

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

**A STRUCTURED REVIEW
OF PATIENT-REPORTED
OUTCOME MEASURES FOR
PEOPLE UNDERGOING
ELECTIVE PROCEDURES FOR
BENIGN GYNAECOLOGICAL
CONDITIONS OF THE UTERUS**

**Report to the Department of Health
2010**



patient-reported

O

U

T

C

O

M

E

measures

**A STRUCTURED REVIEW OF PATIENT-REPORTED
OUTCOME MEASURES FOR PEOPLE UNDERGOING
ELECTIVE PROCEDURES FOR BENIGN GYNAECOLOGICAL
CONDITIONS OF THE UTERUS, 2010**

Elizabeth Gibbons
Anne Mackintosh
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 289402

e-mail: elizabeth.gibbons@dphpc.ox.ac.uk

**A STRUCTURED REVIEW OF PATIENT-REPORTED
OUTCOME MEASURES FOR PEOPLE UNDERGOING
PROCEDURES FOR BENIGN GYNAECOLOGICAL
CONDITIONS OF THE UTERUS, 2010**

CONTENTS

Executive summary	1
1. Introduction, methods, search terms and results	3
2. Preference-based measures	7
3. Generic PROMs	8
4. Condition-specific PROMs	11
5. Discussion and recommendations	14
Appendix A – Methodological criteria	17
Appendix B – Search strategy	19
Appendix C – Details of PROMs	20
Appendix D - Methods of working, membership and conclusions of multi-disciplinary panel	22
References	28

EXECUTIVE SUMMARY

Aims of the report

The aims of this report are to identify Patient-reported Outcome Measures which have been evaluated with patients undergoing an elective uterine procedure for benign gynaecological conditions.

The methods of the review are described and the results of the search including sources and search terms used to identify relevant published research. Details of this evidence are presented firstly for preference-based measures, generic health status and then condition -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 DH.

PREFERENCE-BASED MEASURES

One preference based measure was included:

- a) EQ-5D

GENERIC MEASURES

Two generic measures were included:

- a) SF-36
- b) SF-12

CONDITION-SPECIFIC QUESTIONNAIRES

Four condition-specific measures were included; one specific to women with fibroids:

- a) Uterine Fibroid Symptom and Quality of Life Questionnaire (UFS-QOL).
- b) Ruta Menorrhagia Questionnaire
- c) Menorrhagia Utility Scale
- d) Menorrhagia Outcomes Questionnaire

Recommendations

In light of the different procedures and conditions in the evaluations, the EQ-5D and SF-36 are shortlisted as preference and generic measures depending on purpose of measurement.

Based on appraisal of evidence by the PROM Group, and taking into account ratings and comments from the panel, the UFS-QOL is highlighted as an instrument with good supportive evidence for use with women with fibroids.

There is no instrument with substantive evidence to make clear recommendations for women with menorrhagia (without fibroids). This finding is consistent with other reviews (Clark et al., 2002; Jones et al.,2002).

1. 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. This report expands on this by reviewing the evidence of PROMs for other common elective procedures.

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 (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.

Gynaecological procedures

Two main common procedures are carried out in the UK for women suffering with menorrhagia. The 2008-9 HES data (main procedure) report that there were similar numbers of procedures performed for Therapeutic endoscopic operations on the uterus- 33,546 and Abdominal excision of uterus-32,144. The therapeutic endoscopic interventions include different methods of ablation for the treatment of menorrhagia and embolisation for fibroids. The less invasive techniques are now recommended as an alternative to hysterectomy for some women with heavy menstrual bleeding (NICE 2007).

Aim of the report

The aim of this report is to identify Patient-reported Outcome Measures (PROMs) which have been evaluated with patients undergoing elective procedures for benign gynaecological conditions.

Structure of the report

The methods of the review are described and the results of the search including sources and search terms used to identify relevant published research. Details of this evidence are presented for preference-based measures, generic health status and condition or procedure-specific PROMs. The full body of evidence and a short-list of PROMs were presented to a multidisciplinary panel for comment; details of their comments and ratings are reported in Appendix D. The PROMs Group considered the combination of the review of evidence and the views of the multidisciplinary panel before reaching its own conclusions and recommendations.

Methods

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. Full details of search strategies are found in Appendix A.

The searches were conducted using three main sources.

Records in the bibliography database were searched using the keyword 'gynaecology' up to December 2005. This database was compiled by the PROM group with funding from the Department of Health and the Information Centre, and hosted by the University of Oxford¹.

¹ The PROMs bibliography can be accessed free of charge at <http://phi.uhce.ox.ac.uk/home>.

The Ovid search engine was used to explore a number of relevant databases² from January 2006 until March 2010, using a comprehensive search strategy (see appendix A).

Supplementary searches included scanning the reference lists of review articles and others, checking instrument websites, where found, and drawing on other bibliographic resources. Hand-searching of titles of key journals from October 2009 to April 2010 was also conducted. The following journals were selected:

- Health and Quality of Life Outcomes
- Quality of Life Research
- Obstetrics and Gynaecology
- British Journal of Obstetrics and Gynaecology

The following supplementary sources were searched:

- The National Institute for Health Research: Health Technology Assessment Programme
- The Cochrane library <http://www.thecochranelibrary.com/>
- The EQ-5D website; reference search facility (<http://www.euroqol.org/>)

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

- Women with menorrhagia, dysmenorrhoea, fibroids
- Women undergoing vaginal or abdominal hysterectomy
- Women undergoing endoscopic interventions of the management of the above conditions
- 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.

² Four databases were searched using Ovid: [AMED \(Allied and Complementary Medicine\)](#), [EMBASE](#), [PsycInfo](#), [Ovid MEDLINE\(R\) In-Process & Other Non-Indexed Citations and Ovid MEDLINE\(R\)](#).

Specific inclusion criteria for generic, preference-based and condition-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 the UK, North America, or Australasia;
- the instrument will ideally be multi-dimensional. It is at the reviewer's discretion to include PROMs which are specific to a health condition but have a narrow focus, for example, a specific dimension of health, such as symptoms.

Exclusions

- studies using clinician-rated instruments;
- studies evaluating the performance of non-patient reported measures of functioning or health status where a PROM is used as a comparator;
- studies using incomplete versions of instruments.

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 means of the criteria described in Appendix B.

Searches identified 992 potentially relevant records. When assessed against the review inclusion criteria, 46 articles were included in the review (Table 1).

Table 1: Number of articles identified by the search

<i>Source</i>	<i>Results of search</i>	<i>Number of articles included in review</i>
PROM bibliography: 16,054	292	16
Ovid 2006-2010	700	18
Hand searching	-	12
TOTAL	992	46

2. PREFERENCE-BASED MEASURES

One generic preference-based measure was identified:

a) EQ-5D

a. EQ-5D (The EuroQol Group, 1990)

The European Quality of Life instrument (EuroQol)-EQ-5D), was developed by researchers in five European countries to provide an instrument with a core set of generic health status items (The EuroQol Group, 1990; Brazier et al., 1993). There are two sections to the EuroQol: the EQ-5D or five-dimensional index and the EQ thermometer. The EQ-5D assesses health across five domains: 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; health 'today' is assessed. Weights based upon societal valuations of health states are used to calculate an index score. A score profile can 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.

Five studies were identified, three in the UK.

The EQ-5D scores discriminated women with tiredness which was associated with dysfunctional uterine bleeding, with lower scores than those without tiredness in a study reporting financial and quality of life burden (n=237) (Frick et al., 2009).

Discriminative validity is also reported in a UK study of 139 women undergoing different ablation techniques (Abbott et al., 2003). All women reported significantly lower scores than population norms for both the utility index and the VAS.

The EQ-5D detected significant improvement in the index and VAS scores a small study evaluating different ablation methods for dysfunctional uterine bleeding (Abbott et al., 2003). A UK study provides additional evidence in a similar trial (Hawe et al., 2003).

EQ-5D scores were not discriminative between patients who were satisfied and those who were not following thermal ablation for the treatment of menorrhagia (Clark & Gupta. 2004 UK). This was a small study (50 women).

The EQ-5D has been used as a secondary outcome measure in a UK trial of ablation method, but no important psychometric evidence is reported (Sambrook et al., 2009).

No evaluations were identified with women following hysterectomy.

3. GENERIC PROMs

Two generic measures were identified:

- a) SF-36
- b) SF-12

a. SF-36

The SF-36 is a generic health status instrument with 36 items assessing health across eight domains (Ware, 1997), namely bodily pain (BP: two items), general health perceptions (GH: five items), mental health (MH: five items), physical functioning (PF: ten items), role limitations due to emotional health problems (RE : three items), role limitations due to physical health problems (RP: four items), social functioning (SF: two items), and vitality (VT: four items). An additional health transition item, not included in the final score, assesses change in health. All items use categorical response options (range: 2-6 options). Scoring uses a weighted scoring algorithm and a computer-based programme is recommended. Eight domain scores give a health profile; scores are transformed into a scale from 0 to 100, where 100 denotes the best health. Scores can be calculated when up to half of the items are omitted. Two component summary scores for physical and mental health (PCS and MCS, respectively) can also be calculated.

Twelve studies were included; 8 were from the UK. The SF-36 has been evaluated with women with different conditions (fibroids, menorrhagia) and following different elective procedures (ablation, embolisation and hysterectomy).

Factor analysis confirms the relevance of the eight domains (Garrett et al., 1994).

Discriminative validity of the SF-36 was reported in a large UK study (Garrett et al. 1993) which included patients with different health conditions including 271 women with menorrhagia. Scores for all sub-domains (with PF as an exception) were significantly lower than population norms denoting poorer functioning. Higher scores though were reported for women with menorrhagia compared to people with low back pain for all scales except MH and energy or fatigue. These results were according to hypotheses. Scores were significantly lower for women with menorrhagia than UK population norms and all domains detected significant improvement following ablation (microwave and TCRE) (Cooper et al., 1999 UK).

The SF-36 PCS and MCS scores discriminated women with AUB during the illness trajectory from baseline to short-term post hysterectomy and long term follow-up. Baseline scores were significantly lower than population norms; one month post surgery there were significant improvement in MCS but worsening scores for PCS however, at one year, significant improvement in scores was reported in both component summary scores (Kuppermann et al., 2004).

Discriminative validity is reported with scores significantly lower in women (n=263) with heavy menstrual bleeding prior to treatment than women of similar age from the general population (Cooper et al., 1999 UK). Similarly, lower scores (RP, BP, GHP, VT and SF) were reported in a small study (n=46 patients) of women with fibroids compared to the US female general population (Arleo et al., 2007).

Lower scores for RP, BP, SF and MH were significantly predictive of need for surgery in a group of women attending out-patients with a history of heavy menstrual bleeding. Although these domains were predictive, the MCS and PCS scores were not statistically significantly different (Habiba et al., 2010 UK).

All of the SF-36 scales were sensitive in detecting both differences in severity of symptoms of menorrhagia and change in those who had undergone a surgical procedure (Coulter et al., 1994 UK). Six of the SF-36 sub-scales (excluding PF and GHP) detected change in a group of women post surgical intervention for the treatment of menorrhagia with moderate to large effect sizes reflecting significant improvement (Jenkinson et al. 1994, UK).

The SF-36 sub-scale scores all detected significant improvement from baseline to follow-up at 3, 6, 9 and 12 months. There was though substantial loss to follow-up in this small study of 50 women following uterine fibroid embolisation (Spies et al., 1999). Significant improvement in scores for SF, RP, RE, Energy and Pain has been reported in a postal survey (60 to 80 months post procedure) following microwave endometrial ablation and TCRE. The SF-36 detected within group change (Cooper et al., 2005 UK). An earlier study by Cooper et al. (1999 UK) provides further evidence of responsiveness with statistically significant improvement of scores for women receiving microwave ablation or TCRE (263 patients) and in another study comparing resection with medical treatment with significant changes in scores detected between groups and within (Cooper et al., 2001 UK).

In an international study comparing ultrasound surgery and hysterectomy for the treatment of fibroids reported changed within and between groups suggesting SF-36 responsiveness. However, there were only 19 patients from the UK in this study (Taran et al., 2009).

Significant improvement of scores within group and between groups was detected in a UK study of medical vs. ablation methods for the treatment of women with heavy menstrual bleeding (Cooper et al., 1997). Further discriminative validity and responsiveness is reported in a trial comparing ablation methods with resection. Significant improvement in scores was reported following both procedures but distinctive between group differences in scores per domain discriminated patients (Sambrook et al., 2009 UK).

Ceiling effects have been reported for PF (42%), RP (45%), SF (27%), and RE (49%) in 182 patients at baseline with heavy menstrual bleeding (Habiba et al., 2010, UK).

High patient acceptability is reported in several studies. 90% follow-up has been reported in a postal survey 60 to 80 months post ablation techniques. This was though a well conducted trial of two different techniques and a research nurse contacted participants throughout the study (Cooper et al., 2005 UK). High response rates to postal completion have been reported at baseline and 8 month follow-up in three UK studies (Garrett et al., 1993, Habiba et al., 2010, Jenkinson et al., 1994).

b. SF-12

A shorter 12-item version of SF-36 was developed using regression analysis; 12 items were selected that reproduced 90% of the variance in the overall Physical and Mental Health components of the SF-36. A computer-based scoring algorithm is used to calculate scores; Physical Component Summary (PCS) and Mental (MCS) Component Summary scales are generated using norm-based methods. Scores are transformed to have a mean value of 50, standard deviation (SD) 10, where scores above or below 50 are above or below average physical or mental well-being, respectively.

Six studies were identified; three conducted in the UK. The SF-12 has been evaluated with different populations (fibroids, menorrhagia) and following different elective procedures (ablation, embolisation and hysterectomy).

Discriminative validity is reported in a UK study of 139 women undergoing different ablation techniques (Abbott et al., 2003). All women reported significantly lower scores than population norms for both the PCS and MCS. Scores lower than population norms were reported in a large study (1493) of women suffering from a range of non-cancerous gynaecological conditions such as fibroids and heavy menstrual bleeding (Kuppermann et al., 2007).

Significant score changes were detected in a group of women all receiving a form of ablation for heavy menstrual bleeding (Abbott et al., 2003). The SF-12 PCS and MCS detected significant improvement in scores longitudinally in a long-term follow-up of patients receiving UFE for fibroids (Spies et al., 2007) and embolisation or hysterectomy (Spies et al., 2004). The SF-12 MCS and PCS detected significant improvement scores for women in a small study evaluating different ablation methods for dysfunctional uterine bleeding (Abbott et al., 2003). A UK study provides additional evidence in a similar trial (Hawe et al., 2003).

The SF-12 has been used as a secondary outcome measure in a UK trial of ablation method, but no important psychometric evidence is reported (Sambrook et al., 2009).

A five point increase in score has been reported as predictive of satisfaction with newer techniques in Kuppermann et al. (2007).

4. CONDITION-SPECIFIC

Four condition-specific measures were identified; three for menorrhagia and one specifically for fibroids-UFS-QOL.

- a) Uterine Fibroid Symptom and Quality of Life Questionnaire (UFS-QOL).
- b) Rute Menorrhagia Questionnaire
- c) Menorrhagia Utility Scale
- d) Menorrhagia Outcomes Questionnaire

a) Uterine Fibroid Symptom and Quality of Life Questionnaire (UFS-QOL).

Thirteen studies were included in the review; 3 in the UK.

Literature review, focus groups with patients and clinical opinion informed 37 items in two sub-scales: Symptom Severity Scale (8 items) and HRQL (29 items). The HRQL sub-scales are Concern, Activities, Energy, Mood, control, Self-consciousness, Sexual function. Responses are obtained on a five level Likert scale ranging from 'none of the time' to 'all of the time' for HRQL and 'not at all' to 'a very great deal' for the Symptom severity questions. A recall period of three months is chosen to allow for the variability in the severity and duration of menses.

The development study included 110 patients and 29 normal participants (Spies et al., 2002). Exploratory factor analysis supports the single sub-scale for symptom severity and the six sub-scales for HRQL. Internal consistency of all sub-scales are greater than 0.80.

Reproducibility is reported with ICCs ranging from 0.76 to 0.93 (range 7 to 33 days post-test). Construct validity is supported with moderate correlation of scores for the Activities sub-scale and the SF-36 Bodily Pain scale (0.64) and the Menorrhagia Questionnaire scales (0.44 to 0.76). Discriminative validity is reported with significantly higher level of distress and lower HRQL than normal population scores (Spies et al., 2002).

Statistically significant change (improvement in scores) was reported for all sub-scales in 81 women undergoing uterine artery embolisation for the treatment of symptomatic fibroids. Furthermore, the change in symptom severity was significantly associated with the change in total HRQL and Global satisfaction scores (Smith et al., 2004). Further evidence of responsiveness is reported with statistically significant improvement in scores for women undergoing myomectomy (60) and UAE (149) (Goodwin et al., 2006). Further evidence is reported with significant improvement of scores in women following thermoablation for fibroids in a small sampled study (49 participants) in the UK (Smart et al., 2006). This is replicated in a study evaluating thermoablation (MEgFUS) with significant improvement in mean change in scores which were greater than 10 (Stewart et al., 2006 UK). Significant improvement in scores was reported in one small study (36 patients) comparing different embolic agents (Spies et al., 2005a). Significant change was also detected from baseline to 6 and 12 month follow-up following uterine embolisation for fibroids in a large study (1700 patients) (Spies et al., 2005b).

The Symptom Severity Scale detected substantial improvement in scores in a small study of 35 women undergoing percutaneous radiofrequency thermal ablation for fibroids at one and 6 months post procedure (Kim et al., 2007). Additional evidence of responsiveness is reported in a multicentred study which includes a UK population with significant improvement in symptoms (SSS) (Stewart et al., 2007). The SSS and the HRQL scales detected significant improvement in scores in a large group of women undergoing UAE for fibroids (n=1,782). Mean symptom scores and HRQL scores improved 41 points ($p<0.001$) following the procedure (Goodwin et al., 2008).

A 5-point cut-off for improvement has been specified a priori by Goodwin et al. (2006) based on expert opinion of expected difference following treatment of fibroids by either myomectomy or uterine artery embolisation. Using the same a priori clinically important difference of ≥ 5 points, significant change (improvement) was reported between and within group for patients receiving the same procedures (Sisken et al., 2006).

Responsiveness of the SSS and QOL scales is demonstrated with large effect sizes indicating improvement in scores reported in a study of women of the effectiveness of thermoablation (MEgFUS) for fibroids (Harding et al., 2008).

2112 patients completed baseline data in a follow-up study of women receiving UAE for the treatment of fibroids. Completion at three month follow-up was 85% (Goodwin et al., 2008). Excellent data quality is reported from a large study in the US of postal administration of questionnaires including the UFS-QOL instrument. In a large study of 2,645 women with fibroids presenting for assessment for uterine artery embolisation there were 360 (14%) questionnaires with missing or incomplete data (Myers et al., 2005).

b) Ruta Menorrhagia Questionnaire

The Ruta Menorrhagia Questionnaire was developed in the UK and informed by literature reviews, clinical consultation and patient endorsement both empirically and subjectively. 15 items with a 4 point Likert scale for responses with a maximum score of 47. Responses are summed and converted to a percentage to produce and menorrhagia severity score from 0 to 100 (Ruta et al., 1995). It was specifically developed to aid selection and evaluation of treatments.

Three studies provide evidence; one in the UK. Evaluations are with women with fibroids and those receiving ablation for menorrhagia.

Adequate internal consistency was reported (0.77) and most items item-total correlations were greater than 0.2. Test- retest reliability is also reported as adequate (Ruta et al., 1995). The score detected change in a small sample of women undergoing hydrothermablation with statistically significant improvement in score (Rosenbaum et al., 2005). The Ruta scores in women receiving fibroid embolisation were significantly improved post procedure providing further evidence of responsiveness (Spies et al., 2006).

c) Menorrhagia Utility Scale

Items for this measure were derived from interviews with women referred to a gynaecology clinic for the treatment of menorrhagia. Items within domains were rated using a VAS to describe utility weights for the measure with 100 women from the UK with heavy menstrual bleeding. Four responses are obtained for six items: Practical difficulties, Social Life, Psychological health, Physical health and well-being, work/daily routine, Family life and relationships. Scores are obtained by summation of the scores for each item and represent an overall utility for the patients current health status (0 worst affected to 100 unaffected).

Four studies were included; 2 in the UK. All evaluations were with women with menorrhagia.

Face and content validity is supported from patients evaluation (Shaw et al., 1998 UK).

Low MAS scores were predictive of need for surgery with statistically significantly lower scores in women not needing surgery (Habiba et al., 2010).

Responsiveness is reported with significant improvement in Symptoms scores in a group of women post thermal balloon ablation for menorrhagia (Clark et al., 2004). Significant change in scores was detected in a large UK study evaluating a computerised decision aid for the management of menorrhagia with improvement reported in the intervention group which was consistent with other measures in the trial (Protheroe et al., 2007).

High response rates reported in 200 patients although questionnaires were completed in the clinic prior to consultation (Habiba et al., 2010).

d) Menorrhagia Outcomes Questionnaire

This 26 itemed questionnaire was developed in the UK and has four domains: Symptoms (2), Post-operative complications (3), Quality of Life (7) and Satisfaction with outcome (5), and non scored items for Treatment (8) and a global evaluation of symptoms (1) (Lamping et al., 1998). Two summary scores are obtained for Quality of Life/ Satisfaction (12 items) and Global outcome (17 items).

Two UK studies were included although none are within the last few years.

Internal consistency was adequate with alpha greater than 0.70 for summary scales and item total correlations acceptable for most items greater than 0.3 and retest reliability acceptable (Lamping et al., 1998). Discriminative validity is reported with women with greater severity of symptoms reporting poorer questionnaire scores and convergent validity with moderate correlation of score between the MOQ and the NHP. Discriminative validity is reported in a small study (66 patients) with significantly different scores between women undergoing laser ablation compared historical controls (women who had undergone hysterectomy) (Hindley et al., 2002

UK). The women in the hysterectomy group had significantly higher total scores but the quality of life scores were similar between the two groups.

The developmental study reported good precision with no floor or ceiling effects and patient acceptability with very low missing data.

The questionnaire was developed to evaluate the outcomes of all surgical interventions for the treatment of menorrhagia. It was specifically developed as a single questionnaire which could be administered once following treatment therefore not needing a pre-treatment assessment for comparison. The questions are framed to elicit an assessment of the patient's current symptoms and quality of life compared to prior to the operation. Reliance on post-treatment retrospective items contrasts with more common practice to assess change in health status from PROMs administered before and after an intervention.

Other measures

Some very limited evidence was identified for the Pictorial Blood Loss Assessment Chart (Duleba et al., 2003, Busfield et al., 2005). It was considered not to be an appropriate instrument for use in the NHS due to its narrow dimension of the measurement of menstrual flow and volume.

5. DISCUSSION AND RECOMMENDATIONS

Table 3 shows the appraisals of the evidence for each of the PROMs identified in this review.

One preference based measure was identified in the review: the EQ-5D. Some good evidence of discriminative validity and responsiveness is reported including some studies in the UK. This evidence only relates to women with menorrhagia and following ablation methods of treatment. No evidence was reported with women post hysterectomy.

Two generic measures were included in the review: SF-36 and the SF-12.

The SF-36 presents more evidence than the SF-12 with the majority of studies from the UK. Some positive evidence of construct validity, responsiveness and patient acceptability is presented for the SF-36 but floor and ceiling effects are reported. The evidence for the SF-12 is very limited to some support for construct validity and responsiveness.

Four condition specific measures were identified: Uterine Fibroid Symptom and Quality of Life Questionnaire (UFS-QOL), Rute Menorrhagia Questionnaire, Menorrhagia Utility Scale and the Menorrhagia Outcomes Questionnaire.

The most convincing evidence is reported for the UFS-QOL with good measurement performance and patient acceptability. This includes some UK evidence. Two domain scores are generated for Symptoms and Quality of life. However, it may be considered burdensome for patient completion to some extent as there are 37 items. Although it appears highly relevant in terms of content and focus for women with fibroids, this

does limit its utility for women with menorrhagia without fibroids who undergo either ablation or hysterectomy.

The three other instruments presented (RMO, MUS and MOQ) have been developed in the UK but with very little evidence to support implementations in the NHS. Slightly more persuasive evidence is reported for the Menorrhagia Utility Scale with the added appeal of generating a utility value. It is however a very brief measure in terms of number of items and may not capture domains relevant to patients.

Others have noted that whilst personal concerns and health-related quality of life is of importance in this area, there is a marked paucity of validated condition-specific PROMs to assess outcomes of interventions for women with menorrhagia (Clark et al., 2002; Jones et al.,2002).

Recommendations

In light of the different procedures and conditions in the evaluations, the EQ-5D and SF-36 are shortlisted as preference and generic measures depending on purpose of measurement.

Based on appraisal of evidence by the PROM Group, and taking into account ratings and comments from the panel, the UFS-QOL is highlighted as an instrument with good supportive evidence for use with women with fibroids.

There is no instrument with substantive evidence to make clear recommendations for women with menorrhagia (without fibroids). This finding is consistent with other reviews (Clark et al., 2002; Jones et al.,2002).

Table 3: Appraisal of PROMs included in the review

PROM	Reproducibility	Internal consistency	Validity - content	Validity - construct	Responsiveness	Interpretability	Precision	Acceptability	Feasibility
Preference-based measures									
EQ-5D	0	n/a	0	++	++	0	0	0	0
Generic measures									
SF-36	0	0	0	+++	+++	0	-	++	0
SF-12	0	0	0	++	++	+	0	0	0
Condition-specific measures									
UFS-QOL	+	+	+	++	+++	0	0	++	0
RMQ	+	+	+	0	+	0	0	0	0
MUS	0	0	+	0	+	0	0	+	0
MOQ	+	+	+	+	0	0	+	+	0

Appendix A: Appraisal of the methodological quality of PROMs

A simple rating scale (Table i) was used to rate the sum total of evidence available for each dimension or criterion against which PROMs were assessed. The dimensions or criteria are summarised in Table ii.

Table i: Psychometric and operational criteria

0	<i>not reported (no evaluation completed)</i>
—	<i>Evaluation evidence available indicating poor performance of instrument</i>
+	<i>Some limited evidence in favour</i>
++	<i>Good evidence in favour</i>
+++	<i>Excellent evidence in favour</i>

Table ii: Appraisal criteria

Appraisal component	Definition/test	Criteria for acceptability
Reliability		
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
Responsiveness	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) 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 Guyatt's responsiveness index: mean change among all patients divided by the standard deviation of the changes among patients rating themselves unimproved. Standardised effect sizes: mean change divided by the standard deviation of the baseline score.	Statistically significant differences between known groups and/or a difference of expected magnitude Statistically significant changes on scores from pre to post-treatment and/or difference of expected magnitude
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 B: Search strategy for OVID

(HR-PRO or HRPRO or HRQL or HRQoL or QL or QoL).ti,ab. or quality of life.mp. or (health index* or health indices or health profile*).ti,ab. or health status.mp. or ((patient or self or child or parent or carer or proxy) adj (appraisal* or appraised or report or reported or reporting or rated or rating* or based or assessed or assessment*)).ti,ab. or ((disability or function or functional or functions or subjective or utility or utilities or wellbeing or well being) adj2 (index or indices or instrument or instruments or measure or measures or questionnaire* or profile or profiles or scale or scales or score or scores or status or survey or surveys)).ti,ab.

AND

(dysmenorrhea or dysmenorrhoea or endometriosis or menorrhagia).tw. or hysterectomy.mp. or colpohysterectomy.mp. or ((endometrial or uterine) adj (ablation or embolisation or embolization or excision or resection)).tw. or uterine artery embolisation.mp. or uterine artery embolization.mp. or ((ablation or embolisation or embolization or excision or resection) adj2 (uterus or endometrium)).tw. or (TCRE or ((trans cervical or transcervical) adj (resection adj2 endometrium))).tw. or (endometrial adj (ablation or embolisation or embolization or excision or resection)).tw. or ((ablation or embolisation or embolization or excision or resection) adj2 endometrium).tw. or ((ablation or embolisation or embolization or excision or resection) adj2 (fibroid* or fibroma* or fibromyoma* or myoma*)).tw. or leiomyoma/ or leiomyoma*.tw. or myomectomy.tw.

Appendix C - Summary of content and scoring of PROMs included in the review

<i>Instrument (no. items)</i>	<i>Domains (no. items)</i>	<i>Response options</i>	<i>Score</i>	<i>Administration (completion time, mins.)</i>
PREFERENCE-BASED MEASURES				
European Quality of Life instrument, EuroQol/EQ-5D (5+1)	<i>EQ-5D</i> Anxiety/depression (1), Mobility (1), Pain/discomfort (1), Self-care (1), Usual activities (1) <i>EQ-thermometer</i> Global health (1)	<i>EQ-5D</i> Categorical: 3 options <i>EQ-thermometer</i> VAS <i>Health 'today'</i>	<i>EQ-5D</i> Summation: domain profile Utility index (-0.59 to 1.00) <i>EQ-Thermometer</i> VAS (0-100)	Interview or self
GENERIC MEASURES				
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)	Recall: standard 4 weeks, acute 1 week Categorical: 2-6 options	Algorithm Domain profile (0-100, 100 best health) Summary: Physical (PCS), Mental (MCS) (mean 50, sd 10)	Interview (mean 14-15) Self (mean 12.6)
SF-12: MOS 12-item Short Form Health Survey (12)	Bodily pain (BP) (1), Energy/Vitality (VT) (1), General health (GH) (1), Mental health (MH) (2), Physical functioning (PF) (2), Role limitation-emotional (RE) (2), Role limitation-physical (RP) (2), Social functioning (SF) (1)	Recall: standard 4 weeks, acute 1 week Categorical: 2-6 options	Algorithm Domain profile (0-100, 100 best health) Summary: Physical (PCS), Mental (MCS) (mean 50, sd 10)	Interview or self
CONDITION-SPECIFIC				
Uterine Fibroid Symptom and Quality of Life Questionnaire (UFS-QOL)	Two domains: Symptom Severity Scale (8 items) HRQL (29 items). The HRQL sub-scales are Concern, Activities, Energy, Mood, control, Self-consciousness, Sexual function.	Likert	Five point Likert ranging from 'none of the time' to 'all of the time' for HRQL, and 'not at all' to 'a very great deal' for Symptom severity questions	Self
Ruta Menorrhagia Questionnaire, UK	15 items relating to menorrhagia symptoms	Four point Likert	Maximum score 47. Responses are summed and converted to a percentage to produce a menorrhagia severity score from 0 to 100	Self

Menorrhagia Utility Scale, UK	Six items: Practical difficulties, Social Life, Psychological health, Physical health and well-being, work/daily routine, Family life and relationships.	Likert	Scores are obtained by summation of the scores for each item and represent an overall utility for the patient's current health status (0 worst affected to 100 unaffected).	Self
Menorrhagia Outcomes Questionnaire, UK	Four domains (26 items): Symptoms (2), Post-operative complications (3), Quality of Life (7) and Satisfaction with outcome (5), and non scored items for Treatment (8) and a global evaluation of symptoms (1)	Likert	Two summary scores are obtained for Quality of Life/ Satisfaction (12 items) and Global outcome (17 items).	Self

PROMs - summary of health status domains (after Fitzpatrick et al., 1998)

<i>Instrument</i>	<i>Instrument domains</i>								
	Physical function	Symptoms	Global judgement of health	Psychological well-being	Social well-being	Cognitive functioning	Role activities	Personal constructs	Satisfaction with care
<i>Preference-based measures</i>									
EQ-5D	X	X	X	X	X		X		
<i>Generic measures</i>									
SF-36	X	X	X	X	X		X		
SF-12	X	X	X	X	X		X		
<i>Condition-specific measures</i>									
UFS-QOL		X	X	X	X		X	X	
RMQ		X			X		X		
MUS	X	X		X	X		X		
MOQ		X	X		X				X

APPENDIX F: Methods of working, membership and conclusions of multi-disciplinary panel

Members of the multi-disciplinary panel were invited to participate based on their clinical or research experience of benign gynaecological conditions and interest in Patient-reported Outcome Measures. Four members agreed to participate: a Consultant in Obstetrics and Gynaecology, Senior Lecturer with a research interest in PROMs and gynaecological conditions, a specialist GP and a Nurse Consultant in Gynaecology.

The panel were sent the following documents:

- A structured review of patient-reported outcome measures for elective procedures for benign gynaecological conditions, 2010
- Copies of the PROMs short-listed in the review
 - b) SF-36
 - c) EQ-5D
 - d) Uterine Fibroid Symptom and Quality of Life Questionnaire
 - e) Rute Menorrhagia Questionnaire
 - f) Menorrhagia Utility Scale
 - g) Menorrhagia Outcomes Questionnaire

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 based on four people responding was 16.

The results and comments were then distributed by email to the panel for further rating should they wish to change their vote.

Ratings and comments

Generic measures

SF-36

The panel considered this instrument to have several strengths as well as limitations. Some relevant domains were present for example, psychological factors which were not included in the other specific measures. It was deemed to be a good general measure to provide an overall assessment.

Despite this, it was suggested that it needed to be used with another more specific PROM, as recommended in the review. The timing of administration was highlighted as a potential barrier to understanding short or long term complications as well as other life events not related to the surgery. Its length was considered to impede completion by some patients without support.

EQ-5D

The panel emphasised that as a preference-based measure it was suitable but that there were substantial limitations regarding comprehensiveness and the lack of condition-specific items (as would be expected). There was agreement with the panel and the review that the EQ-5D should be used alongside another more specific measure. The VAS thermometer was thought to be useful. However, the reference time for assessment was also considered to not take into account the variation women have through the menstrual cycle.

Generic Total maximum score= 16

RATING	'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
SF-36				4 (12)		12
EQ-5D			1 (2)	3 (9)		11

Conclusions of scoring

The panel rated each questionnaire using the response categories listed above. Overall, the ratings from the panel considered both measures to be suitable to some extent. Although comments suggested preference for the EQ-5D, total scores slightly favoured the SF-36. Based on this, though, there appears broadly similar support for both the SF-36 and EQ-5D as a generic health measure. However, the EQ-5D may be more favourable on the basis of its brevity of items, response rates and the provision of a preference score.

Condition-Specific PROMs

UFS-QOL

Although this instrument has been developed specifically for women with fibroids, some members thought it had some relevant items for women with menorrhagia and others that were highly specific only to women with fibroids, as its intention. One member was highly positive about its clinical utility and content validity and another noted it contained relevant items in relation to physical and psychological issues.

Its length was considered a potential barrier but it was acknowledged that reducing items would require further evaluation.

These more condition specific PROMs cannot compare outcomes of EA with hysterectomy. Hysterectomy results in amenorrhoea whereas most women following EA will have some ongoing bleeding albeit less problematic

RUTA

Despite some methodological reporting, this instrument was considered to be narrowly focused, focussing mainly on symptoms and the lack of a psychological item/domain a disadvantage. Some members felt it was a useful clinical tool and has potential if used with another measure such as a generic PROM. Comments support the lack of content validity in patients post hysterectomy.

A member pointed out that there were other constructs included in the questionnaire which had not been detailed in the review. This has been noted.

MUS

The specific nature of content in relation to menorrhagia limited its utility for evaluating surgical interventions. Whilst it appears a simple instrument, scoring is actually complex and this was thought to be of little value.

MOQ

One member commented on the lack of recent evidence for this instrument stating that this was not necessarily a negative aspect. Treatments have not changed significantly since the instrument was used. The questionnaire is available as a one off retrospective evaluation was thought to be useful to some extent but may result in biased responses.

Its length and specific time period for assessment was considered a barrier.

Specific Total scores: maximum score = 16

RATING	‘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
UFS-QOL				4 (12)		12
Ruta	1 (0)	1 (1)		2 (6)		7
MUS	2 (0)		1 (2)	1 (3)		5
MOQ	1 (0)		1 (2)	2 (6)		8

Conclusions of scoring

The panel rated each questionnaire using the response categories listed above.

The lack of inclusion of measures of satisfaction with care was highlighted by some members however; this was not within the scope of the review.

The generic measures were scored more or less equally.

In relation to condition-specific measures, no instrument was considered very suitable and all but the UFS-QOL were considered not at all suitable by some members.

However, the members were in agreement with the findings from the review that a generic measure should be supplemented with a condition-specific measure. The UFS-QOL was thought to be useful for women with fibroids and with some modification could have utility for other benign gynaecological conditions.

Recommendations

The SF-36 or EQ-5D is recommended for the measurement of general health status and for women with fibroids, the UFS-QOL could be considered. None of the other instruments were considered superior and further evaluations are recommended.

Patient-reported Outcome Measure Rating Scale

On the basis of the review of evidence and your personal experience, and considering the content of the questionnaire, is this questionnaire suitable for the measurement of the quality and outcomes of services for people undergoing elective procedures for benign gynaecological conditions of the uterus? (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?

Overall, do you agree with the conclusions of the review? Yes No: Please explain

Any additional comments

PROMs in benign gynaecological conditions of the uterus:

CONSENSUS GROUP MEMBERS.

Mr Mike Maresh

Consultant in Obstetrics and Gynaecology
St Mary's Hospital for Women and Children

Dr Georgina Jones

Senior Lecturer in Social Science
University of Sheffield

Prof Mary Ann Lumsden

Professor of Medical Education and Gynaecology
University of Glasgow

Dr Anne Connolly,

GPwSI in Gynaecology and a practising GP
Bradford and Airedale tPCT,

Ms Debra Holloway

Nurse Consultant in Gynaecology,
London Guy's & St Thomas' NHS Trust

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.

Anne Mackintosh

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

REFERENCES

- Abbott J, Hawe J, Hunter D, Garry R. A double-blind randomized trial comparing the Cavaterm and the NovaSure endometrial ablation systems for the treatment of dysfunctional uterine bleeding. *Fertility and Sterility* 2003; 80(1):203-208.
- Abbott JA, Hawe J, Garry R, Corson SL. Quality of life should be considered the primary outcome for measuring success of endometrial ablation. *Journal of the American Association of Gynecologic Laparoscopists* 2003; 10(4):491-495.
- Arleo EK, Masheb RM, Pollak J, McCarthy S, Tal MG. Fibroid volume, location and symptoms in women undergoing uterine artery embolization: does size or position matter? *International Journal of Fertility & Womens Medicine* 2007; 52(2-3):111-120.
- Brazier JE, Jones N, Kind P. Testing the validity of the EuroQol and comparing it with the SF-36 Health Survey questionnaire. *Quality of Life Research* 1993; 2(3):169-180.
- Busfield RA, Farquhar CM, Sowter MC, Lethaby A, Sprecher M, Yu Y et al. A randomised trial comparing the levonorgestrel intrauterine system and thermal balloon ablation for heavy menstrual bleeding. *BJOG: An International Journal of Obstetrics & Gynaecology* 2006; 113(3):257-263.
- Clark TJ, Khan KS, Foon R, Pattison H, Bryan S, Gupta J. Quality of life instruments in studies of menorrhagia: a systematic review. *European Journal of Obstetrics, Gynecology and Reproductive Biology* 2002; 104(2):96.
- Clark TJ, Gupta JK. Outpatient thermal balloon ablation of the endometrium. *Fertility and Sterility* 2004; 82(5):1395-1401.
- Cooper KG, Parkin DE, Garratt AM, Grant AM. A randomised comparison of medical and hysteroscopic management in women consulting a gynaecologist for treatment of heavy menstrual loss. *BJOG: British Journal of Obstetrics and Gynaecology* 1997; 104(12):1360-1366.
- Cooper KG, Bain C, Parkin DE. Comparison of microwave endometrial ablation and transcervical resection of the endometrium for treatment of heavy menstrual loss: a randomised trial. *Lancet* 1999; 354(9193):1859-1863.
- Cooper KG, Jack SA, Parkin DE, Grant AM. Five-year follow up of women randomised to medical management or transcervical resection of the endometrium for heavy menstrual loss: clinical and quality of life outcomes. *BJOG: British Journal of Obstetrics and Gynaecology* 2001; 108(12):1222-1228.
- Cooper KG, Bain C, Lawrie L, Parkin DE. A randomised comparison of microwave endometrial ablation with transcervical resection of the endometrium; follow up at a minimum of five years. *BJOG-An-International-Journal-of-Obstetrics-and-Gynaecology* 2005; 112(4):470-475.

- Coulter A, Peto v, Jenkinson C. Quality of life and patient satisfaction following treatment for menorrhagia. *Family Practice* 1994; 11(4):394-401.
- Darzi A. Our NHS Our Future: NHS Next Stage Review Interim Report. Department of Health, London, October 2007.
- Darzi A. Our NHS Our Future: High Quality Care for All. NHS Next Stage Review Final Report. Department of Health, London, June 2008.
- Department of Health, London, December 2008. Guidance on the routine collection of Patient Reported Outcome Measures (PROMs) for the NHS in England 2009/10.
- Duleba AJ, Heppard MC, Soderstrom RM, Townsend DE. A randomized study comparing endometrial cryoablation and rollerball electroablation for treatment of dysfunctional uterine bleeding. *Journal of the American Association of Gynecologic Laparoscopists* 2003; 10(1):17-26.
- Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment* 1998; 2(14).
- Frick KD, Clark MA, Steinwachs DM, Langenberg P, Stovall D, Munro MG et al. Financial and Quality-of-Life Burden of Dysfunctional Uterine Bleeding Among Women Agreeing To Obtain Surgical Treatment. *Women's Health Issues* 2009; 19(1):70-78.
- Garratt AM, Ruta DA, Abdalla MI, Buckingham JK, Russell IT. The SF-36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? *BMJ: British Medical Journal* 1993; 306(6890):1440-1444.
- Garratt AM, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. *British Medical Journal* 2002; 324(7351):1417-1421.
- Goodwin SC, Bradley LD, Lipman JC, Stewart EA, Noshier JL, Sterling KM et al. Uterine artery embolization versus myomectomy: A multicenter comparative study. *Fertility and Sterility* 2006; 85(1):14-21.
- Goodwin SC, Spies JB, Worthington-Kirsch R, Peterson E, Pron G, Li S et al. Uterine artery embolization for treatment of leiomyomata: Long-term outcomes from the FIBROID registry. *Obstetrics and Gynecology* 2008; 111(1):22-33.
- Habiba M, Julian S, Taub N, Clark M, Rashid A, Baker R et al. Limited role of multi-attribute utility scale and SF-36 in predicting management outcome of heavy menstrual bleeding. *European Journal of Obstetrics Gynecology and Reproductive Biology* 2010; 148(1):81-85.
- Harding G, Coyne KS, Thompson CL, Spies JB. The responsiveness of the uterine fibroid symptom and health-related quality of life questionnaire (UFS-QOL). *Health & Quality of Life Outcomes* 2008; 6:99.

Hawe J, Abbott J, Hunter D, Philips G, Garry R. A randomised controlled trial comparing the Cavaterm endometrial ablation system with the Nd:YAG laser for the treatment of dysfunctional uterine bleeding. *BJOG: British Journal of Obstetrics and Gynaecology* 2003; 110(4):350-357.

Hindley J.T, Law P.A, Hickey M, Smith S.C, Lamping D.L, Gedroyc W.M.W, Regan L. Clinical outcomes following percutaneous magnetic resonance image guided laser ablation of symptomatic uterine fibroids. *Human Reproduction*, Vol. 17, No. 10, 2737-2741, October 2002

Jenkinson CP, Peto V, Coulter A. Measuring change over time: a comparison of results from a global single item of health status and the multi-dimensional SF-36 health status survey questionnaire in patients presenting with menorrhagia. *Quality of Life Research* 1994; 3(5):317-321..

Jones GL, Kennedy SH, Jenkinson C. Health-related quality of life measurement in women with common benign gynecologic conditions: a systematic review. *Am J Obstet Gynecol* (2002) 187:501–511

Kim HS, Kim JW, Kim MK, Chung HH, Lee TS, Jeon YT et al. A randomized prospective trial of the postoperative quality of life between laparoscopic uterine artery ligation and laparoscopy-assisted vaginal hysterectomy for the treatment of symptomatic uterine fibroids: Clinical trial design. *Trials* 2009; 10.

Kuppermann M, Varner RE, Summitt RLJ, Learman LA, Ireland C, Vittinghoff E et al. Effect of hysterectomy vs medical treatment on health-related quality of life and sexual functioning: the medicine or surgery (Ms) randomized trial. *JAMA: Journal of the American Medical Association* 2004; 291(12):1447-1455.

Kuppermann M, Learman LA, Schembri M, Gregorich S, Jacoby A, Jackson RA et al. Effect of noncancerous pelvic problems on health-related quality of life and sexual functioning. *Obstetrics and Gynecology* 2007; 110(3):633-642.

Lamping DL, Rowe P, Clarke A, Black NA, Lessof L. Development and validation of the Menorrhagia Outcomes Questionnaire. *BJOG: British Journal of Obstetrics and Gynaecology* 1998; 105(7):766-779.

McDowell I. *Measuring Health: a guide to rating scales and questionnaires*. 3rd ed. Oxford, UK & New York, USA: Oxford University Press, 2006.

Myers ER, Goodwin S, Landow W, Mauro M, Peterson E, Pron G et al. Prospective data collection of a new procedure by a specialty society: the FIBROID registry. *Obstetrics and Gynecology* 2005; 106(1):44-51.

NICE guideline: Heavy menstrual bleeding (2007)

Protheroe J, Bower P, Chew-Graham C, Peters TJ, Fahey T. Effectiveness of a computerized decision aid in primary care on decision making and quality of life in menorrhagia: results of the MENTIP randomized controlled trial. *Medical Decision Making* 2007; 27(5):575-584.

Rosenbaum SP, Fried M, Munro MG. Endometrial hydrothermablation: a comparison of short-term clinical effectiveness in patients with normal endometrial cavities and those with intracavitary pathology. *Journal of Minimally Invasive Gynecology* 2005; 12(2):144-149.

Ruta DA, Garratt AM, Chadha YC, Flett GM, Hall MH, Russell IT. Assessment of patients with menorrhagia: how valid is a structured clinical history as a measure of health status? *Quality of Life Research* 1995; 4(1):33-40.

Sambrook AM, Bain C, Parkin DE, Cooper KG. A randomised comparison of microwave endometrial ablation with transcervical resection of the endometrium: Follow up at a minimum of 10 years. *BJOG: An International Journal of Obstetrics and Gynaecology* 2009; 116(8):1033-1037.

Sambrook AM, Cooper KG, Campbell MK, Cook JA. Clinical outcomes from a randomised comparison of Microwave Endometrial Ablation with Thermal Balloon endometrial ablation for the treatment of heavy menstrual bleeding. *BJOG: An International Journal of Obstetrics and Gynaecology* 2009; 116(8):1038-1045.

Shaw RW, Brickley MR, Evans L, Edwards MJ. Perceptions of women on the impact of menorrhagia on their health using multi-attribute utility assessment. *BJOG: British Journal of Obstetrics and Gynaecology* 1998; 105(11):1155-1159.

Siskin GP, Shlansky-Goldberg RD, Goodwin SC, Sterling K, Lipman JC, Noshier JL et al. A prospective multicenter comparative study between myomectomy and uterine artery embolization with polyvinyl alcohol microspheres: Long-term clinical outcomes in patients with symptomatic uterine fibroids. *Journal of Vascular and Interventional Radiology* 2006; 17(8):1287-1295.

Smart OC, Hindley JT, Regan L, Gedroyc WG. Gonadotrophin-releasing hormone and magnetic-resonance-guided ultrasound surgery for uterine leiomyomata. *Obstetrics and Gynecology* 2006; 108(1):49-54.

Smith WJ, Upton E, Shuster EJ, Klein AJ, Schwartz ML. Patient satisfaction and disease specific quality of life after uterine artery embolization. *American Journal of Obstetrics and Gynecology* 2004; 190(6):1697-1703.

Spies JB, Cooper JM, Worthington-Kirsch R, Lipman JC, Mills BB, Benenati JF. Outcome of uterine embolization and hysterectomy for leiomyomas: results of a multicenter study. *AJOG: American Journal of Obstetrics and Gynecology* 2004; 191(1):22-31.

Spies JB, Myers ER, Worthington-Kirsch R, Mulgund J, Goodwin S, Mauro M et al. The FIBROID Registry: symptom and quality-of-life status one year after therapy. *Obstetrics and Gynecology* 2005; 106(6):1309-1318.

Spies JB, Allison S, Flick P, Cramp M, Bruno J, Jha RC et al. Spherical polyvinyl alcohol versus tris-acryl gelatin microspheres for uterine artery embolization for leiomyomas: results of a limited randomized comparative study. *Journal of Vascular and Interventional Radiology* 2005; 16(11):1431-1437.

Spies JB, Cornell C, Worthington-Kirsch R, Lipman JC, Benenati JF. Long-term Outcome from Uterine Fibroid Embolization with Tris-acryl Gelatin Microspheres: Results of a Multicenter Study. *Journal of Vascular and Interventional Radiology* 2007; 18(2):203-207.

Spies JB. 'Sustained relief of leiomyoma symptoms by using focused ultrasound surgery' - comment. *Obstetrics and Gynecology* 2007; 110(6):1427-1428.

Spies JB, Warren EH, Mathias SD, Walsh SM, Roth AR, Pentecost MJ. Uterine fibroid embolization: measurement of health-related quality of life before and after therapy. *Journal of Vascular and Interventional Radiology* 1999; 10(10):1293-1303.

Spies JB, Coyne KS, Guaou-Guaou N, Boyle D, Skyrnarz-Murphy K, Gonzalves SM. The UFS-QOL, a new disease-specific symptom and health-related quality of life questionnaire for leiomyomata. *Obstetrics and Gynecology* 2002; 99(2):290-300.

Stewart EA, Rabinovici J, Tempany CMC, Inbar Y, Regan L, Gastout B et al. Clinical outcomes of focused ultrasound surgery for the treatment of uterine fibroids. *Fertility and Sterility* 2006; 85(1):22-29.

Stewart EA, Gastout B, Rabinovici J, Kim HS, Regan L, Tempany CMC. Sustained relief of leiomyoma symptoms by using focused ultrasound surgery. *Obstetrics and Gynecology* 2007; 110(2 I):279-287.

Taran FA, Tempany CMC, Regan L, Inbar Y, Revel A, Stewart EA. Magnetic resonance-guided focused ultrasound (MRgFUS) compared with abdominal hysterectomy for treatment of uterine leiomyomas. *Ultrasound in Obstetrics and Gynecology* 2009; 34(5):572-578.

Ware JE. SF-36 Health Survey. Manual and Interpretation Guide. Boston, MA: The Health Institute, New England Medical Centre, 1997. 2nd edition.