

patient-reported

**PATIENT-REPORTED
OUTCOME MEASUREMENT
GROUP, OXFORD**

**A STRUCTURED REVIEW OF
PATIENT-REPORTED OUTCOME
MEASURES FOR PATIENTS
UNDERGOING
CHOLECYSTECTOMY**

Report to the Department of Health, 2011



**O
U
T
C
O
M
E**

measures

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES FOR PATIENTS UNDERGOING CHOLECYSTECTOMY, 2011

Carolina Casañas i Comabella
Elizabeth Gibbons
Ray Fitzpatrick

Patient-reported Outcome Measurement Group
Department of Public Health
University of Oxford

Copies of this report can be obtained from:

Elizabeth Gibbons
Senior Research Officer
Patient-reported Outcome Measurement Group

tel: +44 (0)1865 289397

e-mail: Elizabeth.gibbons@dph.ox.ac.uk

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES FOR PATIENTS UNDERGOING CHOLECYSTECTOMY

CONTENTS:

EXECUTIVE SUMMARY	1
INTRODUCTION	3
METHODS	5
RESULTS: Generic Patient-reported Outcome Measures	8
RESULTS: Preference-based Patient-reported Outcome Measures	11
RESULTS: Condition-specific Patient-reported Outcome Measures	12
CONCLUSIONS AND RECOMMENDATIONS	15
Appendix A: Search strategy and sources	18
Appendix B: Search Terms	19
Appendix C: Appraisal criteria	20
Appendix D: Generic & cholecystectomy-specific PROMs	23
Appendix E: Methods of working, membership and conclusions of consensus panel	27
References	32

EXECUTIVE SUMMARY

Aim of the report

The aim of this report is to identify Patient-reported Outcome Measures (PROMs) which have been evaluated with patients undergoing an elective cholecystectomy procedure.

The methods of the review are described and the results of the searches including sources and search terms used to identify relevant published research. Details of this evidence are presented for generic health status measures, for preference-based measures and then for condition or procedure-specific PROMs. The review resulted in the identification of a short-list of PROMs which were presented to a multidisciplinary panel for comment. The review of the literature-based evidence and the comments of the multidisciplinary panel underpin final recommendations to the Department of Health.

Results

Two generic instruments, which have been evaluated with people undergoing cholecystectomy, were identified in this review:

1. SF-36
2. Nottingham Health Profile (NHP)

One preference-based measure was identified:

1. European Quality of Life Questionnaire (EQ-5D)

Four condition-specific measures were identified in this review:

1. Gastrointestinal Quality of Life Index (GIQLI)
2. Abdominal Surgery Impact Scale (ASIS)
3. Gallstone Impact Checklist (GIC)
4. Otago Gallstones Condition-Specific Questionnaire (CSQ)

Recommendations

Based on the volume of evaluations and good measurement and operational characteristics, both the SF-36 and the NHP are highlighted as promising PROMs for the evaluation of the quality of NHS services.

One preference-based measure was also identified: the EQ-5D. However promising this instrument has proved in other health conditions, the reader should note there is very limited evidence regarding patients undergoing cholecystectomy in English-speaking populations.

There is very limited evidence available in English-speaking populations and none of the condition-specific instruments clearly stands out as having considerably more supportive evidence. However, based on appraisal of evidence by the PROM Group, and taking into account ratings and comments from the panel, the Otago Gallstones Condition-Specific Questionnaire is worthy of consideration above the other condition-specific measures.

INTRODUCTION

Patient-reported outcome measures (PROMs) offer enormous potential to improve the quality and results of health services. They provide validated evidence of health from the point of view of the user or patient. They may be used to assess levels of health and need in populations, and in users of services they can provide evidence of the outcomes of services for the purposes of audit, quality assurance and comparative performance evaluation. They may also improve the quality of interactions between health professionals and individual service users.

Lord Darzi's Interim Report on the future of the NHS recommends that patient-reported outcome measures (PROMs) should have a greater role in the NHS (Darzi 2007). The new Standard NHS Contract for Acute Services, introduced in April 2008, included a requirement to report from April 2009 on patient-reported outcome measures (PROMs) for patients undergoing Primary Unilateral Hip or Knee replacements, Groin Hernia surgery or Varicose Vein procedures. Furthermore, Lord Darzi's report 'High Quality Care for All' (2008) outlines policy regarding payments to hospitals based on quality measures as well as volume. These measures include PROMs as a reflection of patients' experiences and views. Guidance has now been issued regarding the routine collection of PROMs for the selected elective procedures (Department of Health, 2008) and since April 2009, the routine collection of PROMs for the selected elective procedures has been implemented and is ongoing.

There are three broad categories of PROMs: generic health status, preference-based, and condition- or population-specific-measures. Generic instruments comprise items intended to be relevant to the widest range of patient conditions and the general population. Preference-based measures are also broad in content but additionally provide utilities or values regarding health (for use in, for example, cost-utility analyses of interventions). Condition-specific instruments are often more focused on a particular disease or health condition (for example, diabetes), a patient population (for example, older people), a specific problem or symptom (for example, pain), or a described function (for example, activities of daily living). For any given area of health, condition-specific instruments may have greater clinical appeal due to the inclusion of content specific to particular conditions, and the likelihood of increased responsiveness to interventions.

It has been recommended that a combination of a generic or utility measure with a specific measure be used in the assessment of patient-reported health outcomes, on the grounds that the complementary content of the two types of measure, when combined, should assess a full range of aspects of health relevant to the particular population concerned. However, consensus is often lacking as to which instrument to use for specific purposes and contexts (Garratt et al., 2002). Structured reviews of PROMs for specific health conditions or populations can provide guidance for selection. An evidence-based approach strengthens recommendations from these reviews.

Selection criteria have been defined for assessing the quality of existing PROMs (Streiner and Norman, 2003; McDowell, 2006; Fitzpatrick et al., 1998). These include measurement issues, such as reliability, validity, responsiveness and precision, as well as practical issues, such as acceptability and feasibility.

Cholecystectomy

Cholecystectomy is a common elective procedure carried out in the NHS across England and consists of the surgical removal of the gallbladder. The main reasons for this surgical procedure are:

- *Cholelithiasis*. This refers to the presence of gallstones in the gallbladder; however this condition can be asymptomatic and in this case it does not always require a cholecystectomy.
- *Cholecystitis*. When the gallbladder is inflamed or infected it becomes a painful condition that requires a cholecystectomy.

These conditions are generally painful and directly affect the quality of life (QoL) of the individual. Patients can experience symptoms differently, but most of the people suffering from gallstones who require elective cholecystectomy experience symptoms and limitations such as biliary colic, diet restrictions which impact greatly on daily activities, causing a reduction in Health Related Quality of Life (Chen et al., 2006).

Gallstones are crystalline bodies allocated in the biliary tract (i.e. gallbladder). Gallstones are formed by accretion or concretion of normal or abnormal bile components and can be of two types: *cholesterol gallstones* and *pigment gallstones*. Gallstones consist one of the most common causes of morbidity in the Western world, with an estimated incidence of symptomatic cholelithiasis of up to 2.17 per thousand inhabitants (Keus et al., 2009; Legorreta, 1993; Steiner, 1994).

There are three types of cholecystectomy procedures:

- Open cholecystectomy (OC) (incision of more than 8cm)
- Small-incision cholecystectomy (SIC) (incision of 8cm or less)
- Laparoscopic cholecystectomy (LC) (less than four small incisions)

Cholecystectomy is a very common elective procedure. The Hospital Episode Statistics (HES) reported that 58355 cholecystectomy procedures¹ took place in England during the HES financial year 2007-2008. Of these, 22% were performed during an emergency admission, and 25% were day case cholecystectomies. The mean length of stay for cholecystectomy for patients over 69 years old, or patients who presented complications, was 4.9 days, whereas the mean stay for patients younger than 70 years old or without complications was 2.1 days.

All methods of cholecystectomy affect the patient's QoL both in the pre-operative and post-operative situation, and this needs to be assessed. PROMs can provide valuable information about the patients' perspective of their health and experiences of services.

¹ Based on Health Resource Group (HRG) combining G13 (cholecystectomy >69 or with complication) and G14 (cholecystectomy <70 without complications).

METHODS

Structure of the report

The methods of the review are described, including search strategies and search terms used to identify specific published research regarding PROMs for cholecystectomy operations. The report focuses on evidence and recommendations for PROMs. Evidence is presented for generic, preference based and those PROMs specific to people undergoing elective cholecystectomy.

Methods for the review

Methods adopted were as described in previous reviews performed by the PROM group, Oxford. Comprehensive searches were conducted; articles retrieved were assessed for relevance and evidence of measurement performance and operational characteristics abstracted for each PROM identified.

a) Search sources and terms

Several sources were searched to identify relevant articles.

The primary source of evidence was the bibliographic database compiled by the PROM group in 2002 with funding from the Department of Health and hosted by the University of Oxford². In 2005, it became the property of the NHS Information Centre for Health & Social Care. The PROM database comprises 30,350 records up to the end of 2006 (16,054 online up to December 2005) downloaded from several electronic databases using a comprehensive search strategy (available on request). These records had been assessed as eligible for inclusion in the bibliography and assigned keywords. These records were searched using the keywords 'gallstone', 'gall bladder', 'gallbladder' and 'cholecystectomy', and combinations of keywords ('surgery', 'gastrointestinal', 'hepatology'), and by instrument identified during the review. The search strategy used in this review and the flowchart illustrating the search process are detailed in Appendix A and B, respectively.

Supplementary searches included scanning the reference lists of key articles, checking instrument websites, where found, and drawing on other bibliographic resources.

Hand searching of titles of key journals from May 2009 was conducted. The following journals were selected:

- Health and Quality of Life Outcomes
- Quality of Life Research
- Surgical Endoscopy
- American Journal of Surgery
- British Journal of Surgery
- Surgery

² Available online at <http://phi.uhce.ox.ac.uk>

The National Institute for Health Research: Health Technology Assessment Programme published research was also searched.

In addition, PubMed records for the past three years (2007-2009) were searched using a comprehensive search developed by the PROM group and Outreach Librarian at the University of Oxford. See Appendix B for the detailed search strategy.

b) Inclusion criteria

Published articles were included if they provided evidence of measurement and/or practical properties of relevant PROMs (Fitzpatrick et al., 1998).

Population

- Patients undergoing any kind of open cholecystectomy (OC)
- Patients undergoing any kind of small-incision cholecystectomy (SIC)
- Patients undergoing any kind of laparoscopic cholecystectomy (LC)
- Patients with cholelithiasis
- Patients with cholecystitis
- English speaking populations

Study design selection

- Studies where a principal PROM is being evaluated
- Studies evaluating several PROMs concurrently
- Trials or studies evaluating the effectiveness of interventions; where a PROM is used as an endpoint.
- Prospective studies measuring patient-reported outcomes where data is available for a PROM in terms of measurement performance or operational characteristics

Specific inclusion criteria for generic, preference- and disease-specific instruments

- The instrument is patient-reported
- There is published evidence of measurement reliability, validity or responsiveness following completion in the specified patient population
- Evidence is available from English language publications, and instrument evaluations conducted in populations within UK, North America, or Australasia.
- The PROM ideally will be multi-dimensional. It is at the reviewer's discretion to include instruments which are specific to a health condition but have a narrow focus, for example a specific dimension of health, such as, symptoms.

Exclusion criteria

- Clinician-assessed instruments
- Studies evaluating the performance of non-patient reported measures of functioning or health status where a PROM is used as a comparator indicator

c) Data extraction

For all PROMs included in the review, evidence is reported for the following measurement criteria:

- reliability
- validity
- responsiveness
- precision

Operational characteristics such as patient acceptability and feasibility of administration for staff are also reported.

d) Assessment of methodological quality of PROMs

Assessment and evaluation of the PROMs was performed by using the appraisal criteria described in Appendix C.

Searches identified over 700 potentially relevant records after excluding the duplicates. When assessed against the review inclusion criteria, 14 articles were included in the review (Table 1).

Table 1: Number of articles identified by the literature review

Source	Results of search	Number of articles included in review
PROM database: 30,350	43	5
Pubmed (2007-2009)	671	1
Hand searching	-	8
TOTAL	-	14

RESULTS

A total of seven PROMs for patients undergoing cholecystectomy were identified in this review. Two generic measures were identified:

1. SF-36
2. Nottingham Health Profile (NHP)

The European Quality of Life Questionnaire (EQ-5D) was the only utility measure found.

Finally, four condition specific instruments were included:

1. Gastrointestinal Quality of Life Index (GIQLI)
2. Abdominal Surgery Impact Scale (ASIS)
3. Gallstone Impact Checklist (GIC)
4. Otago Gallstones Condition-Specific Questionnaire (CSQ)

GENERIC MEASURES

1) SF-36

The SF-36 is a generic health status instrument which was developed during the 1970s deriving from the work of the Rand Corporation's Medical Outcomes Study (MOS) (Ware, 2000). It contains 36 items covering 8 dimensions: Physical Functioning (PF: 10 items), Role Limitations due to physical health problems (RP: 4 items), Bodily Pain (BP: 2), Social Functioning (SF: 2), General Mental Health (MH: 5), Role Limitations due to emotional problems (RE: 3), Vitality (V: 4) and General Health Perceptions (GH: 5). Two component scores are derived: Physical Component Score (PCS) and Mental Component Score (MCS). The questionnaire can be self-, interview- or telephone-administered, and it takes five to ten minutes to complete (McDowell, 2006).

Five studies were identified which had used the SF-36 with people undergoing cholecystectomy in English-speaking populations. Two of these are UK-based studies (McMahon et al., 1994; Mallon et al., 2006).

Discriminative validity was reported by Burney & Jones (2002) in a trial with 140 patients undergoing either ambulatory or short-stay laparoscopic cholecystectomy. Patients who were admitted for short-stay LC reported significantly poorer scores in PF ($p = 0.005$), RP ($p = 0.039$) and RE ($p = 0.028$) at two months. At six months, results were similar, with relative functional deficits persisting (Burney & Jones, 2002).

Velanovich (2000) carried out a study assessing various surgical procedures, including 30 patients who underwent cholecystectomy. Statistically significant improvements were noted between the median surgery preoperative scores and the postoperative laparoscopic scores in PF (85 vs. 95, $p < 0.05$), BP (42 vs. 75, $p < 0.05$) and V (47.5 vs. 70, $p < 0.05$). Changes in other dimensions were not statistically significant when comparing pre and postoperative scores. The SF-36 was also

sensitive detecting a significant difference between open and laparoscopic cholecystectomy in BP as would be expected following a more invasive procedure (70 vs 41, $p=0.05$) (Velanovich, 2000).

There were no statistically significant differences in any of the eight dimensions of the SF-36 between patients undergoing OC and LC in the period from 1998 to 2000 (Mallon et al., 2006).

The SF-36 was able to detect change in a study where McMahon et al. (1994) reported an improvement of PF (57 in LC vs. 43 in minilaparotomy; $p<0.05$), SF (74 vs. 67; $p <0.05$), less RP (13 vs. 0; $p < 0.01$) and less BP (67 vs. 44; $p < 0.001$) at one week. At four weeks, SF-36 showed an improvement in the dimension of PF (67 in LC vs. 60 in minilaparotomy; $p < 0.05$).

In a study by Jones et al. (1998), the SF-36 proved to be able to detect statistically significant improvement in the scales of BP (at 2 months after surgery), V and improved SF (6 months), however scores were not provided.

The SF-36 questionnaire has been often used to assess QoL pre and post cholecystectomy and has high patient acceptability. A response rate of 78% was reported in a postal survey by McMahon et al. (1994) one week after cholecystectomy, 88% four weeks after and 81% after 12 weeks (preoperative measurements were not reported). Response rate was 67.7% in a postal survey investigating the increased cholecystectomy rates between 1998-2000 and 1988-1990 (Mallon et al., 2006).

Some studies have used modified versions of the SF-36 questionnaire. Jones et al. (1998) used a large-type version to improve its readability, together with expanded instructions in a study with a 90% participation rate (patients completed the questionnaire preoperatively while they stayed in hospital for the scheduled surgery), and a response rate of 76% (postal survey at 2 months) and 69% (postal survey at 6 months). However in this study, patients often left items, sections or the entire questionnaire blank, or even made notes on it explaining they decided not to complete it because they feared the responses may be interpreted as a failure of the surgical procedure. Instructions had to be re-written to address this problem This study also found two patterns. Both females and males left many of the physical activity items blank because they never performed such activities (i.e. lifting groceries or engage in vigorous activities).

Another study modified the SF-36. McMahon et al. (1994) modified the SF-36 by removing questions 9 and 10 (relating to general mental health) in order to avoid repetition of similar questions in the concurrent questionnaire used, the Hospital Anxiety and Depression Scale (HADS), in a study comparing laparoscopic versus minilaparotomy cholecystectomy with 302 patients.

No evidence was found in English-speaking population studies regarding feasibility, precision, interpretability, appropriateness or reliability of SF-36 in relation to cholecystectomy.

2) Nottingham Health Profile (NHP)

The Nottingham Health Profile (NHP) was developed in the UK with the aim of assessing lay people's perceptions about their health status (Hunt et al., 1986). It is a patient-reported outcome measure which is structured in two parts, which can be used separately. Part I consist of 38 items, each of them being a simple statement regarding subjective health status and with a yes/no answer option. It covers six dimensions: Physical Abilities (PA, 8 items), Pain (P= 8 items), Energy Levels (EL=3), Sleep (S=5), Emotional Reactions (ER=9) and Social Isolation (SI=5), and each dimension has a range of possible scores of 0 (no problems) to 100 (all items checked). Part II consists of seven items assessing different aspects of daily life with a yes/no answer option, and is less widely used because some items referring to work, holidays, etc, may not be applicable (McDowell, 2006).

This questionnaire can be self-, interview-, or telephone-administered, and it takes around 10 minute to complete.

This review identified five studies which used the NHP with people undergoing cholecystectomy in English speaking populations. Four of these studies were UK-based (Bardsley et al., 1992; Nicholl et al., 1992, 1994; Squirrell et al., 1998).

The NHP detected significant changes pre and post cholecystectomy for EL, P, S and ER in 154 patients (Bardsley et al., 1992). NHP was also responsive to change in a group of 70 patients undergoing either LC or mini cholecystectomy one month after surgery and in the mini cholecystectomy group after 3 months (Barkun et al., 1992).

The NHP was used by Squirrell et al. (1998) to assess QoL after LC and SC in 100 patients. The questionnaire was completed at baseline (preoperatively), 3 weeks and 6 months after operation. NHP was responsive to change in the Pain and Sleep scores ($p < 0.001$ for both LC and SC) after operation. The Emotional and Physical mobility dimensions improved significantly within NHP after SC (Emotional $p = 0.013$ and Physical mobility $p = 0.002$). Social isolation scores improved after LC ($p = 0.045$). Scores for energy improved in both groups but not statistically significantly.

One study used the NHP to evaluate effectiveness of not cholecystectomy, but Extracorporeal Shock Wave Lithotripsy (ESWL)³ as a treatment for gallbladder stones in 50 patients (Nicholl et al., 1994). Measurements were taken at baseline (four weeks before treatment) and two weeks, five weeks, three months, six months, and 12 months after treatment. NHP proved to be sensitive to detect change at two weeks and thereafter little change. The dimensions that improved more significantly after surgery were Energy, Pain and Emotional reactions.

Good response rates were obtained in the study by Bardsley et al. (1992): 95% at three months and 90% at 12 months to the follow up postal questionnaire. Squirrell et al. (1998) used only the first part of NHP preoperatively, at 3 weeks and at 6 months after the operation.

³ Extracorporeal Shock Wave Lithotripsy (ESWL) is a non-invasive treatment for gallstones amongst other conditions which was widely used during the late eighties and early nineties, in combination with oral bile salts (Darzi et al., 1994). However, high recurrence of the condition when using this technique, together with the perfecting of cholecystectomy procedures, have limited the use of ESWL to cases that are not suitable for surgery (Paumgartner & Sauter, 2005).

Of the 300 assessments in the Nicholl et al. (1994) study (6 assessments per each of the 50 participants), only 3 were missed. In a few cases patients also failed to answer all the questions at an assessment.

No evidence was found on appropriateness, reliability, validity, precision, interpretability, acceptability or feasibility in English-speaking population studies regarding NHP in relation to cholecystectomy.

PREFERENCE BASED MEASURES

EQ-5D

The European Quality of Life instrument (EuroQol), now generally known as the EQ-5D, was developed by researchers in five European countries as a measure with a core set of generic health status items, based on existing PROMs (The EuroQol Group, 1990; Brazier et al., 1993). It was intended that, in application, the EQ-5D would be supplemented by disease-specific instruments. The developers recommend the EQ-5D for use in evaluative studies and policy research; it can also be used for economic evaluation. The measure can be self or interview-administered.

There are two sections to the EuroQol: the five-dimension index and the EQ ‘thermometer’. The EQ-5D assesses health across five domains, namely Anxiety/Depression (AD), Mobility (M), Pain/Discomfort (PD), Self-Care (SC), and Usual Activities (UA); each domain has one item and a three-point categorical response scale. Weights based upon societal valuations of health states are used to calculate an index score of -0.59 to 1.00 , where -0.59 is a state worse than death and 1.00 maximum well-being; a score profile can also be reported. The EQ thermometer is a single 20-cm vertical visual analogue scale with a range of 0 to 100, where 0 is the worst and 100 the best imaginable health.

This review identified only one study using EQ-5D in relation to cholecystectomy in an English-speaking population. However the reader should note another UK-based study was identified which used EQ-5D in order to calculate QALYs after laparoscopic cholecystectomy (Macafee et al., 2009). The EQ-5D scores provided in this study show no statistically significant difference between early and delayed cholecystectomy in a total of 67 patients.

Some limited evidence of responsiveness of the EQ-5D and VAS was reported in a study comparing quality of life outcomes between conventional laparoscopic cholecystectomy (CLC) and micropuncture laparoscopic cholecystectomy (MPLC) in 40 patients in the United Kingdom (Ainslie et al., 2003). Measurements were made preoperatively, then at 8h, 24h, one week and one month postoperatively. Change was observed within each group at 8 hours and 24 hours postoperatively with decrease in scores for EQ-5D and VAS, but this returned to baseline scores by one month. This would be in accordance with hypothesis as patients within 24 hours of surgery are likely to report worse health status.

CONDITION-SPECIFIC MEASURES

Four condition specific instruments were included:

1. Gastrointestinal Quality of Life Index (GIQLI)
2. Abdominal Surgery Impact Scale (ASIS)
3. Gallstone Impact Checklist (GIC)
4. Otago Gallstones Condition-Specific Questionnaire (CSQ)

1) **Gastrointestinal Quality of Life Index (GIQLI)**

The Gastrointestinal Quality of Life Index (GIQLI) is a Health-Related Quality of Life (HR-QoL) questionnaire which was developed in the early nineties and evaluated in German and Canadian populations (Barkun et al., 1992; Eypasch et al., 1993; Eypasch et al., 1995). It is a 36 item self-administered questionnaire aimed to evaluate QoL of patients with gastrointestinal disease, covering four dimensions: Physical, Emotional, Social and Symptoms. The developmental study of GIQLI (Eypasch et al., 1995), which was carried out in Germany, showed good validity, reliability and internal consistency. The GIQLI also showed good responsiveness: it detected statistically significant change before and after the operation, especially on the first postoperative measurement two weeks after surgery. However the reader should note these results are drawn from non-English speaking population. Despite not being a gallbladder-specific questionnaire, the GIQLI has been used to assess the QoL of patients undergoing cholecystectomy in one study.

Furthermore, the GIQLI was developed simultaneously in German and English, and this review found only one study which used GIQLI to evaluate cholecystectomy in an English-speaking population.

Responsiveness was reported in a clinical trial which evaluated QoL in 70 patients undergoing either laparoscopic cholecystectomy or mini cholecystectomy in Canada (Barkun et al., 1992). Measurements were taken preoperatively, 1 month and 3 months after cholecystectomy. The GIQLI was sensitive enough to detect significant improvement in both patient groups at one month after surgery ($p > 0.05$).

2) **Abdominal Surgery Impact Scale (ASIS)**

The ASIS was recently developed in 2006 by Urbach et al. (2006) in Canada, with the aim of designing a measure of HRQoL specifically after abdominal surgery, for use as an outcome measure in studies comparing laparoscopic and conventional abdominal surgery. After item reduction, the final result was an 18-item questionnaire covering 6 domains (Physical Limitations, Functional Impairment, Pain, Visceral Function, Sleep, Psychological Function).

This review found only one study that used this PROM to evaluate cholecystectomy procedures. Urbach et al. (2006) administered the prototype to 500 patients undergoing both conventional and laparoscopic abdominal surgical procedures. Most of the laparoscopic operations were cholecystectomies (exact numbers not provided).

The hypothesised factor structure of 6 domains was empirically confirmed by factor analysis in Urbach et al. (2006). However the authors admit that further research is also necessary to assess construct validity.

The reliability coefficients were good for five of the scales, with Cronbach's α ranging from 0.80 (Sleep) to 0.90 (Functional impairment) (Urbach et al., 2006). However, internal consistency of the Visceral function subscale was found to be low. The authors explain this by the fact that this scale was constructed from heterogeneous concepts such as gastrointestinal function, urinary function, cough, or thirst. However this does not mean that this scale would not be reliable according to other measures, for example test-retest reliability.

3) Gallstone Impact Checklist (GIC)

The Gallstone Impact Checklist (GIC) was developed in 1996 by Russell et al. in Canada. This is a disease-specific QoL questionnaire covering four dimensions: Pain, Dyspepsia, Emotional Impact and Food and Eating. It contains a 41-item checklist with a yes/no answer option, and, for each relevant item, a visual analogue scale in which to mark the extent to which the item was problematic. Perceived health was assessed using the question "Generally speaking, compared to other persons your age, would you rate your health as excellent, very good, good, fair or poor?". Finally, global perceptions of quality of physical and emotional well-being were measured on a 10-point ladder scale (0 representing the worst state and 10 representing the best).

The questionnaire was designed to be self-administered and it takes around 10 to 15 minutes to complete.

Sixty-seven participants completed the baseline questionnaire in a trial by Russell et al. (1996), and 57 completed the follow-up. The questionnaire was administered at baseline (preoperatively) and four to six weeks later.

Internal consistency was assessed using Cronbach's α , which was overall 0.88 and found to be acceptable. For the Pain subscale, Cronbach's α was 0.60; it was 0.73 for Dyspepsia; 0.78 for Emotional impact, and 0.84 for the Food and eating subscale. Temporal stability was assessed amongst the 9 individuals in this study who did not feel any improvement on their gallstone condition after the second completion of the questionnaire, showing good test-retest reliability. However, this is a small number which does not allow accurate assessment.

The instrument proved to be responsive to change, with an absolute value of effect size of 1.63 in patients who had undergone surgery, in contrast to the 0.40 for those who had not (Russell et al., 1996).

Chen et al. (2006) have pointed out some limitations of the GIC, arguing its lengthiness and complexity for use in routine settings.

No further recent evaluations have been identified.

4) Otago Gallstones Condition-Specific Questionnaire (CSQ)

The Otago Gallstones Condition-Specific Questionnaire was developed by Chen et al. in 2006 in New Zealand, arising from concerns about the GIC (see above), such as feasibility and acceptability issues for use on routine settings (Chen et al., 2006). The authors devised a conceptual model for gallstone-specific QoL, with four underlying domains: Physical Functioning (pain, dyspepsia and diet changes), Systemic Functioning (fatigue), Social Functioning (daily duties, leisure, relationships) and Emotional Functioning (mood).

The Otago gallstones condition-specific questionnaire (CSQ) contains 12 items, each with a 5-point Likert response scale. The final score is obtained using the simple summation method and converting it into a percentage score, where 100 is the most severe disease state. At the end of the questionnaire, a free text area is provided for the patient to add any concerns, comments or suggestions they felt were not captured by the structured reviews (Chen et al., 2006).

The questionnaire was administered to 54 patients (77 % women), alongside with the SF-36 as a generic QoL measure. This was followed by a semistructured interview with a researcher, seeking the patient's opinion on the design, content and suitability of the CSQ. The Gallstone Impact Questionnaire (GIC) was administered as a reference for comparison.

CSQ had good internal consistency, with Cronbach's $\alpha = 0.94$.

Clinical relevance was also evaluated through the relationship of the scale to the surgeon-rated patient priority for surgery (CPAC)⁴, which showed a significant moderate correlation with average $r = 0.62$ ($p < 0.05$).

Patients endorsed the relevance of the CSQ questions, and felt they managed to capture all QoL issues related to gallstone disease (content validity). There was strong correlation of scores with the global condition impact (GR) and the Bodily pain and Social Function dimensions of SF-36.

No floor/ceiling effects were reported; scores were evenly distributed among categories (Chen et al., 2006).

The authors reported a return rate of 93% for the questionnaire mailed at 4 weeks (Chen et al., 2006).

The questionnaire takes less than 15 minutes to complete and patients found it easy to understand and simple to complete, with an average time of 2.7 min (Chen et al., 2006). Completion rate was 100% with no missing values.

⁴ The Clinical Priority Assessment Criteria (CPAC) is a system used by clinicians to prioritise patients. A score is calculated using a range of dimensions (clinical health status, patient-experienced health status, and social), providing a score between 0 to 100, being 100 highest priority).

Comment

The reader should note that this review also found a number of scales measuring pain: Visual Analogue Scale (VAS) and the McGill Pain Questionnaire. However, these two instruments were considered to be non-specific to either abdominal surgery or cholecystectomy and therefore no further comment is made.

Cleary et al. (1991) used a generic self-report instrument that appears to be highly similar to the Functional Status Questionnaire (FSQ) to assess QoL in a sample of 476 cholecystectomy patients, amongst other surgery patients. However this scale seemed to lack sensitivity to change in cholecystectomy patients (Cleary et al., 1991).

SUMMARY OF EVIDENCE

The full-text articles for 72 papers were retrieved and reviewed. Those papers describing studies not used in English-speaking populations or symptom checklists or clinician ratings were excluded. Disappointingly, there were only 14 papers identified that provided useful data for this review.

Two generic, one preference-based and four condition-specific PROMs were identified that had been evaluated with patients undergoing cholecystectomy, or suffering from gallstones, in English-speaking populations (Tables 2 and 3).

The SF-36 (2 UK) and NHP (4 UK) both present good evidence in assessing general QoL in patients undergoing cholecystectomy in the UK. It should be noted that some studies have used modified versions of SF-36 adapted for this population.

The EQ-5D has been evaluated in many different health conditions, however the evidence found in this review regarding cholecystectomy is very limited and extrapolating from the results of the study included may indicate a lack of responsiveness.

This review identified four condition-specific instruments; however there is very limited evidence carried out in English-speaking populations firmly supporting any of them. Two of these instruments (GIQLI and ASIS) are region-specific: specific to gastro intestinal conditions and abdominal surgery. The other two (GIC and CSQ) are gallstone-specific.

The GIQLI does present more evidence in relation to non-English speaking populations, but disappointingly, little for English-speaking, and its general focus limits its sensitivity for patients undergoing cholecystectomy procedures.

RECOMMENDATIONS

For the measurement of comprehensive general health status, the SF-36 has more psychometric supportive evidence and is acceptable to patients and can therefore be recommended.

Despite very little encouraging evidence, the EQ-5D is widely used in the UK context and should be considered further for the assessment of general health in patients undergoing cholecystectomy.

There is very limited evidence available in English-speaking populations and none of the condition-specific instruments clearly stands out as having considerably more supportive evidence. However, based on appraisal of evidence by the PROM Group, and taking into account ratings and comments from the panel, the Otago Gallstones Condition-Specific Questionnaire (CSQ) is worthy of consideration above the other condition-specific measures.

Table 2: Appraisal of psychometric and operational performance of generic PROMs used in cholecystectomy.

PROM (n of studies)	Reproducibility	Internal consistency	Validity: Content	Construct	Responsiveness	Interpretability	Floor/ceiling/precision	Acceptability	Feasibility
SF-36 (5)	0	0	0	+	++	0	0	+	0
NHP (5)	0	0	0	0	++	0	0	+	0
EQ-5D (1)	0	N.A.	0	0	-	0	0	0	0

0 not reported — no evidence in favour + some limited evidence in favour ++ some good evidence in favour +++ good evidence in favour.

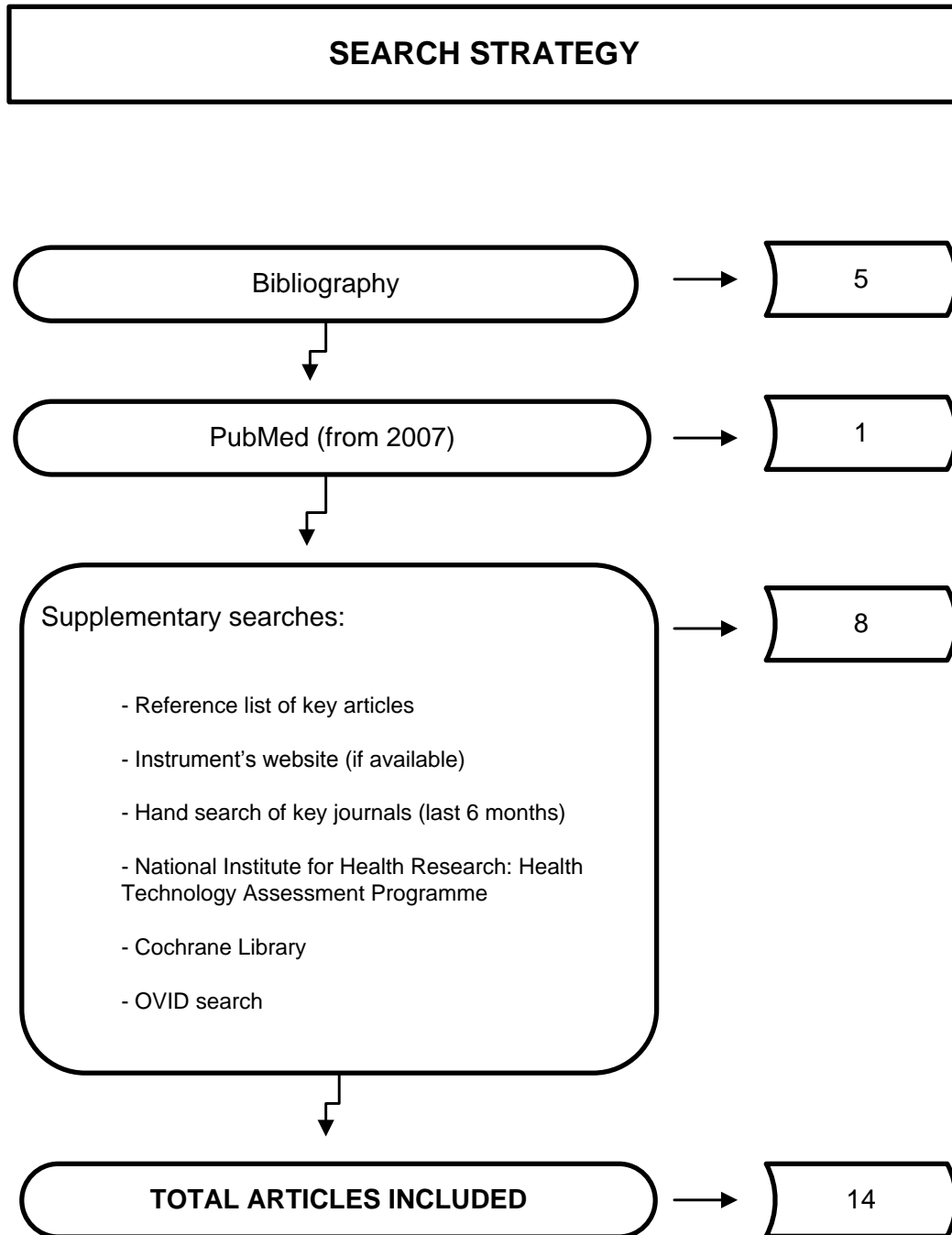
Table 3: Appraisal of psychometric and operational performance of cholecystectomy-specific PROMs.

Instrument (n of studies)	Reproducibility	Internal consistency	Validity: Content	Construct	Responsiveness	Interpretability	Precision	Acceptability	Feasibility
GIQLI (1)	0	0	0	0	+	0	0	0	0
ASIS (1)	0	+	0	+	0	0	0	0	0
GIC (1)	0	+	0	+	+	0	0	+	0
CSQ (1)	0	+	+	+	0	0	+	+	+

0 not reported — no evidence in favour + some limited evidence in favour ++ some good evidence in favour +++ good evidence in favour.

APPENDIX A:

SEARCH STRATEGY AND SOURCES



APPENDIX B:

SEARCH TERMS USED IN PUBMED:

Search ("2007"[Publication Date] : "3000"[Publication Date])

AND

((cholelithiasis[tiab] OR cholecystitis[tiab] OR gallstone*[tiab] OR gall bladder[tiab] OR gallbladder[tiab] OR cholecystectomy[tiab])

AND

(HR-PRO[tiab] OR HRPRO[tiab] OR HRQL[tiab] OR HRQoL[tiab] OR QL[tiab] OR QoL[tiab] OR quality of life[tw] OR life quality[tw] OR health index*[tiab] OR health indices[tiab] OR health profile*[tiab] OR health status[tw] OR ((patient[tiab] OR self[tiab] OR child[tiab] OR parent[tiab] OR carer[tiab] OR proxy[tiab]) AND ((report[tiab] OR reported[tiab] OR reporting[tiab]) OR (rated[tiab] OR rating[tiab] OR ratings[tiab]) OR based[tiab] OR (assessed[tiab] OR assessment[tiab] OR assessments[tiab]))) OR ((disability[tiab] OR function[tiab] OR functional[tiab] OR functions[tiab] OR subjective[tiab] OR utility[tiab] OR utilities[tiab] OR wellbeing[tiab] OR well being[tiab]))

AND

(index[tiab] OR indices[tiab] OR instrument[tiab] OR instruments[tiab] OR measure[tiab] OR measures[tiab] OR questionnaire[tiab] OR questionnaires[tiab] OR profile[tiab] OR profiles[tiab] OR scale[tiab] OR scales[tiab] OR score[tiab] OR scores[tiab] OR status[tiab] OR survey[tiab] OR surveys[tiab])))

Limits: English

APPENDIX C:

APPRAISAL OF THE METHODOLOGICAL QUALITY OF PROMs

The methods that will be used for assessing the performance of PROMs were developed and tested against multi-disciplinary consensus and peer review. They focus on explicit criteria to assess reliability, validity, responsiveness, precision, acceptability and feasibility. A pragmatic combination of the criteria developed and used in previous reports to DH by the Oxford and LSHTM groups will be used.

The appraisal framework focuses on psychometric criteria and PROMs must fulfil some or all to be considered as a short-listed instrument. Practical or operational characteristics are also assessed (acceptability and feasibility) (Appendix D: Appraisal framework).

Psychometric and operational criteria	
0	not reported (no evaluation completed)
-	Evaluation evidence available indicating poor performance of instrument
+	Some limited evidence in favour
++	Good evidence in favour
++ +	Excellent evidence in favour

Appraisal criteria (adapted from Smith et al., 2005 and Fitzpatrick et al., 1998; 2006):

Appraisal component	Definition/test	Criteria for acceptability
Reliability		
Reproducibility/Test-retest reliability	The stability of a measuring instrument over time; assessed by administering the instrument to respondents on two different occasions and examining the correlation between test and re-test scores	Test re-test reliability correlations for summary scores 0.70 for group comparisons
Internal consistency	The extent to which items comprising a scale measure the same construct (e.g. homogeneity of items in a scale); assessed by Cronbach's alpha's and item-total correlations	Cronbach's alphas for summary scores ≥ 0.70 for group comparisons Item-total correlations ≥ 0.20
Validity		
Content validity	The extent to which the content of a scale is representative of the conceptual domain it is intended to cover; assessed qualitatively during the questionnaire development phase through pre-testing with patients. Expert opinion and literature review	Qualitative evidence from pre-testing with patients, expert opinion and literature review that items in the scale represent the construct being measured Patients involved in the development stage and item generation
Construct validity	Evidence that the scale is correlated with other measures of the same or similar constructs in the hypothesised direction; assessed on the basis of correlations between the measure and other similar measures	High correlations between the scale and relevant constructs preferably based on a priori hypothesis with predicted strength of correlation

Construct validity (continued)	The ability of the scale to differentiate known-groups; assessed by comparing scores for sub-groups who are expected to differ on the construct being measured (e.g a clinical group and control group)	Statistically significant differences between known groups and/or a difference of expected magnitude
Responsiveness	The ability of a scale to detect significant change over time; assessed by comparing scores before and after an intervention of known efficacy (on the basis of various methods including t-tests, effect sizes (ES), standardised response means (SRM) or responsiveness statistics	Statistically significant changes on scores from pre to post-treatment and/or difference of expected magnitude. The recommended index of responsiveness is the effect size, calculated by subtracting the baseline score from the follow up score and dividing by the baseline SD. Effect sizes can be graded as small (<0.3), medium (~0.5), or large (>0.8).
Floor/ceiling effects	The ability of an instrument to measure accurately across full spectrum of a construct	Floor/ceiling effects for summary scores <15%
Practical properties		
Acceptability	Acceptability of an instrument reflects' respondents' willingness to complete it and impacts on quality of data	Low levels of incomplete data or non-response
Feasibility/burden	The time, energy, financial resources, personnel or other resources required of respondents or those administering the instrument	Reasonable time and resources to collect, process and analyse the data.

APPENDIX D: GENERIC AND CONDITION-SPECIFIC INSTRUMENTS

GENERIC INSTRUMENTS

Generic patient-reported health instruments

Instrument (no. items)	Domains (no. items)	Response options	Score	Administration (completion time in minutes)
European Quality of Life instrument (EuroQol, EQ-5D) (5+1)	Anxiety/depression (1) Mobility (1) Pain/discomfort (1) Self-care (1) Usual activities (1) EQ-thermometer: Global health (1)	Categorical: 3 options EQ-thermometer VAS Recall: current health	Summation: domain profile Utility index (-0.59 to 1.00) EQ-Thermometer VAS (0-100)	Interview or self
Nottingham Health Profile (NHP) (38)	Bodily pain (BP) (8) Emotional reactions (ER) (9) Energy (E) (3) Physical mobility (PM) (8) Sleep (S) (5) Social isolation (SI) (5)	Yes/no; positive responses weighted Recall: 'general' health	Algorithm Domain profile 0-100, 100 is maximum limitation	Interview Self (10-15)
SF-36: MOS 36-item Short Form Health Survey (36)	Physical functioning (PF) (10) Role limitation-physical (RP) (4) Bodily pain (BP) (2) General health (GH) (5) Vitality (VT) (4) Social functioning (SF) (2) Role limitation-emotional (RE) (3) Mental health (MH) (5) Health transition (1)	Categorical: 2-6 options Recall: standard 4 weeks, acute 1 week	Algorithm Domain profile (0-100, 100 best health) Summary: Physical (PCS), Mental (MCS) (mean 50, sd 10)	Interview (mean values 14-15) Self (mean 12.6)

Summary of generic instruments: health status domains (after Fitzpatrick et al., 1998)

Instrument (no. items)	Instrument domains							
	Physical function	Symptoms	Global judgement of health	Psychological well-being	Social well-being	Cognitive functioning	Role activities	Personal constructs
EQ-5D (5+1)	X	X	X	X	X		X	
NHP (38)	X	X		X	X			
SF-36 (36)	X	X	X	X	X		X	

CONDITION-SPECIFIC INSTRUMENTS

Condition-specific patient-reported health instruments

Instrument (no. items)	Domains (no. items)	Response options	Score	Administration (completion time in minutes)
Gastrointestinal Quality of Life Index (GIQLI) (36)	Four domains: Physical, Emotional, Social, Symptoms	5-point Likert scale Recall: past 2 weeks	0-144 (higher score represent better QoL)	Self (completion time not stated)
Abdominal Surgery Impact Scale (ASIS) (18)	Six domains: Physical limitations (3) Functional impairment (3) Pain (3) Visceral function (3) Sleep (3) Psychological function (3)	Not provided	Not provided	Self (completion time not stated)
Gallstone Impact Checklist (GIC) (41)	Four domains: Pain, Dyspepsia , Emotional impact, Food and eating	Check yes/no. Items applicable have Global well-being measured on 10-point ladder scale	VAS (only in items checked 'yes') 0-10, 10 is best health	Self (10-15)
Otago Gallstones Condition-Specific Questionnaire (CSQ) (12)	Four domains: Physical functioning, Systemic functioning, Social functioning, Emotional functioning	5-point Likert scale Free text area provided at end of questionnaire	Summation converted into % (100% = most severe disease state)	Self (<15)

Summary of condition-specific instruments: health status domains (after Fitzpatrick et al., 1998)

Instrument (no. items)	Instrument domains							
	Physical function	Symptoms	Global judgement of health	Psychological well-being	Social well-being	Cognitive functioning	Role activities	Personal constructs
GIQLI	X	X		X	X			
ASIS	X	X		X				
GIC		X	X	X				
CSQ	X	X		X	X		X	

APPENDIX E: Methods of working, membership and conclusions of consensus panel

Members of the panel were invited to participate based on their clinical or research experience of cholecystectomy and special interest in Patient-reported Outcome Measures. Eight potential members were initially contacted, of which four agreed to participate: all of them were consultant surgeons with expertise in cholecystectomy in NHS hospitals.

The panel were sent the following documents:

- A structured review of patient-reported outcome measures for people undergoing cholecystectomy: A report to the Department of Health.
- Copies of the PROMs short-listed for discussion.

The panel were sent by email rating scales to judge the suitability of the questionnaire for use in the NHS for the evaluation of services. There was a section for comments. The rating scale used the following responses:

‘not at all suitable’ (score 0);
‘to some extent unsuitable’ (score 1);
‘uncertain’ (score 2);
‘to some extent suitable’ (score 3);
‘very suitable’ (score 4).

Scores for each questionnaire were ranked in order of preference. The Total maximum score=16.

Ratings and comments

Generic measures

Overall there was agreement that, of the two generic measures proposed in this review, the SF-36 would be more appropriate to be used with people undergoing cholecystectomy. Its validity and appropriateness for general health issues was acknowledged. However in general it was considered that this questionnaire may not capture issues that are specific in this population, and it may need to be administered alongside another instrument that would be more specific. The recall period of 4 weeks used by the SF-36 was not considered ideal for this condition, due to its acute intermittent symptoms, thus presenting a potential risk of inaccurate assessment.

There seemed to be an agreement that the NHP has a smaller amount of evidence supporting its use with people undergoing cholecystectomy. In addition, its questions were considered too broad and may not capture specific issues.

In general, it was felt that the generic measures may not be appropriate for the current laparoscopic techniques used in cholecystectomy (rather than the traditional open surgery approach, which may have long-term sequelae that may be indeed

appropriately captured by a broad general health questionnaire). It was suggested that they could be used in combination with another questionnaire that would capture the issues specific to the condition.

The EQ-5D was the only utility measure identified in this review. The panel agreed that this is a broad instrument that may not capture cholecystectomy-specific aspects accurately, and also has a very small amount of evidence supporting its use in this population.

Conclusions of scoring

Generic instrument	'not at all suitable' (score 0)	'to some extent unsuitable' (score 1)	'uncertain' (score 2)	'to some extent suitable' (score 3)	'very suitable' (score 4)	Total score
SF-36		1		6	4	11
NHP		1	4	3		8
EQ-5D	0		2	3		5

The panel rated each questionnaire using the response categories listed above, favouring the SF-36. This is consistent with the comments, although all members pointed out that a generic instrument alone may not be sufficient to be used as a PROM in cholecystectomy. EQ-5D was the only utility measure identified in the review, and it was rated relatively low.

Condition-specific measures

The panel commented on four condition-specific instruments:

- GIQLI
- ASIS
- GIC
- CSQ

It was agreed by the panel that the amount of evidence for the GIQLI in English-speaking populations may not be enough to recommend its routine use in the NHS. It was also highlighted that most of the questions in this instrument are too broad and do not cover gallbladder related symptoms.

Most members of the panel felt that the ASIS would not be suitable to assess the effects of laparoscopic cholecystectomy. However one of the members felt it had good feasibility, as it was perceived to be easy to use and understand.

The panel agreed with the conclusions of the report regarding the GIC, in terms of length and lack of validity testing. It was also thought to be unsuitable to discriminate between different methods of cholecystectomy.

The members of the panel agreed that the CSQ could be the most suitable questionnaire of the condition-specific identified by the review, since it was felt it

captures specific issues. However, as expressed in the review, members of the panel noted the lack of evidence and felt that a pilot study would be advised prior to its routine use. A potential limitation of this instrument was mentioned regarding the functional aspects of recovery (“return to normal physical, social and employment function”), which are essential after laparoscopic cholecystectomy.

Conclusions of scoring

Condition-specific instruments	‘not at all suitable’ (score 0)	‘to some extent unsuitable’ (score 1)	‘uncertain’ (score 2)	‘to some extent suitable’ (score 3)	‘very suitable’ (score 4)	Total score
GIQLI		1	2		8	11
ASIS	0	2	2			4
GIC		1	2	3	4	10
CSQ				6	8	14

The panel rated each questionnaire using the response categories listed above. Four questionnaires were scored:

- GIQLI
- ASIS
- GIC
- CSQ

Of all the condition-specific instruments identified in this review, the CSQ was the questionnaire believed to be the most appropriate by the panel to be used with people undergoing cholecystectomy, if a choice were to be made. However limitations were pointed out, such as the lack of evidence supporting its use. This issue is more or less common to all the questionnaires suggested.

Patient-reported Outcome Measure Rating Scale

1. On the basis of the review of evidence and your personal experience, is this questionnaire suitable for the measurement of the quality and outcomes of services for people undergoing cholecystectomy? (please tick one box)

Not at all suitable To some extent unsuitable Uncertain To some extent suitable Very suitable

Do you have another questionnaire you could suggest?

Any additional comments

CHOLECYSTECTOMY OUTCOMES CONSENSUS GROUP MEMBERS

Dr Stepehn Attwood

Consultant Surgeon, Northumbria Healthcare Foundation Trust,
Northumberland.

Dr Avril Chang

Consultant Surgeon, Institute of Minimal Access Surgery, King's College
Hospital, London.

Dr Liam Horgan

Consultant Surgeon, Northumbria Healthcare Foundation Trust, Northumberland.

Dr Zahir Soonawalla

Consultant Surgeon, Nuffield Department of Surgery, Oxford.

PROM Group

Ray Fitzpatrick

Professor of Public Health and Primary Care, Department of Public Health, University
of Oxford.

Elizabeth Gibbons

Senior Research Officer, Department of Public Health, University of Oxford.

Carolina Casañas i Comabella

Research Officer, Department of Public Health, University of Oxford.

REFERENCES

- Ainslie, W. G., Catton, J. A., Davides, D., Dexter, S., Gibson, J., Larvin, M., et al. (2003). Micropuncture cholecystectomy vs conventional laparoscopic cholecystectomy: a randomized controlled trial. *Surgical Endoscopy*, *17*(5), 766-772.
- Bardsley, M. J., Venables, C. W., Watson, J., Goodfellow, J., & Wright, P. D. (1992). Evidence for validity of a health status measure in assessing short term outcomes of cholecystectomy. *Quality in Health Care*, *1*(1), 10-14.
- Barkun, J. S., Barkun, A. N., Sampalis, J. S., Fried, G., Taylor, B., Wexler, M. J., et al. (1992). Randomised controlled trial of laparoscopic versus mini cholecystectomy. The McGill Gallstone Treatment Group. *Lancet*, *340*(8828), 1116-1119.
- Brazier, J., Jones, N., & Kind, P. (1993). Testing the Validity of the Euroqol and Comparing It with the SF-36 Health Survey Questionnaire. *Quality of Life Research*, *2*(3), 169-180.
- Burney, R., & Jones, K. (2002). Ambulatory and admitted laparoscopic cholecystectomy patients have comparable outcomes but different functional health status. *Surgical Endoscopy*, *16*(6), 921-926.
- Chen, T. Y. T., Landmann, M. G., Potter, J. C., & van Rij, A. M. (2006). Questionnaire to aid priority and outcomes assessment in gallstone disease. *ANZ Journal of Surgery*, *76*(7), 569-574.
- Cleary, P. D., Greenfield, S., & McNeil, B. J. (1991). Assessing quality of life after surgery. *Controlled Clinical Trials*, *12*(4 Suppl), 189S-203S.
- Darzi A. (2007). Our NHS Our Future: NHS Next Stage Review Interim Report. London: Department of Health.
- Darzi A. (2008). Our NHS Our Future: High Quality Care for All. NHS Next Stage Review Final Report. London: Department of Health.
- Darzi, A., Geraghty, J. G., Williams, N. N., Sheehan, S. S., Tanner, A. N., & Keane, F. B. (1994). The pros and cons of laparoscopic cholecystectomy and extracorporeal shock wave lithotripsy in the management of gallstone disease. *Annals of the Royal College of Surgeons of England*, *76*(1), 42-46.
- Department of Health (2008). Guidance on the routine collection of Patient-Reported Outcome Measures (PROMs) for the NHS in England 2009/10. London: Department of Health.
- Eypasch, E., Williams, J. I., Wood-Dauphinee, S., Ure, B. M., Schmulling, C., Neugebauer, E., et al. (1995). Gastrointestinal Quality of Life Index: Development, validation and application of a new instrument. *British Journal of Surgery*, *82*(2), 216-222.

Eypasch, E., Wood-Dauphinée, S., Williams, J. I., Ure, B., Neugebauer, E., & Troidl, H. (1993). The Gastrointestinal Quality of Life Index. A clinical index for measuring patient status in gastroenterologic surgery. *Der Chirurg*, 64(4), 264-74.

Fitzpatrick, R., Davey, C., & Buxton, M. (1998). *Evaluating patient-based outcome measures for use in clinical trials*. Health technology assessment, vol. 2/no. 14. Alton (GB): Core Research.

Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. (2002). Quality of life measurement: bibliographic study of patient assessed health outcome measures. *British Medical Journal*, 324:1417.

Hospital Episode Statistics HES, 2007-08. The Health and Social Care Information Centre (2009). Retrieved on October 2009 from <http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=206>

Hunt, S. M., McEwen, J., & McKenna, S. P. (1986). *Measuring health status*. London: Croom Helm.

Jones, K. R., Burney, R. E., Peterson, M., & Christy, B. (1998). Measuring health-status improvement after surgery: experience with the SF-36. *Seminars for Nurse Managers*, 6(3), 139-143.

Keus, F., Gooszen, H. G., & Van Laarhoven, C. J. H. M. (2009). Systematic review: open, small-incision or laparoscopic cholecystectomy for symptomatic cholelithiasis. *Alimentary Pharmacology & Therapeutics*, 29(4), 359-378.

Legorreta, A. P., Silber, J. H., Costantino, G. N., Kobylinski, R. W., & Zatz, S. L. (1993). Increased cholecystectomy rate after the introduction of laparoscopic cholecystectomy. *JAMA*, 270(12), 1429-1432.

Macafee, D. A. L., Humes, D. J., Bouliotis, G., Beckingham, I. J., Whynes, D. K., & Lobo, D. N. (2009). Prospective randomized trial using costutility analysis of early versus delayed laparoscopic cholecystectomy for acute gallbladder disease. *British Journal of Surgery*, 96, 1031-1040.

Mallon, P., White, J., McMenamin, M., Das, N., Hughes, D., & Gilliland, R. (2006). Increased cholecystectomy rate in the laparoscopic era: a study of the potential causative factors. *Surgical Endoscopy*, 20(6), 883-886.

McDowell, I. (2006). *Measuring health A guide to rating scales and questionnaires*. New York: Oxford University Press.

McMahon, A., Russell, I., Baxter, J., Ross, S., Anderson, J., Morran, C., et al. (1994). Laparoscopic versus minilaparotomy cholecystectomy: a randomised trial. *The Lancet*, 343(8890), 135.

Nicholl, J. P., Brazier, J. E., Milner, P. C., Westlake, C., Kohler, B., Williams, B. T., et al. (1992). Randomised controlled trial of cost-effectiveness of lithotripsy and open cholecystectomy as treatments for gallbladder stones. *The Lancet*, 340(8823), 801-807.

Nicholl, J. P., Ross, B., Milner, P. C., Brazier, J. E., Westlake, L., Kohler, B., et al. (1994). Cost effectiveness of adjuvant bile salt treatment in extracorporeal shock wave lithotripsy for the treatment of gall bladder stones. *Gut*, 35(9), 1294-1300.

Paumgartner, G., & Sauter, G. H. (January 01, 2005). Extracorporeal shock wave lithotripsy of gallstones: 20th anniversary of the first treatment. *European Journal of Gastroenterology & Hepatology*, 17(5), 525-7.

Russell, M. L., Preshaw, R. M., Brant, R. F., Bultz, B. D., & Page, S. A. (1996). Disease-specific quality of life: the Gallstone Impact Checklist. *Clinical and Investigative Medicine. Médecine Clinique Et Experimentale*, 19(6), 453-460.

Squirrell, D. M., Majeed, A. W., Troy, G., Peacock, J. E., Nicholl, J. P., & Johnson, A. G. (1998). A randomized, prospective, blinded comparison of postoperative pain, metabolic response, and perceived health after laparoscopic and small incision cholecystectomy. *Surgery*, 123(5), 485-495.

Steiner, C. A., Bass, E. B., Talamini, M. A., Pitt, H. A., & Steinberg, E. P. (1994). Surgical Rates and Operative Mortality for Open and Laparoscopic Cholecystectomy in Maryland. *New England Journal of Medicine*, 330(6), 403-408.

Streiner, D. L., & Norman, G. R. (2003). *Health measurement scales: A practical guide to their development and use*. Oxford medical publications. Oxford: Oxford University Press.

The EuroQol Group. EuroQol: a new facility for the measurement of health related quality of life. *Health Policy* 1990; 16, 199–208.

Urbach, D., Harnish, J., McIlroy, J., & Streiner, D. (2006). A measure of quality of life after abdominal surgery. *Quality of Life Research*, 15(6), 1053-1061.

Velanovich, V. (2000). Laparoscopic vs open surgery. *Surgical Endoscopy*, 14(1), 16-21.

Ware, J. (2000). SF-36 health survey : manual and interpretation guide. Lincoln R.I.: QualityMetric Inc.