Journal of Clinical Epidemiology 114 (2019) 11−21

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Psychometric evaluation supported construct validity, temporal stability, and responsiveness of the Osteoarthritis Questionnaire

Ljoudmila Busija[a,*], Richard H. Osborne[b], Gemma Tatangelo[c], Silvana Niutta[d], Rachelle Buchbinder[e]

[a]*Research Methodology Division, Biostatistics Consulting Platform, Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria 3004, Australia*

[b]*School of Health, Faculty of Health, Arts and Design, Swinburne University, Melbourne, Victoria 3122, Australia*

[c]*School of Psychology, Faculty of Health, Deakin University, Burwood, Victoria 3125, Australia*

[d]*South Australian Cancer Research Biobank, Haematology Directorate, South Australia Pathology, Adelaide, South Australia 5000, Australia*

[e]*Monash Department of Clinical Epidemiology, Cabrini Institute and Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Malvern, Victoria 3144, Australia*

Accepted 3 June 2019; Published online 7 June 2019

Abstract

Objectives: The Osteoarthritis Questionnaire (OA-Quest) is a new measure of individual burden of osteoarthritis, composing seven brief (4−11 items; total 42 items), independent scales. This study describes the psychometric evaluation of the OA-Quest. Specifically, temporal stability, convergent and discriminant validity, and responsiveness of the OA-Quest to joint replacement surgery were examined.

Study Design and Setting: Participants were drawn from rheumatology (n = 62), orthopedic (n = 90), and community (n = 175) settings. Temporal stability of the OA-Quest was evaluated over a 2-week period and was assessed with intraclass correlation coefficients. Responsiveness was assessed with standardized effect sizes and paired samples t-tests. Convergent and discriminant validity was assessed with a multitrait multimethod confirmatory factor analysis (MTMM CFA).

Results: The OA-Quest scales had high temporal stability, with intraclass correlation coefficients between 0.75 (*lost productivity*) and 0.94 (*physical limitations*) and showed improvements of expected magnitude and direction 3 months after joint replacement surgery, supporting their responsiveness. MTMM CFA supported the convergent and discriminant validity of the OA-Quest, demonstrated by adequate model fit (χ^2 = 483.54, df = 184, $P < 0.001$, comparative fit index = 0.95, Tucker-Lewis index = 0.93, root mean square error of approximation = 0.07, standardized root mean square residual = 0.06) and factor loadings of the expected magnitude and direction.

Conclusion: The OA-Quest has strong evidence of temporal stability, construct validity, and is responsive to change following joint replacement surgery.

Keywords: Osteoarthritis; Measurement; Patient-reported outcomes

1. Introduction

Osteoarthritis is a costly and disabling health condition [1]. Its effects on individual range from chronic pain to impaired quality of life [2−4]. To capture the full breadth of the individual burden of osteoarthritis, we developed the Osteoarthritis Questionnaire (OA-Quest) [5]. The OA-Quest compromises seven brief scales (4−11 items; total 42 items), including *physical distress*, *physical limitations*, *psychosocial distress*, *sleep disturbances*, *lost productivity*, *financial hardship*, and *physical deconditioning*. The OA-Quest scales are independent of each other, enabling researchers and clinicians to administer only those that are relevant to their purposes, thus minimizing responder burden.

The development of the OA-Quest was guided by an explicit patient- and clinician-relevant conceptual framework [6] to target issues of most relevance to people with osteoarthritis. The range of concepts measured by the OA-Quest is broader than the currently available

Funding: L.B. was supported by a University of Melbourne Postgraduate Research Scholarship. R.O. was funded in part through a National Health and Medical Research Council of Australia Senior Research Fellowship #APP1059122. R.B. was supported by an NHMRC Senior Principal Research Fellowship #APP1082138.

* Corresponding author. Tel.: +61 3 9903 0644; fax: +61 3 9903 0556.
E-mail address: Lucy.Busija@monash.edu (L. Busija).

https://doi.org/10.1016/j.jclinepi.2019.06.001
0895-4356/

What is new?

Key findings

- The present study demonstrates the psychometric robustness of the patient-centered Osteoarthritis Questionnaire (OA-Quest) scales by presenting the evidence of temporal stability, responsiveness to joint replacement surgery, and construct validity.
- OA-Quest is as psychometrically robust as other measurement tools used in osteoarthritis; however, it is the only questionnaire that measures domains such as financial difficulties and loss of physical fitness, which are of known importance to individuals with osteoarthritis.

What this adds to what was known?

- Patient-reported outcome measures are becoming increasingly important in clinical decision-making.
- OA-Quest is a novel, multidimensional, nonjoint-specific questionnaire developed to comprehensively measure the domains of osteoarthritis burden that are of particular importance to patients and clinicians.

What is the implication and what should change now?

- The breadth of concepts captured by the OA-Quest scales and their psychometric robustness make this questionnaire an expedient addition to measures currently used in osteoarthritis assessment and evaluation.

osteoarthritis-specific measures, such as the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) [7,8] or the Knee injury and Osteoarthritis Outcome Score [9]. The OA-quest is currently the only questionnaire to capture osteoarthritis-related financial burden and loss of fitness [10]. Importantly, the OA-Quest is nonjoint-specific, making it suitable for use in osteoarthritis with multiple joint involvements, which is the most prevalent form of osteoarthritis [11].

Questionnaire validation is an ongoing process of gathering evidence on the meaning and utility of the questionnaire scores in different settings [12,13]. Our initial study provided strong evidence that the OA-Quest has face, content, and factorial validity [5]. However, information on convergent (the extent to which a scale correlates with conceptually related measures) and discriminant validity (the extent to which a scale is uncorrelated with measures of unrelated concepts) is lacking. Similarly, previous work found high estimates of internal consistency of the OA-Quest [5]. Internal consistency provides an indication of how cohesively a set of items measure an underlying concept and is an important prerequisite for construct validity. Temporal stability, the degree to which repeated assessments of stable individuals provide stable results, is another important aspect of reliability and is critical for effective detection of meaningful change. Finally, to support the applications of the OA-Quest in clinical research and practice, it is also necessary to evaluate the extent to which it is responsive to effective interventions, such as joint replacement surgery [14,15].

This study aimed to evaluate the convergent and discriminant validity, temporal stability, and responsiveness of the OA-Quest. To contextualize the interpretation of psychometric properties of the OA-Quest, we also present the corresponding information for the Short-Form Health Survey (SF-36) [16,17] and the WOMAC [7,8]—two established measures frequently used in osteoarthritis research and evaluation [10].

2. Methods

The study sample was designed to represent the full range of osteoarthritis severity and included (1) community-dwelling adults, (2) rheumatology outpatients, and (3) patients undergoing total hip or knee replacement surgery. Data collection took place in 2008 and used postal self-report questionnaires.

The community-dwelling individuals were recruited through Musculoskeletal Australia (MA), who posted questionnaires on one occasion to 500 of their members randomly selected from their database. Eligibility criteria were (1) age 18+ years and (2) self-reported osteoarthritis.

Rheumatology outpatients were recruited from two community-based hospitals in Melbourne (Royal Melbourne Hospital and Cabrini Hospital). Eligibility criteria were (1) 18+ years, (2) recorded diagnosis of osteoarthritis, (3) have consulted a rheumatologist in the previous 2 years, (4) did not have joint replacement surgery in the previous 12 months, and (5) not on a waiting list for joint replacement surgery. Eligibility was established through medical records, with all eligible individuals invited to participate. The participants completed a questionnaire on two occasions, 2 weeks apart.

The joint replacement surgery patients were recruited from an orthopedic clinic at the Repatriation General Hospital, a large community hospital in Adelaide. Eligibility criteria were (1) 18+ years, (2) clinical diagnosis of osteoarthritis, and (3) scheduled to undergo total hip or knee replacement surgery for osteoarthritis. All consecutive eligible patients during the study period were invited to participate. The patients completed a questionnaire immediately before and 3 months postsurgery.

Data from the community-dwelling adults and the baseline data from the rheumatology and orthopedic patients were used to assess convergent and discriminant validity.

Follow-up data from the rheumatology and orthopedic patients were used to assess temporal stability and responsiveness, respectively.

2.1. Ethical approvals

The study was approved by the Human Research Ethics Committees of Cabrini Hospital, the Royal Melbourne Hospital, the University of Melbourne, and the Repatriation General Hospital.

2.2. Measures

2.2.1. Baseline questionnaire

The baseline questionnaire collected demographic and arthritis-related information, including age, gender, employment status, type of arthritis diagnosis, location of joint pain, and severity of joint problems (1 = "none," and 5 = "extremely severe"). In addition to the OA-Quest, SF-36, and WOMAC, the baseline questionnaire included the Center for Epidemiologic Studies Depression (CES-D) Scale, Medical Outcomes Study Sleep Scale (MOSSS), Work Productivity and Activity Impairment Questionnaire General Health Version 2 (WPAI:GH), and single-item measures developed for the present study. These additional measures were selected based on their conceptual concordance with the OA-Quest and were used to assess its convergent and discriminant validity.

The OA-Quest scales (*physical distress*, *physical limitations*, *psychosocial distress*, *sleep disturbances*, *financial hardship*, *lost productivity*, and *physical deconditioning*) and scoring instructions are in Supplementary materials 1. Individuals are asked to indicate how true each statement is of them (0 = "not at all" to 4 = "extremely"). Scale scores are calculated by averaging responses for the corresponding items, with higher scores indicating greater osteoarthritis burden.

The WOMAC has 24 items measuring Pain, Stiffness, and Physical function in hip or knee osteoarthritis (0 = "none," 4 = "extreme") [8]. Higher scores indicate worse outcomes. The WOMAC has evidence of construct validity, reliability, and responsiveness [8,18,19].

The Medical Outcomes Study 36-Item SF-36 Version 2 is a 36-item generic health questionnaire measuring physical functioning, role physical, pain, general health, energy/fatigue, social functioning, role emotional, and emotional well-being [20,21]. Scores range from 0 (worst outcome)

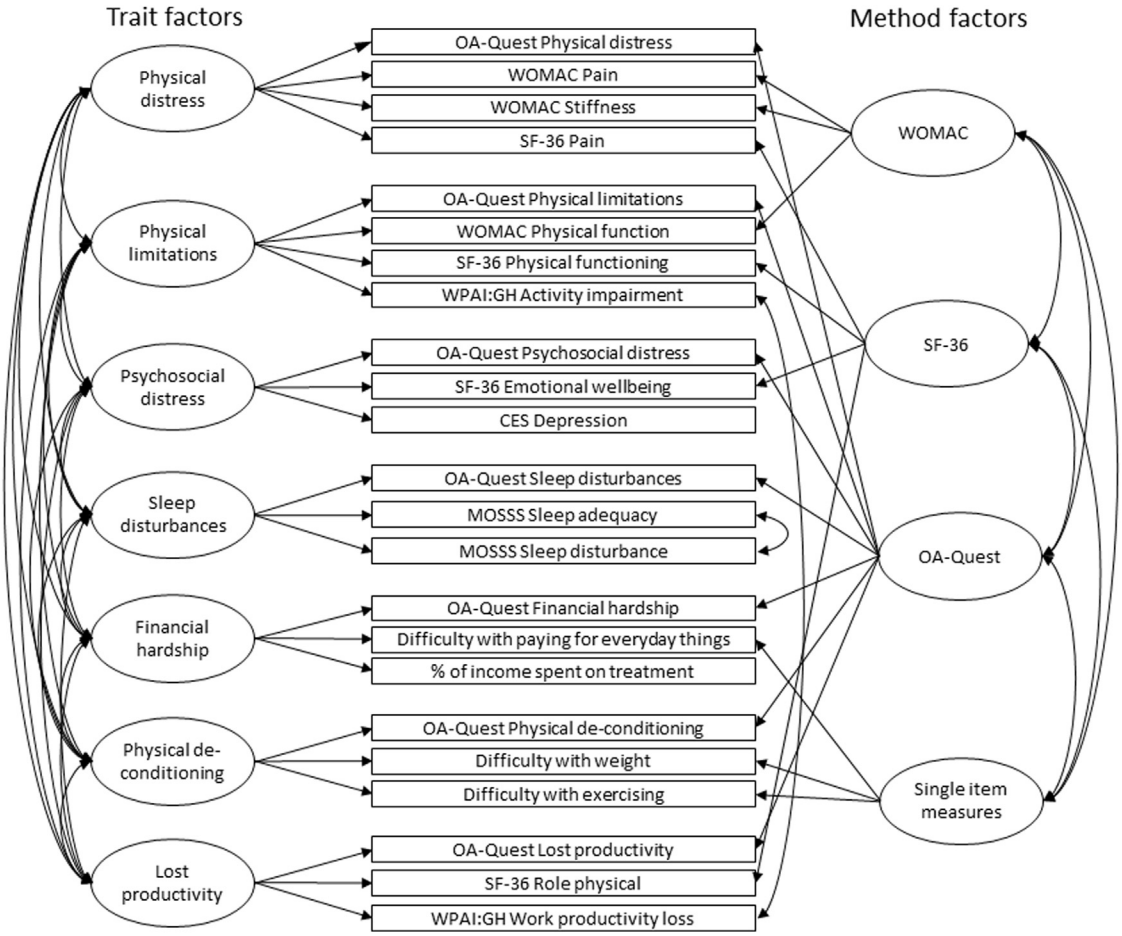

Fig. 1. Hypothesized multitrait multimethod confirmatory factor analysis model for testing convergent and discriminant validity of the OA-Quest.

to 100 (best outcome). Several studies have assessed validity and reliability of the SF-36 in individuals with osteoarthritis [22–24].

CES-D scale measures frequency of depressive symptoms in the general population (20 items; 0 = "rarely or none of the time, less than 1 day", 3 = "most or all of the time, 5 to 7 days") [25]. Higher scores indicate more depression. Previous studies supported reliability and construct validity of the CES-D [25].

The MOSSS [26] measures sleep disturbance (four items; higher scores indicate worse outcome) and sleep adequacy (two items; higher scores indicate better outcome). Items are scored from 1 = "all of the time" to 6 = "none of the time" and converted into 0–100 range. Previous studies supported internal consistency [27] and construct validity [28] of MOSSS.

The WPAI:GH [29] was used to measure daily health-related activity impairment (one item) and work productivity loss (three items; hours worked less productively; hours missed from work, total hours worked). Scores are expressed as percentages, with higher scores indicating worse outcomes. Previous studies supported reliability and construct validity of the WPAI:GH [29].

Single-item measures were developed to target *physical deconditioning* (two items) and *financial hardship* (two items) scales of the OA-Quest, for which no established convergent measures exist [10]. Although single-item measures provide a less reliable assessment of complex concepts than multi-item measures, their use in the assessment of construct validity of multi-item scales is considered acceptable [30]. Three of the four single-item measures captured difficulty in maintaining or achieving healthy weight, exercising, and paying for everyday things (1 = "no difficulty" to 6 = "extreme difficulty"). The fourth measure recorded the percentage of income spent on osteoarthritis treatments (0 = "none," 1 = "1–5%," 2 = "5–10%," 3 = "10–15%," 4 = "15–20%," 5 = "20–25%," 6 ≥ "25%").

2.2.2. Follow-up questionnaire

The follow-up questionnaire comprised the OA-Quest, WOMAC, SF-36, and questions about global changes (compared with baseline) in sleep, emotional well-being, financial situation, and ability to maintain healthy weight, do paid work, perform everyday tasks, and exercise (1 = "a great deal worse," 5 = "a great deal better").

Table 1. Demographic characteristics of study participants

	Group							
	Community (*n* = 175)		Orthopedics (*n* = 90)		Rheumatology (*n* = 62)		Total (*N* = 327)	
Characteristic	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Age, median (interquartile range)	73 (64–79)		68 (61–76)		65 (59–73)		70 (62–78)	
Sex								
Male	31	17.7	41	45.6	19	30.6	91	27.8
Female	144	82.3	49	54.4	43	69.4	236	72.2
Education level								
Primary	18	10.3	10	11.1	9	14.5	37	11.3
Secondary	70	40.0	55	61.1	23	37.1	148	45.3
Tertiary	87	49.7	24	26.7	29	46.8	140	42.8
Missing	0	0.0	1	1.1	1	1.6	2	0.6
Income group								
<$20,000	52	29.7	43	47.8	26	41.9	121	37.0
$20,000–$40,000	65	37.1	25	27.8	11	17.7	101	30.9
$41,000–$60,000	25	14.3	7	7.8	6	9.7	38	11.6
$61,000–$80,000	8	4.6	4	4.4	4	6.5	16	4.9
$81,000–$100,000	5	2.9	3	3.3	7	11.3	15	4.6
>$100,000	7	4.0	4	4.4	6	9.7	17	5.2
Missing	13	7.4	4	4.4	2	3.2	19	5.8
Severity of joint problems								
None	0	0.0	0	0.0	0	0.0	0	0.0
Mild	20	11.4	1	1.1	11	17.7	32	9.8
Moderate	97	55.4	16	17.8	26	41.9	139	42.5
Severe	53	30.3	57	63.3	21	33.9	131	40.1
Extremely severe	3	1.7	15	16.7	3	4.8	21	6.4
Missing	2	1.1	1	1.1	1	1.6	4	1.2

This information was used to describe participants' global perceptions of change in osteoarthritis burden and facilitate interpretation of responsiveness of the OA-Quest scales.

The follow-up questionnaires were posted to rheumatology patients 1 week after the return of their baseline questionnaire and orthopedic patients 2.5 months after their scheduled surgery date. A reminder was posted if the follow-up questionnaire was not returned within 2 weeks.

2.3. Statistical analyses

Descriptive statistics and Cronbach's α were calculated for all scales collected at baseline. Cronbach's $\alpha \geq 0.70$ and ≥ 0.90 was regarded as acceptable for group and individual assessment, respectively [31]. The presence of floor and ceiling effects (>15% of respondents recording the lowest or highest possible score, respectively [15]) was also examined. Temporal stability was assessed with intraclass correlation coefficients (ICC), with values ≥70 considered acceptable

Table 2. Descriptive statistics for the measures used in the study

Measure	Number of items	Score range	% Missing	M (SD)	% Lowest score/ % highest score	Skewness	Cronbach's α
OA-Quest	42						
Physical distress	7	0–4	1.8	1.8 (1.0)	3.7/1.2	0.16	0.92
Physical limitations	11	0–4	0.3	1.3 (0.9)	4.6/0.0	0.64	0.94
Psychosocial distress	7	0–4	2.1	0.9 (1.0)	19.7/0.3	1.28	0.92
Sleep disturbances	5	0–4	0.6	1.5 (1.2)	12.3/3.4	0.46	0.96
Financial hardship	4	0–4	2.1	0.6 (1.0)	57.2/1.3	1.77	0.94
Physical deconditioning	4	0–4	1.5	1.6 (1.1)	10.6/2.5	0.39	0.87
Lost productivity	4	0–4	33.3	0.9 (1.2)	41.3/3.7	1.28	0.92
WOMAC	24						
Pain	5	0–4	2.8	1.7 (0.9)	1.9/0.0	−0.04	0.85
Stiffness	2	0–4	1.2	2.0 (0.9)	3.1/4.0	0.03	0.81
Physical function	17	0–4	1.2	1.8 (0.9)	1.2/0.0	−0.11	0.96
SF-36	36						
Pain	2	0–100	0.3	43.2 (22.4)	4.3/0.6	0.20	0.83
Energy/fatigue	4	0–100	1.2	51.1 (24.6)	2.2/1.2	−0.02	0.81
Physical functioning	10	0–100	0.6	36.4 (23.5)	2.5/0.3	0.54	0.88
Emotional well-being	5	0–100	1.2	63.5 (24.2)	1.2/3.4	−0.61	0.83
Role physical	4	0–100	2.4	56.7 (29.4)	5.0/11.9	−0.25	0.94
General health	5	0–100	2.4	53.9 (21.5)	0.9/0.6	−0.28	0.78
Social functioning	2	0–100	0.3	64.6 (27.8)	2.1/21.8	−0.35	0.86
Role emotional	3	0–100	4.3	73.5 (27.5)	2.6/32.9	−0.87	0.90
CES depression	20	0–60	3.7	15.5 (10.5)	2.5/0.0	0.90	0.89
MOSSS	12						
Sleep adequacy	2	0–100	2.1	48.2 (22.6)	3.4/1.9	−0.03	0.43
Sleep disturbance	4	0–100	3.7	43.9 (24.8)	3.4/1.2	0.15	0.84
WPAI:GH	6						
Activity impairment	1	0–100	2.4	46.7 (26.9)	5.0/1.9	0.06	N/A
Work productivity loss	3	0–100	2.8	4.9 (16.3)	84.4/0.3	3.79	N/A[a]
Single-item measures							
Difficulty with paying for everyday things	1	1–6	4.6	2.15 (1.5)	50.3/4.5	1.16	N/A
% of income spent on treatment	1	0–6	12.8	1.5 (1.1)	9.1/0.4	1.61	N/A
Difficulty with weight	1	1–6	3.1	3.1 (1.71)	24.9/12.3	0.34	N/A
Difficulty with exercise	1	1–6	4.9	3.7 (1.5)	8.0/14.1	−0.13	N/A

SD, standard deviation; CES, Center for Epidemiologic Studies; MOSSS, Medical Outcomes Study Sleep Scale; OA-Quest, Osteoarthritis Questionnaire; SF-36, Short-Form Health Survey; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; WPAI:GH, Work Productivity and Activity Impairment Questionnaire General Health Version 2.

On the SF-36 and MOSS seep adequacy measures, lower scores correspond to worse outcomes. On all other measures, higher scores correspond to worse outcomes.

[a] Cronbach's α is not appropriate because of a nonsummative scoring of the scale.

Table 3. Temporal stability and responsiveness to change of the measures used in the study

	Temporal stability (n = 52)			Responsiveness to joint replacement surgery: Total sample (n = 78)		
	Time 1	Time 2		Time 1	Time 2	
Measure	M (SD)	M (SD)	ICC	M (SD)	M (SD)	ES
OA-Quest						
Physical distress	1.6 (1.1)	1.7 (1.1)	0.92	2.3 (1.0)	1.1 (1.0)	**1.21**
Physical limitations	0.9 (0.9)	1.0 (1.0)	0.94	1.9 (0.8)	1.0 (0.8)	**1.05**
Psychosocial distress	0.8 (0.9)	0.8 (0.9)	0.85	1.3 (1.1)	0.7 (0.9)	**0.58**
Sleep disturbances	1.5 (1.1)	1.5 (1.1)	0.86	2.1 (1.2)	1.0 (1.0)	**0.89**
Financial hardship	0.5 (1.0)	0.6 (1.0)	0.87	0.9 (1.1)	0.6 (0.8)	**0.28**
Physical deconditioning	1.4 (1.2)	1.3 (1.1)	0.88	2.2 (1.0)	1.3 (1.1)	**0.93**
Lost productivity	0.6 (1.0)	0.7 (1.0)	0.75	1.4 (1.3)	1.0 (1.2)	**0.29**
WOMAC						
Pain	1.6 (0.9)	1.6 (0.8)	0.82	2.4 (0.7)	1.2 (0.9)	**1.76**
Stiffness	1.9 (0.9)	1.9 (0.9)	0.75	2.4 (0.8)	1.6 (0.9)	**0.98**
Physical function	1.5 (1.0)	1.4 (0.9)	0.86	2.4 (0.7)	1.4 (0.9)	**1.49**
SF-36						
Pain	48.6 (23.2)	50.9 (25.5)	0.77	29.5 (18.0)	57.0 (23.4)	**1.53**
Energy/fatigue	55.2 (22.9)	49.7 (25.3)	0.78	44.7 (24.6)	58.0 (25.1)	**0.54**
Physical functioning	49.8 (25.2)	46.7 (23.8)	0.82	24.2 (18.7)	41.6 (23.5)	**0.93**
Emotional well-being	64.5 (22.7)	60.6 (27.2)	0.65	61.4 (25.0)	65.5 (22.7)	**0.16**
Role physical	68.4 (29.6)	66.9 (28.6)	0.82	39.3 (27.7)	60.6 (28.1)	**0.77**
General health	54.4 (21.2)	51.2 (22.9)	0.84	58.4 (21.9)	64.0 (21.4)	**0.25**
Social functioning	72.6 (28.8)	66.1 (29.4)	0.77	49.2 (28.9)	68.5 (25.8)	**0.67**
Role emotional	80.7 (24.8)	76.0 (27.6)	0.68	63.6 (28.9)	74.1 (26.9)	**0.36**

ES, effect size; ICC, intraclass correlation coefficient; SD, standard deviation; OA-Quest, Osteoarthritis Questionnaire; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; SF-36, Short-Form Health Survey.
Effect sizes that correspond to significant changes ($P < 0.05$) in scale scores following joint replacement surgery are bolded.

[14]. A two-way mixed effects absolute agreement model was used [32], with single-measure ICCs reported.

The convergent and discriminant validity of the OA-Quest was examined with a general multitrait multimethod confirmatory factor analysis (MTMM CFA) model [33]. The MTMM approach [34] is based on the use of multiple measurement methods to assess multiple characteristics (traits) and postulates that variance in each scale can be partitioned into three main sources: (1) correlation with the underlying concept (trait effect); (2) grouping of scales in the same measure (method effect); and (3) unique variance. The MTMM CFA model specifies N trait factors (conceptual variables of interest) and K method factors (common measurement instrument), with each scale constrained to load on one method factor and one trait factor. High loadings ($\geq$0.7) of scales on their respective trait factors are interpreted as supporting their convergent validity, whereas moderate-to-low ($<$0.7) loadings on a method factor are interpreted as supporting discriminant validity.

The MTMM CFA incorporated 23 variables, including seven OA-Quest scales; three WOMAC scales, four SF-36 scales convergent to the OA-Quest (pain, physical functioning, emotional well-being, and role physical), CES-D, MOSSS sleep adequacy and sleep disturbance scales, WPAI:GH activity impairment and work productivity loss scales, and the four single-item questions. Justification for the selection of convergent measures is in Supplementary Material 2. The model specified seven trait factors that corresponded to the underlying constructs represented by the OA-Quest scales and four method factors that corresponded to the main measures (OA-Quest, SF-36, WOMAC, and single-item measures; Fig. 1). To avoid convergence issues due to modeling factors with small number of indicators ($<$3), method variance in the MOSSS and WPAI:GH scales was captured through correlated residuals [33]. CES-D scale had no common measures and was specified to load only onto a trait factor. Intercorrelations between trait factors and intercorrelations between method factors were allowed, but correlations between trait and method factors were set to zero [33]. Variances of all factors were constrained to one for identification purposes. Good model fit was commensurate with nonsignificant χ^2 ($P > 0.05$), normed χ^2 1–3 [35], Tucker-Lewis index (TLI) and comparative fit index (CFI) > 0.95, and root mean square error of approximation (RMSEA) and standardized root mean square residual (SRMR) $\leq$0.08 [36,37].

Responsiveness to joint replacement surgery					
Hip replacement (n = 30)			Knee replacement (n = 48)		
Time 1	Time 2		Time 1	Time 2	
M (SD)	M (SD)	ES	M (SD)	M (SD)	ES
2.3 (0.9)	0.8 (0.8)	**1.65**	2.3 (1.0)	1.4 (1.1)	**0.95**
1.9 (0.9)	0.8 (0.7)	**1.21**	1.9 (0.8)	1.1 (1.0)	**0.92**
1.2 (1.1)	0.5 (0.7)	**0.62**	1.4 (1.1)	0.9 (1.1)	**0.55**
2.2 (1.0)	0.8 (0.8)	**1.38**	2.0 (1.2)	1.2 (1.1)	**0.63**
0.8 (0.9)	0.5 (0.7)	0.25	1.1 (1.2)	0.7 (0.9)	**0.30**
2.1 (0.9)	1.1 (1.0)	**1.07**	2.3 (1.1)	1.4 (1.1)	**0.86**
1.3 (1.4)	0.7 (0.9)	**0.57**	1.5 (1.4)	1.2 (1.3)	0.10
2.4 (0.5)	0.9 (0.8)	**2.90**	2.4 (0.7)	1.5 (1.0)	**1.28**
2.3 (0.8)	1.3 (0.8)	**1.36**	2.4 (0.8)	1.8 (1.0)	**0.75**
2.6 (0.5)	1.2 (0.8)	**2.60**	2.3 (0.7)	1.6 (1.0)	**1.04**
24.3 (15.4)	64.7 (21.1)	**2.62**	32.9 (18.8)	51.7 (23.8)	**1.03**
45.4 (23.1)	63.2 (25.4)	**0.82**	44.7 (25.5)	53.4 (25.4)	**0.38**
19.8 (16.2)	43.1 (26.0)	**1.41**	26.9 (19.7)	39.9 (22.3)	**0.70**
61.4 (25.8)	67.8 (21.7)	0.24	61.9 (24.6)	62.8 (24.2)	0.11
42.1 (27.3)	67.4 (28.8)	**0.91**	37.8 (27.7)	56.0 (27.5)	**0.69**
62.0 (20.9)	65.5 (23.1)	0.20	56.3 (22.4)	61.6 (20.7)	**0.29**
48.3 (27.0)	73.3 (24.9)	**0.93**	49.7 (30.3)	64.3 (26.9)	**0.52**
63.1 (28.3)	81.2 (24.3)	**0.67**	63.7 (30.6)	68.8 (28.8)	0.21

The responsiveness of the OA-Quest to joint replacement surgery was assessed by comparing scores at baseline and 3-month postsurgery. Statistical significance and magnitude of change were assessed with paired t-tests and standardized effect sizes (ES), respectively [14,38]. Positive values of ES represent improvement, ES $< |0.2|$ interpreted as no change, $|0.20|-|0.49|$ small change, $|0.50|-|0.79|$ moderate change, and $\geq|0.80|$ large change [14]. Given prior research [39,40], we expected large improvements on scales measuring physical aspects of osteoarthritis burden (*physical distress*, *physical limitations*, *sleep disturbances*, and *physical deconditioning*) and small to moderate improvements on scales measuring psychosocial aspects (*psychosocial distress*, *financial hardship*, and *lost productivity*).

Mplus version 7 [41] was used to conduct MTMM CFA. SPSS version 25.00 [42] was used for the remaining analyses.

2.4. Missing data

Data were assumed to be missing at random and were handled with full information maximum likelihood estimation in MTMM CFA and with pairwise deletion in the remaining analyses.

3. Results

3.1. Participants

From the 998 individuals invited to participate (500 MA members, 298 rheumatology outpatients, and 200 orthopedic patients), 327 (33%) responded. The majority (n = 175, 54%) were community-dwelling adults (MA members); 62 (19%) and 90 (28%) were rheumatology outpatients and orthopedic patients, respectively. Follow-up rates were 84% (52/62) for rheumatology outpatients and 87% (78/90) for orthopedic patients.

Participants were aged between 34 and 91 years (median 70 years), and most were female (72%). Only 56 people (17%) were employed. The majority (86%) reported pain in more than one joint, with hands (66%), knees (65%), feet (52%), and hips (48%) affected most frequently (Table 1).

3.2. Descriptive statistics for the scales used in the study

The proportion of missing data on the OA-Quest scales was very low (<3%; Table 2), with the exception of the *lost productivity* scale (33% missing responses). Three OA-Quest scales exhibited floor effects: *Psychosocial*

distress (19.7%), *financial hardship* (57.2%), and *lost productivity* (41.3%). Internal consistency was ≥0.87 for the OA-Quest and ≥0.81 for the WOMAC and SF-36 scales.

3.3. Temporal stability

For the 52 followed-up rheumatology outpatients, the median retest interval was 21 days (interquartile range 18–27 days). Most patients reported no change from baseline in the severity of joint problems (n = 32, 62.7%), emotional well-being (63.5%), sleep (63.5%), financial situation (88.5%), ability to perform everyday tasks (69.2%), maintain weight (78.8%), exercise (71.2%), or do paid work (69.2%). Temporal stability was ≥0.75 for the OA-Quest and WOMAC and ≥0.65 for the SF-36 scales (Table 3).

3.4. Convergent and discriminant validity

The MTMM CFA model had adequate fit with the data (χ^2 = 483.54, df = 184, $P < 0.001$, normed χ^2 = 2.63, CFI = 0.95, TLI = 0.93, RMSEA = 0.07, and SRMR = 0.06; Table 4 and Supplementary Material 3). The results supported both the convergent and discriminant validity of the OA-Quest, with moderately high to high loadings on their trait factors (λ rage 0.66–0.90) and low to moderate loadings on the OA-Quest method factor (λ range 0.27–0.51).

Table 4. Results of multitrait multimethod confirmatory factor analysis testing convergent and discriminant validity of the OA-Quest

		Method factor loadings				
Trait factors	**Trait factor loadings**	**OA-Quest**	**WOMAC**	**SF-36**	**Single item measures**	**Residual variance**
Physical distress						
OA-Quest physical distress	0.87	0.40				0.09
WOMAC pain	0.75		0.56			0.13
WOMAC stiffness	0.58		0.46			0.45
SF-36 pain	−0.82			0.33		0.22
Physical limitations						
OA-Quest physical limitations	0.84	0.46				0.09
WOMAC physical function	0.76		0.59			0.07
SF-36 physical functioning	−0.73			0.30		0.38
WPAI:GH activity impairment	0.83					0.31
Psychosocial distress						
OA-Quest psychosocial distress	0.77	0.51				0.15
SF-36 emotional well-being	−0.82			−0.27		0.25
CES depression	0.90					0.19
Sleep disturbances						
OA-Quest sleep disturbances	0.90	0.27				0.11
MOSSS sleep adequacy	0.70					0.82
MOSSS sleep disturbance	−0.42					0.50
Financial hardship						
OA-Quest financial hardship	0.87	0.42				0.08
Difficulty with paying for everyday things	0.77				0.31	0.32
% of income spent on treatment	0.40					0.84
Physical deconditioning						
OA-Quest physical deconditioning	0.86	0.40				0.11
Difficulty with weight	0.63				0.09[a]	0.59
Difficulty with exercising	0.75				0.27	0.37
Lost productivity						
OA-Quest lost productivity	0.66	0.36				0.44
SF-36 role physical	−0.80			0.16		0.33
WPAI:GH work productivity loss	0.14					0.98

CES, Center for Epidemiologic Studies; MOSSS, Medical Outcomes Study Sleep Scale; QA-Quest, Osteoarthritis Questionnaire; SF-36, Short-Form Health Survey; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; WPAI:GH, Work Productivity and Activity Impairment Questionnaire General Health Version 2.

Blank cells correspond to factor loadings constrained to 0; fully standardized factor loadings are shown.

[a] Statistically nonsignificant loading ($P > 0.05$).

3.5. Responsiveness to change

Of the 78 followed-up orthopedic patients, 30 (38.5%) and 48 (61.5%) underwent total hip and knee replacement, respectively. The majority of procedures were unilateral, with only 1/30 (3.3%) bilateral hip and 6/48 (12.5%) bilateral knee replacements. At 3-month postsurgery (median 3.5 months, interquartile range 2.9–4.3), most participants reported improvements in joint pain (92.3%), sleep (65.3%), emotional well-being (73.1%), and ability to exercise (70.5%) and perform daily tasks (75.6%). However, only 21.8%, 50.0%, and 26.9% reported improvements in financial situation, ability to maintain healthy weight, and do paid work, respectively.

As predicted (and consistent with self-perceptions of global changes), large improvements (ES 0.89—1.21) occurred on the OA-Quest *physical distress*, *physical limitations*, *sleep disturbances*, and *physical deconditioning* scales and small-to-moderate improvements on the *psychosocial distress* (ES = 0.58), *lost productivity* (ES = 0.29), and *financial hardship* (ES = 0.28) (Table 3) scales (all $P < 0.05$). The OA-Quest *physical distress* and *physical limitations* scales showed changes of a slightly lower magnitude (albeit within the large range) that the corresponding disease-specific (WOMAC pain and physical function) and generic (SF-36 pain and physical functioning) scales (Table 3). However, the OA-Quest *psychosocial distress* scale showed moderate change (ES = 0.58), whereas the conceptually related scale of the SF-36 (emotional well-being) showed no change (ES = 0.16) at 3 months follow-up. In contrast, OA-Quest *lost productivity* scale showed a small change (ES = 0.29), whereas a related scale of the SF-36 (role physical) showed moderate change (ES = 0.77). Separate analyses for hip ($n = 30$) and knee ($n = 48$) replacements generally showed greater improvements following hip surgery (Table 3).

4. Discussion

This study provides strong support for the psychometric robustness of the OA-Quest. Our previous work supported unidimensionality and factorial validity [5] of the OA-Quest, both of which are important components of construct validity [43]. The present study provides further support for the construct validity of the OA-Quest by demonstrating convergent and discriminant validity. Internal consistency estimates of the OA-Quest exceeded or approached the 0.90 threshold recommended for individual assessment [31]. Combined with evidence of temporal stability and responsiveness to joint replacement surgery, these results support the use of the OA-Quest in quantifying and monitoring osteoarthritis burden in clinical practice and research, although responsiveness of the OA-Quest to nonsurgical interventions has yet been explored. The comparison data for the WOMAC and SF-36 provide insights into the functioning of the OA-Quest relative to frequently applied osteoarthritis assessments. High internal consistency and temporal stability estimates demonstrate that the OA-Quest is at least as reliable as these measures while different patterns of improvement observed for the OA-Quest and corresponding SF-36 and WOMAC scales support conceptual distinctiveness of domains captured by the OA-Quest.

Our study also identified some weaknesses within the OA-Quest. Namely, floor effects were recorded for the *psychosocial distress*, *financial hardship*, and *lost productivity*, indicating that these scales are targeting higher levels of osteoarthritis burden than was represented in our sample. This is not surprising given that community-dwelling adults comprised the majority of respondents. Although floor effects can affect sensitivity of a measure, floor effects did not compromise the ability of the OA-Quest to detect improvements among individuals with high levels of osteoarthritis burden, as demonstrated by the results of responsiveness tests. An additional limitation was missing data for the *lost productivity* scale. Given that most respondents were retired, this is unsurprising, as this outcome was not applicable to the majority of participants. Future studies among employed individuals are needed to further assess the psychometric performance of this scale.

Despite these limitations, our study supports the psychometric robustness of the OA-Quest. Importantly, the OA-Quest was developed to specifically measure concepts of importance to individuals with this condition. The breadth of concepts captured by the OA-Quest makes it especially suited to situations when it is necessary to determine the full range of osteoarthritis burden [44].

The OA-Quest has high reliability, good evidence of construct validity, and is responsive to change in joint replacement surgery. Overall, the psychometric performance of the OA-Quest compares favorably to that of well-known questionnaires used in osteoarthritis assessment (the WOMAC and SF-36). Because the OA-Quest measures a different set of concepts to those measured by the currently available questionnaires, it will be a useful adjunct to existing measures of osteoarthritis burden.

CRediT authorship contribution statement

Ljoudmila Busija: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Validation, Visualization, Writing - original draft, Writing - review & editing. **Richard H. Osborne:** Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Validation, Writing - review & editing. **Gemma Tatangelo:** Data curation, Formal analysis, Methodology, Validation, Writing - review & editing. **Silvana Niutta:** Data curation, Methodology, Project administration, Resources, Writing - review & editing. **Rachelle Buchbinder:**

Conceptualization, Investigation, Methodology, Resources, Supervision, Validation, Visualization, Writing - review & editing.

Acknowledgments

The authors would like to thank Dr Lerma Ung, formerly from Arthritis Victoria and Ms Gloria Brewster and Professor Graham Mercer, formerly from the Repatriation General Hospital, Adelaide, South Australia for their assistance with recruiting participants for this project. They also wish to acknowledge all participants in this study and their valuable contribution to this work.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2019.06.001.

References

[1] Vina ER, Kwoh CK. Epidemiology of osteoarthritis: literature update. Curr Opin Rheumatol 2018;30:160−7.
[2] Hunter DJ, Schofield D, Callander E. The individual and socioeconomic impact of osteoarthritis. Nat Rev Rheumatol 2014;10:437−41.
[3] Neogi T, Zhang Y. Epidemiology of osteoarthritis. Rheum Dis Clin North Am 2013;39:1−19.
[4] Sharif B, Garner R, Hennessy D, Sanmartin C, Flanagan WM, Marshall DA. Productivity costs of work loss associated with osteoarthritis in Canada from 2010 to 2031. Osteoarthritis Cartilage 2017;25:249−58.
[5] Busija L, Buchbinder R, Osborne RH. Development and preliminary evaluation of the OsteoArthritis Questionnaire (OA-Quest): a psychometric study. Osteoarthritis Cartilage 2016;24:1357−66.
[6] Busija L, Buchbinder R, Osborne RH. A grounded patient-centered approach generated the personal and societal burden of osteoarthritis model. J Clin Epidemiol 2013;66:994−1005.
[7] Bellamy N. WOMAC osteoarthritis index. User guide VIII. Brisbane: The University of Queensland; 2007.
[8] Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833−40.
[9] Roos EM, Lohmander LS. The knee injury and osteoarthritis outcome score (KOOS): from joint injury to osteoarthritis. Health Qual Life Outcomes 2003;1:64.
[10] Busija L, Osborne RH, Roberts C, Buchbinder R. Systematic review showed measures of individual burden of osteoarthritis poorly capture the patient experience. J Clin Epidemiol 2013;66:826−37.
[11] Badley EM, Millstone DB, Perruccio AV. Nature of joint involvement in osteoarthritis in the population: multi-joint osteoarthritis, the rule not the exception? Osteoarthr Cartilage 2017;25:S198−9.
[12] Hawkins M, Elsworth GR, Osborne RH. Application of validity theory and methodology to patient-reported outcome measures (PROMs): building an argument for validity. Qual Life Res 2018;27:1695−710.
[13] Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Am Psychol 1995;50:741−9.
[14] Fayers P, Machin D. Quality of life: assessment, analysis and interpretation. Chichester: Wiley; 2000.
[15] Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34−42.
[16] Kosinski M, Keller SD, Hatoum HT, Kong SX, Ware JE Jr. The SF-36 Health Survey as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis: tests of data quality, scaling assumptions and score reliability. Med Care 1999;37:MS10−22.
[17] Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473−83.
[18] Bellamy N, Kean WF, Buchanan WW, Gerecz-Simon E, Campbell J. Double blind randomized controlled trial of sodium meclofenamate (Meclomen) and diclofenac sodium (Voltaren): post validation reapplication of the WOMAC osteoarthritis index. J Rheumatol 1992;19:153−9.
[19] McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities osteoarthritis index (WOMAC): a review of its utility and measurement properties. Arthritis Rheum 2001;45:453−61.
[20] Ware JE, Kosinski M, Gandek B. SF-36 health survey: manual and interpretation guide. 2000th ed. Lincoln: Quality Metric Inc; 1993.
[21] Ware JE, Kosinski MA, Gandek B. SF-36 health survey: manual and interpretation guide. Lincoln: Quality Metric Inc; 2005.
[22] Hawker G, Melfi C, Paul J, Green R, Bombardier C. Comparison of a generic (SF-36) and a disease specific (WOMAC) (Western Ontario and McMaster Universities Osteoarthritis Index) instrument in the measurement of outcomes after knee replacement surgery. J Rheumatol 1995;22:1193−6.
[23] Kiebzak GM, Campbell M, Mauerhan DR. The SF-36 general health status survey documents the burden of osteoarthritis and the benefits of total joint arthroplasty: but why should we use it? Am J Manag Care 2002;8:463−74.
[24] Kosinski M, Keller SD, Ware JE Jr, Hatoum HT, Kong SX. The SF-36 health survey as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis: relative validity of scales in relation to clinical measures of arthritis severity. Med Care 1999;37:MS23−39.
[25] Sawyer Radloff L. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977;1:385−401.
[26] Spritzer KL, Hays RD. MOS sleep scale user's manual, version 1.0. Los Angeles, CA: RAND; 2003.
[27] Hays RD, Martin SA, Sesti AM, Spritzer KL. Psychometric properties of the medical outcomes study sleep measure. Sleep Med 2005;6:41−4.
[28] Viala-Danten M, Martin S, Guillemin I, Hays RD. Evaluation of the reliability and validity of the medical outcomes study sleep scale in patients with painful diabetic peripheral neuropathy during an international clinical trial. Health Qual Life Outcomes 2008;6:113.
[29] Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics 1993;4:353−65.
[30] U.S. Department of Health and Human Services Food and Drug Administration. Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims. Rockville, MD: U.S. Department of Health and Human Services; 2009.
[31] Scientific Advisory Committee of the Medical Outcomes T. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res 2002;11:193−205.
[32] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420−8.
[33] Marsh HW. Confirmatory factor analyses of multitrait-multimethod data: many problems and a few solutions. Appl Psychol Meas 2016;13:335−61.

[34] Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol Bull 1959;56:81−105.
[35] Schumacker RE, Lomax RG. A beginner's guide to structural equation modeling. 2nd ed. Mahwah: Lawrence Erlbaum; 2004.
[36] Brown TA. Confirmatory factor analysis for applied research. New York: The Guilford Press; 2006.
[37] Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model 1999;6:1−55.
[38] Cohen J. Statistical power analysis for social and behavioural sciences. New York: Academic Press; 1977.
[39] Busija L, Osborne RH, Nilsdotter A, Buchbinder R, Roos EM. Magnitude and meaningfulness of change in SF-36 scores in four types of orthopedic surgery. Health Qual Life Outcomes 2008;6:55.
[40] March LM, Cross MJ, Lapsley H, Brnabic AJM, Tribe KL, Bachmeier CJM, et al. Outcomes after hip or knee replacement surgery for osteoarthritis - a prospective cohort study comparing patients' quality of life before and after surgery with age-related population norms. Med J Aust 1999;171:235−8.
[41] Muthén LK, Muthén BO. Mplus. 7th ed. Los Angeles, CA: Muthén & Muthen; 1998-2012.
[42] IBM Corp. IBM SPSS statistics for windows. 25th ed. Armonk, NY: IBM Corp; 2017.
[43] Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull 1955;52:281−302.
[44] Australian Commission on Safety and Quality in Health Care. Osteoarthritis of the knee clinical care standard. Sydney: ACSQHC; 2017.