BJR

HIP

Can pain and function be distinguished in the Oxford Hip Score in a meaningful way?

AN EXPLORATORY AND CONFIRMATORY FACTOR ANALYSIS

K. K. Harris, A. J. Price, D. J. Beard, R. Fitzpatrick, C. Jenkinson, J. Dawson

From University of Oxford, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Oxford, United Kingdom

K. K. Harris, BSc, MSc, DPhil, Postdoctoral Research Associate
A. J. Price, MA, DPhil, FRCS(Orth), Professor of Orthopaedic Surgery
D. J. Beard, MA, MSc, DPhil, Professor of Musculoskeletal Sciences
University of Oxford, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Windmill Road, Oxford OX3 7LD, UK.

R. Fitzpatrick, MA, MSc, PhD, Professor of Public Health and Primary Care
C. Jenkinson, MA, MSc, DPhil, Professor of Health Services Research
J. Dawson, MA, MSc, DPhil, University Research Lecturer
University of Oxford, Nuffield Department of Population Health, Oxford OX3 7LF, UK.

Correspondence should be sent to A. J. Price; e-mail: andrew.price@ndorms.ox.ac.uk

10.1302/2046-3758.311.2000313 $2.00

Bone Joint Res 2014;3:305–9.
Received 11 April 2014; Accepted 6 August 2014

Objective
The objective of this study was to explore dimensionality of the Oxford Hip Score (OHS) and examine whether self-reported pain and functioning can be distinguished in the form of subscales.

Methods
This was a secondary data analysis of the UK NHS hospital episode statistics/patient-reported outcome measures dataset containing pre-operative OHS scores on 97 487 patients who were undergoing hip replacement surgery.

Results
The proposed number of factors to extract depended on the method of extraction employed. Velicer's Minimum Average Partial test and the Parallel Analysis suggested one factor, the Cattell's scree test and Kaiser-over-1 rule suggested two factors. Exploratory factor analysis demonstrated that the two-factor OHS had most of the items saliently loading either of the two factors. These factors were named 'Pain' and 'Function' and their respective subscales were created. There was some cross-loading of items: 8 (pain on standing up from a chair) and 11 (pain during work). These items were assigned to the 'Pain' subscale. The final 'Pain' subscale consisted of items 1, 8, 9, 10, 11 and 12. The 'Function' subscale consisted of items 2, 3, 4, 5, 6 and 7, with the recommended scoring of the subscales being from 0 (worst) to 100 (best). Cronbach's alpha was 0.855 for the 'Pain' subscale and 0.861 for the 'Function' subscale. A confirmatory factor analysis demonstrated that the two-factor model of the OHS had a better fit. However, none of the one-factor or two-factor models was rejected.

Conclusion
Factor analyses demonstrated that, in addition to current usage as a single summary scale, separate information on pain and self-reported function can be extracted from the OHS in a meaningful way in the form of subscales.

Cite this article: ***Bone Joint Res*** **2014;3:305–9.**

Keywords: Patient-reported outcomes; Osteoarthritis; Hip replacement; Outcomes assessment

Article focus
- This study aims to explore if self-reported pain and functioning can be distinguished from the Oxford Hip Score (OHS) in the form of subscales.

Key messages
- Exploratory factor analysis and confirmatory factor analysis demonstrated that the OHS can be used as a summary scale and in the form of pain and functional subscales.
- 'Pain' subscale consists of items 1, 8, 9, 10, 11 and 12. The 'Function' subscale consisted of items 2, 3, 4, 5, 6 and 7.
- The recommended scoring of the subscales is from 0 (worst) to 100 (best).

Strengths and limitations
- To our knowledge, the OHS is the only hip-specific instrument that has been subjected to such a high level of scrutiny in the population of patients undergoing hip replacement surgery.

- Consistent factor-analytic results, based on large-scale data, provide convincing evidence in favour of the use of the OHS and its subscales.
- Further research could usefully focus on evaluating their construct validity and responsiveness.

Introduction

Hip replacement surgery is an effective treatment for hip osteoarthritis (OA), resulting in improved mobility, pain relief, and overall health-related quality of life (HRQoL). In the US more than 400 000 hip replacements are performed per year,[1] with more than 86 000 patients undergoing this procedure per year in England and Wales.[2] The success of hip replacement is often measured using patient-reported outcome measures (PROMs). In this context, PROMs aim to offer a valid and reliable representation of patients' perceptions of their quality of life in relation to their hip problem.

The Oxford Hip Score (OHS) is a 12-item PROM developed to assess patients' perceptions of their HRQoL in those undergoing hip replacement surgery. It was designed to be used as a single composite scale, which reflected patients' perception of pain and functional impairment arising from their hip. In this form, it has proven to be valid, reliable and responsive.[3,4] Originally, Likert-type responses for each item were scored 1 to 5, with a summary score of 12 (best) to 60 (worst). Subsequently, the scoring method changed, with the recommendation made to score each question from 0 to 4, with a summary score of 0 (worst) to 48 (best).[5] The OHS items were generated by conducting qualitative interviews with patients before and after undergoing hip replacement surgery, which suggested that pain and functional disability were generally inextricably linked. In 2009, the OHS[3] and the EQ-5D[6] (a generic measure of health status) were adopted as a part of the UK national patient-reported outcome measures programme (NHS PROMs) as a primary outcome measure for patients undergoing hip replacement.

A decade ago, a study suggested that the OHS could be analysed in the form of pain and functional subscales.[7] However, these findings were based on data from a single centre and the exploratory factor analysis (EFA) was based on a Pearson correlation matrix. Both EFA and confirmatory factor analysis (CFA) assume normally distributed data when using Pearson-product moment correlation, but these are not robust when instruments with Likert-type responses are used.[8,9] In such situations, it is now recognised that EFA and CFA should be based on the matrix comprising polychoric correlations,[10] which is also robust to underlying non-normality.[11]

In this paper we explore the factor structure of the OHS using a large national dataset and using the same methodology that we applied in our recent publication, which investigated pain and functional subscales in the Oxford Knee Score.[12] We employed a polychoric correlation matrix in conducting both an EFA and a CFA to explore whether pain and function can be distinguished in the OHS in a meaningful way.

Table I. Procedures undertaken

Procedure*	N (%)
Primary THR	76 009 *(78)*
Primary TPR of the head of the femur	257 *(0.3)*
Primary hybrid prosthetic hip replacement	11 166 *(11.5)*
Other primary hip replacement	7 *(0)*
Revision total hip replacement	7203 *(7.4)*
Hip resurfacing	2179 *(2.2)*
Missing or unclear what type of procedure was performed	666 *(0.7)*

*Procedure field was coded according to the relevant classification of interventions and procedures codes[13]
Patients who had more than one procedure were classified as 'mixed'
TPR, total prosthetic replacement

Materials and Methods

A secondary data analysis of the NHS hospital episode statistics/PROMS dataset on 97 487 patients who underwent hip replacement from April 2009 to December 2011 was performed. The sample consisted of 39 969 men (41%) and 57 518 women (59%) with a mean age of 68 (14 to 100). An EFA was performed using IBM SPSS 20 (Armonk, New York). LISREL (Chicago, Illinois) software was used to conduct the CFA. Available information on procedures undertaken is presented in Table I. Procedures were coded according to the relevant Classification of Interventions and Procedures codes.[13] Where observations did not contain any procedure codes or contained contradictory codes (i.e. codes for both primary and revision procedures, or codes for THR and hybrid replacements), these observations were classed as missing or unclear surgical procedures.

Statistical analysis. Factor analysis is a procedure that is widely recommended and used in the construction and validation of PROMs.[14-16] The main goal of factor analysis is to explain the observed variables (in the case of PROMs, items on a scale) by a smaller number of latent variables (factors).[14,17]

EFA and CFA are two general techniques for conducting a factor analysis and the method used depends on the purpose of the study. Normally, EFA may be used to identify the underlying structure of a measure or to discard redundant items. If, on the other hand, the underlying structure of the measure is already known and the goal is to check if this structure holds across groups (invariance), CFA is the method of choice. When conducting EFA, there is often no *a priori* knowledge about the relationships between the latent and observed variables, and the purpose of EFA is to identify latent factor solutions that are able to explain the pattern of correlations or covariances between the observed variables. Alternatively, CFA can be used statistically to test the fit of an *a priori* hypothesised structure of an instrument. Usually, several competing models that are based on theory and/or empirical

Table II. Results of two-factor exploratory factor analysis (abbreviated item content next to question number)

	Factor 1	Factor 2
Q5 (Shopping)	0.783	0.021
Q3 (Transport)	0.771	0.030
Q4 (Dressing)	0.758	-0.070
Q7 (Stairs)	0.750	0.075
Q2 (Washing)	0.733	-0.003
Q6 (Walking)	0.445	0.289
Q10 (Sudden pain)	-0.124	0.833
Q12 (Night pain)	-0.084	0.779
Q1 (Pain)	0.157	0.637
Q8 (Standing up)	0.363	0.484
Q11 (Work)	0.428	0.473
Q9 (Limping)	0.283	0.422

research are tested for good fit. Nevertheless, when both EFA and CFA are conducted, it is important to consider the type of measurement scale represented by the instrument. Measures that use Likert-type responses (such as the OHS) provide a categorical (ordinal) description of an underlying continuous variable. In this case, the EFA and CFA should be based on the matrix of polychoric correlations (rather than Pearson), which is also robust to underlying non-normality.[11]

EFA. As the goal of EFA was to identify the number of factors that the measure was assessing, principal axis factoring (PAF) was chosen as the extraction method.[14,18] The decision on the number of factors to extract was assessed by using several methods: Kaiser-over-1 (K-over-1) rule,[19] the scree test,[20] Velicer's minimum average partial (MAP) test[21] and Horn's parallel analysis (PA).[22] Factors were rotated using the oblique rotation method (promax). Items were assigned to a factor if their loading on a factor was > 0.3.[23]

CFA. CFA was conducted to test the fit of the two hypothesised factor models.

Model 1 hypothesised that all 12 items characterise the single underlying factor. This model was tested as the one-factor model corresponds to the conceptual basis of the OHS.[3] The acceptability of this model was further confirmed by evidence of its high internal consistency and on the basis of the number of extracted factors in this study using some of the most commonly recommended methods, namely Horn's PA[22] and Velicer's MAP test.[21]

Model 2 tested two first-order correlated factors as indicated by other commonly recommended methods: the scree test[20] and K-over-1 rule.[19]

As the data were ordinal and non-normal, the diagonally-weighted least squares (DWLS) method, based on polychoric correlations and asymptomatic covariances, was used for extraction.[11] No modification indices were considered. The DWLS method was used to estimate relationships between items and factors. This method works best with large datasets containing ordinal data. The following fit indices were considered satisfactory: root mean square error of approximation (RMSEA) < 0. 05 close fit, < 0.08 good fit, < 0.1 satisfactory fit; comparative fit index (CFI) > 0.95, and standard root mean square residual (SRMR) < 0.08 good, < 0.05 close fit.[24] Cronbach's alpha[25] was used to test the internal consistency of the subscales.

Results

EFA. Depending on the method employed, one- or two-factor models of the OHS were suggested. Velicer's MAP test[21] and the PA[22] suggested one-factor. The scree test[20] and K-over-1[19] rule suggested two factors, with the second eigenvalue of 1.02. The first two factors explained 64% of the variance. Table II demonstrates the factor loadings for the two-factor solution.

The two-factor EFA revealed that items 2 (have you had any trouble with washing and drying yourself (all over) because of your hip?), 3 (have you had any trouble getting in and out of a car or using public transport because of your hip? (whichever you tend to use)), 4 (have you been able to put on a pair of socks, stockings or tights?), 5 (could you do the household shopping on your own?), 6 (for how long have you been able to walk before pain from your hip becomes severe? (with or without a stick)) and 7 (have you been able to climb a flight of stairs?) loaded saliently on Factor 1. This factor was labelled 'Function'. Items 1 (how would you describe the pain you usually have from your hip?), 9 (have you been limping when walking, because of your hip?), 10 (have you had any sudden, severe pain – 'shooting', 'stabbing' or 'spasms' – from the affected hip?) and 12 (have you been troubled by pain from your hip in bed at night?) loaded significantly on the Factor 2. This factor was labelled 'Pain'. Items 8 (have you been able to put on a pair of socks, stockings or tights?) and 11 (how much has pain from your hip interfered with your usual work (including housework)?) were markedly cross-loading. These items were assigned to the 'Pain' factor.

Cronbach's alpha was 0.861 for the 'Function' subscale and 0.855 for the 'Pain' subscale.

CFA. CFA (Table III) indicated that the two-factor model of the OHS demonstrated marginally better fit than the one-factor model. However, neither of the models was rejected. The results of EFA and CFA demonstrate that the OHS can be used both as a single summary score and in the form of Pain and Function component subscales. Items 1, 8, 9, 10, 11 and 12 can be grouped into a 'Pain' component and items 2, 3, 4, 5, 6 and 7 can be grouped into the 'Function' component. We recommend scoring the two component subscales on a scale 0 (worst) to100 (best).

Discussion

The aim of this study was to explore if pain and function can be distinguished in the OHS in a meaningful way, by conducting both EFA and CFA. EFA and CFA demonstrated that the OHS can be considered as consisting of either one or two factors.

Table III. Summary of confirmatory factor analysis fit measures for one- and two-factor model

Factor	χ^2	df	CFI	SRMR	RMSEA	RMSEA 90% CI
1	6251	54	1.00	0.052	0.034	(0.034 to 0.035)
2	4114	53	1.00	0.043	0.028	(0.027 to 0.029)

χ^2, chi squared; df, degrees of freedom; CFI, comparative fit index; SRMR, standard root mean square residual; RMSEA, root mean square error of approximation; 90% CI, 90% confidence interval

In our previous paper,[12] we have demonstrated that the OKS, which was developed in a similar way to the OHS, can be used both as a summary scale and in the form of pain and functional subscales. As with the OHS, the OKS had items that loaded significantly (above 0.3) on both factors.[12] This is expected, as in certain contexts (such as advanced OA or around the time of arthroplasty), pain and function have been shown to have considerable overlap, although some distinction can still be made between the two.[3,26-29] As stated in our previous paper, the cross-loading of the items supports this interpretation as the items demonstrate that they are tapping into these different (yet overlapping) concepts.[12]

The findings in our study are, in fact, broadly similar to those from a previous study by Norquist et al[7] where data were analysed from patients from one institution undergoing routine hip replacement surgery. The EFA in that study, with varimax rotation, demonstrated the same subscale structure to our own EFA analysis. Due to the large study sample, the CFA demonstrated that the chi-square value was high and statistically significant ($p < 0.05$) and alternative fit indices (CFI, SRMR, RMSEA) were considered.[24] As with the OKS analysis, the CFA demonstrated excellent fit for both one- and two-factor models and, if anything, slightly favoured the two-factor model.

The work in this paper provides further evidence that contributes towards the construct validity of the OHS. Furthermore, the two derived subscales allow for additional data analysis to be conducted with the OHS in terms of self-reported pain and function. Clinical studies specifically focused on assessing either pain or function could use these subscales as primary outcome measures of interest and to calculate required sample sizes accordingly. However, while these subscales have demonstrated good construct validity and high internal consistency, further research could usefully focus on evaluating their construct validity and responsiveness.

References

1. **No authors listed.** Agency for Healthcare Research and Quality, 2012, Facts and Figures. Statistics on Hospital-Based Care in the United States, 2009. www.hcup-us.ahrq.gov/reports/factsandfigures/2009/exhibit3_1.jsp. (date last accessed 25 July 2014).
2. **No authors listed** National Joint Registry for England and Wales, 2013. 10th Annual Report. Hemel Hempstead, Hertforshire, UK. http://www.njrcentre.org.uk/njrcentre/Portals/0/Documents/England/Reports/10th_annual_report/NJR%2010th%20Annual%20Report%202013%20B.pdf (date last accessed 25 July 2014).
3. **Dawson J, Fitzpatrick R, Carr A, Murray D.** Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg [Br]* 1996;78-B:185–190.
4. **Dawson J, Fitzpatrick R, Churchman D, Verjee-Lorenz A, Clayson D.** User Manual for the Oxford Hip Score (OHS). Isis Outcomes. http://isis-innovation.com/wp-content/uploads/2014/09/User-Manual-OHS-Contents.pdf (date last accessed 18 October 2014).
5. **Murray D, Fitzpatrick R, Rogers K, et al.** The use of the Oxford hip and knee scores. *J Bone Joint Surg [Br]* 2007;89-B:1010–1014.
6. **Brooks, R.** EuroQol: the current state of play. *Health Policy* 1996;37:53–72.
7. **Norquist JM, Fitzpatrick R, Dawson J, Jenkinson C.** Comparing alternative Rasch-based methods vs raw scores in measuring change in health. *Med Care* 2004;42(1 Suppl):I25–I36.
8. **Bollen KA.** Structural equations with latent variables. Wiley-Interscience, 1989.
9. **Nunnally JC, Bernstein, IH.** *Psychometric theory.* New York: McGraw-Hill, 1994.
10. **Olsson U.** Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika* 1979;44:443–460.
11. **Flora DB, Curran PJ.** An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol Methods* 2004;9:466–491.
12. **Harris K, Dawson J, Doll H, et al.** Can pain and function be distinguished in the Oxford Knee Score in a meaningful way? An exploratory and confirmatory factor analysis. *Qual Life Res* 2013;22:2561–2568.
13. **No authors listed.** Health and social care information centre. http://systems.hscic.gov.uk/data/clinicalcoding/codingstandards/opcs4 (date last accessed 24 October 2014).
14. **Floyd FJ, Widaman KF.** Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment* 1995;7:286–299.
15. **Mokkink LB, Terwee CB, Patrick DL, et al.** The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–745.
16. **Streiner DL, Norman GR.** *Health measurement scales: a practical guide to their development and use.* Fourth ed. Oxford: Oxford University Press, 2008.
17. **Tabachnick BG, Fidell LS.** Using multivariate statistics. Sixth ed. Pearson Education Limited, 2001:659-730.
18. **Norman GR, Streiner DL.** *Biostatistics: the Bare Essentials.* PMPH USA Ltd, 2008.
19. **Kaiser HF.** The application of electronic computers to factor analysis. *Educational and Psychological Measurement* 1960;20:141–151.
20. **Cattell RB.** The scree test for the number of factors. *Multivariate Behavioral Research,* 1966;1:245–276.
21. **Velicer WF.** Determining the number of components from the matrix of partial correlations. *Psychometrika* 1976;44:321–327.
22. **Horn JL.** A rationale and test for the number of factors in factor analysis. *Psychometrika* 1965;30:179–185.
23. **Kline P.** *An easy guide to factor analysis.* Routledge, 1993.
24. **Kline RB.** *Principles and practice of structural equation modeling*: The Guilford Press, 2010.
25. **Cronbach LJ.** Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334.
26. **Dawson J, Fitzpatrick R, Murray D, Carr A.** Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg [Br]* 1998;80-B:63–69.
27. **Gooberman-Hill R, Woolhead G, MacKichan F, et al.** Assessing chronic joint pain: lessons from a focus group study. *Arthritis Rheum* 2007;15;57:666–671.
28. **Heuts PH, Vlaeyen JW, Roelofs J, et al.** Pain-related fear and daily functioning in patients with osteoarthritis. *Pain* 2004;110:228–235.
29. **Boersma K, Linton SJ.** How does persistent pain develop? An analysis of the relationship between psychological variables, pain and function across stages of chronicity. *Behav Res Ther* 2005;43:1495–1507.

Funding statement:

- D. J. Beard reports a grant received by the University of Oxford from the Health Technology Assessment National Institute for Health Research that is related to this article, as well as receiving personal consultancy and lecture fees from Biomet and Smith & Nephew that are not related to this article. J. Dawson reports grants from the Department of Health, Health Technology Assessment and The European Brain Council to the University of Oxford, as well as personal consultancy fees from Isis Innovations, neither of which are related to this article. A. J. Price reports educational consultancy fees received from Biomet, Stryker, DePuy and Smith & Nephew, none of which is related to this article.

Author contributions:

- K. K. Harris: Obtained and cleaned the dataset, Performed data analysis, Wrote the first draft of the paper
- A. J. Price: Obtained and cleaned the dataset, Supervised the project
- D. J. Beard: Obtained and cleaned the dataset, Supervised the project, Contributed to the interpretation of the findings
- R. Fitzpatrick: Contributed to the methodological development
- C. Jenkinson: Contributed to the methodological development
- J. Dawson: Contributed to the methodological development

ICMJE Conflict of Interest:

- None declared