

Using WOMAC Index scores and personal characteristics to estimate Assessment of Quality of Life utility scores in people with hip and knee joint disease

Dr Ilana N Ackerman: BPhysio(Hons), PhD

Melbourne EpiCentre, Department of Medicine (Royal Melbourne Hospital)

The University of Melbourne

Mark A Tacey: BSc

Melbourne EpiCentre, Department of Medicine (Royal Melbourne Hospital)

The University of Melbourne and Melbourne Health

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

Dr Megan A Bohensky: BA, MPH, PhD

Melbourne EpiCentre, Department of Medicine (Royal Melbourne Hospital)

The University of Melbourne

Prof Danny Liew: MBBS(Hons), FRACP, PhD

Melbourne EpiCentre, Department of Medicine (Royal Melbourne Hospital)

The University of Melbourne

A/Prof Caroline A Brand: MBBS, BA, MPH, FRACP

Melbourne EpiCentre, Department of Medicine (Royal Melbourne Hospital)

Melbourne Health

Corresponding author

Dr Ilana N Ackerman

Senior Research Fellow

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

Accepted manuscript

Abstract

Purpose: To determine whether Assessment of Quality of Life (AQoL) utility scores can be reliably estimated from Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) scores in people with hip and knee joint disease (arthritis or osteoarthritis).

Methods: WOMAC and AQoL data were analysed from 219 people recruited for a national population-based study. Generalised linear models were used to estimate AQoL utility scores based on WOMAC total and subscale scores and personal characteristics. Goodness of fit was assessed for each model and plots of prediction errors versus actual AQoL utility scores were used to gauge bias.

Results: Each model closely predicted the average AQoL utility score for the overall sample (actual mean AQoL 0.64, range of predicted means 0.63-0.64; actual median AQoL 0.71, range of predicted medians 0.68-0.69). No clear preferred model was identified and overall, the models predicted 40-46% of the variance in AQoL utility scores. The WOMAC function subscale model performed similarly to the total score model. The models functioned best at the mid-range of AQoL scores, with greater bias observed for extreme scores. Inaccuracies in individual-level estimates and low/high HRQoL subgroup estimates were evident.

Conclusion: Reliable overall group-level estimates were produced, supporting the application of these techniques at a population level. Using WOMAC scores to predict individual AQoL utility scores is not recommended and the models may produce inaccurate estimates in studies targeting patients with low/high HRQoL. Where pain and stiffness data are unavailable, the WOMAC function subscale can be used to generate a reasonable utility estimate.

Word count: 250

Keywords (MeSH headings)

Arthritis; Osteoarthritis; Outcome Measures; Quality of Life

Introduction

As in many countries, osteoarthritis (OA) is a major public health problem in Australia and a leading cause of disability. The hip and knee joints are commonly affected, with over 86,000 joint replacements performed in 2012 for severe joint disease [1]. As part of our national study to explore the broader burden of hip and knee joint disease (arthritis and OA) in Australia [2, 3], we collected health status and health-related quality of life (HRQoL) data using the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and Assessment of Quality of Life (AQoL) instruments, respectively. Disease-specific measures of wellbeing such as the WOMAC Index are valuable for health-related research as they incorporate domains considered relevant to the condition of interest (for example, pain and function). These measures can also be highly responsive to change [4-6], which is an important consideration when assessing intervention effectiveness or patient deterioration over time. On the other hand, generic (or non-disease-specific) health measures provide a more holistic overview of HRQoL by incorporating broader constructs such as social relationships and psychological health. A key advantage of generic HRQoL measures, such as the AQoL, is that they enable comparisons to be made between different health conditions and treatments. This is particularly relevant for funders of healthcare services (for example, governments and health insurers) who must determine priorities for allocating limited resources. Instruments such as the AQoL also generate utility scores which can be used to calculate quality-adjusted life years (QALYs) and estimates of cost-utility [7]. These data are paramount for health economic evaluations and assist policy decision-making about the value of new and existing interventions.

As disease-specific and generic measures of wellbeing provide complementary information, it is common for both types of instruments to be included in epidemiological and intervention studies. However, in studies where participant burden is a concern or for completed studies where only disease-specific data have been collected, there is great interest in statistical methods for estimating utility scores based on disease-specific scores. Using disease-specific data to estimate utility scores could also reduce bias from missing HRQoL data in longitudinal studies or randomised controlled trials. In the literature, this technique is commonly referred to as 'mapping', and the majority of research in this field has focussed on estimating EQ-5D (EuroQoL) HRQoL utility scores from other self-reported data [8]. Mapping data to utility scores also has important health policy implications, and these methods have been used in

submissions to the Pharmaceutical Benefits Advisory Committee in Australia [9] and the National Institute for Health and Care Excellence in the UK [10] to facilitate the evaluation of pharmaceutical and other interventions.

In rheumatology research, studies have mapped disease-specific WOMAC scores collected from people with hip or knee OA or knee pain to generic measures of HRQoL such as the EQ-5D [11, 12] and the Health Utilities Index Mark 3 (HUI3) [13, 14]. Researchers have also used data from joint-specific instruments such as the Oxford Hip [15] and Knee Scores [16] to estimate EQ-5D utility scores. Although the AQL instrument has been used to assess HRQoL in a range of clinical and research settings including OA studies [17], we are not aware of any studies that have mapped disease-specific WOMAC data to AQL utility scores. Whether reliable utility estimates can be generated remains unknown. Using data from a population-based study of hip and knee joint disease, this study aimed to determine whether WOMAC scores can be used to predict AQL utility scores.

Materials and Methods

Study design

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 extracted for the present study from a national survey. The sample selection, recruitment and data collection methods for the population-based survey have been reported previously [2, 18]. In brief, following approval from the Australian Electoral Commission (AEC), we obtained an extract from the federal electoral roll which was used to sample people aged ≥ 39 years from all 8 Australian states and territories. As electoral enrolment is compulsory for Australians aged ≥ 18 years, the electoral roll provides comprehensive coverage of the Australian adult population. Name, age group, sex and address information was available from the extract. 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. In addition to screening for 4 doctor-diagnosed conditions (hip arthritis, hip OA, knee arthritis and knee OA) [18], the study questionnaire was used to collect self-reported data including date of birth, country of birth, language spoken, marital status, highest level of education, height, weight and previous joint replacement surgery.

Body mass index (BMI) was calculated using height and weight data and classified into underweight (BMI < 18.5), normal weight (BMI 18.5-24.99 kg/m²), overweight (BMI 25-29.99) and obese categories (BMI ≥ 30), according to World Health Organisation definitions [19]. Residential location was classified as metropolitan or provincial/rural based on AEC ratings for each federal electoral division [20]. Socioeconomic status was approximated using postcodes to link to the Australian Socio-Economic Indexes for Areas (SEIFA) 2006 Index of Relative Socio-Economic Advantage and Disadvantage [21]. The lowest SEIFA decile represents geographical areas with the greatest socioeconomic disadvantage while the highest decile represents areas with the greatest advantage, based on census data including level of education, occupation, housing status, single-parent homes and car ownership.

The WOMAC Index and AQL-4D instruments were also included in the study questionnaire. The WOMAC Index is a disease-specific measure of health status [22] which contains 24 items covering pain (5 items), stiffness (2 items) and function (17 items). It produces pain, stiffness and function subscale scores which are summed to produce a total WOMAC score. Scores range from 0 (least pain) to 20 (highest pain) for pain, 0 (least stiffness) to 8 (greatest stiffness) for stiffness, 0 (best function) to 68 (worst function) for function, and 0 (best health) to 96 (worst health) for the total score. The AQL-4D is a generic (non-disease-specific) measure of HRQoL. It contains 12 items and produces a utility score ranging from -0.04 (worst possible HRQoL) to 1.00 (full HRQoL). Negative AQL utility scores indicate a health state worse than death [23]. Three additional items (relating to illness) do not contribute to utility score calculation and were not administered. Australian normative data are available for the AQL instrument [7]. AQL and WOMAC scores were not available for 18 individuals (7.6%) due to missing responses, leaving data from 219 participants available for analysis.

Statistical analysis

Descriptive statistics and chi-squared tests were used to identify personal characteristics which were likely to be associated with AQL utility scores. To assist with the descriptive analysis, the AQL utility scores were categorised according to tertiles (low, middle and high HRQoL). To allow for the skewed distribution and ceiling effect of the AQL utility scores, we used the AQL disutility (i.e. 1 - AQL utility) as a continuous variable when exploring its relationship with WOMAC scores and other potential covariates. Generalised linear models (GLM) with log link, Gaussian and Gamma family were explored to identify the model with the best fit. Mean absolute error (MAE), root mean squared error (RMSE), Akaike's information criterion (AIC) and Bayesian information criterion (BIC) statistics were used to assess model fit. The MAE is the average of the absolute difference between observed and predicted values. The RMSE is the squared root of the average squared prediction error, and is more sensitive to larger errors. Optimal fit was inferred through the minimisation of these two measures. The AIC and BIC were used to assess the quality of the models based on a balance of goodness of fit and model complexity. Pseudo R-squared (R^2) values provided an indication of the unexplained variance in each model and were used for comparison with published studies.

The Gaussian log link provided a slightly lower RSME, MAE and BIC than the Gamma family with log link and was therefore adopted as the base model type (Gamma distribution data are available upon request). Total WOMAC score was considered in Model A (Table 2), while each of the three WOMAC subscale scores were considered in Model B. Model C considered interactions between WOMAC subscales, and Model D considered the function subscale score only. Personal characteristics identified from the descriptive analysis or hypothesised to be associated with HRQoL were included in Model E, while personal characteristics significantly associated with the AQoL utility score were included in Model F.

Pitman's tests for differences in variance were used to examine agreement between each model and the base model (Model A). Prediction errors (predicted AQOL scores minus actual AQOL scores) were plotted against actual AQOL scores to assess model performance and bias across the full range of AQOL scores. Model accuracy was also assessed by comparing actual AQoL utility scores (mean, median and range) with predicted values for the overall sample and for the low, middle and high HRQoL tertiles. All statistical tests were two-sided and conducted at a significance level of 0.05. Statistical analysis was performed using Stata version 12 (StataCorp, Texas, USA).

Results

Participants

The demographic characteristics of the sample are presented in Table 1. The mean (SD) age of the overall sample was 64 (12) years and 62% of participants ($n=136$) were female. Although the low HRQoL tertile group was older, this difference was not statistically significant ($p=0.16$), and there were no significant differences in sex distribution ($p=0.78$) or in the proportion of Australian-born ($p=0.81$) or English speaking participants ($p=0.09$) across the tertiles. The low HRQoL tertile had the lowest proportion of people who had completed university education (13.7%, compared with 27.4% and 32.9% for the middle and high tertiles, respectively, $p=0.02$). Obesity was more common for the low HRQoL tertile (49.3% versus 36.4% and 32.8% for the middle and high tertiles, respectively, $p=0.03$). There was no significant difference in marital status or residential location according to AqoL score. However, there was a trend towards a greater proportion of the low HRQoL tertile living in areas with the greatest socioeconomic disadvantage. WOMAC scores (subscale scores and total score) decreased significantly across the tertiles, indicating that low HRQoL was associated with higher pain, more stiffness and worse function (Table 1).

Relationship between AqoL utility scores and WOMAC scores

AqoL utility scores for the sample ranged from -0.04 to 1.00. The mean (SD) AqoL score for the overall sample was 0.64 (0.25), indicating that average HRQoL was considerably lower than Australian population norms (mean 0.83, SD 0.20) [7]. Total WOMAC scores ranged from 0 to 96 (mean 27.3, SD 21.3). Mean (SD) WOMAC pain, stiffness and function subscales scores were 5.5 (4.5), 2.6 (2.0) and 19.2 (15.6), respectively. Figure 1 illustrates the relationship between AqoL utility scores and WOMAC scores and shows a clear negative correlation (Spearman correlation coefficient -0.57). Moderate negative correlations between AqoL utility scores and WOMAC pain, stiffness and function subscale scores were also evident (Spearman correlation coefficients of -0.55, -0.48 and -0.56, respectively).

Performance of the models

Table 2 presents the variables included in Models A to F and goodness of fit statistics. There was no clear preferred model, with comparable RMSE and MAE statistics evident across the range of models. Additionally, models A, D and F demonstrated better fit according to the AIC statistics, while Models A and D showed better fit according to the BIC statistic. Overall, the models explained 40-46% of the variance in AQL utility scores.

When the WOMAC pain, stiffness and function subscales were included separately (Model B), only the function subscale was significantly associated with the AQL disutility score ($p=0.006$). For Model C, none of the included variables was significantly associated with the AQL disutility score and no improvement to the goodness of fit statistics was evident. No improvement in model fit was observed for Model D (WOMAC function subscale only). However, the BIC value for Model D was slightly lower than for Model A, suggesting that using the WOMAC function score may be more efficient than using the total WOMAC score. Personal characteristics that were associated with AQL utility scores (education, obesity and socioeconomic status) were also included in Model E, together with the total WOMAC score, age, sex and previous joint replacement surgery status. However, only previous joint replacement was found to be significantly associated with the AQL disutility score. After excluding non-significant variables, 3 variables were found to be significantly associated with the AQL disutility score (Model F): total WOMAC score, previous joint replacement and sex.

Pitman's tests for differences in variance indicated that Models C and E provided inferior models, compared to Model A ($p=0.001$ and $p<0.001$, respectively). Comparisons of Models B, D and F with Model A revealed no significant differences ($p=0.197$, $p=0.390$ and $p=0.081$, respectively), indicating that these models produced similar predicted utility values. Figure 2 plots actual AQL utility scores versus prediction errors for Models A and F and shows that errors were lowest in the mid-range of AQL utility scores and highest for the extremes (low and high AQL scores). A similar pattern was found for each model. Consideration of personal characteristics in combination with the total WOMAC score (Model F)

did not significantly change the overall relationship between predicted errors and actual AQoL utility scores.

Accuracy of the models

Table 3 shows that each model accurately predicted the mean and median AQoL utility scores for the overall sample. However, the predicted range varied considerably and in some cases the models produced a minimum value beyond the lowest possible AQoL score, further indicating prediction errors particularly at the lower end of the scale. Table 3 also demonstrates that when the analyses were limited to the low HRQoL tertile, the mean and median utility scores were over-estimated by all of the models. When limited to the high HRQoL tertile, the mean and median utility scores were considerably underestimated by each model. In contrast, for the middle HRQoL tertile, the predicted mean and median scores were comparable to actual AQoL data.

Using the models to predict individual AQoL scores

Although each model accurately predicted the overall sample mean, the models were unable to reliably predict AQoL utility scores at the individual level. This is demonstrated using individual data from 2 study participants who differed according to age, gender, joint replacement surgery status and WOMAC scores:

Example 1

A 70-year old male with a history of previous joint replacement surgery had a total WOMAC score of 55.3 and pain, functioning and stiffness subscale scores of 11, 39.3 and 5, respectively. In comparison to his actual AQoL utility score of **0.47**, his predicted AQoL utility score based on:

- the total WOMAC score only (Model A) was **0.45**
- the WOMAC function subscale score only (Model D) was **0.45**
- the total WOMAC score, sex and previous joint replacement (Model F) was **0.32**

Example 2

A 54-year old female with no history of previous joint replacement surgery had a total WOMAC score of 48 and pain, functioning and stiffness subscale scores of 11, 33 and 4, respectively. In comparison to her actual AQoL utility score of **0.61**, her predicted AQoL utility score based on:

- the total WOMAC score only (Model A) was **0.53**
- the WOMAC function subscale score only (Model D) was **0.53**
- the total WOMAC score, sex and previous joint replacement (Model F) was **0.56**

The first example shows that for an individual with a low actual AQoL utility score, Models A and D produced close estimates of the utility score, while Model F considerably under-estimated this. In contrast, the second example shows that for an individual with a moderate actual AQoL score, each model under-estimated the utility score. For both examples, using either the total WOMAC score (Model A) or the function subscale score (Model D) produced similar results.

Discussion

As the healthcare costs of OA continue to grow worldwide [24], consideration of cost-effectiveness is critical for ensuring the appropriate distribution of limited healthcare resources [25]. In this context, it is clear that generic measures of HRQOL, such as the AQL instrument, are essential for the comprehensive economic evaluation of interventions in clinical and policy settings. Mapping techniques can inform this process by using individual disease-specific data to estimate utility scores, if primary utility data are not available. This study is the first to consider mapping disease-specific WOMAC data to the AQL utility score, and our methods have potential application across a range of rheumatology research settings given that WOMAC data are frequently collected in studies of OA and joint replacement surgery.

Using a population-based sample of people with hip or knee joint disease, we found that WOMAC scores could reliably predict average AQL utility scores at the overall sample level. There was no clear preferred model, according to goodness of fit statistics and error plots, and the models explained between 40% and 46% of the variance in AQL scores. We also found that a model based on the WOMAC function subscale alone (representing 17 of 24 WOMAC items) was reasonably efficient and demonstrated similar fit to the total WOMAC score model. This is encouraging for situations where WOMAC pain and stiffness data are not collected or are missing. Overall, the models appeared to function best at the mid-range of AQL scores, with evidence of bias for people with low or high AQL scores. This is consistent with earlier work which mapped WOMAC scores to EQ-5D utility scores [11]. However, as the application of these techniques relate to group-level or population-level quality of life or economic analysis (rather than estimating individual patient outcomes), it is likely that individual prediction errors will be balanced across the sample [13] and this was confirmed by our data. Prediction errors may relate in part to the different constructs covered by the WOMAC and AQL instruments [14]. Our subgroup analyses also suggest that mapping WOMAC scores to AQL utility scores would not be appropriate for studies targeting people with low HRQoL (for example, those with end-stage joint disease) or very high HRQoL (for example, individuals with early disease) as it would produce biased group estimates. The reason for this finding is unclear, as the WOMAC Index has been used for samples with low [26] and high health status [27], but could relate to subgroup differences in prioritising HRQoL.

In relation to model performance, our results were comparable to previous studies that used similar statistical techniques to estimate utility scores from WOMAC data [11-13]. However, direct comparison of findings is limited as none of these studies mapped data to the AQL instrument. Our models demonstrated slightly better predictive performance than the models reported by Grootendorst and colleagues [13] (predictive values between 39-40% for the HUI3 measure), and similar performance to the model reported by Xie et al [12] (predictive value of 45% for the EQ-5D measure). Similar to our findings, Grootendorst et al [13] found their derived model did not reliably predict individual utility scores, but was capable of estimating overall group scores. This was confirmed in a subsequent validation study [14]. Grootendorst et al's preferred model included WOMAC scores, age, sex and OA severity to estimate HUI3 utility scores. This is similar to our Model F, although we included previous joint replacement surgery which could be considered a proxy for joint disease severity. In contrast, Xie et al included only the total WOMAC score in their preferred model, while Barton et al [11] included age and sex in addition to WOMAC scores. Differences in study populations and disease severity may partially account for between-study variation as clinical trials participants [11, 13, 14] and patients from hospital orthopaedic departments [12] were used previously.

There are several limitations to this study which should be noted. The diagnosis of hip or knee joint disease was based on self-reported, doctor-diagnosed arthritis or OA, similar to the methods used for other population-based studies [28, 29]. The generalisability of the findings to other clinical populations is not known. Furthermore, although the WOMAC Index and AQL instrument are available in a range of translated languages, our analyses relate only to the English versions. We also acknowledge that additional demographic and clinical variables (such as doctor-diagnosed co-morbidities) may have improved model performance and accuracy, but a limited number of variables were collected in the survey to minimise participant burden. Additionally, we have not tested our models using an independent dataset containing WOMAC and AQL data and this is an important area for subsequent research. Similar to the two-stage methods used previously [14], we plan to soon undertake a validation study which will examine the accuracy of these models when applied to different patient populations. Specifically, this external validation will use other OA research datasets to determine whether the models yield similar results to those obtained from the original dataset. Using independent datasets which have not been part of the

models' development is a critical step in the mapping process and provides a more robust test of predictive performance. Finally, we acknowledge that the current cross-sectional design only allows prediction of utility scores at a single time point and we cannot confirm model performance over time. Future use of longitudinal datasets will enable us to determine whether WOMAC scores can be used to predict AQL utility scores at different time points.

In conclusion, our study indicates that total WOMAC and function subscale scores can be used to reliably estimate average AQL utility scores for a population-based sample with hip or knee joint disease. Although a range of models were tested, each produced similar results in terms of model performance and accuracy. Using WOMAC data to predict individual-level AQL utility scores is not recommended and greater prediction errors were evident for people with low or high HRQL, suggesting that mapping techniques are unlikely to be accurate for these groups.

Word count: 3493

Acknowledgment

We gratefully acknowledge the assistance of Alexandra Gorelik (Senior Statistician, Melbourne EpiCentre) for her statistical advice and constructive feedback. 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).

Conflict of interest

There are no conflicts of interest to declare.

Accepted manuscript

References

1. Australian Orthopaedic Association National Joint Replacement Registry. (2013). Annual report. Adelaide: Australian Orthopaedic Association.
2. Ackerman, I. N., & Osborne, R. H. (2012). Obesity and increased burden of hip and knee joint disease in Australia: Results from a national survey. *BMC Musculoskeletal Disorders*, 13, 254.
3. Ackerman, I. N., Ademi, Z., Osborne, R. H., & Liew, D. (2013). Comparison of Health-Related Quality of Life, work status and health care utilization and costs according to hip and knee joint disease severity: Results from a national Australian study. *Physical Therapy*, 93(7), 889-899.
4. McConnell, S., Kolopack, P., & Davis, A. M. (2001). The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): A review of its utility and measurement properties. *Arthritis Care and Research*, 45(5), 453-461.
5. Ackerman, I. N., Graves, S. E., Bennell, K. L., & Osborne, R. H. (2006). Evaluating quality of life in hip and knee replacement: Psychometric properties of the WHOQOL-BREF instrument. *Arthritis Care and Research*, 55(4), 583-590.
6. Collins, N. J., Misra, D., Felson, D. T., Crossley, K. M., & Roos, E. M. (2011). Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Outcome Survey Activities of Daily Living Scale (KOS-ADL), Lysholm Knee Scoring Scale, Oxford Knee Score (OKS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Activity Rating Scale (ARS), and Tegner Activity Score (TAS). *Arthritis Care & Research*, 63(S11), S208-228.
7. Hawthorne, G., & Osborne, R. H. (2005). Population norms and meaningful differences for the Assessment of Quality of Life (AQoL) measure. *Australian and New Zealand Journal of Public Health*, 29(2), 136-142.
8. Dakin, H. (2013). Review of studies mapping from quality of life or clinical measures to EQ-5D: An online database. *Health and Quality of Life Outcomes*, 11(1), 151.
9. Australian Government Department of Health. Appendix 6: Utility valuation of health outcomes. <http://www.pbs.gov.au/info/industry/listing/elements/pbac-guidelines/e-appendixes/f-appendix-6>. Accessed 8 January 2014.

10. National Institute for Health and Care Excellence (NICE). Published technology appraisals. <http://guidance.nice.org.uk/TA/Published>. Accessed 8 January 2014.
11. Barton, G., Sach, T., Jenkinson, C., Avery, A., Doherty, M., & Muir, K. (2008). Do estimates of cost-utility based on the EQ-5D differ from those based on the mapping of utility scores? *Health and Quality of Life Outcomes*, 6(1), 51.
12. Xie, F., Pullenayegum, E. M., Li, S.-C., Hopkins, R., Thumboo, J., & Lo, N.-N. (2010). Use of a disease-specific instrument in economic evaluations: mapping WOMAC onto the EQ-5D utility index. *Value In Health*, 13(8), 873-878.
13. Grootendorst, P., Marshall, D., Pericak, D., Bellamy, N., Feeny, D., & Torrance, G. W. (2007). A model to estimate Health Utilities Index Mark 3 utility scores from WOMAC index scores in patients with osteoarthritis of the knee. *Journal of Rheumatology*, 34(3), 534-542.
14. Marshall, D., Pericak, D., Grootendorst, P., Gooch, K., Faris, P., Frank, C., Bellamy, N., Torrance, G., & Feeny, D. (2008). Validation of a prediction model to estimate Health Utilities Index Mark 3 utility scores from WOMAC index scores in patients with osteoarthritis of the hip. *Value in Health*, 11(3), 470-477.
15. Pinedo-Villanueva, R. A., Turner, D., Judge, A., Raftery, J. P., & Arden, N. K. (2013). Mapping the Oxford hip score onto the EQ-5D utility index. *Quality Of Life Research*, 22(3), 665-675.
16. Dakin, H., Gray, A., & Murray, D. (2013). Mapping analyses to estimate EQ-5D utilities and responses based on Oxford Knee Score. *Quality of Life Research*, 22(3), 683-694.
17. Busija, L., Pausenberger, E., Haines, T. P., Haymes, S., Buchbinder, R., & Osborne, R. H. (2011). Adult measures of general health and health-related quality of life. *Arthritis Care & Research*, 63, S383-S412.
18. Ackerman, I. N., Busija, L., Tacey, M. A., Bohensky, M. A., Ademi, Z., Brand, C. A., & Liew, D. (in press). 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*.
19. World Health Organization. (2000). Obesity: Preventing and managing the global epidemic Geneva: World Health Organization.
20. Australian Electoral Commission. Current federal electoral divisions. <http://www.aec.gov.au/profiles/>. Accessed 17 December 2013.

21. Australian Bureau of Statistics. 2033.0.55.001 - Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia - Data only, 2006. <http://www.abs.gov.au/AUSSTATS/abs@.nsf/allprimarymainfeatures/356A4186CCDDC4D1CA257B3B001AC22C?opendocument>. Accessed 8 January 2014.
22. Bellamy, N. (2009). WOMAC Osteoarthritis Index. User guide IX. Brisbane: Australia.
23. Hawthorne, G., Richardson, J., Day, N., & McNeil, H. (2000). Construction and utility scaling of the Assessment of Quality of Life (AQoL) instrument. Melbourne: Monash University.
24. Chen, A., Gupte, C., Akhtar, K., Smith, P., & Cobb, J. (2012). The global economic cost of osteoarthritis: How the UK compares. *Arthritis*. doi:10.1155/2012/698709.
25. Hilgsmann, M., Cooper, C., Arden, N., Boers, M., Branco, J., Luisa Brandi, M., Bruyère, O., Guillemin, F., Hochberg, M., Hunter, D., Kanis, J., Kvien, T., Laslop, A., Pelletier, J., Pinto, D., Reiter-Niesert, S., Rizzoli, R., Rovati, L., Severens, J., Silverman, S., Tsouderos, Y., Tugwell, P., & Reginster, J. (2013). Health economics in the field of osteoarthritis: An expert's consensus paper from the European Society for Clinical and Economic Aspects of Osteoporosis and Osteoarthritis (ESCEO). *Seminars in Arthritis and Rheumatism*. doi:10.1016/j.semarthrit.2013.07.003.
26. Ackerman, I. N., Graves, S. E., Wicks, I. P., Bennell, K. L., & Osborne, R. H. (2005). Severely compromised quality of life in women and those of lower socioeconomic status waiting for joint replacement surgery. *Arthritis Care and Research*, 53(5), 653-658.
27. Bellamy, N., Wilson, C., & Hendrikz, J. (2010). Population-based normative values for the Western Ontario and McMaster Osteoarthritis Index and the Australian/Canadian (AUSCAN) Hand Osteoarthritis Index functional subscales. *Inflammopharmacology*, 18(1), 1-8.
28. Grotle, M., Hagen, K., Natvig, B., Dahl, F., & Kvien, T. (2008). Prevalence and burden of osteoarthritis: Results from a population survey in Norway *Journal of Rheumatology*, 35(4), 677-684.
29. Australian Bureau of Statistics. 2007-2008 National Health Survey Users' Guide - Electronic. [http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/CC0FB5A08570984ECA25762E0017CF2B/\\$File/4363055001_2007-08.pdf](http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/CC0FB5A08570984ECA25762E0017CF2B/$File/4363055001_2007-08.pdf). Accessed 8 January 2014.

Table 1. Characteristics of the sample

Characteristic	Overall sample (n=219)	Low HRQoL tertile^a (n=73)	Middle HRQoL tertile^a (n=73)	High HRQoL tertile^a (n=73)	p
Age, mean (SD)	64.2 (12.3)	66.5 (13.0)	63.0 (12.5)	63.0 (11.1)	0.16
Female, n (%)	136 (62.1)	43 (58.9)	46 (63.0)	47 (64.4)	0.77
Country of birth, n (%)					0.81
Australia	169 (77.0)	54 (74.0)	58 (79.5)	57 (78.1)	
Other	49 (22.4)	18 (24.7)	15 (20.6)	16 (21.9)	
Unknown	1 (0.5)	1 (1.4)	0 (0.0)	0 (0.0)	
Language spoken, n (%)					0.09
English	207 (94.5)	65 (89.0)	71 (97.3)	71 (97.3)	
Other	11 (5.0)	7 (9.6)	2 (2.7)	2 (2.7)	
Unknown	1 (0.5)	1 (1.4)	0 (0.0)	0 (0.0)	
Highest level of education, n (%)					0.02
Primary school or less	19 (8.7)	13 (17.8)	3 (4.1)	3 (4.1)	
Year 7-10	75 (34.3)	25 (34.3)	24 (32.9)	26 (35.6)	
Year 11-12	28 (12.8)	11 (15.1)	11 (15.1)	6 (8.2)	
Trade/TAFE	42 (19.2)	13 (17.8)	15 (20.6)	14 (19.2)	
University	54 (24.7)	10 (13.7)	20 (27.4)	24 (32.9)	
Unknown	1 (0.5)	1 (1.4)	0 (0.0)	0 (0.0)	
Marital status, n (%)					0.30
Married/de facto	156 (71.6)	47 (64.4)	56 (77.8)	53 (72.6)	
Single/divorced/widowed	60 (27.5)	24 (32.9)	16 (22.2)	20 (27.4)	
Unknown	2 (0.9)	2 (2.7)	0 (0.0)	0 (0.0)	
BMI category, n (%)					0.03
Underweight	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Normal weight	57 (28.2)	12 (17.4)	26 (39.4)	19 (28.4)	
Overweight	65 (32.2)	23 (33.3)	16 (24.2)	26 (38.8)	
Obese	80 (39.6)	34 (49.3)	24 (36.4)	22 (32.8)	
Living in provincial/rural area, n (%)	70 (31.9)	23 (31.5)	29 (39.7)	18 (24.7)	0.15
Lowest 2 SEIFA deciles, n (%)	30 (13.7)	15 (20.6)	10 (13.7)	5 (6.9)	0.06
Previous hip or knee replacement, n (%)	39 (17.8)	16 (21.9)	13 (17.8)	10 (13.7)	0.43
WOMAC score, mean (SD)					
Pain subscale	5.5 (4.5)	8.7 (4.9)	4.6 (3.2)	3.2 (3.1)	<0.001
Stiffness subscale	2.6 (2.0)	3.8 (2.1)	2.3 (1.6)	1.8 (1.6)	<0.001
Function subscale	19.2 (15.6)	31.2 (16.2)	15.5 (11.1)	10.8 (10.8)	<0.001
Total WOMAC score	27.3 (21.3)	43.8 (22.5)	22.5 (14.9)	15.7 (14.7)	<0.001

Totals for each characteristic may not equal n=219 due to missing responses

^aLow HRQoL tertile: AqoL score <0.588; middle HRQoL tertile: AqoL between 0.588-0.796; high HRQoL tertile: AqoL>0.796

Table 2. Results of parameter and diagnostic tests used to predict AQOL disutility scores

	Model					
	A	B	C	D	E	F
	Total WOMAC	WOMAC subscales	WOMAC subscales + interactions	WOMAC function subscale only	Total WOMAC + other variables ^a	Total WOMAC + sex + previous JRS
Variables						
Intercept	-1.571 ^b	-1.591 ^b	-1.706 ^b	-1.569 ^b	-1.534 ^b	-1.510 ^b
Total WOMAC score	0.018 ^b				0.017	0.017 ^b
Pain subscale score		0.007	0.023			
Stiffness subscale score		0.043	0.009			
Function subscale score		0.018 ^c	0.025	0.025 ^b		
Pain score squared			0.010			
Stiffness score squared			0.004			
Function score squared			0.000			
Pain x stiffness interaction			-0.009			
Pain x function interaction			0.004			
Stiffness x function interaction			0.003			
Age					0.001	
Sex					-0.147 ^c	-0.144 ^c
Previous joint replacement surgery					0.138	0.168 ^c
Completed trade/technical/university education					-0.115	
Obesity (Body Mass Index ≥30)					0.004	
Lowest two SEIFA deciles					0.039	
Statistics						
Goodness of fit measures ^d						
Root Mean Squared Error (RMSE)	1.481	1.485	1.510	1.488	1.484	1.484
Mean Absolute Error (MAE)	1.452	1.455	1.474	1.458	1.453	1.455
Akaike's Information Criterion (AIC)	-0.367	-0.353	-0.333	-0.367	-0.359	-0.388
Bayesian Information Criterion (BIC)	-1160.7	-1150.0	-1117.9	-1186.1	-1015.9	-1150.3
Pseudo R-squared	0.398	0.400	0.421	0.395	0.458	0.421
Pitman's test for difference in variance ^e		p=0.197	p=0.001	p=0.390	p<0.001	p=0.081

^a Personal characteristics included in this model: age, sex, previous joint replacement surgery, completed trade/technical/university education, obesity and lowest two SEIFA deciles

JRS: joint replacement surgery; SEIFA: Australian Socio-Economic Indexes for Areas 2006 Index of Relative Socio-Economic Advantage and Disadvantage

^bp<0.05 ^cp<0.01 ^dLower goodness of fit values indicate better model fit ^eCompared to Model A

Table 3. Accuracy of predicted AQOL utility scores

Sample ^a	Actual AQoL	Model A	Model B	Model C	Model D	Model E	Model F
Overall sample							
Mean (SD)	0.64 (0.26)	0.64 (0.68)	0.64 (0.16)	0.64 (0.17)	0.64 (0.16)	0.63 (0.18)	0.64 (0.16)
Median (IQR)	0.71 (0.51-0.84)	0.68 (0.57-0.75)	0.69 (0.57-0.75)	0.68 (0.55-0.77)	0.69 (0.57-0.75)	0.69 (0.56-0.75)	0.69 (0.57-0.75)
Range ^b	-0.04-1.00	-0.12-0.79	-0.12-0.80	0.00-0.82	-0.10-0.79	-0.12-0.82	-0.18-0.81
Low HRQoL tertile							
Mean (SD)	0.33 (0.18)	0.52 (0.19)	0.52 (0.19)	0.51 (0.20)	0.52 (0.19)	0.49 (0.21)	0.51 (0.20)
Median (IQR)	0.32 (0.18-0.51)	0.56 (0.42-0.66)	0.54 (0.42-0.67)	0.53 (0.41-0.66)	0.55 (0.41-0.66)	0.52 (0.36-0.66)	0.54 (0.40-0.67)
Range ^b	-0.04-0.59	-0.12-0.79	-0.12-0.80	0.00-0.82	-0.10-0.79	-0.12-0.80	-0.18-0.81
Middle HRQoL tertile							
Mean (SD)	0.70 (0.06)	0.68 (0.09)	0.68 (0.09)	0.68 (0.10)	0.68 (0.09)	0.68 (0.10)	0.68 (0.09)
Median (IQR)	0.71 (0.65-0.76)	0.70 (0.64-0.75)	0.69 (0.64-0.75)	0.70 (0.63-0.76)	0.71 (0.64-0.76)	0.71 (0.63-0.75)	0.70 (0.64-0.75)
Range ^b	0.59-0.80	0.42-0.79	0.44-0.80	0.45-0.82	0.42-0.79	0.42-0.80	0.44-0.81
High HRQoL tertile							
Mean (SD)	0.88 (0.06)	0.72 (0.09)	0.72 (0.10)	0.73 (0.10)	0.72 (0.09)	0.72 (0.08)	0.72 (0.09)
Median (IQR)	0.87 (0.84-0.92)	0.74 (0.69-0.77)	0.74 (0.69-0.78)	0.75 (0.69-0.80)	0.74 (0.69-0.78)	0.74 (0.69-0.79)	0.74 (0.68-0.78)
Range ^b	0.80-1.00	0.24-0.79	0.20-0.80	0.19-0.82	0.27-0.79	0.48-0.82	0.31-0.81

SD: Standard deviation; IQR: interquartile range; HRQoL: Health-Related Quality of Life

^a Low HRQoL tertile: AQoL score<0.588; middle HRQoL tertile: AQoL between 0.588-0.796; high HRQoL tertile: AQoL>0.796

^b Range for AQoL instrument: -0.04 to 1.00

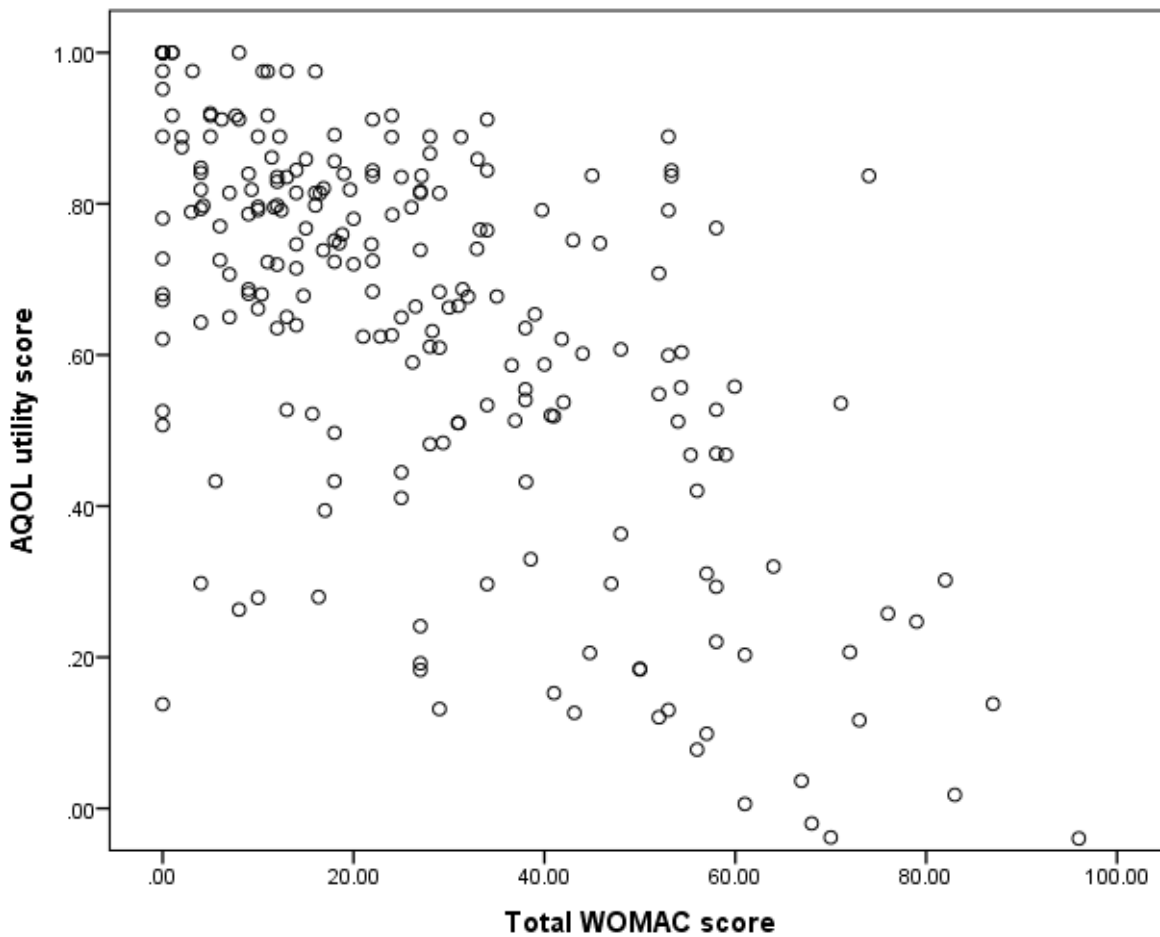


Fig 1 Relationship between total WOMAC scores and actual AQOL utility scores

Lower WOMAC score indicates better health status and higher AQoL score indicates better Health-Related Quality of Life. This figure shows a moderate negative correlation between the instrument scores (Spearman's correlation -0.57).

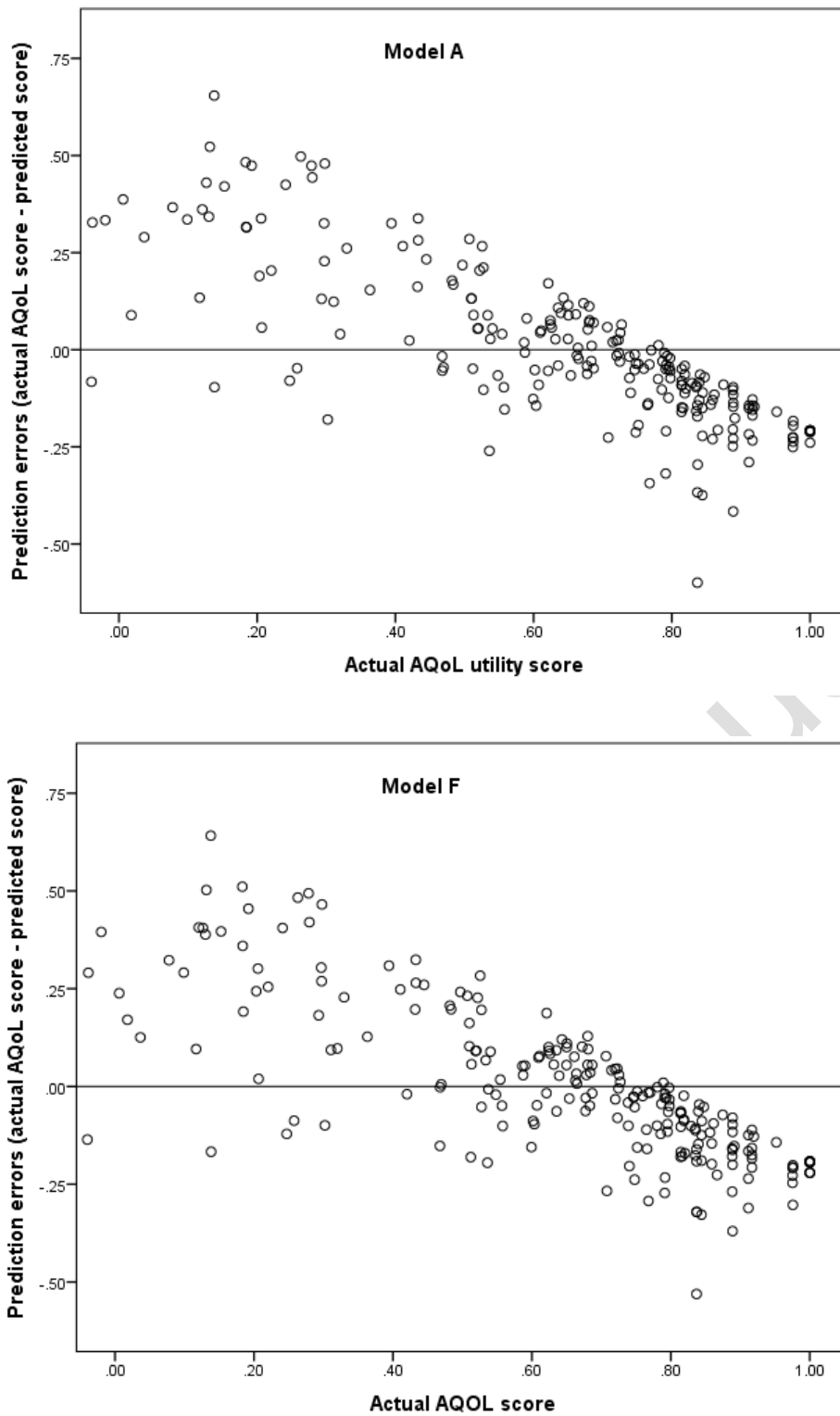


Fig 2 Actual AQoL utility scores versus prediction errors: Comparison between Model A (Total WOMAC score) and Model F (total WOMAC + sex and previous joint replacement surgery)

This figure shows that both models performed similarly; prediction errors were lowest in the mid-range of AQoL utility scores and highest for the extremes, indicating greater bias in relation to low and high AQoL utility scores. Each model tended to over-predict low AQoL scores and under-predict high AQoL scores.

Figure captions

Fig 1 Relationship between total WOMAC scores and actual AQOL utility scores

Lower WOMAC score indicates better health status and higher AQoL score indicates better Health-Related Quality of Life. This figure shows a moderate negative correlation between the instrument scores (Spearman's correlation -0.57).

Fig 2 Actual AQOL utility scores versus prediction errors: Comparison between Model A (Total WOMAC score) and Model F (total WOMAC + sex and previous joint replacement surgery)

This figure shows that both models performed similarly; prediction errors were lowest in the mid-range of AQoL utility scores and highest for the extremes, indicating greater bias in relation to low and high AQoL utility scores. Each model tended to over-predict low AQoL scores and under-predict high AQoL scores.



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Ackerman, IN;Tacey, MA;Ademi, Z;Bohensky, MA;Liew, D;Brand, CA

Title:

Using WOMAC Index scores and personal characteristics to estimate Assessment of Quality of Life utility scores in people with hip and knee joint disease

Date:

2014-10

Citation:

Ackerman, I. N., Tacey, M. A., Ademi, Z., Bohensky, M. A., Liew, D. & Brand, C. A. (2014). Using WOMAC Index scores and personal characteristics to estimate Assessment of Quality of Life utility scores in people with hip and knee joint disease. *QUALITY OF LIFE RESEARCH*, 23 (8), pp.2365-2374. <https://doi.org/10.1007/s11136-014-0667-y>.

Publication Status:

Accepted manuscript

Persistent Link:

<http://hdl.handle.net/11343/41027>