

Journal Club: 18th March 2010
Organiser: Mr Adrian Wilson
Basingstoke & North Hampshire Hospital
Orthopaedic Department

Sam Yasen, MBBS, BSc (hons), MRCS; Joel Melton, MBBS, MSc, FRCS (Tr & Orth)

Introduction

There has been considerable recent debate at the North Hampshire Hospital regarding the choice and utilisation of knee-based scoring systems in assessing outcomes after surgical intervention. This journal club was an opportunity to critically review some of the available scoring systems and make suggestions regarding which are the most appropriate for use in the clinical setting. The meeting was attended by consultants and registrars from the North Hampshire Hospital. A reading list was created and distributed one week prior to the journal club along with a summary document which was prepared by one of the registrars as an overview to guide the meeting. Each paper was presented in brief by one of the registrars, followed by general discussion and critique. Recommendations to practice were subsequently made.

Appropriate questionnaires for knee arthroplasty: results of a survey from 3600 patients from the Swedish knee arthroplasty register.

Dunbar MJ, Robertsson O, Ryd L, Ligren L
J Bone Joint Surg [Br] 2001; 83-B:339-44.

Review

Introduction

The Swedish knee arthroplasty register holds over 70 000 records, but the only outcome measure previously recorded has been revision status, which yields data on operations that fail only. The authors sought to identify which of the available scoring tools would be best in establishing more comprehensive endpoints.

Methods

Four general health questionnaires, the Nottingham Health Profile (NHP), the 12-item and the 36-item short form health surveys (SF-12, SF-36) and the Sickness Impact Profile (SIP); and three disease/site specific questionnaires, the Lequesne, the Oxford 12-item knee score (OKS), and the

Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) were evaluated. 3600 patients with a primary diagnosis of osteoarthritis were randomly selected and split into 12 groups each receiving a different combination of health and disease/site specific questionnaire, along with a third questionnaire on time taken and need for assistance to complete these. Three weeks after the first mailing, 420 were randomly selected to repeat the questionnaires to test reproducibility. Data was collected on content and construct validity, reliability, responsiveness and feasibility.

Results

3052 questionnaires were returned (84.8%). NHP, SF-12 and SF-36 had significantly higher response rates than SIP, whereas no difference between response rates for the disease/site specific questionnaires was identified. The SF-12 had the highest percentage being completed of the general health questionnaires, versus the OKS for the disease/site specific. Mean length of time to complete the questionnaires was shortest for the SF-12 at 7.7 minutes, while the OKS ranked second in class at 9.6 minutes. Patients reported significantly greater frequency (28.7%) requiring assistance with the SF-36 to complete compared with the other generic questionnaires, whereas similar frequencies were observed amongst the disease/site specific tools (around 24%). The SF-12 was the only general questionnaire with no floor or ceiling effect. The WOMAC had the highest floor effect (18.3%), with the OKS and Lequesne being lower (6.8% and 6.4% respectively), and the highest ceiling effect (0.8%) versus the OKS (0.1%) and Lequesne (nil). Reliability as assessed by the intraclass correlation coefficient and Cronbach's alpha were varied amongst the health questionnaires but ranked highest for the OKS in the disease/site specific group.

Discussion

The authors subsequently avoided comparing construct validity, as there is no established standard for knee arthroplasty, and constructs are simply validated against other questionnaires or a surgeon's perceptions. The responsiveness of the scores was also not evaluated as there is no consistently established means for this. However, findings from this study, including floor/ceiling effects and reliability estimates were of a similar magnitude to that from other cited studies.

Conclusion

The questionnaires were consequently ranked according to the described parameters, with the SF-12 rating best of the general health questionnaires, and the OKS best of the disease/site specific scores. The SIP scored worst over multiple parameters and its use was discouraged.

Recommendation

This paper suggests that the Oxford Knee Score and the SF-12 are the most appropriate questionnaires in the evaluation of patients undergoing knee arthroplasty. They are therefore recommended for incorporation into clinical use at the North Hampshire Hospital.

The use of the Oxford hip and knee scores

Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, Dawson J.
J Bone Joint Surg [Br] 2007; 89-B:1010-14.

Review

Background

The Oxford hip and knee scores were published in 1996 and 1998 respectively. They were devised as patient-reported, joint-specific tools designed to minimise the influence of co-morbidity. Their reliability, validity and responsiveness have been evaluated in numerous prospective studies, and they have been widely adopted.

Aims

This article seeks to explain the post-publication modifications that these scores have undergone, clarify areas of confusion and recommend ways in which they are best employed. The discussion that follows focuses on the knee questionnaire where applicable.

Discussion

Wording and response categories

Certain questions cause more difficulty than other for patients to answer. Question 4 on the OKS which asks how long patients are able to walk for before pain becomes severe, has a possible response of 'not at all', which can be wrongly interpreted as the opposite meaning to that intended. Question 7 asks if patients 'could' kneel down and get up again afterwards. As some patients will have been told not to kneel this can cause confusion, and should be interpreted as a hypothetical question in this instance. The word 'could' should be in bold and italicised to identify this.

Scoring

Previously questions were scored from 1-5, with 1 representing the best response, giving total scores of between 12 (best) and 60 (worst). This has been modified to come reflect the scoring pattern of the majority of other questionnaires, such that a higher score is a more favourable one. Questions are now score between 0-4 with 4 being the best response, thus giving a total of between 0 and 48.

Missing data

If only one or two questions are unanswered, it is feasible to use the mean value from the other responses to calculate the total score. If more than two questions are unanswered then the overall score should not be defined.

Use of the scores

The OKS is principally a scoring tool for patients undergoing arthroplasty surgery. It has, however, been applied to other disorders of the joint as well in the evaluation of interventions such as pharmacological treatment, after osteotomy or rehabilitation. The largest determinant of the outcome score is the pre-operative score, and absolute scores also tend to decrease with age. Therefore, change in score should be analysed in addition to the post-operative score. Most improvement in function is noted within the first year, so it is reasonable to score patients at one year post-operatively. As the score is joint and disease specific, it is sensible to use it alongside a general health assessment such as the SF-12.

Recommendation

The OKS is an appropriate score for patients undergoing knee arthroplasty, but may also be useful in evaluating results after osteotomy for osteoarthritis. The score should be used in conjunction with a general health measure such as the SF-12.

Review

Knee injury rating scales

Lysholm J, Tegner Y

Acta Orthopaedica 2007; 78:445-453.

Background

Loss of body function or structure (such as loss of ligament or integrity) leads to impairment (WHO 2001). On an individual basis, activity limitation or restriction in participation can arise as a result

of impairment. This can be expressed via a scoring scale. The combination of impairment and activity limitation is disability, while functioning is the absence of disability. An ideal score should have good validity, reliability and responsiveness with no floor or ceiling effect.

Aims

To evaluate the most commonly used knee injury rating scales according to their validity, reliability, responsiveness and mode of administration.

Methods

The five most common knee injury scales were identified through a PubMed search of available abstracts.

Results

The selected rating scales were as follows: Lysholm-Tegner , 498 abstracts; IKDC (International Knee Documentation Committee), 173; Cincinnati knee rating system, 73; KOOS (Knee osteoarthritis outcome score), 35; and Marshall/HSS (Hospital for Special Surgery), 22 abstracts.

Lysholm - Tegner rating system

First published in 1982 and refined in 1985 as a 100-point scale covering 8 different parameters, used in conjunction with a separate 10-level activity grading system. Widely cited and analysed on documented parameters and now validated as a patient-administered tool.

IKDC rating system

Developed by a panel of orthopaedic surgeons to initially include 7 parameters assessing impairment and disability around the knee. Later modified to include subjective factors such as symptoms, sports activity and ability to function. Validity, reliability and responsiveness reported to be good.

Cincinnati rating system

Initially published in 1983, but undergoing several post-publication modifications. Currently comprising 11 components which include physical examination, instrumented knee stability, testing and radiographic findings. Remains a clinician-administered scale.

KOOS

Developed in part from the WOMAC scale to assess knee injury and post-traumatic arthritis. It contains five separate dimensions (with 42 questions) regarding pain, symptoms, function in daily living, function in sport and recreation, and knee-related quality of life, which are scored

separately. It has not undergone any post-publication modification and has been shown to have good validity, reliability and responsiveness with no floor or ceiling effects.

Marshall / HSS score

First presented in 1972 and modified in a step-wise fashion into the HSS score. Validity and reliability are poorly documented and therefore it is not further discussed.

Discussion

The authors emphasise that injury rating scales should not present subjective and objective variables as a composite score as these should be individually reported. The IKDC and Cincinnati fail on this front. They also cite literature identifying subjective findings as being more important than objective parameters, which favours the Lysholm-Tegner and KOOS scores. There is increasing acceptance that scores should be patient-administered. The Lysholm-Tegner has recently been validated to this effect and the KOOS was designed to be completed by the patient. The KOOS is the only score that reports health-related quality of life, but as this is specifically addressing the knee, it is still necessary to combine it with a generic score to facilitate cross-study comparison. Of the four scores, the Lysholm-Tegner is the easiest to use.

Conclusion

Although the authors do not identify an indisputable 'best' score, they note that the Lysholm-Tegner score is the most widely accepted and has been cited by other authors as the 'gold standard' to be adopted. The use of any knee score, however, is suggested as best in combination with a generic questionnaire.

Recommendation

The Lysholm-Tegner score should be used for all knee injury evaluation and following soft tissue reconstruction surgery at the North Hampshire Hospital.

Review

Knee injury and osteoarthritis outcomes score (KOOS) - validation and comparison to the WOMAC in total knee replacement

Roos EM, Toksvig-Larsen S

Health & Qual of Life Outcomes 2003; 1:17

Aims

The KOOS was developed as an extension of the WOMAC score for younger or more active patients with knee injury and osteoarthritis. The objective of this study was to evaluate the usefulness of the score in elderly patients eligible for total joint replacement.

Methods

Patients with knee osteoarthritis about to undergo primary total joint arthroplasty at Lund University Hospital, Sweden were recruited over a 15-month period for inclusion in the study. To assess reliability, questionnaires were mailed one week apart at two intervals (pre-operatively and at 6-month follow up) with a pre-paid return envelope. Scores were calculated for each of the five subscales separately, with no score being calculated if more than two items were omitted.

Content validity was assessed by asking patients at baseline to rate the importance of improvement in each subscale, and construct validity by comparison of the sports subscale to the SF-36 physical functioning subscale. Responsiveness was assessed by effect size, standardised response mean and relative efficiency.

Results

105 patients (66 women) with a mean age of 71.3 years. One died before follow up and 7 did not return questionnaires leaving 97 (92%) patients for analysis at 6 months (and 90 (86%) patients at 12 months). Subscale scores for 'sports and recreation function' were calculable for only 58/105 patients, but 103-105/105 patients in other subscales.

Reliability

54 patients completed the questionnaire twice within 1-23 days. There was no statistical difference between scores except for the 'symptoms' subscale where patients reported more symptoms on the second test occasion (60/100 vs. 58/100 points, $p=0/04$).

Validity

Content validity was demonstrated by over 90% of patients reporting improvements in four of the KOOS subscales as important to them, and 51% in the 'sports' subscale. Construct validity was shown by high correlation with the SF-36 subscale.

Responsiveness

A significant improvement ($P < 0.001$) was seen post-operatively in all subscales. The most responsive subscale was 'knee-related quality of life', followed by 'pain'. The 'sports' subscale was least responsive. Pre-operatively there was no notable floor or ceiling effect. Post-operatively 15% reported the best possible 'pain' score at 6 months, and 16% the best 'sports' score. At 12 months this was 22% for 'pain' and 17% for 'quality of life'.

Comparison to WOMAC

The above findings are similar to that for the WOMAC, but the WOMAC does not assess the 'sports' function subscale and has poorer floor and ceiling effects.

Discussion

The KOOS has previously been validated for assessing operative treatments following knee injury including arthroscopy and anterior cruciate ligament reconstruction. This study now validates it as a tool for the assessment of older individuals with osteoarthritis with good responsiveness. The authors argue that the score can be deemed reliable as there was no difference in test-retest scores for four of the subscales, and a clinically insignificant difference (of 2 points only) for the 'symptoms' subscale. Construct validity is rather tenuously demonstrated by comparison to the SF-36, which other authors have argued has its own short-comings; whereas content validity is assessed on the basis of patients expectations.

Advantages of the KOOS over the WOMAC are suggested by the inclusion of the 'sports' subscale which was rated as important by the majority of patients, and the 'knee-related quality of life' subscale which proved to be the most responsive measure.

Conclusion

The authors conclude that the KOOS is a useful, reliable, valid and responsive instrument for use in elderly patients with osteoarthritis, and may have advantages over the WOMAC in assessing younger patients with higher expectations.

Recommendation

The KOOS is a detailed score with 5 subscales which is more cumbersome to use than either the OKS for evaluating patients with knee arthropathy or the Lysholm-Tegner score for soft tissue knee injury. Therefore although this paper demonstrates its clinical efficacy, it is not recommended for scoring patients at the North Hampshire Hospital.