The value of short and simple measures to assess outcomes for patients of total hip replacement surgery

Ray Fitzpatrick, Richard Morris, Shakoor Hajat, Barnaby Reeves, David W Murray, David Hannen, Marianne Rigge, Olwen Williams, Paul Gregg

Abstract

Objectives—To evaluate the performance of a patient assessed outcome measure, the Oxford Hip Score, in a national study of primary hip replacement surgery.

Design—A survey of patients’ health status before undergoing primary hip replacement surgery and three months and one year after surgery.

Setting—143 hospitals in three NHS English regions.

Patients—7151 patients admitted for primary total hip replacement surgery over a period of 13 months from September 1996.

Main measures—For patients, Oxford Hip Score and satisfaction with hip replacement and, for surgeons, American Anesthesiologists’ Society (ASA) classification of physical status.

Results—The response rates to the postal questionnaire at three and 12 months follow-up were 85.2% and 80.7%, respectively. Including all three administrations of the questionnaire, all except two items of the Oxford Hip Score were completed by 97% or more respondents and only one item at one administration appeared marginally to reduce the reliability of the score. The effect sizes for changes in the score from baseline to three months was 2.50 and to 12 months was 3.05. Patients rated by surgeons as being healthy preoperatively by the ASA classification were somewhat more likely to return a completed questionnaire at three months (79.4% versus 75.3%) and 12 months (72.4% versus 70.3%) than those rated as having poorer health.

Conclusions—Overall there was little evidence of difficulties for patients in completing the Oxford Hip Score or of unreliable data, except in relation to one questionnaire item. The instrument was very responsive to change over time and score changes for the Oxford Hip Score related well to patients’ satisfaction with their surgery. The instrument is an appropriate measure in terms of validity, responsiveness, and feasibility for evaluating total hip replacement from the perspective of the patient. (Quality in Health Care 2000;9:146–150)

Keywords: total hip replacement surgery; Oxford Hip Score; questionnaire; assessment

Primary total hip replacement surgery is an effective and very commonly performed procedure to reduce pain and improve physical function in patients with arthritis. Nevertheless, it is widely recognised that the quality of evidence to inform surgical choices—for example, with regard to type of prosthesis used—is modest, with few randomised trials or other forms of well designed studies to evaluate variations in practice.1 2 Two recently published NHS R&D systematic reviews of primary total hip replacement concurred in evaluating the quality of evidence in this field as poor, and specifically argued for the need for greater use of validated patient assessed outcome measures to be incorporated into future evaluations of total hip replacement.3 4 To date, orthopaedic evidence on effectiveness has relied excessively on indicators such as whether revision of surgery occurs, a potentially misleading approach to outcomes since patients whose hip replacements are judged to have “failed” do not always receive revision surgery.5

A number of patient assessed outcome measures have recently been developed in the form of questionnaires for use in evaluative health research, assessing issues such as functional status and health related quality of life. Several such measures have been applied to total hip replacement.6 7 Because of the need to measure specific forms of pain and the problems of mobility in older patients who may experience substantial symptoms for other reasons than hip disease, the Oxford Hip Score was developed specifically to assess the patient’s perceptions of pain, mobility, and function in relation to problems of the hip.3

An opportunity to examine the usefulness of the Oxford Hip Score was provided by the National Total Hip Replacement Project (NTHRP). This study, coordinated by the Clinical Effectiveness Unit of the Royal College...
Assessment of outcomes of total hip replacement surgery

of Surgeons and carried out in three English health regions, is the first attempt to describe in detail current practice in primary total hip replacement in England on a large sample of patients combining evidence from both surgeons and patients. The Oxford Hip Score was selected as the primary method to assess patients’ perceptions of their hip problems before surgery and their outcomes at three and 12 months postoperatively.

The emergence of patient based outcome measures has led to concerns that overly simple standardised questionnaires may fail to capture matters of importance to patients and therefore produce misleading evidence of outcomes of health care interventions. The need for more elaborate and detailed methods of eliciting patients’ experiences has also been expressed specifically in the field of orthopaedics. It has also been argued that patients can have difficulties completing questionnaires if their simple and standardised format fails to address nuances and complexities of personal experience. This paper examines the measurement properties of the Oxford Hip Score in the context of a large study of outcomes of primary total hip replacement. Specifically, the paper examines (1) the extent of patients’ difficulties in completing the Oxford Hip Score as evidenced by the response rate for the questionnaire and by the frequency of missing values and unreliable data; (2) the sensitivity to change (responsiveness) of the Oxford Hip Score; and (3) some potential biases from this method of assessing outcome.

Methods
A total of 390 consultant firms from 143 hospitals in three English health regions agreed to participate in a national study of NHS and private primary total hip replacement, recruitment for which began in September 1996 for NHS patients and in October 1996 for private patients, continuing for one year in both groups. The design of the study required the surgeon to complete a questionnaire for each operation performed for primary total hip replacement. The patient was invited to complete a questionnaire at a point during their hospital stay before surgery. This questionnaire also asked for permission to send a postal follow up questionnaire three months and one year after surgery. Regional research coordinators collated patients’ and surgeons’ questionnaires and returned them to the Clinical Effectiveness Unit (CEU), Royal College of Surgeons, London. The group of patients whose preoperative questionnaires were incorporated into the central database of the CEU were subsequently followed up by post from the CEU.

The surgeon’s questionnaire collected data on details of surgical approach, type of prosthesis, cement, anaesthesia, thromboprophylaxis, and a rating of the patient’s overall health status by means of the American Society of Anesthesiologists’ classification of physical status (ASA score). Questionnaires from both surgeon and patient were obtained for 5038 cases.

The patient’s preoperative questionnaire included the 12 item Oxford Hip Score, questions about other major health problems, pain in other joints and duration of outpatient and inpatient wait. The three and 12 month follow up questionnaires included the Oxford Hip Score and other questions about further admission for problems or complications of their hips, satisfaction with hospital care and with results of their hip operation. For non-respondents at both follow up surveys one reminder was sent at two weeks and a further reminder together with another copy of the questionnaire was sent two weeks later.

The Oxford Hip Score is a 12 item self-completed questionnaire addressing the patient’s perceptions of pain and disability arising from their hip. It is intended to be used as a single summed score with the total score reflecting the severity of problems that the respondent has with his or her hip. The range of scores is from 12 to 60 with a high score indicating a poor perceived state of health. From prospective evidence of patients before and six months after hip replacement surgery it has been shown to have very satisfactory reproducibility and has been validated with reference to arthritic-specific and generic health status measures and the surgeon’s assessment in terms of the Charnley Hip Score. Items in a summed score constituting a scale should be internally consistent, as measured by Cronbach’s alpha, with higher values of alpha reflecting higher reliability. The alpha value should normally be above 0.70. In the current study internal reliability was assessed for all three administrations by Cronbach’s alpha for the score as a whole and when individual items were removed to examine their effect on reliability. Effect sizes for the Oxford Hip Score were calculated to assess responsiveness as the difference between the mean preoperative and follow up scores (three and 12 months) divided by the standard deviation (SD) of the preoperative scores.

Results
A total of 7151 completed preoperative questionnaires were returned to the CEU. The mean (SD) age of the patients was 67.8 (10.9) years; 4284 (61.4%) of those who reported their sex were women and 4285 (87.5%) of patients whose diagnosis was available had osteoarthritis.

A total of 6174 patients replied to the follow up questionnaire at three months. In order to estimate the three month mortality rate following primary total hip replacement, the vital status of all patients in the baseline sample was ascertained from the National Office of Statistics. This process identified 80 patients as having died before the three month follow up. When these were removed from the denominator a response rate to the questionnaire of 85.2% at three months was obtained. In response to the 12 month follow up 5854 questionnaires were returned. It was not possible to ascertain further deaths so the response rate (80.7%) was calculated from the
denominator of those returning preoperative questionnaires with those who died in the first three months again removed.

An analysis was performed of the extent of missing values for items of the Oxford Hip Score for the three administrations (preoperative, three months and 12 months). Most of the questionnaires returned had no missing items from the Oxford Hip Score (91.9% at preoperative assessment, 88.3% at three month follow up, and 87.3% at 12 month follow up). At all three assessments, for respondents whose questionnaires were incomplete, most commonly only one item was left unanswered. When analysed by individual items of the Oxford Hip Score, fewer than 3% of respondents omitted any item. The only exceptions were item 6 (distance can walk before pain severe) which was left blank by 3.2% and 4.1% of respondents at three and 12 months, respectively, and item 9 (limping when walking) which was left incomplete by 3.1% at 12 months (table 1).

Table 1: Missing values for the Oxford Hip Score at three administrations

<table>
<thead>
<tr>
<th>Item Description</th>
<th>Pre-surgical N (%)</th>
<th>3 month N (%)</th>
<th>12 month N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. How would you describe the pain you usually had from your hip?</td>
<td>120 (1.7%)</td>
<td>125 (2.0%)</td>
<td>134 (2.3%)</td>
</tr>
<tr>
<td>Q2. Have you had any trouble with washing and drying yourself (all over) because of your hip?</td>
<td>97 (1.4%)</td>
<td>94 (1.5%)</td>
<td>126 (2.2%)</td>
</tr>
<tr>
<td>Q3. Have you had any trouble getting in and out of a car or using public transport because of your hip?</td>
<td>101 (1.4%)</td>
<td>123 (2.0%)</td>
<td>164 (2.3%)</td>
</tr>
<tr>
<td>Q4. Have you been able to put on a pair of socks, stockings or tights?</td>
<td>100 (1.4%)</td>
<td>130 (2.1%)</td>
<td>118 (2.0%)</td>
</tr>
<tr>
<td>Q5. Could you do the household shopping on your own?</td>
<td>143 (2.0%)</td>
<td>133 (2.2%)</td>
<td>174 (3.0%)</td>
</tr>
<tr>
<td>Q6. For how long have you been able to walk before pain from your hip became severe? (with or without a stick)</td>
<td>142 (2.0%)</td>
<td>195 (3.2%)</td>
<td>240 (4.1%)</td>
</tr>
<tr>
<td>Q7. Have you been able to climb a flight of stairs?</td>
<td>141 (2.0%)</td>
<td>183 (3.0%)</td>
<td>162 (2.8%)</td>
</tr>
<tr>
<td>Q8. After a meal (sat at a table) how painful has it been for you to stand up from a chair because of your hip?</td>
<td>98 (1.4%)</td>
<td>113 (1.8%)</td>
<td>129 (2.2%)</td>
</tr>
<tr>
<td>Q9. Have you been limping when walking because of your hip?</td>
<td>111 (1.6%)</td>
<td>143 (2.3%)</td>
<td>182 (3.4%)</td>
</tr>
<tr>
<td>Q10. Have you had any sudden severe pain (“shooting”, “stabbing”, or “spasms”) from the affected hip?</td>
<td>117 (1.9%)</td>
<td>110 (1.8%)</td>
<td>111 (1.9%)</td>
</tr>
<tr>
<td>Q11. How much has pain from your hip interfered with your usual work (including housework)?</td>
<td>145 (2.0%)</td>
<td>173 (2.8%)</td>
<td>171 (2.9%)</td>
</tr>
<tr>
<td>Q12. Have you been troubled by pain from your hip in bed at night?</td>
<td>88 (1.2%)</td>
<td>121 (2.0%)</td>
<td>159 (2.6%)</td>
</tr>
</tbody>
</table>

Potential sources of bias were examined in relation to the Oxford Hip Score arising from incomplete evidence being provided by older patients or those with poorer general health status. From the denominator of patients who returned a questionnaire, those who completed every item were somewhat younger than those who omitted at least one item (mean (SD) age 67.7 (11.0) versus 69.9 (10.3), difference 2.3 (95% CI 3.4 to 1.1), p<0.001). Also, patients whose general health status was more favourable—that is, fit and healthy with no co-morbidity as rated by the surgeon on the ASA score—were slightly more likely to complete the questionnaire fully than those rated as having minor or severe medical problems, whether preoperatively (2205/2362 (93.4%) versus 2113/2313 (91.4%), difference 2% (95% CI 0.5 to 3.5), p<0.01), at three month follow up (91.2% versus 86.1%, difference 5.1% (95% CI 3.2 to 7.1), p<0.0001), or at 12 month follow up (89.5% versus 85.9%, difference 3.6% (95% CI 1.6 to 5.7), p<0.001).

Older patients were somewhat less likely to return a questionnaire at all before surgery; the mean age of those who returned a questionnaire was 67.8 (10.9) years compared with 69.1 (11.5) for those not returning a questionnaire (95% CI 3.3 to 4.9, p=0.001). However, the association between older age and non-return of questionnaires was not observed for the three and 12 month follow up assessments. Similarly, patients rated fit and healthy on the ASA score were more likely to return a questionnaire before surgery than those rated as having minor or severe medical problems (2862/4973 (57.6%) versus 2313/4675 (49.5%), difference 8% (95% CI 6.0 to 10). No significant differences between patients grouped by ASA score were observed for rate of return at three month and 12 month follow up. Age and health status were in turn found to be weakly associated with outcomes

www.qualityhealthcare.com
on the Oxford Hip Score. Older patients were somewhat more likely to report poorer scores at three month and 12 month follow up ($r=0.06$ and 0.11, respectively, both $p<0.01$). Similarly, at three month follow up those who were rated fit and healthy on the ASA score preoperatively reported a mean (SD) Oxford Hip Score of 25.0 (9.3) compared with 26.5 (9.4) reported by those with minor or severe medical problems (difference $-1.5$ (95% CI $-2.1$ to $-0.9$), $p<0.001$). At 12 months follow up those rated fit and healthy on the Oxford Score reported a mean score of 20.5 (8.8) compared with 22.8 (9.3) for those with minor or severe medical problems (difference $-2.3$ (95% CI $-2.9$ to $-1.7$), $p<0.001$).

**Discussion**

The use of outcome measures focused on patients' perceptions to evaluate health care is still relatively novel and therefore requires careful appraisal. This study provides an assessment of the use of such a measure, the Oxford Hip Score, in a pragmatic survey recruiting patients across a diverse range of 143 hospitals in three NHS regions. The study firstly provides several kinds of evidence of the extent of changes that the questionnaire may detect, the reliability of the questionnaire in two waves of follow up postal survey at three and 12 months after discharge (85.2% and 80.7%, respectively) are very satisfactory for this method of administration. It is not possible to disentangle the effects of the Oxford Hip Score on the response rate from other questionnaire items included in the two postal surveys. Other analyses therefore considered the amount of missing data in the returned questionnaires and the contribution to the reliability of each item of the questionnaire. It is extremely encouraging that between 87% and 92% of returned questionnaires in the three waves of the study were returned with no missing data. Questionnaire items were filled out by at least 97% of respondents for most items; only one item (question 6 about distance respondent can walk without severe pain) produced markedly different levels of missing data, but even for this item only 3% and 4% of respondents failed to complete it in the postal surveys. The analysis of reliability showed that, at all three administrations, the reliability of the instrument was affected by only one item. A qualitative study by McMurray and colleagues suggested that difficulties in response to this item may be produced by a lack of clarity in the response categories. There are no universally agreed criteria for assessing responsiveness of instruments, although most approaches in some way assess the degree of intra-individual change over time observed by an instrument in patients expected to experience change. The problem with all such approaches is that it is difficult to be precise about the extent of expected change. The Oxford Hip Score proved highly sensitive to change (responsive) in the study; patients reported major improvements in pain and function between preoperative assessment and three month postoperative follow up, with effect sizes comparable to other studies of outcomes of total hip replacement. The instrument also provides evidence of the further improvements that are believed to occur in the course of the rest of the first year after surgery. Furthermore, these improvements were consistently and significantly associated with patients' more direct global judgements of satisfaction with their hip replacement.

Evidence was obtained of small potential biases from this approach to assessing outcomes. Older and less healthy patients were somewhat less likely to complete the Oxford Hip Score. This is a potential problem found with patient assessed outcome measures more generally. This evidence reinforces the need for short instruments that minimise the burden to patients of assessing outcomes.

The performance of the Oxford Hip Score needs to be compared with available measures such as the widely used SF-36. In a direct comparison of the Oxford Hip Score and SF-36, both completed by the same series of patients, the Oxford Hip Score resulted in a higher completion rate and higher responsiveness. This is consistent with other evidence that older respondents have difficulties with the SF-36. The SF-36 also assesses broad aspects of pain that may be difficult to relate to hip problems.

A central aspect of appraising measures such as the Oxford Hip Score is whether they prove useful in detecting differences between patients that are relevant to evaluating health care. Variations in outcome are not expected to occur between patients who have received different forms of total hip replacement—for example, different kinds of prosthesis—until at least five years after surgery. It is therefore premature to judge the Oxford Hip Score in this respect in the National Hip Replacement Project. However, evidence has already been obtained in other applications of the instrument to indicate that it detects significant differences in the threshold of pain and disability at which private and NHS patients receive total hip replacement surgery and also significant differences in the outcomes of primary compared with revision surgery. The Oxford Hip Score is intended for use in any context based on samples of patients such as a randomised controlled trial or well designed observational surveys or audits of orthopaedic surgery where it is possible to take account of potential confounding factors. It is not intended for use in decision making regarding individual patients.

Critiques of the excessive simplicity of patient assessed outcome measures, especially shorter instruments, are effectively emphasising one aspect of their measurement properties—namely, validity. They argue that, given more time and more in depth questioning, patients are capable of providing more detailed information about their health status and perceptions of the benefits of health care interventions. Whilst this may be true, such critiques do not address the need for outcome measures to be adequate in a number of other respects, in particular with regard to responsiveness, acceptability, and...
feasibility. There is a trade off that has to be made between these properties as evaluative instruments for health care interventions such as total hip replacement. It is widely recognised that large sample sizes, almost inevitably from multicentre studies, are required to detect the modest differences between surgical strategies in hip replacement surgery. It is not feasible to collect detailed in depth information from patients about their experiences. Qualitative evidence has an important and distinctive role in the evaluation of health care. Indeed, in the NTHHRP reported here, investigators have collaborated with the College of Health to analyse answers to open ended questions about their experiences. In the context of patient assessed outcomes, qualitative evidence is essential in initially identifying issues of concern to patients that need to be included in outcome measures. McMurray and colleagues also used qualitative evidence to suggest reasons for difficulties respondents may have with an instrument. However, it is less clear how qualitative evidence can contribute to identifying the modest but important benefits that may be associated with different surgical strategies. Health service researchers need to be able to detect such differences in order to improve the quality of total hip replacement surgery.