Original Article

The validity and reliability of the modified forgotten joint score

Patrick G. Robinson^a, Conor S. Rankin^{b,*}, Jonathan Lavery^c, Iain Anthony^c, Mark Blyth^c, Bryn Jones^c

^a Trauma and Orthopaedics Department Royal Infirmary Edinburgh, UK

^b University of Edinburgh, 3/2f2 Warrender Park Terrace, Edinburgh, EH9 1JA, UK

^c Glasgow Royal Infirmary Orthopaedic Research Department, Glasgow, UK

ARTICLE INFO

Keywords:

Arthroplasty

Hip

Knee

Patient reported outcome measures

ABSTRACT

We aim to validate the “Modified Forgotten Joint Score” (MFJS) as a new patient-reported outcome measure (PROM) in hip and knee arthroplasty, against the UK's gold standard Oxford Hip and Knee Scores (OHS/OKS).

The original Forgotten Joint Score (FJS) (12 items) was created to assess post-arthroplasty joint awareness. We modified the FJS to 10-items to improve its reliability.

Postal questionnaires were sent out to 400 total hip or knee replacement (THR/TKR) patients who were 1–2 years' post-op, along with the OHS/OKS. Data, collected from the 212 returned questionnaires (53% response rate), was analysed in relation to construct and content validity. A sub-cohort of 77 patients took part in a test-retest repeatability study, to assess reliability of the MFJS.

The MFJS proved to have an increased discriminatory power in high-performing patients in comparison to the OHS and OKS. 30.8% of TKR patients ($n = 131$) scored highly (87.5% or more) in the OKS compared to just 7.69% in the MFJS TKR patients. The MFJS proved to have increased test-retest repeatability, based upon its intra-class correlation coefficient of 0.968 compared to the Oxford's 0.845, $p < 0.001$.

The MFJS is a more relevant tool, compared to the FJS, with greater discrimination in the assessment of well performing hip and knee arthroplasties in comparison to the OHS/OKS.

1. Introduction

Hip and Knee Joint replacement surgery has proven to be highly successful in improving patient's pain and function.¹ Surgical techniques and advancements in technology have led to joint arthroplasty evolving to the benefit of the patients. However, it is vital to evaluate patient's outcome post-operatively to assess the levels of improvement in joint function after joint replacement in order to demonstrate efficacy and monitor patient progress. As well as measuring simple surgical parameters, assessment of function has been recognised as an essential evaluation tool which has led to the development of patient-reported outcome measures (PROMS). Technological advancements in surgical techniques or implant design need to be evaluated against current standards, but often gains in performance are small and these may not be able to be picked up by current scoring systems/PROMS which may be restricted by a ceiling effect.^{2,3}

The Forgotten Joint Score (FJS) was developed by Behrend et al in 2007. This new PROM measures a very appealing concept; the ability for a patient to forget about their artificial joint in everyday life.⁴

Behrend et al believed the optimal outcome after a total knee or total hip replacement (TKR/THR) was for a patient to be “unaware” that they had a prosthetic joint.

In the UK, the gold-standard PROM for knee and hip arthroplasties are the Oxford Knee and Hip Scores (OKS/OHS). The Oxford 12-item Knee and Hip Questionnaires were developed in 1998 by Dawson et al as a self-administered, disease and site specific questionnaire, specifically for hip and knee arthroplasty patients.⁵ Since then, the OKS/OHS have been used extensively throughout the UK and have been translated into several languages for use globally.^{2,3,5–10}

The objective of our study was to assess the usefulness in everyday orthopaedic practice of our modification of the Forgotten Joint Score – the Modified Forgotten Joint Score (MFJS) by validating it against the OHS and OKS. In our evaluation we assessed the construct and content validity of both questionnaires along with the reliability by comparing the test-retest repeatability of the two scores.¹¹

* Corresponding author.

E-mail addresses: probins2@ed.ac.uk (P.G. Robinson), conorrankin@doctors.org.uk (C.S. Rankin), jonathan.lavery@nhs.net (J. Lavery), iain.anthony@ggc.scot.nhs.uk (I. Anthony), mark.blyth@ggc.scot.nhs.uk (M. Blyth), bryn.jones@ggc.scot.nhs.uk (B. Jones).

<https://doi.org/10.1016/j.jor.2018.03.029>

Received 3 February 2018; Accepted 25 March 2018

Available online 27 March 2018

0972-978X/ © 2018 Prof. PK Surendran Memorial Education Foundation. Published by Elsevier, a division of RELX India, Pvt. Ltd. All rights reserved.

2. Background

2.1. The Oxford hip and knee questionnaires

The Oxford Knee and Hip Questionnaires contain 12 items, which assesses a patient's pain and function by asking them to answer a range of questions. Each question is scored on five point scale from 0 to 4. The score for each question is added together to give an overall score out of 48.

In the Oxford Score, high scores indicate good outcomes. A high score indicates higher level of function and less pain. For ease of interpretation we use the term “ceiling-effect” to refer to the best possible score (48) and “floor-effect” to refer to the worst possible score (0).

2.2. The original forgotten joint score

The original FJS is a 12-item questionnaire which asked patients to answer questions based upon their “awareness” of their artificial joint during everyday activities. The questionnaire differs from that of the OKS and OHS as it is not site specific, covering both hip and knee arthroplasty patients in the one questionnaire. The FJS scales answers from 1 to 5. These scores add up to give a score out of 60, which is then converted into a percentage.

Behrend et al stated that in an initial validation study of the FSJ-12, it outperformed the Western Ontario and McMaster Universities (WOMAC) osteoarthritis index in several areas, including discriminatory power and combating the ceiling effect.⁴

2.3. Reliability and validity

The reliability of a questionnaire is defined as the ability of a test to “yield the same results on repeated trials under the same conditions”.^{4,5}

Validity in our case refers to the ability of a questionnaire to measure the construct it is intended to measure. It can be determined by measuring the correlation between two study groups as well as determining the frequency distribution of scores, along with the ceiling and floor effects.¹² A ceiling effect occurs when a patient achieves a very high score in a questionnaire and would be unable to show improvement in subsequent questionnaires despite improving clinically.¹³

2.4. The pilot study

We performed an initial pilot study in 2013 comparing the FJS to the OKS/OHS. The FJS proved to have increased sensitivity, especially in the well performing patients in comparison to the OHS and OKS. However, some areas of missing data in the FJS responses were observed (Table 1).

We used a further pilot group of TKR and THR patients (n = 25) to gain feedback with regards to their understanding of the original questions, along with suggestions on proposed alternative questions.

We put into place a number of modifications which are summarised below:

• Removal of Q.11. . . awareness taking a walk/hiking?

In 5.71%(THR) and 3.91%(TKR) of respondents the answer to this question was missing, while also showing a strong correlation with Q.3 – implying patients often left this question out or answered it the same as Q.3. Both factors lead to the belief that the question was redundant and therefore should be removed.

• Removal of Q.12. . . awareness when playing your favourite sport?

This question posed a significant problem based on the large percentage of missing data associated with it; 39.13% (THR) and

Table 1

FJS – Overall Missing Data.

FJS-12	Missing Data THR	Missing Data TKR
Overall Missing Data	6.82%	6.69%
1. Awareness in bed at night?	0.29%	0.69%
2. Awareness sitting on a chair for more than 1 h?	0.28%	0.92%
3. Awareness when you are walking for more than 15 min?	1.63%	0.69%
4. Awareness taking a bath/shower?	0.82%	1.61%
5. Awareness travelling in a car?	1.91%	1.84%
6. Awareness climbing stairs?	0.82%	1.84%
7. Awareness walking on uneven ground?	1.91%	2.07%
8. Awareness squatting?	23.10%	25.75%
9. Awareness standing for longer?	1.09%	1.61%
10. Awareness doing housework or gardening?	5.16%	2.99%
11. Awareness taking a walk/hiking?	5.71%	3.91%
12. Awareness when playing your favourite sport?	39.13%	47.82%

47.82%(TKR) patients failed to answer this question. This, along with the fact that playing sport is not a popular activity within the arthroplasty population, meant it was decided to omit this question from the modified questionnaire.

• Rewording of Q.8. . . awareness when squatting?

This question again was completed poorly with 23.10% (THR) and 25.75% (TKR) of patients failing to answer it. In discussions with our patient group (n = 25), feedback showed that many patients did not fully understand the intended activity being asked and therefore failed to answer the question. With this in mind we decided to amend the question to;

• ○ . . . awareness when squatting/crouching?

• Rewording of Q.10. . . awareness when doing housework/gardening?

This question had a percentage of missing data of 5.16% (THR) and 2.99% (TKR), with almost twice as many males failing to answer it as females. Therefore, it was decided to modify this question to;

• ○ Awareness when doing housework or gardening or the most strenuous activity you do around the home?

2.5. Scoring system change

The original FJS scored each answer on a range from 1 to 5. This was added up and converted to a percentage to give an overall score (20–100%). The higher the percentage the better the outcome. The new, Modified FJS (MFJS) is now scored on a range from 0 to 4. This means it gives a more easily understood percentage range of 0–100%.

These changes created a Modified Forgotten Joint Score (MFJS) with an aim to maintain the FJS's increased discriminatory power while ameliorating the large amount of missing data seen in the pilot study.

3. Methods

The study population, to assess the reliability and validity of the MFJS, consisted of 400 consecutive patients who had received either a THR or TKR in a university teaching hospital. 200 patients underwent THR and 200 underwent a TKR. Patients were between 1–2 years post-arthroplasty. These patients were sent out the following postal

Fig. 1. Frequency of Scores against Knee Oxford Knee Score.

questionnaires:

- Modified Forgotten Joint Score
- Oxford Hip or Knee Score
- Visual pain analogue scale

Of the 400 questionnaires sent out to patients, 212 completed forms were returned, (53% response rate) consisting of 131 TKR (61.8%) and 81 THR (38.2%).

3.1. Validity

Validity was assessed by comparison of the MFJS to the OKS/OHS and by determining the frequency distribution of scores, as well as the ceiling and floor effects associated with each questionnaire. The ceiling effect for the MFJS is a 100% (40/40), while the ceiling effect for the OHS/OKS is 48/48. It was hypothesised that the OHS/OKS had a low sensitivity in the well performing patients, based upon its high ceiling effect observed in the pilot study in 2013 (Figs. 1–3).

We classified an “excellent” result in the OKS/OHS as per Kalairajah et al’s method, where 42–48 is referred to as excellent. The equivalent scoring bracket in the MFJS was 87.5–100%.¹⁴

3.2. Reliability

All patients who returned questionnaires were sent additional identical questionnaires to be completed 30 days after responding. In total 140 patients (80 TKR and 60 THR) were involved in the test-retest repeatability assessment. Of the 140 patients, 77 (43 TKR/34 THR) patients completed and returned forms, which represents a response

rate of 55%. To assess the reliability, or the test-retest repeatability, of the MFJS, we calculated the intraclass-correlation coefficient (ICC).

Following the removal and amendment of questions when forming the MFJS, the internal consistency of the items needed to be assessed. Cronbach’s Alpha assesses the homogeneity of the items within the questionnaire and measures reliability along with the ICC.⁵ By calculating this, the inter-item correlation for each question could be determined and this highlighted any question that seemed to deviate from the area of interest. To group answers together, we used the parameters stipulated in Dunbar’s paper: good, very good and excellent were used.¹⁵

To minimise the potential subjectivity in asking patients about their “awareness” of their artificial joint, all patients completed a visual pain analogue scale on a range from 0 to 10, with 0 being no pain at all and 10 being unbearable pain. With these analogue scores the Pearson’s correlation was calculated between the overall MFJS scores and the corresponding pain scores.

4. Results

The percentage of missing data for the MFJS is shown in Table 4 for both TKR and THR. The MFJS had 6.12% less missing data in the TKR and 5.42% less in the THR compared to the FJS. The frequency of distribution of the OKS and OHS had a strong negative skew, towards the ceiling of the score, which can be seen in Figs. 1 and 3. This was confirmed by 30.77% of TKRs scoring 42–48 in the OKS and 42.25% of THRs scoring 42–48 in the OHS. This was compared to 7.69% (TKR) and 12.68% (THR) scoring 87.5–100 in the MFJS. Figs. 2 and 4 illustrate the more evenly distributed nature of the MFJS.

When using Pearson’s Correlation of the MFJS and OKS against

Fig. 2. Frequency of Scores against Knee Oxford Hip Score.

Fig. 3. Frequency of Scores against Knee MFJS Score.

pain, we reported less correlation between the MFJS and pain compared to OKS and pain; $R = 0.602$ ($p < 0.0001$) (Fig. 5) and $R = -0.88$ ($p < 0.0001$) (Fig. 6) respectively. Although theoretically asking a patient how aware they are of their prosthetic joint is equivalent to asking them if they have pain in their joint, using Pearson's correlation has disproved this for the MFJS and shows a differentiation between awareness and pain. The OKS correlated more with pain, which may have been predicted, as five out of the 12 questions in the OKS directly ask about pain.

The internal consistency of the MFJS was 0.952 compared with 0.943 in the OKS ($p < 0.0001$) and these were both classified as “excellent” in Cronbach's Alpha calculation (Fig. 7). This confirmed our prediction that the OKS had impressive internal consistency. Encouragingly, it also demonstrated that the MFJS had “excellent” results, despite the removal of two questions.

Analysis of reliability of the MFJS with intra-class coefficient showed the MFJS scored within the “excellent” range for reliability compared to the OHS/OKS, which scored within the “good” bracket (Fig. 8). This finding of the 10-item MFJS being more repeatable than the 12-item Oxford score is somewhat surprising as generally the questionnaire with the greater number of items has a greater level of reliability.

5. Discussion

The original FJS was created as a means of providing a new assessment tool after hip or knee arthroplasty. It explored an exciting concept of a patient's “awareness” of their prosthetic joint, as opposed to merely concentrating on pain or function of the joint. It aimed to

provide a more sensitive tool, especially in the assessment of well-performing patients. From our pilot study in 2013, the original FJS seemed to provide an increased discriminatory tool based upon its frequency of distribution and ceiling effects.

However, there were problems with recurring incomplete data that inhibited the FJS. A study by Jenkinson et al using a computerised method of imputing values, which can be applied to a variety of questionnaires, suggested that if three or more questions were unanswered then that patient's score should be omitted.¹⁶ We addressed this issue by modifying some questions slightly to improve understanding and removing unnecessary questions.

The more even distribution of values shown with the new MFJS and lower ceiling effect, compared to the OHS suggest that it provides a more sensitive tool in the assessment of hip and knee arthroplasty patients post-operatively. Although we found the same ceiling effect for the OKS and the MFJS (Table 2), there was a negative skew of scores in the OKS shown with 30% of patients scoring between 42–48. This would suggest the OKS is less effective in differentiating the better performing patients. Takeuchi et al also stated that “notable ceiling effects of the OKS were reported”,² while Jenny et al observed a ceiling effect of 7% in the OKS.³

The literature suggests that the reliability of the OHS/OKS is credible; Eun et al showed the OHS to have a high level of reliability, with an ICC of 0.85.¹⁰ The same ICC (0.85) was observed in the assessment of the OKS by Takeuchi et al.² We also had similar results, however the MFJS was shown to be more reliable and the new modifications didn't alter the internal consistency of the questionnaire based upon the Cronbach's Alpha of the 10-questions (Table 3).

The percentage of unanswered questions dramatically decreased

Fig. 4. Frequency of Scores against Hip MFJS Score.

Fig. 5. Pain Scores against Oxford Knee Score.

Fig. 6. Pain scores against Modified Forgotten Joint Score.

Fig. 7. Cronbach's Alpha of MFJS and OKS.

Fig. 8. Intraclass Correlation Coefficient of MFJS and OKS.

with the modifications put in place and our modifications have reduced the flaws which hindered the original FJS, whilst maintaining the positive aspects of the questionnaire.

We acknowledge that our response rate for the questionnaires is below what some would regard as adequate. Unfortunately, this is a recognised pitfall of questionnaires including the OKS.^{17, 18} The range of accepted response rates in the literature is vast. Both Draugalis et al and Babbie state the general consensus for an acceptable response rate is for at least half the intended population to respond.^{19, 20}

We feel the MFJS tests the concept of “awareness” of a prosthetic

Table 2
Ceiling and Floor effects for TKR.

Knee Patient Group (n = 131)	Percentage (%)
Ceiling Effect for OKS	1.54
Ceiling Effect for MFJS Knees	1.54
Scores from 42 to 48 in OKS	30.77
Scores from 87.5 to 100 in MFJS Knees	7.69
Floor Effects for OKS	

joint in a more complete way than the FJS. It is an exciting new PROM, which is exclusively a post-operative assessment tool. With this in mind we feel it should be used in conjunction with the OHS/OKS to further

Table 3
Ceiling and Floor effects for THR.

Hip Patient Group (n = 81)	Percentage (%)
Ceiling Effects for OHS	11.27
Ceiling Effects for MFJS Hips	8.45
Scores from 42 to 48 in OHS	42.25
Scores from 87.5 to 100 in MFJS Hips	12.68
Floor Effects for OHS	1.41

Table 4
Missing Data in the MFJS.

MFJS – are you aware of your artificial joint.....	% of Missing Items (TKR)	% of Missing Items (THR)
Overall Missing Data	0.70	1.27
Q.1.....in bed at night?	0.00	0.00
Q.2.....when sitting on a chair for more than an hour?	0.00	1.41
Q.3.....when you are walking for more than 15 min?	0.78	1.41
Q.4.....when you are taking a bath/shower?	0.78	1.41
Q.5.....when you are travelling in a car?	0.00	1.41
Q.6.....when you are climbing stairs?	1.55	0.00
Q.7.....when you are walking on uneven ground?	0.78	0.00
Q.8.....when you are squatting/crouching?	2.33	7.04
Q.9.....when you are standing for long periods of time?	0.78	0.00
Q.10.....when you are doing housework/gardening or the most strenuous activity you do around the home?	0.00	0.00

discriminate high scoring patients following lower limb joint arthroplasty.

Disclaimer

Views presented in this article are our own and not an official position of our associated institutions.

Conflicts of interest

None.

Financial support

None.

References

- Ostendorf M, Van Stel HF, Buskens E, et al. Patient-reported outcome in total hip replacement. *Bone Joint J.* 2004;86:801–808.
- Takeuchi R, Sawaguchi T, Nakamura N, Ishikawa H, Saito T, Goldhahn S. Cross-cultural adaptation and validation of the Oxford 12-item knee score in Japanese. *Arch Orthop Trauma Surg.* 2011;131:247–254.
- Jenny JY, Diesinger Y. Validation of a French version of the Oxford knee questionnaire. *Orthop Traumatol: Surg Res.* 2011;97:267–271.
- Behrend H, Giesinger K, Giesinger JM, Kuster MS. The “forgotten joint” as the ultimate goal in joint arthroplasty: validation of a new patient-reported outcome measure. *J Arthroplast.* 2012;27:430–436.
- Haverkamp D, Breugem SJ, Siersevelt IN, Blankevoort L, Dijk CN. Translation and validation of the Dutch version of the Oxford 12-item knee questionnaire for knee arthroplasty. *Acta Orthop.* 2005;76:347–352.
- Charoencholvannich K, Pongcharoen B. Oxford knee score and SF-36: translation & reliability for use with total knee arthroscopy patients in Thailand. *J Med Assoc Thai.* 2005;88:1194–1202.
- Gonçalves RS, Tomás AM, Martins DI. Cross-cultural adaptation and validation of the Portuguese version of the Oxford Knee Score (OKS). *Knee.* 2012;19:344–347.
- Naal FD, Sieverding M, Impellizzeri FM, von Knoch F, Mannion AF, Leunig M. Reliability and validity of the cross-culturally adapted German Oxford hip score. *Clin Orthop Relat Res.* 2009;467:952–957.
- Dunbar MJ, Robertsson O, Ryd L, Lidgren L. Translation and validation of the Oxford-12 item knee score for use in Sweden. *Acta Orthop Scand.* 2000;71:268–274.
- Eun IS, Kim OG, Kim CK, Lee HS, Lee JS. Validation of the Korean version of the Oxford Knee Score in patients undergoing total Knee arthroplasty. *Clin Orthop Relat Res.* 2013;471:600–605.
- Karros DJ. Statistical methodology: II. Reliability and validity assessment in study design, Part B. *Acad Emerg Med.* 1997;4:144–147.
- Roos EM, Toksvig-Larsen S. Knee injury and osteoarthritis outcome score (KOOS)—validation and comparison to the WOMAC in total knee replacement. *Health Qual Life Outcomes.* 2003;1:17.
- Dunbar MJ, Robertsson O, Ryd L, Lidgren L. Appropriate questionnaires for knee arthroplasty. *Bone Joint J.* 2001;83:339–344.
- Kalairajah Y, Azurza K, Hulme C, Molloy S, Drabu KJ. Health outcome measures in the evaluation of total hip arthroplasties—a comparison between the Harris hip score and the Oxford hip score. *J Arthroplast.* 2005;20:1037–1041.
- Dunbar MJ. Subjective outcomes after knee arthroplasty. *Acta Orthop Scand.* 2001;72:1–63.
- Jenkinson C, Heffernan C, Doll H, Fitzpatrick R. The Parkinson's disease questionnaire (PDQ-39): evidence for a method of imputing missing data. *Age Ageing.* 2006;35:497–502.
- Dawson J, Fitzpatrick R, Murray D, Carr A. A response to issues raised in a recent paper concerning the Oxford knee score. *Knee.* 2006;13:66–68.
- Whitehouse SL, Blom AW, Taylor AH, Pattison GT, Bannister GC. The Oxford knee score; problems and pitfalls. *Knee.* 2005;12:287–291.
- Draugalis JR, Coons SJ, Plaza CM. Best practices for survey research reports: a synopsis for authors and reviewers. *Am J Pharm Educ.* 2008;72:11.
- Babbie ER. *Survey research methods*. 2nd ed. Belmont, California: Cengage Learning; 1990.