Comparison of measures to assess outcomes in total hip replacement surgery

Jill Dawson, Ray Fitzpatrick, David Murray, Andrew Carr

Abstract

***Objectives*—To compare the performance of a disease specific and a general health questionnaire in assessing changes resulting from total hip replacement.**

***Design*—Two stage prospective study of patients undergoing total hip replacement surgery involving an assessment at a clinic before and six months after surgery. 60(32%) patients were followed up by post.**

***Setting*—Outpatient departments at a specialist orthopaedic hospital and peripheral clinics within Oxfordshire.**

***Patients*—188 patients admitted for unilateral total hip replacement between February and mid-August 1994.**

***Main measures*—Patients' self assessed scores with the 12 item Oxford hip score and SF-36 general health questionnaire together with surgeons' assessment with Charnley hip score obtained before and again at six months after surgery.**

***Results*—186 patients were followed up six months after total hip replacement; a subsample (n = 60) by post. Of the 60 postal patients, 59(98·3%) fully completed the Oxford hip score compared with 44(73·3%) who fully completed the SF-36. For the follow up sample as a whole, postoperative changes in scores produced a large effect size of 2·75 on the Oxford hip score, compared with −1·89 physical function (SF-36), −2·13 pain (SF-36). With the exception of physical function and role (physical), postoperative SF-36 scores were shown to be similar to or better than those found by two population surveys on patients of comparable age. The responsiveness of a disease specific questionnaire, the Oxford hip score, and relevant sections of a general questionnaire, SF-36, were found to be similar as assessed by three different criteria.**

***Conclusions*—A disease specific questionnaire, the Oxford hip score, and a general state of health questionnaire, SF-36, performed similarly in assessing outcomes of total hip replacement except that the disease specific questionnaire resulted in a higher completion rate and greater responsiveness in some sections. On the other hand the general health questionnaire drew attention to broader problems of physical function not considered by the Oxford hip score. The health questionnaires examined here offer a valid and practical means of monitoring outcomes of hip replacement surgery.**

Department of Public Health and Primary Care, Gibson Building, Radcliffe Infirmary, Oxford
Jill Dawson, research officer

Nuffield Orthopaedic Centre, Windmill Road, Headington, Oxford
Jill Dawson, clinical audit assistant (outcomes)
Andrew Carr, consultant orthopaedic surgeon

Department of Public Health and Primary Care, University of Oxford
Ray Fitzpatrick, university lecturer

Nuffield Department of Orthopaedic Surgery, University of Oxford
David Murray, honorary consultant and senior research fellow

Correspondence to:
Dr R Fitzpatrick, Department of Public Health and Primary Care, Gibson Building, Radcliffe Infirmary, Woodstock Road, Oxford OX2 6HE

Accepted for publication 28 February 1996

(*Quality in Health Care* 1996;5:81–88)

Keywords: health questionnaires, total hip replacement, patient satisfaction, outcomes

Introduction

Total hip replacement has rapidly expanded to become one of the commonest forms of major surgery performed on elderly people.[1] Successful hip replacement can produce major benefits to the patient by reducing pain and improving mobility and function. Although total hip replacement is regarded as a very effective intervention, it is commonly remarked how little evidence has been produced in the form of trials or other systematic observational data to examine the extent and sources of variations in outcomes.[1–3] As a result, little is known of such issues as the health consequences of a proliferation by manufacturers of alternative prostheses for use in total hip replacement. One main concern that has prompted more attention to be focused on quality assurance in hip surgery is the growing rate of revisions that have to be performed, with one study estimating that 13% of hip replacements performed are revisions of previous surgery.[4] It has also been suggested that because of the additional surgical complexities of revision surgery, outcomes are less favourable than for primary surgery.[5]

One of the problems that has inhibited more systematic and accurate monitoring of total hip replacement has been the lack of appropriate outcome measures. Most studies of outcome of hip replacement rely solely on surgical failure defined as the need to perform revision surgery.[6] The need to revise surgery is a rather crude measure of outcome and the decision by the patient and the surgeon to revise may be determined by various factors including health service resources, the patient's illness, behaviour, and general fitness for further surgery. The main alternative measure in use is the surgeon's assessment of outcome. This approach involves several problems. Either such assessments are not standardised or use is made of one of a proliferating range of systems that are of questionable reliability and validity.[6 7] All such scoring systems require surgeons' judgements, which may relate only modestly with patients' perceptions.[7 8]

Another important consideration is that results of total hip replacement need to be monitored over very long time periods as important variations in outcomes may not emerge until a decade after surgery.[9] Surgeons' assessment systems are quite time consuming and are not normally performed in the course of a clinic. Methods of monitoring outcomes

therefore need to be feasible as well as reliable, valid, and responsive.

Patient based outcomes such as health status or quality of life measures are obvious candidates to examine for such requirements and a start has been made in the United States to examine the role of such questionnaires for total hip replacement.[10 11] There are various methods to assess state of health from short and simple self completion forms of questionnaire to longer and more complex interview methods.[12]

Options exist among health status measures. Measures specific to a disease or condition are intended to be maximally sensitive to particular areas of the specific health problem and intervention being monitored whereas general health questionnaires are designed to detect a wider range of consequences of illness and make comparisons between different disease and treatment groups.[13 15]

In view of the many difficulties surrounding the monitoring of outcomes by means of clinical assessment by a surgeon, a study was designed to compare the performance of a disease specific and a general health measure of outcome for total hip replacement in a routine surgical setting. To examine performance comparatively, the study considered two questions:

(1) Are there differences in the completion rates of a disease specific questionnaire compared with a general health questionnaire for patients receiving total hip replacement?

(2) How responsive is a disease specific questionnaire compared with a general health questionnaire in assessing changes resulting from total hip replacement?

The responsiveness of a questionnaire is its sensitivity to important changes over time. Questionnaires may be valid in the sense of making accurate distinctions between people or groups at a point in time but fail to detect important changes within a person.[16] There is less consensus about methods for assessing the responsiveness of a questionnaire than is the case with reliability and validity and it is necessary to compare changes in scores from the same person with a range of other relevant evidence of temporal trends.[17 18] The responsiveness of the two questionnaires were examined in the current study in four ways.

Firstly the overall effect sizes (defined below) of the two questionnaires were examined. There is considerable evidence that patients receiving total hip replacement experience substantial improvements in areas of health such as pain and mobility even within a six month period.[19 20] It was therefore predicted that the effect sizes of scores compared before and six months after total hip replacement should be large.

Secondly, changes in scores for questionnaires were examined against the evidence of patients' own retrospective judgements. Retrospective judgements in the form of transition questions (comparing current health with a specified past state) have been shown to be accurate and valid methods of assessing outcomes in several areas of medicine[21 18] as well as specifically in assessments of patients with arthritis.[22 23] It was predicted that changes in scores for the questionnaires in the current study should be significantly greater for patients who retrospectively report the most improvement.

Thirdly, changes in scores for the questionnaires under study were compared with changes assessed independently. The level of agreement can be examined between changes over time in the study questionnaires and other relevant longitudinal data available such as clinical or laboratory variables.[24] In the current study orthopaedic surgeons made independent judgements in comparable areas of state of health to those considered by the two questionnaires. It was predicted that directions of change indicated by patients' self reported state of health should agree with directions of change over time indicated by clinicians' judgements.

Fourthly, the two questionnaires were examined in relation to two groups of patients expected to experience different levels of improvement. There is clinical, radiological, and patient reported evidence that patients receiving revision hip surgery receive less substantial improvements than patients receiving total hip replacement for the first time.[25 27] It was therefore predicted that, in the questionnaires under study, changes in scores for patients receiving revision surgery should be less favourable than for patients receiving primary surgery.

It is important to recognise that there can be no single gold standard against which to judge questionnaires' responsiveness and as many different pieces of relevant evidence as possible should be examined.

Methods

Between the end of February and mid-August 1994, a prospective study of patients about to undergo hip replacement surgery (on one hip) was carried out at the Nuffield Orthopaedic Centre, Oxford. Consecutive patients were recruited, by a researcher, in a preadmission assessment clinic, and 188 patients who proceeded to surgery completed various assessments before surgery and once again at routine appointments in outpatient clinics six months after surgery. These assessments included the Charnley hip score,[28] carried out by an orthopaedic surgeon, the SF-36 general health survey questionnaire,[29] and the Oxford hip score (figure).[30] The second and third of these questionnaires are designed to be completed by the patient. We selected the SF-36 as a measure of general state of health for two reasons; firstly, because it is becoming the most widely used measure of general health and secondly, because a wide range of comparative data are available for several studies of the general population. The Oxford hip score was chosen because it was developed specifically to assess outcomes of hip replacement surgery.

Patients were often accompanied by a partner, relative, or friend and it was not unusual for them to receive some assistance

PATIENT No.....

Problems with your hip

During the past 4 weeks.. ✓tick one box for every question.

1 *During the past 4 weeks.....*
How would you describe the pain you usually had from your hip?
None ☐ | Very mild ☐ | Mild ☐ | Moderate ☐ | Severe ☐

2 *During the past 4 weeks.....*
Have you had any trouble with washing and drying yourself (all over) because of your hip?
No trouble at all ☐ | Very little trouble ☐ | Moderate trouble ☐ | Extreme difficulty ☐ | Impossible to do ☐

3 *During the past 4 weeks.....*
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)*
No trouble at all ☐ | Very little trouble ☐ | Moderate trouble ☐ | Extreme difficulty ☐ | Impossible to do ☐

4 *During the past 4 weeks.....*
Have you been able to put on a pair of socks, stockings or tights?
Yes, Easily ☐ | With little difficulty ☐ | With moderate difficulty ☐ | With extreme difficulty ☐ | No, Impossible ☐

5 *During the past 4 weeks.....*
Could you do the household shopping on your own?
Yes, Easily ☐ | With little difficulty ☐ | With moderate difficulty ☐ | With extreme difficulty ☐ | No, Impossible ☐

6 *During the past 4 weeks.....*
For how long have you been able to walk before pain from your hip became severe? *(with or without a stick)*
No pain/ More than 30 minutes ☐ | 16 to 30 minutes ☐ | 5 to 15 minutes ☐ | Around the house only ☐ | Not at all ☐

7 *During the past 4 weeks.....*
Have you been able to climb a flight of stairs?
Yes, Easily ☐ | With little difficulty ☐ | With moderate difficulty ☐ | With extreme difficulty ☐ | No, Impossible ☐

8 *During the past 4 weeks.....*
After a meal (sat at a table), how painful has it been for you to stand up from a chair because of your hip?
Not at all painful ☐ | Slightly painful ☐ | Moderately painful ☐ | Very painful ☐ | Unbearable ☐

9 *During the past 4 weeks.....*
Have you been limping when walking, because of your hip?
Rarely/ never ☐ | Sometimes, or just at first ☐ | Often, not just at first ☐ | Most of the time ☐ | All of the time ☐

10 *During the past 4 weeks.....*
Have you had any sudden, severe pain - 'shooting', 'stabbing' or 'spasms' - from the affected hip?
No days ☐ | Only 1 or 2 days ☐ | Some days ☐ | Most days ☐ | Every day ☐

11 *During the past 4 weeks.....*
How much has pain from your hip interfered with your usual work *(including housework)*?
Not at all ☐ | A little bit ☐ | Moderately ☐ | Greatly ☐ | Totally ☐

12 *During the past 4 weeks.....*
Have you been troubled by pain from your hip in bed at night?
No nights ☐ | Only 1 or 2 nights ☐ | Some nights ☐ | Most nights ☐ | Every night ☐

Oxford hip score questionnaire

from this person in reading through their questionnaires.

A proportion of consultants at the Nuffield Orthopaedic Centre hold outpatient clinics on other hospital sites throughout the Oxford Region. Patients treated at the Nuffield Orthopaedic Centre are seen after surgery at these clinics – nearer to their own homes. For this reason, some patients (60(32%)) in the current study received all follow up assessment forms by post. This included the postoperative clinical assessment form, contained in an envelope, which patients were asked to take with them to the next appointment with their surgeon. Their clinical assessment was then completed at a peripheral clinic. An accompanying letter made the suggestion that if a relative or friend had helped the patient with reading their questionnaire on the first occasion, that the same person might be asked to provide help again if at all possible. Once completed, all questionnaires were returned in prepaid envelopes. For the few patients (11(5·9%)) who did not receive a postoperative clinical assessment, information was gathered from patients' hospital notes by the researcher.

The Charnley hip score requires a surgeon to grade patients' pain, mobility, and walking (one question for each) on six point scales in relation to one hip, with lower scores indicating greater disability. Walking may only be assessed on patients who have no other condition – than the one hip – which might undermine walking ability. Although some effort has been made to analyse the construct validity of the Charnley hip score less attention has been given to how reliably surgeons assess the score.[31 32] Surgeons completed the Charnley score without knowledge of patients' state of health scores.

The SF-36 contains 36 items and measures health on eight dimensions covering functional status, wellbeing, and overall evaluation of health. Its reliability and validity have been widely studied and reported.[33 34] Scale scores have a range of 0–100. A low score indicates poor perceived state of health. The Oxford hip score (figure) is a 12 item questionnaire which assesses pain and function of the hip. The range of scores is from 12 to 60 with higher scores denoting worse pain and function. It has been shown to be highly reliable (internal consistency: Cronach's α 0·84; test-retest reliability over 24 hours: $r = 0{\cdot}89$, $P < 0{\cdot}001$; paired t tests $P > 0{\cdot}05$), and construct validity has been established by levels of agreement with the arthritis impact measurement scales[35] – a validated measure widely used to assess arthritis – of pain ($r = 0{\cdot}56$ $P < 0{\cdot}01$), mobility ($r = 0{\cdot}48$ $P < 0{\cdot}01$), and physical activity ($r = 0{\cdot}55$ $P < 0{\cdot}01$).

As well as completing the health status measures, patients were asked a few demographic questions on a front sheet at the preoperative stage. Three retrospective transition questions were asked on the questionnaire six months after surgery. The questions asked patients to estimate changes (better, same, or worse) which had occurred in

general problems arising from their hip, pain, and daily function compared with their condition before surgery. Additional clinical data on diagnosis, perioperative details, and postoperative complications were gathered on a structured form as part of the six month postoperative clinical assessment.

Effect sizes were calculated for the Charnley, SF-36, and Oxford hip score to assess the extent of change in patients' clinical condition between the preoperative and six month postoperative assessments. Effect size is a method of calculating the extent of change measured by a questionnaire in a standardised way that allows comparison between questionnaires.[36] This is calculated as the difference between the sample's mean preoperative and postoperative scores, divided by the SD of the preoperative scores.

t Tests have been used to compare changes in scores for subgroups on different state of health questionnaires. The distribution of changes in scores generally approximated to the normal and under such conditions the *t* test is robust – particularly given the size of the sample.

Results

STUDY POPULATION

In the study period 188 patients proceeded to surgery. Twenty nine patients (15·6%) received revision of a previous total hip replacement, but most (157(84·4%)) received primary hip surgery. Two patients (1%) died after surgery before their follow up assessment.

The median age of patients was 70·5 (range 37·5 to 89·6, mean(SD) 69·4(3·5)), the sample comprised 117(62·9%) women and 69(37·1%) men. Most of the patients (161(86·6%)) were diagnosed as having advanced, primary osteoarthritis as the main cause of their hip disease, but 10 patients (5·4%) had osteoarthritis secondary to congenital dysplasia or dislocation. Nine patients (4·8%) had presented with a failed previous repair of a fractured head or neck of a femur and the remaining six patients (3·2%) had either a history of inflammatory arthritis or osteonecrosis. One hundred and fifty of the total hip replacements performed (81·1%) involved a fully cemented prosthesis, 21(11·4%) were uncemented, and 14(7·6) were hybrid (one of the two components cemented).

POSTOPERATIVE PERIOD TO SIX MONTHS

Forty patients (21·7% of 184 patients for whom information was available) had a major postoperative complication. This included 11(6·0%) who either required additional unexpected transfusion for heamorrhage or treatment with antibiotics for a wound infection, nine (4·9%) who were treated for deep vein thrombosis or pulmonary embolism – despite standard prophylactic treatment used in this hospital – three(1·6%) required treatment for a severe cardiorespiratory event, and 11(6·0%) dislocated the new hip joint. The remaining six (3·3%) were treated for various conditions – for example, urinary retention, and intestinal obstruction. Also, a further 28 patients were recorded as having some other major related or unrelated medical event during the six month postoperative period. Not all postoperative complications occurred during the original period in hospital, and overall, 22 patients (11·8%) required readmission to hospital due to a postoperative complication.

HEALTH STATUS SCORES BEFORE AND AFTER SURGERY

All 186 patients eligible for six month assessment attempted to complete and return the Oxford hip score and SF-36. All but one (postal) patient completed the Oxford hip score fully, and 167 patients (89·7%) completed all questions on the SF-36; eight patients (4·3%) missed out more than three questions. Patients who completed the SF-36 by post were significantly less likely than other patients to return fully completed questionnaires (44(73%) *v* 123(98%), 1 df, $P < 0·0001$).

Table 1 shows the scores for the relevant sections of the SF-36 before and after surgery. The scores of patients older than 64 are also presented and compared with similar age groups from two population surveys.[34 37] At six months after surgery, patients receiving total hip replacement produced scores that were similar to or better than those obtained from patients of similar age derived from the general population on most of the SF-36 sections: bodily pain, mental health, social functioning, role limitations due to emotional problems, vitality, general health perceptions and change in health. However, scores for physical functioning were somewhat poorer and role limitation due to physical problems were

Table 1 Comparison between scores obtained on the SF-36 dimensions before and at six months after total hip replacement and scores obtained by other studies based on patients of comparable age

	Preoperative scores		*Six months after surgery*			
SF-36 sections	*All Patients (n = 186) mean score*	*Patients > 64 y (n = 147) mean score*	*All patients (n = 186) mean score*	*Patients > 64 y (n = 126) mean score*	*General population survey age 65–74[33] (n = 103) mean score*	*General practice population > 65 y[34] (n = 42) mean score*
Physical functioning	17·8	17·3	49·5	46·2	60	66·1
Bodily pain	30·1	30·7	65·9	66·8	67	71·1
Mental health	70·8	71·5	78·5	77·9	73	77·3
Social functioning	51·8	51·1	79·1	78·0	80	79·3
Role limitation: physical problems	13·7	14·8	44·3	43·5	59	63·1
Role limitation: emotional problems	55·7	56·2	73·8	70·8	73	54·1
Vitality	46·7	47·5	61·7	61·1	57	55·8
Perception of general health	74·0	72·4	75·3	76·5	58	60·5
Change in health	42·9	42·2	75·9	76·6	not recorded	47·6

considerably poorer than those obtained in the population surveys of older people.

Preoperative Oxford hip scores produced a median value of 44 (range 16–59) mean(SD) 43·6(7·0). These scores compared with a median value of 22 (range 12–51) mean(SD) 24·3(9·4) at six months after surgery. The mean(SD) change in score was 19·2(9·6) producing an effect size of 2·8 (table 2). This may be compared with effect sizes for the physical (–1·89) and pain (–2·13) sections of the SF-36. The distribution of changes in scores approximated to normal. The surgeons' Charnley assessment of pain produced a slightly larger effect size than that produced by either of the patient based measures, although smaller effect sizes resulted from surgeons' assessments of walking ability and, in particular, range of movement of the replaced hip.

RESPONSIVENESS AND RETROSPECTIVE JUDGEMENTS

Three questions considered retrospective judgements of change in state of health. In response to these, 160 patients (86%) said that their hip pain was "much better", and 20 patients (10·8%) said that their pain was "slightly better" after surgery; but two patients (1·1%) reported "no change", leaving four patients (2·2%) reporting pain that was "slightly worse" or "much worse" than before the operation. One hundred and fifty patients (80·6%) recorded that overall, problems related to their hip were much better, 25(13·4) as slightly better since surgery. Six patients (3·2%) reported no change and five patients' (2·7%) responsed that they were slightly worse or much worse than before their operation. Finally, 116 patients (62·4%) said that their day to day life was much better since their operation, 48(25·8%) conceded that it was slightly better, with 12(6·5%) reporting no change; this left 10 patients (5·2%) who said that their day to day life was now slightly worse or much worse than before the hip operation.

Table 2 Changes in scores and effect sizes, as measured by three measures of state of health before and six months after total hip replacement

Questionnaire: Section	*Mean change (SD)*	*Effect size*
Charnley hip score:		
Pain	–2·65(1·3)	–3·27
Walking*	–2·07(1·5)	–1·67
Movement	–1·21(1·4)	–1·03
SF-36:		
Physical	–31·57(27·4)	–1·89
Pain	–35·84(27·9)	–2·13
Mental health	–7·82(17·8)	–0·42
Social function	–27·05(34·4)	–0·88
Role: physical	–30·40(42·8)	–1·18
Role: mental	–17·23(50·6)	–0·40
Vitality	–14·86(22·6)	–0·66
Perceptions of health	–1·37(16·0)	–0·07
Change in health	–33·06(30·1)	–1·80
Oxford hip score	19·24(9·6)	2·75

*Excludes patients with a second symptomatic hip or impaired ability to walk due to another condition (n = 104).

Patients' retrospective judgements were compared with independent assessments provided by changes in scores (preoperative minus postoperative score) resulting from surgeons' Charnley ratings, the nine sections of the SF-36, and the Oxford hip score (table 3). In each case the changes in scores for the group of patients presenting the most favourable retrospective judgements were compared with changes in scores for all other patients. Patients who made the most favourable retrospective judgements reported significantly larger changes in scores for the pain component of the Charnley hip score (but not the walking and range of movement components), the Oxford hip score, and the sections of SF-36 more relevant to total hip replacement – such as pain and physical function – than did those with less favourable retrospective judgement. The changes in scores of the Charnley pain component were significant at the $P < 0·05$ level. By comparison, the Oxford hip score and pain and physical function sections of SF-36 showed high levels of significance ($P < 0·001$). However, for less relevant SF-36 sections there were small or no differences found in changes in scores of patients with more favourable compared with less favourable retrospective judgements.

The level of agreement between surgeon's independent assessment and patient based questionnaires was found to be highest between the Charnley pain component and the Oxford hip score (table 4). Changes in scores relating to the pain section of the SF-36 agreed less well with those of the Charnley pain score. Changes in scores for other relevant sections of the SF-36 correlated significantly with the

Table 3 Mean changes in scores for Charnley hip score, Oxford hip score, and SF-36 related to patients' retrospective judgements of change in health

Retrospective questions	*Charnley*		*SF-36*										*Oxford hip score*
	Pain	*Walking*[1]	*Range of movement*	*Physical*	*Pain*	*Mental health*	*Social function*	*Role physical*	*Role mental*	*Vitality*	*Perceptions of health*	*Change in health*	
How have the problems related to your hip changed since you had your operation?:													
Much better	–2·82*	–2·17NS	–1·24NS	–36·2†	–39·0***	–8·4NS	–30·8***	–34·0***	–18·5NS	–17·7***	–2·9***	–37·0†	21·2†
Other responses	–1·91	–1·50	–1·06	–12·8	–22·8	–5·6	–11·1	–15·4	–11·8	–3·3	4·9	–16·7	11·0
Have you experienced a change in the level of pain from your hip since having the operation?:													
Much better	–2·80†	–2·20*	–1·20NS	–35·0†	–38·6***	–8·2NS	–29·8***	–34·2†	–20·0NS	–15·7NS	–2·3NS	–35·6***	21·2
Other responses	–1·71	–1·21	–1·26	–9·8	–18·8	–5·4	–9·8	–5·4	–1·4	–9·8	4·2	–17·3	7·6
How has your hip operation changed your day to day life so far?:													
Much better	–2·85*	–2·25NS	–1·23NS	–41·1†	–42·8†	–9·2NS	–30·5NS	–35·8*	–23·9*	–18·2*	–39·0†	28·4†	22·6†
Other responses	–2·32	–1·72	–1·18	–15·8	–24·3	–5·5	–21·3	–21·2	–6·1	–9·4	–23·2	30·5	13·7

*P < 0·05; ***P < 0·001; †P < 0·0001.

Table 4 Correlation coefficient between change scores obtained on the Charnley hip score and those measured by the Oxford hip score and the SF-36 after total hip replacement

	Charnley hip score		
	Pain	*Walking†*	*Movement*
SF-36:			
Physical	0·24**	0·27**	0·05
Pain	0·22**	0·21*	0·13
Mental health	0·14	0·03	0·06
Social function	0·26**	0·40**	0·07
Role: physical	0·24**	0·22*	0·08
Role: mental	0·03	0·26**	0·12
Vitality	0·20**	0·36**	< 0·01
Health perceptions	0·01	0·09	0·01
Health change	0·30**	0·15	0·10
Oxford hip score	0·51**	0·30**	0·06

*P < 0·05; **P < 0·01.
†Excludes patients with a second symptomatic hip or impaired ability to walk due to another condition (n = 104).

Charnley pain assessment. Significant correlation was found between changes in scores for the Charnley walking assessment and both the Oxford hip score and relevant sections of the SF-36. By contrast, the changes in scores relating to the Charnley assessment of range of movement of the hip was not significantly correlated with any of the patient based measures.

COMPARISON BETWEEN PATIENTS WITH PRIMARY AND REVISION HIP REPLACEMENT

No significant differences were found between the 157 patients with primary hip replacement and the 29 who underwent revision surgery for any of the following variables: patients' age or sex, presence of symptomatic disease affecting other joints, surgeons' assessment of post-operative hip *x* ray film, occurrence of post-operative complications or duration of stay in hospital. Also, no significant differences were found between the preoperative scores obtained for the two groups with the Oxford hip score or any of the SF-36 sections.

In retrospective assessments of changes in their condition 140(89%) of the patients with primary hip replacement said that the pain from their hip was much better after surgery compared with 20(69%) of those who had had a previous hip replacement revised. This difference was significant ($P < 0·01$). Also, 131(83%) of the patients with primary hip replacement said that the problems related to their hip were much better after surgery compared with 19(66%) of those having revision surgery. This too was significant ($P < 0·05$). No difference was found between the two groups in perceived change in day to day life so far resulting from the hip surgery.

Differences were examined in changes in scores for the surgeon's Charnley scores and two health status questionnaires obtained from patients receiving primary compared with revision hip replacement (table 5). Although differences were not significant for the surgeon's Charnley scores, patients receiving revision surgery experienced less improvement than other patients as measured by the Oxford hip score and the two physical function scales of the SF-36. Comparatively, the SF-36 and Oxford hip score seemed equally sensitive to the different experiences of patients undergoing primary and revision total hip replacement. However, the pain scale of the SF-36 did not register a significant difference between groups.

Table 5 Change scores and effect sizes, as measured by three health status measures before surgery and 6 months after surgery, comparing primary hip replacement patients with those who underwent revision surgery

Questionnaire: section	*Primary hip surgery*		*Revision surgery*		*Primary v revision surgery (t tests comparing change in scores)*
	Mean change (SD)	*Effect size*	*Mean change (SD)*	*Effect size*	
Charnley hip score:					
Pain	−2·72(1·2)	−3·53	−2·22(1·6)	2·28	−1·57NS
Walking†	−2·04(1·5)	−1·57	−2·21(1·6)	−1·35	0·38NS
Movement	−1·19(1·3)	−1·06	−1·30(1·7)	−0·91	0·35NS
SF-36:					
Physical	−34·00(26·2)	−2·18	−18·52(30·6)	−0·85	−2·74**
Pain	−36·73(28·5)	−2·20	−31·03(24·2)	−1·78	1·01NS
Mental health	−7·53(18·2)	−0·41	−9·48(15·7)	−0·43	0·52NS
Social function	−28·03(34·1)	−0·94	−21·84(35·9)	−0·62	−0·89NS
Role: physical	−33·83(42·6)	−1·37	−10·58(38·8)	−0·48	−2·60*
Role: mental	−16·89(51·2)	−0·39	−19·05(48·4)	−0·44	0·21NS
Vitality	−15·10(23·2)	−0·69	−13·62(18·9)	−0·53	−0·32NS
Health perceptions	−2·10(16·0)	−0·10	−2·48(15·6)	−0·13	−1·42NS
Health change	−33·44(30·7)	−1·86	−31·03(27·3)	−1·49	−0·39NS
Oxford hip score	20·08(9·6)	2·95	14·72(8·5)	1·84	2·80**

*P < 0·05; **P < 0·01.

Discussion

Orthopaedic surgeons are not normally able to include in routine clinics the performance of standard clinical assessments such as the Charnley hip score, which involve time consuming measures of – for example, range of movement.[31] As a result, regular assessment of outcome is problematic. It is therefore vital to examine alternative methods of measuring outcomes of orthopaedic surgery, not only for measurement properties such as validity and responsiveness, but also for practical feasibility. Patient based outcomes are now being widely considered as such an option.[38]

Although surgeons were not invited to assess the value of information on health status in this study, a recent consensus conference of orthopaedic surgeons made it clear that methods of assessing patients' views of orthopaedic surgery are essential.[39] Studies that have been carried out to examine the value of such data on outcomes to doctors suggest that they generally find such measures helpful, but it is not clear that their use leads to changes in clinical management.[40] Rather than being used in decisions about care of individual patients, it is more likely that patient based outcome measures will find a role in clinical trials and the continual process of quality assurance.[41]

Overall, the results of the current study are consistent with other evidence in suggesting that patient based data on outcomes can be collected with very good response rates and minimal disruption to routine care.[42] Patients were prepared to complete and return questionnaires in assessment clinics before surgery and, excluding two patients who died, 100% of patients returned some information about health status at the six month follow up. However, it required most of one researcher's time over an 18 month period to recruit patients into the study, ensure the completion of assessments before surgery and at follow up, and process the data. As others have found,[43 44]

the use of patient based outcome measures requires a considerable level of staff commitment and resources without which poor response rates and uninformative results are likely. Questionnaire based monitoring of outcomes should not be viewed as a costless option.

The study firstly compared the performance of the two types of questionnaire, disease specific and general, for rates of response and completion. Although response rates were satisfactory for the shorter (12 item) Oxford hip score, a few patients, particularly in the subgroup surveyed by post, did not fully complete the SF-36. Another study has found that older respondents may have difficulties with the SF-36 and it is suggested that this is in part due to inappropriateness of particular items such as questions which relate to work or vigorous activities.[37] In view of the mean age of patients receiving hip replacements, the Oxford hip score avoids questions in these areas and is deliberately brief to achieve high reponse rates among elderly people.

The study also examined the capacity of the questionnaires to measure benefits of total hip replacement. As indicated by the very high effect sizes, both the Oxford hip score and SF-36 show substantial improvements for pain, function, and mobility. However, a comparison of SF-36 scores with population data suggests, as have other studies,[20] that state of health may not be restored to the same level as is found among elderly people in general. There still may be substantial scope for further improvement in hip replacement. Reasons for the variability found in the results are not clear but include the possibility that patients are advised not to place undue strain on the hip by physical exertion. Patients may also have lost confidence in their ability to be active or become adjusted to their low levels of function before surgery.

Four different analyses were selected to compare the responsiveness of the two questionnaires. Firstly, although both questionnaires on state of health showed substantial improvements after total hip replacement, the effect size of the Oxford hip score was larger and more consistent with the extent of changes suggested by surgical judgement. These results suggest that the disease specific questionnaire was somewhat more responsive. The Oxford hip score was designed specifically for this application, has fewer redundant questions, and therefore produced larger effect sizes to assess the extent of improvement associated with surgery. This is consistent with other evidence that shorter questionnaires may perform just as effectively as more detailed ones if the content is appropriate and relevant, and that they have additional potential advantages of good response rates and efficiency of processing.[27 45 46]

Patients' retrospective judgements of outcome were used as a second bench mark against which to compare changes in scores of the two kinds of questionnaire on state of health. This has been shown to be a useful method of evaluating the reponsiveness of these questionnaires.[23] On relevant sections of SF-36 (physical function and pain) the general questionnaire was as responsive as the disease specific questionnaire.

The third method of comparing questionnaires was to compare their changes in scores with those obtained from independent clinical evidence. For some dimensions of the Charnley hip score, particularly pain, the Oxford hip score correlated more strongly than the SF-36. Patterns were less clear for other dimensions.

The fourth method of examining responsiveness of questionnaires on state of health was to compare their sensitivity to differences in outcome that were expected to occur for primary surgery compared with revision surgery. Clinical and radiological evidence indicates that revision surgery is associated with more problems of loosening and less favourable outcomes than occur for patients receiving primary surgery.[27 26] Moreover, levels of patient satisfaction with total hip replacement are generally lower in patients undergoing revision surgery.[25] In the current study patients' retrospective judgements of change in their physical condition differed according to whether they had received primary hip replacement or revision surgery. When relevant sections of the SF-36 were compared with the Oxford hip score, they seemed equally sensitive to these differences of outcome.

Thus for overall comparisons of a general and a disease specific measure, the Oxford hip score provided somewhat higher response rates in the postal format and fewer missing items. In terms of responsiveness, by two criteria, overall effect size and agreement with independent evidence of change, the Oxford hip score seemed more sensitive to change. By the other two criteria, patients' retrospective judgements and detection of differences between primary and revision surgery, the two questionnaires showed similar levels of sensitivity to the changes provided by total hip replacement.

The Oxford hip score is therefore more appropriate for examining the main outcomes of pain and functioning after total hip replacement than is a general questionnaire. This result is consistent with a similar study of orthopaedic knee surgery in which a disease specific questionnaire proved more sensitive to outcomes of knee surgery.[47] On the other hand the general questionnaire may be useful in drawing attention to other problems experienced by patients. In the knee surgery study, the SF-36 identified significant levels of pain not directly related to the knee. In this study limitations in physical function were found in patients at follow up that were not identified by the disease specific measure. The general health measure is also useful in making comparisons across treatments and conditions for resource allocation and other comparative judgements.

As it becomes increasingly necessary to conduct multicentred studies recruiting many patients to detect the various sources of variations in outcomes of total hip replacement, so standardised patient based measures will offer relative advantages compared with

surgical judgement.[6 38] It needs to be examined whether patient based measures are a meaningful and practical method of monitoring outcomes over the much longer time periods in which hip replacements are expected to function. Current results suggest that simple measures may provide appropriate means to supplement the evidence of outcomes provided by conventional measures.

Financial support was provided by a grant from Oxford Regional Health Authority (audit). We thank all the orthopaedic consultants and surgical staff at the Nuffield Orthopaedic Centre for allowing their patients to be included in this study, Kim Clipsham for her invaluable assistance in outpatient clinics, and the nursing staff in the clinics, as well as individual patients for giving their valuable time.

1 Ivory J, Summerfield J, Thorne S, Lowdon I, Williamson D. Total hip replacement. *Quality in Health Care* 1994; **3**:114–9.
2 Laupacis A, Rorabeck C, Bourne R, Feeny D, Tugwell P, Sim D. Randomized trials in orthopaedics: why, how and when? *J Bone Joint Surg* 1989;**71**:535–43.
3 Gross M. Innovations in surgery: a proposal for phased clinical trials. *J Bone Joint Surg* 1993;**75**:351–4.
4 Williams M, Newton J, Frankel S, Braddon F, Barclay E, Gray J. Prevalence of total hip replacement: how much demand has been met? *J Epidemiol Community Health* 1994;**48**:188–91.
5 Christel P, Djian P. Recent advances in adult hip joint surgery. *Curr Opin Rheumatol* 1994;**6**:161–71.
6 Murray D, Carr A, Bulstrode C. Survival analysis of joint replacements. *J Bone Joint Surg* 1993;**75**:697–704.
7 Liang M, Katz J, Phillips C, *et al.* The total hip arthroplast evaluation form of the American Academy of Orthopaedic Surgeons. *J Bone Joint Surg* 1991;**73**:639–46.
8 Wright J, Rudicel S, Feinstein A. Ask patients what they want: evaluation of individual complaints before total hip replacement. *J Bone Joint Surg* 1994;**76**:229–34.
9 Bulstrode C, Murray D, Carr A, Pynsent P, Carter S. Designer hips. *BMJ* 1993;**306**:732–3.
10 Liang M, Fossel A, Larson M. Comparison of five health status instruments for orthopaedic evaluation. *Med Care* 1990;**28**:632–42.
11 Cleary P, Reilly D, Greenfield S, Mulley A, Wexler L, Frankel F, *et al.* Using patient reports to assess health-related quality of life after total hip replacement. *Quality of Life Research* 1993;**2**:3–11.
12 O'Boyle C, McGee H, O'Malley K, Joyce C. Individual quality of life in patients undergoing hip replacement. *Lancet* 1992;**339**:1088–91.
13 Fletcher A, Gore S, Jones D, Fitzpatrick R, Spiegelhalter D, Cox D. Quality of life measures in health care II: design, analysis, and interpretation. *BMJ* 1992;**305**:1145–8.
14 Jenkinson C, Bardsley M, Lawrence K. Measurement in subjective health assessment: themes and prospects. In: C Jenkinson, ed. *Measuring health and medical outcomes.* London: UCL Press 1993;176–86.
15 Patrick D, Deyo R. Generic and disease-specific measures in assessing health status and quality of life. *Med Care* 1989;**27**:S217–32.
16 Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruements. *J Chron Dis* 1987;**40**:171–8.
17 Fitzpatrick R, Ziebland S, Jenkinson C, *et al.* Importance of sensitivity to change as a criterion for selecting health status measures. *Quality in Health Care* 1992;**1**:89–93.
18 Garratt A, Ruta D, Abdalla M, Russell I. SF 36 health status questionnaire II. Responsiveness to changes in health status in four common clinical conditions. *Quality in Health Care* 1994;**3**:186–92.
19 Liang M, Larson C, Cullen K, Schwartz J. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum* 1985;**28**:542–7.
20 Wiklund I, Romanus B. A comparison of quality of life before and after arthroplasty in patients who had arthrosis of the hip joint. *J Bone Joint Surg* 1991;**73**:765–9.
21 Feinstein A, Wells C. A new clinical taxonomy for rating change in functional activities of patients with angina. *Am Heart J* 1977;**93**:172–82.
22 Deyo R, Inui T. Towards clinical applications of health status measures: sensitivity of scales to clinically important changes. *Health Serv Res* 1984;**19**:275–89.
23 Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A, Mowat A. Transition questions to assess outcomes in rheumatoid arthritis. *Br J Rheumatol* 1993;**32**:807–11.
24 Meenan R, Gertman P, Mason J, Dunaif R. The arthritis impact measurement scales: further investigations of a health status measure. *Arthritis Rheum* 1982;**25**:1048–53.
25 Kay A, Davison B, Badley E, Wagstaff S. Hip arthroplasty: patient satisfaction. *Br J Rheumatol* 1983;**22**:243–9.
26 Steele R, Dempster D, Smith M. Total hip replacement of failed surface arthroplasty. *Acta Orthop Scand* 1985; **56**:133–4.
27 Stromberg C, Herberts P, Palmertz B. Cemented revision hip arthroplasty. A multicentre five to nine year study of 204 first revisions for loosening. *Acta Orthop Scand* 1992;**63**:139–42.
28 Charnley J. Long term results of low friction arthroplasty of the hip performed as a primary intervention. *J Bone Joint Surg* 1972;**54**:61.
29 Ware J, Sherbourne C. The MOS 36 item short form health survey: 1 conceptual framework and item selection. *Med Care* 1992;**30**:473–83.
30 Dawson J, Fitzpatrick R, Carr A, Murray D. Development and validation of a questionnaire to assess patients' perceptions in relation to total hip replacement surgery. *J Bone Joint Surg* 1996;**78**:185–90.
31 Murray D. The hip. In: P Pynsent, J Fairbank, A Carr, eds. *Outcome measures in orthopaedics 1993.* Oxford: Butterworth-Heinemann, 1993:198–227.
32 Callaghan J, Dysart S, Savory C, Hopkinson W. Assessing the results of hip replacement. A comparison of five different rating systems. *J Bone Joint Surg* 1990; **72**:1008–9.
33 Ruta D, Abdalla M, Garratt A, Coutts A, Russell I. SF 36 health survey questionnaire: 1. Reliability in two patient based studies. *Quality in Health Care* 1994;**3**:180–5.
34 Brazier J, Harper R, Jones N, O'Cathain A, Thomas K, Usherwood T, Westlake L. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* 1992;**305**:160–4.
35 Meenan RF, Gertman PM, Mason JH. Measuring health status in arthritis: the arthritis impact measurement scales. *Arthritis Rheum* 1980;**23**:146–52.
36 Kazis L, Anderson J, Meenan R. Effect sizes for interpreting changes in health status. *Med Care* 1989;**27**(suppl): 178–89.
37 Hayes V, Morris J, Wolfe C, Morgan M. The SF-36 health survey questionnaire: is it suitable for use with older adults? *Age Ageing* 1995;**24**:120–5.
38 Amadio P. Outcome measurements: more questions, some answers. *J Bone Joint Surg* 1993;**75**:1583–4.
39 Radford P. In: P Pynsent, J Fairbank, A Carr, eds. *Outcome measures in orthopaedics 1993.* Oxford: Butterworth-Heinemann, 1993:59–80.
40 Kazis L, Callaghan R, Meenan R, Pincus T. Health status reports in the care of patients with rheumatoid arthritis. *J Clin Epidemiol* 1990;**43**:1243–53.
41 Shanks J, Frater A. Health status, outcome and attributability: is a red rose red in the dark. *Quality in Health Care* 1993;**2**:259–62.
42 Bardsley M, Venables C, Watson J, Goodfellow J, Wright P. Evidence for validity of a health status measure in assessing short term outcomes of cholecystectomy. *Quality in Health Care* 1992;**1**:10–4.
43 Wasson J, Keller A, Rubenstein L, Hays R, Nelson E, Johnson D. Benefits and obstacles of health status assessment in ambulatory settings. *Med Care* 1992;**30**: MS42–9.
44 Lansky D, Butler J, Waller F. Using health status measures in the hospital setting. *Med Care* 1992;**30**:MS57–73.
45 Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A, Mowat A. A comparison of the sensitivity to change of several health status instruments in rheumatoid arthritis. *J Rheumatol* 1993;**20**:429–36.
46 Katz J, Larson M, Phillips C, Fossel A, Liang M. Comparative measurement sensitivity of short and longer health status instruments. *Med Care* 1992;**30**:917–25.
47 Bombadier C, Melfi C, Paul J, Green R, Hawker G, Wright J, *et al.* Comparison of a generic and a disease-specific measure of pain and physical function after knee replacement surgery. *Med Care* 1995;**33**:5131–44.