Performance of the Assessment of Quality of Life measure in people with hip and knee joint disease and implications for research and clinical use.

Dr Ilana N Ackerman: BPhysio(Hons), PhD
Melbourne EpiCentre, Department of Medicine (Royal Melbourne Hospital)
The University of Melbourne

Dr Lucy Busija: BA(Hons), MSc, PhD
Melbourne EpiCentre and Melbourne Brain Centre, Department of Medicine (Royal Melbourne Hospital)
The University of Melbourne and
Biostatistics Unit, Faculty of Health, Deakin University

Mark A Tacey: BSc
Melbourne EpiCentre, Department of Medicine (Royal Melbourne Hospital)
The University of Melbourne and Melbourne Health

Dr Megan A Bohensky: BA, MPH, PhD
Melbourne EpiCentre, Department of Medicine (Royal Melbourne Hospital)
The University of Melbourne

Dr Zanfina Ademi: MPharm, MPH, PhD
Melbourne EpiCentre, Department of Medicine (Royal Melbourne Hospital)
The University of Melbourne and
Department of Epidemiology and Preventive Medicine, Monash University

A/Prof Caroline A Brand: MBBS, BA, MPH, FRACP
Melbourne EpiCentre, Department of Medicine (Royal Melbourne Hospital)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1002/acr.22129
© 2013 American College of Rheumatology
Received: Jan 31, 2013; Revised: Jul 15, 2013; Accepted: Aug 13, 2013
Melbourne Health

Prof Danny Liew: MBBS(Hons), FRACP, PhD
Melbourne EpiCentre, Department of Medicine (Royal Melbourne Hospital)
The University of Melbourne

Corresponding author

Dr Ilana N Ackerman
Melbourne EpiCentre
Department of Medicine (Royal Melbourne Hospital)
The University of Melbourne
Parkville Victoria 3050
Australia
Phone: +61 3 9342 2153 Fax: +61 3 9312 8780
Email: i.ackerman@unimelb.edu.au

Funding support

This research was supported in part by a Physiotherapy Research Foundation and United Pacific Industries Thermoskin Research Grant (#T09-THE003). Dr Ackerman is supported by a National Health and Medical Research Council of Australia Public Health (Australian) Early Career Fellowship (#520004). Dr Busija was supported by a postdoctoral fellowship jointly funded by the Melbourne EpiCentre and the Centre for Research Excellence in Translational Neuroscience (Royal Melbourne Hospital and The University of Melbourne).

Word count: 3796
Abstract

Objective: To comprehensively evaluate the performance of the Assessment of Quality of Life (AQoL) instrument for measuring Health-Related Quality of Life (HRQoL) in people with hip and knee joint disease (arthritis or osteoarthritis).

Methods: Data from 237 individuals were available for analysis from a national cross-sectional, population-based study of hip and knee joint disease in Australia. AQoL-4D data were evaluated using Rasch analysis. A range of measurement properties was explored including model and item fit, threshold ordering, differential item functioning and targeting.

Results: Good overall fit of the AQoL with the Rasch model was demonstrated across a range of tests, supporting internal validity. Only 1 item (relating to hearing) showed evidence of misfit. Most AQoL items showed logical sequencing of response option categories, with threshold disordering evident for only 2 of the 12 items (items 4 and 9). Minor issues with potential clinical and research implications include limited options for reporting pain, and some evidence of measurement bias between demographic subgroups (including age and sex). Participants’ HRQoL was generally better than that represented by the AQoL items (mean for person abilities -2.50, SD=0.66; mean for item difficulties 0.00, SD 0.67), indicating ceiling effects which could impact on the instrument’s ability to detect HRQoL improvement in population-based studies.

Conclusion: The AQoL is a competent tool for assessing HRQoL in people with hip and knee joint disease, although researchers and clinicians should consider the caveats identified when selecting appropriate HRQoL measures for future outcome assessment involving this patient group.

Word count: 247

Keywords

Osteoarthritis, Outcome Measures, Quality of Life, Questionnaire Design
Significance and Innovations

- Although the AQoL instrument has been used to assess patient outcomes and treatment effectiveness in several studies involving people with hip and knee joint disease, its psychometric properties have not been fully evaluated in this setting.

- Overall, the AQoL items worked well as a unidimensional measure of HRQoL and only 1 of the 12 items (relating to hearing) demonstrated misfit with the Rasch model.

- The AQoL offers several benefits which support its use in research and clinical settings, but issues relevant for hip and knee joint disease include limited options for classifying milder pain, and some evidence of ceiling effects and measurement bias between demographic subgroups.
Introduction

Osteoarthritis (OA) affects over 8% of the Australian population (1), with the hip and knee joints commonly involved. As in many countries, OA is the most common reason for hip and knee replacement surgery in Australia (2) and a growing concern in terms of disability, lost productivity and health care costs. In Australia, over $AUD 1 billion is spent by the government annually on OA-related treatment including primary care, specialist medical services, medications and joint replacement surgery (3). Reliable measurement of health outcomes in OA is of increasing importance for funding providers, health professionals and patients. In particular, assessment of the efficacy and cost-effectiveness of interventions for hip and knee OA is used to support advocacy, program planning and service provision.

An important patient-centred health outcome is Health-Related Quality of Life (HRQoL) (4). While there is no single definition, this construct is often used to describe the personal impact of an illness and any related treatment (5). Adopting a broader approach, the World Health Organization (WHO) describes quality of life as “an individual’s perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns (6)”. HRQoL instruments capture the impact of health status on life, including physical, emotional and social health domains. Although disease-specific measures play an important role, an advantage of generic HRQoL measures is that they permit comparisons between different chronic diseases and patient populations.

The Assessment of Quality of Life (AQoL) scale was developed to provide a simple, comprehensive measure of HRQoL for evaluating health care interventions (7). Its development was based on the WHO’s concepts of health and quality of life, review of other measures, and consultation with stakeholders including patients and health professionals. The AQoL is a multi-attribute measure which encompasses several key concepts including independent living, social relationships, physical senses and psychological wellbeing (8). As a generic measure which is easy to administer, it has potential applications in research and clinical settings and its utility properties enable health economic evaluation. Based on its use in rheumatology research, the AQoL was selected for inclusion in a recent review of
adult measures of general health and HRQoL (9). Although psychometric evaluations of the AQoL have been undertaken in several patient populations (10-14), only one study (15) has utilised item response theory (IRT), in the form of the Rasch measurement model, with the primary aim of constructing a shortened version of the AQoL (AQoL-8). However, as this study (15) used data from the original validation study, information about how well the AQoL works in external populations (beyond those used to develop the instrument) is not available.

The Rasch measurement model (16) allows an objective assessment of scale attributes that are implicit in measurement, including unidimensionality of the item set, assessment of the ordering of response categories, and assessment of possible measurement bias (17). This makes the Rasch model a useful adjunct to traditional measurement approaches. Given the ongoing use of the AQoL in rheumatology research, the purpose of this study was to apply the Rasch model to evaluate the measurement properties of the AQoL instrument in hip and knee joint disease (arthritis or OA).
Materials and Methods

Study design

This study was a secondary analysis of data collected in a population-based survey.

Participants and procedure

Data from 237 individuals with hip arthritis, hip OA, knee arthritis and/or knee OA were available for analysis from a national study of hip and knee joint disease. The sample selection and recruitment procedures have been reported previously (18). Briefly, sampling of people aged ≥39 years from all 8 Australian states and territories was undertaken using an extract from the federal electoral roll obtained from the Australian Electoral Commission. As electoral enrolment and voting is compulsory for Australians aged ≥18 years, the electoral roll provides comprehensive coverage of the Australian adult population. The selected sample was mailed an introductory letter, plain language statement and study questionnaire. Return of a completed questionnaire was deemed to constitute consent, as approved by The University of Melbourne Human Research Ethics Committee. The participants included in the present study had responded affirmatively to at least one of the following screening questions:

- “have you ever been told by a doctor or other health professional that you have hip arthritis?”;
- “have you ever been told by a doctor or other health professional that you have hip osteoarthritis?”;
- “have you ever been told by a doctor or other health professional that you have knee arthritis?”; and
- “have you ever been told by a doctor or other health professional that you have knee osteoarthritis?”

Study questionnaire

In addition to screening for doctor-diagnosed hip and knee joint disease, the study questionnaire was used to collect demographic and HRQoL data. The Western Ontario and McMaster Universities Osteoarthritis (WOMAC) Index, a disease-specific measure of health status (19), was used to classify hip and knee joint disease severity. It contains 24 items covering pain, stiffness and physical function. Total WOMAC scores were transformed to a 0 (best health) to 100 (worst health) scale. Similar to methods reported previously (20), a WOMAC score <7 was considered to be asymptomatic joint disease, 7-38 was considered mild-moderate disease and ≥39 was considered severe joint disease.
HRQoL was assessed using the 12 item AQoL-4D instrument. Three additional illness items do not contribute to utility score calculation and were not administered. Item response options range from 1 (indicating best HRQoL state) to 4 (indicating worst HRQoL state) (8). Utility scores range from -0.04 (worst possible HRQoL) to 1.00 (full HRQoL). Negative AQoL utility scores indicate a health state worse than death (21). The AQoL has demonstrated good psychometric properties (9) and has been validated for self-administration (22). It has been used in several arthritis studies (9) and is comparable to the WOMAC Index and Short Form-36 for detecting HRQoL differences in OA populations (13). Australian normative data are available (23).

**Statistical analysis**

Descriptive analysis of demographic data was performed using IBM SPSS Statistics 20 (IBM, Armonk, NY). Rasch analysis was conducted using Winsteps® Rasch measurement software (Winsteps.com, Beaverton, OR). Winsteps® deals with missing data on an item-by-item basis (all available data for each AQoL item are included in the analyses).

The Rasch model is a mathematical model that summarises the relationship between the level of a characteristic (e.g., HRQoL) represented by an item (referred to as ‘item difficulty’) and the level of the same characteristic possessed by a person (referred to as ‘person ability’) (16). The Rasch model predicts that more able individuals will endorse items that are more difficult. When the observed pattern of individual responses to scale items conforms to the pattern of responses predicted from the Rasch model, a scale is functioning as intended and internal validity is supported (24). Fit between the AQoL scale and the Rasch model was evaluated at the individual item and scale levels. The criteria used to assess fit of the AQoL with the Rasch model are summarised in Figure 1.

Although the Rasch model was originally developed for use with dichotomous items (16), extensions of the model to polytomous (multi-category) items have since been proposed. In this study, we applied a partial credit model (25) to analyse the AQoL’s measurement properties. This model was selected as it
allows the locations of item thresholds to differ across items and is appropriate when each item is scored
using a different response scale. Additionally, we performed a likelihood ratio test comparing a partial
credit model (chi-square = 3465.65, df=2385, p<0.001) with a rating scale model (26) (chi-square =
3759.48, df=2405, p<0.001), which constrains threshold locations to be the same across all items. This
test was significant (chi-square = 293.83, df=20, p<0.001), indicating that a partial credit model was
more appropriate for this study.

Overall fit between the data and the Rasch model is traditionally evaluated using the chi-square
goodness of fit test (27); this tests the null hypothesis that the model has perfect fit with the Rasch model
in the population. However, given its sensitivity to sample size, the chi-square test is often statistically
significant. To overcome this, Kline (28) proposed a normed chi-square statistic, which divides the chi-
square value by the degrees of freedom. For good model fit, the normed chi-square should be <2.5 (29).

The Rasch model was also used to assess item threshold ordering, differential item functioning (DIF)
and targeting. Thresholds define the boundaries between item response categories and represent the
transition from one response category to the next (30). Evaluation of item threshold ordering can confirm
whether response categories within each AQoL item are ordered hierarchically, with a progressive
sequence of worsening HRQoL states. For example, item 12 describes a progressive sequence of
worsening pain: response option 1 ‘None at all’; option 2 ‘I have moderate pain’; option 3 ‘I suffer from
severe pain’; option 4 ‘I suffer from unbearable pain’. When response categories are functioning as
intended, individuals with greater pain are expected to endorse response options that represent higher
pain levels (31). Disordered thresholds can occur when individuals have difficulty discriminating
between response categories, when there are too many response categories or when items are unclear.

Differential item functioning is a form of measurement bias which occurs when population subgroups
with comparable HRQoL utility scores differ in their pattern of responses to an item (30). In this study,
DIF was assessed across subgroups based on age (≤65 years vs ≥65 years), sex, education (high
school or less vs further education), residential location (metropolitan vs provincial/rural), country of birth
(Australia vs other), primary language (English vs other) and previous hip or knee surgery (surgery vs no surgery). The criteria used to evaluate DIF in this study are summarised in Figure 1.

Targeting refers to the ability of the AQoL items to adequately represent the full continuum of ability levels of the sample, providing an assessment of the extent of construct coverage. Targeting was assessed by examining the distribution of AQoL item difficulties versus person abilities, and comparing the means and standard deviations (SD) of these distributions. The Rasch model converts person abilities and item difficulties into logits, placing items and persons onto a common measurement scale. The mean location of items is always fixed at 0 and when the distribution of item difficulties presents a good match for the distribution of person abilities, the latter also has a mean of 0. A scale’s ability to discriminate between levels of a construct (such as HRQoL) is maximised when there is a good match between person abilities and item difficulties. The criteria used to assess the targeting ability of the AQoL are summarised in Figure 1.

Finally, the Rasch model assumes that all items in a scale measure a single concept (unidimensionality). Unidimensionality was examined using the procedure described by Smith (32). Principal components analysis of the residuals (PCAR) was conducted to identify potential subsets of items within the scale. For each participant, ability estimates derived from each subset were then compared using independent t-tests. The assumption of unidimensionality is supported when <5% of participants differ significantly on these tests. Related to unidimensionality is the assumption of local response independence, which assumes that the response to one item is independent from the response to other items. Correlations of ≥0.3 between item residuals indicate response dependency, potentially violating the unidimensionality of a scale (24). The criteria used to evaluate unidimensionality are summarised in Figure 1.
Results

Participants

The demographic characteristics of the sample are presented in Table 1. The median (interquartile range) age was 66 (56-74) years and 62% of participants ($n=147$) were female. Sixteen per cent ($n=39$) had hip arthritis or hip OA, 60% ($n=142$) had knee arthritis or knee OA, and 23% ($n=54$) reported arthritis or OA affecting the hip and knee joints. Based on participants’ WOMAC scores, 16% ($n=37$) were classified as having asymptomatic joint disease, 51% ($n=120$) as mild to moderate, and 27% ($n=64$) had severe joint disease. Severity could not be classified for 16 participants (7%) due to missing WOMAC responses.

AQoL scores

AQoL utility scores for the sample ranged from -0.04 to 1.00. The mean (SD) AQoL score for the sample was 0.64 (0.25), indicating that average HRQoL for people with hip or knee joint disease was well below Australian population norms (mean 0.83, SD 0.20, minimal important difference ~0.06 (23)). Those with severe joint disease had the lowest HRQoL (mean 0.42, SD 0.26) which was similar to scores reported for people with severe hip and knee joint disease awaiting joint replacement (mean 0.39, SD 0.24 (33)). Individuals with mild to moderate disease reported higher AQoL scores (mean 0.71, SD 0.18), and those with asymptomatic disease had the highest HRQoL (mean 0.80, SD 0.21).

The proportion of missing scores for each item was low, ranging from 2% for items 2 (difficulty with household tasks) and 6 (relationships with family) to 5% for item 12 (pain or discomfort).

Fit with the Rasch model

Good overall fit of the AQoL with the Rasch model was demonstrated across a range of tests, supporting internal validity. The normed chi-square was 1.45, demonstrating very good fit between the data and the Rasch model. Good model fit was also supported by the mean-square indices, with all values within the acceptable range (Table 2). The Person Separation Index was 4.79, indicating that the AQoL items
could distinguish between at least four levels of HRQoL (27). Only one item showed evidence of misfit with the Rasch model. Item 8 (hearing) had higher than expected values on 3 of the 4 item fit indices (Table 2), indicating a high amount of random error (greater unpredictability) in the responses to this item. There was no evidence of misfit with the Rasch model for the remaining 11 AQoL items. The Test Characteristic Curve and the Test Information Function for the AQoL are presented in the Supplementary materials (Figures S1 and S2).

Threshold ordering

The higher response category options on the AQoL items are intended to represent worse HRQoL (7). However, the distribution of item threshold locations revealed evidence of threshold dis ordering for 2 items (Figure 2). Category reversal was seen for item 9, where the difficulty of the third response option (‘I am only understood by people who know me well…’) was greater than the difficulty of the fourth response option category (‘I cannot adequately communicate with others’) (item difficulties 0.00 and -0.42, respectively). However, the fourth response option category was only reported by 1 individual. Additionally, the third and fourth response option categories for item 4 (relating to relationships) had an equivalent level of difficulty, with threshold locations of -0.56 and -0.57. This may reflect their conceptual similarity (‘My relationships are seldom close and warm’, compared with ‘I have no close and warm relationships’).

Differential item functioning

Evidence of DIF was found for several AQoL items. Significant age-related bias was identified for item 2 (difficulty with household tasks) by the Mantel-Haenszel test (Bonferroni-corrected alpha = 0.004) and the DIF contrast test (≥0.50 logits). Older individuals reported higher scores (indicating a lower HRQoL state) for this item, compared with younger individuals with similar HRQoL scores (DIF contrast = -0.89, p <0.001). The DIF contrast test also indicated that older people had higher scores on item 3 (difficulty getting around the home and community; DIF contrast = -0.63, p=0.058) and item 8 (hearing; DIF contrast = -0.67, p=0.007), but lower scores on item 11 (anxiety, worry and depression; DIF contrast = 0.55, p=0.014).
Both tests identified sex-related bias for item 2 (household tasks; DIF contrast = -1.08, p < 0.001), item 8 (hearing; DIF contrast = 0.87, p = 0.002) and item 9 (communication with others; DIF contrast = 0.96, p = 0.002). Females reported higher scores (indicating a lower HRQoL state) for item 2 but lower scores for items 8 and 9. The DIF contrast test also showed that females scored higher on item 11 (anxiety, worry and depression; DIF contrast = -0.52, p=0.028) but scored lower on item 7 (sight; DIF contrast = 0.62, p=0.038) than males with comparable utility scores.

No significant DIF was detected for any AQoL items according to residential location. However, the DIF contrast test identified a number of items that displayed measurement bias according to country of birth (item 12, pain and discomfort), primary language (item 2, household tasks; item 3, difficulty getting around the community; item 4, quality of relationships; item 5, relationships with other people; item 10, sleep; item 12, pain and discomfort), education (item 7, sight) and previous hip or knee surgery (item 9, communication with others).

Targeting

The ability of the AQoL items to target the full continuum of HRQoL in the sample was fair (Figure 3). Item difficulties ranged from -1.70 to 1.21 and person abilities ranged from -5.77 to 1.72. The mean for person abilities (Z=2.15, SD=0.66) was well below the mean of item difficulties (0.00, SD=0.67), indicating that participants’ HRQoL was generally better than that represented by the items. In addition, the fourth response category (indicating the worst HRQoL state) was not endorsed by any participants for items 3 (difficulty getting around community) and 7 (sight), further indicating that these items represented worse HRQoL than experienced by the sample.

Unidimensionality

Although a residual correlation of 0.37 between items 1 and 2 was identified (relating to self-care and activities of daily living, respectively), further analyses supported unidimensionality of the AQoL. PCAR identified two potential sub-dimensions within the AQoL: the first sub-dimension contained items 1, 2, 3,
6 and 12, and the second contained items 4, 5, 7, 8, 9, 10, 11 (see Table 2 for item descriptors). However, independent t-tests showed that ability estimates derived from these subsets differed significantly for only 16 of the 237 individuals (6.8%, 95% CI 4.0%-10.9%), supporting unidimensionality of the AQoL items.
Discussion

This study has provided detailed information on the performance of the AQoL instrument for evaluating HRQoL in people with hip and knee joint disease. The AQoL has previously been found to be an effective tool for assessing HRQoL among people with joint disease (12, 13) and our analyses support its use in this setting. However, we identified several pertinent issues which should be considered when selecting an appropriate HRQoL measure for research or clinical purposes.

This research represents the first evaluation of the 12-item AQoL instrument using Rasch analysis in a population external to that used for its original development. Overall, the AQoL items worked well as a single, unidimensional measure of HRQoL and our analyses revealed reasonable fit with the Rasch model, supporting internal validity. This is consistent with the findings of the Rasch analysis for the shortened AQoL-8 instrument (15). Only 1 item demonstrated misfit (item 8, relating to hearing) and this should be viewed in the context of other measures used in OA such as the WOMAC Index (34), the OAQoL instrument (35) and the Intermittent and Constant Osteoarthritis Pain (ICOAP) scale (36), which all showed evidence of misfit with the Rasch model, necessitating removal of several items to achieve model fit.

From a pragmatic standpoint, the AQoL offers several benefits which support its use in research and clinical settings (37). The instrument and scoring algorithm are freely available and the AQoL can be used to generate descriptive HRQoL data and evaluate cost-effectiveness. Participant burden is minimal, as evidenced by the low proportion of missing data across the 12 items. However, our findings have several implications for use of the AQoL in these settings. There was some evidence of ceiling effects, meaning that participants tended to choose the best HRQoL state. This indicates that while there is potential for detecting HRQoL deterioration, the instrument may not detect improvement among samples selected from the general population because many people score highly. This would not be a concern for studies involving patients with lower baseline HRQoL; for example, those awaiting joint replacement surgery, where the AQoL was found to be efficient in detecting post-operative improvement (12). We also found evidence of measurement bias according to particular personal characteristics (for example, age and sex), indicating the instrument would not be appropriate for comparing HRQoL...
between demographic subgroups. However, there may be valid explanations for why subgroups responded differently to individual items. For example, psychological distress is more prevalent among females in the population (1), and among females with arthritis, compared to males with arthritis (38). This may have contributed to the DIF observed for item 11. Similarly, age-related differences in item 3 scores (relating to difficulty getting around the home and community) would also be expected.

Perhaps the most important consideration is that the AQoL offers limited response categories for describing pain severity. Item 12 (“How much pain or discomfort do you experience?”) only offers the response options “none at all”, “I have moderate pain”, “I suffer from severe pain”, and “I suffer from unbearable pain.” There are no options for classifying mild or intermittent pain, which may explain why this item demonstrated the greatest missing responses. Several participants annotated this item with additional information to describe their pain, such as “sometimes it comes and goes”, “occasionally”, “small amount”, and “mild”. This is particularly relevant as pain of variable intensity and frequency is common among this patient group, with the potential for significant HRQoL impact. While longer versions of the AQoL (e.g., AQoL-8D) contain items for reporting the frequency of serious pain and interference with activities, there are no options for reporting less than moderate pain (39). Additionally, as the AQoL is neither disease-specific nor joint-specific, pain assessment is not limited to the hip or knee. Finally, we found some evidence of threshold disordering for 2 items, meaning these items did not offer a progressive sequence of worsening HRQoL states. If similar findings are observed in other settings then revision of these item response scales may be warranted.

A key strength was our use of a population-based sample which should improve generalisability. The study utilised a relatively large sample (samples of ≥100 are recommended for Rasch analysis (24)) of people with a range of joint disease severity who reported the full spectrum of HRQoL scores. We acknowledge that our evaluation is limited to hip and knee joint disease and that model stability over time could not be evaluated as only cross-sectional data were available.

In conclusion, the AQoL is a competent tool for assessing HRQoL in people with hip and knee joint disease. Our analyses demonstrated good overall fit with the Rasch model across a range of tests,
supporting internal validity. Only 1 item showed evidence of misfit and most items showed logical sequencing of response option categories. Issues for researchers and clinicians to consider include limited options for reporting pain, some evidence of measurement bias between demographic subgroups, and ceiling effects which could impact on the ability to detect HRQoL improvement in population-based studies.
References


18. Ackerman IN, Osborne RH. Obesity and increased burden of hip and knee joint disease in Australia: Results from a national survey. BMC Musculoskelet Disord 2012;13:254.


Acknowledgments

The authors wish to thank Dr Andrew Briggs for his constructive feedback on the manuscript, and the study participants for their involvement in this research.
Figure 1. Criteria used to evaluate the psychometric properties of the AQoL within the Rasch measurement model

**Evidence for good fit with the Rasch model**
- unstandardised infit and outfit mean-square statistics between 0.5-1.5 (27)
- standardised infit and outfit residuals between -2.5-2.5 (24)
- non-significant chi-square goodness of fit test
- normed chi-square ≤2.5 (ratio of chi-square to degrees of freedom) (29)
- Person Separation Index ≥2.0 (27)

**Evidence for threshold ordering**
- sequential ordering of item threshold locations from lowest to highest (i.e., 1, 2, 3, 4)

**Evidence for differential item functioning (27)**
- significant Mantel-Haenszel test
- absolute difference of ≥0.5 logits in between-group scores for an item

**Evidence for adequate targeting**
- mean for person abilities and mean for item difficulties are comparable
- the range of logits for item difficulties covers the full range of logits for person abilities

**Evidence for unidimensionality**
- <5% of all participants differ significantly on potential subsets identified using PCAR (32)
- residual correlations between items <0.3 (24)
Table 1. Demographic characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age group</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;50 years</td>
<td>38 (16)</td>
</tr>
<tr>
<td>50-59 years</td>
<td>51 (22)</td>
</tr>
<tr>
<td>60-69 years</td>
<td>56 (24)</td>
</tr>
<tr>
<td>70-79 years</td>
<td>63 (27)</td>
</tr>
<tr>
<td>≥80 years</td>
<td>28 (12)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>147 (62)</td>
</tr>
<tr>
<td>Male</td>
<td>90 (38)</td>
</tr>
<tr>
<td><strong>Diagnosis</strong></td>
<td></td>
</tr>
<tr>
<td>Hip arthritis or hip OA</td>
<td>39 (16)</td>
</tr>
<tr>
<td>Knee arthritis or knee OA</td>
<td>142 (60)</td>
</tr>
<tr>
<td>Hip arthritis / hip OA and knee arthritis / knee OA</td>
<td>54 (23)</td>
</tr>
<tr>
<td><strong>Severity of hip and knee joint disease</strong></td>
<td></td>
</tr>
<tr>
<td>Asymptomatic (WOMAC score &lt;7)</td>
<td>37 (16)</td>
</tr>
<tr>
<td>Mild to moderate (WOMAC score 7-38)</td>
<td>120 (51)</td>
</tr>
<tr>
<td>Severe (WOMAC score ≥39)</td>
<td>64 (27)</td>
</tr>
<tr>
<td><strong>Previous hip or knee surgery</strong></td>
<td>99 (42)</td>
</tr>
<tr>
<td><strong>Body Mass index</strong></td>
<td></td>
</tr>
<tr>
<td>Underweight / normal weight (≤24.99 kg/m²)</td>
<td>60 (25)</td>
</tr>
<tr>
<td>Overweight (25-29.99 kg/m²)</td>
<td>69 (29)</td>
</tr>
<tr>
<td>Obese (≥30 kg/m²)</td>
<td>84 (35)</td>
</tr>
<tr>
<td><strong>Married or living with partner</strong></td>
<td>167 (71)</td>
</tr>
<tr>
<td><strong>Born in Australia</strong></td>
<td>181 (76)</td>
</tr>
<tr>
<td><strong>English as primary language spoken at home</strong></td>
<td>220 (93)</td>
</tr>
<tr>
<td><strong>Residential location</strong></td>
<td></td>
</tr>
<tr>
<td>Metropolitan</td>
<td>161 (68)</td>
</tr>
<tr>
<td>Provincial / rural</td>
<td>76 (32)</td>
</tr>
<tr>
<td><strong>Highest level of education completed</strong></td>
<td></td>
</tr>
<tr>
<td>Primary school or less</td>
<td>21 (9)</td>
</tr>
<tr>
<td>Year 7-10</td>
<td>81 (34)</td>
</tr>
<tr>
<td>Year 11-12</td>
<td>32 (14)</td>
</tr>
<tr>
<td>Further education (trade/technical or university)</td>
<td>101 (43)</td>
</tr>
<tr>
<td><strong>Paid work</strong></td>
<td></td>
</tr>
<tr>
<td>Paid employment</td>
<td>86 (36)</td>
</tr>
<tr>
<td>Retired</td>
<td>130 (55)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>11 (5)</td>
</tr>
<tr>
<td>Stopped work due to hip or knee arthritis / OA</td>
<td>7 (3)</td>
</tr>
</tbody>
</table>

OA: Osteoarthritis

Totals for each characteristic may not equal n=237 due to missing responses
Table 2. Item fit and overall fit statistics

<table>
<thead>
<tr>
<th>Item*</th>
<th>Item difficulty (SE)</th>
<th>Infit statistics</th>
<th>Outfit statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean-square</td>
<td>Standardised fit residual</td>
</tr>
<tr>
<td>1. Self-care</td>
<td>0.44 (0.14)</td>
<td>0.81</td>
<td>-1.10</td>
</tr>
<tr>
<td>2. Activities of daily living</td>
<td>-0.10 (0.11)</td>
<td>1.00</td>
<td>0.10</td>
</tr>
<tr>
<td>3. Mobility</td>
<td>0.22 (0.16)</td>
<td>0.66</td>
<td>-2.30</td>
</tr>
<tr>
<td>4. Relationships</td>
<td>-0.09 (0.12)</td>
<td>1.35</td>
<td>2.20</td>
</tr>
<tr>
<td>5. Social isolation</td>
<td>-0.08 (0.12)</td>
<td>0.90</td>
<td>-0.90</td>
</tr>
<tr>
<td>6. Family role</td>
<td>0.35 (0.12)</td>
<td>0.79</td>
<td>-1.80</td>
</tr>
<tr>
<td>7. Sight</td>
<td>0.33 (0.15)</td>
<td>1.19</td>
<td>2.30</td>
</tr>
<tr>
<td>8. Hearing</td>
<td>0.30 (0.12)</td>
<td>1.30</td>
<td>2.80</td>
</tr>
<tr>
<td>9. Communication</td>
<td>1.21 (0.18)</td>
<td>1.20</td>
<td>0.90</td>
</tr>
<tr>
<td>10. Sleep</td>
<td>-1.70 (0.09)</td>
<td>1.06</td>
<td>0.70</td>
</tr>
<tr>
<td>11. Anxiety</td>
<td>-0.27 (0.11)</td>
<td>0.94</td>
<td>-0.60</td>
</tr>
<tr>
<td>12. Pain</td>
<td>-0.60 (0.15)</td>
<td>0.93</td>
<td>-0.60</td>
</tr>
<tr>
<td>Overall fit with the Rasch model</td>
<td>0.00 (0.67)</td>
<td>1.01</td>
<td>0.10</td>
</tr>
</tbody>
</table>

*Item descriptors are as reported by the developers of the AQoL instrument (8)

SE: Standard error
Figure 2. Assessment of threshold ordering

The horizontal boxes indicate lack of separation for the third and fourth response option categories for AQoL item 4, and evidence of category reversal for item 9.
Figure 3. Assessment of targeting

Item difficulties represent the level of HRQoL based on the AQoL items, and person abilities represent levels of HRQoL of the study participants. Person ability estimates are based on total scores calculated across all AQoL items, with higher scores representing worse HRQoL. Item difficulty is estimated by summing the responses of all participants to that item, with higher values representing items that are more difficult. This figure demonstrates that the HRQoL of participants was generally better than that represented by the AQoL items, with the AQoL items targeting poor HRQoL.
Figure legends

Figure 1. Criteria used to evaluate the psychometric properties of the AQoL within the Rasch measurement model

Figure 2. Assessment of threshold ordering

The horizontal boxes indicate lack of separation for the third and fourth response option categories for AQoL item 4, and evidence of category reversal for item 9.

Figure 3. Assessment of targeting

Item difficulties represent the level of HRQoL based on the AQoL items, and person abilities represent levels of HRQoL of the study participants. Person ability estimates are based on total scores calculated across all AQoL items, with higher scores representing worse HRQoL. Item difficulty is estimated by summing the responses of all participants to that item, with higher values representing items that are more difficult. This figure demonstrates that the HRQoL of participants was generally better than that represented by the AQoL items, with the AQoL items targeting poor HRQoL.
Minerva Access is the Institutional Repository of The University of Melbourne

Author/s: Ackerman, IN; Busija, L; Tacey, MA; Bohensky, MA; Ademi, Z; Brand, CA; Liew, D

Title: Performance of the Assessment of Quality of Life Measure in People With Hip and Knee Joint Disease and Implications for Research and Clinical Use

Date: 2014-03-01

Citation: Ackerman, IN; Busija, L; Tacey, MA; Bohensky, MA; Ademi, Z; Brand, CA; Liew, D, Performance of the Assessment of Quality of Life Measure in People With Hip and Knee Joint Disease and Implications for Research and Clinical Use, ARTHRITIS CARE & RESEARCH, 2014, 66 (3), pp. 481 - 488

Publication Status: Accepted manuscript

Persistent Link: http://hdl.handle.net/11343/58911