Patient Reported Outcome Measures (PROMs) in Elective Surgery

Report to the Department of Health

Health Services Research Unit, London School of Hygiene & Tropical Medicine & Clinical Effectiveness Unit, Royal College of Surgeons of England

Dr John Browne

Dr Liz Jamieson

Dr James Lewsey

Dr Jan van der Meulen

Health Services Research Unit, London School of Hygiene & Tropical Medicine

Professor Nick Black

Professor John Cairns

Dr Donna Lamping

Dr Sarah Smith

Clinical Effectiveness Unit, Royal College of Surgeons of England

Miss Lynn Copley

Mrs Jackie Horrocks

12 December 2007

EXECUTIVE SUMMARY

EXECUTIVE SUMMARY

1. FEASIBILITY

Where and when should patients be recruited?

According to both the preferences of patients and the recruitment rates achieved, there is little difference between pre-op assessment clinics and on admission. However, there are some disadvantages to using the pre-op assessment clinic so **we recommend recruitment on admission**. Around 80% recruitment can be achieved irrespective of surgical procedure or type of facility.

How can recruitment be maximised?

The main cause of non-recruitment was a failure to invite patients to participate. Overall, about 25% of potentially eligible patients were not invited to take part. Failure to invite in many centres is understandable given the short period of time in which staff had to adapt their routine practices to the study protocol. **We recommend a longer period in which to develop local arrangements, staff training to identify eligible patients and demonstration of the benefits of auditing patients' care.** We believe these steps will lead to high invitation rates. However, the active support of senior clinicians and managers is also needed, as are clear responsibilities and disciplined procedures.

Which staff are best suited to the task?

Nursing staff are better placed than clerical staff to administer questionnaires, possibly because of their experience in interacting with patients. With correct training, however, there seems no reason why clerical staff should not be equally proficient.

How can ineligibility be minimised?

The ineligibility rate (3.9%) was low for hernia, vein, hip and knee surgery. It was rare for a patient to be incapable of completing a written questionnaire in English (though few centres were in areas with a high proportion of ethnic minorities). However, up to 30% of cataract patients at some centres were deemed unable to participate due, not surprisingly, to vision problems. **We recommend investigation of the use of an interviewer to administer questionnaires to some cataract patients.**

How can patient agreement to participate be maximised?

There was a wide range of results across centres (33% to 100%) with a few very poorly performing centres skewing the overall mean rate (87%). **We recommend the questionnaires be shortened by excluding some questions that have been shown to be redundant, eliminating the research consent form which would not be required, and supplying the information sheet separately**. We believe that the agreement rate can be increased to an acceptable level in any future routine PROMs programme.

How can deceased patients be identified in a timely fashion?

To avoid causing bereaved relatives and friends any distress, we used the NHS Strategic Tracing Service to identify any post-operative deaths. This was successful in eight of the eleven deaths that occurred. **We recommend use of the NSTS** which enabled us to limit the sending of post-operative questionnaires to only three deceased patients out of a total sample of 2310 patients.

How can a high post-operative survey response be achieved?

In the light of the high post-operative response rates achieved (80-90%), **we recommend that post-operative questionnaires be dispatched and returned to a central facility, separate from the centre performing the surgery.**

What is the overall cost?

Based on necessary (see below) and achievable recruitment (80%) and response (80%) rates, the best estimate of the cost per patient was about £6.50. About 22% was met locally (staff time recruiting patients and using NSTS to trace post-operative deaths), making the central cost about £5 per patient successfully recruited. Almost 60% the central cost was for data entry, a factor that could be reduced using different technology.

2. CENTRE COMPARISONS

Which PROMs should be used?

This project has confirmed the responsiveness of four disease-specific PROMs for assessing the impact of surgery. Concerns about the content validity and responsiveness of the VF-14 mean that a better instrument needs to be developed. **We recommend the use of:**

- **cataract surgery: VF-14** (until further research on the content validity of the VF-14 and treatment appropriateness in this area has been conducted)
- **hernia repair:** none (until a disease-specific instrument has been developed and tested)
- **hip surgery: OHS**
- **knee surgery: OKS**
- **vein surgery: AVVQ**

Although it is possible to map changes in disease-specific scores onto a generic PROM (EQ-5D), the association for some procedures (cataract, vein surgery) is weak or only moderate (hernia repair). If the benefits of different procedures are to be compared, **we recommend the inclusion of a generic utility measure.** There is little difference in responsiveness between EQ-5D and SF-6D. Given the greater time involved in completing the SF-6D (36 items compared with 5 items), the higher incidence of missing data and the higher cost (given that use of the SF-36 scale normally requires payment of a fee to the copyright owners) **we recommend the use of the EQ-5D**. This does not apply to cataract surgery where both generic measures were unresponsive.

In addition, given their importance to patients and clinicians, **we recommend the inclusion of patient-reported complications.** It is important to recognise that the absolute incidence of complications reported by patients may include post-operative experiences that clinicians would define as 'normal' (eg wound discomfort) and also events that may not relate to the surgical operation (eg a urinary infection). In addition, the reliability and validity of complications measurement has not been tested as rigorously as other PROMs. The use of patient-reported complications should therefore be restricted to comparisons (such as between providers or procedures).

What are adequate recruitment and response rates?

Patients who are less healthy (as determined by the EQ-5D) are less likely to be recruited and less likely to respond to post-operative questionnaires. Post-operative non-responders also tend to be younger than responders. Considering the recruitment and response rates achieved at some of the participating centres, **we recommend 80% recruitment and 80% response rates should be sought to reduce any biases to insignificant levels**.

How should data be analysed to compare centres?

We recommend imputing missing items for all PROMs. The advantage of imputing is that it retains data and maintains sample sizes that would otherwise be reduced. The clustering of patients nested within centres did not affect the centre comparisons. **We recommend reconfirming the lack of clustering with a larger data set.**

Given the marked differences in case-mix (patients' pre-operative characteristics) between centres, any comparison of outcomes must take such differences into account if it is to be meaningful. We developed risk adjustment models for each PROM which explained a large proportion of the variance in post-operative scores observed. **We recommend the following confounders be included to risk-adjust post-operative PROM scores: pre-operative PROM score, age, sex, general health status, comorbidity (eight systemic conditions) and previous similar surgery.** Duration of symptoms is not required. The inclusion of the IMD (Index of Multiple Deprivation) was useful only for knee surgery. **We recommend investigation of the impact that ocular comorbidity might have on risk adjustment for cataract surgery**

The models included all characteristics that it is feasible to collect from patients (rather than from clinicians), with the exception of height and weight**. We recommend improvement in these models by collecting much larger samples of centres and patients, and by the inclusion of some clinical factors (e.g. ocular morbidity for cataract surgery) either collected from clinicians or derived from existing clinical databases.**

Should 'minimally important differences' be employed to compare centres?

Both the anchor-based and distribution-based approach to generating MID values for PROMs are problematic: the former because a stable value, representative of the continuum of pre-op severity, cannot be generated; the latter because the value generated has no known relationship with patient experience. **We recommend that**

MID values be used with great caution when interpreting PROMs data and that statistical significance testing should remain the main approach to the comparison of centres.

What power and statistical significance levels should be used?

Given the undesirable consequences of both false positives and false negatives in quality of care research, stringent criteria should be used with respect to both statistical significance and power when comparing centre performance on PROMs. **We recommend that statistical comparisons of centre performance should be based upon 95% power and a significance level of $p < 0.002$ (i.e. 99.8% confidence intervals).**

What is an adequate sample size?

We recommend a sample of about 150 patients from each centre to make meaningful comparisons. This would be sufficient to detect a difference of 1 Standard Error of the Measure (for both disease-specific and generic measures) with 95% power and statistical significance of $p < 0.002$ (equivalent to 99.8% confidence intervals) for cataract surgery, hernia repair and varicose vein surgery. For hip and knee surgery this sample would be adequate for the generic measure but for the disease-specific measures the power would be slightly less.

3. SUMMARISING AND PRESENTING FINDINGS

What is the best way of presenting data?

We recommend the use of funnel plots that show both the unadjusted and adjusted data. This approach also allows more than one control limit to be shown, thus allowing the observer to impose their own level of certainty. This does not preclude also providing access to more detailed tabulated data, particularly for surgeons.

What is the best indicator for comparing centres?

We recommend using the mean post-operative PROM score adjusted for patients' pre-operative characteristics as the indicator of health gain. In addition, the proportion of patients reporting at least one complication (risk adjusted) should be used.

At what geographical level should centres be compared?

In view of the travel distance that patients are likely to find acceptable for common surgical operations, **we recommend that the performance of providers in the local 'region' be provided together with the national average**. Data for all providers in the country could also be made available, particularly for surgeons.

At what level of provision should comparisons be made?

We had sufficient data for making comparisons of centres, not individual surgeons. Despite patients and provider managers expressing an interest in comparisons of surgeons, current NHS policy on choice is limited to centres

and the outcome of surgery relates to the performance of a whole team, not just the operating surgeon. Therefore, **we recommend that at present the unit of comparison is the centre**.

How frequently should performance information be updated?

As data collection systems become more sophisticated, it may be possible to provide continuously updated information. The feasibility of setting up such a system should not be underestimated. In the meantime, **we recommend 'batch processing' that updates every six months.** Allowing for three month post-operative outcomes, comparative information on providers for hernia, vein and cataract surgery would relate to performance 5 to 10 months previous. For hip and knee replacement, which currently have a six month follow-up, the information would be 8 to 13 months previous.

1. INTRODUCTION

1.1 Patient Reported Outcome Measures

Patient Reported Outcome Measures (PROMs) are measures of health status or health-related quality of life (HrQL) that come directly from patients. As such the term, PROM, is a misnomer as they don't attempt to determine the outcome (impact) of a health care intervention but instead they assess a person's health status or HrQL at one point in time. The impact of a health care intervention is determined by comparing the patient's self-reported health status at two points in time (in surgery, before and after the operation). Despite being a misnomer, PROM has become an internationally accepted term so we adopt it in this report.

A wide range of reliable and valid PROMs, usually administered in questionnaire format, have been developed in recent years. Some of these were developed for particular conditions or treatments (disease-specific measures) while others were developed to facilitate comparisons between conditions or treatments (generic measures). The use of PROMs will ensure that, in addition to clinical measures of outcome, patient perspectives will be taken into account. In addition, there is interest in using utility-weighted generic PROMs to determine the relative cost-utility of different interventions to inform commissioning decisions in the NHS.

1.2 PROMs use in assessing quality of care in the UK

While PROMs are commonly included in studies of treatment effectiveness they are less often included in assessing health care quality. To learn more about the methods and experiences of those who have included PROMs in auditing elective surgery we interviewed 15 UK-based experts (Appendix 1). We also interviewed representatives of two organisations using PROMs in the USA. The methods used by these organisations are outlined in two case studies in Appendix 2. These interviews revealed that the use of PROMs in audit is still in its infancy and that research is required to resolve a number of important issues. The main findings were:

- Most PROMs use turned out not to be for comparative audit (e.g. for the purpose of comparing performance across institutions) but to monitor the care of individual patients (e.g. monitoring pain and discomfort following hernia surgery) or for research.
- BUPA has the most well developed, routine PROMs collection system for elective surgery in the UK at present.
- Preoperative PROMs are usually administered at either a pre-operative assessment visit (usually around two weeks before surgery) or on admission for surgery. In contrast, BUPA mail questionnaires to patients together with the letter inviting the patient for surgery.
- Postoperative PROMs are usually administered at a routine post-operative assessment clinic. However, not all patients are invited to attend such clinics, leading to incomplete follow-up. Instead, BUPA and the NHS Patient Satisfaction Survey use postal survey methods to collect post-treatment PROMs.
- Written consent to allow the administration of pre- or postoperative PROMs is not usually obtained from the patient. Completion of the questionnaire is taken as a sign of consent to participate.

- Where PROMs data are collected primarily for assessing quality of care (e.g. at BUPA hospitals), data are anonymised wherever possible and are not available in an identifiable format to the clinicians in charge of the patients care.
- Other than the NHS Patient Satisfaction Survey, there is little attempt to cater for patients at high risk of not being recruited or of attrition (e.g. patients with literacy problems). Most surveys do not send reminder letters to non-responders and the systematic tracing of patients who have died (e.g. by using the NHS Strategic Tracing Service) is rare.
- Statistical methods to deal with missing data in questionnaires (e.g. imputation) or to 'risk-adjust' outcomes are rarely employed. There is little agreement about how to discriminate between statistical and clinical significance when interpreting results.
- There is little familiarity with different methods for presenting PROMs data, though BUPA do use control charts to compare the performance of different health care providers.

We also worked with a private survey organisations, Clinical & Health Knowledge Services (CHKS) Ltd, to evaluate their use of PROMs in the NHS. By 2006, CHKS had established a PROMs database in seven NHS acute hospitals that covers a broad range of patients receiving acute care. As well as covering elective care, some hospitals are monitoring the outcomes of emergency patients. This large and broadly based database is growing and may extend to other trusts who have shown an interest in the initiative. The database provides a useful resource to address a number of substantive questions about the use of PROMs in the NHS. The proportion of elective patients for whom CHKS has managed to collect complete pre- and post-treatment PROMs data is very low, varying from 4% to 17% between hospitals. Post-treatment response rates for emergency patients are better at 31% to 34%.

1.3 Treatment Centres

Treatment Centres (TCs) are providers dedicated providing planned surgery and diagnostic procedures for a range of conditions. TCs may be housed in purpose-built facilities (which may be on a stand-alone site or on the campus of an existing hospital), mobile units which move from site to site, or use existing hospital buildings refurbished to suit their needs. Some TCs are 'virtual', in that they make use of existing hospital theatres, wards and/or staff and differ from their host hospital only in the nature of the contractual arrangements with Primary Care Trusts (Department of Health, 2006). TCs may be run by NHS Trusts or by independent sector providers. As the Department of Health (DoH 2007) explains on its website, the capital investment needed to establish TCs can come through NHS public capital, the NHS Private Finance Initiative, independent sector funds or a combination of all of these (e.g. public-private partnerships).

NHS TCs and Independent Sector TCs (ISTCs) have been introduced over the last four years to reduce waiting times for routine procedures and increase the availability of such treatments. The DoH decided that the quality of care in TCs needs to be carefully monitored and compared to the performance of traditional NHS delivery models. A range of Key Performance Indicators (KPIs) for ISTCs have been developed, one of which (KPI 15) is based on the use of Patient Reported Outcome Measures (PROMs).

In December 2005 we completed a review of the international literature on the most appropriate PROMs to use for five high-volume elective procedures. The final report for that review is freely accessible at: (http://www.lshtm.ac.uk/hsru/staff/documents/PROMsreport1205.pdf). Meanwhile, this feasibility study of implementing the routine use of PROMs had commenced in June 2005.

1.4 Aim and Objectives

The aim was to determine the feasibility of collecting pre- and post-operative PROMs from patients undergoing elective surgery and to investigate how such data could best be analysed and presented. The project had five objectives:

Determine the feasibility of data collection methods:

- proportion of ineligible patients (such as those who cannot complete a written questionnaire in English)
- proportion of eligible patients who are not invited before surgery to participate by local data collectors
- proportion of patients who decline to participate when invited
- proportion of patients who do not respond to a follow-up, post-operative questionnaire

Compare performance between centres:

- explore differences in data quality among centres
- adjust for pre-operative patient characteristics
- explore measures of the uncertainty in the centre-specific estimates (e.g. confidence intervals)

Methodological considerations

- provide estimates of the mean and standard deviation of the pre- and post-operative PROMs
- produce definitions of 'minimally important differences' for each of the chosen measures
- estimate the minimum volume of patients needed to discriminate between centres
- provide information on the psychometric properties of the PROMs where this has not previously been reported
- compare the responsiveness of generic and disease-specific measures and of different generic measures.

Summarising and presenting findings

- establish the effectiveness of different techniques for the summary, presentation and feedback of results to patients, clinicians, provider managers, commissioners, DoH and government agencies.

Enhancing the use of PROMs

- monitoring treatment appropriateness
- monitoring secular trends in utilisation and equity
- evaluating clinical and cost-effectiveness of operations

2. METHODS

2.1 Overall design

The project had a quantitative and a qualitative component. The quantitative part comprised five prospective cohort studies, each covering a particular elective surgical procedure: unilateral hip replacement, unilateral knee replacement, groin hernia repair, varicose vein surgery and cataract surgery. These procedures were selected by the DoH because they represent a variety of surgical complexity and speciality. Recruited patients were assessed before surgery and then at 3 or 6 months after surgery, depending on the procedure performed. Consecutive patients at each participating centre were potentially eligible for inclusion.

The qualitative component included interviews with a randomly selected group of patients and with local data collectors about the feasibility of routine PROMs use. In addition, two focus groups were held to explore stakeholders' views of the best means of presenting comparisons between surgical centres.

2.2 Quantitative study

Patient inclusion and exclusion criteria

Consecutive patients listed for unilateral hip or knee replacement, cataract removal, groin hernia repair and varicose vein surgery were invited to take part. The study was limited to those aged 16 years or over, as the PROMs chosen were designed for adults. We excluded patients undergoing bilateral joint replacements as the measures used (Oxford Hip and Knee Scores) were designed for unilateral surgery (alternative versions designed for bilateral surgery have not yet undergone full psychometric evaluation). Patients undergoing knee and hip resurfacing operations were eligible for inclusion, as were patients undergoing revisions of any of the five procedures.

Patients judged to be incapable of completing a written questionnaire in English (because of cognitive impairment, poor sight, literacy or language comprehension problems) were ineligible.

Sample size

We chose the response rate to the post-operative questionnaire as the determinant of sample size. We aimed to estimate a response rate for each surgical procedure with a confidence interval of +/-4%. To achieve this we needed 320 patients for each procedure which in turn required the pre-operative recruitment of approximately 450 patients per procedure (assuming that the response rate was 70%), making a total sample size of 2250.

Development of questionnaires

Each pre-operative questionnaire included an information sheet to be read by patients before consenting to participate and a consent form. Two carbon copies of both the information sheet and the consent form were included so that one set could be retained by the patient and one set could be retained by the centre in the patients' notes. The original set (still contained within the questionnaire booklet) was to be returned to the

research team by the local data collectors. Patients who consented to take part in the study were also asked to write their home address in the questionnaire so that they could be sent a follow-up questionnaire after surgery.

Sociodemographic data collected from each patient consisted of age, sex and postcode. The latter was used to determine an Index of Deprivation for each patient. All pre-operative questionnaires contained the first part of the EQ-5D, a generic measure of health-related quality of life (EuroQol Group 1990). The second part of the EQ5D, a Visual Analogue Scale, was not included as we did not require patient valuations of HrQL. Each questionnaire also included a disease-specific measure of health-related quality of life:

- VF-14 for cataract removal (Steinberg et al. 1994)
- Aberdeen Varicose Vein Questionnaire (AVVQ) for varicose vein surgery (Garratt et al. 1993)
- Oxford Hip Score for hip replacement (Dawson et al.1996)
- Oxford Knee Score for knee replacement (Dawson et al. 1998)

The literature review we conducted revealed that no disease-specific measure for hernia surgery has yet undergone appropriate psychometric development. We judged, on the basis of usage in the literature, that the SF-36 (UK version 2), although a generic measure, was an appropriate substitute for a disease-specific measure. When analysing the SF-36 results we derived both the Physical Component Summary score (PCS) and the Mental Component Summary score (MCS).

We also used this version of the SF-36 in patients undergoing hip replacement surgery (in addition to the Oxford Hip Score). This was to allow a comparison of the relative responsiveness of the EQ-5D and the SF-6D (a utility measure derived from the SF-36v2 using an algorithm supplied by John Brazier of the University of Sheffield) in patients undergoing moderately severe surgery (a similar comparison was also performed in hernia patients – i.e. patients undergoing surgery of minor severity). The pre-operative questionnaire also contained questions relating to the patient's age, sex, duration of symptoms, history of previous similar surgery and general health status. In addition, patients were asked to indicate which of 21 co-morbidities they suffered from. The list was based on work by Bayliss et al (2005):

- Angina, heart disease or irregular heart beat
- Poor blood circulation in your legs
- High blood pressure
- Nerve condition such as Parkinson's disease, multiple sclerosis or neuropathy
- Chronic headaches
- Psychological problem such as depression or anxiety
- Asthma
- Chronic bronchitis or emphysema
- Osteoarthritis
- Rheumatoid arthritis
- Osteoporosis
- Chronic back pain or sciatica

- Cancer (within the past 5 years)
- Diabetes
- Rheumatic disease such as fibromyalgia or lupus
- Stomach problem such as an ulcer or reflux disease
- Irritable bowel, colitis or Crohn's disease
- Thyroid disorder
- Kidney disease
- Urinary problems
- Liver disease

The post-operative questionnaires contained the same PROMs as those used in the pre-operative period. In addition, patients were asked to confirm the date that surgery took place and which side of the body the operation was on. There was also a question relating to four generic post-operative complications commonly experienced (allergy or reaction to drug, and urinary, bleeding or wound problems) which had been developed for auditing day surgery in the 1990s (Black & Sanderson 1993). The aim of this question was to elicit patients' perceptions of their post-operative experience. The implication is that if a patient experienced some discomfort from a wound, they may report that as a 'wound problem' even though a surgeon may deem their experience as 'normal and expected' rather than an abnormal, adverse outcome. Also, if a patient experiences urinary symptoms during the post-operative period, they may ascribe this to the operation whereas a clinician may perceive it as being unrelated. For these reasons, we would, therefore expect the incidence of patient reported 'complications' to be higher than that reported by clinicians. Whilst these factors need to be taken into account when considering the absolute incidence of complications, it doesn't effect the validity of inter-centre comparisons.

Finally, patients were asked to rate the overall success of the operation on a five-point scale and to rate how their current symptoms compared to those they experienced before surgery. Although these 'global' questions were included, experience suggests that they are of limited value (Streiner & Norman 2003). In most studies, about 85% of patients tend to report that they feel better and have no regrets about having undergone surgery regardless of the actual impact an operation has had. This probably reflects partly their delight in having come through surgery without encountering any serious adversity and partly an understandable need to justify to themselves that the decision to undergo surgery had been the correct one. Despite such reservations we included these questions given the pilot nature of this study.

Identifying ineligible patients

Local staff were requested to judge whether they considered a patient was capable of completing a written questionnaire in English on their own, before inviting the patient to participate. This judgement was based on their assessment of the patient during the clerking process. Staff were requested to provide some information on ineligible patients on a 'patient participation sheet' (Appendix 3) by placing a tick against one of the following:

- patient does not understand English (primary language to be noted)

- patient has literacy problem
- patient visually impaired
- patient cognitively impaired
- other reason

Consent to participate

Eligible patients were invited to participate in the study. They were asked to read the information sheet and sign the consent form. If a patient declined this was noted on the patient participation sheet.

Using the NHS Strategic Tracing Service

Before post-operative questionnaires were mailed, we attempted to confirm that the patient was still alive. Each participating centre was requested to use the NHS Strategic Tracing Service (NSTS) to ascertain whether a patient was recorded as having died. At most NHS Trusts permission to use NSTS was arranged through the Caldicott Guardian. The NSTS is a free service available to all NHS Trusts. Patients can be traced either in batches or individually on-line over the NHS net. Batch tracing requires the patient's NHS number. Using the patient's name and postcode it was possible to search for a death record using the advanced search facility on the NSTS software.

The time between a patient's death and a record of that death appearing on the system varies and depends on how the NSTS is notified. When a patient dies the NSTS can be notified by either the Primary Care Trust (PCT) in which the patient is registered, the NHS Central Register (NHSCR), or both. Information relating to a patient's death is provided to the NHSCR by the Registrar of Births and Deaths. The majority of deaths will be notified to the NSTS by NHSCR within six weeks of the death occurring. The PCT may be informed by the GP when the patient dies and this information can be available to the NSTS more quickly, often within a week. This had important implications for our ability to stop an inappropriate follow-up questionnaire being sent out to a recently deceased patient.

Lists of patients whose post-operative questionnaires were due to be sent out were provided to the centres monthly by the research team. Patients' names and their postcodes, in the form of a password-protected Excel table, were e-mailed to the appropriate person (usually someone from information services suggested by the Caldicott Guardian).

All but one NHS Trust was registered with the NSTS. The local contact at the non-registered Trust checked their own systems and alerted the research team if any deaths were found. Formal procedures for death notification could not be arranged with the private sector providers.

Post-operative survey

Post-operative questionnaires were mailed to patients' homes by the research team from the Royal College of Surgeons of England. The questionnaires were sent three months after surgery to patients who had undergone

cataract, hernia and varicose vein procedures, and six months after surgery to patients who had undergone joint replacement. Non-responders were sent a reminder letter and replacement questionnaire five weeks after the original mailing. No further attempt was made to contact non-responders.

Patients were advised in the covering letter sent out with the follow-up questionnaire that they could call the research team on a freephone number if they had any queries about the questionnaire.

2.3 Qualitative study

Interviews with local data collectors

Semi-structured interviews with local data collection staff (either as individuals or in groups) were conducted as soon as possible after patient recruitment was completed at each centre. The interview schedules used are shown in Appendix 4. At least one representative per operation group who had been involved in administering questionnaires was included. The purpose of the interviews was to explore the efficiency of the data collection process, to discover any problems that had arisen, and for staff to suggest ways the process could be improved. All staff interviewed were asked to sign a consent form and give details of their grade and title and the approximate number of patients in the operation group(s) they had recruited. The interviews took about one hour. Most staff invited were willing to attend and contribute to the discussion. Shorthand notes were taken and these were written up, usually within 24 hours of the discussion.

Interviews with patients

Patients who had completed both a pre- and post-operative questionnaire were eligible for inclusion. We aimed to select randomly at least one patient in each operation group per centre. The selected patients were contacted by letter to request their consent to participating in a short, semi-structured telephone interview with a member of the research team. Those consenting were requested to offer three suitable dates and times for the interview. If a patient had not responded within four weeks, another patient was randomly selected. We aimed to continue this process until one patient from each operative group had consented or thematic saturation had been reached.

The focus of these interviews was the feasibility of data collection (e.g. what aspects of the questionnaire survey patients liked and disliked) and how the methods used could be improved. The interview schedule used is shown in Appendix 5.

Review of published methods of data presentation

A search of the Internet was conducted to identify examples of ways of presenting data that compares the surgical performance of health care providers. All of the examples found were from the USA. The search continued until no further presentations methods were found. Three types of presentation were found: categorisation - results divided into three or more categories and often presented in terms of number of stars; bar charts - graphs with bars indicating the outcome for each centre, displayed either vertically or horizontally; tables - providing the actual numerical results. Examples found include:

Categorisation:	HealthGrades - five 'star' categories of complication rates
	Maryland Health Care Commission - three 'star' categories of readmission rates
	PHC4 Hospital Performance Report - three categories of readmission rates and mortality
Bar charts:	Solucient 100 Top Hospitals - mortality and complications compared with 'peers'
	New Jersey DHSS - 30-day post-op mortality compared with rest of State
	Pacific Business Group - mortality compared with rest of region
Tables:	Pacific Business Group - mortality compared with rest of region

We decided to also consider a fourth method, a funnel plot. This is a statistical process control technique that seeks to measure deviation from expected outcomes. The centres are plotted on a scatter chart with the number of operations on the x axis and the outcome on the y axis. Control limits denoting two and three standard deviations above and below the overall mean outcome are used to detect outliers. The funnel plot owes its name to the characteristic shape of the control limits, which flare as they approach the y-axis. This reflects the reduced confidence levels associated with low activity.

Focus groups of stakeholders

To obtain the views of potential stakeholders about how to best present PROMs data we ran two focus groups. One group considered cataract surgery, hernia repair and varicose vein surgery; the other group considered hip and knee replacement. The aim was to include in each group: six surgeons, three nurses, three provider managers, two commissioners, and two patients. Members of the research team were in attendance. The meetings lasted for two hours. After an initial presentation of the study, the groups were asked their views of:

- how data should be presented (tables, bar charts, categories, funnel plots)
- which indicators should be used to compare providers (average improvement in PROM score; proportion achieving a minimally important difference; complication rate; overall self-assessment)
- which providers should be included in comparisons (local ones only, regional, national)
- who should be compared (operating surgeons, consultant surgeons, hospital/treatment centres, Trusts)
- how frequently comparisons should be up-dated (monthly, quarterly, six-monthly, annually)

2.4 Administration

Research ethics approval

Ethical approval was obtained from the Wales MREC in October 2005.

Recruitment of centres

Although the study was initially to cover only TCs, the DoH decided that it would be useful to also investigate the feasibility of PROM use in acute NHS hospitals as any future performance assessment system should be extended to all providers. In order to represent the range of organisational designs in each of the three sectors under study (NHSTCs, ISTCs and NHS acute hospitals) we aimed to recruit 30 centres, 10 from each sector. We

also tried to ensure that the different geographical regions of England were represented. It was also our intention to recruit 10 providers of each of the five procedures.

From the outset it was agreed with the DoH that only centres which had demonstrated a previous enthusiasm to participate in outcome assessment projects would be approached to participate. The DoH identified NHSTCs (via NHS Elect) and NHS acute hospitals (mainly 'champions' of quality improvement in day surgery) who might be willing to participate. There was difficulty in obtaining contact details for ISTCs, as there was considerable confusion about whether or not this sector was already collecting PROMs data. Eventually, contact details were obtained for only two ISTCs (both managed by the same company) and one private hospital whose workload is entirely elective surgical work for the NHS. As a result we are not able to compare ISTCs with NHS providers.

All centres identified as possible participants were contacted by the research team. Those expressing an interest were provided with more details of the protocol. The means of obtaining local research governance approval was also discussed. Average monthly volumes for the relevant operations performed were obtained to allow us to estimate the likely periods of patient recruitment at each centre.

The research team was contacted by representatives of an NHS General Practice in which groin hernia surgery is performed. It was agreed with the DoH that it would be interesting to examine the feasibility of PROMs use in this setting.

Local research governance approval

For each participating NHS provider, application for research governance approval was made to the R&D department of the relevant NHS Trust. The procedure and requirements varied between Trusts. Most R&D departments required a copy of the MREC application and approval letter, copies of the research protocol and all supporting documentation. Some R&D departments required the completion of the R&D form on the COREC website. Nine Trusts required an honorary contract to be issued for the lead researcher (LJ) on the grounds that she would be talking to hospital staff and patient representatives on hospital premises. No Trust required her to have an honorary contract because of the use of patient-completed questionnaires. Approval took between two weeks and two months to receive, depending on the regularity of committee meetings to consider applications and the timing of receipt of our documents.

It was often not until local R&D approval was obtained that our contact person at each centre could discuss the centre's participation with clinical staff and others likely to be affected. This caused a delay in the start of recruitment at some centres. At one centre, a consultant surgeon explicitly requested that his patients not be invited to participate in the study. At another centre the study was started and then stopped for a few weeks because two consultants and one specialist nurse were already carrying out PROMs assessment and did not want to confuse or overburden their patients. These were the only instances where explicit non-participation requests from clinical staff were made. We are aware, however, that patients under the care of some other consultants were excluded from the study on a systematic basis at a small number of centres.

Research governance approval at the participating ISTCs was obtained from the local medical advisory committee and local clinical governance staff. Such approval was not deemed necessary by the private hospital that participated.

Pre-study site visits

Pre-study site visits were made to all participating centres by LJ. These varied in formality and attendance. All sites were given a short presentation outlining the study objectives and methods. Staff were given a study handbook (Appendix 6) which outlined the methods and provided answers to anticipated questions. Short summaries of the study protocol (Appendix 7) were also provided so that these could be handed out to those who were not able to attend the meeting. Draft questionnaires were available for inspection during the site visit. The procedures for identifying eligible patients and obtaining patient consent were explained.

At each centre a contact person was identified for LJ to liaise with on a regular basis during the patient recruitment period. Each centre was contacted at least once a fortnight by telephone or e-mail during the patient recruitment phase and any problems arising were noted and, if possible, solved.

Recruitment of patients: who, how and when

The exact time and location at which patients were identified and, if appropriate, recruited varied by centre. Some centres chose to use the pre-op assessment clinic and some chose on admission for surgery. A number of considerations influenced the choice, including time pressures in each setting, availability of staff to administer the process, and the amount of physical space available.

2.5 Analyses: Feasibility of data collection

Patient recruitment and post-op response rates

At the end of data collection at each centre, the contact person was asked for the number of patients per procedure treated during the recruitment period (N1). Theatre management systems, theatre logbooks and other hospital administrative databases were suggested as possible sources of data. Additional information was collected or derived in the following ways:

Nx	Patient did not attend for surgery or operation/pre-op assessment was cancelled.
N1	Baseline figure of total number of patients meeting criteria going through pre-op assessment clinic/admission for surgery within data collection period who had surgery.
n1	Patients who were unable to participate: visual impairment, cognitive impairment, language difficulty, literacy problems, other reasons.
N2	Number of potentially eligible patients (N1 – n1).
n2	Number of patients not invited .
N3	Patients invited to take part (N2 - n2).
n3	Number of patients invited who declined to take part.

N4 Patients participated (N3 - n3).

N4/N2 Proportion of eligible patients participating.

N4/N3 Proportion of invited patients participating.

We compared the recruitment rates by procedure and by the method of recruitment (pre-operative assessment clinic versus on admission). To investigate recruitment bias, we compared the case-mix of centres by their recruitment rate, though this did not allow us to make an accurate estimate of any bias as we had no data on the characteristics of non-recruited patients. In contrast, we were able to assess the extent of response bias in the post-operative information as we had detailed data on non-responders from their pre-operative questionnaires.

Cost-effectiveness of data collection

We estimated the average cost per successful unit of data collection (i.e. a patient completing both the pre- and post-operative questionnaire) for each procedure, and for the two main approaches to patient recruitment (pre-operative assessment or on admission). Costs included questionnaire printing, postage to and from the centres and patients, staff time recruiting patients, data entry and analysis. We explored the relationship between the recruitment and post-operative response rates and the costs in a sensitivity analysis. We also collected estimates from other organisations engaged in similar audits to give a comparative context to our cost estimates.

Interview data

Responses in the semi-structured interviews were examined and a simple thematic analysis resulted in the identification of issues of concern to respondents. As the interviews were not audio-taped, no attempt was made to include verbatim quotes in the analyses.

2.6 Analyses: Comparison of performance between centres

Descriptive analyses

Summary statistics of age, sex, duration of symptoms, previous surgery, comorbidity and pre-operative health status are reported by centre. Similarly, post-operative health status and complications are described.

Funnel plots

Funnel plots were used to compare the performance between centres. In these, the mean post-op PROM scores for each centre are plotted against the number of patients in each centre (volume), with a benchmark (and its 95% and 99.8% control limits) at different volumes superimposed. The benchmark is the mean post-op PROM score across all centres.

Funnel plots were also used to compare the performance between centres in terms of post-operative complications.

Risk-adjustment of outcomes was conducted using linear regression (logistic regression in the case of incidence of complications) with pre-operative PROM scores and information on patient characteristics (age, sex, general

health status, co-morbidity, index of multiple deprivation) being potential risk factors. A significance level of 0.1 was used to develop a risk-adjusted model to strike a balance between model parsimony and missing out important risk factors due to low power.

2.7 Analyses: Methodological considerations

Dealing with missing items

The SF-36 rule for imputing missing items was followed for the generic and the disease-specific PROMs. Where there are missing items on a questionnaire, the mean response from the items the patient has completed were imputed if the patient had completed at least half the total number of items. For the EQ-5D we imputed the mode rather than the mean. To check the suitability of these methods, overall mean PROM scores were compared when missing item imputation was and was not conducted.

Dealing with missing questionnaires

No attempt was made to impute missing questionnaires. The post-operative response rate was calculated for each centre. If the patients who responded were systematically different from the non-responders and the response rate was low, the comparison of performance between centres may be biased. To assess this the pre-operative characteristics of patients who did and did not respond post-operatively were compared for each operation.

Non-independence of data due to clustering effects

Most standard statistical techniques assume that data are independent. This assumption would be violated for PROMs data if patients within a centre were more similar than patients across centres. To test this, multilevel regression models measured the amount of variance in outcome between centres.

Minimally important differences

Following the recommendations of Crosby et al. (2003) we derived numerical values for a minimally important difference (MID) for each PROM using both an anchor-based method and a distribution-method that takes the reliability of the instrument into account. With this integrated approach, one can define a MID considering the changes that are recognised as being important from the point of view of the patients while also considering whether the defined MID is beyond the measurement error of the PROM instrument.

The anchor-based method involved calculating the mean change in PROM scores for patients who reported that their symptoms were "a little better" after surgery (Jaeschke et al. 1989). An advantage of anchor-based methods is that they can provide an explicit link with patient-reported changes in health status. However, this method may lack stability across the pre-operative severity continuum, with lower MID values applying to patients with lower severity. This phenomenon has been observed in previous research with patients with back pain and in obese patients (Kolotkin et al. 2002; Stratford et al. 1998). We used Pearson's correlation coefficient to examine whether the anchor-based definition of a minimally important difference was consistent across pre-operative PROM scores. We also conducted a simple comparison of MID values produced by patients with

above average (median) and below average pre-operative severity. In these analyses we used only patients who reported that their symptoms were "a little better".

The Standard Error of the Measurement (SEM) method has been suggested to produce a distribution-based MID that is also dependent on the measurement's reliability (Crosby et al. 2003). An important characteristic of the SEM method is that it is to a large extent sample-independent, or in other words, it is relatively constant when measured in different samples of patients. The SEM was calculated using the following formula (Wyrich et al. 1999):

$$SEM = SD\ (1\text{-}r)^{½}$$

where SD is the standard deviation of the pre-operative PROM values and r is a measure of the reliability of the measure. When calculating the SEM we used the published intra-class correlation coefficient (ICC) for each PROM, or other reliability measures if the ICC was unavailable, as our best estimate of the instrument's reliability.

A limitation of the estimation of SEM values is that the literature on reliability measures (i.e. class correlation coefficient or ICC) for generic measures is poor at the level of individual surgical procedures. We therefore used population-based estimates of test-retest reliability for the EQ-5D and the SF-36 PCS and MCS scales. For the EQ-5D an ICC of 0.84 was used (van Agt et al. 1994) and for the SF-36 an ICC of 0.89 (Brazier et al. 1992) was used for the PCS. For the SF-6D, which is derived from the SF-36 and reflects both physical and mental health status, we used a simple average (0.845) of the ICCs for the PCS (0.89) and MCS (0.80) scales (Brazier et al. 1992).

Reliability measures for disease-specific measures (other than the Oxford Hip Score) were outlined in our earlier report (see section 1.3). A Pearson correlation coefficient of 0.88 was found for the Oxford Hip Score (Dawson, personal communication) and we used this as our best estimate of test-retest reliability for this measure.

Responsiveness of PROMs

The literature on responsiveness for all the PROMs used was poor with respect to the surgical procedures and treatment settings under study. The responsiveness of both generic and disease-specific PROMs was therefore calculated and compared. To do this we used 'effect sizes' to characterise the magnitude of change following surgery. Effect sizes were calculated by dividing the mean change score by the standard deviation of the pre-operative score. By convention an effect size of 0.2 is considered small, 0.5 is moderate and 0.8 or greater is large (Kazis et al. 1989)

Sample size estimates

Using the estimates for MID and the standard deviation, we produced sample size estimates specific to each PROM and surgical procedure. In all these calculations we assumed that a relative benchmark of performance (e.g. the national average) will be used. There is an ongoing debate about the level of 'false positives' that are

acceptable auditing quality of care and what level of statistical significance should be used. For example, the Association of Cardiothoracic Surgeons of Great Britain and Ireland use 99.8% confidence intervals in their comparative audit. We therefore provide sample size estimates both for significance levels of 0.05 and 0.002 to illustrate the impact the choice of the significance level would have. We also demonstrate the impact that different approaches to avoiding false negatives (e.g. avoiding failure to detect an underperforming centre) have on sample size estimates by using both 80% and 95% power levels.

Mapping disease-specific onto generic measures

One of the main advantages of generic measures such as the EQ-5D and the SF-36 is the possibility of comparing the benefits accrued by patients undergoing different treatments. It would seem valuable therefore to be able to map changes in disease-specific measures onto changes in generic measures. Specifically, it would be valuable to know the magnitude of change in a generic measure that corresponds to a one point change in a disease-specific measure. To provide these values we used simple linear regression to model the relationship between change scores for disease-specific measures and change scores for generic measures, using the generic change scores as the dependent variable.

2.8 Analyses: Establishing the best ways of summarising and presenting findings

Members of the research team made notes during the focus group discussions. In addition the proceedings were audio-taped, transcribed and analysed to identify the views of each category of stakeholder and to obtain verbatim comments.

3. RESULTS: Feasibility of data collection

3.1 Centre recruitment

Nineteen NHS acute hospitals, 15 NHS TCs, two ISTCs, one private hospital (exclusively treating NHS funded patients) and one NHS General Practice were contacted regarding participation in the study. In practice, three of the NHS TCs had not yet opened and one did not perform any of the target procedures. Four of the remaining NHS TCs and six of the NHS acute hospitals declined to participate. The most common reasons given for not taking part were lack of time and resources and previous participation in a pilot for the NHS patient survey. Both ISTCs, the private hospital and the NHS General Practice took part.

At two of the participating NHS acute hospitals, the eye units responsible for cataract surgery initially agreed to take part in the study and were provided with all relevant training and materials but subsequently did not attempt data collection because of time and resource constraints. At a third NHS acute hospital the local data collection staff stated that patients were resistant to taking part in the study because they found it difficult to read the questionnaires. Recruitment was therefore abandoned after a brief period.

Patient recruitment was therefore undertaken at 24 centres: 7 NHS TCs, 13 NHS acute hospitals, two ISTCs, one NHS General Practice and one private hospital (Table 3.1). The 13 NHS acute hospitals contained three types of organisational unit (day surgery units, eye units and orthopaedic units). Where relevant, we present results for these separately, as they presented particular feasibility challenges.

Table 3. 1: Type of participating centre by procedure

Procedure	NHS TC	NHS acute hospital	ISTC	Private hospital	General Practice
Hernia	6	9	1	1	1
Cataract	3	5	1	0	0
Varicose vein	6	8	1	1	0
Hip replacement	2	7	1	1	0
Knee replacement	2	7	1	1	0

3.2 Concerns raised by staff at pre-study site visits

Major concerns about the use of PROMs for audit were rare among local data collectors and clinicians. Only one centre voiced concerns about the validity of PROMs, and this related more to the use of PROMs in isolation from clinical data rather than a concern about their inherent value. Staff at one NHS centre commented that they did not want their centre to be compared to TCs that ”cherry pick” patients as they felt this would be an unfair comparison.

The following concerns about the methods used in this study were raised:

Criticism of questionnaires

- Staff at some centres felt that the questionnaires were too long.

- The colour of the type face used on the cataract questionnaire (blue) was not appropriate for cataract patients and black type on yellow or white paper would be preferable.
- It would be helpful to patients to have a staff member stick an address label from the patient file rather than have them fill it in themselves. It was agreed that this could be done.

Patients discussing questionnaires with others

- Patients might ask relatives to assist them in completing the questionnaire, or might discuss the questionnaires with other patients present in a waiting area. LJ advised that all that could reasonably be done was to ask patients to complete the questionnaire on their own and that staff should not help answer or interpret any of the questions.

Feasibility for certain patient groups

- Patients undergoing cataract surgery are usually considerably older, more anxious and less able to complete questionnaires by themselves than other patients undergoing surgery. These patients also have their pupils dilated prior to surgery, which might make the day of surgery an inappropriate time for questionnaire completion. In addition, some patients would be unable to hold a pen due to arthritis, or might forget to bring their reading glasses. In response to all the above LJ advised that one purpose of the study was to gauge the prevalence of such difficulties.

Existing use of PROMs

- At four centres, some or all of the patients were already being asked to complete PROMs as part of a research or audit project. At one centre it was agreed that the data collected in this study would substitute for the data normally collected. At a second centre the patients continued to complete the PROMs used locally in addition to those used in this study. At a third centre patients undergoing a particular type of knee replacement who were enrolled in a research study which required the use of PROMs, were excluded from this study.
- A small number of centres interrupted their routine use of patient satisfaction questionnaires for their duration of participation in our study.

3.3 Staff and time point used to recruit patients

No constraints were imposed by the research team as to who should recruit the patients. Either administrative or nursing staff were given responsibility for this task.

Patients were recruited either at a pre-op assessment clinic or on admission. The choice was constrained by the requirement that the time between questionnaire completion and the day of surgery should be minimised (to ensure responses were not out-of-date). In some centres this gap was lengthy (particularly for patients undergoing hip or knee replacement) which meant recruitment at the pre-op assessment clinic was inappropriate. One centre found a way to deal with this problem by forwarding only those questionnaires completed by patients who had a date of surgery booked within a month of their attendance at the pre-op

assessment clinic. A further problem with using the pre-op assessment clinic was that some centres did the assessment by telephone for the 'fitter' patients.

3.4 Difficulties experienced in recruiting patients

Most centres experienced no problems recruiting patients. In general, it did not take patients very long to complete their questionnaires. However, some problems were identified:

Physical re-location problems

Two centres experienced difficulties with the location at which recruitment was supposed to take place. The day surgery unit at one NHS hospital was decorated during the recruitment period and was temporarily re-housed at another site. It moved back towards the end of the recruitment period. At another NHS hospital the pre-op assessment unit was re-located. At both sites staff did not feel these moves greatly affected recruitment.

Giving questionnaires to patients to complete at home

In two centres patients were given questionnaires to take home following attendance at a pre-assessment clinic. This occurred because several nurses on different shifts were involved in recruitment and information about the study may not have filtered down to all staff. This procedure was stopped immediately it came to light. Despite this, patients treated at some centres continued to take the questionnaire home.

Filing consent forms

At some centres staff said that they had difficulty filing hospital copies of patient consent forms because patients' records were returned to the medical records department very quickly.

Interaction between patients

At one centre, two patients who declined to participate set off a chain reaction of refusals: "a vociferously unpleasant patient who refused to do the questionnaire and put the whole open-planned waiting room in an edgy mood, which had an impact on increasing the refusal rate". At another centre, one cataract patient became extremely distressed by the process and tore up the booklet.

Overwhelming demand on some patients

At a small number of centres staff felt that some patients who declined when invited to participate at a pre-op assessment clinic may have been overwhelmed by the pre-operative assessment process and the other questionnaires or paper work they were being asked to complete.

At one centre a patient did not understand that the consent process related to research rather than an aspect of their care. This may have been because of the number of other forms the patient had to complete before surgery.

Incomplete consent forms/addresses

A small number of patients signed the consent form in the questionnaire but did not tick the box which stated that they had read and understood the information sheet. More seriously, a small number of patients completed the questionnaire but did not sign the consent form at all. The data collected from these patients could not, therefore, be used.

There were approximately 10 instances of questionnaires returned to the research team with no home address for the patient. In these cases, the centres were approached and addresses were supplied.

Change of venue for surgery

A few patients did not have their surgery in the same centre as their pre-operative assessment. This arose when the assessment uncovered a condition that required an overnight stay rather than day surgery.

Difficulty of providing additional information to patients

At one centre staff did not feel that they had sufficient time to discuss the project with patients.

Lack of compliance with surgical inclusion criteria

A few cases of umbilical hernia operations were included. In contrast, at two centres patients undergoing hip/knee replacement revision procedures and resurfacings were excluded. The main reason for these deviations occurring was that information about the study was not always passed on fully to all participating staff.

Staff assisting patient to complete questionnaires

One centre had a member of staff who helped patients complete the questionnaire. Once spotted, this was stopped.

Problems completing the patient participation sheets

Most centres completed the patient participation sheets although a few pages in some centres were “lost” at the end of the recruitment period. In some instances mistakes were made. For example, certain patients were labelled as ‘non-consenters’ when in fact they had simply not brought their glasses and could not read the questionnaire.

Lower volumes of patients undergoing surgery than had been expected

At many centres, the number of patients undergoing varicose vein, hip and knee surgery were lower than expected. At these centres recruitment had to continue for longer than expected. The lower volumes seem to have been due to a combination of financial cutbacks (leading to reduced numbers of patients listed for surgery, centres no longer taking referrals, closed wards and staff reductions) and more restrictive criteria for surgery being introduced (particularly for varicose vein surgery).

3.5 Recruitment rates by centre and procedure

Recruitment rates by centre and by procedure are shown in Table 3.2. Each row provides data on one procedure at one centre. In the text, these are referred to as a 'centre/procedure'. Despite repeated attempts, it was not possible to obtain some data (designated Not Available in the table). For six of the 59 centre/procedures, the amount of missing data was so great as to exclude them from the table. There were three reasons for not recruiting patients: unable to participate, not invited by staff or declined to participate.

Patients unable to participate (n1)

The proportion judged unable to complete a written questionnaire in English was generally low, with a mean rate of around 5%. However, this figure is skewed by the high rates (up to 30%) recorded at some centres for patients undergoing cataract surgery. For other procedures, the mean proportion was 3.9%.

The most common ineligibility reasons noted on the patient participation sheets were problems with vision (often due to patients forgetting their glasses) and lack of manual dexterity (e.g. because of arthritis in the hands). An inability to understand written English was rarely noted. The most commonly mentioned languages were Polish, Punjabi and Urdu. There were very few instances of cognitive impairment. Other reasons mentioned included anxiety about the operation and anxiety about completion of a questionnaire.

Patients not invited to participate (n2)

The average proportion of patients not invited to participate was 25%. This varied from 0% to 70% depending on the centre and the operation. For the 45 centre/procedures for which information on failure to invite were available, 11 managed to invite over 90% of eligible patients. In contrast, in 12 instances the staff invited fewer than 50%.

There were several reasons for patients not being invited: ran out of questionnaires; staff forgetfulness, staff resistance to the study, delays in getting one or more surgeons' agreement to participate, lack of staff time and lack of time for patients (particularly if the patient was the first on a list).

The proportion was similar whether patients were recruited during a pre-op assessment clinic visit (26%) or on admission (25%). There was also no association with procedure. While high invitation rates (over 90%) were achieved in all types of facility, the 12 poorest rates (less than 50%) all occurred in NHSTCs or NHS Acute Hospitals

Table 3.2: Recruitment statistics by centre and procedure[**]

Centre	Surgery	Time of recruitment[+]	DNA/Canc (Nx)	Patients meeting criteria (N1)	Unable to participate* (n1)		Potentially eligible (N2)	Not invited (n2 and n2/N2%)	Invited to take part (N3)	Declined + (no consent) (n3)***	Patients participated (N4)	Proportion eligible participating (N4/N2)	Proportion invited participating (N4/N3)
					Reason known	Reason not known							
A	Hernia	b	0	66	1	0	65	23 (35.4)	42	16	26	40.0	61.9
	Vein	b	0	56	1	0	55	NA	NA	NA	45	81.8	NA
	Cataract	b	0	40	8	0	32	18 (56.3)	14	8	6	18.8	42.9
C	Hernia	b	0	NA	0	0	NA	NA	21	0 (1)	20	NA	95.2
	Vein	b	0	NA	1	0	NA	NA	19	1	18	NA	94.7
D	Hernia	b	2	75	3	0	72	NA	NA	NA	71	98.6	NA
	Vein	b	0	44	1	0	43	1 (2.3)	42	1	41	95.3	97.6
	Cataract	b	1	159	13	0	146	38 (26)	108	11	97	66.4	89.8
	Hip	b	0	45	1	0	44	19 (43.2)	25	3 (1)	21	47.7	84.0
	Knee	b	0	57	0	0	57	27 (47.4)	30	4	26	45.6	86.7
E	Hernia	b/a	0	34	1	0	33	0 (0)	33	6	27	81.8	81.8
	Vein	b/a	0	20	0	0	20	1 (5.0)	19	1	18	90.0	94.7
F	Hip	a	0	90	1	3	86	43 (50.0)	43	2 (4)	37	43.0	86.0
	Knee	a	0	109	0	6	103	39 (37.9)	64	3 (3)	58	56.3	90.6
G	Cataract	a	4	142	14	6	122	11 (9.0)	111	14 (2)	95	77.9	85.6
H	Cataract	b	0	100	35	0	65	16 (24.6)	49	8 (1)	40	61.5	81.6
I	Hernia	a	0	76	2	0	74	44 (59.5)	30	4	26	35.1	86.7
	Vein	a	0	37	1	0	36	24 (66.7)	12	4	8	22.2	66.7

Table 3.2 continued

Centre	Surgery	Time of recruitment[+]	DNA/Canc (Nx)	Patients meeting criteria (N1)	Unable to participate* (n1)		Potentially eligible (N2)	Not invited (n2 and n2/N2%)	Invited to take part (N3)	Declined + (no consent) (n3)***	Patients participated (N4)	Proportion eligible participating (N4/N2)	Proportion invited participating (N4/N3)
					Reason known	Reason not known							
J	Hernia	b	0	79	1	1	77	39 (50.6)	38	3 (1)	34	44.2	89.5
	Vein	b	0	52	0	0	52	29 (55.8)	23	3 (1)	19	36.5	82.6
	Hip	b	0	156	3	3	150	89 (59.3)	61	15 (1)	45	30.0	73.8
	Knee	b	0	87	3	0	84	20 (23.8)	64	18	46	54.8	71.9
K	Hernia	b	1	NA	0	0	NA	NA	40	19 (1)	20	NA	50.0
	Vein	b	0	NA	1	0	NA	NA	15	5	10	NA	66.7
L	Hernia	b	0	43	1	0	42	25 (59.5)[#]	17	0	17	40.5	100.0
	Vein	b	0	25	0	0	25	11 (44.0)	14	0	14	56.0	100.0
	Cataract	a	1	159	29	0	130	21 (16.2)	109	16	93	71.5	85.3
M	Cataract	a	1	94	14	1	79	0 (0)	79	5	74	93.7	93.7
N	Hernia	a	0	58	0	1	57	29 (50.8)	28	0	28	49.1	100.0
	Vein	a	0	37	1	0	36	5 (13.9)	31	1	30	83.3	96.8
	Hip	b	0	36	3	0	33	11 (33.3)	22	4	18	54.5	81.8
	Knee	b	0	32	3	0	29	16 (55.2)	13	3	10	34.5	76.9
O	Hernia	a	0	37	1	0	36	17 (47.2)	19	2	17	47.2	89.5
	Hip	b	3	59	5	4	50	5 (10.0)	45	14 (1)	30	60.0	66.7
	Knee	b	1	67	12	5	50	0 (0)	50	15	35	70.0	70.0
P	Hernia	b	0	90	4	0	86	23 (26.7)	63	5 (1)	57	66.3	90.5
Q	Vein	b	0	26	2	0	24	3 (12.5)	21	1	20	83.3	95.2
	Hip	b	0	40	2	0	38	18 (47.4)	20	5 (1)	14	36.8	70.0
	Knee	b	0	93	0	0	93	65 (70.0)	28	3	25	26.9	89.3
	Cataract	b	0	248	28	32	188	65 (34.6)	123	27	96	51.1	78.0

Table 3.2 continued

Centre	Surgery	Time of recruitment[+]	DNA/Canc (Nx)	Patients meeting criteria (N1)	Unable to participate* (n1)		Potentially Eligible (N2)	Not invited (n2 and n2/N2%)	Invited to take part (N3)	Declined + (no consent) (n3)***	Patients participated (N4)	Proportion eligible participating (N4/N2)	Proportion invited participating (N4/N3)
					Reason known	Reason not known							
R	Hernia	a	0	84	0	0	84	40 (47.6)	44	9 (1)	34	40.5	77.3
	Vein	a	0	42	NA	NA	NA	NA	20	0 (1)	19	NA	95.0
	Hip	b	0	83	3	1	79	NA	NA	17 (1)	61	77.2	NA
	Knee	b	0	90	2	0	88	8 (9.1)	80	15	65	73.9	81.3
	Cataract	b	0	154	23	3	128	10 (7.8)	118	24 (5)	89	69.5	75.4
S	Hernia	a	0	37	0	0	37	7 (18.9)	30	5	25	67.6	83.3
	Vein	a	0	43	0	0	43	14 (32.6)	29	4 (1)	24	55.8	82.8
T	Hip	b	3	68	5	4	59	3 (5)	56	12 (2)	42	71.2	75.0
	Knee	b	3	55	1	0	54	0 (0)	54	6 (1)	47	87.0	87.0
U	Hernia	b	2	113	2	0	111	34 (30.6)	77	8	69	62.2	89.6
	Vein	b	1	65	2	0	63	5 (7.9)	58	5	53	84.1	91.4
	Hip	a/b	1	NA	1	0	NA	NA	77	5 (3)	69	NA	89.6
	Knee	a/b	4	NA	1	0	NA	NA	64	2 (1)	61	NA	95.3
V	Cataract	a	5	155	19	7	129	17 (13.2)	112	23 (2)	87	67.4	77.7
W	Hernia	a	2	45	1	0	44	2 (4.5)	42	1	41	93.2	97.6
	Vein	a	1	46	NA	NA	NA	NA	NA	NA	41	NA	NA
	Hip	a	0	41	NA	NA	NA	NA	38	0 (1)	37	NA	97.4
	Knee	a	1	19	NA	NA	NA	NA	NA	NA	16	NA	NA
X	Hernia	a	1	39	1	0	38	0 (0)	38	0	38	100.0	100.0

NA indicates data were Not Available

*The number 'unable to participate' may be an under-estimate as some patients 'Not invited' may have proved to be unable to participate. Conversely, some centres may have minimised the number they failed to invite by recording them as unable to participate.

+indicates that patients were recruited on admission (a) or at the pre-op assessment clinic (b).

[#]Centre L applied a 3 week rule of a time lag between pre-op and surgery and for a month or so hernia patients were coming through without a date for surgery due to an initiative to get people off the waiting list which is why these patients were "missed".

[**]Inadequate information provided for hernias (N4 = 9) and veins (N4 = 1) at Centre B, for hernias (N4 = 11) and veins (N4 = 2) at Centre F, and for hips (N4 =23) and knees (N4 =11) at Centre K.

***The number in parentheses refers to patients for whom we did not receive a signed consent form,

Patients declining to participate (n3)

The mean proportion of invited patients who declined to take part was about 13%. In addition, about 1.5% completed a pre-operative questionnaire but did not sign a consent form. This masks some variation between centre/procedures: in 38 of the 53 centre/procedures for which we obtained data, over 80% agreed to participate whereas in 6 of the 53 the proportion was less than 70%. Agreement to participate was not related to procedure or type of facility. It was, however, associated with when patients were recruited: those centres that recruited at a pre-operative assessment clinic (34%) were twice as likely to have a low participation rate (less than 70% acceptance) than those centres that recruited on admission (15%).

Number of patients recruited (N4)

The recruitment target for each of the five surgical procedures was 450 patients. Because of a combination of unexpectedly high volume and the need to collect data at each centre for at least two weeks (to get a thorough understanding of local feasibility) this target was exceeded for groin hernia surgery (570 patients) and cataract surgery (677 patients). In contrast, fewer patients undergoing hip replacement (397), knee replacement (400) and varicose vein surgery (363) were recruited.

Recruitment rate (N4/N2)

Overall the proportion of eligible patients recruited was 60.3%. It varied by procedure: hip replacement 49.1%; knee replacement 55.4%; hernia repair 59.6%; cataract surgery 66.3%; and vein surgery 68.5%. Recruitment rate varied by centre/procedure from 18.8% to 100%: only 21.3% achieved a recruitment rate of over 80% and 36.2% of centre/procedures recruited less than 50% of those eligible. There was no association with the time of recruitment (pre-op assessment clinic versus on admission). In contrast, the type of facility was associated with poor recruitment: 7 of the 8 lowest proportions were in NHS Acute Hospitals.

The above recruitment rates are based on the number of eligible patients able to participate. An alternative denominator that could be used is the number eligible regardless of their ability to complete the questionnaire, which would result in a substantially lower recruitment rate for cataract patients.

3.6 Use of NHS Strategic Tracing Service

Of the 2350 patients recruited to the study, 11 (0.5%) died before their post-operative follow-up questionnaire. Eight of these deaths were discovered using the NSTS thus avoiding the dispatch of a questionnaire. But three deaths were only discovered when relatives of the deceased, who had opened the questionnaire mailed to the deceased's address, informed the research team.

3.7 Post-operative response rates

The overall response rate to the post-operative questionnaires was 85.7%. It varied between procedures: varicose vein surgery 75.3%; hernia repair 79.4%; cataract surgery 90.0%; knee replacement 90.2%; hip replacement 91.6%.

Of the 61 centre/procedures, 23 achieved over 90% response and 24 achieved 80 - 89% response. In only 11 instances (4 hernia, 6 veins and one knee) was the response less than 70%.

The effect of sending a reminder letter at five weeks after non-receipt of a returned questionnaire was to boost the response rate by about 10%.

3.8 Local data collectors' views of data collection

Semi-structured interviews with one or more staff were carried out at 23 of the 24 centres. The following summarises the main findings. A more detailed account appears in Appendix 8. Data collectors made several suggestions as to how recruitment of eligible patients could be maximised.

Whose responsibility?

The number of staff involved in recruiting patients varied from one to twenty between centres. It usually involved nursing staff but in some cases Health Care Assistants, student nurses or receptionists undertook the task. Several felt that it did not matter who undertook recruitment. However, many felt that it should be nursing staff because they were already interacting with the patients, building up trust and they would know how to answer questions. Reception staff were felt to be too busy and may not be able to answer questions about the study. There was also a confidentiality issue as reception staff are not always privy to the type of operation a patient was to have. One centre invested some time in educating the reception staff in regard to the names of various operations to be included so that they would correctly identify eligible patients. Also, sometimes reception staff are volunteers.

Most respondents agreed that the most important factor in recruitment was staff attitude and that it was important for staff to build up a rapport with the patients. It took a while to establish recruitment procedures and for everyone to know what they were doing. One problem from a continuity and communication perspective was that staff often worked on a rotation basis, particularly on the wards, and some people job shared.

Best time and location to recruit patients

Staff were divided as to whether the appropriate place to give out the questionnaires was at pre-op assessment or on admission. The disadvantages of using the pre-op assessment clinic were: already too much paperwork handed out at pre-operative assessment; unknown time interval until surgery, which varied between centres from a couple of weeks to a few months; patients might take the questionnaires home with them; if given the questionnaires at the beginning of the assessment, it could cause delays; if at the end, patients wanted to go home and didn't want to hold up whoever was transporting them; and patients might exaggerate their symptoms to ensure they got an operation.

The benefits of using the pre-op assessment clinic were: more time than on admission; patients less anxious than on the day of surgery; calmer atmosphere; it is usually a dedicated area within the centre, so easier to organise; might be completed while awaiting diagnostic investigations; if recruited on admission the operating

lists might be delayed or a patient might not have time; and patients might want to give something back to the nursing staff who had just spent time with them at pre-op assessment.

Identifying eligible patients

It was easier for staff to remember which operations to include if the clinic largely dealt with the specific operations covered in the study. Staff had few problems identifying patients who could not understand English or who suffered from cognitive impairment. Occasionally a patient would tell the nurse privately that he/she was unable to read and/or write. It was often written on a patient's pre-admission sheet whether or not they understood English. Some staff used "body language" to determine whether a patient was too anxious to be eligible. Many cataract patients were not offered a questionnaire because their vision was too poor. Most patients with severe visual problems said themselves that their vision was too poor for them to take part. Some centres prepared a list of eligible patients the night before from the theatre/assessment list for the following day and put a questionnaire in the patients' notes.

Failure to invite

Some staff failed to invite patients to participate due to forgetfulness, time pressures, or a key person being on leave. Several staff claimed not to know which patients to include, which explained why patients having hip and knee revisions were not always included at some centres.

Reasons that patients chose not to participate

Most patients wanted to help out with the study. Often patients declined to participate because of a lack of time. However, some patients were adversely influenced by other patients: one patient declining could lead to a chain of refusals. Sometimes patients just left the questionnaire blank and put them to one side rather than specifically state that they did not want to take part. Sometimes staff felt that a patient would decline rather than admit to having a literacy problem.

Workload

Staff felt that it was the initial setting up of the study locally that was time-consuming. Recruiting patients did not significantly impact on staff time but there were some concerns about workload if it became routine. Most centres said it took about five (could be up to ten) minutes per patient.

Questionnaire design

Generally staff felt that the questionnaires were user friendly but that they "looked like a book" although they were not as long as initially thought. They reported that a few patients commented on the length or thickness of the booklet, that a number of the questions are similar, and those requiring varicose vein surgery had commented that the drawing of the veins had put them off.

The length of time patients spent completing the questionnaires varied from five to thirty minutes. A small number of patients failed to complete the questionnaires due to time pressures. Some completed them straight

away, others would leave them to complete later. Those who took longer were older, the less educated, patients with visual or cognitive recall problems, or those with arthritis. Patients for joint replacement surgery tend to have a lot of other paperwork, e.g. NJR forms. Patients often asked staff for help completing the questionnaires and would discuss the questions with their relatives and with other patients.

A small number of patients were unsure how to answer particular questions. These included the co-morbidity question (perhaps because they did not understand the names of certain conditions or were uncertain if they had a certain condition). There was also a question relating to driving in the cataract questionnaire which sometimes caused confusion.

In regard to the cataract surgery questionnaire, it was suggested that it might be better to use black print on white or yellow paper although it was agreed that the font size (14) was appropriate. It was also agreed that the questionnaires were big enough for the patients to hold comfortably.

Value of pre-study visit and regular contact

All staff who had been present at the pre-study presentation said that they had found it useful, not only in term of understanding the aims but also to "put a face to a name" and to feel they could get in touch with the lead researcher with any queries. Adherence to the study protocol by a centre was better when the lead researcher met staff at the initial meeting. Most people appreciated the regular contact and many said they would not have felt comfortable about participation if the initial visit had not taken place. In some centres the presentation was repeated by local staff to explain the methodology to colleagues who had not been present at the meeting.

Study materials

The study handbook was felt to be moderately useful by those people who had seen it but it was not read by all staff. Some centres would have appreciated a supply of pens as they often had to stop and find pens for patients to use which they found irritating. Some felt that a script should be provided to help staff recruit patients.

3.9 Patients' views of recruitment and follow-up

86 patients were approached to participate in a telephone interview: 28 patients following hernia surgery, 15 patients following cataract surgery, 21 patients following varicose vein surgery, 11 patients following hip replacement/revision surgery and 11 patients following knee replacement/revision surgery.

Interviews were carried out with 36 patients (42% of those approached): 10 hernia surgery (all male), 2 cataract surgery (both male), 12 varicose vein surgery (6 male), 6 hip replacement surgery (2 male) and 6 knee replacement surgery (1 male). There was a poor response from cataract surgery patients, probably due to these patients being older. The patients had been treated in a variety of facilities and some had been recruited at pre-op assessment clinics and others on admission.

Time at which patients recruited

Most patients were happy with the time at which they had been recruited, whether at pre-op assessment clinic or on admission. The majority said that it would not have mattered when they had been recruited. Most patients who had been recruited on admission felt it had given them something to do. A few patients who had been recruited at the pre-op assessment clinic said they might have felt too nervous to complete the questionnaire on admission and another couple said that there might be time pressures on the day.

Taking the pre-operative questionnaires home

Patients were also asked whether or not they should be able to take the pre-operative questionnaires home. Almost all felt that the questionnaires would be "lost" and never returned. Two patients mentioned the fact that they had children and would not have time to complete the form at home. Others expressed a view that if a patient was too nervous at the time or required help completing the questionnaire, then the patient should be allowed to take it home to complete.

Questionnaire design

Most of the patients felt that the questionnaires were fine and generally thought that the study was a good idea. Indeed, some believed that collecting PROMs data was "a necessity". Most felt that other patients would be happy to complete the questionnaires, whilst recognising that there will always be some people who will refuse. They did not take long to complete - 22 of the 37 patients taking ten minutes or less. A further seven patients said they spent up to 20 minutes, two spent 30 minutes and one spent an hour on it. Four patients did not specify a completion time saying that time "was not a problem" or stated that the time for completion was "fair enough".

The majority of patients said the questionnaires were easy to understand and very unlikely to cause any upset ("unless someone is particularly sensitive"). The only criticisms were: some questions were "too generalised"; the applicability of questions on driving in the VF-14 to non-drivers was unclear; repetition in the hip replacement questionnaire which arose from including both the SF-36 and the EQ-5D; put off by what they referred to as the "psychological context", questions relating to mental state.

Some patients would have liked the opportunity to write some free text to mention factors such as additional complications and other factors they felt may have hindered their recovery, and about the hospital care they had received, whether this was to criticise or praise. (Questionnaires were often returned with notes on them). A couple of patients felt the questions were not appropriate if the operation had gone wrong, they were "only useful if it is all plain sailing".

How could the post-operative response rate be increased?

Most patients felt that there was little more that could be done to increase the response rate. The only suggestions were: the use of an incentive; warn patients that the questionnaires were coming; have nursing staff stress the importance of completing the questionnaires; and make completion compulsory somehow.

3.10 Patients' use of freephone line

Fifty-eight patients called the freephone line between 23 May 2006 and 31 March 2007, the details of which appear in Appendix 9. The principal concerns were:

- not yet 3/6 months post-op
- waiting for a follow-up appointment or further treatment
- had or about to have a second operation
- impact of other medical conditions
- operation had not taken place
- queries about questions
- operative complications
- patient treated elsewhere
- patient too ill to respond
- wanted to include additional feedback comments
- difficulties remembering date of operation
- wanting to comment adversely about the Consultant
- claimed not to have agreed to participate

3.11 Cost-effectiveness of data collection

The model for the costs of data collection is presented in Table 3.3. All of the costs included vary with respect to the number of patients invited to participate (N). The proportion of these patients completing a pre-operative questionnaire is represented by a and the proportion of those participating who subsequently complete a post-operative questionnaire is represented by ß. Thus, the 'completion rate' is a ß (that is, the proportion of those approached from whom a pre-op and post-op questionnaire are obtained).

The cost of data analysis has not been included because it is not currently known:

- how frequently analyses would take place
- what analyses would be undertaken
- whether the same pre-specified analyses would be repeated on each occasion or whether it might vary.

If a standard analysis is repeated across centres and over time the overall addition to the cost per patient is likely to be small.

The assumed print runs of 5000 for pre-printed envelopes and questionnaires secures quite substantial quantity discounts. If purchasing were to be co-ordinated nationally rather than organised locally there may be opportunities to further reduce costs. Note, however, that envelopes and questionnaires make up only about 10% of the overall costs.

Table 3.3: Cost model and assumptions

Activity	Cost (£)	Comment
Recruitment of patients	0.748221 N	5 minutes per patient, assuming Band 1 staff
Questionnaires	0.81 a N	36 pence per questionnaire, assuming average of 12 pages and purchasing batches of 5000
Envelopes	0.094875 a N	7.59 pence per item if buying in batches of 5000
Postage	(0.4625 + 0.37 ß) a N	37 pence per item, 25% receive reminder
NSTS check	0.215713 a N	1 minute per patient, assuming Band 4 staff
Data entry	(1.292942 + 1.292942 ß) a N	7 minutes per patient, assuming Band 3 staff
Total	0.748221 N + 2.87603 a N + 1.662942 a ß N	

The total cost is calculated by adding together all the component costs (last row). The mean cost per pair of completed questionnaires is thus the total cost divided by the number of patients completing both a pre and post-op questionnaire, a ß N:

0.748221/ a ß + 2.87603/ ß + 1.662942

The average cost per patient based on the observed values of a and ß are presented in Table 3.4 for each surgical treatment.

Table 3.4: Average cost per matching pre-operative and follow-up questionnaire by surgical procedure and time of recruitment

Surgical treatment	Recruited at pre-operative assessment clinic	Recruited on admission	Overall (irrespective of time of recruitment)
Hernia repair	£6.44	£6.33	£6.39
Varicose vein surgery	£6.57	£6.61	£6.58
Cataract surgery	£5.90	£5.84	£5.87
Hip replacement	£5.90	£5.70	£5.81
Knee replacement	£5.89	£5.77	£5.85

In order to provide a feel for the precision of these estimates the standard error of the completion rate was estimated and this was used to provide lower and upper estimates for the cost per pair of completed questionnaires (reported in Table 3.5).

Table 3.5: 95% confidence interval for cost per pair of completed questionnaires by surgical procedure

Surgical treatment	Lower estimate	Upper estimate
Hernia repair	£6.17	£6.63
Varicose vein surgery	£6.26	£6.95
Cataract surgery	£5.75	£5.99
Hip replacement	£5.65	£5.98
Knee replacement	£5.69	£6.01

The relationship between response rates and the cost per patient is explored in Table 3.6. Clearly there are many different combinations of a and ß which can give rise to any given completion rate.

Table 3.6: Predicted mean cost per questionnaire for a range of completion rates

a	ß	Completion rate (a ß)	Mean cost
0.85	0.4706	0.40	£9.64
0.85	0.5882	0.50	£8.05
0.85	0.7059	0.60	£6.98
0.85	0.8235	0.70	£6.22
0.95	0.8421	0.80	£6.01
0.95	0.9474	0.90	£5.53

To provide a comparative context for the estimated costs described above we contacted a number of organisations involved in the collection of routine health care information on patients. Two organisations involved exclusively in the collection of clinical data provided estimates of the total cost per complete patient record. The Intensive Care National Audit and Research Centre (ICNARC) estimates the cost to be between £12.97 and £55.16 depending on the centre involved (this is based upon a sample of 9% of the units contributing data to ICNARC). The UK Renal Registry estimates the average cost to be £16 per patient.

We also obtained estimates of the costs per patient successfully recruited by three organisations involved in the collection of patient-reported data. CHKS Ltd, which is involved in the collection of PROMs at seven acute NHS hospitals, estimates the cost to be £11.07 for elective patients (this involves both pre- and post-treatment questionnaires) and £3.15 for emergency admissions (this involves post-treatment postal surveys only). The estimated cost for approved survey contractors (http://www.nhssurveys.org/default.asp) involved in NHS cross-sectional surveys ranges is £3 to £3.50. Finally, BUPA Ltd estimates their cost as about £3 per completed pre- and post-operative questionnaires. The costs for BUPA and the NHS survey contractors are somewhat lower than the estimates for the methods in our study because they rely exclusively on mailed surveys and do not involve local staff in recruiting the patients. The latter represents about 20% of the costs of the method we used.

4. RESULTS: Comparison of performance between centres

4.1 Cataract surgery

4.1.1 Pre-operative characteristics of patients

Overall, 677 patients were recruited in nine centres (though inter-centre comparisons exclude Centre A where only six patients were recruited). Table 4.1 shows that the age and sex mix varied between centres. The mean age ranged from 70 to 77 years and the percentage of women from 46% to 65%. This is similar to all patients treated by the NHS in 2005-6: 75 years and 60.8% female.

There was consistency between the centres in terms of the median length of time that patients had experienced symptoms. However, the percentage of patients who had previously undergone surgery for cataract varied from 21.9% to 49%. (Note: Centre H chose to recruit only 'first eye' patients to the study).

Table 4.1: Patient demographics and cataract history

Centre	n	Age (yrs), mean (SD)	Female n (%)	Duration of symptoms (yrs), median (IQR)	Previous cataract surgery n (%)
D	97	75.9 (8.4)	45 (46.4)	2 (1 - 3.5)	40 (41.2)
G	95	70.5 (10.8)	57 (60.6)	2 (1 - 4)	46 (48.9)
H	40	74.1 (10.2)	24 (60.0)	2 (1 - 4)	1 (2.6)
L	93	73.5 (11.2)	59 (63.4)	2 (1 - 3.8)	37 (40.2)
M	74	76.1 (9.9)	42 (56.8)	2 (2 - 3.5)	31 (42.5)
Q	96	73.8 (9.3)	53 (55.2)	2 (1 - 5)	21 (21.9)
R	89	73.7 (10.2)	58 (65.2)	2 (1 - 4)	25 (28.1)
V	87	76.9 (8.1)	55 (63.2)	2 (1 - 3)	42 (48.3)
Total	671	74.3 (9.9)	393 (58.7)	2 (1 - 4)	243 (36.4)

Note: age and sex was not known for 1 patient; duration of symptoms not known for 82 patients

The number of systemic comorbidities per patient did not vary greatly between centres (Table 4.2). When restricted to those comorbidities which might affect surgical outcome, the median number of comorbidities was 1 for each centre.

Table 4.2: Patient comorbidities

Centre	No. of comorbidities per patient, median (IQR)	No. of comorbidities which could affect surgical outcome* per patient, median (IQR)
D	2 (1 - 3)	1 (1 - 2)
G	1 (0 - 3)	1 (0 - 2)
H	2 (1 - 3)	1 (0 - 2)
L	1 (1 - 2)	1 (0 - 1)
M	2 (1 - 3)	1 (0 - 2)
Q	2 (1 - 3)	1 (0 - 2)
R	2 (1 - 3)	1 (0 - 2)
V	1 (0 - 3)	1 (0 - 1)
Total	2 (1 - 3)	1 (0 - 2)

* Angina, asthma, bronchitis, diabetes, high BP, kidney disease, liver disease, poor circulation

Before surgery, almost a third of patients rated their general health status as 'Excellent' or 'Very good' (Table 4.3). The percentage varied between centres from 22% to 47% (while those only 'fair' or poor' varied from 14% - 34%).

Table 4.3: Pre-operative general health status

Centre	In general, would you say your health is:				
	Excellent, n (%)	Very good, n (%)	Good, n (%)	Fair, n (%)	Poor, n (%)
D	3 (3.1)	22 (22.7)	39 (40.2)	29 (29.9)	4 (4.1)
G	3 (3.2)	25 (26.6)	49 (52.1)	16 (17.0)	1 (1.1)
H	3 (7.5)	13 (32.5)	18 (45.0)	6 (15.0)	0 (0)
L	9 (9.8)	34 (37.0)	36 (39.1)	12 (13.0)	1 (1.1)
M	2 (2.7)	14 (18.9)	39 (52.7)	17 (23.0)	2 (2.7)
Q	3 (3.2)	27 (28.4)	48 (50.5)	15 (15.8)	2 (2.1)
R	8 (9.0)	25 (28.1)	32 (36.0)	21 (23.6)	3 (3.4)
V	6 (7.0)	20 (23.3)	46 (53.5)	12 (14.0)	2 (2.3)
Total	37 (5.6)	180 (27.0)	307 (46.0)	128 (19.2)	15 (2.3)

Note: Pre-operative general health status not known for 4 patients

Table 4.4 shows the distribution of pre-operative VF-14 and EQ-5D scores by centre. The mean VF-14 scores varied between centres from 77.0 to 87.8. There was little variation in the EQ-5D scores, although Centre M had a noticeably lower score than the other centres.

Table 4.4: Pre-operative VF-14 and EQ-5D scores

Centre	Pre-operative VF-14 mean (SD)	Pre-operative EQ-5D mean (SD)
D	77.0 (19.2)	0.77 (0.24)
G	80.7 (19.8)	0.79 (0.24)
H	78.4 (16.9)	0.84 (0.25)
L	86.2 (12.3)	0.82 (0.22)
M	80.4 (21.5)	0.73 (0.26)
Q	83.5 (15.8)	0.83 (0.22)
R	84.0 (17.5)	0.80 (0.24)
V	87.8 (14.5)	0.82 (0.25)
Total	82.5 (17.6)	0.80 (0.24)

Note: Pre-operative VF-14 and EQ-5D was not known for 5 and 16 patients, respectively

4.1.2 Post-operative health

Of the 671 patients recruited, 648 were eligible for follow-up (18 did not undergo surgery, 2 had died and the whereabouts of three was unknown). Of these, 583 (90.0%) patients returned a questionnaire three months after surgery. Just over a third of patients rated their general health status as 'Excellent' or 'Very good' (Table 4.5), a similar proportion to that before surgery. However, slightly more patients rated their general health status as only 'Fair' or 'Poor' after surgery (27% vs. 21.5%).

Table 4.5: Post-operative general health status

Centre	In general, would you say your health is:				
	Excellent, n (%)	Very good, n (%)	Good, n (%)	Fair, n (%)	Poor, n (%)
D	2 (2.7)	15 (20.0)	33 (44.0)	25 (33.3)	0 (0)
G	10 (13.3)	21 (28.0)	24 (32.0)	17 (22.7)	3 (4.0)
H	1 (2.9)	12 (34.3)	15 (42.9)	7 (20.0)	0 (0)
L	4 (4.4)	28 (30.8)	28 (30.8)	26 (28.6)	5 (5.5)
M	0 (0)	18 (29.0)	24 (38.1)	17 (27.0)	4 (6.4)
Q	3 (3.8)	19 (23.8)	39 (48.8)	13 (16.3)	6 (7.5)
R	7 (9.0)	26 (33.3)	28 (35.9)	11 (14.1)	6 (7.7)
V	5 (6.0)	26 (31.3)	34 (41.0)	13 (15.7)	5 (6.0)
Total	32 (5.5)	165 (28.5)	225 (38.8)	129 (22.2)	29 (5.0)

Note: Post-operative general health status not known for 3 patients

Table 4.6 shows the distribution of pre- and post-operative VF-14 scores by centre. Overall, the mean change in score was 9.9 units (83.0 - 92.9), indicating an improvement in vision. The mean improvement varied between centres (7.4 to 15.7).

Table 4.6: Pre- and post-operative VF-14 scores

Centre	n	Pre-operative VF-14 mean (SD)	Post-operative VF-14 mean (SD)
D	75	76.6 (19.6)	90.7 (17.9)
G	75	79.6 (19.4)	94.0 (13.0)
H	35	78.9 (16.5)	94.6 (8.2)
L	90	86.0 (12.4)	93.7 (12.8)
M	62	80.7 (22.3)	89.0 (21.3)
Q	78	85.3 (14.8)	93.1 (13.8)
R	79	84.9 (16.8)	92.3 (13.9)
V	83	88.1 (17.9)	95.8 (8.1)
Total	577	83.0 (17.4)	92.9 (14.3)

Note: Post-operative VF-14 score not known for 6 patients

The risk adjustment model contained only two variables: pre-operative VF-14 score and general health status. These two variables explained 22% of the variation between patients in post-operative VF-14 score.

In the funnel plot shown in Figure 4.1, the hollow circles represent each centre's unadjusted post-operative VF-14 score and the solid circles represent each centre's adjusted score. The horizontal line is the mean post-operative VF-14 score for all centres combined (92.9). The solid curved lines represent 99.8% control limits and the broken curved lines represent 95% control limits. The plot indicates that there is no statistical difference between each centre's adjusted post-operative VF-14 score and target performance at the $p<0.05$ significance level.

Figure 4.1: Funnel plot of post-operative VF-14 scores in cataract surgery

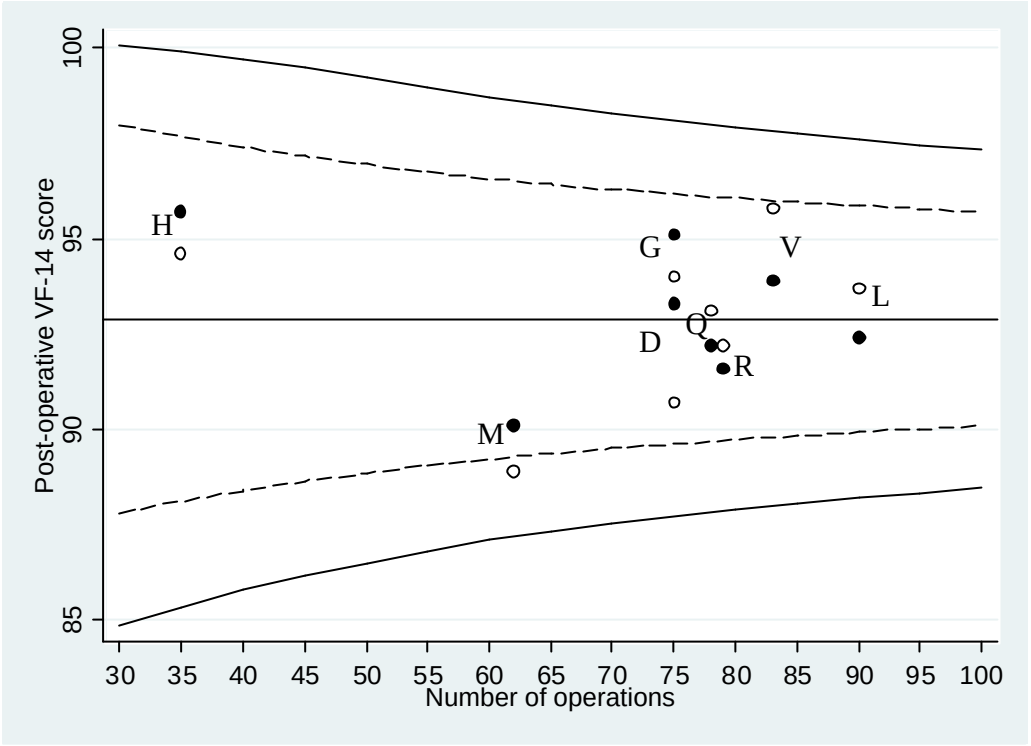


Table 4.7 shows the distribution of pre- and post-operative EQ-5D scores by centre. Overall, the mean change in score was 0.03 units (0.81 – 0.78) indicating little change in health state.

Table 4.7: Pre- and post-operative EQ-5D scores

Centre	n	Pre-operative EQ-5D mean (SD)	Post-operative EQ-5D mean (SD)
D	74	0.81 (0.22)	0.76 (0.28)
G	72	0.80 (0.22)	0.79 (0.32)
H	35	0.83 (0.27)	0.80 (0.25)
L	86	0.82 (0.21)	0.78 (0.25)
M	61	0.72 (0.27)	0.69 (0.30)
Q	81	0.84 (0.21)	0.81 (0.27)
R	76	0.81 (0.23)	0.76 (0.29)
V	81	0.83 (0.25)	0.83 (0.25)
Total	566	0.81 (0.23)	0.78 (0.28)

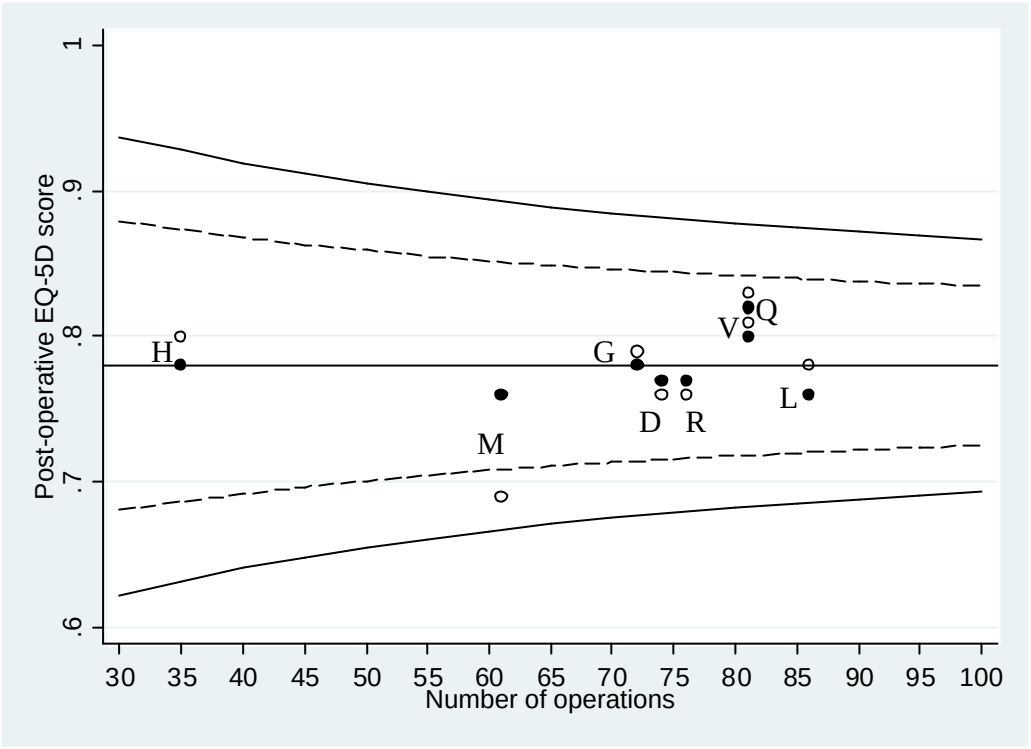
Note: Limited to patients with pre- and post-operative EQ-5D scores

The risk adjustment model contained age, pre-operative EQ-5D score, previous cataract surgery, general health status and comorbidity. These five variables explained 45% of the variation between patients in post-operative EQ-5D score.

In the funnel plot shown in Figure 4.2 the hollow circles represent each centre's unadjusted post-operative EQ-5D score and the solid circles represent each centre's adjusted post-operative EQ-5D score. The horizontal line is the mean post-operative VF-14 score for all centres combined (0.78). The plot indicates that there is no

statistical difference between each centre's adjusted post-operative EQ-5D score and target performance at the $p < 0.05$ significance level.

Figure 4.2: Funnel plot of post-operative EQ-5D scores in cataract surgery



Overall, 9.8% of patients reported one or more generic post-operative complications. The proportion varied between centres from 1.2% to 16.5% (Table 4.8).

Table 4.8: Post-operation complications following cataract surgery

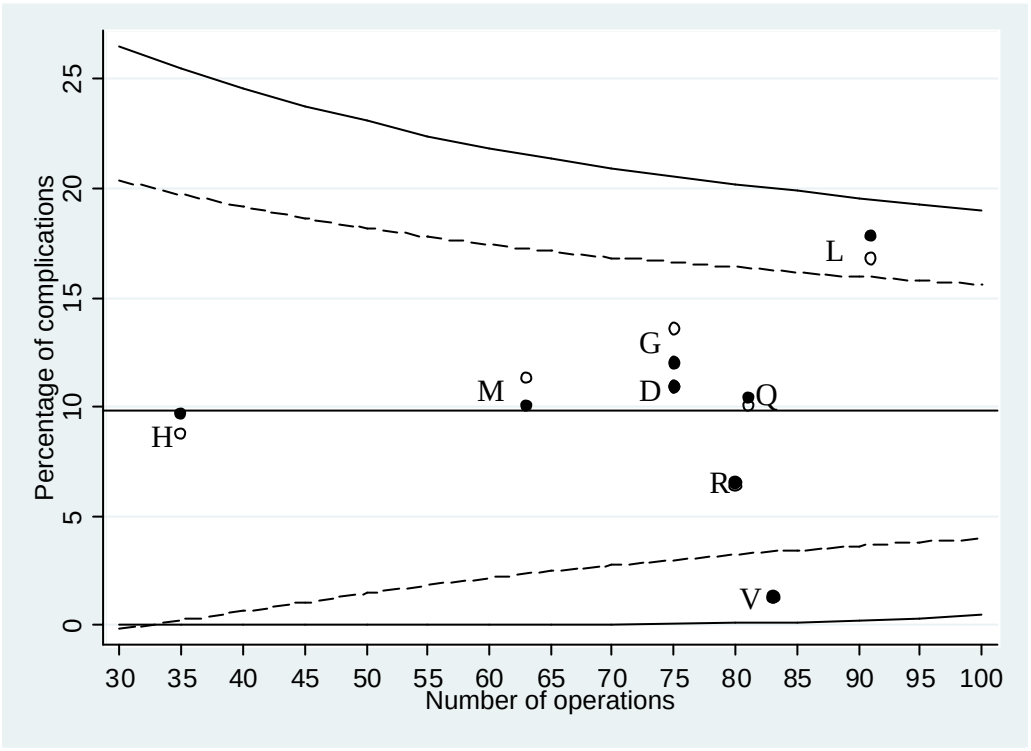
Centre	Allergy or reaction to drug n (%)	Urinary problems n (%)	Bleeding n (%)	Wound problems n (%)	Any complication n (%)
D	2 (2.7)	0 (0)	2 (2.7)	5 (6.7)	8 (10.7)
G	2 (2.7)	3 (4.0)	4 (5.3)	3 (4.0)	10 (13.3)
H	2 (5.7)	2 (5.7)	2 (5.7)	1 (2.9)	3 (8.6)
L	5 (5.5)	3 (3.3)	3 (3.3)	4 (4.4)	15 (16.5)
M	4 (6.4)	3 (4.8)	0 (0)	1 (1.6)	7 (11.1)
Q	1 (1.2)	0 (0)	5 (6.2)	2 (2.5)	8 (9.9)
R	2 (2.5)	1 (1.3)	1 (1.3)	1 (1.3)	5 (6.3)
V	0 (0)	0 (0)	1 (1.2)	0 (0)	1 (1.2)
Total	18 (3.1)	12 (2.1)	18 (3.1)	17 (2.9)	57 (9.8)

The risk adjustment model contained only pre-operative EQ-5D score. The model discriminated between those patients who did and who did not have complications only moderately (area under ROC = 0.61).

In the funnel plot shown in Figure 4.3 the hollow circles represent each centre's unadjusted percentage of complications and the solid circles represent each centre's adjusted percentage of complications. The horizontal

line represents target performance – the overall percentage of complications for all centres combined (9.8%). The plot indicates that the percentage of complications at Centre L is statistically higher than expected while the percentage of complications at Centre V is statistically lower than expected at the $p < 0.05$ significance level.

Figure 4.3: Funnel plot of post-operative complications in cataract surgery



The majority of patients reported that their vision was better following surgery (93.1%) with little variation between centres (Table 4.9).

Table 4.9: Estimation of change in visual functioning following cataract surgery

Centre	Overall, how are the problems in the eye(s) that had surgery now compared to before your operation?				
	Much better n (%)	A little better n (%)	About the same n (%)	A little worse n (%)	Much worse n (%)
D	61 (82.4)	8 (10.8)	1 (1.4)	2 (2.7)	2 (2.7)
G	59 (78.7)	13 (17.3)	1 (1.3)	1 (1.3)	1 (1.3)
H	28 (80.0)	5 (14.3)	2 (5.7)	0 (0)	0 (0)
L	72 (79.1)	11 (12.1)	3 (3.3)	2 (2.2)	3 (3.3)
M	47 (74.6)	10 (15.9)	2 (3.2)	2 (3.2)	2 (3.2)
Q	67 (82.7)	5 (6.2)	4 (4.9)	1 (1.2)	4 (4.9)
R	65 (82.3)	8 (10.1)	5 (6.3)	1 (1.3)	0 (0)
V	74 (89.2)	8 (9.6)	1 (1.2)	0 (0)	0 (0)
Total	473 (81.4)	68 (11.7)	19 (3.3)	9 (1.6)	12 (2.1)

4.1.3 Summary and discussion

- Mean age of patients was similar to that for the NHS overall and to that reported in studies from several countries (UK, USA, Denmark, Canada, Spain, Sweden) over the past 15 years and did not vary greatly between centres.
- The proportion of patients undergoing second eye surgery (36.4%) varied considerably between centres: for five centres it was 40-50%, for two centres it was 20-30% but for Centre H it was only 2.6% because they chose to recruit only 'first eye' patients to the study.
- The general health status of patients varied considerably between centres: the proportion reporting excellent or very good health was 22-47% while the proportion fair or poor was 14-34%.
- Visual function (mean VF-14 score 83.1) was high before surgery but at a similar level to that reported for British private patients in 2001 (83.2) and in 2005/6 (84.4) (BUPA - personal communication). These scores are considerably higher than the value of 68.1 reported for NHS patients in 1994/5 (Desai 1996). The proportion of patients with a score of 90 or more rose from 18.1% in 1994/95 to 43.1% in this study. A similar though less dramatic change has been reported in Canada: 1992/93 mean VF-14 was 71.0 but in 1999/2000 it was 79.0 (Norregaard et al. 1998; Wright et al. 2002). Several factors may have contributed to this secular change:
  - a higher proportion of operations are now conducted on 'second eyes' (partly to restore binocular vision) which inevitably will be less severely affected than the 'first eye' (though second eye surgery was reported as at least as beneficial as first eye surgery by Lundstrom et al. in 2001)
  - operations are being conducted earlier before cataracts become more severe ('anticipatory surgery')
  - the threshold for operating has fallen as the availability of surgery has increased
- Risk adjustment explained only 22% of the variation in change in VF-14 scores. Despite this, none of the outcomes in the eight centres were statistically significantly different from what would be expected (at the $p < 0.05$ significance level). It should be noted that the adjustment may not have taken one potential confounder, that of ocular comorbidity, fully into account as accurate information on this cannot be obtained directly from patients. Data were restricted to generic complications that commonly occur following most surgery. Data on specific ocular complications were not included as it was not felt possible to obtain these from patients.
- The EQ-5D was largely unresponsive, a finding that is consistent with other studies that have considered changes in generic measures (EQ-5D, SF36) following interventions to improve vision (and hearing). This may reflect the lack of any items relating to sensory functions in those instruments.
- The incidence of complications (as perceived by patients) varied considerably between centres from 1.2% to 16.5%. The latter centre requires more detailed investigation to see if this is a cause of concern. These incidences are much higher than those reported by clinicians. For example, Patwardham et al (2006) reported only 0.36% of patients experienced endophthalmitis following surgery.

4.2 Hernia surgery

4.2.1 Pre-operative characteristics of patients

Overall, 561 patients were recruited from 17 centres (excluding Centre B where only nine patients were recruited). Table 4.10 shows that the mean age ranged from 46 to 64 years and the percentage of women from 0% to 18%. This is similar to all patients treated by the NHS in 2005-6: 52 years and 8.1% female. There was consistency between the centres in terms of the median length of time that patients had experienced symptoms. However, the percentage of patients who had previously undergone hernia surgery varied considerably, from 4% to 55%.

Table 4.10: Patient demographics and hernia history

Centre	n	Age (yrs), mean (SD)	Female, n (%)	Duration of symptoms (yrs), median (IQR)	Previous hernia surgery, n (%)
A	26	48.8 (11.3)	5 (19.2)	2 (1 – 5)	4 (16.0)
C	20	64.1 (14.6)	0 (0)	1 (0.9 – 1.5)	4 (20.0)
D	71	56.9 (16.0)	11 (15.5)	1 (1 – 4)	12 (17.4)
E	27	53.1 (14.9)	2 (7.4)	1 (0.8 – 2)	4 (14.8)
F	11	61.7 (21.4)	2 (18.2)	1 (1 – 9)	6 (54.6)
I	26	52.7 (16.8)	2 (7.7)	1 (1 – 2)	1 (3.9)
J	34	54.2 (18.5)	1 (2.9)	1 (0.3 – 2)	8 (24.2)
K	20	58.6 (12.7)	2 (10.0)	1.3 (0.8 – 3.8)	5 (25.0)
L	17	55.3 (13.5)	2 (11.8)	1.3 (1 – 2)	1 (5.9)
N	28	46.0 (13.8)	2 (7.1)	1 (1 – 3)	6 (21.4)
O	17	49.2 (17.7)	1 (5.9)	1 (0.7 – 3)	1 (5.9)
P	57	61.1 (15.8)	7 (12.3)	1 (1 – 2)	17 (29.8)
R	34	50.7 (16.0)	6 (17.7)	1.3 (1 – 4)	8 (25.0)
S	25	45.9 (16.7)	1 (4.0)	1 (0.5 – 2)	5 (20.8)
U	69	48.7 (14.1)	3 (4.4)	1 (0.6 – 2)	9 (13.4)
W	41	56.7 (15.3)	3 (7.3)	1 (1 – 3)	17 (41.5)
X	38	58.8 (13.0)	3 (7.9)	2 (1 – 4)	3 (8.1)
Total	561	54.3 (15.9)	56 (9.8)	1 (1 – 2)	113 (20.2)

Note: age not known for 9 patients; duration of symptoms not known for 30 patients; previous hernia surgery not known for 10 patients

Patients undergoing hernia surgery had little comorbidity (Table 4.11). When restricted to those comorbidities which might affect surgical outcome, the median number was zero for all centres.

Table 4.11: Patient comorbidities

Centre	No. of comorbidities per patient, median (IQR)	No. of comorbidities which affect surgical outcome* per patient, median (IQR)
A	1 (0 – 2)	0 (0 – 1)
C	1 (0 – 2)	0 (0 – 1)
D	1 (0 – 2)	0 (0 – 1)
E	0 (0 – 1)	0 (0 – 1)
F	1 (0 – 2)	0 (0 – 1)
I	0 (0 – 1)	0 (0 – 0)
J	0 (0 – 1)	0 (0 – 0)
K	1 (0 – 1)	0 (0 – 0.5)
L	1 (0 – 1)	0 (0 – 1)
N	0 (0 – 1)	0 (0 – 0.5)
O	0 (0 – 1)	0 (0 – 0)
P	1 (0 – 2)	0 (0 – 1)
R	0 (0 – 1)	0 (0 – 1)
S	0 (0 – 0)	0 (0 – 0)
U	0 (0 – 1)	0 (0 – 0)
W	1 (0 – 1)	0 (0 – 1)
X	1.5 (0 – 2)	0 (0 – 1)
Total	0 (0 – 1)	0 (0 – 1)

* Angina, asthma, bronchitis, diabetes, high BP, kidney disease, liver disease, poor circulation

Before surgery, 56.2% of patients rated their general health status as 'Excellent' or 'Very good' (Table 4.12) (inter-centre range 35% to 77%) and only 8.1% reported 'Fair' or 'Poor' health (range 0-15.4%).

Table 4.12: Pre-operative general health status

Centre	In general, would you say your health is:				
	Excellent, n (%)	Very good, n (%)	Good, n (%)	Fair, n (%)	Poor, n (%)
A	2 (7.7)	7 (26.9)	13 (50.0)	3 (11.5)	1 (3.9)
C	1 (5.0)	12 (60.0)	6 (30.0)	1 (5.0)	0 (0)
D	5 (7.0)	35 (49.3)	25 (35.2)	5 (7.0)	1 (1.4)
E	3 (11.1)	13 (48.2)	11 (40.7)	0 (0)	0 (0)
F	0 (0)	7 (63.6)	3 (27.3)	1 (9.1)	0 (0)
I	5 (19.2)	15 (57.7)	5 (19.2)	1 (3.9)	0 (0)
J	6 (17.7)	15 (44.1)	12 (35.3)	1 (2.9)	0 (0)
K	4 (20.0)	11 (55.0)	3 (15.0)	2 (10.0)	0 (0)
L	0 (0)	10 (58.8)	7 (41.2)	0 (0)	0 (0)
N	1 (3.6)	10 (35.7)	14 (50.0)	3 (10.7)	0 (0)
O	3 (17.7)	6 (35.3)	8 (47.1)	0 (0)	0 (0)
P	8 (14.0)	23 (40.4)	18 (31.6)	7 (12.3)	1 (1.8)
R	6 (17.7)	15 (44.1)	11 (32.4)	2 (5.9)	0 (0)
S	6 (24.0)	11 (44.0)	7 (28.0)	1 (4.0)	0 (0)
U	7 (10.1)	34 (49.3)	24 (34.8)	3 (4.4)	1 (1.5)
W	4 (9.8)	17 (41.5)	17 (41.5)	3 (7.3)	0 (0)
X	2 (5.3)	13 (34.2)	18 (47.4)	4 (10.5)	1 (2.6)
Total	63 (11.1)	257 (45.1)	204 (35.8)	41 (7.2)	5 (0.9)

Table 4.13 shows the pre-operative SF-36 Physical Component Summary (PCS), SF-36 Mental Component Summary (MCS) and EQ-5D scores by centre. Mean PCS scores varied between centres from 41.5 to 50.4, the mean MCS scores from 48.3 to 55.1, and mean EQ-5D scores from 0.75 to 0.84.

Table 4.13: Pre-operative SF-36 and EQ-5D scores

Centre	Pre-operative SF-36 PCS mean (SD)	Pre-operative SF-36 MCS mean (SD)	Pre-operative EQ-5D mean (SD)
A	46.1 (8.4)	48.5 (10.7)	0.75 (0.12)
C	44.4 (8.8)	51.6 (10.8)	0.78 (0.11)
D	46.4 (8.5)	51.8 (10.2)	0.76 (0.19)
E	49.4 (6.9)	52.1 (9.5)	0.81 (0.13)
F	41.5 (12.3)	50.1 (9.3)	0.69 (0.34)
I	50.4 (8.0)	52.9 (7.8)	0.84 (0.14)
J	45.8 (9.9)	51.4 (13.2)	0.77 (0.17)
K	49.1 (9.5)	55.1 (9.2)	0.81 (0.13)
L	48.5 (9.2)	53.3 (9.4)	0.82 (0.13)
N	49.9 (8.7)	49.3 (7.6)	0.77 (0.15)
O	46.8 (12.5)	53.4 (7.8)	0.82 (0.14)
P	47.6 (10.8)	50.7 (11.8)	0.79 (0.20)
R	48.2 (8.0)	52.5 (8.5)	0.77 (0.24)
S	46.4 (6.3)	48.3 (12.1)	0.76 (0.17)
U	46.6 (9.2)	51.3 (10.3)	0.76 (0.19)
W	50.4 (7.4)	49.2 (11.2)	0.75 (0.21)
X	45.3 (11.0)	52.1 (7.8)	0.78 (0.19)
Total	47.3 (9.2)	51.3 (10.1)	0.78 (0.18)

Note: Pre-operative SF-36 and EQ-5D was not known for 42 and 18 patients, respectively

4.2.2 Post-operative health

Of the 561 patients recruited, 548 were eligible for follow-up (12 did not undergo surgery and the whereabouts of one was unknown). Of these, 435 (79.4%) patients returned a questionnaire three months after surgery. Overall, 57.6% of patients rated their general health status as 'Excellent' or 'Very good' (Table 4.14), similar to the proportion before surgery (56.2%).

Table 4.14: Post-operative general health status

Centre	In general, would you say your health is:				
	Excellent, n (%)	Very good, n (%)	Good, n (%)	Fair, n (%)	Poor, n (%)
A	3 (23.1)	4 (30.8)	4 (30.8)	2 (15.4)	0 (0)
C	3 (17.7)	7 (41.2)	7 (41.2)	0 (0)	0 (0)
D	6 (10.9)	23 (41.8)	19 (34.6)	4 (7.3)	3 (5.5)
E	8 (33.3)	9 (37.5)	7 (29.2)	0 (0)	0 (0)
F	0 (0)	2 (25.0)	5 (62.5)	0 (0)	1 (12.5)
I	5 (26.3)	11 (57.9)	2 (10.5)	0 (0)	1 (5.3)
J	5 (17.9)	6 (21.4)	16 (57.1)	1 (3.6)	0 (0)
K	4 (23.5)	11 (64.7)	1 (5.9)	1 (5.9)	0 (0)
L	1 (6.3)	8 (50.0)	5 (31.3)	2 (12.5)	0 (0)
N	1 (5.6)	9 (50.0)	6 (33.3)	1 (5.6)	1 (5.6)
O	3 (30.0)	6 (60.0)	0 (0)	1 (10.0)	0 (0)
P	9 (18.4)	21 (42.9)	13 (26.5)	5 (10.2)	1 (2.0)
R	3 (11.1)	14 (51.9)	6 (22.2)	3 (11.1)	1 (3.7)
S	1 (6.3)	7 (43.8)	8 (50.0)	0 (0)	0 (0)
U	5 (10.6)	18 (38.3)	21 (44.7)	2 (4.3)	1 (2.1)
W	2 (6.1)	18 (54.6)	12 (36.4)	1 (3.0)	0 (0)
X	4 (11.1)	12 (33.3)	17 (47.2)	3 (8.3)	0 (0)
Total	63 (14.6)	186 (43.0)	149 (34.4)	26 (6.0)	9 (2.1)

Note: Post-operative general health status not known for 2 patients

Table 4.15 shows the distribution of pre-operative and post-operative SF-36 PCS by centre. Overall, the mean change in score was 2.7 units (47.7 - 50.4), indicating an improvement in physical health. The mean change varied between centres (–0.6 to 6.8).

Table 4.15: Pre- and post-operative SF-36 PCS score

Centre	n	Pre-operative SF-36 PCS mean (SD)	Post-operative SF-36 PCS mean (SD)
A	8	48.6 (7.4)	49.4 (5.0)
C	16	45.7 (8.8)	52.5 (8.1)
D	54	46.6 (8.8)	46.9 (10.7)
E	22	50.9 (5.6)	50.3 (8.0)
F	8	42.8 (12.4)	44.5 (11.3)
I	17	50.8 (7.2)	54.4 (6.3)
J	27	46.9 (8.1)	47.2 (8.6)
K	17	49.4 (9.6)	52.4 (8.3)
L	16	48.0 (9.2)	54.0 (5.9)
N	17	49.9 (9.3)	52.5 (8.0)
O	10	47.9 (14.2)	51.7 (12.9)
P	42	48.6 (9.4)	52.6 (8.3)
R	23	48.7 (7.1)	49.8 (9.9)
S	12	46.2 (7.0)	49.0 (12.5)
U	46	46.6 (9.0)	51.6 (10.6)
W	30	49.7 (7.7)	49.1 (8.8)
X	35	45.9 (10.9)	50.2 (9.8)
Total	400	47.7 (8.8)	50.4 (9.4)

Note: Data restricted to patients with pre- and post-operative SF-36 scores

The risk adjustment model contained pre-op SF-36 PCS score, EQ-5D score, general health status, comorbidity and previous hernia surgery. These five variables explained 47% of the variation between patients in post-operative SF-36 PCS score.

In the funnel plot shown in Figure 4.4, the hollow circles represent each centre's unadjusted post-operative SF-36 PCS score and the solid circles represent each centre's adjusted score. The horizontal line represents target performance – mean post-operative PCS score for all centres combined (50.4). There was no statistically significant difference ($p<0.05$) between each centre's adjusted post-operative SF-36 PCS and target performance.

Figure 4.4: Funnel plot of post-operative SF-36 PCS in hernia surgery

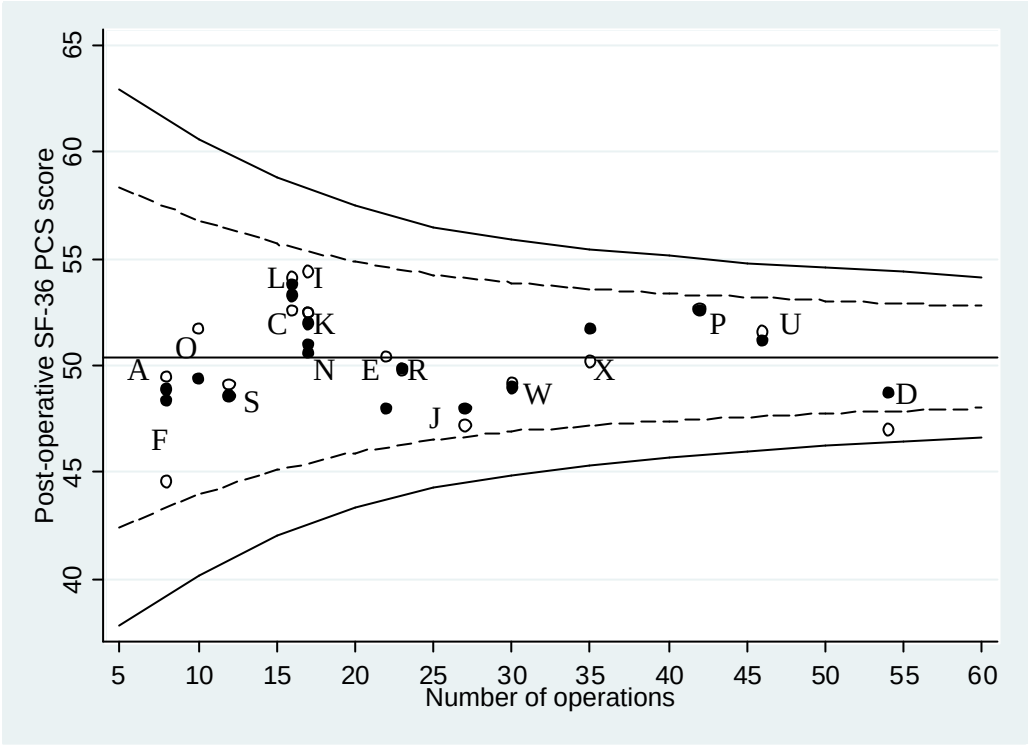


Table 4.16 shows the distribution of pre-operative and post-operative SF-36 MCS scores by centre. Overall, there was no change on the MCS, indicating that hernia surgery has no impact on mental health status as measured by the SF-36. Therefore, in all subsequent analyses of outcome after hernia repair we only show data for the PCS score on the SF-36.

Table 4.16: Pre- and post-operative SF-36 MCS scores in hernia surgery

Centre	n	Pre-operative SF-36 MCS mean (SD)	Post-operative SF-36 MCS mean (SD)
A	8	51.6 (10.9)	53.8 (13.0)
C	16	52.1 (11.4)	53.9 (8.1)
D	54	51.5 (10.1)	50.8 (11.8)
E	22	53.2 (7.0)	55.0 (5.7)
F	8	51.0 (9.5)	53.7 (10.1)
I	17	53.2 (8.2)	52.3 (10.1)
J	27	50.7 (13.1)	48.9 (10.1)
K	17	56.8 (7.8)	55.8 (6.1)
L	16	53.8 (9.5)	53.9 (5.8)
N	17	49.4 (7.7)	49.1 (9.3)
O	10	56.5 (3.7)	56.2 (2.7)
P	42	51.2 (9.2)	53.5 (8.4)
R	23	52.8 (8.6)	53.3 (8.1)
S	12	46.5 (13.3)	46.9 (10.2)
U	46	50.8 (10.8)	49.9 (10.2)
W	30	52.7 (7.5)	53.4 (6.6)
X	35	52.4 (7.9)	52.0 (9.4)
Total	400	52.0 (9.5)	52.1 (9.3)

Note: Data for patients with pre- and post-operative SF-36 scores

Table 4.17 shows the distribution of pre- and post-operative EQ-5D scores by centre. Overall, the mean change in score was 0.06 units (0.79 – 0.85), indicating an improvement in health state. The mean improvement varied between centres (–0.01 to 0.14).

Table 4.17: Pre- and post-operative EQ-5D scores in hernia surgery

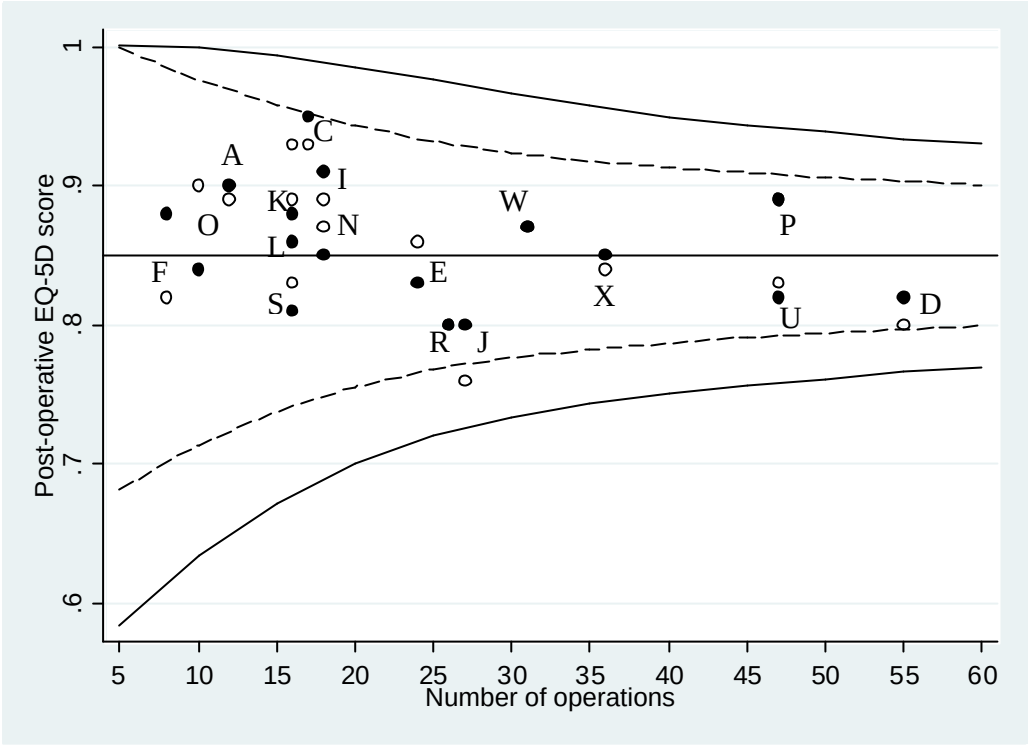
Centre	n	Pre-operative EQ-5D mean (SD)	Post-operative EQ-5D mean (SD)
A	12	0.76 (0.08)	0.89 (0.12)
C	17	0.79 (0.12)	0.93 (0.10)
D	55	0.75 (0.21)	0.80 (0.27)
E	24	0.82 (0.13)	0.87 (0.12)
F	8	0.76 (0.34)	0.83 (0.35)
I	18	0.83 (0.14)	0.88 (0.23)
J	27	0.77 (0.17)	0.76 (0.23)
K	16	0.82 (0.12)	0.93 (0.12)
L	16	0.83 (0.14)	0.90 (0.22)
N	18	0.76 (0.15)	0.89 (0.12)
O	10	0.82 (0.16)	0.90 (0.13)
P	47	0.80 (0.20)	0.89 (0.12)
R	26	0.75 (0.25)	0.80 (0.20)
S	16	0.74 (0.19)	0.83 (0.26)
U	47	0.78 (0.16)	0.84 (0.22)
W	31	0.82 (0.12)	0.88 (0.15)
X	36	0.79 (0.19)	0.84 (0.16)
Total	424	0.79 (0.18)	0.85 (0.20)

Note: Data on patients with pre- and post-operative EQ-5D scores

The risk adjustment model contained pre-op EQ-5D score, SF-36 PCS and MCS scores, general health status, comorbidity and previous hernia surgery. These five variables explained 38% of the variation between patients in post-operative EQ-5D scores.

In the funnel plot shown in Figure 4.5 the hollow circles represent each centre's unadjusted post-operative EQ-5D score and the solid circles represent adjusted scores. The horizontal line represents target performance – the mean post-operative EQ-5D score for all centres combined (0.85). There was no statistically significant difference between adjusted post-operative EQ-5D score and target performance at the $p < 0.05$ significance level.

Figure 4.5: Funnel plot of post-operative EQ-5D scores in hernia surgery



Overall, 24.1% of patients had post-operative complications. The percentage varied between centres from 5.9% to 39.4% (Table 4.18).

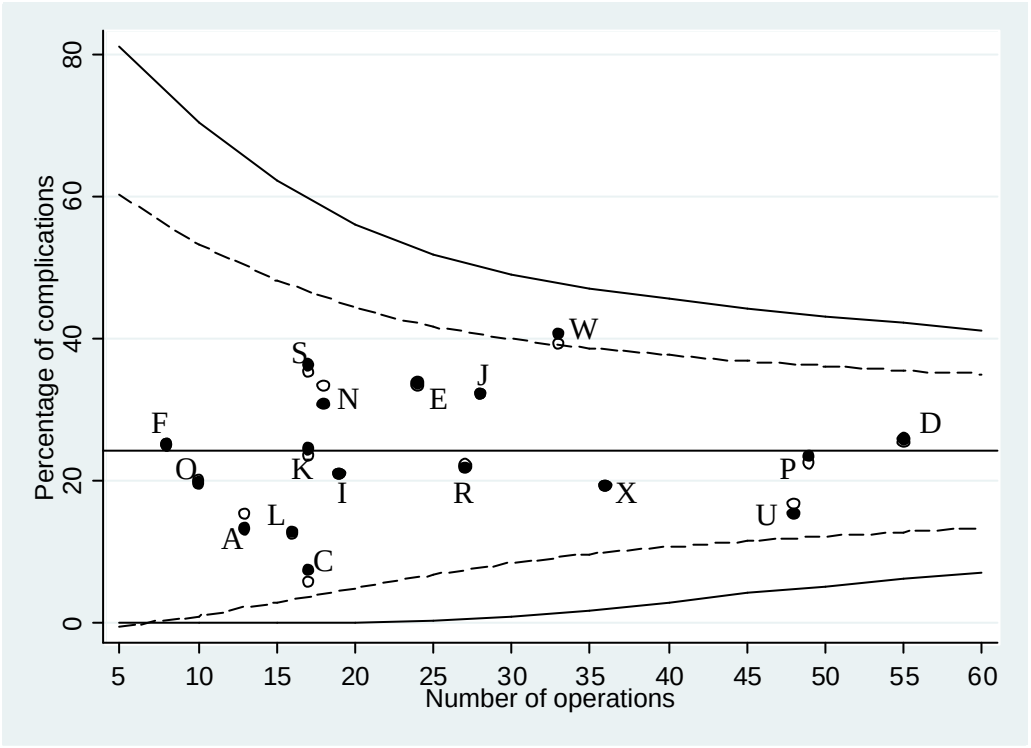
Table 4.18: Incidence of post-operative complications in hernia surgery

Centre	Allergy or reaction to drug, n (%)	Urinary problems, n (%)	Bleeding, n (%)	Wound problems, n (%)	Any complication, n (%)
A	2 (15.4)	0 (0)	0 (0)	0 (0)	2 (15.4)
C	1 (5.9)	0 (0)	1 (5.9)	0 (0)	1 (5.9)
D	3 (5.5)	4 (7.3)	6 (10.9)	7 (12.7)	14 (25.5)
E	4 (16.7)	2 (8.3)	1 (4.2)	1 (4.2)	8 (33.3)
F	2 (25.0)	0 (0)	0 (0)	0 (0)	2 (25.0)
I	1 (5.3)	0 (0)	0 (0)	3 (15.8)	4 (21.1)
J	1 (3.6)	4 (14.3)	2 (7.1)	3 (10.7)	9 (32.1)
K	0 (0)	3 (17.7)	1 (5.9)	1 (5.9)	4 (23.5)
L	1 (6.3)	0 (0)	0 (0)	1 (6.3)	2 (12.5)
N	1 (5.6)	2 (11.1)	3 (16.7)	2 (11.1)	6 (33.3)
O	0 (0)	0 (0)	1 (10.0)	1 (10.0)	2 (20.0)
P	1 (2.0)	1 (2.0)	2 (4.1)	9 (18.4)	11 (22.5)
R	2 (7.4)	0 (0)	2 (7.4)	3 (11.1)	6 (22.2)
S	2 (11.8)	2 (11.8)	0 (0)	3 (17.7)	6 (35.3)
U	1 (2.1)	3 (6.3)	2 (4.2)	4 (8.3)	8 (16.7)
W	1 (3.0)	3 (9.1)	3 (9.1)	9 (27.3)	13 (39.4)
X	4 (11.1)	1 (2.8)	2 (5.6)	0 (0)	7 (19.4)
Total	27 (6.2)	25 (5.6)	26 (6.0)	47 (10.8)	105 (24.1)

The risk adjustment model contained age and sex. The model discriminated between those patients who did and who did not have complications only slightly (area under ROC = 0.57).

In the funnel plot shown in Figure 4.6 the hollow circles represent each centre's unadjusted percentage of complications and the solid circles represent adjusted percentages. The horizontal line represents target performance – the overall percentage of complications for all centres combined (24.1%). The plot indicates that there is no statistically significant difference between each centre's adjusted percentage of complications and target performance at the $p < 0.05$ significance level with the exception of centre W where the complication rate was higher than expected.

Figure 4.6: Funnel plot of incidence of post-operative complications in hernia surgery



The majority of patients reported that their problems were better following surgery (93.8%) with little variation between centres (range 85.2-100%) (Table 4.19).

Table 4.19: Estimation of change in functioning following hernia surgery

Centre	Overall, how are you now compared to before your operation?				
	Much better, n (%)	A little better, n (%)	About the same, n (%)	A little worse, n (%)	Much worse, n (%)
A	10 (83.3)	1 (8.3)	0 (0)	0 (0)	1 (8.3)
C	15 (88.2)	0 (0)	2 (11.8)	0 (0)	0 (0)
D	48 (87.3)	4 (7.3)	2 (3.6)	0 (0)	1 (1.8)
E	21 (87.5)	0 (0)	1 (4.2)	1 (4.2)	1 (4.2)
F	8 (100)	0 (0)	0 (0)	0 (0)	0 (0)
I	19 (100)	0 (0)	0 (0)	0 (0)	0 (0)
J	20 (71.4)	6 (21.4)	0 (0)	1 (3.6)	1 (3.6)
K	14 (82.4)	3 (17.7)	0 (0)	0 (0)	0 (0)
L	13 (81.3)	3 (18.8)	0 (0)	0 (0)	0 (0)
N	17 (94.4)	0 (0)	1 (5.6)	0 (0)	0 (0)
O	7 (77.8)	2 (22.2)	0 (0)	0 (0)	0 (0)
P	40 (81.6)	7 (14.3)	2 (4.1)	0 (0)	0 (0)
R	20 (74.1)	3 (11.1)	1 (3.7)	2 (7.4)	1 (3.7)
S	13 (86.7)	0 (0)	1 (6.7)	0 (0)	1 (6.7)
U	41 (85.4)	4 (8.3)	2 (4.2)	0 (0)	1 (2.1)
W	25 (75.8)	5 (15.2)	1 (3.0)	2 (6.1)	0 (0)
X	34 (94.4)	1 (2.8)	1 (2.8)	0 (0)	0 (0)
Total	365 (84.7)	39 (9.1)	14 (3.3)	6 (1.4)	7 (1.6)

4.2.3 Summary and discussion

- Hernia surgery patients are predominantly male (90.2%), with a mean age of 54.3 years, and suffering from few, if any, comorbid conditions (median = 0).
- Surgery made little or no difference to patients general health status (the proportion reporting their general health as only fair or poor was 8.1% both before and after surgery). However, both the 'disease-specific' measure used (SF36-PCS) and the EQ-5D showed an improvement in patients' physical health.
- The overall self-reported complication rate was very high (24.1%). The proportion of patients who reported problems with their wound (10.8%) was twice that recently reported from Scotland (5.3%) (Taylor et al 2003). However, the latter was restricted to wound infections confirmed, via telephone interviews, by clinical staff.
- At none of the 17 centres did patients experience a statistically significant difference in outcome from that expected (at the $p < 0.05$ significance level) as regards improvement in health status (SF36-PCS, EQ-5D). One centre had a complication rate higher than expected, but given the number of centres, the existence of one outlier might have been due to chance at the $p < 0.05$ significance level.
- As expected, the mean SF36-MCS score did not change following surgery.
- There has been little change over the past 12 years in the pre-operative health of patients undergoing hernia surgery: the mean SF-36 PCS score for NHS patients in Oxford in 1992/94 was 46.4 (Jenkinson et al. 1997), 46.8 and 45.5 for private patients in, respectively, 1998 and 2005/6 (BUPA, personal communication).

4.3 Hip replacement surgery

4.3.1 Pre-operative characteristics of patients

Overall 397 patients were recruited in 11 centres. Table 4.20 shows that, between centres, the mean age ranged from 62 to 72 years and the percentage of women from 43% to 76%. This is similar to all patients treated by the NHS in 2005-6: 70 years and 62.3% female. The median duration of symptoms varied at the centre level from 2.0 to 4.5 years; and the percentage of patients who had previously had a hip replacement varied from 13% to 39%.

Table 4.20: Patient demographics and history of hip problems

Centre	n	Age (yrs), mean (SD)	Female, n (%)	Duration of symptoms (yrs), median (IQR)	Previous hip replacement, n (%)
D	21	71.9 (7.4)	16 (76.2)	2 (1.5 – 3)	4 (19.1)
F	37	63.0 (11.1)	17 (46.0)	3 (2 – 5)	9 (24.3)
J	45	70.4 (9.6)	26 (57.8)	3 (2 – 5)	12 (26.7)
K	23	69.4 (7.6)	14 (60.9)	2 (2 – 4)	7 (30.4)
N	18	66.8 (8.0)	8 (44.4)	2.3 (1.3 – 7.3)	7 (38.9)
O	30	67.7 (10.8)	15 (50.0)	2 (1.5 – 5)	6 (20.7)
Q	14	61.7 (10.0)	6 (42.9)	4.5 (2.5 – 5.5)	5 (35.7)
R	61	69.6 (10.1)	36 (60.0)	2.3 (2 – 6)	10 (16.7)
T	42	67.1 (12.1)	19 (45.2)	2 (2 – 4)	13 (31.0)
U	69	68.7 (7.9)	44 (63.8)	3 (2 – 5)	9 (13.2)
W	37	67.5 (10.4)	22 (59.5)	2.5 (2 - 4)	9 (24.3)
Total	397	68.0 (10.0)	223 (56.3)	3 (2 - 5)	91 (23.1)

Note: Data missing on age (8 patients); gender (1); duration of symptoms (31); previous knee replacement surgery (3).

The number of comorbidities per patient did not vary greatly between centres (Table 4.21).

Table 4.21: Patient comorbidities

Centre	No. of comorbidities per patient, median (IQR)	No. of comorbidities which affect surgical outcome* per patient, median (IQR)
D	2 (1 - 2)	1 (0 – 1)
F	1 (0 – 2)	1 (0 – 1)
J	2 (1 – 3)	1 (0 – 1)
K	2 (1 – 3)	1 (1 – 1)
N	2 (1 – 3)	0.5 (0 – 1)
O	2 (0 - 3)	0 (0 - 1)
Q	2 (1 - 3)	1 (0 - 1)
R	2 (1 – 3)	0 (0 – 1)
T	2 (1 – 4)	1 (0 – 1)
U	1 (1 – 2)	1 (0 – 1)
W	2 (1 - 3)	1 (0 – 1)
Total	2 (1 - 3)	1 (0 – 1)

* Angina, asthma, bronchitis, diabetes, high blood pressure, kidney disease, liver disease, poor circulation

Before surgery, 34.7% of patients rated their general health status as ‘Excellent’ or ‘Very good’ (Table 4.22), (21% to 51% between centres) and those rating their health ‘Fair’ or ‘Poor’ was 20.2% (6.6 - 33.3%).

Table 4.22: Pre-operative general health status of patients

Centre	In general, would you say your health is:				
	Excellent, n (%)	Very good, n (%)	Good, n (%)	Fair, n (%)	Poor, n (%)
D	2 (9.5)	5 (23.8)	11 (52.4)	3 (14.3)	0 (0)
F	3 (8.1)	8 (21.6)	22 (59.5)	3 (8.1)	1 (2.7)
J	6 (13.3)	15 (33.3)	21 (46.7)	2 (4.4)	1 (2.2)
K	2 (8.7)	7 (30.4)	8 (34.8)	3 (13.0)	3 (13.0)
N	0 (0)	4 (22.2)	11 (61.1)	3 (16.7)	0 (0)
O	1 (3.3)	7 (23.3)	12 (40.0)	9 (30.0)	1 (3.3)
Q	0 (0)	3 (21.4)	8 (57.1)	2 (14.3)	1 (7.1)
R	5 (8.3)	13 (21.7)	23 (38.3)	16 (26.7)	3 (5.0)
T	1 (2.5)	8 (20.0)	20 (50.0)	8 (20.0)	3 (7.5)
U	1 (1.5)	26 (38.8)	27 (40.3)	12 (17.9)	1 (1.5)
W	3 (8.1)	16 (43.2)	14 (37.8)	2 (5.4)	2 (5.4)
Total	24 (6.1)	112 (28.6)	177 (45.2)	63 (16.1)	16 (4.1)

Table 4.23 shows the distribution of pre-operative Oxford Hip Scores (OHS) and EQ-5D scores by centre. The mean OHS varied between centres from 41.1 to 48.3 and the mean EQ-5D scores from 0.22 to 0.39.

Table 4.23: Pre-operative Oxford Hip and EQ-5D scores

Centre	Pre-operative OHS mean (SD)	Pre-operative EQ-5D mean (SD)
D	48.3 (7.0)	0.22 (0.27)
F	43.8 (7.0)	0.32 (0.30)
J	44.9 (7.2)	0.31 (0.29)
K	44.8 (7.5)	0.29 (0.37)
N	41.3 (6.9)	0.32 (0.31)
O	41.1 (7.3)	0.38 (0.32)
Q	41.2 (9.5)	0.39 (0.36)
R	44.1 (8.6)	0.27 (0.32)
T	44.8 (7.3)	0.31 (0.32)
U	41.1 (8.9)	0.37 (0.32)
W	41.6 (7.0)	0.34 (0.33)
Total	43.3 (8.0)	0.32 (0.32)

Note: Oxford Hip and EQ-5D was not known for 1 and 4 patients, respectively

4.3.2 Post-operative health

Of the 397 patients recruited, 383 were eligible for follow-up after 6 months (2 had died, 11 had not undergone surgery and one could not be traced). Of these, 351 (91.6%) returned a post-operative questionnaire. Overall, 49% rated their general health status as 'Excellent' or 'Very good' 6-months after surgery (Table 4.24), compared with 35% before surgery.

Table 4.24: Post-operative general health status

Centre	In general, would you say your health is:				
	Excellent, n (%)	Very good, n (%)	Good, n (%)	Fair, n (%)	Poor, n (%)
D	1 (5.6)	10 (55.6)	5 (27.8)	2 (11.1)	0 (0)
F	6 (16.7)	15 (41.7)	10 (27.8)	5 (13.9)	0 (0)
J	7 (18.9)	12 (32.4)	12 (32.4)	5 (13.5)	1 (2.7)
K	2 (10.5)	10 (52.6)	5 (26.3)	2 (10.5)	0 (0)
N	1 (5.9)	4 (23.5)	10 (58.8)	2 (11.8)	0 (0)
O	3 (11.5)	8 (30.8)	8 (30.8)	5 (19.2)	2 (7.7)
Q	2 (16.7)	5 (41.7)	5 (41.7)	0 (0)	0 (0)
R	5 (9.3)	20 (37.0)	17 (31.5)	10 (18.5)	2 (3.7)
T	0 (0)	11 (32.4)	11 (32.4)	8 (23.5)	4 (11.8)
U	7 (11.9)	24 (40.7)	21 (35.6)	7 (11.9)	0 (0)
W	3 (9.1)	13 (39.4)	12 (36.4)	4 (12.1)	1 (3.0)
Total	37 (10.7)	132 (38.3)	116 (33.6)	50 (14.5)	10 (2.9)

Table 4.25 shows the distribution of pre-operative and post-operative OHS by centre. Overall, the mean change in score was 20 units (43.2 – 23.2). The mean improvement varied between centres (15.6 to 24.9).

Table 4.25: Pre- and post-operative Oxford Hip Scores (OHS)

Centre	n	Pre-operative OHS mean (SD)	Post-operative OHS mean (SD)
D	18	47.6 (7.3)	27.7 (9.8)
F	36	43.8 (7.1)	19.8 (6.6)
J	36	44.9 (7.6)	25.0 (8.8)
K	19	43.4 (7.3)	18.5 (5.3)
N	17	41.1 (7.0)	23.1 (9.1)
O	28	41.0 (7.3)	23.4 (8.8)
Q	12	41.0 (8.9)	25.2 (9.6)
R	54	44.3 (8.4)	23.4 (10.9)
T	35	45.0 (7.2)	29.4 (12.3)
U	60	41.4 (8.6)	20.4 (7.7)
W	34	41.5 (7.1)	22.8 (10.2)
Total	349	43.2 (7.9)	23.2 (9.6)

The risk adjustment model contained pre-op OHS, previous hip replacement, comorbidity and general health status. These four variables explained 24% of the variation in post-operative Oxford Hip scores.

In the funnel plot shown in Figure 4.7 the hollow circles represent each centre's unadjusted post-operative Oxford Hip score and the solid circles represent each centre's adjusted post-operative Oxford Hip score. The horizontal line represents the mean post-operative Oxford Hip Score for all centres combined (23.2). The adjusted scores for Centres K and F were significantly better than expected at the $p < 0.05$ significance level while that for Centre T was significantly worse.

Figure 4.7: Funnel plot of post-operative Oxford Hip Score in hip replacement surgery.

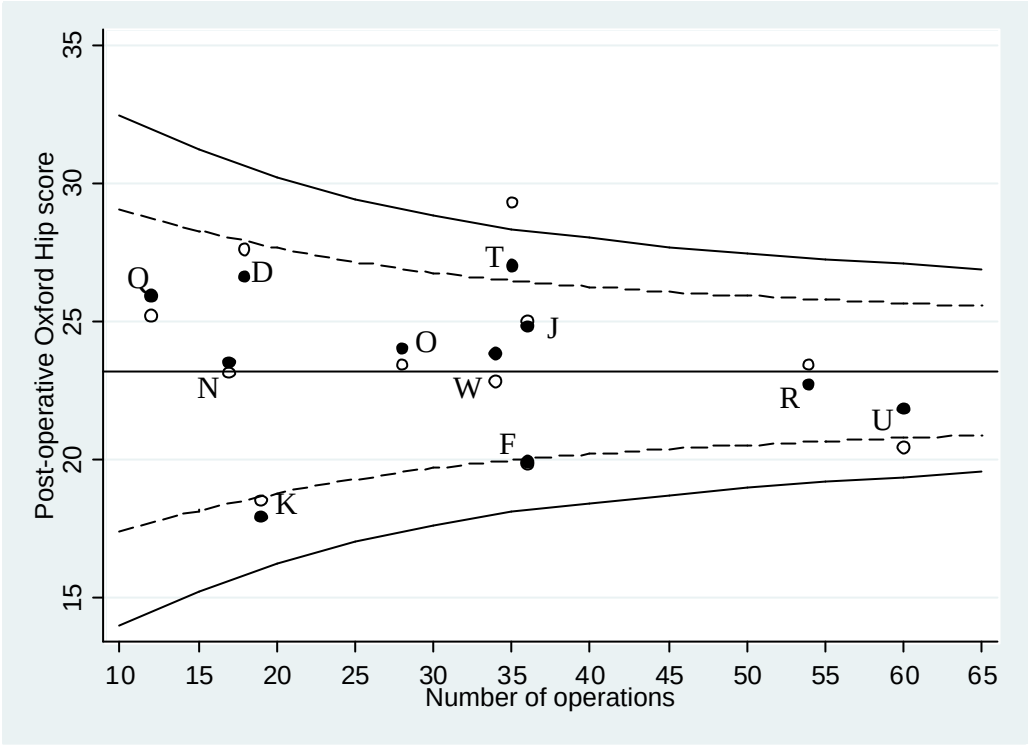


Table 4.26 shows the distribution of pre-operative and post-operative EQ-5D scores by centre. Overall, the mean change in score was 0.42 units (0.32 – 0.74). The improvement in health varied between centres (0.30 to 0.50).

Table 4.26: Pre- and post-operative EQ-5D scores in hip replacement surgery.

Centre	n	Pre-operative EQ-5D mean (SD)	Post-operative EQ-5D mean (SD)
D	18	0.24 (0.28)	0.67 (0.24)
F	36	0.31 (0.31)	0.79 (0.24)
J	37	0.28 (0.29)	0.69 (0.24)
K	19	0.35 (0.35)	0.79 (0.22)
N	17	0.34 (0.31)	0.77 (0.23)
O	27	0.38 (0.32)	0.73 (0.25)
Q	12	0.43 (0.33)	0.80 (0.20)
R	54	0.26 (0.31)	0.76 (0.24)
T	34	0.30 (0.33)	0.60 (0.35)
U	59	0.35 (0.31)	0.82 (0.16)
W	33	0.35 (0.33)	0.71 (0.28)
Total	346	0.32 (0.31)	0.74 (0.25)

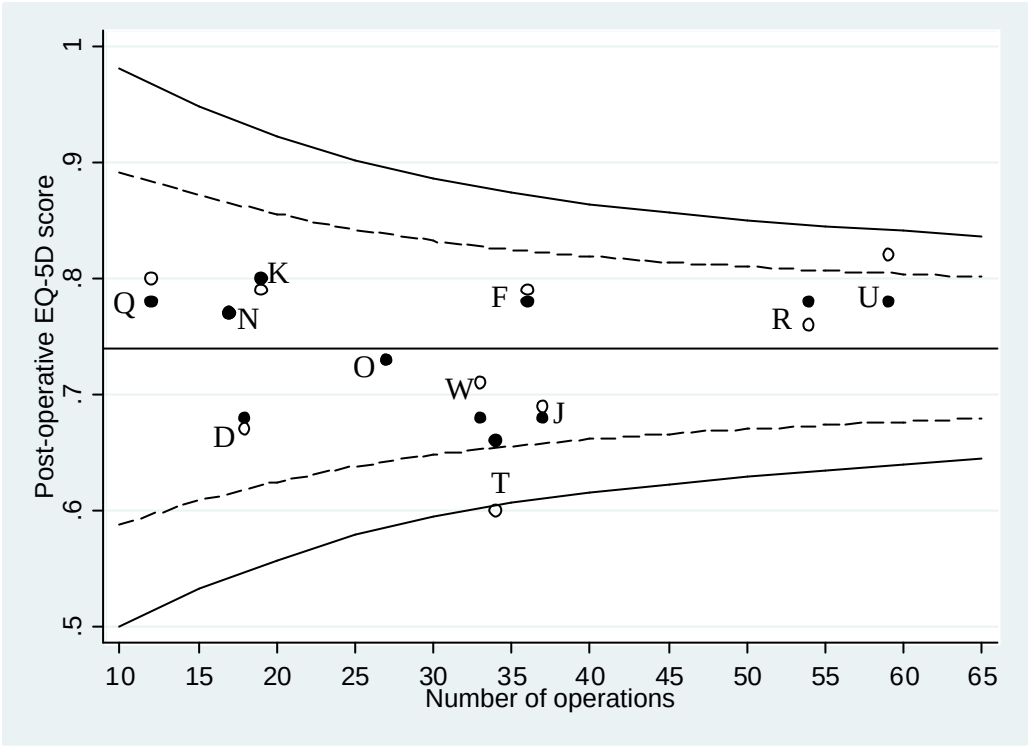
Note: Data on patients with pre- and post-operative EQ-5D scores

The risk adjustment model contained pre-op EQ-5D score, OHS, previous hip replacement, general health status, and comorbidity. These five variables explained 26% of the variation between patients' post-operative EQ-5D scores.

In the funnel plot shown in Figure 4.8 the hollow circles represent each centre's unadjusted post-operative EQ-5D score and the solid circles represent each centre's adjusted post-operative EQ-5D score. The horizontal line

represents target performance – the mean post-operative EQ-5D score for all centres combined (0.74). There is no statistical difference between each centre's adjusted post-operative EQ-5D score and target performance.

Figure 4.8: Funnel plot of post-operative EQ-5D scores in hip replacement surgery.



Overall, 32.5% of patients had post-operative complications. The percentage varied between centres from 21.1% to 61.1% (Table 4.27).

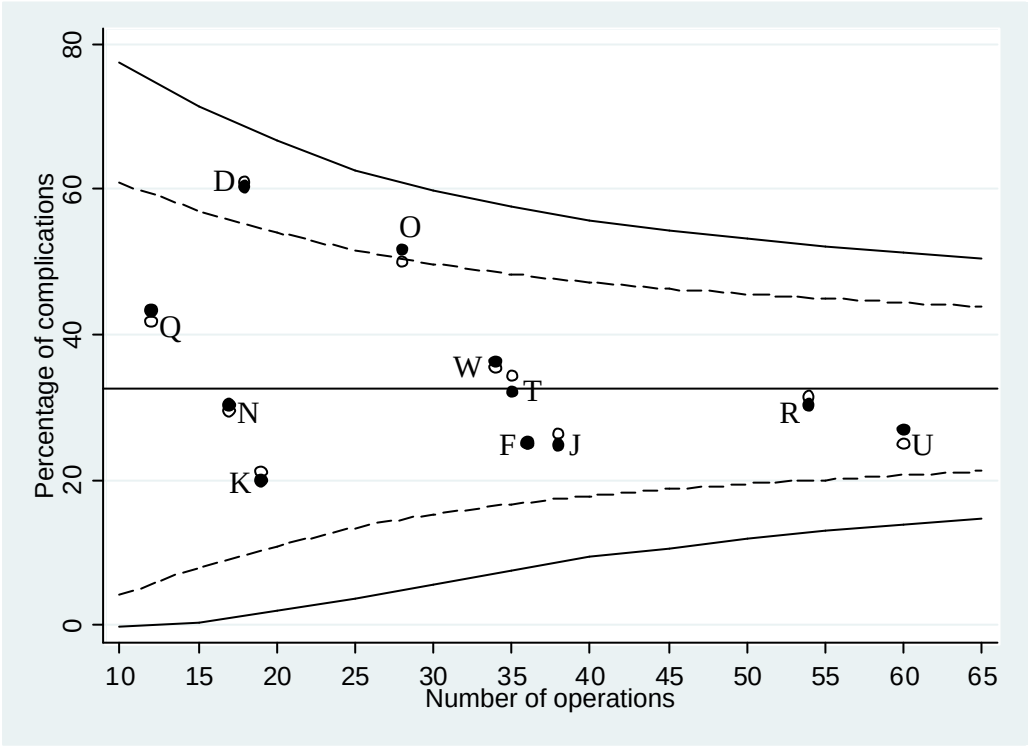
Table 4.27: Incidence of post-operative complications in hip replacement surgery.

Centre	Allergy or reaction to drug, n (%)	Urinary problems n (%)	Bleeding, n (%)	Wound problems n (%)	Any complication n (%)
D	5 (27.8)	2 (11.1)	3 (16.7)	4 (22.2)	11 (61.1)
F	2 (5.6)	3 (8.3)	6 (16.7)	2 (5.6)	9 (25.0)
J	5 (13.2)	4 (10.5)	3 (7.9)	1 (2.6)	10 (26.3)
K	0 (0)	2 (10.5)	1 (5.3)	2 (10.5)	4 (21.1)
N	2 (11.8)	1 (5.9)	1 (5.9)	2 (11.8)	5 (29.4)
O	3 (10.7)	6 (21.4)	2 (7.1)	5 (17.9)	14 (50.0)
Q	1 (8.3)	3 (25.0)	1 (8.3)	1 (8.3)	5 (41.7)
R	5 (9.3)	8 (14.8)	5 (9.3)	6 (11.1)	17 (31.5)
T	6 (17.1)	4 (11.4)	5 (14.3)	4 (11.4)	12 (34.3)
U	4 (6.7)	4 (6.7)	3 (5.0)	6 (10.0)	15 (25.0)
W	4 (11.8)	4 (11.8)	4 (11.8)	3 (8.8)	12 (35.3)
Total	37 (10.5)	41 (11.7)	34 (9.7)	36 (10.3)	114 (32.5)

The risk adjustment model contained pre-op Oxford Hip score and comorbidity. The model discriminated between those patients who did and who did not have complications only slightly (area under ROC = 0.59).

In the funnel plot shown in Figure 4.9 the hollow circles represent each centre's unadjusted percentage of complications and the solid circles represent each centre's adjusted percentage of complications. The horizontal line represents target performance – the percentage of complications for all centres combined (32.5%). The incidence of complications in Centres D and O were significantly higher than expected at the $p < 0.05$ significance level.

Figure 4.9: Funnel plot of incidence of post-operative complications in hip replacement surgery.



The majority of patients reported that they were better following surgery (94.6%) with little variation between centres (range 85.7 -100%) (Table 4.28).

Table 4.28: Estimation of change in functioning following hip replacement surgery.

Centre	Overall, how are you now compared to before your operation?				
	Much better, n (%)	A little better, n (%)	About the same, n (%)	A little worse, n (%)	Much worse, n (%)
D	15 (83.3)	2 (11.1)	0 (0)	0 (0)	1 (5.6)
F	34 (94.4)	2 (5.6)	0 (0)	0 (0)	0 (0)
J	33 (86.8)	4 (10.5)	1 (2.6)	0 (0)	0 (0)
K	19 (100)	0 (0)	0 (0)	0 (0)	0 (0)
N	16 (94.1)	1 (5.9)	0 (0)	0 (0)	0 (0)
O	23 (85.2)	2 (7.4)	2 (7.4)	0 (0)	0 (0)
Q	10 (83.3)	1 (8.3)	0 (0)	1 (8.3)	0 (0)
R	47 (87.0)	4 (7.4)	2 (3.7)	1 (1.9)	0 (0)
T	25 (71.4)	5 (14.3)	1 (2.9)	1 (2.9)	3 (8.6)
U	51 (87.9)	4 (6.9)	1 (1.7)	2 (1.7)	0 (0)
W	29 (85.3)	2 (5.9)	1 (2.9)	0 (0)	2 (5.9)
Total	302 (86.8)	27 (7.8)	8 (2.3)	5 (1.4)	6 (1.7)

4.3.3 Summary and discussion

- Overall, the general health of patients undergoing hip replacement was not good: 20.2% rated their health as fair or poor and this was reflected in a mean EQ-5D score of 0.32.
- There was considerable variation between centres in patients' age and sex mix, duration of their symptoms, previous hip surgery and general health status.
- The mean pre-operative OHS (43.3) was similar to that reported in a recent national audit (44.7) (RCS/BOA 2000), as was the mean post-operative score (23.2 compared with 26.6 at 3 months and 21.9 at 12 months). This suggests that hip replacement could be assessed at the same time as other procedures, three months post-op, in audits.
- Surgery led to a considerable improvement in their primary problem (mean OHS improved from 43.2 to 23.2) and this was reflected in an improvement in their general health status (49% rated excellent or very good compared to 35% before surgery) and EQ-5D (mean score from 0.32 to 0.74).
- Although none of the 11 centres differed significantly from their expected outcome in terms of the patients' post-operative EQ-5D scores, in one centre the improvement in OHS was significantly less than expected (at the $p < 0.05$ significance level).
- The reported incidence of wound problems (10.3%) was higher than that previously reported (4.6%) though the latter was based on GPs' reports rather than (patients' reports (RCS/BOA 2000). In two centres, the incidence of complications was higher than expected.

4.4 Knee replacement surgery

4.4.1 Pre-operative characteristics of patients

Overall, 400 patients were recruited in 11 centres. Table 4.29 shows that the mean age ranged from 66.0 to 74.5 years, the percentage of women from 40% to 90%, the median length of symptoms varied from 4 to 6 years, and the percentage of patients who had previously had a knee replacement varied from 36% to 67%. This is similar to all patients treated by the NHS in 2005-6: 70 years and 57.7% female.

Table 4.29: Patient demographics and duration of symptoms

Centre	n	Age (yrs), mean (SD)	Female, n (%)	Duration of symptoms (yrs), median (IQR)	Previous knee replacement n (%)
D	26	67.1 (7.8)	12 (46.2)	6 (3 - 10)	9 (36.0)
F	58	68.4 (8.5)	27 (46.6)	6 (3 - 12)	28 (50.0)
J	46	69.7 (8.7)	20 (43.5)	5 (2 - 9)	18 (40.0)
K	11	74.5 (6.6)	7 (63.6)	6 (2 – 10)	4 (36.4)
N	10	71.0 (8.2)	9 (90.0)	4.5 (2.5 - 6)	6 (66.7)
O	35	71.6 (6.5)	17 (50.0)	5 (3 - 10)	17 (50.0)
Q	25	66.1 (12.4)	10 (40.0)	5 (2.5 - 10)	10 (40.0)
R	65	68.8 (11.7)	38 (58.5)	4 (2 – 10)	31 (48.4)
T	47	67.6 (9.5)	27 (58.7)	5 (2 – 10)	19 (41.3)
U	61	66.9 (8.7)	35 (58.3)	6 (3 - 10)	24 (39.3)
W	16	67.2 (16.8)	10 (62.5)	5 (3 - 7)	6 (37.5)
Total	400	68.6 (9.8)	212 (53.4)	5 (3 - 10)	172 (43.9)

Note: Missing data: age (14 patients); gender (3); duration of symptoms (27); previous knee replacement surgery (8)

The number of comorbidities per patient did not vary greatly between centres (Table 4.30).

Table 4.30: Patient comorbidities

Centre	No. of comorbidities per patient, median (IQR)	No. of comorbidities which affect surgical outcome* per patient, median (IQR)
D	2.5 (1 - 4)	1 (0 - 2)
F	2 (1 - 3)	1 (0 – 1)
J	2 (1 - 3)	0.5 (0 – 1)
K	3 (1 – 3)	1 (0 – 2)
N	3.5 (1 - 5)	2 (0 – 3)
O	2 (1 - 3)	1 (0 - 1)
Q	2 (1 - 3)	1 (0 - 1)
R	2 (1 – 3)	1 (0 – 1)
T	3 (2 – 4)	1 (0 – 2)
U	2 (1 – 2)	1 (0 – 1)
W	2 (1.5 - 4)	1 (0 - 2)
Total	2 (1 - 3)	1 (0 - 1)

* Angina, asthma, bronchitis, diabetes, high BP, kidney disease, liver disease, poor circulation

Before surgery, 31.9% of patients rated their general health status as 'Excellent' or 'Very good' (Table 4.31) (varied between centres from 10% to 48%) and 20.1% as 'Fair' or 'Poor' (range 9.1 to 33.4%).

Table 4.31: Pre-operative general health status

Centre	In general, would you say your health is:				
	Excellent, n (%)	Very good, n (%)	Good, n (%)	Fair, n (%)	Poor, n (%)
D	0 (0)	3 (11.5)	19 (73.1)	4 (15.4)	0 (0)
F	1 (1.9)	11 (20.4)	24 (44.4)	15 (27.8)	3 (5.6)
J	1 (2.3)	17 (38.6)	22 (50.0)	4 (9.1)	0 (0)
K	0 (0)	1 (10.0)	6 (60.0)	2 (20.0)	1 (10.0)
N	0 (0)	2 (22.2)	4 (44.4)	3 (33.3)	0 (0)
O	1 (2.9)	11 (31.4)	19 (54.3)	4 (11.4)	0 (0)
Q	2 (8.0)	10 (40.0)	9 (36.0)	4 (16.0)	0 (0)
R	2 (3.2)	19 (30.2)	28 (44.4)	12 (19.1)	2 (3.2)
T	0 (0)	11 (23.4)	22 (46.8)	11 (23.4)	1 (1.7)
U	4 (6.7)	23 (38.3)	27 (45.0)	5 (8.3)	1 (1.7)
W	0 (0)	5 (31.3)	7 (43.8)	4 (25.0)	0 (0)
Total	11 (2.8)	113 (29.1)	187 (48.1)	68 (17.5)	10 (2.6)

Table 4.32 shows the distribution of pre-operative Oxford Knee Scores (OKS) and EQ-5D scores by centre. Mean OKS (37.2 to 45.8) and EQ-5D scores (0.23 to 0.62) varied considerably between centres.

Table 4.32: Pre-operative Oxford Knee Scores and EQ-5D scores

Centre ID	Pre-operative OKS mean (SD)	Pre-operative EQ-5D mean (SD)
D	41.8 (6.3)	0.43 (0.30)
F	42.5 (8.1)	0.37 (0.35)
J	41.2 (8.5)	0.38 (0.33)
K	45.8 (7.6)	0.28 (0.29)
N	40.4 (6.7)	0.36 (0.28)
O	37.2 (6.9)	0.50 (0.26)
Q	43.6 (6.2)	0.33 (0.33)
R	43.1 (7.3)	0.35 (0.33)
T	45.4 (6.2)	0.23 (0.31)
U	40.5 (7.5)	0.43 (0.29)
W	37.9 (7.4)	0.62 (0.22)
Total	41.9 (7.6)	0.38 (0.32)

Note: Pre-operative Oxford Knee and EQ-5D was not known for 1 patient

4.4.2 Post-operative health

Of the 400 patients recruited, 387 were eligible for follow-up 6 months after surgery (11 had not undergone surgery, 2 could not be traced). Of these, 349 (90.2%) patients returned a post-operative questionnaire. There was some improvement in people's general health status: 38.1% of patients rated their general health status as 'Excellent' or 'Very good' 6-months after surgery (Table 4.33) compared with 31.9% before surgery.

Table 4.33: Post-operative general health status

Centre	In general, would you say your health is:				
	Excellent, n (%)	Very good, n (%)	Good, n (%)	Fair, n (%)	Poor, n (%)
D	1 (4.6)	6 (27.3)	11 (50.0)	4 (18.2)	0 (0)
F	1 (2.0)	11 (21.6)	22 (43.1)	14 (27.5)	3 (5.9)
J	2 (5.1)	19 (48.7)	14 (35.9)	3 (7.7)	1 (2.6)
K	1 (9.1)	1 (9.1)	6 (54.6)	3 (27.3)	0 (0)
N	1 (11.1)	2 (22.2)	2 (22.2)	4 (44.4)	0 (0)
O	1 (3.2)	14 (45.2)	13 (41.9)	1 (3.2)	2 (6.5)
Q	1 (8.3)	6 (50.0)	3 (25.0)	2 (16.7)	0 (0)
R	1 (1.6)	22 (36.1)	26 (42.6)	9 (14.8)	3 (4.9)
T	1 (2.4)	12 (29.3)	20 (48.8)	6 (14.6)	2 (4.9)
U	2 (3.7)	21 (38.9)	26 (48.2)	5 (9.3)	0 (0)
W	0 (0)	5 (38.5)	6 (46.2)	2 (15.4)	0 (0)
Total	12 (3.5)	119 (34.6)	149 (43.3)	53 (15.4)	11 (3.2)

Table 4.34 shows the distribution of pre-operative and post-operative OKS by centre. Overall, the mean improvement in score was 14.7 units (41.7 – 27.0). The improvement varied between centres (12.7 to 19.8).

Table 4.34: Pre- and post-operative Oxford Knee Scores

Centre	n	Pre-operative OKS mean (SD)	Post-operative OKS mean (SD)
D	22	40.6 (5.9)	24.7 (7.9)
F	52	42.6 (8.1)	28.4 (10.3)
J	41	40.9 (8.7)	27.2 (10.3)
K	11	45.8 (7.6)	26.0 (8.1)
N	9	41.4 (6.4)	26.0 (7.3)
O	31	36.5 (7.0)	22.6 (7.5)
Q	12	43.8 (5.4)	25.2 (9.7)
R	61	43.5 (7.1)	26.2 (8.2)
T	41	45.0 (6.4)	32.3 (8.8)
U	54	40.1 (7.7)	26.7 (9.1)
W	13	38.8 (5.4)	25.9 (9.5)
Total	347	41.7 (7.6)	27.0 (9.2)

Note: Data for patients with pre- and post-operative Oxford Knee scores

The risk adjustment model contained pre-op Oxford Knee Score, index of multiple deprivation (based on postcode of residence), comorbidity and general health status. These four variables explained 27% of the variation in post-operative scores.

In the funnel plot shown in Figure 4.10 the hollow circles represent each centre's unadjusted post-operative Oxford Knee score and the solid circles represent each centre's adjusted post-operative OKS. The horizontal line represents the mean post-operative OKS for all centres combined (27.0). There were no differences in adjusted outcomes at the $p < 0.05$ significance level.

Figure 4.10: Funnel plot of post-operative Oxford Knee Score in knee replacement surgery.

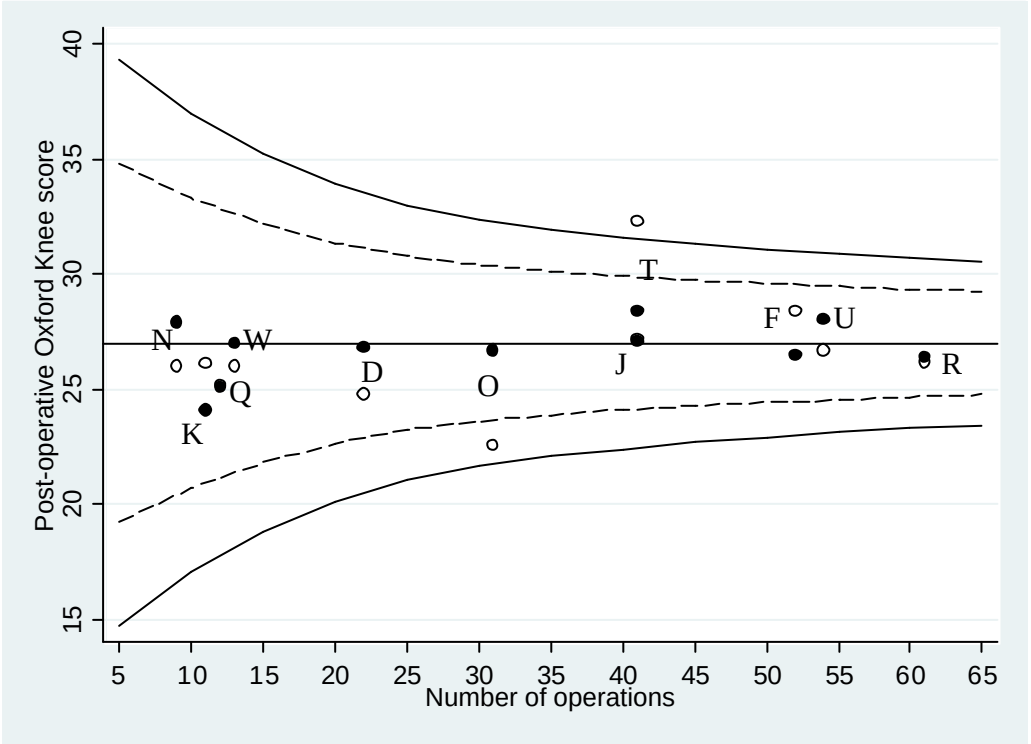


Table 4.35 shows the distribution of pre-operative and post-operative EQ-5D scores by centre. Overall, the mean change in score was 0.31 units (0.39 – 0.70). The improvement varied between centres (0.16 to 0.43).

Table 4.35: Pre- and post-operative EQ-5D scores in knee replacement surgery.

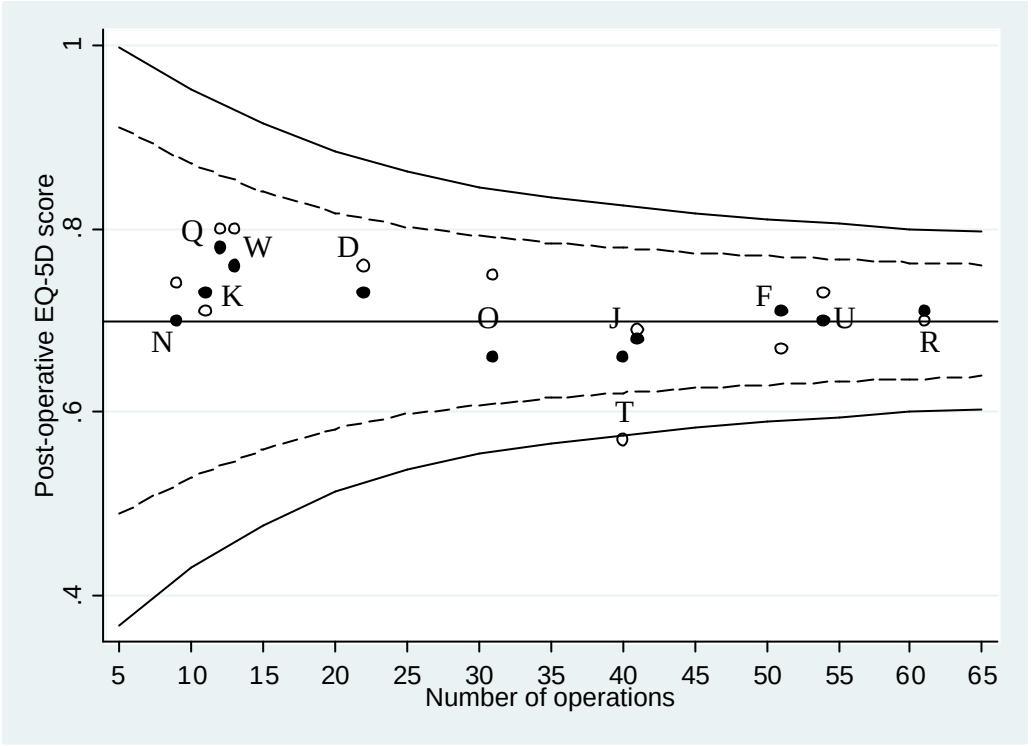
Centre	n	Pre-operative EQ-5D mean (SD)	Post-operative EQ-5D mean (SD)
D	22	0.49 (0.27)	0.77 (0.16)
F	51	0.37 (0.35)	0.67 (0.26)
J	41	0.39 (0.33)	0.69 (0.27)
K	11	0.28 (0.29)	0.71 (0.29)
N	9	0.37 (0.30)	0.75 (0.24)
O	31	0.49 (0.27)	0.75 (0.23)
Q	12	0.38 (0.29)	0.81 (0.15)
R	61	0.34 (0.33)	0.70 (0.25)
T	40	0.25 (0.31)	0.57 (0.30)
U	54	0.45 (0.29)	0.73 (0.20)
W	13	0.64 (0.16)	0.80 (0.21)
Total	345	0.39 (0.31)	0.70 (0.25)

Note: Data on patients with pre- and post-operative EQ-5D scores

The risk adjustment model contained pre-op EQ-5D score, Oxford Knee score, index of multiple deprivation, general health status, age and comorbidity. These six variables explained 27% of the variation between patients in post-operative EQ-5D scores.

In the funnel plot shown in Figure 4.11 the hollow circles represent each centre's unadjusted post-operative EQ-5D score and the solid circles represent each centre's adjusted post-operative EQ-5D score. The horizontal line represents the mean post-operative EQ-5D score for all centres combined (0.70). None of the centres adjusted outcomes were significantly different from expected at the $p < 0.05$ significance level.

Figure 4.11: Funnel plot of post-operative EQ-5D scores in knee replacement surgery.



Overall, 35.4% of patients had post-operative complications. The percentage varied considerably between centres from 7.7% to 75.0% (Table 4.36).

Table 4.36: Incidence of post-operative complications in knee replacement surgery.

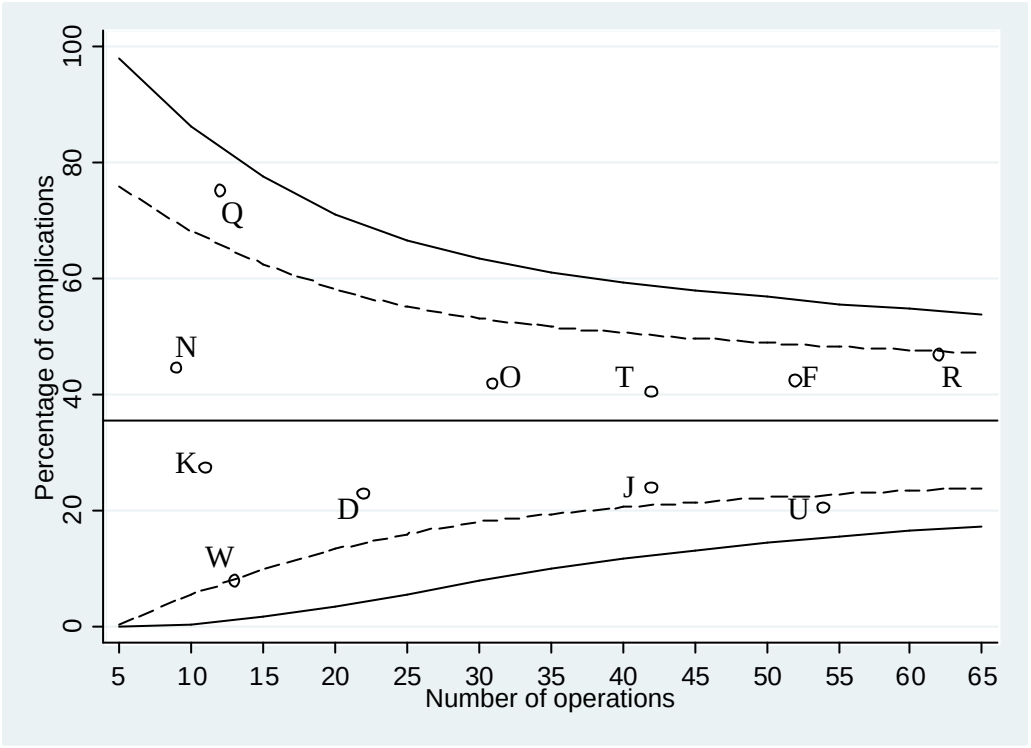
Centre	Allergy or reaction to drug, n (%)	Urinary problems n (%)	Bleeding n (%)	Wound problems n (%)	Any complication n (%)
D	1 (4.6)	2 (9.1)	1 (4.6)	2 (9.1)	5 (22.7)
F	5 (9.6)	11 (21.2)	6 (11.5)	8 (15.4)	22 (42.3)
J	4 (9.5)	5 (11.9)	0 (0)	3 (7.1)	10 (23.8)
K	1 (9.1)	1 (9.1)	0 (0)	1 (9.1)	3 (27.3)
N	1 (11.1)	0 (0)	3 (33.3)	1 (11.1)	4 (44.4)
O	8 (25.8)	7 (22.6)	2 (6.5)	2 (6.5)	13 (41.9)
Q	3 (25.0)	4 (33.3)	0 (0)	4 (33.3)	9 (75.0)
R	9 (14.5)	6 (9.7)	10 (16.1)	16 (25.8)	29 (46.8)
T	7 (16.7)	6 (14.3)	3 (7.1)	5 (11.9)	17 (40.5)
U	4 (7.4)	3 (5.6)	0 (0)	4 (7.4)	11 (20.4)
W	1 (7.7)	0 (0)	0 (0)	0 (0)	1 (7.7)
Total	44 (12.6)	45 (12.9)	25 (7.1)	46 (13.1)	124 (35.4)

No risk factor was significant at the 10% level so only unadjusted results are shown for post-operative complications.

In the funnel plot shown in Figure 4.12 the hollow circles represent each centre's unadjusted percentage of complications. The horizontal line represents the overall percentage of complications for all centres combined

(35.4%). The incidence of complications in Centre Q was higher than expected and that in Centre U lower than expected (p<0.05).

Figure 4.12: Funnel plot of post-operative complications in knee replacement surgery.



The majority of patients reported that they were better following surgery (86.9%) with some variation between centres (range 76.9 -100%) (Table 4.37).

Table 4.37: Estimation of change in functioning following knee surgery.

Centre	Overall, how are you now compared to before your operation?				
	Much better, n (%)	A little better, n (%)	About the same, n (%)	A little worse, n (%)	Much worse, n (%)
D	19 (86.4)	0 (0)	1 (4.6)	2 (9.1)	0 (0)
F	38 (76.0)	8 (16.0)	0 (0)	1 (2.0)	3 (6.0)
J	28 (70.0)	4 (10.0)	4 (10.0)	3 (7.5)	1 (2.5)
K	8 (72.7)	3 (27.3)	0 (0)	0 (0)	0 (0)
N	8 (88.9)	0 (0)	1 (11.1)	0 (0)	0 (0)
O	22 (71.0)	6 (19.4)	0 (0)	1 (3.2)	2 (6.5)
Q	9 (75.0)	3 (25.0)	0 (0)	0 (0)	0 (0)
R	43 (70.5)	7 (11.5)	7 (11.5)	4 (6.6)	0 (0)
T	28 (66.7)	11 (26.2)	2 (4.8)	0 (0)	1 (2.4)
U	32 (59.3)	13 (24.1)	5 (9.3)	2 (3.7)	2 (3.7)
W	10 (76.9)	0 (0)	1 (7.7)	1 (7.7)	1 (7.7)
Total	245 (71.0)	55 (15.9)	21 (6.1)	14 (4.1)	10 (2.9)

4.4.3 Summary and discussion

- The case-mix varied considerably between centres in terms of age, sex, duration of symptoms, previous surgery, general health status and EQ-5D scores.
- Surgery produced considerable improvement in patients' health status as determined by both the disease-specific (OKS) and the generic measure (EQ-5D). In contrast there was little change in patients' rating of their general health status with 13% of patients reporting they were no better or worse than before surgery.
- Unlike the four other procedures studied, the risk adjustment model for OKS and EQ-5D included the index of deprivation of the patient's area of residence. This may have been acting as a proxy for some other factor such as body mass index.
- None of the centres' performance differed significantly from that expected as judged by improvement in health status.
- The incidence of complications varied dramatically though only one centre was significantly higher than expected. The overall incidence (35.4%) was substantially higher than the 12% reported by patients recently (Singh et al 2004). Data for the latter audit was collected 12 months after surgery, which may partly explain the lower rate.

4.5 Varicose vein surgery

4.5.1 Pre-operative characteristics of patients

Overall 363 patients were recruited in 16 centres (though inter-centre comparisons exclude centres B, F and I as only 1, 2, and 8 patients were recruited at these centres, respectively, leaving 352). Table 4.38 shows that the mean age ranged from 42 to 53 years, the percentage of women from 40% to 83%, the median duration of symptoms from 8 to 20 years, and the percentage of patients who had previously undergone surgery for varicose veins from 25% to 42%. This is similar to all patients treated by the NHS in 2005-6: 50 years and 65.2% female.

Table 4.38: Patient demographics and varicose vein history

Centre	n	Age (yrs), mean (SD)	Female, n (%)	Duration of symptoms (yrs), median (IQR)	Previous varicose vein surgery, n (%)
A	45	48.2 (14.1)	27 (60.0)	12 (5 – 20)	11 (24.4)
C	18	48.6 (12.8)	12 (66.7)	16 (5 – 20)	6 (33.3)
D	41	47.7 (11.2)	27 (65.9)	11 (5 – 20)	14 (35.0)
E	18	46.1 (11.0)	8 (44.4)	10 (8 – 20)	7 (41.2)
J	19	49.1 (13.3)	11 (57.9)	13 (5 – 30)	6 (35.3)
K	10	50.9 (9.4)	4 (40.0)	20 (8 – 20)	4 (40.0)
L	14	53.4 (10.0)	9 (64.3)	15 (5 – 32)	4 (28.6)
N	30	47.0 (12.6)	23 (76.7)	10 (5 – 20)	8 (26.7)
Q	20	44.4 (15.5)	15 (75.0)	9 (5 – 20)	8 (42.1)
R	19	42.2 (9.6)	12 (63.2)	8 (5 – 15)	5 (27.8)
S	24	44.5 (14.7)	20 (83.3)	10 (6 – 16)	6 (25.0)
U	53	44.7 (13.2)	33 (62.3)	12 (6 – 15)	15 (28.9)
W	41	47.0 (13.4)	30 (73.2)	10 (5 – 18)	10 (25.6)
Total	352	46.8 (12.8)	231 (65.6)	10 (5 – 20)	104 (30.3)

Note: Data missing on age (6 patients); duration of symptoms (11); previous varicose vein surgery (9)

The number of comorbidities per patient did not vary greatly between centres (Table 4.39).

Table 4.39: Patient comorbidities

Centre	No. of comorbidities per patient, median (IQR)	No. of comorbidities which affect surgical outcome* per patient, median (IQR)
A	1 (0 – 2)	0 (0 – 1)
C	0 (0 – 1)	0 (0 – 1)
D	1 (0 – 1)	1 (0 – 1)
E	1 (0 – 1)	0 (0 – 1)
J	1 (0 – 1)	0 (0 – 1)
K	1.5 (1 – 2)	1 (0 – 1)
L	1.5 (1 – 2)	1 (0 – 1)
N	1 (0 – 2)	0.5 (0 – 1)
Q	1 (0 – 2.5)	0 (0 – 1)
R	0 (0 – 1)	0 (0 – 1)
S	1 (0 – 2)	0 (0 – 1)
U	0 (0 – 1)	0 (0 – 1)
W	1 (0 – 2)	0 (0 – 1)
Total	1 (0 – 2)	0 (0 – 1)

* Angina, asthma, bronchitis, diabetes, high BP, kidney disease, liver disease, poor circulation

Before surgery, 58.8% of patients rated their general health status as 'Excellent' or 'Very good' (Table 4.40) (varied between centres from 42% to 72%) while 9.3% reported 'Fair' or 'Poor' health (0 – 17.8%).

Table 4.40: Pre-operative general health status

Centre	In general, would you say your health is:				
	Excellent, n (%)	Very good, n (%)	Good, n (%)	Fair, n (%)	Poor, n (%)
A	4 (8.9)	15 (33.3)	18 (40.0)	7 (15.6)	1 (2.2)
C	6 (33.3)	6 (33.3)	5 (27.8)	1 (5.6)	0 (0)
D	8 (19.5)	20 (48.8)	8 (19.5)	5 (12.2)	0 (0)
E	4 (22.2)	9 (50.0)	3 (16.7)	2 (11.1)	0 (0)
J	1 (5.3)	10 (52.6)	6 (31.6)	2 (10.5)	0 (0)
K	0 (0)	5 (50.0)	4 (40.0)	1 (10.0)	0 (0)
L	3 (21.4)	5 (35.7)	5 (35.7)	1 (7.1)	0 (0)
N	4 (14.3)	15 (53.6)	6 (21.4)	3 (10.7)	0 (0)
Q	4 (21.1)	5 (26.3)	8 (42.1)	2 (10.5)	0 (0)
R	2 (11.1)	7 (38.9)	9 (50.0)	0 (0)	0 (0)
S	2 (8.3)	11 (45.8)	9 (37.5)	1 (4.2)	1 (4.2)
U	8 (15.1)	29 (54.7)	13 (24.5)	3 (5.7)	0 (0)
W	3 (7.5)	18 (45.0)	17 (42.5)	2 (5.0)	0 (0)
Total	49 (14.1)	155 (44.7)	111 (32.0)	30 (8.7)	2 (0.6)

Table 4.41 shows the distribution of pre-operative AVVQ and EQ-5D scores by centre. Mean AVVQ scores varied between centres from 14.5 to 24.0 and for EQ-5D scores from 0.64 to 0.81.

Table 4.41: Pre-operative AVVQ and EQ-5D scores

Centre	Pre-operative AVVQ mean (SD)	Pre-operative EQ-5D mean (SD)
A	19.0 (8.8)	0.74 (0.21)
C	14.6 (5.9)	0.77 (0.23)
D	15.6 (7.1)	0.78 (0.17)
E	18.4 (9.0)	0.78 (0.16)
J	15.8 (8.8)	0.78 (0.20)
K	24.0 (7.7)	0.66 (0.34)
L	17.1 (9.3)	0.78 (0.07)
N	14.5 (7.0)	0.81 (0.11)
Q	19.4 (9.8)	0.64 (0.29)
R	15.9 (7.6)	0.73 (0.21)
S	16.4 (8.6)	0.78 (0.17)
U	15.8 (8.4)	0.78 (0.16)
W	16.0 (6.9)	0.77 (0.21)
Total	16.7 (8.2)	0.76 (0.19)

Note: AVVQ and EQ-5D scores was not known for 5 and 6 patients, respectively

4.5.2 Post-operative health

Of the 352 patients recruited, 344 were eligible for follow-up (8 had not undergone surgery). Of these, 259 (75.3%) returned a post-operative questionnaire. Overall, 60.1% rated their general health status as 'Excellent' or 'Very good' (Table 4.42), a similar proportion to that before surgery.

Table 4.42: Post-operative general health status in varicose vein surgery.

Centre	In general, would you say your health is:				
	Excellent, n (%)	Very good, n (%)	Good, n (%)	Fair, n (%)	Poor, n (%)
A	2 (6.7)	8 (26.7)	15 (50.0)	5 (16.7)	0 (0)
C	2 (13.3)	11 (73.3)	2 (13.3)	0 (0)	0 (0)
D	4 (11.4)	13 (37.1)	16 (45.7)	2 (5.7)	0 (0)
E	3 (18.8)	10 (62.5)	2 (12.5)	1 (6.3)	0 (0)
J	0 (0)	6 (54.6)	3 (27.3)	1 (9.1)	1 (9.1)
K	1 (12.5)	3 (37.5)	4 (50.0)	0 (0)	0 (0)
L	2 (16.7)	6 (50.0)	4 (33.3)	0 (0)	0 (0)
N	4 (20.0)	10 (50.0)	6 (30.0)	0 (0)	0 (0)
Q	2 (16.7)	6 (50.0)	2 (16.7)	2 (16.7)	0 (0)
R	3 (25.0)	4 (33.3)	5 (41.7)	0 (0)	0 (0)
S	2 (10.5)	8 (42.1)	8 (42.1)	0 (0)	1 (5.3)
U	4 (11.4)	22 (62.9)	9 (25.7)	0 (0)	0 (0)
W	5 (15.2)	14 (42.4)	13 (39.4)	1 (3.0)	0 (0)
Total	34 (13.2)	121 (46.9)	89 (34.5)	12 (4.7)	2 (0.8)

Table 4.43 shows the distribution of pre-operative and post-operative AVVQ scores by centre. Overall, the mean change in score was 6.6 units (16.6 – 10.0). The extent of improvement varied between centres (3.2 to 16.2).

Table 4.43: Pre- and post-operative AVVQ scores in varicose vein surgery.

Centre	n	Pre-operative AVVQ mean (SD)	Post-operative AVVQ mean (SD)
A	30	20.6 (9.1)	14.0 (11.1)
C	15	14.4 (6.1)	6.3 (5.1)
D	35	16.5 (6.8)	11.1 (8.7)
E	16	17.0 (7.6)	7.8 (5.9)
J	11	15.2 (7.9)	12.0 (10.2)
K	8	24.6 (8.5)	8.4 (5.4)
L	12	15.9 (9.6)	8.7 (11.3)
N	19	15.0 (7.3)	10.4 (8.3)
Q	12	18.1 (8.7)	9.8 (11.3)
R	12	15.0 (8.2)	9.9 (8.4)
S	19	17.4 (9.3)	9.6 (8.5)
U	36	14.9 (6.5)	8.6 (7.9)
W	33	15.5 (7.2)	9.7 (7.7)
Total	258	16.6 (7.9)	10.0 (8.7)

Note: These sample sizes refer to those patients with pre- and post-operative AVVQ scores

The risk adjustment model contained pre-op AVVQ score and sex. These two variables explained 36% of the variation between patients in post-operative AVVQ score.

In the funnel plot shown in Figure 4.13 the hollow circles represent each centre's unadjusted post-operative AVVQ score and the solid circles represent each centre's adjusted score. The horizontal line represents the mean post-operative AVVQ score for all centres combined (10.0). For no centre was the adjusted score statistically significantly different from the expected score at the $p < 0.05$ significance level.

Figure 4.13: Funnel plot of post-operative AVVQ score in varicose vein surgery.

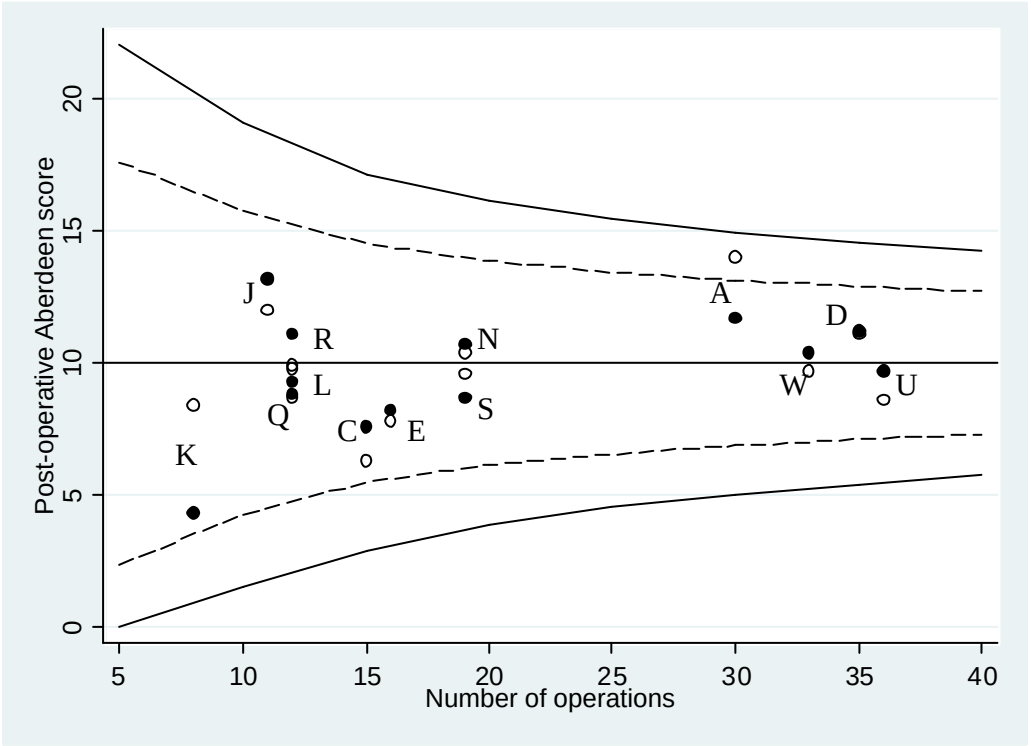


Table 4.44 shows the distribution of pre-operative and post-operative EQ-5D scores by centre. Overall, the mean change in score was 0.10 units (0.77 – 0.87). The extent of improvement varied between centres (0.03 to 0.26).

Table 4.44: Pre- and post-operative EQ-5D scores in varicose vein surgery.

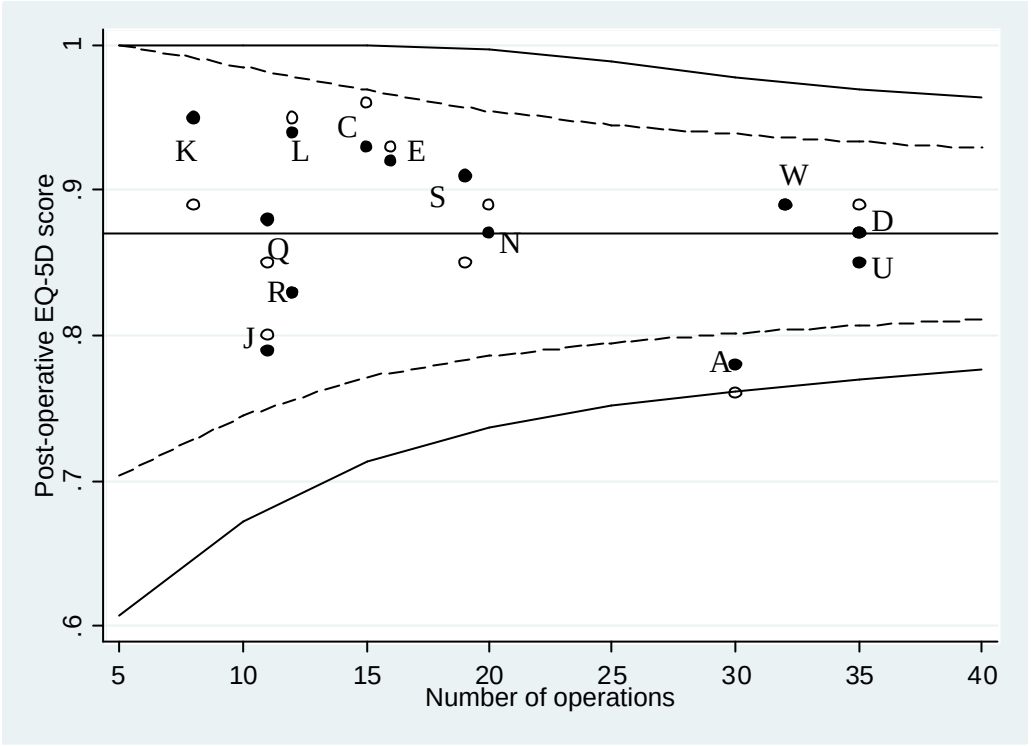
Centre	n	Pre-operative EQ-5D mean (SD)	Post-operative EQ-5D mean (SD)
A	30	0.73 (0.23)	0.76 (0.24)
C	15	0.79 (0.21)	0.96 (0.08)
D	35	0.77 (0.18)	0.87 (0.21)
E	16	0.78 (0.17)	0.93 (0.10)
J	11	0.73 (0.23)	0.80 (0.15)
K	8	0.63 (0.37)	0.89 (0.16)
L	12	0.78 (0.08)	0.95 (0.09)
N	20	0.81 (0.12)	0.90 (0.11)
Q	11	0.67 (0.32)	0.85 (0.30)
R	12	0.76 (0.22)	0.84 (0.23)
S	19	0.79 (0.19)	0.86 (0.28)
U	35	0.82 (0.11)	0.89 (0.20)
W	32	0.78 (0.19)	0.90 (0.13)
Total	256	0.77 (0.20)	0.87 (0.19)

Note: These sample sizes refer to those patients with pre- and post-operative EQ-5D scores

The risk adjustment model contained pre-op EQ-5D score, AVVQ score, general health status, age and comorbidity. These five variables explained 29% of the variation between patients in post-operative EQ-5D scores.

In the funnel plot shown in Figure 4.14 the hollow circles represent each centre's unadjusted post-operative EQ-5D score and the solid circles represent each centre's adjusted post-operative EQ-5D score. The horizontal line represents the mean post-operative EQ-5D score for all centres combined (0.87). For Centre A the adjusted score is significantly worse than expected at the $p < 0.05$ significance level but not at the $p<0.002$ level.

Figure 4.14: Funnel plot of post-operative EQ-5D scores in varicose vein surgery.



Overall, 40.9% of patients had post-operative complications. The percentage varied between centres from 20.0% to 75.0% (Table 4.45).

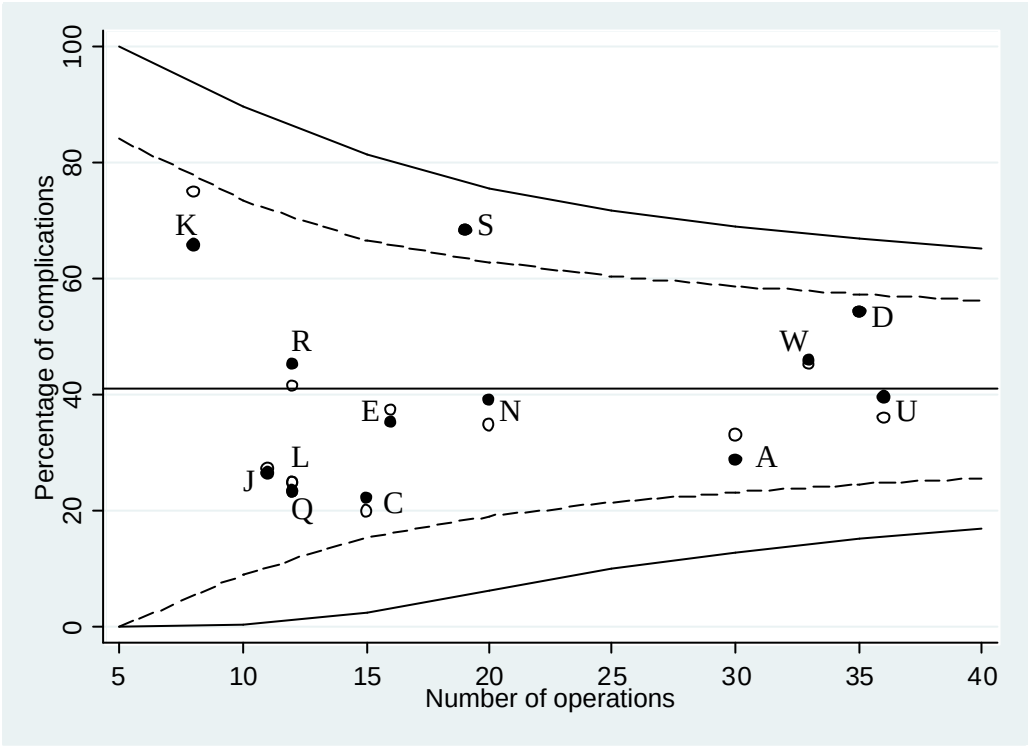
Table 4.45: Incidence of post-operative complications in varicose vein surgery.

Centre	Allergy or reaction to drug, n (%)	Urinary problems, n (%)	Bleeding, n (%)	Wound problems, n (%)	Any complication, n (%)
A	0 (0)	1 (3.3)	5 (16.7)	9 (30.0)	10 (33.3)
C	0 (0)	1 (6.7)	2 (13.3)	1 (6.7)	3 (20.0)
D	4 (11.4)	1 (2.9)	12 (34.3)	15 (42.9)	19 (54.3)
E	2 (12.5)	2 (12.5)	2 (12.5)	3 (18.8)	6 (37.5)
J	0 (0)	1 (9.1)	2 (18.2)	0 (0)	3 (27.3)
K	0 (0)	0 (0)	2 (25.0)	5 (62.5)	6 (75.0)
L	0 (0)	0 (0)	2 (16.7)	2 (16.7)	3 (25.0)
N	3 (15.0)	1 (5.0)	2 (10.0)	1 (5.0)	7 (35.0)
Q	0 (0)	0 (0)	3 (25.0)	2 (16.7)	3 (25.0)
R	0 (0)	0 (0)	4 (33.3)	3 (25.0)	5 (41.7)
S	0 (0)	0 (0)	10 (52.6)	6 (31.6)	13 (68.4)
U	0 (0)	1 (2.8)	8 (22.2)	9 (25.0)	13 (36.1)
W	0 (0)	0 (0)	10 (30.3)	12 (36.4)	15 (45.5)
Total	9 (3.5)	8 (3.1)	64 (24.7)	68 (26.3)	106 (40.9)

The risk adjustment model contained pre-op AVVQ score, sex and comorbidity. The model discriminated between those patients who did and who did not have complications only moderately (area under ROC = 0.63).

In the funnel plot shown in Figure 4.15 the hollow circles represent each centre's unadjusted percentage of complications and the solid circles represent each centre's adjusted percentage of complications. The horizontal line represents the overall percentage of complications for all centres combined (40.9%). The adjusted incidence of complications in Centre S is significantly higher than expected at the $p < 0.05$ significance level but not at the $p<0.002$ level.

Figure 4.15: Funnel plot of post-operative complications in varicose vein surgery.



The majority of patients reported that they were better following surgery (90.3%) with some variation between centres (range 72.4 -100%) (Table 4.46).

Table 4.46: Estimation of change in functioning following varicose vein surgery.

Centre	Overall, how are you now compared to before your operation?				
	Much better, n (%)	A little better, n (%)	About the same, n (%)	A little worse, n (%)	Much worse, n (%)
A	13 (44.8)	8 (27.6)	7 (24.1)	1 (3.5)	0 (0)
C	12 (80.0)	2 (13.3)	1 (6.7)	0 (0)	0 (0)
D	22 (64.7)	10 (29.4)	0 (0)	0 (0)	2 (5.9)
E	14 (87.5)	2 (12.5)	0 (0)	0 (0)	0 (0)
J	8 (72.7)	3 (27.3)	0 (0)	0 (0)	0 (0)
K	6 (75.0)	0 (0)	1 (12.5)	1 (12.5)	0 (0)
L	11 (91.7)	1 (8.3)	0 (0)	0 (0)	0 (0)
N	14 (70.0)	4 (20.0)	1 (5.0)	0 (0)	1 (5.0)
Q	8 (66.7)	4 (33.3)	0 (0)	0 (0)	0 (0)
R	9 (75.0)	2 (16.7)	1 (8.3)	0 (0)	0 (0)
S	19 (100)	0 (0)	0 (0)	0 (0)	0 (0)
U	25 (71.4)	4 (11.4)	3 (8.6)	3 (8.6)	0 (0)
W	22 (66.7)	8 (24.2)	2 (6.1)	1 (3.0)	0 (0)
Total	183 (71.5)	48 (18.8)	16 (6.3)	6 (2.3)	3 (1.2)

4.5.3 Summary and discussion

- Case-mix varied considerably between centres in terms of the proportion of women and the duration of symptom history. Aside from their primary condition, patients' general pre-op health was good.
- The pre-operative severity was slightly less than that previously reported (16.6 versus 20.2) (Critchley et al 1997) though the extent of improvement following surgery was similar.
- Given this, surgery had little impact on their general health. However, they reported considerable improvement as measured by the disease-specific instrument (AVVQ) and, surprisingly, the generic instrument (EQ-5D).
- None of the centres' performance differed significantly from that expected as determined by the AVVQ. However, one centre performed worse than expected according to the EQ-5D.
- The incidence of complications (bleeding, wound problems) varied considerably with one centre being significantly worse than expected. The incidence of wound complications (26.3%) was much higher than that reported (4%) in a previous study (Critchley et al 1997) though the latter was based on surgeons' rather than patients' reports.

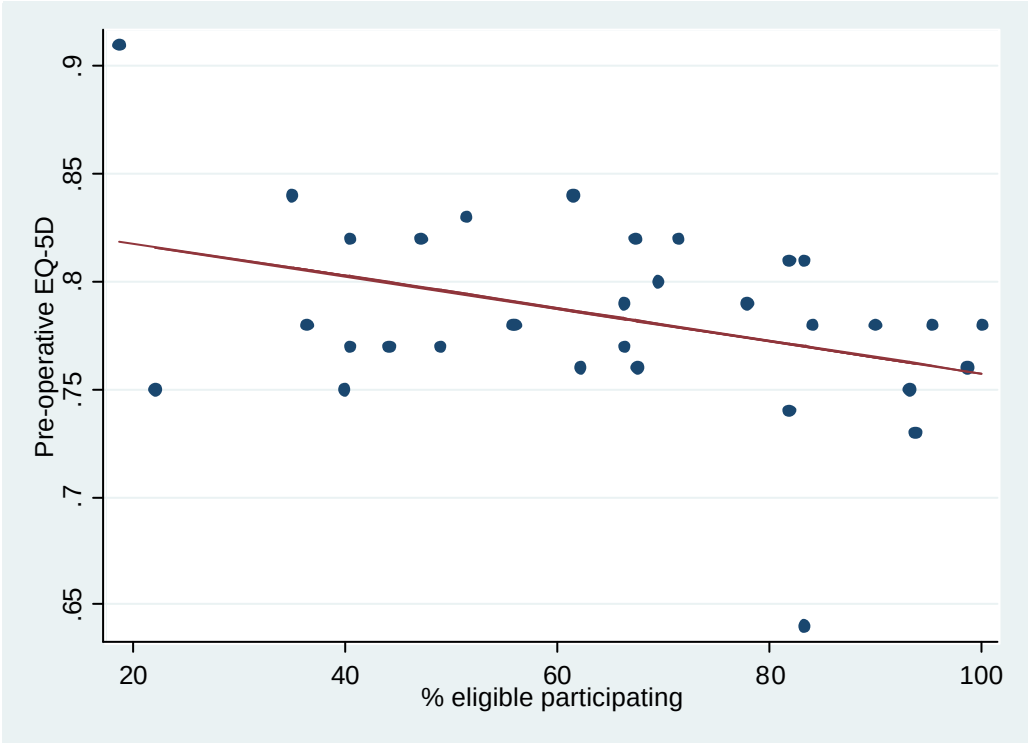
5. RESULTS: Methodological considerations

5.1 What is an adequate recruitment rate?

There has been little research assessing recruitment bias. Most research studies and audits fail to report the proportion of eligible patients invited to participate, let alone assess the impact that failure to recruit might have had on the results. In this project, the overall recruitment rate was 60.3% (see section 3.5), ranging from 18.8% to 100% for centre/procedures. To determine whether failure to recruit introduced any bias, we compared the mean pre-operative health status of patients (measured by EQ-5D) with the proportion participating. We did this separately for the orthopaedic procedures from the other three procedures as the former patients had a much lower mean EQ-5D score (about 0.35) than the latter (about 0.8).

Figure 5.1 shows the relationship for patients undergoing cataract, hernia or vein surgery. Each dot represents the mean pre-op EQ-5D score for a centre. The superimposed regression line indicates that an increase of 20% of those eligible participating is associated with a decrease in pre-operative mean EQ-5D scores of 0.015 units (the Pearson correlation coefficient is –0.38). There is strong evidence that this association cannot be attributed to chance (p = 0.03). Adjusting for surgical procedure does not alter the association or the p value.

Figure 5.1: Mean pre-operative EQ-5D score for patients undergoing cataract, hernia or vein surgery by the recruitment proportion.



Although higher recruitment was associated with a lower mean EQ-5D score in orthopaedic procedures, it was not as marked (Pearson correlation coefficient -0.14) and was not statistically significant. This may partly be due

to less data (n=16 compared with n=32) and to a floor effect (patients for hip and knee surgery report very poor pre-operative general health).

These findings demonstrate the potential bias that low recruitment could introduce when comparing centres. There is a visual suggestion in Figure 5.1 that the pre-operative EQ-5D score is unchanged above 80% recruitment. Based on these data we suggest a recruitment rate of 80% should be aimed for.

5.2 What is an adequate post-op response rate?

The post-operative response rates were excellent (see section 3.7) with cataract, knee and hip patients achieving about 90% response and hernia and vein patients about 75-80%. However, even at these high rates there were differences as regards two pre-operative patient characteristics - age and general health. The extent of the differences was not related to the response rates.

For all five procedures, non-responders were younger than responders (Table 5.1). While the difference was slight for cataract and hip surgery, the mean age differed by almost 15 years for hernia repair.

Table 5.1: Mean age (SD) of responders and non-responders to post-operative questionnaire.

Procedure	Responders' mean age (SD)	Non responders' mean age (SD)
Cataract	74.4 (9.7)	71.9 (12.9)
Hernia	57.0 (14.6)	42.4 (14.9)
Veins	48.7 (12.9)	40.0 (9.5)
Knee	69.4 (9.1)	60.0 (12.6)
Hip	68.2 (9.9)	64.9 (9.6)

The only other difference was that non-responders tended to have poorer general health beforehand (Table 5.2). This was reflected in the patients' EQ-5D scores for cataract (0.81 v 0.72) and knee (0.39 v 0.32) surgery but was not as marked for the other three procedures.

Table 5.2: Proportion of responders and non-responders reporting 'Fair' or 'Poor' health before surgery.

Procedure	Responders (%)	Non-responders (%)
Cataract	19.5	32.2
Hernia	8.2	10.6
Vein	7.1	12.0
Knee	19.0	29.4
Hip	18.8	37.0

There was little difference between responders and non-responders in terms of sex, duration of symptoms, previous surgery, comorbidity and disease-specific PROM. In view of the existence of the observed differences even with high response rates, it is clear that rates should be as high as possible. Experience from this project suggest rates of about 80% are attainable for minor surgery and about 90% for major surgery.

5.3 What is the impact of imputing missing PROM items?

Inevitably, some patients failed to answer every item (question) contained within the PROMs in the pre- and post-operative questionnaires. Despite the proportion of missing items being very low, if every patient with even one missing item in a PROM were to be excluded from the analyses, a significant proportion of patients would be lost to the study (Table 5.3), thus reducing the power of the comparisons.

Table 5.3: Proportion of PROMs with at least one missing item before and after surgery.

PROM	Pre-operative	Post-operative
VF-14	17.7	10.2
SF36-PCS	13.4	10.5
AVVQ	8.3	4.8
OKS	4.8	7.8
OHS	5.6	7.7

The proportion of EQ-5D responses with missing items was less. To overcome this problem, missing data values were imputed, using the SF-36 rule. The impact of imputation was examined by comparing the mean scores with and without imputation. Table 5.4 shows what little impact imputation has on the mean score. However, imputation prevents a significant loss of statistical power through a diminution in the sample sizes.

Table 5.4: Impact of imputation of missing PROM items on the mean (SD) scores

PROM	Mean (SD) no. of missing items	Mean (SD) with imputation*	Mean (SD) no imputation**
Cataract surgery			
pre-op VF-14	0.4 (1.2)	82.5 (17.6)	82.2 (18.1)
post-op VF-14	0.2 (0.8)	92.8 (14.4)	93.2 (14.1)
pre-op EQ-5D	0.2 (0.7)	0.80 (0.24)	0.80 (0.23)
post-op EQ-5D	0.1 (0.5)	0.78 (0.28)	0.78 (0.28)
Hernia repair			
pre-op SF-36 PCS	1.0 (3.9)	47.2 (9.3)	47.5 (9.2)
post-op SF-36 PCS	0.4 (2.3)	50.2 (9.6)	50.4 (9.5)
pre-op EQ-5D	0.2 (0.9)	0.77 (0.19)	0.78 (0.19)
post-op EQ-5D	0.1 (0.8)	0.85 (0.20)	0.85 (0.20)
Varicose veins			
pre-op AVVQ	0.4 (1.6)	16.6 (8.2)	16.4 (8.1)
post-op AVVQ	0.1 (0.4)	10.0 (8.7)	9.9 (8.7)
pre-op EQ-5D	0.1 (0.6)	0.76 (0.19)	0.76 (0.19)
post-op EQ-5D	0.01 (0.1)	0.87 (0.19)	0.87 (0.19)
Knee replacement			
pre-op OKS	0.1 (0.5)	41.9 (7.6)	42.0 (7.4)
post-op OKS	0.2 (0.9)	26.9 (9.3)	26.9 (9.1)
pre-op EQ-5D	0.04 (0.3)	0.38 (0.32)	0.38 (0.32)
post-op EQ-5D	0.1 (0.5)	0.70 (0.25)	0.70 (0.25)
Hip replacement			
pre-op OHS	0.1 (0.5)	43.3 (8.0)	43.4 (8.0)
post-op OHS	0.2 (0.9)	23.2 (9.6)	22.9 (9.7)
pre-op EQ-5D	0.1 (0.5)	0.32 (0.32)	0.32 (0.32)
post-op EQ-5D	0.1 (0.5)	0.74 (0.25)	0.74 (0.25)

5.4 Does clustering affect our estimates of centre performance?

Multilevel regression models (patients nested within centres) for the PROM and post-operative complication outcomes, and for each surgical procedure, indicated that the variance between centres was small and not statistically different from zero. This indicates for these data clustering is not something that needs to be taken into account in the analyses. With data from a larger number of centres, this should be checked for again.

5.5 Is risk adjustment adequate to identify poor performers?

The risk-adjusted regression models accounted for between 22% and 47% of the total variation in PROM outcomes. Although these results are reasonable it does not rule out the possibility that unmeasured risk factors bias the associations between centre and outcome.

The risk-adjusted logistic regression models for complication outcomes had area under ROC values ranging from 0.57 to 0.63. These results indicate only slight discrimination such that the risk-adjustment for this outcome is less reliable than for the PROM outcomes. It is possible that unmeasured risk factors bias the associations between centre and outcome.

The capacity to develop robust risk adjustment models was limited by the relatively small numbers of patients and the absence of clinical data.

5.6 Which generic measure: EQ-5D or SF-6D?

In Table 5.5 the relative responsiveness of the two generic measures was compared for hernia repair and hip replacement. There wasn't much difference: the EQ-5D was more responsive than the SF-6D for hernia and less responsive for hip replacement. Given the greater cost of the SF-6D both in terms of respondents' time (36 items compared with 5 items) and in terms of administration (through having to pay a licence fee), we do not report further on the SF-6D.

Table 5.5: Relative responsiveness of EQ-5D and SF-6D

PROM/surgical procedure	Effect size	Classification of responsiveness
SF-6D (hernia)	0.31	Small
EQ-5D (hernia)	0.39	Small
SF-6D (hip)	1.4	Large
EQ-5D (hip)	1.3	Large

5.7 What is a minimally important difference?

Table 5.6 shows the distribution and anchor-based MID values derived for each PROM and surgical procedure. The anchor-based values are derived from those patients who reported their symptoms were "a little better" after surgery. For cataract, hernia and varicose vein surgery the anchor-based values for generic and disease-specific PROMs (apart from the AVVQ) are lower than the distribution based value, whereas the opposite is true for hip and knee replacement.

Table 5.6: Distribution and anchor-based MID values for different PROMs and surgical procedures

PROM (surgical procedure)	Distribution-based method					Anchor-based method	
	N	SD of pre-op scores	Mean change score	Reliability coefficient	SEM	N	Mean (SD) change score in reference group
VF-14	582	17.4	9.8	0.79	8.0	68	5.4 (18.3)
EQ-5D (cataract)	571	0.23	-0.03	0.84	0.09	68	0.02 (0.29)
SF-36 PCS (hernia)	405	9.0	2.6	0.89	2.9	36	0.70 (6.5)
EQ-5D (hernia)	430	0.18	0.07	0.84	0.07	38	0.02 (0.11)
Oxford Hip Score	349	7.9	20.0	0.88	2.7	26	13.1 (6.8)
EQ-5D (hip)	346	0.31	0.42	0.84	0.12	26	0.32 (0.25)
Oxford Knee Score	347	7.6	14.8	0.92	2.1	55	9.7 (2.1)
EQ-5D (knee)	345	0.31	0.31	0.84	0.12	55	0.25 (0.29)
AVVQ	268	7.9	6.5	0.79	3.6	53	5.0 (9.0)
EQ-5D (varicose)	266	0.19	0.10	0.84	0.08	52	0.06 (0.22)

The standard deviations associated with the anchor-based MID values are large, indicating a large amount of variation within the reference group patients with respect to what constitutes an MID. The analyses shown in Table 5.7 confirm this observation and indicate that for all surgical procedures the anchor-based values derived vary according to pre-operative severity: higher values are seen in patients with higher pre-operative severity and lower values in patients with lower severity. As a consequence, Pearson correlations between pre-op scores and change scores in patients who reported their symptoms were "a little better" after surgery (the reference group) show a clear association. These findings demonstrate that the anchor-based MID values are difficult to interpret. The sample size calculations presented below were therefore only based on distribution-based MID values.

Table 5.7: Stability of anchor-based MID values across the severity of pre-op scores in patients who reported their symptoms were "a little better" after surgery

PROM/surgical procedure	Correlation between pre-op score and change score	Mean MID value for higher pre-op severity patients		Mean MID value in lower pre-op severity patients	
		N	Mean	N	Mean
VF-14	-0.51	34	13.1	34	-2.3
EQ-5D (cataract)	-0.32	33	0.003	35	-0.05
SF-36 PCS (hernia)	-0.20	18	2.5	18	-1.1
EQ-5D (hernia)	-0.40	19	0.05	19	-0.006
Oxford Hip Score	*0.19	12	13.6	12	11.5
EQ-5D (hip)	-0.61	13	0.41	13	0.24
Oxford Knee Score	*0.54	28	11.9	27	7.4
EQ-5D (knee)	-0.65	28	0.43	27	0.07
AVVQ	*0.38	27	7.3	26	2.6
EQ-5D (varicose)	-0.75	22	0.15	30	-0.01

*these positive correlations are due to the scoring direction for the measures involved.

5.8 What is an adequate sample size?

The distribution-based method can be used to define a general rule as illustrated in the following hypothetical but typical example. Given a PROM with a population mean of 50, standard deviation of 10, range of 0 to 100 and ICC of 0.85, using 80% power, 53 patients per centre would be required if a statistical significance level of 0.05 is used (equivalent to 95% confidence intervals), and 104 patients if significance level of 0.002 is used (equivalent to 99.8% confidence intervals). Using 95% power the respective sample sizes would be 87 and 150. These calculations assume that the results of each centre are to be compared with a single reference value (e.g. the mean post-op score for all centres).

Table 5.8 shows the sample sizes required based upon the distribution-based MID values that are presented in Table 5.6. The centre-level sample sizes required reflect both the high power associated with continuous data and the high MID values used in the calculations. If smaller MID values were used the sample sizes required to detect such a difference would increase commensurately. For example if we again consider the hypothetical example of a PROM with a population mean of 50, standard deviation of 10, range of 0 to 100 and ICC of 0.85, to detect a difference of 0.5 SEM rather than 1 SEM the following sample sizes per centre would be needed: 210 (0.05 significance level and 80% power), 413 (0.002 significance level and 80% power), 347 (0.05 significance level and 95% power) and 599 (0.002 significance level of and 95% power).

Table 5.8: Number of patients required to detect whether centre value differs from the national average

PROM (surgical procedure)	Distribution-based			
	0.05 significance level		0.002 significance level	
	80% power	95% power	80% power	95% power
VF-14	38	62	74	107
EQ-5D (cataract)	52	85	101	147
SF-36 PCS (hernia)	76	126	149	216
EQ-5D (hernia)	52	86	103	149
Oxford Hip Score	68	112	133	192
EQ-5D (hip)	53	87	104	150
Oxford Knee Score	103	171	203	294
EQ-5D (knee)	53	87	104	150
AVVQ	38	63	75	108
EQ-5D (varicose)	45	74	88	127

5.9 Is the responsiveness of the PROMs adequate?

Table 5.9 is based on the pre-op standard deviations and mean change scores shown in Table 5.6. It demonstrates that the effect sizes produced by disease-specific measures are always higher than those for the generic measure (apart from for hernia repair).

Table 5.9: Relative responsiveness of different PROMs

PROM (surgical procedure)	Effect size	Classification of responsiveness
VF-14	0.56	Moderate
EQ-5D (cataract)	0.13	Not responsive
SF-36 PCS (hernia)	0.29	Small
EQ-5D (hernia)	0.39	Small
Oxford Hip Score	2.5	Large
EQ-5D (hip)	1.3	Large
Oxford Knee Score	1.9	Large
EQ-5D (knee)	1.0	Large
AVVQ	0.82	Large
EQ-5D (varicose)	0.53	Moderate

5.10 Can disease-specific measures be mapped to generic measures?

Table 5.10 shows the changes in the EQ-5D which equate to a 1 point change in different disease-specific measures. These figures are based on simple linear regression models which show the variance explained by each model. (For the purpose of these analyses the SF-36 PCS in hernia surgery is treated as a disease-specific measure). The coefficients shown in the table represent the amount of change in the EQ-5D equivalent to a 1 point change in the disease-specific measure.

Table 5.10: Mapping disease-specific measures to the EQ-5D

Disease-specific measure	Pearson correlation with EQ-5D	Linear regression coefficient	95% confidence intervals around coefficient	Variance explained by model
VF-14	0.17	0.002	0.001-0.003	3%
SF-36 PCS (hernia)	0.44	0.01	0.008-0.12	20%
Oxford Hip Score	0.65	0.02	0.019-0.024	42%
Oxford Knee Score	0.62	0.02	0.018-0.023	38%
AVVQ	0.23	0.006	0.003-0.009	5%

In cataract and varicose vein surgery there is a weak association between changes in the disease-specific measures and the EQ-5D. In hernia repair the association between SF-36 PCS change and EQ-5D change is moderate (20% variance explained) and indicates that a 1 point change in the PCS score equates to a 0.01 point change in the EQ-5D score. In hip and knee replacement surgery the association between the Oxford Hip and Knee Scores and the EQ-5D is quite strong. For both Oxford measures a 1 point change is equivalent to a 0.02 point change in the EQ-5D.

6. RESULTS: Summarising and presenting findings

Of the 32 people invited to participate in the two focus groups, six were unable to attend on the day (two surgeons, one nurse, two provider managers, one commissioner). Before considering the specific issues the groups were asked to address, there are four general observations regarding the method and the validity of the results:

First, we experienced considerable difficulty in finding commissioners from PCTs prepared to participate, and of those who agreed only one attended. Two other commissioners did participate but they were from the Department of Health and a general practice involved in Practice-based Commissioning. We are therefore cautious as to the representativeness of commissioners' views.

Second, while surgeons (and to a lesser extent, provider managers) accepted the need for and desirability of performance measures and comparisons being made available privately, they were anxious about public dissemination, fearing that such data would be used in a 'misleading' way by journalists and the mass media. This reluctance about the whole enterprise coloured their views.

Third, although we were seeking participants' views based on their own perspective, there was a tendency for people to pronounce on behalf of other groups, presuming to know the opinion of those groups and despite representatives of those groups being present. Most striking was surgeons' tendency to present patients' views:

> *If I was a manager or a patient, I would be interested in re-admission rates, length of stay in a hospital after hip or knee replacements to give an indication of quality of the pathway and complications. If I was a patient with insight, that is what I would want to know. (Surgeon)*

At times this extended to offering contradictory views to those that had just been expressed by patients themselves:

> *Patient: I think the average lay person would find it easier to cope with a histogram. Star ratings do not give as much information as they would really like… Histograms seem to be a good in-between.*
>
> *Surgeon: It depends who you are reading the data. If you are a hospital manager or a surgeon you would want more information and breakdown of the figures but if you are reading the Sunday Times in your armchair a star rating for a patient may be all they want.*

Offering the views of others was not confined to surgeons:

> *if I was running a hospital, I would want to know and engage with clinicians if there was a disparity in performance. (Commissioner)*

These comments reflected a prevailing view that the main reason for collecting data on performance was to inform patients in choosing a provider. There was little recognition of the need for such data for clinicians and managers to improve the quality of care. We have distinguished between participants' own views and their 'projected' views in the analysis.

And fourth, despite five experienced, senior surgical nurses attending the meetings, they did not offer any distinctive views, independent of those expressed by the surgeons. This may reflect concordance between the two clinician groups or a reluctance to express divergent views in the presence of their medical colleagues.

We present the results of the discussions of the five principal questions which the groups were asked to consider.

6.1 What is the best way of presenting data?

Simple categories ('star ratings')

Patients favoured this means of presentation as it is 'understandable' and 'differences between providers matter'. They felt more than three categories might be needed.

> *I like the star ratings because they are simple…I know people that could not spell histogram. They are simple – above, average or below average. (Patient)*

Commissioners were divided, some thought the simplicity of categorisation was useful for making choices as long as it was made clear what the indicator was measuring.

> *Star ratings are a description of performance. From the purchasers' point of view, it is extremely important and even more from the providers' point of view. Why are they not doing well and how can they improve themselves? (Commissioner)*

> *You need to be very clear about the criteria you are using in this example so that people do not misappropriate the basis of the star and the key is to be clear about what it is. (Commissioner)*

Surgeons, nurses and provider managers were firmly opposed to this approach. Apart from the 'bad reputation' that the DoH star ratings had deservedly attracted, they felt categorisation: told people little; was associated with government oversight; can be manipulated by the mass media; were too labile (i.e. loss of one star suggested a major decline in performance); and threatened the viability of providers in a market driven system.

> *It is dangerous. If neighbouring hospital has better star than you, commissioners decide where work is going to go. A hospital may not be failing but will get less work to do. May go down the pan. (Surgeon)*

Bar charts

Patients thought bar charts showing the ranking of providers' performance was legitimate and helpful as (unlike simple categories) it shows the extent of any differences between providers. The preference was for a vertical rather than a horizontal display, with the inclusion of confidence intervals. Overall it was a good compromise between categories (insufficient detail) and tables (too much information).

> *Histogram is a good idea – it gives it more statistical appearance without being too confusing. Shows quite clearly that differences between providers are not great. More helpful that the rather perceptually large differences between one group of stars and another which may not be helpful. (Patient)*

Surgeons preferred bar charts to categories but felt it was not legitimate (or helpful) to rank providers by performance and would prove to be divisive.

> *the histogram where they are not ranked emphasises the variability to people. For people who do not understand, this may be divisive – better left without ranking. (Surgeon)*

This view was shared by some commissioners though others felt ranking was needed.

> *All things being equal. We have to make a decision on a daily basis. The ways it is presented in histogram: great. That is the way. (Commissioner)*

Tables

Surgeons favoured tables as they provide 'raw' data, including information on the case-mix of patients and the volume of patients treated. Commissioners also valued having numerical values.

> *Those sort of raw data are therefore very important. For some groups they would want that information. It is more hard hitting than star ratings, the actual numbers. (Surgeon)*

Patients found tables too complex and detailed for their needs.

Funnel plots

In response to the views expressed at the first focus group, the second group was also offered funnel plots as a way of presenting comparative data. Commissioners and provider managers were both enthusiastic though surgeons and patients were less enamoured. The inclusion both of the unadjusted and the adjusted measure of performance was favoured.

> *Funnel plot is excellent. Great. (Commissioner)*

> *If you have not got much time, then funnel plots are excellent. (Manager)*

There was no consensus as to the control limits that should be applied in judging performance. Not surprisingly, surgeons and provider managers favoured 99.8% which would minimise the likelihood of any providers performance being viewed as unacceptable. In contrast, commissioners and patients felt 95% was appropriate so as to ensure under-performers were detected (even at the increased risk of falsely labelling some providers).

6.2 What is the best indicator for comparing centres?

Change in the mean PROM score

Some surgeons questioned the validity of a measure that was derived from two measures (pre-op and post-op PROMs), the validity of which they were not convinced.

> *Problem is it's your decision about what they have written - not what the patients have told you is in the scores… So it is not fair to take your assumption on it. (Surgeon)*

However, others in the groups did not express any reservations about the use of this indicator.

Proportion achieving a 'minimally important difference (MID)'

Patients, commissioners and some surgeons favoured use of the proportion achieving a MID as it gives patients contemplating surgery an indication of their likelihood of gaining a useful benefit. This also enhances informed, shared decision-making and helps patients have a realistic expectation of outcomes.

> *I think it challenges you as a patient to say I thought this cataract operation was successful, and now you are saying that only 68% can expect to get improvement…Might make me ask about having the operation and taking the risk. We ought to be trying to move to give patients details about relative risk. (Commissioner)*

> *Very valid point about raising expectation – saying yes we can do this but the outcome is not guaranteed. Interesting step forward. There is an expectation that if you are going in for surgery you are going to be better. (Patient)*

Other surgeons were concerned about the basis of defining MIDs and felt it should only be used if there was a consensus as to what constitutes a MID.

Change in patients' rating of their overall health

While some surgeons felt this to be a useful additional measure, there was no great enthusiasm for it from other members of the groups.

> *You would get a better set of numbers if you said to the patients "are you better from surgery?"'… isn't the point to make people happy. (Surgeon)*

Proportion reporting a post-operative complication

Surgeons, patients and commissioners felt this was a valuable and necessary measure to include. For the surgical operations being considered, it was felt an indication of the overall proportion of patients reporting complications was sufficient, rather than information on specific complications. The need for additional data on longer term complications was also recognised.

> *Every patient should know in advance of an operation that there is a 75% chance of a significant clinical improvement but you should also know and be able to discuss what the possibilities are as far as post op complications. Every patient should have access to that. (Patient)*

> *From the patient and surgical view complication rates are very important. This only concentrates on immediate post op complications. Certainly later complications such as a high dislocation rate. That is very important. (Surgeon)*

6.3 At what geographical level should centres be compared?

The groups were asked to consider which other providers they would want any particular provider compared with: other 'local' ones; those in the same 'region' of the country; all those in England.

Patients wanted to see comparisons with other providers in the same region as these were the other services they could choose from (assuming they didn't want to travel longer distances).

> *I think a patient is interested in the region and particularly in the Choose and Book system. They can look beyond the normal. A region is easy to grasp. (Patient)*

In contrast, surgeons wanted to be compared nationally with all providers.

> *As a surgeon you need to know about the country – how you compare across the country. (Surgeon)*

Commissioners wanted local comparisons (reflecting current NHS policy on offering a limited choice to patients) together with a regional and a national average as benchmarks.

> *It would be great to see your local ones as a whole region and compare it to an average. You know your patients are not going to go to Leeds from Hackney. But you can say what can we do to get up to that level? (Commissioner)*

6.4 At what level of provision should comparisons be made?

Operating surgeon

There was support from patients (who perceived they were the principal determinant of performance) and from provider managers (who wanted this level of assessment to enable them to detect inadequate performance).

I would like to know who the operating surgeon is and what their results are. I would like to know at least who is running a team. (Patient)

Operating surgeon level - yes. We need to see the surgeon and monitor our surgeons and audit…The operating surgeon has responsibility – he sees the patients clinically reads the notes and makes the judgement. It is his call. (Manager)

However, surgeons were less supportive, feeling this level was inappropriateas the responsibility for performance rested with the consultant. Commissioners also were unsupportive as they contracted with organisations not individual surgeons.

Would it be reasonable to make that information public? Is it a right that it should be? In the aircraft industry, it has made strides to be more reliable. You do not ask who the pilot is. They have moved the system of reliability in the industry…Whether it is reasonable to share that information because it is very complex I am not so sure. (Commissioner)

Consultant surgeon

Patients and provider managers wanted comparisons at this level (in addition to operating surgeon) as did commissioners.

From someone looking at the data from running the hospital and working with the consultants, it would be meaningless unless you had it at consultant level. (Manager)

Although there was greater justification than comparing operating surgeons, clinicians were not enthusiastic about comparisons at this level.

Provider (hospital/treatment centre/day surgery unit)

This was seen to be the appropriate level by surgeons, patients and commissioners.

From a practical point of view, people are referred to a provider – even if a patient is referred to a consultant for cataract, most hospitals work with a common waiting list- so the patient does not have the choice of nominating their operating surgeon unless they go privately. So it would have to be at the provider level. (Surgeon)

I think with a purchasing hat on you would want provider level. (Commissioner)

6.5 How frequently should performance information be updated?

The focus groups were invited to comment on how up-to-date performance information needed to be, bearing in mind the post-operative data would be collected 3-6 months after surgery. There was a consensus that, if the surgical technology was not changing and there wasn't a fast turnover of surgeons, the information need not be updated more frequently than every 6 – 12 months.

> *That would be ideal. Surgeons may have changed in that time. You would be suspicious of old data. Last six months. (Patient)*

> *Annual report… is all that is required. 2005 acceptable. 2004 would be going back a bit. (Patient)*

Surgeons and provider managers would also like to have systems that would allow them to monitor performance continuously using mechanisms such as CUSUM plots.

7. RECOMMENDATIONS

7.1 Feasibility

Where and when should patients be recruited?

According to both the preferences of patients and the recruitment rates achieved, there is little difference between pre-op assessment clinics and on admission. However, there are some disadvantages to using the pre-op assessment clinic.

At the time of attending a pre-op assessment clinic, a patient's date for surgery has usually not been fixed. In some cases there may be a considerable length of time between the pre-op assessment appointment and the date of surgery. This creates a number of problems:

- The more time between completion of the pre-op questionnaire and date of surgery, the more 'redundant' are the responses on the questionnaire.
- If surgery is cancelled, the patient will have completed an unnecessary questionnaire.
- Without knowing the exact date of surgery, or whether the surgery was ever performed, it is not possible to know when a post-operative questionnaire should be sent. Given that the follow-up period should be the same for all patients undergoing a particular procedure, this is a significant problem.
- If surgery is performed at a different centre then inaccurate estimates of each centre's performance will arise.

One solution to these problems might be for participating centres to regularly inform a co-ordinating centre of the surgery dates for all patients who complete pre-operative questionnaires. This would incur additional costs and extra work for local data collection staff. It would also not solve the problem of cancelled operations.

There are other problems associated with recruiting patients at pre-op assessment clinics:

- Patients receive a large amount of information both written and verbal at these clinics and are often asked to complete other questionnaires and consent forms. This leads to high burden and may be confusing to the patient.
- A firm decision on the need for surgery is usually only taken after all test and assessment results are known. The patient may believe that the response they give to a questionnaire about their health and quality of life may influence the decision, invalidating their responses.
- Pre-op assessment is done by telephone for certain procedures at certain centres and assessment by e-mail/internet is also possible. This makes patient recruitment impossible.
- Patients often attend a pre-op assessment clinic with a relative, carer or partner who may 'help' complete the patient's questionnaire.

In contrast to the pre-op assessment clinic, the identification and recruitment of patients during the surgical admission period is relatively straightforward. The main problem is the time pressure surrounding the surgical episode, particular as a large number of patients are booked each day and a large number of tasks have to be

completed for each patient. In particular, the first patient of the day has to be processed quickly. An obvious solution to this problem is to ask the first patient to arrive slightly earlier than is normal current practice.

A separate logistical problem is the relatively high number of nursing and administrative staff involved in day surgery provision. Unlike the pre-op assessment clinic which involves a small number of staff, some day surgery units employ a high number of part-time staff. This may make it difficult to ensure all staff recognise their responsibility to identify appropriate patients. Again, this problem is not insurmountable, but does require a commitment among staff to organise and adhere to a particular protocol.

Another argument against recruitment on admission is the heightened levels of anxiety and stress that the patient may experience on the day of surgery. However, we found little or no evidence that this was a significant problem. If anything, the reverse was true and patients were grateful to have something to occupy their time while waiting.

Finally, we found that in centres that recruited on admission, those patients that were invited to participate were more likely to agree.

Given all the disadvantages of trying to recruit staff in pre-assessment clinics, we recommend recruitment on admission. While the recruitment proportions at some centres was disappointing, it is clear that high levels (over 80%) can be achieved irrespective of surgical procedure or type of facility.

How can recruitment be enhanced?

The main cause of non-recruitment was a failure to invite patients to participate. Overall, about 25% of potentially eligible patients were not invited to take part. This is similar to the level seen in the recent National Sino-Nasal Audit (30%). Failure to invite in many centres is understandable given the short period of time in which staff had to adapt their routine practices to the study protocol. It is likely that given a longer period in which to develop local procedures, rates of recruitment would be higher. In addition, staff training to identify eligible patients and efforts to motivate through demonstration of the benefits to staff of auditing their patients' care should lead to higher invitation rates.

Which staff are best suited to the task?

Interviews with local data collectors suggested that nursing staff are better placed than clerical staff to administer questionnaires, possibly because of their experience in interacting with patients. With correct training, however, there seems no reason why clerical staff should not be equally proficient.

Is the ineligibility rate acceptable?

The ineligibility rate (3.9%) was low for hernia, vein, hip and knee surgery. It was rare for a patient to be incapable of completing a written questionnaire in English. This may reflect the type of centres recruited to the study, as very few were located in geographical areas with a high proportion of ethnic minorities. It was also rare

for patients to be judged cognitively incompetent to complete a written questionnaire. Some elderly patients lacked sufficient manual dexterity to complete the questionnaire without help. In contrast, up to 30% of cataract patients at some centres were deemed unable to participate due, not surprisingly, to vision problems. One possible solution would be for an interviewer-administered questionnaire but we were not able to test the feasibility of this in this project.

Will patients agree to participate?
Overall, the average rate at which patients declined to take part in the study (13%) was higher than we have seen in previous research (usually less than 5%). However, there was a wide range of results across centres (0% to 67%) with a few very poorly performing centres skewing the overall mean rate.

It seems as though time pressures on both patients and staff at some centres increased the likelihood of a patient declining to take part. In many of these cases it is likely that patients would have consented to take part if they had sufficient time. It is also possible, however, that the increasing burden of paperwork on NHS patients has increased their general resistance to questionnaire completion.

It is essential that sufficient time is allowed for questionnaire administration. There may also be some value in redesigning the questionnaires to make them appear less lengthy by keeping information and consent forms separately. We believe that the consent rate can be increased to an acceptable level in any future routine PROMs programme.

Can deceased patients be identified in a timely fashion?
To avoid causing bereaved relatives and friends any distress, we used the NHS Strategic Tracing Service to identify any post-operative deaths. This was successful in eight of the eleven deaths that occurred. We therefore recommend this mechanism which, while not perfect, did enable us to limit the sending of post-operative questionnaires to only three out of 2310 patients.

Is the post-operative survey response rate high enough?
In the light of the high post-operative response rates achieved, we recommend the administrative system we adopted in which post-operative questionnaires were dispatched and returned to a central facility, separate from the centre performing the surgery.

What is the overall cost?
Based on what was achieved in this project and estimates of what we believe could be achieved (80% recruitment rate; 80% response rate) the best estimate of the cost per patient is about £6.50. This includes collecting and entering data from a pre- and post-operative questionnaire. Analyses, preparation of reports and other output, and dissemination would be additional.

Overall conclusion

We are confident that any future programme of routine PROMs administration is feasible, but have identified a number of challenges that will need to be addressed. The most important of these are the time pressures on staff and patients, and the need to develop administration methods and questionnaires that ensure high consent rates for patients undergoing cataract surgery.

7.2 Centre comparisons

What data should be collected?

This project has confirmed the responsiveness of four disease-specific PROMs (SF-36 PCS, AVVQ, OHS, OKS) for assessing the impact of surgery. With the exception of hernia repair, the effect size produced by disease-specific measures are higher than the EQ-5D. In contrast, the responsiveness of the VF-14 was only moderate. Given that there was also a concern expressed about its content validity (which, along with inappropriate treatment, would explain the ceiling effect observed) we recommend that further research is conducted on PROMs in cataract surgery.

Although it is possible to map changes in disease-specific scores onto a generic PROM (EQ-5D), the association for some procedures (cataract, vein surgery) is weak or only moderate (hernia repair). This suggests that if the benefits of different procedures are to be compared, a generic utility measure such as the EQ-5D is needed, in addition to disease-specific instruments. As regards which generic instrument, there is little difference in effect size between EQ-5D and SF-6D. Given the greater time involved in completing the SF-6D (36 items compared with 5 items), the higher incidence of missing data and the possibly higher cost (given that use of the SF-36 scale normally requires payment of a fee to the copyright owners) we recommend the use of the EQ-5D. This recommendation does not apply to cataract surgery where the EQ-5D is unresponsive to the impact of surgery.

Exploration of confounders that need to be included in risk-adjustment of various outcome measures revealed some common variables (Table 7.1): age, sex, general health status, comorbidity (eight systemic conditions) and previous surgery were all included in various models. By far the strongest predictors of post-op PROM scores were the relevant pre-op PROM scores.

Table 7.1: Confounding factors included in risk-adjustment models

Post-op outcome used as dependent variable	Significant confounders in risk-adjustment models							Variance explained by models (%)	
	Pre-op PROMs	General health status	No. of comorbidities	Prev. surgery	Age	Sex	IMD	Using relevant pre-op PROM only	Full model
VF-14	VF-14	✔						21	22
PCS (hernia)	PCS, EQ-5D	✔	✔	✔				31	47
Oxford Hip Score	OHS	✔	✔	✔				16	24
Oxford Knee Score	OKS	✔	✔				✔	15	27
AVVQ	AVVQ					✔		33	36
EQ-5D (cataract)	EQ-5D	✔	✔	✔	✔			40	45
EQ-5D (hernia)	EQ-5D, PCS	✔	✔	✔				24	38
EQ-5D (hip)	EQ-5D, OHS	✔	✔	✔				13	26
EQ-5D (knee)	EQ-5D, OKS	✔	✔		✔		✔	14	27
EQ-5D (v. vein)	EQ-5D, AVVQ	✔	✔		✔			18	29
Complications (cataract)	EQ-5D								61*
Complications (hernia)	None				✔	✔			57*
Complications (hip)	OHS		✔						59*
Complications (knee)									
Complications (v. vein)	AVVQ		✔			✔			63*

*****area under the ROC curve

Duration of symptoms was not required in any model. The inclusion of the IMD (Index of Multiple Deprivation) appeared to be useful only for knee surgery. As predicted, the global question comparing how a patient felt after surgery to beforehand proved to lack much discriminatory power and should not be included.

We recommend the use of the following PROMs in future quality of care studies:

- cataract surgery: VF-14 (until further research on the content validity of the VF-14 and treatment appropriateness in this area has been conducted)
- hernia repair: EQ-5D (until a disease-specific instrument has been developed and tested)
- hip surgery: OHS and EQ-5D
- knee surgery: OKS and EQ-5D
- vein surgery: AVVQ and EQ-5D

We do not recommend using the SF-36 PCS for hernia because: it's a generic measure; the full SF-36 has to be administered which is a considerable burden for respondents; the effect size of the SF-36 PCS is less than that of the EQ5D; and payment has to be made for its use.

In addition, the following data should be collected pre-operatively: age, sex, general health status, comorbidity and previous surgery. For knee surgery, the patient's postcode is also needed to generate an IMD score.

What are adequate recruitment and response rates?

Patients who are less healthy (as determined by the EQ-5D) are less likely to be recruited and less likely to respond to post-operative questionnaires. Post-operative non-responders also tend to be younger than responders. Considering the recruitment and response rates achieved at some of the participating centres, we recommend that 80% recruitment and 80% response rates should be sought to reduce any biases to insignificant levels.

How should data be analysed to compare centres?

The results in section 5.3 indicate that imputing missing items using the SF-36 rule is appropriate for all PROMs. The mean PROM scores were very similar when imputation and no imputation approaches were compared. The advantage of imputing is that it retains data that would otherwise be lost.

In these data, the clustering of patients nested within centres did not affect our centre comparisons (see section 5.4). However, if the project is rolled out to include all providers and data are collected from large numbers of centres this should be checked for again.

Given the marked differences in case-mix (patients' pre-operative characteristics) between centres, any comparison of outcomes must take such differences into account if it is to be meaningful. We developed risk adjustment models for each outcome which explained a large proportion of the variance in post-operative PROM scores observed.

Risk adjustment for complication rates was less convincing. This may reflect the greater importance of physical and physiological factors as determinants of the risk of complications than is true for changes in functional health status and quality of life. It may also be evidence that variation in surgical performance has a greater impact on adverse outcomes (complications) than on the extent of health gain for a patient. This can only be resolved by the addition of data on potential clinical confounding factors in future analyses.

The models included all characteristics that it is feasible to collect from patients (rather than from clinicians), with the exception of height and weight. While we recommend use of these models, it is likely that they can be improved in the future both by the availability of much larger samples of centres and patients, and by the inclusion of some clinical factors (e.g. ocular morbidity for cataract surgery) possibly derived from existing clinical datasets such as the National Joint Registry.

Should 'minimally important differences' be employed to compare centres?

Both the anchor-based and distribution-based approach to generating MID values for PROMs are problematic. The former because a stable value, representative of the continuum of pre-op severity, cannot be generated. The latter because the value generated has no known relationship with patient experience. We recommend that MID values be used with great caution when interpreting PROMs data and that statistical significance testing should remain the main approach to the comparison of centres.

What is an adequate sample size?

We recommend a sample of about 150 patients from each centre to make meaningful comparisons. This would be sufficient to detect a difference of 1 Standard Error of the Measure (for both disease-specific and generic measures) with 95% power and statistical significance of $p < 0.002$ (equivalent to 99.8% confidence intervals) for cataract surgery, hernia repair and varicose vein surgery. For hip and knee surgery this sample would be adequate for the generic measure but for the disease-specific measures the power would be slightly less.

How should centres be compared?

We compared centres in three ways: change in disease-specific PROM, change in EQ-5D and proportion of patients reporting complications. Comparisons were clearly limited by the relatively small number of centres undertaking any one of the procedures. As Table 7.2 shows, it was only the last of these that revealed outliers (at $p < 0.05$ significance level) for all five procedures. The use of disease-specific PROMs resulted in outliers for only one procedure, hip replacement. Given our concerns about the adequacy of the risk adjustment we were able to perform with the available data, acceptance of outlier status should be treated cautiously.

While we recommend disease-specific PROMs and complication rates for comparing centres, the EQ-5D does not appear to be as responsive. (For example, it did not detect any outliers for hip replacement). We do, however, recommend inclusion of the EQ-5D to enable comparisons of the utility of different surgical procedures.

Table 7.2: Number of outlying centres for different outcome measures (at $p < 0.05$ significance level).

Surgical procedure	No. of centres	Outcome measure		
		Disease-specific PROM	EQ-5D	Complications
Cataract surgery	8	0	0	1 overperformer 1 underperformer
Hernia repair	17	0	0	1 overperformer
Hip replacement	11	2 overperformers 1 underperformer	0	2 overperformers
Knee replacement	11	0	0	1 overperformer 1 underperformer
Varicose veins	13	0	1 underperformer	1 overperformer

7.3 Summarising and presenting findings

For most of the issues considered by the two focus groups, there was a lack of unanimity as how best to summarise and present findings. We have, therefore, sought to identify and recommend approaches for which we believe there is consensus (i.e. no stakeholder group disagrees strongly).

What is the best way of presenting data?

We recommend the use of funnel plots that show both the unadjusted and adjusted data. This approach also allows more than one control limit to be shown, thus allowing the observer to impose their own level of certainty.

This might, legitimately, depend on someone's objective – an individual patient might want to use a higher level of certainty than a commissioner.

The use of funnel plots does not preclude also providing access to more detailed tabulated data, particularly for surgeons.

What is the best indicator for comparing centres?

Although there was much interest in and support for the use of 'proportion of patients achieving a MID', the current methodological uncertainties about defining MIDs precludes its use at present. We therefore recommend using the mean post-operative PROM score adjusted for patients' pre-operative characteristics as the indicator of health gain. In addition, the proportion of patients reporting at least one complication (risk adjusted) should be used.

At what geographical level should centres be compared?

In view of the travel distance that patients are likely to find acceptable for common surgical operations, it is recommended that the performance of providers in the local 'region' be included in any comparison. In addition, the national average should be shown.

Data for all providers in the country could also be made available, particularly for surgeons.

At what level of provision should comparisons be made?

We had sufficient data for making comparisons of centres, not individual surgeons. Despite patients and provider managers expressing an interest in comparisons of surgeons, current NHS policy on choice is limited to centres and the outcome of surgery relates to the performance of a whole team, not just the operating surgeon. Therefore, we recommend that at present the unit of comparison is the centre.

How frequently should performance information be updated?

As data collection systems become more sophisticated and automated, it may be possible to provide continually updated information as it becomes available. The feasibility of setting up such a system should not be underestimated. In the meantime, we recommend 'batch processing' that updates every six months. Allowing for three month post-operative outcomes, comparative information on providers for hernia, vein and cataract surgery would relate to performance 5 to 10 months previous. For hip and knee replacement, which currently have a six month follow-up, the information would be 8 to 13 months previous.

7.4 Enhancing the use of PROMs

Monitoring treatment appropriateness

Much of modern elective surgery is targeted primarily at improvement in patient perceived health-related quality of life. One would therefore suppose that the patient's pre-operative health-related quality of life should be below a certain threshold otherwise, by definition, the primary objective of treatment cannot be achieved. There is

evidence from this study that many patients undergoing elective surgery have high self-reported health-related quality of life using disease-specific measures and have little capacity to benefit according to these measures. It would seem possible therefore to set thresholds in pre-operative PROMs data at which surgery is indicated. Prognostic modelling could be used to measure the degree of benefit (i.e. improvement in PROMs score) experienced by patients with differing levels of pre-op impairment in health-related quality of life, and pinpoint the pre-op PROM score at which a minimally important improvement is not likely to occur. More simply, one could subtract the minimally important difference from the maximum (i.e. best) pre-op PROM score and state that patients scoring above (i.e. with better health-related quality of life than) this level should not receive surgery. There are two strong arguments against this approach. First, the PROMs we have used in this study, and PROMs in general (given that, by definition, they are concerned with outcome) were not designed as measures of appropriateness. They are unlikely therefore, to have sufficient content validity for the task of identifying inappropriate surgery. The VF-14, for example, does not include questions about glare, which may be an important indication for cataract surgery.

A second argument is the possibility that patients would learn about the thresholds being used to determine appropriateness of surgery and 'game' their responses to ensure they receive surgery.

A third, much weaker argument is the finding that patients with little capacity for improvement report their surgery has been worthwhile. For all five procedures in this study, patients were asked to report on how they felt after surgery compared to beforehand. As predicted (see section 2.2), the vast majority of patients reported they were a little or much better: cataract surgery 93.1%, hernia repair 93.8%, hip replacement 94.6%, knee replacement 86.9% and vein surgery 90.3%. In Table 7.3 we compare the views of the patients with the best pre-op health-related quality of life (defined as the best 25% of patients) with those with the worst pre-op health-related quality of life (defined as the worst 25% of patients). The table shows that the patient groups are not radically different with respect to their views on the benefits of surgery. Indeed in cataract, hip and varicose vein surgery it is the patients with the least capacity to improve (as defined by their pre-op PROMs score) who are more likely to report that they have benefited from surgery (as defined by their post-op response to the global question on improvement).

Table 7.3: Number and proportion who report that their problems are 'much better' after surgery in patients with the highest pre-op severity versus those with the lowest pre-op severity.

PROM	High severity patients		Low severity patients	
	N	%	N	%
VF-14	104	75.9	131	82.4
PCS (hernia)	87	87.9	110	84.6
Oxford Hip Score	64	82.1	79	89.8
Oxford Knee Score	59	68.6	56	68.3
AVVQ	38	55.9	54	79.4

This analysis confirms what we observed when trying to derive an anchor-based MID: global measures of benefit from surgery (e.g. "Overall how are the problems in the eye(s) that had surgery now compared to before your

operation?") are deeply problematic. Such questions are subject to known psychometric biases such as retrospective inflation of previous problems and response shift (Streiner and Norman, 2003) and should not be used to justify surgery in patients with low pre-op severity or to generate anchor-based MIDs.

Monitoring secular trends in utilisation and equity

While there are many problems with the use of pre-op (or more generally pre-treatment) PROMs as tools in clinical decision making, the data they produce may still be of use in provoking and contextualising important health policy questions.

For example, they might be used to track changing severity thresholds for elective surgery. Why, for example, have pre-op VF-14 scores improved so dramatically in recent years.? Is it because we have begun to treat different clinical groups (e.g. second eye surgery) or is it a more general trend towards preventive surgery (i.e. treating patients who may go on to develop symptoms but have not yet done so)? If the latter is true, why is this occurring? Is this a response to changes in our evidence-base (e.g. a RCT showing the cost-effectiveness of preventive surgery) or is it related to an expansion in capacity (e.g. because of the introduction of ISTCs) and a consequent need to supply patients for this expanded capacity?

Pre-op PROMs data might also be used to identify the existence of possible inequities across sectors (e.g. is it easier to get surgery if being treated by the independent versus NHS sector) and geographical regions.

Evaluating clinical and cost-effectiveness

The routine use of PROMS will quickly lead to a large database that could be used to address clinical and cost-effectiveness questions. In our study we have only collected data from patients. In the future, however, it would be useful to link PROMs data to existing clinical databases. For example, linking Oxford Hip and Knee Scores to the data held by the National Joint Registry (ensuring appropriate consent and confidentiality procedures have been followed) would allow us to explore the comparative effectiveness of different prostheses in terms of health-related quality of life and not just short-term complications and revision rates.

REFERENCES

Bayliss EA, Ellis JL, Steiner JF (2005). Subjective assessments of comorbidity correlate with quality of life health outcomes: Initial validation of a comorbidity assessment instrument. Health and Quality of Life Outcomes 3:51-58.

Black NA, Sanderson CFB (1993). Day surgery - development of a questionnaire for eliciting patients' experiences. Quality in Health Care 2:157-161.

Brazier JE, Harper R, Jones NMB, O'Cathain A, Thomas KJ, Usherwood T, Westlake L. (1992). Validating the SF-36 health survey questionnaire: new outcome measure for primary care. BMJ 1992;305:160-164.

Critchley G et al (1997). Complications of varicose vein surgery. Annals R Coll Surg Engl 79:105-110.

Crosby RD, Kolotkin RL, Williams RG (2003). Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 56:395-407.

Dawson J et al. (1996). Questionnaire on the perceptions of patients about total hip replacement. J Bone Joint Surg Br 78:185-90.

Dawson J et al. (1998). Questionnaire on the perceptions of patients about total knee replacement. J Bone Joint Surg Br 80:63-9.

Department of Health (2007). General Information about Treatment Centres **http://www.dh.gov.uk/PolicyAndGuidance/OrganisationPolicy/SecondaryCare/TreatmentCentres/TreatmentCentresArticle**

Desai P (1996). The outcomes of cataract surgery: the relationships between visual acuity, visual function and quality of life. PhD thesis. University of London.

EuroQol Group (1990). A new facility for the measurement of health-related quality of life. Health Policy 1990;16:199-208.

Garratt AM et al. (1993). Towards measurement of outcome for patients with varicose veins. Qual Health Care 2:5-10.

Jaeschke R, Singer J, Guyatt GH (1989). Measurement of health status: ascertaining the minimal clinically important difference. Control Clin Trials 10:407–415.

Jenkinson C, Layte R, Lawrence K (1997). Development and testing of the Medical Outcomes Study 36-item Short Form health survey summary scale scores in the United Kingdom: results from a large-scale survey and clinical trial. Med Care 35:410-6.

Kazis LE, Anderson JJ, Meenan RF (1989). Effect sizes for interpreting changes in heath status. Med Care 27:S178-89.

Kolotkin RL, Crosby RD, Williams GR (2002). Integrating anchor-based and distribution-based methods to determine clinically meaningful change in obesity-specific quality of life. Qual Life Res 11:670.

Lundstrom M, Stenevi U, Thorburn W (2001). Quality of life after first- and second-eye cataract surgery: five year data collected by the Swedish National Cataract Register. J Cataract Refract Surg 27:1553-9.

Norregaard JC, Bernth-Petersen P, Alonso J et al. (1998). Variation in indications for cataract surgery in United States, Denmark, Canada, and Spain: results from the International Cataract Surgery Outcomes Study. Br J Ophthalmol 82:1107-1111.

Patwardhan A et al (2006). Incidence and outcomes evaluation of endophthalmitis management after phacoemulsification and 3-piece silicone intraocular lens implantation over 6 years in a single eye unit. J Cataract Refract Surg 32:1018-21.

Royal College of Surgeons of England and the British Orthopaedic Association (2000). National Total Hip Replacement Outcome Study. Final Report to the Department of Health

Singh AK, Esler CNA, Harper WM (2004). Complications and re-operations following primary total knee replacement from the Trent arthroplasty joint register. J Bone Joint Surg - (Br), Orthopaedic Proceedings 86-B:443.

Stratford PW, Binkley JM, Riddle DL, et al. (1998) Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 1. Phys Ther 78:1186–96.

Steinberg EP et al. (1994). The VF-14: An index of functional impairment in patients with cataract. Arch Ophthalm 112:630-638.

Streiner DL, Norman GR (2003). Health Measurement Scales: A Practical Guide to their Development and Use. Oxford University Press: Oxford.

Taylor EW et al. (2003). Surgical site infection after groin hernia repair. British Journal of Surgery 91: 105-111.

van Agt H, Essinck-Bot M, Krabbe P et al. (1994). Test-retest reliability of health state valuations collected with the EuroQol questionnaire. Social Science and Medicine 39:1537-1544.

Wight CJ, Chambers GK, Robens-Paradise Y (2002). Evaluation of indications for and outcomes of elective surgery. Can Med Ass J 167:461-6.

Wyrich K, Nienaber N, Tierney W, et al. (1999). Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. J Clin Epidemiol 52:861–73.

ACKNOWLEDGEMENTS

We thank: the many patients who participated by completing questionnaires and being interviewed; the staff of the surgical centres who recruited patients, assisted in supplying data on the activity of their centres, and provided valuable insights and information about the process; the members of the two focus groups; our surgical advisors - John Sparrow, Paul Gregg, Andrew Kingsnorth, Paul Edwards; Andrew Vallance-Owen and Claire Minett for data on BUPA members; and advice from Bernard Ribeiro, Sophie Staniszewska, Kirstie Haywood, John Brazier, David Nuttall, Donald Franklin, Aileen Simpkins and Alan Glanz.

To preserve confidentiality we have not identified the centres that participated in this project. However, they know who they are and we are indebted to their commitment and hardwork in helping make this project a success.

Appendix 1. PROMsUK User Survey Results

All interviews were conducted by telephone and all participants provided informed consent to participate. The following participants were interviewed:

General experts on PROMs (N = 3)

Professor Crispin Jenkinson, University of Oxford
Dr Andrew Valance-Owen, BUPA Foundation
Ms Rachel Reeves, Picker Institute Europe

PROM users in patients undergoing varicose vein surgery (N = 2)

Ms Sara Baker, Royal Bournemouth Hospital
Professor Paul Edwards, Countess of Chester Foundation Hospital

PROM users in patients undergoing hernia surgery (N = 3)

Professor Andrew Kingsnorth and Ms Christine Porter, Derriford Hospital Plymouth
Professor Paddy O'Dwyer, Western Infirmary Glasgow
Mr Martin Kurzer, British Hernia Centre

PROM users in patients undergoing hip or knee replacement surgery (N = 6)

Professor Paul Gregg, National Joint Registry
Mr Colin Esler, Leicester Warwick Medical School
Mr Fares Haddad, Middlesex Hospital London
Dr Paul Pynsent, Royal Orthopaedic Hospital Birmingham
Professor David Murray, Nuffield Orthopaedic Hospital Oxford
Professor Andrew Carr, Nuffield Orthopaedic Hospital Oxford

PROM users in patients undergoing cataract surgery (N = 1)

Dr John Sparrow, University of Bristol, Division of Ophthalmology

A further nine experts were contacted but interviews could not be arranged within the time frame of the study.

PART 1 -- Practical and data collection issues:

1. Experience of the interviewees with PROMs

Most interviewees reported that they collected PROMs data on all their patients indefinitely, but this was not for comparative audit purposes (e.g. for the purpose of comparing performance across institutions). In most instances data was collected to monitor the care of individual patients (e.g. monitoring pain and discomfort following hernia surgery), for specific research purposes (e.g. psychometric research) or simply to build a large database that might be useful for miscellaneous research projects. A specific comparative audit purpose is a part of PROMs use in BUPA hospitals. It was noted by one interviewee that the National Joint Registry was considering the routine use of PROMs in patients undergoing joint replacement. Decisions regarding the exact timing of administration (e.g. whether both pre- and postoperative questionnaires should be administered) and sampling methods had not yet been taken.

The interviewees generally did not consider themselves qualified to express an opinion about the value of sampling versus census methods in quality of care studies using PROMs, nor did they report that sample size calculations had been performed when determining the design of their PROMs data collection systems. One interviewee reported that a sampling rather than a census approach was used with the NHS patient satisfaction survey.

2. Patient recruitment methods

Preoperative PROMs are usually administered in questionnaire format (usually written but occasionally using touch screen computer terminals) at either a pre-surgery assessment visit (usually around two weeks before surgery) or on admission for surgery. However, BUPA report that preoperative questionnaires are sent to patients by post with the letter inviting the patient for surgery.

A variety of postoperative recruitment methods were reported by those using PROMs in a routine fashion. A key determinant is the care protocol used at a particular institution. For example, where patients are invited to a routine follow-up examination after surgery, this visit can be used as an opportunity to collect postoperative PROMs data. For example, both of the interviewees working with varicose vein surgery patients reported that routine postoperative hospital appointments are the norm and are conducted at a time point (around 6 weeks after surgery) that allows for the administration of postoperative PROMs to all patients.

However, in most instances routine postoperative patient assessments at an appropriate time point are not the norm. In hip/knee replacement and in cataract surgery at least one outpatient postoperative appointment is offered to the patient but this is usually too early (6 to 8 weeks) to allow for the collection of useful PROMs data. Thus, postal surveys are usually used to collect postoperative PROMs data with these patients. Some interviewees reported that postoperative PROMs data was collected from all patients who attend postoperative clinics, but after the first visit only patients with ongoing problems were invited to attend. Thus, in these centres, 100% postoperative follow-up of the cohort was not possible at the most interesting postoperative time points.

Postoperative hospital appointments are not normally offered to patients undergoing hernia surgery and postoperative PROMs are therefore usually collected via telephone or postal surveys.

The NHS Patient Satisfaction Survey has recently commissioned a piece of research on the effectiveness of different methods for recruiting patients into its survey. This research will compare the response rates obtained from a telephone survey, a postal survey and a survey using face-to-face questionnaire administration.

3. Consent

Most interviewees reported that formal written informed consent to allow the administration of pre- or postoperative PROMs was not obtained from the patient. Where informed consent was not obtained it was because the interviewees considered the use of PROMs at their institution to be a part of the routine care process. However, some interviewees did state that written informed consent was a part of the PROMs data collection process at their institution, particularly with respect to permission to send a postoperative questionnaire to the patient's home address. Where informed consent was obtained, the person supervising the consent process was a designated local staff member such as a junior clinician, research nurse or clerk.

Written informed consent is not currently obtained in the NHS Patient Satisfaction Survey. Completion of the questionnaire is taken as a sign of consent to participate. The National Joint Registry has previously sent postoperative PROM questionnaires to a sample of 10,000 patients. Consent particular to PROMs was not obtained in this exercise, but the overall consent process for the National Joint Registry does cover the possibility that the patient will have a questionnaire sent to their home address at some point after surgery. Written informed consent is not obtained by BUPA hospitals before sending questionnaires to their patients. A covering letter explains the purposes to which the data will be put and the patients are given the option of not completing the questionnaire.

4. Confidentiality

There were mixed responses regarding confidentiality provisions. All interviewees stressed that they followed best practice in ensuring the security of any data collected. However, in some instances clinicians have access to patient identifiable PROMs data for patients under their care. In these instances the interviewees noted that PROMs data collection was considered a part of the process of care for individual patients, and was not used to monitor quality of care at the institution.

Where PROMs data was collected primarily for quality of care purposes (e.g. at BUPA hospitals) there was recognition of the need to ensure patients could freely express their opinions about their health and quality of life, both before and after surgery, without fear of this affecting their care. Data was therefore stored in an

anonymous fashion wherever possible and was not available in an identifiable format to the clinicians in charge of the patients care.

In the NHS Patient Satisfaction Survey, the patients name and address are released to a third party (Picker Institute Europe). To facilitate this, employees at the Institute obtain honorary contracts at the various Trusts who participate in the Survey

5. Catering for patients at high risk of attrition

A. Patients with literacy problems. The NHS Patient Satisfaction Survey has a telephone 'hotline' where patients can respond to the postal questionnaire over the phone should they have problems with literacy. Other interviewees reported that their were no provisions for patients with literacy problems when questionnaires were sent by post, other than to ensure that they were written with the lowest possible reading age in mind. In instances where questionnaires are administered face-to-face (e.g. preoperative questionnaires to be completed on admission for surgery in the hospital) most interviewees reported that the person administering the questionnaire was available to answer patient queries and even to read the questions to the patient if necessary.

B. Patients whose first language is not English. The NHS Patient Satisfaction Survey has a 'language line' service that allows for the live translation of questionnaires into alternative languages. Alternative language versions of written questionnaires are also available. All other interviewees reported that alternative language versions of written questionnaires were not used in the projects with which they were familiar. However, they did report that *ad hoc* translation of the questionnaires by friends/relatives or the local hospital translation service was often used when questionnaires were administered in hospital. They presumed that friends/relatives also translated questionnaires for the patient when they were completed at home.

C. Patients with cognitive impairment. None of the interviewees reported specific arrangements to cater for patients with a cognitive impairment. Most considered it to be a very rare problem. For patients with a minor cognitive impairment it was common to give the patient some assistance when completing the questionnaire (e.g. read it to the patient rather than have them read it themselves) or to ask a friend/relative to give the patient assistance. The interviewees reported that patients with severe cognitive impairment would be excluded from PROMs studies at their institution, but they noted that it would be quite rare for these patients to have elective surgery. The NHS Patient Satisfaction Survey is considering the need for specific arrangements to ensure that they can cater for all patients availing of mental health services.

6. Dealing with non-responders

Some interviewees reported that they made no attempt to 'chase' patients who did not respond to a postal survey. Otherwise one reminder letter was normally used. The NHS Patient Satisfaction Survey uses two reminder letters.

7. Checking to ensure that patients are still alive before contacting them at their home

The systematic tracing of patients who have died (e.g. by using the NHS Strategic Tracing Service) was only performed by the NHS Patient Satisfaction Survey.

PART 2 -- Data analysis and interpretation

1. Dealing with missing data

Most interviewees were not aware of the specific methods used to deal with missing data from written questionnaires in the projects with which they were familiar. One interviewee reported that the policy was to ignore any measures that did not have all the items completed on the psychometric instrument of interest. Another interviewee reported that imputation was used as long as only "1 or 2 items" were missing. A third interviewee reported that patients were rung at home to request the answers to missing items.

Missing data was not considered to be a large problem for questionnaires completed in the presence of hospital staff, as it was common practice to check the questionnaires following completion and ask the patients for the answers to any missing items.

2. Risk adjustment of PROMs data to ensure fair comparisons between healthcare providers

Most interviewees did not express a strong opinion about this issue. It was agreed that this was an important issue and most felt it was important to use the pre-operative scores on a particular health outcome measure to in some way 'adjust' the post-operative scores. It was also felt that adjustment for age and sex should be done as a matter of course. Beyond this, most interviewees felt that the development of trustworthy risk adjustment models was a task for future research. One interviewee was concerned that important aspects of the patient's treatment history (e.g. referral from another hospital, time on waiting list) might be very difficult to record.

3. Defining a clinically important difference in performance?

Interviewees were generally cautious when asked about what would constitute a clinically important difference between healthcare providers (in the context of PROMs data). Some expressed surprise at the idea that PROMs could ever be used for this purpose. Most argued that PROMs data should only be used to rate healthcare provider performance in conjunction with data on clinical outcomes. Most interviewees were familiar with the idea of distribution based methods for defining a minimally important difference (e.g. effect size) but felt that these approaches lacked face validity. BUPA have a 'control chart' approach to monitoring the performance of their hospitals. This approach allows for the detection of hospitals whose mean score on a particular PROM deviates (e.g. +/- 3 SD) from a national average. The NHS Patient Satisfaction Survey project encourages participating NHS Trusts to aim to reach the 'top' 20% on key indices such as overall satisfaction with care.

4. Useful methods for presenting PROMs data to improve clinical practice

Most interviewees were not familiar with formal methods for presenting PROMs data. As mentioned above, BUPA use control chart presentations to compare the performance of different healthcare providers. Some interviewees reported that they routinely compare the raw PROMs data recorded for different consultants within the same institution, using simple bar charts.

One interviewee suggested that a simple comparison across consultants or hospitals might be the (risk-adjusted) proportion of patients who achieve a minimally important improvement in outcome on the PROM in question.

5. Can PROMs detect under-performing centres in a timely way

Most interviewees were concerned that any comparative audit which only concentrates on performance over a particular period may take an unacceptable amount of time to detect poor performance (because of the need to accumulate a large number of cases), or miss poor performance occurring at other times. It was felt that this problems was particularly acute for audits that used PROMs as the primary outcome measure, because of the need to collect long-term outcomes (e.g. up to 12 months following joint replacement).

Some interviewees were worried that a focus on 'early' detection of underperformance might lead to excessive false accusations. There was a concern that poor results for a particular hospital in a particular year might be due to chance, and that over time any particular 'outlier' would assume a less extreme position within the distribution.

6. Other useful applications of this type of data

Almost all the interviewees agreed that the collection of preoperative PROMs data could allow for interesting comparisons across healthcare providers. There was much anecdotal evidence that patients had different preoperative severity profiles at different hospitals. Many clinicians stated that they thought the hospitals with more deprived patient constituencies were those with the most conservative 'listing' policies (i.e. only those patients with the highest disease severity were listed for surgery). It was also stated that there was a widespread belief that it was easier to get listed for surgery at a private hospital.

There was a mixed reaction to the idea of using preoperative PROMs scores to identify patients who should and should not be listed for surgery. The idea of a 'cut-off point' below which surgery is not indicated had been explored in some areas (e.g. cataract surgery) but interviewees were generally uncomfortable with this approach. It was suggested that PROMs were not designed for this use and were therefore not capable of capturing the variety of clinical indications for surgery (which might not necessarily manifest themselves as problems perceived by patients). It was also suggested that patients might eventually 'game' this system, if they

became aware of the purpose to which the preoperative questionnaire was put (e.g. by exaggerating their symptoms).

All interviewees recognised the value that a large, trustworthy database of patient outcomes would have in answering research questions outside the quality of care arena. These include prognostic questions (e.g. which patients benefit the most from surgery), equity questions (e.g. do patients from some geographical areas benefit more), effectiveness questions (e.g. do some surgical techniques provide better patient-reported outcomes than others) and cost-effectiveness questions (e.g. are some models of service provision more cost-effective than others).

Appendix 2. Case studies of international PROMs use in quality of care assessment

Case Study 1: Medicare (USA) Health Outcome Survey - Interview with Dr Chris Haffer

Managed care plans are an important source of health care services for Medicare beneficiaries (i.e. insured over age 65 or permanently disabled under age 65) in the United States. At present, 5.3 million beneficiaries receive care in these settings, of whom 87% are enrolled in Medicare+Choice plans (renamed Medicare Advantage plans in 2004). The number enrolled in these plans is projected to increase to 13.6 million by 2010, given the recent passage of the Medicare Prescription Drug, Improvement and Modernization Act. Thus, an estimated 30% of Medicare beneficiaries will be enrolled in managed care settings by 2010.

The Medicare Health Outcomes Survey (HOS, previously the Health of Seniors survey) for managed care is the first patient-reported outcomes measure to be used in the Medicare population. It is a longitudinal, self-administered survey which - prior to 2006 - utilized the SF-36 and additional case-mix adjustment variables. Effective 2006, this survey uses the VR-12, which is very similar in content to the SF-12. The VR-12 is being used because the instrument and all scoring algorithms are fully in the public domain.
One of the main objectives of the HOS is to compare the 'performance' of the nearly 300 Medicare Advantage Plans (MAPs). It is important to note that Medicare recipients are eligible for inclusion in the HOS regardless of whether they have recently received care (i.e. patients are not linked to specific care episodes). Thus, the purpose is not to make a direct evaluation of the quality of care delivered by MAPs, but to compare the overall performance of MAPs in maintaining and improving the health and quality of life of the Medicare patients enrolled in their plans.

Since the inception of the Medicare HOS program in 1998, CMS has, on average, spent more than $2.1 million per year to fund the HOS program.

PART 1 -- Practical and data collection issues:

1. Experience of the interviewees with PROMs

Dr. Chris Haffer is currently Director of the Medicare Health Outcomes Survey and a Technical Advisor in the Office of Research, Development, and Information at the Centers for Medicare and Medicaid Services (CMS). Ms. Bowen is a program administrator. Mr. Long is a research analyst.

2. Patient recruitment methods

For each of the MAPs, one thousand randomly sampled Medicare beneficiaries, who were continuously enrolled for a six month period in the plan are surveyed every Spring. Two years later, these same respondents are surveyed again. In addition, a new baseline survey is administered to a new cohort each year. The first baseline cohort was surveyed in 1998 and was resurveyed in March of 2000. Since its inception over 1 million beneficiaries have been surveyed. Response rates have averaged 66% for baseline surveys and 81% for follow-up surveys since the inaugural year.

3. Consent

Informed consent is achieved using a cover letter included with the first HOS questionnaire mailing. There is no penalty for refusal and recipients are fully informed of the data uses by CMS and their health plan. Survey vendors are not allowed to use incentives of any kind.

4. Confidentiality

Patient identifiers are collected and retained by those carrying out the HOS. Patient identifiable data is made available to the MAPs after each completed cohort as part of the process of feeding back data on performance. This should, in theory, allow each MAP to match each of the surveyed patients under their care to administrative data, should they wish to explore further the MAP-specific predictors of performance.

5. Catering for patients at high risk of attrition

A. Patients with literacy problems. Proxies are allowed and encouraged to assist with the HOS.

B. Patients whose first language is not English. Versions of the HOS are available in Spanish and Chinese.

C. Patients with cognitive impairment. A modified version of the HOS has been developed for vulnerable patients at greatest risk for poor outcomes, which may include cognitively impaired.

6. Dealing with non-responders

The survey process involves (i) a pre- survey notification postcard informing recipients that they are about to receive a postal questionnaire (ii) the questionnaire (iii) a standard 'thank you' postcard sent to all those surveyed, which also acts as a reminder for those recipients that have not yet responded (iv) a second questionnaire mailing for non-responders (v) a second reminder postcard (vi) at least six telephone follow-up calls to act as a final reminder for non-responders.

There are small average differences in some of these characteristics between follow-up responders and non-responders. Respondents who do not respond, on average, are older, less educated, have lower incomes, are in poorer mental and physical health, are more likely to be Medicaid beneficiaries, and are more likely to be African American and male. Although these differences exist, the impact of attrition on analyses using the HOS is likely to be relatively minor. On many of these dimensions, average differences between follow-up respondents and attriters are small, even if they were statistically significant.

7. Checking to ensure that patients are still alive before contacting them at their home

The HOS uses data from the U.S. Social Security Administration to verify that Medicare recipients are still alive when they are sent the HOS questionnaires.

PART 2 -- Data analysis and interpretation

1. Dealing with missing data

The HOS has developed regression based imputation algorithms that allow them to impute from completed items to missing items.

2. Risk adjustment of PROMs data to ensure fair comparisons between healthcare providers

The HOS takes account of socio-demographic variables (including poverty status), baseline health status measures, and individual chronic conditions to perform case-mix adjustment. The expected probability that a beneficiary's Physical Component Summary (PCS) score and Mental Component Summary (MCS) were worse, the same or better over the two-year follow-up period is calculated for all beneficiaries who have both baseline and follow-up PCS/MCS scores. The calculation of the expected probability of PCS/MCS worse/same/better is based on the case-mix of the beneficiary at baseline. A series of cascading logistic regression models has been developed (logistic regression was used because the probability of PCS same or better is a dichotomous variable where 1=PCS same or better and 0=PCS worse). To get the best available model for all of the data, the most comprehensive model available is used for each beneficiary. Thus, if all of the independent variables are present for Model A, this model is used to calculate an expected probability of PCS same or better for the beneficiary. If not, then Model B is used if all of the independent variables for Model B are not missing, and so on. Only one model is used for each beneficiary, and an expected probability is calculated for every beneficiary.

3. Defining a clinically important difference in performance?

The performance of different MAPs is assessed by calculating the difference between actual and expected results over two years for three major outcome variables: death, change in physical health as measured by the PCS score, and change in mental health as measured by the MCS score. For reporting purposes, death and PCS scores are combined to measure change in physical health. Scores are reported as the percentage of respondents within a plan who were better, same, and worse in physical health and in mental health over the two-year period, after adjustment for case-mix. The modelling and parameter estimates utilized in the calculation of all HOS performance measurement results were derived from the original Cohort I analysis.

If statistical analyses reveal that the MAPs differed significantly, outlier plans are identified. A *t* statistic, expressing the significance of the MAP differences from the average national results, is calculated by dividing the MAP deviation by the standard error of the deviation. Plans with a *t* statistic > 2 are designated as significantly better than expected, while plans with a *t* statistic < -2 are designated as significantly worse than expected, compared to the average national results.

4. Useful methods for presenting PROMs data to improve clinical practice

A key function of the HOS is to provide health status data to MAPs in order to facilitate improvements in quality of care provided to their Medicare populations. Each year, a baseline report and a performance measurement report are produced for each MAP participating in the Medicare HOS. Each participating MAP receives plan-specific baseline and performance measurement reports, which present aggregated results for its plan, the state total, and the HOS national total. After the administration of each baseline cohort, a cohort-specific baseline report is produced. This report presents PCS and MCS scores and includes data on resource utilization predictors, health status indicators, comparative results, and respondent demographics. The sub-scales also provide useful information, such as bodily pain and vitality scores, that can be used to guide quality improvement plans. The performance measurement report provides information similar to that of the baseline report plus MAP performance data.

Five HOS conferences have been sponsored, targeting various stakeholders. The conferences have provided training related to maximizing the use of the HOS, explaining the role of Medicare HOS in CMS's quality improvement strategy, and presenting interventions designed from HOS results. In addition, the HOS partners present the HOS findings and activities at professional conferences on an ongoing basis.

The area of CMS responsible for using the HOS data for plan performance monitoring is the Division of Health Plan Accountability. CMS influences the performance of health plans by: (1) providing an incentive for health plans to provide and maintain quality of care (i.e., the exemption from portions of the biennial site audit) and (2) imposing a penalty for health plans with poor performance in the form of possible contract non-renewal of health plans with a persistent history of poor performance. Few plans are actually non-renewed due to performance issues.

Although CMS is using the HOS data for performance monitoring and assessment, CMS is not using them to encourage quality of care through market pressure brought to bear by publicly reporting the HOS data. CMS has developed and rigorously tested the module to publicly report the HOS data but has yet to implement the tool.

5. Can PROMs detect under-performing centres in a timely way?

HOS performance measurement data are available to CMS at the end of the data collection year for use in monitoring the performance of MAPs, i.e. data collected in Spring 2006 will be available for CMS use at the end of calendar year 2006. In addition, plan-level HOS performance measurement data are made available to MAPs approximately one year after follow-up survey data are collected for each cohort.

6. Other useful applications of this type of data

The HOS has been used as a vehicle to explore emerging and/or geriatric quality care–related health issues and concerns. The HOS instrument has included questions about retirement community living, smoking frequency and cessation, management of urinary incontinence, and healthy days.

HOS data are used by CMS in calculating a "frailty" payment adjustment. That is, an adjustment to a health plan's payment to account for the increased cost of caring for the extremely frail elderly.

The HOS data have also contributed to the development of a database of measures of functional status and health outcomes of disadvantaged populations. CMS also plans to establish a chronic disease database of health-related data by chronic diseases, as mandated by the Medicare Modernization Act of 2003, using the HOS as one of the data sources.

Case study 2: Veteran's Administration (USA) Health Outcome Survey - Professor Lewis Kazis

PART 1 -- Practical and data collection issues:

1. Experience of the interviewees with PROMs

In the United States, a health care system designed specifically for members of the military and their relatives is in existence (the Veteran's Health Administration or VHA). In 1995, the VHA began surveying its patients using a standardized instrument modelled from the Picker Institute. The Performance Analysis Center for Excellence (PACE) of the Office of Quality &Performance (OQP) is the analytical, methodological, and reporting staff for Surveys of the Health Experiences of Patients (SHEP).

The VHA is divided into 21 regional Veteran Integrated Service Networks (VISNs), individual Medical Centers (VAMCs), and Community Based Outpatient Clinics (CBOCs). The information provided by the surveys is designed to drive improvement at all levels of the organization (i.e., National, VISN, VAMC, and CBOC) including the development of VHA's strategic plan.

PACE/OQP currently administer three different surveys. In decreasing order by sample size these include Ambulatory Care SHEP, Inpatient Care SHEP, and the Veteran Satisfaction Survey for Prosthetics/Sensory Aids. The Ambulatory and Inpatient Care SHEP are directly comparable to the Picker Institute's surveys allowing for a private sector referent. The SHEP contains the VR-12, a questionnaire similar in content to shorter versions of the SF-36.

2. Patient recruitment methods

Beginning in 2002 the Ambulatory Care and Inpatient surveys have been administered monthly with quarterly follow-up for outpatients and semi-annual follow-up for inpatients. The Prosthetics/Sensory Aids survey is administered every other year. All surveys are administered by mail to a sample of veterans meeting qualifying criteria. PACE/OQP is responsible for designing the survey instrument, drawing the sample, performing the analysis, and reporting results. Data collection (mail, data scan and comment capture) is contracted to an external organisation.

3. Consent

Return of questionnaire is taken as consent to participate.

4. Confidentiality

To protect the confidentiality of patients, percentages of respondents are not reported when the sample size for a particular healthcare provider is less than 10. There are no exceptions to this rule.

5. Catering for patients at high risk of attrition

A. Patients with literacy problems. No arrangements are in place at present for respondents with literacy problems.

B. Patients whose first language is not English. A Hispanic version of the survey is available.

C. Patients with cognitive impairment. No arrangements are in place at present for respondents with literacy problems.

6. Dealing with non-responders

A reminder letter is sent to non-responders.

7. Checking to ensure that patients are still alive before contacting them at their home

Care is taken to ensure that sample frames are as up to date as possible with respect to respondents' mortality status.

PART 2 -- Data analysis and interpretation

1. Dealing with missing data

The normal SF-36 style rules are used to deal with missing data.

2. Risk adjustment of PROMs data to ensure fair comparisons between healthcare providers

There are differences in patient characteristics (age and health status) across networks that are known to influence patient satisfaction scores. In an effort to "level the playing field" for all sites of care, the VHSS scores are adjusted for those patient characteristics.

3. Defining a clinically important difference in performance?

Statistical methods only are used to detect differences in performance.

4. Useful methods for presenting PROMs data to improve clinical practice

SHEP Survey Reports are published electronically. These reports give VISN, facility, bed section and CBOC level data (where applicable) and comparisons, including statistical variance from the National mean and an external referent, where applicable. Tools for performing detailed analyses for the surveys (e.g., Medical Center, Clinic, Question Level, etc.) are provided through a custom designed web-based query form.

The five main SHEP performance measures (Inpatient Overall Quality, Outpatient Overall Quality, Provider Wait Time, Primary Care Appointment As Soon As Wanted – New Patients, and Primary Care Appointment As Soon As Wanted – Established Patients) are produced on a monthly basis.

Outpatient and Inpatient reports are produced quarterly and semi-annually, respectively. The purpose of these reports is to produce clinic, facility, and network scores, and to identify those locations that are performing significantly better (or worse) than the national average.

5. Can PROMs detect under-performing centres in a timely way?

Certain reports are available via the internet to those healthcare providers who wanted a more rapid data cycle. However, this rapid data cycle comes at the risk of presenting biased results. The data available via these rapid reports can be used for local trending and quality improvement purposes, but is not reported in any official capacity.

6. Other useful applications of this type of data

The data produced in these surveys has the potential to be used in effectiveness and prognostic studies.

Appendix 3: Patient Participation Sheet

NAME OF HOSPITAL:

			Patient given questionnaire		**Patient not given questionnaire**				
Patient hospital identifier	Date	Operation type HIP, KNEE VEINS HERNIA CATARACT	Patient completed (or partially completed) form	Patient refused	Patient does not understand English (please note primary language)	Patient has literacy problem	Patient visually impaired	Patient cognitively impaired	Other reason

Appendix 4: Schedule for semi-structured interviews with local data collectors

GROUP CONSENT

- Was the process BETTER or WORSE than you envisaged?
- Why were you successful? If not, why did it not work?

Staff

- How many people here were actually involved in giving out the questionnaires?
- Grades of staff
- Timing
- Talk me through the process of capturing patients for this study

Recruiting and consent of patients

- Problems in identifying patients eligible for the study? How did you deal with this?
- Did you handle the recruitment and consent process yourself or did others help you?
- Where and when in the patient care pathway did recruitment and consent happen?
- Did you have enough time to properly manage the recruitment and consent process?
- What were the most common queries from patients about participating in the study?
- How did you go about identifying patients who could not complete a written questionnaire in English? Was this easy to do? Do you think you made any mistakes?
- What do you think were the most common reasons for refusing consent to participate?
- Problems with consent and recruitment particular to a surgical group (e.g. hips, knees, cataracts, hernia, varicose veins)?

Completion of questionnaires

- Main problems in overseeing the completion of questionnaires? How did you deal with this? How much time did it take?
- How long did each of the 5 questionnaires take to complete?
- What were the most common queries from patients while filling in the questionnaires?
- What did they find most difficult or confusing about the questionnaires?
- What do you think were the most common reasons for failing to complete a questionnaire?
- Problems with questionnaire completion that was particular to a surgical group?

Problems with communicating with the CEU at RCS.

- Were the site visits useful?
- What could have been improved?
- Were the study handbook and other materials from the CEU useful?
- How could they have been improved?

Suggestions about what a future system should look like that uses PROMs to monitor quality of care?

- Questionnaires
- Mode of administration
- Staff level

Appendix 5: Schedule for semi-structured interviews with patients

1. Overall what do you think about the questionnaires we asked you to complete?

2. Do you think they are a nuisance for patients or will patients be happy to complete them? Please consider
   a. Time to complete
   b. Comprehension
   c. Upset they cause

3. How could we increase the likelihood that patients will respond to questionnaires?

4. Do you think the actual questions we asked in the pre-op questionnaire were appropriate for the condition? If not, why not?

5. Do you think the actual questions we asked in the post-op questionnaire were appropriate to the recovery of the condition? If not, why not?

6. Mode of administration
   a. What point should the questionnaires be given out in the patient care pathway?
   b. Should patients be able to take them home?
   c. Would it be better to have them read out over the telephone?

Appendix 6: Study Handbook for centres

The Royal College of Surgeons of England

Patient Reported Outcome Measures in Elective Surgery

1. Who are we and how can you contact us?

We are researchers based at the Clinical Effectiveness Unit (CEU) of The Royal College of Surgeons of England (RCS) and the London School of Hygiene and Tropical Medicine. Staff at the Unit have expertise in carrying out this sort of work. I am the Research Fellow working on this project (Dr. Liz Jamieson). The Principal Investigator is Dr. John Browne (Senior Lecturer). We can both be contacted at the Clinical Effectiveness Unit, The Royal College of Surgeons of England, 35-43 Lincoln's Inn Fields, London WC2A 3PE.
Telephone: 020 7869 6603 (John Browne); 020 7869 6607 (Liz Jamieson).
E-mail: jbrowne@rcseng.ac.uk ; ljamieson@rcseng.ac.uk.

2. Background to the study

Treatment Centres (TCs) are dedicated units which provide planned surgery and diagnostic procedures for a range of conditions. Both NHS and Independent Sector TCs have been introduced over the last two years to reduce waiting times for routine procedures and increase the capacity of spaces available for such treatments. Rigorous, high quality assessment of outcomes is essential for any new treatment or method of service provision. The Department of Health in England has decided that the quality of care of TCs needs to be carefully monitored and compared to the performance of mainstream elective surgery units. The use of *Patient Reported Outcome Measures (PROMs)* in addition to clinician-based assessments will ensure that patient perspectives on the performance of elective surgery units will be taken into account.

2.1 Aims

The aim of the study is to discover how easy it is to find out what patients think about their surgery. We are also interested in how we can use patient opinions to improve healthcare services.

2.2 Objectives

The overall aim of this pilot study will be to establish the feasibility of routine PROMs use with patients undergoing elective surgery. 'Feasibility' will be evaluated through examination of case ascertainment rates (that is, have we included all eligible patients?), consent rates, 'missing data' levels and postal survey response rates to the follow-up questionnaire.

2.3 Ethical Issues

The study has received multi-centre ethics committee (MREC) approval and is supported and funded by the Department of Health. It has also been reviewed and approved by the local Research & Development Department.

3. Appointment of yourself as the local staff member

We are grateful that you have agreed to help us with the pre-operative component of the study. Your involvement will be explained to you in this pre-site visit. We will also use this visit to deliver the patient pre-operative questionnaires and SAEs to you.

4. Duration of study

Please start data collection immediately after this site visit and continue until we contact you to say that data collection can stop (we do not envisage this taking any longer than three months).

5. Procedure

(1) Before you start the study, it would be very useful if you had **clear definition of when and where you will be distributing the questionnaires.**

(2) Adult patients (16 years or older) undergoing the following types of surgery are

eligible for inclusion:

- hip replacement (not bilateral)
- knee replacement (not bilateral)
- cataract removal
- groin hernia repair
- varicose vein surgery

(3) Before approaching patients to take part in our study, please will you consider whether they are **able to complete the questionnaire.**

(4) **If you judge that a patient would be unable to complete the questionnaire**, please:

- do not pass the questionnaire booklet to the patient
- keep a record of whether one of the following applied:
    - patient has literacy problems;
    - patient has eyesight problems;
    - does not understand English (patient's primary language to be noted);
    - patient has sufficient cognitive impairment to make it impossible for them to complete a written questionnaire;
    - other reason.
- A form which will help you keep track of which patients did and did not complete the questionnaires is attached to this handbook.

(5) **If you consider that a patient would be able to complete the questionnaire,** please pass them the booklet and explain that you have been requested on behalf of the Royal College of Surgeons to kindly ask patients to complete a questionnaire to find out how they are before surgery.

(6) **If a patient informs you that he/she does not wish to take part, please take the booklet back from them** and record the fact that the 'Patient refused'. The booklet can be handed out to someone else (provided it is unmarked). However, all partially completed booklets must be returned to us.

(7) **If a patient decides to take part in the study and returns the completed questionnaire to you:**

- please check that the consent form has been completed
- give the patient a copy of the consent form (a detachable carbon copy is available in the questionnaire);
- keep one copy of the consent form in the patient's clinical notes (a detachable carbon copy is available in the questionnaire);
- please check that the patient's name and address have been completed in the booklet;
- return the completed booklets to us as soon as possible in the stamped, addressed envelopes provided to you.
- record the fact that 'Patient consented'.

6. Frequently asked questions about this study (staff questions)

What should I do if a patient does not consent to take part in the study?
Make a tick in the table to record that a patient listed for a particular operation refused to take part.

What should I do if I know that a patient cannot read/ is visually impaired, has cognitive difficulties or does not speak or understand English?
Please do not pass a booklet to the patient and, again, please would you place a tick in the relevant column on the attached table to record the fact that an eligible patient was unable to take

part because of one of the above reasons. In cases where the patient does not understand/ speak English, it would be useful to know the patient's primary language.

What should I do if the patient consents but then it appears that he/she cannot fill in the questionnaire by himself/ herself?
Please do not help them complete the questionnaire. Simply wait until they have finished with the questionnaire booklet and make a note of the problem by ticking the relevant column in the table (for example, visual or cognitive impairment).

What should I do if a patient asks me for help regarding a question(s) in one of the questionnaires?
Please suggest to the patient that he/she completes the answer he/she feels is most appropriate. Please do not discuss the actual questions with the patient.

How many patients are being studied?
It is hoped that we will have results from 2,250 patients from around 22 different hospitals.

How can I reassure a patient that his/her confidentiality will be observed?
All information about individuals, units or hospitals will be strictly confidential and will not be disclosed or released to others outside the research team. All reports and publications will contain summary data to prevent identification of individuals, units or hospitals.

How will we find out about the patients' outcome after surgery?
The Royal College of Surgeons will send each consenting patient a questionnaire through the post a few months after their surgery.

How will the results be used?
The results will be disseminated through presentations at relevant conferences and by submitting papers to peer reviewed journals.

I would like to know more about the study/ find out about the results
Please contact John Browne at the Clinical Effectiveness Unit, The Royal College of Surgeons of England, 35-43 Lincoln's Inn Fields, London WC2A 3PE (telephone: 020 7869 6603). We hope the results will be published in about 18 months' time.

7. Frequently asked questions by patients

I don't understand a question- can a member of staff help me?
If a patient does not understand a question, please request that he/ she leaves the question blank rather than you explaining it to them.

Can a member of staff read out the questions to me?
Please do not read out questions to patients. Apart from the fact that this would be time-consuming for you, the questionnaires are not designed to be read out.

Will I have to complete any more questionnaires at any time?
As part of the pre-operative set of questionnaires, the patients will consent to being sent a very similar follow-up questionnaire at some point after their operation. The patient will also provide his/ her home address in order that the questionnaire may be sent there

- 3 months after hernia, varicose vein and cataract operations
- 6 months after a knee or hip replacement operation

Will my Doctor see my answers?
No. Please explain that only research staff will see the responses. A patient's treatment will not be affected in any way.

I would like to know more about the study/ find out about the results
Please contact John Browne at the address above. We hope the results will be published in about 18 months' time.

Thank for your time and help with this study.

Appendix 7: Summary of study for centres

Piloting the routine use of patient reported outcome measures in elective surgery patients

Dr John Browne, Senior Lecturer in Outcome Assessment,
Health Services Research Unit,
London School of Hygiene and Tropical Medicine

Aim
This project will establish the feasibility of collecting patient opinions (patient reported outcome measures, known as PROMs) routinely in Treatment Centres and NHS mainstream elective surgery units.

The primary aims of the study will specifically address the following questions:

- Will we pick up all eligible patients?
- What is the consent rate?
- What is the refusal rate?
- We will also estimate how many eligible patients were not recruited into the study, i.e. how many patients could not be given a questionnaire because of eyesight, comprehension, literacy or language difficulties.
- What is the response rate to the follow up questionnaire?

Sample
Consecutive patients listed for cataract surgery, hernia repair, hip replacement, knee replacement or varicose vein surgery at participating Treatment Centres or NHS mainstream surgery units over the duration of the study will be invited to participate

Method

Part 1: Pre and post op questionnaires
Patients will be asked to complete a questionnaire pre-operatively and at an appropriate post-operative time point. A nominated local staff member at each participating centre will take responsibility for consenting and recruiting eligible patients, noting if a patient cannot complete a questionnaire due to literacy problems, cognitive defects, language or eyesight problems. Each centre will be visited once before the study begins by a member of our research team. All procedures will be explained in detail at this visit. The centres will be given a telephone number at the Royal College of Surgeons (the co-ordinating centre) which they can ring during normal working hours should they have any queries about data collection procedures.

A literature review has recommended the use of the following questionnaires:

- All patients: EQ-5D
- Cataract patients: VF-14
- Hip replacement patients: Oxford Hip Score + SF-36
- Knee replacement patients: Oxford Knee Score
- Varicose vein procedure patients: AVVQ
- Hernia repair patients: SF-36

In the pre-operative questionnaires we will ask patients to provide information on their age and sex and aspects of their disease history (previous surgery for the target condition, comorbidities, rating of general health, length of time with the condition) so that we can provide descriptive data on the samples. This will allow policy makers and the research community to draw conclusions about the generalisability of our data.

Varicose vein, hernia and cataract patients will be sent postal questionnaires 3 months after their surgery. Hip and knee replacement patients will receive the questionnaires 6 months after their surgery.

Prior to sending out the follow-up questionnaires, each participating NHS centre will be asked to search using the NHS Strategic Tracing Service for all patients from their centre who have agreed to participate in the study. This will identify which patients have died since they had their operation. The names of any deceased patients will be communicated to us and these will be removed from the sample.

Part 2: Interviews

Interview data will be used to inform the study recommendations by focussing on improving the data collection process and evaluating different feedback methods.

The data collection process will be discussed in semi-structured interviews with:

(a) the local staff member responsible for carrying out the study at each centre. These interviews will focus on the time and problems involved with collecting the data.

(b) two clinicians (one nurse and one surgeon). These interviews will focus on the usefulness of feedback methods and the acceptability to surgeons and nurses of the statistical techniques proposed.

(c) a patient representative from each of the participating sites (selected through the existing Patient Advisory and Liaison Services (PALS). These interviews will focus on the usefulness of PROMs for quality of care purposes.

Appendix 8: Local data collectors' views of data collection

General feedback
The number of staff involved in administering the pre-operative questionnaires varied between one and twenty depending on the centre involved. Staff who gave out the questionnaires were usually nursing staff and, in some cases Health Care Assistants, student nurses or receptionists. Usually, the pre-op assessment nursing staff or ward sisters took the lead. One problem from a continuity and communication perspective was that staff often worked on a rotation basis, particularly on the wards and some people "job shared".

Most staff felt that patients were compliant and there was more positive feedback than negative feedback from patients about the study. No staff member found the process worse than they had envisaged when they had initially been invited to participate, although some said that they had foreseen time difficulties. They felt the most important factor in recruitment was staff attitude and that it was important for staff to build up a rapport with the patients and to work as a team. Staff often commented that it took a while to settle into the data collection procedure and for everyone to know what they were doing. Quite a few staff had previous experience of giving out questionnaires and they felt that patients were used to being given forms to complete (e.g. the health screening questionnaire). On the whole nursing staff were very supportive of the PROMs study but some did admit "others would not see the giving out of questionnaires as the most important thing they had to do". The fact that staff were not required to read out the questions or to assist the patients with completion of the questionnaires was mentioned as a positive factor by some. Tearing out hospital copies of the consent forms and information sheets (and in some cases tearing out a patient's copies) was an issue for many centres.

Some centres admitted that they probably failed to invite some patients to participate due to forgetfulness, time pressures, or a key person being on leave. Consequently if our main contact was absent for any reason, the system would often break down.

Reasons that patients chose not to participate
Often patients declined to participate because of a lack of time, because of the rushed nature of pre-operative assessment clinic procedures and the fact that patients often wanted to get home if the questionnaire was offered at the end of the assessment, or due to the tasks required on the actual day of surgery. A few patients suffered from pre-operative stress and were not given a questionnaire on the day of surgery. Patients also affected one another in terms of compliance, either positively or negatively. If one patient refused, this led to a chain of refusals in a couple of instances. This did not happen so much when patients' appointments were more staggered, there being less opportunity for patients to overhear one another.

Patients were not unduly concerned about giving out their home addresses, and were not upset that they could not complete the questionnaires at home, even though some questionnaires did go home with patients inadvertently from some centres. In general, patients did not raise many queries about participation. In fact most patients wanted to help out with the study and were extremely compliant. A few patients made specific comments such as "I do not want anything going to my house" and "can't you go anywhere without forms?" or "this is too private" but generally such remarks were few. Again, only a few patients commented on the length or thickness of the booklet – "too big", "too many boxes" or "too many questions" (hip patient). A patient requiring hernia surgery commented that "a number of the questions are very similar". Several of the patients requiring varicose vein surgery had commented that the drawing of the veins had put them off. Sometimes patients just left the questionnaires blank and put them to one side rather than specifically state that they did not want to take part. Also, some patients would skip certain questions.

One nurse said that she felt the key factor regarding completion was whether patients understood what they had to do. Many of the older patients had very poor memories. A receptionist from one

centre recruiting cataract surgery patients said she had noticed men were more reluctant than women to take part.

Local innovations
Some centres devised their own methods for ensuring all potentially eligible patients were identified, for example, preparing a list of eligible patients the night before from the theatre/ assessment list for the following day and putting a questionnaire in the patients' notes. However, staff from one centre mentioned that trauma cases would take priority over elective hip/knee surgery patients and therefore the list could change at short notice.

It was easier for staff to remember which operations to include if the clinic largely dealt with the specific operations to be covered in the PROMs study, e.g. a hip/knee/spine clinic or an eye clinic mainly doing cataract surgery rather than a pre-operative assessment clinic dealing with a range of surgical conditions at which only a couple of hernia/varicose vein surgery patients would attend on any one day.

Problems identifying ineligible patients
Staff had few problems identifying patients who could not understand English or who suffered from cognitive impairment. Occasionally a patient would tell the nurse privately that he/she was unable to read and/or write. Sometimes staff felt that a patient would refuse rather than admit to having a literacy problem but nursing staff said that it was usually down to common sense whether or not a patient would be able to participate. It was often written on a patient's pre-admission sheet whether or not they understood English. Staff from many centres stated that they had extremely few non-English patients and, even if English was not the patient's first language, they were still able to complete the questionnaires. One centre in the South West mentioned that they had a lot of Somali patients for whom they often had to book interpreters and another centre in the North West said they had a growing Polish community. A couple of nurses mentioned that they used body language to determine whether to offer a questionnaire to a patient or not and, again, it was obvious if a patient was too anxious. Many cataract patients were not offered a questionnaire because their vision was too poor, "they had cataracts that were just too hard". Most patients with severe visual problems said themselves that their vision was too poor for them to take part. One blind lady wanted to take part and so staff allowed her companion to read out the questions to her.

Staff from Treatment Centres admitted that they probably had extremely few patients with any sort of impairment as they tended to treat fitter patients who did not require any additional care.

Time pressures
Most centres felt that that giving out the questionnaires did not significantly impact on staff time, but there were some concerns about staff time should PROMs data collection become routine. Staff felt that it was the initial setting up of the PROMs data collection locally that was time-consuming, i.e. deciding the procedure and training staff. Most centres said it took about five (could be up to ten) minutes per patient to administer a questionnaire but occasionally staff felt that they did not have this time to spare. "If you are busy, five extra minutes is still time". In some cases, operating and clinic lists were delayed as a result of distributing the questionnaires. Certainly in day surgery, the first patient on the list would often be missed. In pre-operative assessment clinics, giving the questionnaires out to patients whilst they were waiting for test results or having an X-ray worked well. Staff often noticed this themselves and changed their method accordingly. If patients were given the questionnaires at the beginning of the assessment, it could delay the scheduling of appointments; if at the end, patients would just want to get home or they might be waiting for a lift. However, there is always the danger that if the questionnaires are given out at the start of the pre-op assessment, the operation may be postponed or cancelled further to the test results.

Another problem mentioned by several staff was actually remembering to offer a questionnaire to an eligible patient.

The length of time patients spent completing the questionnaires varied from five to thirty minutes. Some patients completed the questionnaires straight away, others would leave them to complete later. Those who took longer were older, the "not so well educated" or patients with visual or cognitive recall problems. The orthopaedic staff also pointed out that some patients were very arthritic. Also, orthopaedic patients tend to have a lot of other paperwork, e.g. NJR forms. It was pointed out that hip patients tend to be more elderly, whereas the knee patients tend to be younger and fitter. The extra length of the hip questionnaire was only occasionally raised as an issue by staff or patients. It was mentioned that a small number of patients failed to complete the questionnaires due to time pressures. Patients often asked staff for help completing the questionnaires and would discuss the questions with their relatives and with other patients.

Sometimes staff had no idea how long it took a patient to complete a questionnaire as patients were often sent back to a waiting area in reception to complete it and asked to hand it in to the receptionist before they left.

Confusion over questionnaire content
A small number of patients were sometimes unsure how to answer particular questions. These included the co-morbidity question (perhaps because they did not understand the names of certain conditions) or they had queries such as "I had this medical problem five years ago, should I write it down?" There was also a question relating to driving in the pre-operative cataract questionnaire which caused confusion. Several patients did not like having to draw their veins on the form as required in the AVVQ.

Several people mentioned that patients did not seem to want to keep their copy of the information sheet and consent form. However, it was said that patients generally were used to filling in forms.

Questionnaire design
Most staff felt that the questionnaires looked very professional and were very easy to understand. It was felt to be important that the Information Sheet was limited to one page as "patients will not read too many pages". Generally it was felt that the questionnaires were user friendly but that they "looked like a book" although it was accepted that the questionnaires were not as long as initially thought. Several people commented that some patients had not understood that they would be followed up in a few months' time and thought this should be made even clearer.

One nurse from an Eye Clinic felt that the lines were quite near together on the questionnaires. Another nurse commented that she thought it was quite a complicated form, and mentioned that the patients were often on drugs and had complicated histories. Her feeling was that even though patients agreed to complete the questionnaires, they found it irritating. All the staff at this centre said that very few patients could remember dates and histories. It was mentioned by a staff member that the questionnaires looked very long, although she admitted that none of the patients had mentioned the length of the questionnaire and that they had coped better than she had thought they would.

Value of pre-study visit, regular contact and study materials
All staff who had been present at the initial pre-study presentation said that they had found it very useful, not only in term of understanding the aims of the study, but also to "put a face to a name" and to feel they could get in touch with the lead researcher with any queries. Adherence to the study protocol by a centre was better when the lead researcher met staff at the initial meeting, but until local staff fully understood what they were required to do a decision had not always been made about the most appropriate place and people to distribute the questionnaires. Most people appreciated the regular contact and many said they would not have felt comfortable about participation if the initial visit had not taken place. A few people used the presentation to explain the methodology to other staff members who had not been present at the meeting. Staff who had not met the lead researcher initially said that they would have preferred to have been at the original meeting and gone through things at the presentation.

The study handbook had been useful to those people who had read or seen it but it was certainly not read by all staff. Several people mentioned that they had not seen the list of operations to be included, which explained why patients having hip and knee revisions were not always included at some centres. Some centres would have appreciated a supply of pens as they often had to stop and find pens for the patients to use and they found this irritating. One nurse thought a checklist for the patients would have been useful – e.g. (a) sign, (b) complete, (c) take away information sheet and consent form; another nurse suggested that a script should be provided of what staff should say to patients when they handed out the questionnaires. Two nurses felt the Handbook was too long and that a sheet of key bullet points would mean that staff would be more likely to read it. However, generally it was felt that there was nothing more the CEU could have done to support staff, except perhaps to provide a carbonated copy of the patient participation sheets for their records.

In regard to the cataract surgery questionnaire, it was suggested that it might be better to use black print on white or yellow paper although it was agreed that we had used the correct font size (14) according to RNIB recommendations. It was also agreed that the questionnaires were big enough for the patients to hold comfortably.

Are questionnaires the best way to obtain patient opinion?
Most staff felt there was no better way than to use questionnaires to obtain patient opinions. Some felt that questions also helped to prompt older people to think about specific issues and that the use of tick boxes was fairer as more people would be able to participate rather than just targeting those who could write more. However, one person mentioned that she did not think we would have had so many pre-operative questionnaires completed if they had just been posted out to patients.

Interviews were felt to be too time consuming and difficult and it would be hard to be precise. One nurse felt that for patients to give their opinion to a Consultant would probably be quite intimidating for them. There was a suggestion that some patients could complete the questionnaires on-line or use a palm top or a comments box. Tape recordings of the questionnaires could be made for the visually impaired.

Best time point and location to administer questionnaires
Staff were divided as to whether the appropriate place to give out the questionnaires at was pre-op assessment or around the surgical admission episode. Four people said they would change their mind if they did the study again, three from doing it 'on the day' to 'pre-op assessment' and one from 'pre-op assessment' to 'on the day'. Some people felt that there was already too much paperwork handed out at pre-operative assessment; others felt that there was more time then and patients would be less anxious than on the day of surgery. Pre-op was said to be "calmer" and it was usually a dedicated area within the centre. One factor against giving out the questionnaires at pre-operative assessment was the unknown time interval until surgery, which varied enormously between centres, from a couple of weeks to a few months. Time per patient allowed for pre-operative assessments varied between centres and depended on the type of operation. One sister of an Eye Clinic felt that pre-op assessment would most definitely be the wrong place to give out the cataract questionnaires as patients would not take it in then. It was also felt that more patients would take the questionnaires home with them from pre-op assessment (and this did seem to be borne out). Against giving out the questionnaires on the day of surgery was the fact that the first patient on the list would often be missed (because the list could not be delayed) and also that some patients might be too anxious.

It was also suggested that patients would be more likely to complete the questionnaires at pre-op assessment because they might be worried they might not get the operation if they did not. Therefore patients might exaggerate their symptoms or feel that if they were too negative in their answers it could affect their surgery. Patients might want to give something back to the nursing staff who had just spent time with them at pre-op assessment.

There was one suggestion that it might be better to give out the questionnaires at outpatients when patients are listed for surgery.

Staff best placed to administer questionnaires
Many interviewees felt that nursing staff should administer the questionnaires because they were already interacting with the patients and building up trust which would provide reassurance for the patient. Patients tend to respect nursing staff and nurses would know how to answer questions.

Reception staff were generally felt to be too busy to be involved in giving out questionnaires and it was thought that they may not be able to answer questions about the study. There was also a confidentiality issue as reception staff are not always privy to the type of operation a patient is to have. One centre actually invested some time during our study in educating the reception staff in regard to the names of various operations to be included in our study so that they would correctly identify eligible patients. Also, some times reception staff are volunteers. However, an official receptionist, who knew patients and staff, was felt to be fine. Sometimes, though, there is no receptionist in the pre-op assessment area.

Several others felt that it did not matter who gave out the questionnaires; what mattered was the manner/ attitude of the person giving out the questionnaires.

However, one staff member commented that if patient opinions were to be obtained from interviews, this would depend on the choice of interviewer. If it was a healthcare worker patients would give answers they thought staff wanted to hear. Therefore interviews had to be conducted by an independent person in order to be fair.

Should PROMs be administered continuously or on an intermittent basis?
The majority of staff felt that they would prefer to do the data collection for a short time each year, rather than continuously, even though they accepted that it would probably just become routine if it were built into their everyday schedules. However, there could be objections if there was a fast throughput of patients and several people were still concerned that time would be an issue. Staff would probably be more enthusiastic if they knew it was for a short period and this was confirmed by their willingness to participate in this pilot (because it was for a limited time). One Matron did not feel that short periods of data collection would be worthwhile and another nurse felt that 6-8 weeks should be the minimum period. There were a lot of comments that nurses were already used to giving out forms to patients, although many felt that there would too many audits being carried out. One centre carried out a surgical site infection audit for three months per year in trauma and orthopaedics, another said they carried out orthopaedic surveillance and another mentioned a pain audit. A few centres said that extra payment for participation would help.

Would there be any biases if recruitment was for a short period?
Most staff did not feel that there would be any biases if patients were recruited for only a short period each year, although issues like staff annual leave and sickness, and patients getting colds and/or not wanting to turn out in the winter time, were raised. One nurse felt that arthritic patients feel worse in the winter as they are affected by the weather. It was also mentioned that there was more capacity in the summer. It was also mentioned that labourers may put off hernia operations until after the summer and that patients may not come in during Ramadan. One other issue concerned the decline in volume of operations towards the end of the financial year.

Appendix 9: Patients' use of freephone line

No operation
Several patients called to say that their operation had not taken place, some because of factors concerning themselves and some because of cancellations by the centre. For example, three patients were due to have varicose vein surgery: one was found to have sugar in his blood, another had not had her operation as her husband had been ill and a third patient called to say her operation had been cancelled and she had now moved away from the area. One patient's knee operation had been postponed five times due to illness and cancellation and one patient's hip surgery was cancelled because it was thought she had a mass. (It turned out that she did not have a mass but it was still cancelled). One patient's hernia operation was cancelled due to a ward closure at the centre.

Patients sent elsewhere
One lady patient called to say that her varicose vein operation had taken place at another Hospital (rather than at the NHS Treatment centre) as she had asthma.

Another patient had an ECG at his pre-op assessment prior to his hernia operation which showed a blip on the heart and, as there was no resuscitation equipment at the ISTC, the operation did not go ahead there.

Operative complications
Some patients called to inform the CEU about post operative complications or what they felt was a lack of progress after the operation.

One patient enquired whether she was 'normal' in that she still did not feel she had recovered fully since her hernia operation. After her operation she spent over eight weeks in bed following a haematoma. Our survey had made her think about her lack of progress.

The surgeon performing a hernia operation on another patient found twisted veins in his lower gut during the operation. Consequently, the patient said he still had tenderness at 12 weeks. He wanted to point out that he thought our questionnaire was excellent but he would have liked space to comment on how impressed he was with the surgeon, anaesthetist and Hospital and to have been able to explain about the specific complications he had experienced.

A cataract patient called to say he had some leakage and had three stitches in his eye which were not due to come out yet.

One patient's blood pressure dropped during his hip operation and his heart had to be re-started. He had a catheter fitted and "they made a mess of it". He thinks this is what caused his hernia. He was happy with his hip operation and was doing fine until he had a hernia and since then his walking distance had been limited.

One patient said that after her hip operation she became agitated during the night and had a seizure and an embolism but she survived. She later saw her surgeon and he said "these things happen". She was not given any medication. She does not understand why it happened and nothing was said for nearly two months. Her hip operation, however, had been successful.

A patient undergoing knee replacement surgery said that she was in hospital with hyponatremia for three weeks and her condition had not been picked up by hospital staff.

One male hip patient had a transient ischaemic attack after the operation which impeded his progress.

Impact of other medical conditions

Some patients were concerned that other conditions they suffered from might affect their responses. Patients were advised to answer the questionnaires as accurately as they could without trying to relate their responses to a particular health condition or type of surgery.

For example, an ankle operation which might affect kneeling ability but had nothing to do with the patient's knee operation. This patient also mentioned that he had peritonitis after his knee operation and was in intensive care for a week. Another patient said she had her cataract operation but she didn't think she should complete the form as her sight was getting worse because of other eye problems and therefore it would give a misleading picture of the success of her operation. One male hernia patient was unsure how to answer the questions as he had a brain haemorrhage 18 years ago and was now getting new symptoms. He also has problems with arthritis. He was worried his answers would reflect these conditions and would be nothing to do with his hernia operation.

Waiting for a follow-up appointment or further treatment before responding
Some patients wanted to wait until after their follow up appointment to complete the questionnaire. They were advised not to do this and it was explained to them that we were interested in their (the patient's) current opinion rather than the consultant's.

One hernia patient said he had received our survey and hoped to complete it after he had more detailed information following his visit to the consultant in 3 weeks.

Three cataract patients rang, one to ask if she should wait until after her follow up appointment to fill in the post op questionnaire and another who mentioned that she was still having treatment. She had had a cold sore and had a viral infection. She was due to have her follow up appointment in 3-4 weeks time to get the all-clear. The other patient still had stitches in one eye and intended to wait until she had seen the Consultant before returning the questionnaire.

One patient explained that her hip had been fine until very recently and she was going back to see the Consultant. She was unsure about whether she should wait until then.

A knee patient said he was due to have manipulation on it and was worried that if he completed the questionnaire prior to that he would have to say the results were poor then but after the manipulation he might then feel better.

Time lag since operation
In about a dozen cases the full post-operative period (either three or six months) had not passed since the patient's surgery when they received the post-operative questionnaire. The first few patients were advised to complete the questionnaires and return them but when it was realised that this was becoming a common problem, patients were advised not to complete the form until the correct period had passed. We set up a reminder system and asked the patients who called to tell us about a delay to hold onto the forms and they then received a reminder telephone call/ letter at the appropriate time point.

One patient called in relation to his hernia questionnaire. He was enquiring about completion time. His operation was not a success and he would like to fill it in but he was extremely busy at the moment.

No consent
A lady telephoned on behalf of her father re our hernia questionnaire. She said he had not agreed to take part in the questionnaire and requested that we did not write to him again.

Patient too ill to respond
One patient's relative used the free phone to inform the CEU that the patient was too unwell to complete the cataract questionnaire. She explained that her Uncle had had a successful cataract

operation but he was very old and had been diagnosed with cancer of the oesophagus at around the same time and was now very ill. His sight has deteriorated further.

Putting additional notes/ other conditions
One knee patient said he would like to put a note at the bottom of his questionnaire as he felt there was something we should know.

One patient's husband called in relation to his wife's hip operation. He said he had sent his form back with a detailed note of what had happened.

Operation having taken place/ due on second eye/hip/knee/vein
One patient called to say that he was going in to have his second eye done shortly.

One patient explained that his first cataract operation had been a success but he had since had another since that time which had not been a success. He said he was having trouble reading the form. The patient called again and said he had thought about things and, as he had two follow up questionnaires, he was going to complete one for his left eye (his dominant eye) and one for his right eye. He did not think we would get a correct picture otherwise.

One patient said he has only had his varicose veins on his left leg done and was due to have the right leg done. Therefore he was unsure how to answer some of the questions.

Another patient said her hip operation was a success but she now had to have a knee operation and was worried that some of her answers would indicate that her hip operation was not a success but it was to do with her knee.

One male patient had his hip surgery done in April and was due to have his other hip done in the near future. He was concerned that his some of his answers in regard to mobility would be affected by the second hip.

One lady said she had had her hip replacement operation (and she was happy with it) but in the interim period she had also had a knee replacement operation. Therefore some of the questions in regard to her mobility would have nothing to do with her hip operation.

One patient said she was due to have another hip replacement in January.

In all cases, patients were advised to complete the questionnaires as best they could in regard to questions on mobility etc. but it was explained that in relation to the success of the operation question, we would just be referring to the how successful they felt the operation to which the pre-operative questionnaire related had been. They were advised that they could put a note on the questionnaire explaining the situation if they wished to.

Difficulties remembering date of operation
One male patient telephoned to say he could not remember the exact date of his cataract operation. [In this case, as the questionnaires were given out on the day of surgery at his centre, the consent date would be the day of the operation.]

Queries about questions
One patient called about the knee questionnaire. He did not feel there were enough questions relating to bending of the knee. When he went to see his Consultant afterwards, he only had a 90% bend and he was told not to overdo the exercises. He later saw another Doctor who said he should do exercises with a Physiotherapist as the bend had deteriorated. Therefore he felt he had been given the wrong advice initially. He later rang up again to ask about the anxiety and depression question. He said over the last two weeks he had been anxious because his wife had had a cancer scare, although she had now been given the all-clear. This was nothing to do with his knee and he was unsure how to answer this.

One patient called in relation to his hernia follow-up questionnaire. He was not sure what time point the previous month referred to. Also, he said he would like to have had a question referring to aftercare as he had not had any.

One patient called as she was initially unsure how to fill out the hip questionnaire because we were asking her how she felt 3-4 weeks ago and she had had her operation about six months ago but she said her husband had since explained it to her.

Wanting to comment on the Consultant
A patient called in relation to her varicose vein operation. She had returned our form but she has been very “disappointed” with her operation and still has to wear stockings for the rest of her life. “Still looks the same”. Did not feel her consultant had much time for her.