PA instrument used. Data extraction was conducted
16 by two different reviewers independently (the primary reviewer and one of the
17 secondary reviewers). As the aim of Stage One was simply to identify studies and
18 instruments no further data extraction or quality assessment was conducted.

19 **Stage Two**

20 **Selection Criteria**

21 The selection criteria for Stage Two were studies that performed an evaluation of the
22 at least one measurement property of the instruments identified in Stage One in

1 populations with joint pain attributable to OA, or community dwelling adults of a similar
2 age (aged 45 years and over). For purposes of describing all instruments included in
3 Stage Two, articles that described the instruments attributes (the settings, recall
4 period, purpose) were also retrieved. The search strategy for Stage Two was
5 constructed using a high sensitivity search term filter for identifying articles on
6 measurement tool properties (20). This filter was combined with the name of the
7 instrument identified in Stage One of this review. The full search strategy for Stage
8 Two is shown in appendix 2.

9 ***Data extraction and quality assessment***

10 In Stage Two, the Quality Assessment of Physical Activity Questionnaires checklist
11 (QAPAQ) was used to extract data and conduct a preliminary quality assessment (21).
12 The QAPAQ is a comprehensive checklist of all the measurement properties and
13 qualitative attributes of self-report PA instruments and has been used in previous
14 systematic reviews evaluating measurement properties of self-report PA (7, 11, 12). A
15 comprehensive quality assessment of the articles identified in Stage Two was
16 conducted using the COnsensus-based Standards for the selection of health
17 Measurement Instruments (COSMIN) checklist (22). The COSMIN checklist has been
18 used in previous systematic reviews that have assessed the quality of other self-
19 reported instruments (23-26). To reduce reviewer burden within this systematic review,
20 the COSMIN was modified by removing items on generalisability and interpretability
21 already covered in the QAPAQ (21).

22 Following quality assessment, a previously used grading system was conducted to
23 assign a quantitative score to the evidence of each instrument's measurement
24 properties and the quality of that evidence (23-25). The grading system combined the

1 strength of evidence (using the COSMIN checklist) (Appendix 3) to a criteria for each
2 measurement property (10) (Appendix 4), which was extracted using the QAPAQ (21).
3 For the purposes of this systematic review construct validity was defined in terms
4 convergent construct validity in which the self-reported instrument reflects PA
5 measured objectively, such as accelerometers or heart rate monitoring. In criterion
6 validity the gold standard measurement for PA in the review was considered as
7 double-labelled water (DLW). Measurement error was not formally assessed as a
8 COSMIN criterion as we could not identify a minimal important change reported for
9 any of the instruments, measurement error has been reported when evaluated by
10 studies.

11

12 **Results**

13 ***Stage One***

14 From the search of the electronic databases and hand searching of reference lists of
15 included studies, 20,292 articles were identified which reduced to 20,116 following
16 removal of duplicates. Ninety-one studies comprising 23 unique self-reported PA
17 instruments met the inclusion criteria and were included in the review. This is indicated
18 by a PRISMA flowchart (Figure 1). Included studies focused on knee OA (n=52), knee
19 and/or on hip OA (n=22), hip OA (n=8), general joint pain or multiple sites of OA (n=4)
20 foot pain or foot OA (n=3) and knee pain (n=2) populations. Thirty-two of the studies
21 were longitudinal cohort studies, 29 were randomized controlled trials, 18 were cross-
22 sectional studies, 9 studies examined the measurement properties of instruments and
23 3 were systematic reviews. Seventeen studies were conducted in the United States
24 (USA), 13 in Australia and the United Kingdom (UK), 12 in the Netherlands, 5 in

1 Canada and Germany, 4 in Switzerland and Denmark, 3 in Sweden, Brazil and
2 Portugal, and Norway each and 1 in Greece, Spain, Japan, and Iran, two studies were
3 multi-country studies across Europe.

4 ***add Figure 1 here***

5 ***PA instruments identified***

6 The self-reported instruments of PA (n=23) used in the included studies identified in
7 Stage One are listed in appendix 3. The most common PA instruments used were the
8 Physical Activity Scale for the Elderly (PASE) (used in 34 studies), and the
9 International Physical Activity Questionnaire- Short Form (IPAQ-SF) (used in 17
10 studies). Nineteen of the instruments identified were multi-item self-reported PA
11 questionnaires and 5 were single item PA instruments.

12 ***Stage Two***

13 Within Stage Two of the systematic review, 3,661 articles were identified, with 54
14 meeting the inclusion criteria (Figure 2). Of those, nine (16%) evaluated the
15 measurement properties of one or more of the identified PA instruments in adults with
16 joint pain attributable to OA (knee =3; hip =3; combined hip and knee =3).
17 Forty-five articles (84%) evaluated the measurement properties of the PA instruments
18 in community dwelling adult populations aged 45 and over (adults aged 65 years and
19 over = 20; aged 45-64 years = 25). The majority of studies were conducted in Australia
20 (n=9), USA (n=8), the Netherlands (n=5), Japan (n=4) and China (n=4). Thirty-five
21 studies evaluated construct validity, 36 evaluated reliability or measurement error, two
22 studies examined content validity, two examined criterion validity, two evaluated

1 internal consistency and one evaluated responsiveness. A summary of the
2 characteristics of the articles included in Stage Two have been included (Appendix 6).

3 ***add Figure 2 here***

4 Of the 23 instruments identified in Stage One, 13 (56.5%) had a least one
5 measurement property evaluated in either a population with joint pain attributable to
6 OA or a community dwelling adult population aged 45 years and over. Table 2
7 describes the characteristics of these instruments.

8 ***add Table 2 here***

9 ***Measurement properties of the PA instruments in populations with joint pain***
10 ***attributable to OA***

11 There were no instruments identified in Stage One and evaluated in Stage Two which
12 demonstrated full adequacy across all measurement property domains in populations
13 with joint pain attributable to OA (Table 3). Criterion validity, internal consistency,
14 content validity, structural validity and responsiveness were not assessed in any of the
15 instruments. There was no evidence of any measurement properties for the Active
16 Australia Survey (AAS), modified Baecke, Incidental And Planned Activity
17 Questionnaire For Older People (IPEQ), Short Questionnaire To Assess Health
18 Enhancing Physical Activity (SQUASH), Short Telephone Activity Recall
19 Questionnaire (STAR) or Zutphen Physical Activity Questionnaire in populations with
20 joint pain attributable to OA.

21 In terms of reliability, the only multi-item instruments with correlations or ICC above
22 0.7 in studies deemed to be of good-to-excellence methodological quality were the
23 Beacke, Human Activity Profile (HAP), IPAQ-SF and PASE in populations with joint

1 pain attributable to OA (27-30). While the quality evidence rated as fair, all the single
2 scale instruments (Activity Rating Scale (ARS), Tegner scale and University Of
3 California, Los Angeles Activity (UCLAA) scale) demonstrated correlations above 0.7
4 in populations with joint pain attributable to OA for reliability (29). The measurement
5 error of HAP, IPAQ-SF and PASE has been evaluated, while there is no minimally
6 important change index to assess the adequacy of measurement error in these
7 instruments. The proportion of error in IPAQ-SF and PASE were large compared to
8 their maximal possible scoring range, while the HAP was small. Suggesting large
9 measurement error in populations with joint pain attributable to OA in the IPAQ-SF
10 and PASE (28, 30-33) (Table 3).

11 For construct validity in populations with joint pain attributable to OA, the Baecke,
12 IPAQ-SF and PASE demonstrated only low to moderate correlations (0.06-0.49) with
13 accelerometers (30-33) (Table 3).

14 ***Measurement properties of the PA instruments in community dwelling adult***
15 ***aged 45 and over***

16 There were no instruments identified in Stage One and evaluated in Stage Two
17 which demonstrated full adequacy across all measurement property domains in
18 community dwelling adult aged 45 and over (Table 3). Structural validity was not
19 assessed in any of the instruments (Table 4.).

20 In terms of reliability, the AAS displayed adequate reliability in one study (34) but
21 inadequate reliability in two studies (35, 36). The modified Baecke demonstrated
22 reliability in three studies above and below adequate reliability(37-39). The HAP,
23 IPEQ and STAR demonstrated adequate reliability in three studies (40-42). The
24 IPAQ-SF in 7 studies (43-50), and the PASE in 8 studies both demonstrated

1 reliability above and below adequate reliability (13, 51-56). Measurement error had
2 been assessed in the PASE in one study; finding a relatively small standard error
3 measurement (SEM) (3.3-8.5) to the maximal scoring range of the PASE (0-400)
4 (56).

5 The PASE and modified Baecke were the only instruments to have criterion validity
6 evaluated and this was in community dwelling older adults aged 45 and over. Both
7 demonstrated a moderate correlation to DLW, in another study the PASE also
8 demonstrated a non-significant correlation to DLW (51, 57, 58).

9 For construct validity, the AAS correlation with accelerometers was assessed in 5
10 studies and ranged from 0.39-0.61, all demonstrating some moderate correlations
11 (34, 36, 43, 59, 60). The Modified Baecke demonstrated non-significance in a
12 correlation with heart rate monitoring (37). The HAP showed moderate correlations
13 to accelerometers in a single study(40). IPEQ showed a low correlation to
14 accelerometers in a single study (61). The IPAQ-SF was evaluated for construct
15 validity in 9 studies, correlations to accelerometers ranged from non-significant to
16 moderate correlations(44, 46-49, 62-65). The PASE was evaluated for construct
17 validity in 5 studies, correlations to accelerometers ranged from low to moderate
18 correlations (51-53, 66, 67). The SQUASH demonstrated high agreement with heart
19 monitoring in a single study (68). The STAR demonstrated low correlations with
20 accelerometers in a single study (42). The Zutphen demonstrated moderate
21 correlations with accelerometers (69).

22 The IPAQ-SF and PASE were evaluated for internal consistency, each in a single
23 study. In both the IPAQ-SF and PASE internal consistency was deemed adequate.
24 The AAS and IPAQ-SF were assessed for their content validity by cognitive

1 interviews about the understanding of the items in the instrument (50, 56). In both the
2 AAS and IPAQ-SF terminology used in items were confusing or unclear to
3 participants, making recall difficult (70, 71). Responsiveness was evaluated in the
4 IPEQ and was evaluated to be less responsive to changes in PA levels compared to
5 accelerometer (61).

6 ***add Table 3 here***

7 ***add Table 4 here***

8 ***Methodological quality of the included studies***

9 For reliability, eight studies were evaluated as poor quality as a small sample size was
10 used ($n < 50$) (29, 31-33, 38, 39, 46, 72), sample sizes below 50 are considered too
11 small for evaluating measurement properties (10). Five studies that assessed reliability
12 were evaluated as fair quality as their sample size was above 50, but they used a
13 correlation rather than test for agreement (intra-class correlation) (35-37, 43, 54).
14 Fourteen studies were evaluated as good quality with sample sizes larger than 50 but
15 smaller than 100 (10), and seven studies were evaluated as excellent quality with
16 sample sizes greater than 100 (10). One good quality study evaluated measurement
17 error in a sample size < 100 (56).

18 The two studies that evaluated criterion validity were evaluated as poor quality due to
19 their sample size (57, 58). Of the studies evaluating construct validity: seven were
20 evaluated as poor quality due to sample size (31-33, 38, 39, 46, 60, 67); three were
21 evaluated as fair quality (45, 59, 68), as while the sample size was deemed
22 appropriate, these studies used pedometers or heart monitors rather than
23 accelerometers; twelve studies were evaluated as good quality with sample sizes

1 larger than 50 but small than 100 (27, 30, 34, 36, 42, 48, 49, 51, 52, 66, 69, 73); and
2 10 studies were evaluated as excellent quality with sample sizes greater than 100 (35,
3 40, 43, 47, 53, 61-64, 74), only one of the studies in this review used hypothesis testing
4 to evaluate construct validity(49). Responsiveness was assessed in one study, which
5 was evaluated as excellent quality due to a large sample size above 100 participants
6 and a comparison with an accelerometer. Two studies of excellent quality assessed
7 content validity using cognitive interviews (70, 71).

8

9 **Discussion**

10 Stage One of this systematic review identified 23 self-reported PA instruments that
11 have been used previously in populations with joint pain attributable to OA. However,
12 based on the findings from Stage Two of this systematic review, it is still not clear
13 which instrument is most appropriate for use in those with OA. This is due to the lack
14 of evidence of adequate measurement properties for all the instruments identified.

15 *Reliability and internal consistency*

16 In both populations, most self-report instruments demonstrated adequate test-retest
17 reliability. Although methodological quality ranged from poor to excellent. This
18 suggests that these self-report instruments are reliable in measuring levels of PA in
19 test re-test evaluations. Two studies evaluated internal consistency, one in the IPAQ-
20 SF and one in the PASE, both were of good methodological quality and indicated
21 adequate consistency of all the items (Cronbach's $\alpha \geq 0.70$).

22 *Criterion validity and construct validity*

1 None of the instruments demonstrated strong correlations (above 0.70) with direct
2 measures of PA, such as accelerometers or heart monitors, in those with joint pain
3 attributable to OA or community dwelling older adults aged 45 years and over. Two
4 studies evaluated criterion validity using the gold standard measurement of PA (DLW)
5 (57, 58), but these only demonstrated low or not statistically significant correlations
6 and were based on small samples below 50 participants. The implication of low to
7 moderate criterion and construct validity of these instruments is that researchers
8 cannot be certain the degree to which instruments reflect actual PA levels, particularly
9 as there were no clear pattern in the self-report instruments regarding over-or-
10 underestimating PA level compared to direct measures (75).

11 *Content validity*

12 Notably, only two studies evaluated content validity. Both were conducted on
13 community dwelling adult populations aged 45 years and over and examined AAS and
14 IPAQ-SF [15, 34]. These studies highlighted participant misinterpretation of both PA
15 definitions and the questions used within these instruments. Gaining a clearer
16 understanding of the difficulties demonstrated with interpreting definitions of PA and
17 the questions contained within self-report PA instruments more generally would be
18 useful.

19 *Responsiveness*

20 None of the studies examined the responsiveness of the instruments in those with joint
21 pain attributable to OA, and only one study evaluated responsiveness (using the IPEQ)
22 in community dwelling older adults aged 45 years and over. It is therefore unclear how
23 sensitive the self-report PA instruments identified are to detecting changes in PA levels
24 in populations with joint pain attributable to OA. This is a major limitation when

1 evaluating PA interventions aimed at increasing PA levels in these populations (76).
2 None of the studies identified in this review formally addressed structural
3 validity or cross-cultural validity in any of the instruments in any of our populations of
4 interest.

5 The studies that evaluated measurement properties in populations with joint pain
6 attributable to OA identified in this review were limited to only those in the knee and
7 hip. None of the studies in Stage Two included those with joint pain in the foot or hand
8 attributable to OA. This lack of evidence also limits comparisons of the measurement
9 properties between different joints of pain attributable to OA.

10 *Strengths and limitations*

11 This systematic review used a comprehensive search strategy including multiple
12 electronic databases, reference list screening from included studies. It is also original
13 in its inclusion of studies of populations with joint pain attributable to OA and
14 community dwelling older adults aged 45 and over. This study has used the gold
15 standard tool for assessing study quality in outcome measures (22), as well as a
16 previously published standardized form for extracting data on measurement properties
17 of PA instruments (21).

18 Despite identifying many studies in Stage Two (n=54), it is difficult to determine to what
19 degree the findings in community dwelling adults aged 45 years and over are
20 generalisable to similar aged adults with OA or joint pain attributable to OA. The review
21 focused on the most common sites of OA for the review in adults aged 45 and over,
22 where the prevalence of OA is most common (1), the findings of this review may not
23 be generalisable to younger people with post-traumatic OA.

1 *Conclusion*

2 This systematic review has demonstrated that there is limited evidence for the
3 measurement properties of previously used self-report PA instruments in populations
4 with joint pain attributable to OA. Further high methodological quality evaluation of
5 additional measurement properties is required for commonly used instruments for this
6 population. It is particularly recommended that such studies use larger sample sizes
7 of at least 50, or ideally larger than 100 participants (10). Such studies will allow
8 researchers to make appropriately informed decisions when selecting self-reported PA
9 instruments in OA research. While the evidence that was identified demonstrated
10 adequate test re-test reliability in a couple of instruments, overall the evidence on
11 validity and responsiveness was lacking. Investigations into content validity may
12 particularly help researchers to identify areas within self-reported PA instruments that
13 may cause participants to misinterpret the questions and therefore report PA
14 inaccurately. Evaluation of the responsiveness of PA instruments commonly used in
15 randomized controlled trials focused on OA is highly recommended (76), especially if
16 PA is the primary outcome. Future studies should also consider building the evidence
17 base focused on reliability of PA instruments by examining correlations with direct
18 measures of PA in OA populations.

19

20 **Acknowledgements**

21 We would like to thank Jo Jordan for her protocol feedback and systematic review
22 methodological advice.

23 **Author contributions**

1 RS was the overall lead for the work for the systematic review and was involved at all
2 stages of the paper. RS, ELH, GAM and KSD conceived and designed the review. All
3 authors were involved in study searching, quality assessment and data extraction
4 checking and editing drafts of the paper. All authors have approved the final version.

5 **Role of the funding source**

6 The lead author (RS) was funded by a Keele University ACORN studentship. The
7 funder had no involvement in in the design, collection, analysis, writing or publishing
8 of this paper. This work was also supported by a National Institute for Health Research
9 (NIHR) Programme Grant (RP-PG-0407-10386). This research was also funded by
10 the Arthritis Research UK Centre in Primary Care grant (Grant Number 18139). KD
11 and ELH are part-funded by the National Institute for Health Research (NIHR)
12 Collaborations for Leadership in Applied Health Research and Care West Midlands.
13 KD is also part-funded by a Knowledge Mobilisation Research Fellowship (KMRF-
14 2014-03-002) from the NIHR. MAH is supported by the National Institute for Health
15 Research (NIHR) School for Primary Care Research. JQ is funded by a National
16 Institute for Health Research (NIHR) Academic Clinical Lectureship in Physiotherapy,
17 awarded as part of Professor Christian Mallen’s NIHR Research Professorship (NIHR-
18 RP-2014-026). The views expressed in this publication are those of the author(s) and
19 not necessarily those of the NHS, the NIHR, HEE or the Department of Health and
20 Social Care.

21 **Conflict of interest**

22 There is no conflict of interest for any of the authors.

23 **References**

- 1 1. National Institute for Health and Care Excellence. Osteoarthritis: care and
2 management. Clinical guideline [CG177]. 2014.
- 3 2. McAlindon TE, Bannuru RR, Sullivan M, Arden N, Berenbaum F, Bierma-
4 Zeinstra S, et al. OARSI guidelines for the non-surgical management of knee
5 osteoarthritis. *Osteoarthritis and Cartilage*. 2014;22(3):363-88.
- 6 3. Fernandes L, Hagen KB, Bijlsma JW, Andreassen O, Christensen P,
7 Conaghan PG, et al. EULAR recommendations for the non-pharmacological core
8 management of hip and knee osteoarthritis. *Annals of the Rheumatic Diseases*.
9 2013;72(7):1125-35.
- 10 4. Plooij B, Scherder EJ, Eggermont LH. Physical inactivity in aging and
11 dementia: a review of its relationship to pain. *Journal of Clinical Nursing*. 2012;21(21-
12 22):3002-8.
- 13 5. Herbolzheimer F, Schaap LA, Edwards MH, Maggi S, Otero Á, Timmermans
14 EJ, et al. Physical Activity Patterns Among Older Adults With and Without Knee
15 Osteoarthritis in Six European Countries. *Arthritis Care & Research*. 2016;68(2):228-
16 36.
- 17 6. Holden MA, Nicholls EE, Young J, Hay EM, Foster NE. Exercise and physical
18 activity in older adults with knee pain: a mixed methods study. *Rheumatology*.
19 2015;54(3):413-23.
- 20 7. Van Poppel MN, Chinapaw MJ, Mokkink LB, Van Mechelen W, Terwee CB.
21 Physical activity questionnaires for adults. *Sports Medicine*. 2010;40(7):565-600.
- 22 8. Helmerhorst HHJ, Brage S, Warren J, Besson H, Ekelund U. A systematic

- 1 review of reliability and objective criterion-related validity of physical activity
2 questionnaires. *International Journal of Behavioral Nutrition and Physical Activity*.
3 2012;9(1):1.
- 4 9. Warren JM, Ekelund U, Besson H, Mezzani A, Geladas N, Vanhees L.
5 Assessment of physical activity—a review of methodologies with reference to
6 epidemiological research: a report of the exercise physiology section of the
7 European Association of Cardiovascular Prevention and Rehabilitation. *European*
8 *Journal of Cardiovascular Prevention & Rehabilitation*. 2010;17(2):127-39.
- 9 10. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al.
10 Quality criteria were proposed for measurement properties of health status
11 questionnaires. *Journal of Clinical Epidemiology*. 2007;60(1):34-42.
- 12 11. Forsén L, Loland NW, Vuillemin A, Chinapaw MJ, van Poppel MN, Mokkink
13 LB, et al. Self-administered physical activity questionnaires for the elderly. *Sports*
14 *Medicine*. 2010;40(7):601-23.
- 15 12. Terwee C, Bouwmeester W, van Elstrand S, de Vet H, Dekker J. Instruments
16 to assess physical activity in patients with osteoarthritis of the hip or knee: a
17 systematic review of measurement properties. *Osteoarthritis and Cartilage*.
18 2011;19(6):620-33.
- 19 13. Washburn RA, McAuley E, Katula J, Mihalko SL, Boileau RA. The physical
20 activity scale for the elderly (PASE): evidence for validity. *Journal of Clinical*
21 *Epidemiology*. 1999;52(7):643-51.
- 22 14. Dunlop DD, Song J, Semanik PA, Sharma L, Chang RW. Physical activity
23 levels and functional performance in the osteoarthritis initiative: a graded

- 1 relationship. *Arthritis & Rheumatology*. 2011;63(1):127-36.
- 2 15. Foster NE, Healey EL, Holden MA, Nicholls E, Whitehurst DG, Jowett S, et al.
3 A multicentre, pragmatic, parallel group, randomised controlled trial to compare the
4 clinical and cost-effectiveness of three physiotherapy-led exercise interventions for
5 knee osteoarthritis in older adults: the BEEP trial protocol (ISRCTN: 93634563).
6 *BMC Musculoskeletal Disorders*. 2014;15(1):254.
- 7 16. Fransen M, Su S, Harmer A, Blyth FM, Naganathan V, Sambrook P, et al. A
8 longitudinal study of knee pain in older men: Concord Health and Ageing in Men
9 Project. *Age and Ageing*. 2013;43(2):206-12.
- 10 17. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al.
11 The COSMIN study reached international consensus on taxonomy, terminology, and
12 definitions of measurement properties for health-related patient-reported outcomes.
13 *Journal of Clinical Epidemiology*. 2010;63(7):737-45.
- 14 18. Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL.
15 Exercise for osteoarthritis of the knee. *Cochrane Database of Systematic Reviews*.
16 2015(1).
- 17 19. Baker PRA, Francis DP, Soares J, Weightman AL, Foster C. Community wide
18 interventions for increasing physical activity. *Cochrane Database of Systematic*
19 *Reviews*. 2015(1).
- 20 20. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a
21 methodological PubMed search filter for finding studies on measurement properties
22 of measurement instruments. *Quality of Life Research*. 2009;18(8):1115-23.

- 1 21. Terwee CB, Mokkink LB, van Poppel MN, Chinapaw MJ, van Mechelen W, de
2 Vet HC. Qualitative attributes and measurement properties of physical activity
3 questionnaires. *Sports Medicine*. 2010;40(7):525-37.
- 4 22. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al.
5 The COSMIN checklist for assessing the methodological quality of studies on
6 measurement properties of health status measurement instruments: an international
7 Delphi study. *Quality of Life Research*. 2010;19(4):539-49.
- 8 23. Schellingerhout JM, Verhagen AP, Heymans MW, Koes BW, Henrica C,
9 Terwee CB. Measurement properties of disease-specific questionnaires in patients
10 with neck pain: a systematic review. *Quality of Life Research*. 2012;21(4):659-70.
- 11 24. Saether R, Helbostad JL, Riphagen II, Vik T. Clinical tools to assess balance
12 in children and adults with cerebral palsy: a systematic review. *Developmental
13 Medicine & Child Neurology*. 2013;55(11):988-99.
- 14 25. Weldam SW, Schuurmans MJ, Liu R, Lammers J-WJ. Evaluation of Quality of
15 Life instruments for use in COPD care and research: a systematic review.
16 *International Journal of Nursing Studies*. 2013;50(5):688-707.
- 17 26. Gilchrist F, Rodd H, Deery C, Marshman Z. Assessment of the quality of
18 measures of child oral health-related quality of life. *BMC Oral Health*. 2014;14(1):1.
- 19 27. Merom D, Delbaere K, Cumming R, Voukelatos A, Rissel C, Van der Ploeg
20 HP, et al. Incidental and planned exercise questionnaire for seniors: validity and
21 responsiveness. *Medicine & Science in Sports & Exercise*. 2014;46(5):947-54.
- 22 28. Ono R, Hirata S, Yamada M, Nishiyama T, Kurosaka M, Tamura Y. Reliability

- 1 and validity of the Baecke physical activity questionnaire in adult women with hip
2 disorders. *BMC musculoskeletal disorders*. 2007;8(1):1.
- 3 29. Bennell KL, Hinman RS, Crossley KM, Metcalf BR. Is the Human Activity
4 Profile a useful measure in people with knee osteoarthritis? *Journal of Rehabilitation*
5 *Research and Development*. 2004;41(4):621.
- 6 30. Naal FD, Impellizzeri FM, Leunig M. Which is the best activity rating scale for
7 patients undergoing total joint arthroplasty? *Clinical Orthopaedics and Related*
8 *Research*. 2009;467(4):958-65.
- 9 31. Casartelli NC, Bolszak S, Impellizzeri FM, Maffiuletti NA. Reproducibility and
10 Validity of the Physical Activity Scale for the Elderly (PASE) Questionnaire in
11 Patients After Total Hip Arthroplasty. *Physical Therapy*. 2015;95(1):86-94.
- 12 32. Blikman T, Stevens M, Bulstra SK, van den Akker-Scheek I, Reininga IH.
13 Reliability and validity of the Dutch version of the International Physical Activity
14 Questionnaire in patients after total hip arthroplasty or total knee arthroplasty.
15 *Journal of Orthopaedic & Sports Physical Therapy*. 2013;43(9):650-9.
- 16 33. Bolszak S, Casartelli NC, Impellizzeri FM, Maffiuletti NA. Validity and
17 reproducibility of the Physical Activity Scale for the Elderly (PASE) questionnaire for
18 the measurement of the physical activity level in patients after total knee
19 arthroplasty. *BMC Musculoskeletal Disorders*. 2014;15(1):1.
- 20 34. Svege I, Kolle E, Risberg MA. Reliability and validity of the Physical Activity
21 Scale for the Elderly (PASE) in patients with hip osteoarthritis. *BMC Musculoskeletal*
22 *Disorders*. 2012;13(1):1.

- 1 35. Hertogh EM, Monninkhof EM, Schouten EG, Peeters PH, Schuit AJ. Validity
2 of the Modified Baecke Questionnaire: comparison with energy expenditure
3 according to the doubly labeled water method. *International Journal of Behavioral*
4 *Nutrition and Physical Activity*. 2008;5(1):1.
- 5 36. Schuit AJ, Schouten EG, Westerterp KR, Saris WH. Validity of the Physical
6 Activity Scale for the Elderly (PASE): according to energy expenditure assessed by
7 the doubly labeled water method. *Journal of Clinical Epidemiology*. 1997;50(5):541-
8 6.
- 9 37. Colbert LH, Matthews CE, Havighurst TC, Kim K, Schoeller DA. Comparative
10 validity of physical activity measures in older adults. *Medicine and Science in Sports*
11 *and Exercise*. 2011;43(5):867.
- 12 38. Fjeldsoe BS, Winkler EA, Marshall AL, Eakin EG, Reeves MM. Active adults
13 recall their physical activity differently to less active adults: test–retest reliability and
14 validity of a physical activity survey. *Health Promotion Journal of Australia*.
15 2013;24(1):26-31.
- 16 39. Brown WJ, Burton NW, Marshall AL, Miller YD. Reliability and validity of a
17 modified self-administered version of the Active Australia physical activity survey in a
18 sample of mid-age women. *Australian and New Zealand Journal of Public Health*.
19 2008;32(6):535-41.
- 20 40. Pettee GK, McClain JJ, Lee CD, Swan PD, Alvar BA, Mitros MR, et al.
21 Evaluation of physical activity measures used in middle-aged women. *Medicine and*
22 *Science in Sports and Exercise*. 2009;41(7):1403-12.
- 23 41. Pols MA, Peeters P, Kemper HC, Collette H. Repeatability and relative validity

- 1 of two physical activity questionnaires in elderly women. *Medicine and Science in*
2 *Sports and Exercise*. 1996;28(8):1020-5.
- 3 42. Pols MA, Peeters PH, Bueno-de-Mesquita HB, Ocke MC, Wentink CA,
4 Kemper HC, et al. Validity and repeatability of a modified Baecke questionnaire on
5 physical activity. *International Journal of Epidemiology*. 1995;24(2):381-8.
- 6 43. Voorrips LE, Ravelli AC, Petra C, Dongelmans A, Deurenberg P, van
7 Staveren WA. A physical activity questionnaire for the elderly. *Diet and Physical*
8 *Activity as Determinants of Nutritional Status in Elderly Women*. 1991:43.
- 9 44. de Carvalho Bastone A, de Souza Moreira B, Alvarenga Vieira R, Noce
10 Kirkwood R, Domingues Dias JM, Corrêa Dias R. Validation of the human activity
11 profile questionnaire as a measure of physical activity levels in older community-
12 dwelling women. *Journal of Aging & Physical Activity*. 2014;22(3).
- 13 45. Delbaere K, Hauer K, Lord SR. Evaluation of the incidental and planned
14 activity questionnaire for older people. *British Journal of Sports Medicine*.
15 2010;44(14):1029-34.
- 16 46. Matthews CE, Ainsworth BE, Hanby C, Pate RR, Addy C, Freedson PS, et al.
17 Development and testing of a short physical activity recall questionnaire. *Medicine*
18 *and Science in Sports and Exercise*. 2005;37(6):986-94.
- 19 47. Brown W, Trost S, Bauman A, Mummery K, Owen N. Test-retest reliability of
20 four physical activity measures used in population surveys. *Journal of Science and*
21 *Medicine in Sport*. 2004;7(2):205-15.
- 22 48. Booth ML, Ainsworth BE, Pratt M, Ekelund U, Yngve A, Sallis JF, et al.

- 1 International physical activity questionnaire: 12-country reliability and validity.
2 *Medicine and Science in Sports and Exercise*. 2003;195(9131/03):3508-1381.
- 3 49. Deng H, Macfarlane D, Thomas G, Lao X, Jiang C, Cheng K, et al. Reliability
4 and validity of the IPAQ-Chinese: the Guangzhou Biobank Cohort study. *Medicine
5 and Science in Sports and Exercise*. 2008;40(2):303.
- 6 50. Mäder U, Martin BW, Schutz Y, Marti B. Validity of four short physical activity
7 questionnaires in middle-aged persons. *Medicine and Science in Sports and
8 Exercise*. 2006;38(7):1255-66.
- 9 51. Tomioka K, Iwamoto J, Saeki K, Okamoto N. Reliability and validity of the
10 International Physical Activity Questionnaire (IPAQ) in elderly adults: the Fujiwara-
11 kyo Study. *Journal of Epidemiology*. 2011;21(6):459-65.
- 12 52. Cerin E, Barnett A, Cheung M-c, Sit CH, Macfarlane DJ, Chan W-m.
13 Reliability and validity of the IPAQ-L in a sample of Hong Kong urban older adults:
14 does neighborhood of residence matter? *Journal of Aging and Physical Activity*.
15 2012;20(4):402-20.
- 16 53. Ryan DJ, Wullems JA, Stebbings GK, Morse CI, Stewart CE, Onambele-
17 Pearson GL. Reliability and validity of the international physical activity questionnaire
18 compared to calibrated accelerometer cut-off points in the quantification of sedentary
19 behaviour and physical activity in older adults. *PloS one*. 2018;13(4):e0195712.
- 20 54. Ayyat E, Kilinc M, Kirdi N. The Turkish version of the physical activity scale for
21 the elderly (PASE): Its cultural adaptation, validation, and reliability. *Turkish Journal
22 of Medical Sciences*. 2017;47(3):908-15.

- 1 55. Dinger M, Oman F, Taylor E, Vesely S, Able J. Stability and convergent
2 validity of the Physical Activity Scale for the Elderly (PASE). *Journal of Sports*
3 *Medicine and Physical Fitness*. 2004;44(2):186.
- 4 56. Hagiwara A, Ito N, Sawai K, Kazuma K. Validity and reliability of the Physical
5 Activity Scale for the Elderly (PASE) in Japanese elderly people. *Geriatrics &*
6 *Gerontology International*. 2008;8(3):143-51.
- 7 57. Ngai SP, Cheung RT, Lam PL, Chiu JK, Fung EY. Validation and reliability of
8 the Physical Activity Scale for the Elderly in Chinese population. *Journal of*
9 *Rehabilitation Medicine*. 2012;44(5):462-5.
- 10 58. Vaughan K, Miller WC. Validity and reliability of the Chinese translation of the
11 Physical Activity Scale for the Elderly (PASE). *Disability and Rehabilitation*.
12 2013;35(3):191-7.
- 13 59. Alqarni AM, Vennu V, Alshammari SA, Bindawas SM. Cross-cultural
14 adaptation and validation of the Arabic version of the physical activity scale for the
15 elderly among community-dwelling older adults in Saudi Arabia. *Clinical*
16 *Interventions in Aging*. 2018;13:419-27.
- 17 60. Heesch KC, Hill RL, Van Uffelen JG, Brown WJ. Are Active Australia physical
18 activity questions valid for older adults? *Journal of Science and Medicine in Sport*.
19 2011;14(3):233-7.
- 20 61. Freene N, Waddington G, Chesworth W, Davey R, Cochrane T. Validating
21 two self-report physical activity measures in middle-aged adults completing a group
22 exercise or home-based physical activity program. *Journal of Science and Medicine*
23 *in Sport*. 2014;17(6):611-6.

- 1 62. Grimm EK, Swartz AM, Hart T, Miller NE, Strath SJ. Comparison of the IPAQ-
2 Short Form and accelerometry predictions of physical activity in older adults. *Journal*
3 *of Aging and Physical Activity*. 2012;20(1):64-79.
- 4 63. Dyrstad SM, Hansen BH, Holme IM, Anderssen SA. Comparison of self-
5 reported versus accelerometer-measured physical activity. *Medicine and Science in*
6 *Sports and Exercise*. 2014;46(1):99-106.
- 7 64. Steene-Johannessen J, Anderssen SA, van der Ploeg HP, Hendriksen IJ,
8 Donnelly AE, Brage S, et al. Are Self-report Measures Able to Define Individuals as
9 Physically Active or Inactive? *Medicine and Science in Sports and Exercise*.
10 2016;48(2):235-44.
- 11 65. Lipert A, Jegier A. Comparison of Different Physical Activity Measurement
12 Methods in Adults Aged 45 to 64 Years Under Free-Living Conditions. *Clinical*
13 *Journal of Sport Medicine*. 2017;27(4):400-8.
- 14 66. Harada ND, Chiu V, King AC, Stewart AL. An evaluation of three self-report
15 physical activity instruments for older adults. *Medicine and Science in Sports and*
16 *Exercise*. 2001;33(6):962-70.
- 17 67. Washburn R, Ficker J. Physical Activity Scale for the Elderly (PASE): the
18 relationship with activity measured by a portable accelerometer. *Journal of Sports*
19 *Medicine and Physical Fitness*. 1999;39(4):336.
- 20 68. de Hollander EL, Zwart L, de Vries SI, Wendel-Vos W. The SQUASH was a
21 more valid tool than the OBiN for categorizing adults according to the Dutch physical
22 activity and the combined guideline. *Journal of Clinical Epidemiology*. 2012;65(1):73-
23 81.

- 1 69. Harris TJ, Owen CG, Victor CR, Adams R, Ekelund U, Cook DG. A
2 comparison of questionnaire, accelerometer, and pedometer: measures in older
3 people. *Medicine and Science in Sports and Exercise*. 2009;41(7):1392-402.
- 4 70. Heesch KC, Van Uffelen J, Brown WJ. How do older adults respond to active
5 Australia physical activity questions? Lessons from cognitive interviews. *Journal of*
6 *Aging & Physical Activity*. 2014;22(1).
- 7 71. Heesch KC, Van Uffelen JG, Hill RL, Brown WJ. What do IPAQ questions
8 mean to older adults? Lessons from cognitive interviews. *International Journal of*
9 *Behavioral Nutrition and Physical Activity*. 2010;7(1):1.
- 10 72. Bilek LD, Venema DM, Camp KL, Lyden ER, Meza JL. Evaluation of the
11 human activity profile for use with persons with arthritis. *Arthritis Care & Research*.
12 2005;53(5):756-63.
- 13 73. Hurtig-Wennlöf A, Hagströmer M, Olsson L. The International Physical Activity
14 Questionnaire modified for the elderly: aspects of validity and feasibility. *Public*
15 *Health Nutrition*. 2010;13(11):1847-54.
- 16 74. Winkler E, Waters L, Eakin E, Fjeldsoe B, Owen N, Reeves M. Is
17 measurement error altered by participation in a physical activity intervention?
18 *Medicine and Science in Sports and Exercise*. 2013;45(5):1004-11.
- 19 75. Washburn RA, Smith KW, Jette AM, Janney CA. The Physical Activity Scale
20 for the Elderly (PASE): development and evaluation. *Journal of Clinical*
21 *Epidemiology*. 1993;46(2):153-62.
- 22 76. Prince SA, Adamo KB, Hamel ME, Hardt J, Gorber SC, Tremblay M. A

- 1 comparison of direct versus self-report measures for assessing physical activity in
2 adults: a systematic review. *International Journal of Behavioral Nutrition and*
3 *Physical Activity*. 2008;5(1):1.
- 4 77. Quicke JG, Foster NE, Croft PR, Ogollah RO, Holden MA. Change in physical
5 activity level and clinical outcomes in older adults with knee pain: a secondary
6 analysis from a randomised controlled trial. *BMC Musculoskeletal Disorders*.
7 2018;19(1):59.
- 8 78. Armstrong T, Bauman AE, Davies J. Physical activity patterns of Australian
9 adults: results of the 1999 National Physical Activity Survey: Australian Institute of
10 Health and Welfare; 2000.
- 11 79. Baecke JA, Burema J, Frijters J. A short questionnaire for the measurement
12 of habitual physical activity in epidemiological studies. *The American Journal of*
13 *Clinical Nutrition*. 1982;36(5):936-42.
- 14 80. Moore DS, Ellis R, Allen PD, Monroe PA, Cherry KE, O'Neil CE, et al.
15 Construct validation of physical activity surveys in culturally diverse older adults: a
16 comparison of four commonly used questionnaires. *Research Quarterly for Exercise*
17 *and Sport*. 2008;79(1):42-50.

Tables

Table 1: Selection criteria for articles in Stage One.

Inclusion	Exclusion
Age range that includes participants 45 years old or over(1).	
At least 50% of the study participants have OA or joint pain attributable to OA in the foot, knee, hip and hand(1).	Over 50% of the study participants with inflammatory arthritis.
Measurement instrument of PA using a reproducible self-reported questionnaire.	A measure of physical fitness rather than a measure of daily PA participation.
Self-reported PA used as a primary or secondary outcome measure.	Direct measures of PA. For example, accelerometers and calorimetry.
All research settings (hospital, primary care, community settings, etc.)	Not written in English.
All quantitative research methodologies (RCTS, cross-sectional, etc.)	Case study research design of a single subject.

Table 2: Characteristics of the PA instruments included in Stage Two.

Instrument and associated study	Construct	Setting	Recall Period	Purpose	Target population	Justification	Format	Interpretability	Ease of use
<i>Multi-item</i>									
Active Australia Survey (AAS) (77)	Leisure time PA	leisure time activities at different intensities	7 Days	To assess knowledge of health benefits for PA in adult populations	Developed for adults aged 18-65, can be used internationally	Offers data on PA that can be implemented into self-report survey or interviewing	9 items, self-report on time spend during activities or frequency	Total score in time spent physically active during a week and time spent sedentary	Short time taken to complete

							of		
							activities		
Baecke (78)	Habitual PA across three domains; work related, leisure time and sport	Activities in: occupatio n, sport and leisure time	Usual week	To assess habitual physical activities for epidemiologica l studies	Young adults	At the time of development, no appropriate instrument was available for use in epidemiologic al studies	16 items, Self-report questionn aire with closed answered questions	Scores are given in three indices; work, sport, leisure time. These scores are not interpretable outside of the Baecke	Small number of multiple choice questions
Modified Baecke (39)	Physical activities in household and leisure	Househol d activities and leisure	One year	Modified to better suite elderly population from the	Elderly adults, aged 65 years and over	Original Baecke not appropriate for elderly populations.	Interviewe r administer ed, not self-report	Time spent PA in hours for one week. Scores can be compared to recommendatio	Interview er required, takes 30 minutes

	sporting activities	sporting activities		original Baecke				ns on PA levels for health benefits	to complete.
Human Activity Profile (HAP) (28)	Energy expenditure or physical fitness	Daily activities	Same day	Originally developed as indicator of quality of life in pulmonary rehabilitation	Clinical and healthy populations	Previously developed instruments had floor and ceiling effects	94 items in a list, each one a daily activity	Scores give average levels of activity and maximal achievable activity	Closed answer questions, time taken to complete: 1-2 minutes
Incidental And Planned Activity Questionnaire For	Incidental and planned physical activities	Gym or home, activities in daily life	7 days or 3 months	Used in longitudinal epidemiology studies to	Frailer populations	Other instruments for adults aged 45 years and over have too	10 items, on planned or structured exercises	Scores are interpretable to time spent physically active	self-complete instrument, quick

Older People (IPEQ) (41)				assess levels of PA		many items for survey use	and activities in daily living		to complete
International Physical Activity Questionnaire (IPAQ-SF & IPAQ-LF) (44)	Energy expenditure in a week. There is a long version and short version	Long version includes; different settings Short version does not separate settings	Two versions; last week and usual week	Research to compare populations in levels of PA	Adults, 18-65 years old. Different languages available	A generic outcome measure of PA to be used in any adult population internationally	Short version: 4 items, Long version: 27 items. Closed questions, some with continuous scale answer	Scores given in energy expenditure per week, scores can be compared to recommendations on PA levels for health benefits	Short version requires minimal time and effort. Long version takes longer and requires recall in

									different aspects of PA
Physical Activity Scale For The Elderly (PASE) (79)	Time spent participating in PA	PA in various settings at work, home and leisure time	Leisure activities, occupational and household activities	Research to assess PA in elderly adults	Elderly adults, aged 65 years and over	None of the generic measures of PA are appropriate for elderly adults	32 items within the six different domains	Scores given as a total score, total score not interpretable in a meaningful way	Questions are easy to fill out with full instructions, short recall period, 32 items is a high number

Short Questionnaire To Assess Health Enhancing Physical Activity (SQUASH) (68)	Habitual activities	Leisure activities, travelling activities, househol d activities, activities at work	Normal week over past few months	A self-report measure with comparable scores to recommendati ons of levels of physical activities for health benefits	All adult populations	Required a measurement where scores were interpretable to quantify weekly PA levels	11 items asking questions on PA in different settings	Scores can be classified for recommended PA levels	Very short, simple to complete
Short Telephone Activity Recall Questionnaire (STAR) (42)	Classificati on of PA in moderate and vigorous levels of PA	All PA	Last 7 days	A telephone administered short instrument to classify individuals in	All adult populations	A need for a quick-to- complete measure of PA over the telephone	3 items, two versions available; open responses and	Responders can be classified into different levels of PA	Very quick to administe r

				different levels			closed		
				of PA			responses		
Zutphen (69)	Daily physical activities	Leisure- time, walking, househol d activities, sporting activities and hobbies.	7 days, although some items differ	Used to assess levels of PA in a longitudinal study	Designed for a study in older male adults, but has been used in male and female adults since	Developed as an appropriate measure of PA over time for a longitudinal study	17 items, open and closed questions	Total score given as energy expenditure	Short with minimal requirem ents for completi on

Single item

Activity Rating Scale (ARS) (29)	Physical activities	All physical activities	Past year	To assess level of PA in one item	Patient with knee disorder	No valid single item measure of PA	1 item: with a 5-point scale	Scoring range from 0-4	Only one item
Tegner Scale (29)	Physical activities	All physical activities	Past week	To assess level of PA in one item	Knee injury	No valid single item measure of PA	1 item: with a 10-level response	Each value on the scale identifies individuals at an interpretable level of PA	Only one item
University Of California, Los Angeles Activity Scale	Physical activities	All physical activities	Past week	To assess level of PA in one item	Joint replacement surgery	No valid single item measure of PA	1 item: with a 10-level response	Each value on the scale identifies individuals at an interpretable level of PA	Only one item

(UCLAA)

(29)

Table 3: Summary of each instruments measurement properties included in Stage Two.

Instrument	Reliability	Measurement error	Criterion validity	Construct validity	Other measurement properties
Populations with joint pain attributable to OA					
<i>Multi-item</i>					
Active Australia Survey (AAS)	0	0	0	0	0
Baecke	ICC=0.87, good quality(27)	0	0	Convergent construct validity, correlation to	0

accelerometer= 0.49,
good quality(27)

Modified Baecke	0	0	0	0	0
Human Activity Profile (HAP)	ICC= 0.95, 0.96, excellent quality(28). ICC=0.60, 0.83, poor quality(72)	SEM=3, excellent quality(28)	0	0	0
Incidental And Planned Activity Questionnaire For Older People (IPEQ)	0	0	0	0	0

International Physical Activity Questionnaire Short Form (IPAQ-SF)	ICC= 0.76, 0.87, Excellent quality(29). ICC=0.5, fair quality(31).	SEM=2487, SDC=1039, fair quality(31).	0	Convergent construct validity, correlation to accelerometer= 0.29, fair quality(31).	0
Physical Activity Scale For The Elderly (PASE)	ICC=0.77, poor quality(33). ICC=0.58, 0.77, poor quality(32). ICC=0.77, fair quality(30).	SEM= 23-35%, SDC= 63- 97%, fair quality(30, 32). SEM= 31, SDC= 87, poor quality(33).	0	Convergent construct validity, correlation to accelerometer=0.3, poor quality(33). correlation to accelerometer=0.06, 0.45, poor quality(32). correlation to accelerometer=0.27, good quality(30)	0
Short Questionnaire	0	0	0	0	0

To Assess

Health

Enhancing

Physical

Activity

(SQUASH)

Short

0

0

0

0

0

Telephone

Activity Recall

Questionnaire

(STAR)

Zutphen

0

0

0

0

0

Single item

Activity Rating Scale (ARS)	Kappa=0.65, 0.88, fair quality(29)	0	0	0	No floor or ceiling effect, fair quality(29)
Tegner	Kappa=0.54, 0.84, fair quality(29)	0	0	0	No floor or ceiling effect, fair quality(29)
University Of California, Los Angeles Activity Scale (UCLAA)	Kappa=0.80, 0.86, fair quality(29)	0	0	0	No floor or ceiling effect, fair quality(29)

Community dwelling adults aged 45 and over

Multi-item

Active Australia Survey (AAS)	Spearman's rank=0.58, 0.64, good quality(35). Spearman's rank=0.32, fair quality(36). Spearman's rank=0.76, fair quality(34)	0	0	Correlation to accelerometer=0.48, 0.52, good quality(35). Correlation to pedometers=0.42, good quality(59). Correlation to accelerometer=0.39, 0.49, good quality(36). Correlation to accelerometer=0.49, 0.56, good quality(60). Correlation to accelerometer=0.45, 0.61, good quality(34).	Wide range of limitations in items in terms of content validity, excellent quality(70)
-------------------------------------	--	---	---	--	--

Baecke	0	0	0	0	0
Modified Baecke	Spearman's rank=0.65, 0.89, fair quality(38). Correlation=0.73, 0.82, poor quality(37). Spearman's rank=0.86, poor quality(39).	0	Correlation with DLW, r=0.54, poor quality(57).	Correlation to heart rate monitoring= NS, poor quality(37). Correlation to PASE, good quality(80).	0
Human Activity Profile (HAP)	ICC=0.79, 0.94, good quality(40)	0	0	Correlation to accelerometer=0.52, 0.55, good quality(40)	0

Incidental And Planned Activity Questionnaire For Older People (IPEQ)	ICC=0.80, 0.84, good quality(41).	0	0	Correlation to accelerometer=0.17, excellent quality(61)	IPEQ responsiveness index=0.31, ActiGraph responsiveness index=0.65, excellent quality(61)
International Physical Activity Questionnaire Short Form (IPAQ-SF)	ICC=0.68, excellent quality(43). Spearman's rank=0.46-0.96, good quality(44). ICC=0.84, excellent quality(45). Spearman's	0	0	Correlation to accelerometer= 0.30- 0.33, good quality(44). Correlation to accelerometer= NS, good quality(62). Correlation to accelerometer= 0.30- 0.33, poor quality(46). Correlation to	Content validity showed that definitions were confusing and recall was difficult, good quality(71). Internal consistency, Cronbach alpha=0.70, good quality(50).

rank=0.54, poor

quality(46).

ICC=0.5, 0.65,

excellent

quality(47).

ICC=0.86, good

quality(48).

Spearman's rank=

0.26, good

quality(49).

ICC=0.99, good

quality(50).

accelerometer= 0.38-

0.56, good quality(47).

Correlation to

accelerometer= 0.39,

good quality(48).

Correlation to

accelerometer= 0.33,

excellent quality(63).

Correlation to

accelerometer= 0.29,

good quality(49).

Correlation to

accelerometer=NS,

excellent quality (64). Sig

difference to

accelerometer, poor

quality(65).

Physical Activity Scale For The Elderly (PASE)	ICC=0.60, good quality(51). ICC=0.60, good quality(52). ICC=0.65, good quality(53). ICC=0.75, good quality(66). ICC=0.68-0.84, good quality(13). ICC=0.81, fair quality(54). ICC=0.79, good quality(55). ICC=0.90-0.98, good quality(56).	SEM= 3.3–8.5, good quality(56).	Correlation with DLW=NS, good quality(51). Correlation with DLW=0.58, poor quality(58).	Correlation to accelerometer= 0.36, good quality(51). Correlation to accelerometer= 0.43, good quality(52). Spearman's rank correlation to accelerometer= 0.16, fair quality(53). Correlation to accelerometer=0.52, 0.59, good quality(66). Correlation to accelerometer= 0.49, poor quality(67).	Internal consistency, Cronbach alpha=0.71-0.75, good quality(56).
--	--	---------------------------------	--	--	---

Short Questionnaire To Assess Health Enhancing Physical Activity (SQUASH)	0	0	0	0	Agreement with heart monitors= 97.6%, fair quality(68).	0
Short Telephone Activity Recall Questionnaire (STAR)	Kappa= 0.57-0.76, excellent quality(42).	0	0	0	Correlation to accelerometer= 0.14- 0.15, good quality(42).	0
Zutphen	0	0	0	0	Correlation to accelerometer= 0.34, good quality(69).	0

Table 4: Grading of each instruments' measurement properties using COSMIN checklist and QAPAQ.

	Reliability and measurement error		Criterion validity		Construct validity using objective measure		Internal consistency		Content validity		Structural validity		Responsiveness	
	Joint pain	Older adults	Joint pain	Older adults	Joint pain	Older adults	Joint pain	Older adults	Joint pain	Older adults	Joint pain	Older adults	Joint pain	Older adults
Active Australia Survey (AAS)	0	++	0	0	0	---	0	0	0	---	0	0	0	0
Activity Rating Scale (ARS)*	+	0	0	0	0	0	0	0	0	0	0	0	0	0
Baecke	++	0	0	0	0	0	0	0	0	0	0	0	0	0
Modified Baecke	0	?	0	?	0	?	0	0	0	0	0	0	0	0
Human Activity Profile (HAP)	+++	0	0	0	0	--	0	0	0	0	0	0	0	0

International Physical Activity Questionnaire Short Form (IPAQ-SF)	+++	±	0	0	?	---	0	+	0	---	0	0	0	0
Incidental And Planned Activity Questionnaire For Older People (IPEQ)	0	++	0	0	0	---	0	0	0	0	0	0	0	+++
Physical Activity Scale For The Elderly (PASE)	++	±	0	--	--	---	0	+	0	0	0	0	0	0
Short Questionnaire To Assess Health Enhancing Physical Activity (SQUASH)	0	0	0	0	0	+	0	0	0	0	0	0	0	0

Short Telephone Activity Recall Questionnaire (STAR)	0	±	0	0	0	--	0	0	0	0	0	0	0	0
Tegner*	+	0	0	0	0	0	0	0	0	0	0	0	0	0
(University Of California, Los Angeles Activity Scale) UCLAA*	+	0	0	0	0	0	0	0	0	0	0	0	0	0
Zutphen	0	0	0	0	0	--	0	0	0	0	0	0	0	0

Key: '?' indicates unclear findings due to study quality; '±' indicates conflicting findings. * indicates single scale items. Strength of the evidence was given based on quality of articles assessed by the COSMIN [6]. 'Joint pain' refers to joint pain attributable to OA. Instruments were given a positive, negative or zero score for the corresponding measurement property based on criteria (10) (Appendix 3 & 4).

Figure 1: PRISMA flowchart of included articles from Stage One.

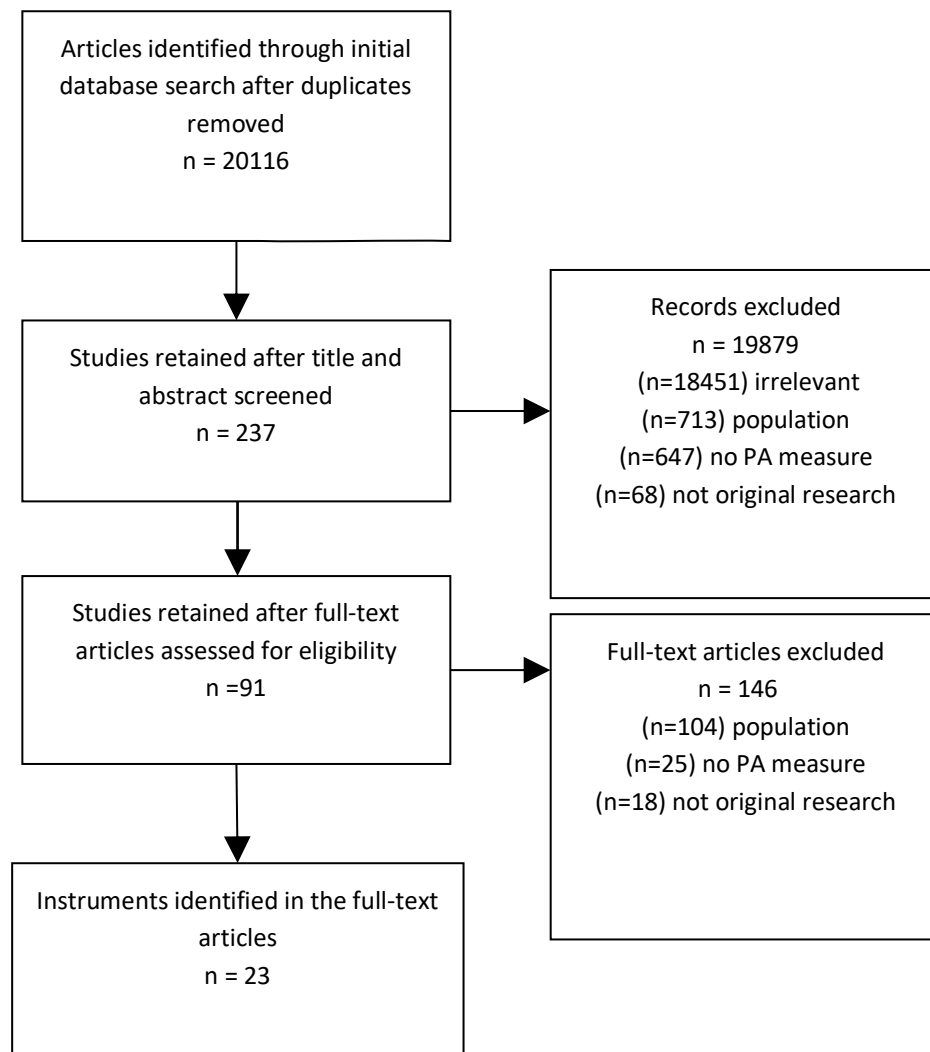


Figure 2: PRISMA Flowchart of included articles from Stage Two.

