PATIENT-REPORTED OUTCOME MEASUREMENT GROUP, OXFORD

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES FOR PEOPLE UNDERGOING ELECTIVE PROCEDURES FOR BENIGN PROSTATIC HYPERPLASIA

Report to the Department of Health
2011

patient-reported

OUTCOME

measures

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES FOR PEOPLE UNDERGOING ELECTIVE PROCEDURES FOR BENIGN PROSTATIC HYPERPLASIA, 2011

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 289402
e-mail: elizabeth.gibbons@dph.ox.ac.uk

Alternatively, it can be downloaded free of charge from the PROM Group website:

http://phi.uhce.ox.ac.uk/

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES FOR PEOPLE UNDERGOING PROCEDURES FOR BENIGN PROSTATIC HYPERPLASIA

CONTENTS

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 prostatectomy procedure.

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 condition -specific PROMs evaluated with men with prostatic hyperplasia undergoing an elective prostatectomy. 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
Two preference based measures were included:

a) EQ-5D
b) HUI

GENERIC MEASURES
Three generic measures were included:

a) SF-36
b) SF-12
c) NHP

CONDITION-SPECIFIC QUESTIONNAIRES
Seven condition-specific measures were included:

a) AUA/IPSS
b) BPH HR-QoL
c) BPH II
d) ICS*male*
e) NIH CPSI
f) PPSM
g) POQ

Recommendations
Based on appraisal of the evidence by the PROM group and taking into account the ratings and comments from the panel, the EQ-5D is recommended for the measurement of general health status and the AUA/IPSS for specific quality of life despite the lack of multi-dimensionality.

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; Streiner and Norman, 2003). These include measurement

issues, such as reliability, validity, responsiveness and precision, as well as practical issues, such as acceptability and feasibility.

BENIGN PROSTATIC HYPERPLASIA

Benign Prostatic enlargement is secondary to prostatic hyperplasia and causes bladder outlet obstruction. Medical therapies are often used to reduce symptoms (under review from NICE) but surgical excision is still frequently performed. Surgical interventions are principally the following: transurethral vaporisation of the prostate (TUVP) and transurethral resection of the prostate (TURP). During the period of 2007/2008 there were just over 25, 000 transurethral resections of the outlet of male bladder (TURP) performed as an elective procedure in England[1].

Symptoms are often referred under three sub-domains- voiding (obstruction etc), storage (hypersensitive bladder), and postmicturation (dribbling etc). Prevalence increases with age.

Treatment guidelines exist and NICE are currently developing guidelines for the management of lower urinary tract symptoms (LUTS) in men. No QOF indicators are specified for LUTS in men and it is considered to be under diagnosed and subsequently under treated (Kirby & Fitzpatrick. 2009).

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 prostatectomy for benign prostatic hyperplasia.

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 report concludes with discussion 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 B.

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 Oxford. In 2005, it became the property of the NHS Information Centre for Health & Social Care. The PROM database comprises over 16,000 records (available online at http://phi.uhce.ox.ac.uk) downloaded from several electronic databases using a comprehensive search strategy (available on request); these records have been assessed as

[1] See Hospital Episode Statistics, Headline figures 2007-8 at http://www.hesonline.org.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=193

eligible for inclusion in the bibliography and assigned keywords. The titles and abstracts of these, as well as a further 14,000 records identified as potential inclusions, were searched.

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 2007 to January 2010 was conducted. The following journals were selected:

- Health and Quality of Life Outcomes
- Quality of Life Research
- Journal of Urology
- British Journal of Urology

The following supplementary sources were searched:

- The National Institute for Health Research: Health Technology Assessment Programme
- The EQ-5D website; reference search facility (http://www.euroqol.org/)
- PubMed records for the period 2007 to 2010

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 with benign prostatic hyperplasia (BPH)
- patients receiving medical therapies for BPH
- patients undergoing transurethral resection or vaporisation of the prostate;
- 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-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 with very small samples, i.e. fewer than 50 participants;
- 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 A.

Searches identified 3246 potentially relevant records. When assessed against the review inclusion criteria, 53 articles were included in the review (Table 1).

Table 1: Number of articles identified by the search

Source	***Results of search***	***Number of articles included in review***
PROM bibliography: 30,350	619	38
PubMed 2007-2010	2627	16
Hand searching	-	0
TOTAL	**3246**	**54**

2. PREFERENCE-BASED MEASURES

Two preference-based measures were identified:

a) EQ-5D
b) Health Utilities Index

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.

Four studies were identified, three in the UK. No recent studies were found (post 2005).

Discriminative validity was reported in a group of men following TURP in the UK. Significantly different scores were reported according to age, ASA grade and number of symptoms: the men who were younger, fitter and had high symptom burden appeared to have a greater benefit from surgery (MacDonagh et al., 1997). Weak correlation of scores (0.32) was reported in a small study with the HUI and EQ-5D (Schulz et al., 2002).

The EQ-5D was responsive to change in a group of men undergoing TURP with significant improvement in scores post surgery at 6 weeks, 6 months and one year (MacDonagh et al., 1997).

The EQ-5D did not detect change in a trial comparing TUVP vs. TURP (213) (Fowler et al., 2005 UK). A ≥5 point change on IPSS was the criterion of success and was achieved in both intervention groups. No score changes were observed for the SF-36 or the EQ-5 D in this study. Similarly, in a UK trial of TURP and contact laser prostatectomy, the EQ-5D did not detect significant change in either arm of the trial with effect sizes all less than 0.40 (Jenkinson et al., 1997 UK). The authors note that any change in general HRQoL result was not detectable by these measures.

Some evidence of patient acceptability is reported with short completion times (ten minutes) reported in a small study (29 men) (Schulz et al., 2002). This was interview administered alongside the HUI and IPSS as well as Time trade off techniques. Response rates of greater than 69% have been reported in a postal survey at three time points: 6 weeks, 6 months and one year post TURP in the UK (n=314) (MacDonagh et al., 1997).

b. Health Utilities Index

The Health Utilities Index (HUI) was designed as a comprehensive framework for describing health status and health-related quality of life for use in clinical studies, population health surveys, and economic evaluations; the original HUI has been largely superseded by HUI2 and HUI3. The Health Utilities Index Mark 3 (HUI3) consists of eight dimensions, rated by members of the general population as the most important attributes or dimensions of health status. For each attribute, there are five or six levels of functioning, ranging from highly impaired to normal, defined in terms of capacity rather than performance, to avoid confounding abilities with preferences. A combination of levels across the attributes constitutes a health state; utility scores, based on community preferences, can be obtained for each health state using an algorithm, with 0 representing death and 1 perfect health.

Population norm data have been obtained from several large general population surveys. Over 15 different language versions of the HUI are available, and it has been used in more than 25 countries.

One study was identified with very limited and weak evidence.

Weak correlation of scores (0.32) was reported in a small study with the HUI and EQ-5D (Schulz et al., 2002).

Long completion times (30 minutes) have been reported in a small study (29 men) (Schulz et al., 2002). This was interview administered alongside the EQ-5D and IPSS as well as Time trade off techniques.

3. GENERIC PROMs

Three generic measures were identified:

a) SF-36
b) SF-12
c) Nottingham Health Profile (NHP)

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.

Five studies were included; all were UK/Ireland). No recent studies were identified.

Two factor solutions are reported using principal components analysis in a large UK study which included a sub-group of men with BPH (Jenkinson & Layte 1997).

Generally, very little evidence of responsiveness is presented. Only a small study provides some empirical evidence of the detection of score changes, but only for specific domains. The RP, BP, SF and PCS detected a significant change on scores (improvement in health status) in a small study of 24 patients, three months following spinal anaesthesia TURP (O'Sullivan et al., 2004 Ireland).

Other evidence is less favourable. Keoghane et al. (1996 UK) report the SF-36 scores to be insensitive to detecting change in patients participating in a trial of TURP compared to laser surgery for BPH compared to condition-specific measures. Other measures in the trial were responsive to change both pre and post-operatively within treatment groups and statistically significant differences between groups, with a greater reduction of symptoms in the TURP group.

The SF-36 also did not detect change in a UK trial comparing TUVP vs. TURP (213) (Fowler et al., 2005). A ≥5 point change on IPSS was the criterion of success and was achieved in both groups. No score changes were observed for the SF-36 or the EQ-5D in this study. The authors note that any change in general HRQoL result was not detectable by these measures. Similarly, in a UK trial of TURP and contact laser prostatectomy, the SF-36 did not detect significant change in either arm of the trial with effect sizes all less than 0.40 (Jenkinson et al., 1997).

An indicator of patient acceptability is provided by a high response rate (72%) to postal administration in a large UK study (13,042) which included a sub-group of men with BPH (116) (Jenkinson & Layte. 1997).

No recent evaluations have been identified.

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.

Two studies were identified; one of which was conducted in the UK (Jenkinson & Layte. 1997).

Discriminative validity is proven in a large survey of the general population (30,000 US) with scores significantly higher for those men (and women) with no LUTS compared to those with high symptom burden (Coyne et al., 2009). Men with BPH were included in the group with LUTS.

Convergent validity is supported in a large study with high correlation of scores (≥0.94) from PCS-36 and PCS-12 and MCS-36 and MCS-12 F-12 (Jenkinson & Layte. 1997). The component scores (PCS, MCS) discriminated between known groups; for example those with chronic illnesses reporting poorer health.

c. Nottingham Health Profile

The Nottingham Health Profile (NHP) was developed in the UK during the 1970s for use in the evaluation of medical or social interventions. Instrument content was derived from over 2000 statements given by 768 patients with a variety of chronic ailments and other lay people.

Part I of the instrument has 38 items across six domains: bodily pain (BP), emotional reactions (ER), energy (E), physical mobility (PM), sleep (S), and social isolation (SI), as shown in Table 1. All items are statements that refer to departures from normal functioning, and relate to feelings and emotional state rather than change in behaviour. Respondents answer ‘yes’ or ‘no’ according to whether or not they feel the item applies to them in general. Positive responses are weighted and summed to give six domain scores between 0 and 100, where 100 denotes maximum limitation.

Part II of the NHP is less widely used and provides a brief indicator of handicap. The instrument may be self-, interview-, or telephone-administered.

Four studies conducted in the UK were identified although none in the last ten years.

Known group’s validity is reported; scores from patients with BPH and undergoing surgery had comparable scores to patients receiving minor surgery (Doll et al., 1993a).

Responsiveness is reported by Doll et al. (1993c) with patients having higher scores pre-operatively compared to scores post surgery which was associated with high symptom burden. Additionally, the Pain and Social Isolation domains detected differences in patients

post-operatively who had remaining prostatism, with significantly higher scores than those patients without post-operative symptom burden (Doll et al., 1993b). Most of the statements in the NHP detected change following surgery in a group of men undergoing TURP with significant improvements in scores in each section (Doll et al., 1993a). Further evidence is provided with significant improvement in scores post TURP surgery at 6 weeks, 6 months and one year (MacDonagh et al., 1997).

Patient acceptability is supported with some good response rates to postal surveys. Response rates of greater than 69% have been reported in a postal survey at three time points: 6 weeks, 6 months and one year post TURP in the UK (n=314) (MacDonagh et al., 1997). Higher response rates are reported from a postal survey in the UK with over 90% return rates at one year post-operatively (Doll et al., 1993b).

A 70 % response rate is reported from a postal survey including the NHP but some patients failed to complete this part of the questionnaire package. Age was strongly associated with non-response (Doll et al., 1993a UK).

4. CONDITION-SPECIFIC

Seven condition-specific measures were identified

a) American Urological Association (AUA) Symptom Index-/International Prostate Symptom Score (IPSS) (Barry et al., 1992)
b) Benign Prostatic Hyperplasia HR-QoL (BPH HR-QoL)
c) BPH Impact Index (BII)
d) International Incontinence Study male Questionnaire (ICSmale)
e) National Institute of Health Chronic Prostatitis Symptom Index (NIH CPSI)
f) Patient Perception of Study Medication (PPSM) questionnaire
g) Prostate Outcomes Questionnaire (POQ)

a. American Urological Association Symptom Index /International Prostate Symptom Score (IPSS)

The AUA Symptom Index was developed by Barry et al. (1992) and forms the basis for the IPSS. The latter instrument includes the seven items from the AUA Symptom Index but an addition question relating to degree of bother of the symptoms present. Both instruments have specific evaluations and evidence and will be presented as reported in the article.

Fifteen studies were identified evaluating the AUA Symptom index, 5 in the UK (Ahmed et al., 1997; Emberton et al., 1995; Jenkinson et al., 1997; Keoghane et al., 1996; Mishriki et al., 2008).

Most evidence is reported pre-2003, prior to the addition of the quality of life item comprising IPSS.

AUA symptom Index: This 7 itemed instrument was developed in the US to categorise BPH symptoms. The seven items considered important to patients by a multidisciplinary measurement committee are: frequency, nocturia, weak urinary stream, hesitancy, intermittence, incomplete emptying and urgency. Symptom responses are obtained on a 6 point Likert scale of frequency of symptoms; scores are summed for a total score. Scores ranged from 0 to 35 with high scores (20 to 35) representing severe symptoms. Scores between 0 and 7 represent Mild; Moderate= 8 to 19 (Barry et al., 1995a).

The developers Barry et al. (1992) report high internal consistency and test re-test reliability. A lower but acceptable ICC (0.74) was reported in a large trial of different pharmacological and surgical interventions (n= 1229) by the authors (Barry et al., 1995a). Good test-retest reliability has been reported in Emberton et al. (1995) in terms of agreement (correlation of scores and weighted kappa statistic).

High correlation of scores is reported between the newly developed scale and patients global ratings of the magnitude of symptoms (Barry et al., 1992). The scale also discriminated patients with BPH and controls without the condition (Barry et al., 1995a).

Barry et al. (2000), performed factor analysis to assess the empirical validity of two hypothetical concepts (‘filling’ and ‘voiding’). Results indicated that psychometrically two factors were valid. When correlating these factors statistically with physiological values results were weak.

Some support for convergent validity is reported with moderate correlation of scores reported between AUA Symptom Index and BPH impact index (0.54) (Barry et al., 1995a). The authors acknowledge that there are important differences in concepts being measured in both of these measures.

Several studies present evidence of responsiveness. Patient AUA Symptom scores significantly improved post-operatively (Barry et al., 1992). Both effect size and responsiveness index (Guyatt's) were of greater magnitude that the BPH impact index in Barry et al. (1995a).

Verbal ratings of perceived change correlated empirically (Barry et al., 1995a). Those who had improved moderately had greater decrease of scores (5.1). These results were different though depending on baseline scores. A minimal score change (2 points) was reported as slight improvement if baseline scores were less than 20, but a greater change score was reported (6 points) if baseline scores were greater than 20.

ROC curves were constructed and sensitivity and specificity of cut-off scores were analysed to explore thresholds for improvement and aid determining clinically significant change. AUC for the AUA Symptom Index was 0.74 and significantly greater than the BPH impact index score (Barry et al., 1995a). Cut-off values for sensitivity and specificity analysis were determined by score changes and not absolute scores. Results showed that the sensitivity was 68% and specificity 70% for score decreases of 3 or more or 3 or less respectively.

Keoghane et al. (1996 UK) reported the AUA Symptom Index scores to be more responsive than other measures in a trial of TURP compared to laser surgery for BPH. A Bothersome score detected change in both groups but the effect sizes were greater for the AUA (1.34 vs. 0.97 for laser surgery and 1.98 vs. 1.52 for TURP). The SF-36 in this study did not detect change in scores possibly indicating greater relevance of items in symptom scores. Other studies provide evidence of the AUA symptoms scores detecting change (improvement) following both TURP and laser prostatectomy (Cowles et al., 1995; Keoghane et al., 2000 UK) and microwave thermotherapy vs. TURP (Ahmed et al., 1997 UK). Large effect sizes (2.07) are reported in a UK trial of TURP vs. laser prostatectomy compared to small effect sizes for the SF-36 and EQ-5D (Jenkinson et al., 1997 UK). Statistically significant change was detected in patient scores (one month) following introduction of alternative alpha blocker therapy for the control of BPH symptoms (Kaplan et al., 2008b).

Responsiveness has been reported in studies evaluating the efficacy of photoselective vaporization of the prostate. Statistically significant improvement in scores was reported during follow-up (from one month to one year). Scores reduced (symptom improvement) by a minimum of 10 points from baseline (Te et al., 2004; Araki et al., 2008). Similar reduction of scores was reported in Mishriki et al. (2008) in a group of patients (280) following TURP in the UK. Other measures in this study demonstrated improvement such as QoL, Bother scores and urinary flow rate. Smaller but statistically significant changes in scores were reported in a medical therapy intervention study (3000+ participants). Score improvement was by 6 units (Gittelman et al., 2006). The AUA symptom score detected change in several studies of combination therapy reported in a review by Greco & McVary. (2008)

Good completion rates have been reported in a trial context 68% (Gittelman et al., 2006).

Different methods of administration have been evaluated in the UK. Emberton et al. (1995) evaluated the effect of contemporaneously or recollected responses to a postal survey including the AUA symptom index and a Symptom impact index. There was an overall response rate of 79%. No statistically significant differences of recollected group mean scores obtained three months after surgery were reported. However, agreement of scores at the individual level was poor and the direction of differences inconsistent. Patient's perceptions of pre-operative health status had changed.

International Prostate Symptom Score (IPSS)-The IPSS contains the same seven items as the AUA symptom index but an additional question related to quality of life- expressed as the degree of bother associated with these symptoms with scoring ranging from 0 to 6. Responses for the seven AUA items are scored on a 6 point Likert scale of frequency of symptoms and scores are summed for a total score. Symptoms for the IPSS are classified as mild (≤ 0); moderate (8-9); severe ($\geq$20) (Rosen et al., 2003a). It has been referred to as a useful tool for diagnosing bladder outlet obstruction, specifically BPH.

Seventeen studies were identified evaluating the IPSS with men with BPH, one in the UK (Fowler et al., 2005). All studies were published post 2003 when the additional quality of life item was incorporated into the AUA Symptom Index.

Some evidence of predictive validity is reported. Total and bother scores of the IPSS were independently predictive of ejaculatory problems in a large international study (Rosen et al., 2009); as symptom prevalence and bother increased, the severity of ejaculatory problems worsened. Based on hypotheses, the IPSS symptom scores have been found to increase with age ($p<0.001$).

Convergent validity is supported with moderate correlation of scores with IPSS and BPH-II (0.44 to 0.67) in a large study of men with BPH receiving watchful waiting or medication therapy (O'Leary et al., 2008).

Discriminative validity is reported in a large survey of the general population (30,000 US) with scores on the QoL sub-scale higher for those men (and women) with no LUTS compared to those with high symptom burden (Coyne et al. 2009).

Several studies present evidence of responsiveness. The IPSS detected a significant change in scores (improvement in symptoms) in a small study of patients (24) three months following spinal anaesthesia TURP (O'Sullivan et al., 2004 Ireland). Responsiveness of the IPSS items is also reported with statistically significant different scores in a medication trial for the control of BPH symptoms. Those in the intervention group had a significant decrease in scores. This was consistent with other measures in the trial (BPH-Impact Index and the IIEF) (McVary et al., 2007; Roehrborn et al., 2008b). Similarly, the IPSS items and the QoL item all detected change in a mono and combination therapy trial (4844 patients) with statistically significant improvement in scores in patients in the combination therapy group. Other measures included physiological measures of prostate size and urinary flow; all showing improvement (Roehrborn et al., 2008a, 2009a, 2009b). Similar findings are reported in a trial of combination medical therapy (Kaplan et al., 2008a) and monotherapy (Nickel et al., 2008; Roehrborn et al., 2009c). Question 8 (QOL) detected significant change (improvement) in a large trial (4844) of different medical therapy regimes for men with moderate to severe BPH (IPSS score of $\geq$12). This was consistent with another patient-reported scale (BII) (Barkin et al., 2009).

Different change scores have been reported as clinically significant across studies. Roehrborn et al. (2008b) specify 2 units. McCallister et al. (2003) and Fowler et al. (2005 UK) specified a priori that a score change of ≥5 was a satisfactory outcome in a trial of TUVP vs. TURP (213). The IPSS scores detected change in both groups greater than the criterion of success (≥5 point change).

The IPSS is used in several trials as a primary endpoint of different surgical procedures for BPH such as microwave thermotherapy and laser prostatectomy. The IPSS detected change in these studies included in systematic reviews both within groups and to some extent between groups (Hoffman et al., 2000; 2007).

The variability of scores was tested in 210 patients with stable BPH by comparing questionnaire results from office based completion and then one week later postal administration. The response rate from the postal questionnaire was 60%. Strong correlation of scores was reported (0.81) but a small proportion of patients scores improved clinically significantly as defined by the author's classification. Some selection bias may be present in this study as the response rate was only 60% (O Connor et al., 2003).

High in unit completion rates (93%) have been reported (O'Leary et al., 2008).
The IPSS has been used in several invention evaluation studies as a primary end-point. It is considered an objective, diagnostic tool and for assessment of severity of symptoms.

b. Benign Prostatic Hyperplasia HR-QoL (BPH HR-QoL)
This instrument was developed incorporating patient views and items and domains identified from a literature review (Epstein et al., 1992). The original questionnaire comprise of 60 items ordered in 12 domains. Items are scaled on a 7 point Likert with responses indicating the intensity of impact of symptoms on activities (none of the time, to all of the time). The final questionnaire contains 49 items in six domains. Domains include Urinary symptoms (12), Bother (12), Psychological Well-being Index (10), BPH interferences with activities (BSIA) (7), Worry and concern (4), Sexual satisfaction (4).

Two studies were identified.

High internal consistency has been reported for those scales with multiple items (0.81 to 0.96). Reproducibility of domains was variable with low to moderate ICCs (0.89 for PF to 0.41 for US) (Epstein et al.,1992).

Some evidence of construct validity is reported. Scores discriminated between men with and with out BPH symptoms and between patients post surgery and non-surgical patients (Epstein et al., 1992). The domains with the best combination of sensitivity and specificity have been reported as the Urinary Symptoms and the Worry and concern domain (Epstein et al., 1991).

The Urinary Symptoms and Bother scales were highly responsive to change following a surgical intervention in a small sample of men (23). Other scales did not detect change in this group (Epstein et al., 1992).

No recent studies were identified.

c. BPH Impact Index (BII)
The BII is a 4-item instrument which assesses the overall impact of BPH on patient's general well-being with 4 to 5 point Likert scales for responses. A total score of all items ranges from 0 to 13 with higher scores denoting greater impact.

Four studies were identified, one form the UK (Jenkinson et al., 1997).

Limited evidence of construct validity is reported. Moderate correlation of scores has been reported between the BPH impact index and AUA symptom index (0.54) (Barry et al., 1995). However, there are important differences in concepts being measured in both of these measures.

Verbal ratings of perceived change correlated with score changes in Barry et al. (1995). Those who had improved moderately had greater decrease of scores (1.1).

ROC curves were constructed and sensitivity and specificity of cut-off scores were analysed to explore thresholds for improvement and aid determining clinically significant change. The AUC for the BPH impact index score was 0.67 but the AUC for the AUA symptom index was significantly greater (Barry et al., 1995a).

Cut-off values for sensitivity and specificity analysis were determined by score changes and not absolute scores. Results showed that the sensitivity was 79% and specificity 70% for score decreases of 3 or more or 3 or less respectively.

Some good evidence of responsiveness has been reported with both effect size and responsiveness index (Guyatt's) of greater magnitude than the AUA symptom index (Barry et al., 1995a). Responsiveness of the BPH-impact index is reported with statistically significant different scores in a medication trial for the control of BPH symptoms. Those in the intervention group had a significant decrease in scores and thus symptoms. This was consistent with other measures in the trial (IPSS and the IIEF) (McVary et al., 2007; Roehrborn et al., 2008b). Large effect sizes (1.38) are reported in a UK trial of TURP vs. laser prostatectomy compared to smaller effect sizes for the SF-36 and EQ-5D (<0.4) (Jenkinson et al., 1997).

d. International Incontinence Study male Questionnaire (ICS*male)*
The ICS*male* questionnaire contains 28 items relating to symptoms and was derived from existing instruments, in-depth interviews with patients and urologists' expert opinion in the UK. Responses are obtained with reference to the prevalence and degree of occurrence of a urinary symptom followed by a question asking about the degree of bother that problem causes obtained on a 4 point Likert scale of 'no problem' to 'a serious problem'.

Four studies were identified, all conducted in the UK. No recent studies were identified (post 2000).

Psychometric evidence during development was good with high patient acceptability from unaided completion from patients with suspected BPO. The instrument discriminated patients who reported lower levels of symptom burden from a community dwelling sample and from

those with higher symptom burden in a clinical setting. High internal consistency (greater than 0.8) and test re-test reliability is reported (0.78, 0.91) (Donovan et al., 1996 UK).

The ICS*male* questionnaire has evidence of sensitivity to expected change following interventions. Results from an international study including UK participants indicate significant improvement and greater magnitude of change in scores for patients following TURP with less change but still significant following drug therapies and minimally invasive treatments (Donovan et al., 1999).

A short version was developed and factor analysis revealed two domains and 11 items: Voiding (ICS*male*VS) (5) with a Cronbach's alpha (0.76) and Incontinence (6) (alpha 0.78). Normal distribution of scores has been reported for the Voiding scale with a minimum score of 0 and maximum 20. The Incontinence scale (ICS*male*IS) observable range is 0 to 18. A positive skew has been reported.

High correlation of scores is reported between the ICS*male* VS and IPSS and lower correlation of scores between ICS*male*IS and IPSS supporting construct validity.
The SF instrument detected expected change following invasive procedures (TURP and laser therapy) compared to conservative management (Donovan et al., 2000).
Following development of the SF, an additional question relating to QoL is included.

A ICS*QoL* instrument has been developed which has six items relating to general and specific problems. Limited data is available for this measure despite good acceptability, discriminative validity and construct validity (Donovan et al., 1997).

e. National Institute of Health Chronic Prostatitis Symptom Index (NIH CPSI)
The items for this instrument were derived from literature reviews, focus groups and cognitive interviewing. Its primary focus is for men with chronic prostatitis with or without BPH. The instrument has nine items in three components: Pain (4) focusing on severity, location and frequency; Urinary function (2 items) concerning irritative and obstructive; Quality of Life (3) about the effect of symptoms on QoL (Litwin et al., 1999).
One study reports very limited evidence with men with BPH.

High test re-test reliability (0.83 to 0.93) and internal consistency (0.86 to 0.91) (Litwin et al., 1999) is reported. The instrument has good discriminative properties with significantly different scores between men with and without chronic prostitis.

High acceptability has been reported during development with a response rate of 70% in a large study (747 participants) which included men with and without BPH but with prostatitis (Litwin et al., 1999).
.
Several studies report good measurement performance and operational characteristics with populations of men with chronic prostatitis and/or chronic pelvic pain syndrome. Although men with BPH often experience similar symptoms, this instrument has not been evaluated extensively with this population specifically.

f. Patient Perception of Study Medication (PPSM) questionnaire
This instrument was developed by GlaxoSmithKline for use in a clinical trial. This 12-item instrument was designed to quantify patients satisfaction with the effects of treatment in 4 areas: control of urinary symptoms (2), strength of urinary stream (2), pain of urination (4),

effect on usual activities (2) and two other items relating to overall satisfaction and the other about use of medication. Responses are obtained on a 7 point Likert satisfaction scale.

One UK study of 879 patients with BPH receiving medication therapy has been identified which provides good psychometric evidence (Black et al., 2009).

Exploratory factor analysis suggests that two concepts are present: PPSM-Global and PSSM-Pain with a Total score range from7 to 49. High internal consistency is reported (0.95 to 0.97) and all item to item correlations were above 0.7. Convergent validity is supported with moderate correlation of scores between the scores and IPSS (0.48 to 0.58) and 0.31 to 0.45 for the BII. Known group's validity is reported with scores significantly different according to BPH severity as measured by the IPSS and BII. PSSM-Global scores detected change with improvement following treatment.

High completion rates and patient acceptability is reported with no floor or ceiling effects.

g. Prostate Outcomes Questionnaire (POQ)
The POQ was developed in the UK (Lamping et al., 1998) and 27 items are located in four domains (Urinary symptoms, Complications after surgery, Quality of life and Patient satisfaction with outcome). High internal consistency and item total correlations have been reported. Convergent validity is supported and it is sensitive to change in patients following surgery. High response rates have been reported.

No further recent evaluations have been identified.

Sexual functioning
Problems with sexual functioning for patients with BPH and following TURP have been reported Muntener et al., 2007). Generally, these measures have been developed for people with sexual problems relating to LUTS including patients with BPH but not exclusive to other conditions such as prostatitis. The methodological evaluation of these instruments is beyond this review but the reader may wish to be aware of them. Table 2 provides basic information.

Table 2

Instrument	Details
Danish Prostatic Symptom Score (DAN-PSS-Sex) (Rosen et al., 2003)	The DAN-PSS-Sex consists of six questions added to the DAN-PSS 12 items) which address urinary symptoms) dealing with erectile problems, ejaculation problems, pain/or discomfort during ejaculation and their respective bothersomeness.
International Index of Erectile Function (IIEF) (Rosen et al., 1997)	This is a 15 item, scale of male sexual function that assesses separate domains of erectile function (EF; 6 items), orgasmic function (OF; 2 items), sexual desire (SD; 2 items), intercourse satisfaction (IS; 3 items) and overall satisfaction (OS; 2 items).
Urolife Scale (Urolife TM BPH QOL) Benign Prostatic Hypertrophy Health-Related Quality of Life Questionnaire (Lukacs et al., 1993)	20 questions, which were scored on a 10 cm visual analogue scale. Six questions are included on each of three dimensions: physical/functional status, mental health and social life; two additional questions relate to overall quality of life dimension
Male Sexual Health Questionnaire (Rosen et al., 2004)	25 items addressing sexual function and satisfaction
Brief Sexual Functioning Inventory (O'Leary et al., 1995)	12 items relating to sexual functioning, perception of problems and overall satisfaction

5. DISCUSSION AND RECOMMENDATIONS

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

Two preference based measures were identified in the review: EQ-5D and the HUI. More favourable evidence is presented for the EQ-5D compared to the HUI but despite good discriminative validity and patient acceptability, there is considerable evidence of a lack of responsiveness with men with BPH following surgery. It is conceivable that the EQ-5D is too broad to be sensitive to specific problems men with BPH experience. The evidence for the HUI is disappointing; weak construct validity and long completion times.

Three generic measures were included in the review: SF-36, SF-12 and NHP. The SF-36 presents slightly more evidence (all UK) than the SF-12 and NHP. Some positive evidence of construct validity and patient acceptability is presented for the SF-36 but there is less favourable evidence of responsiveness with all studies reported lack of sensitivity to change. The evidence for the SF-12 is very limited to some support for construct validity. The NHP on the other hand is reported as responsive to change in groups of patients post surgery. However, to note, no recent studies were identified (last ten years).

The AUA/IPSS symptom index is both the most widely evaluated condition-specific instrument of the seven included in the review. The American Urological Association clinical guideline on the management of BPH (2003) refers to the AUA Symptom Index/International Prostate Symptom Score (IPSS) to be used in an initial assessment of patients presenting with BPH and also as a guide for therapy. For example, a score of ≤ 7 would indicate watchful waiting. This review provides strong and supportive evidence of good psychometric performance and high patient relevance and acceptability. Of particular significance is the considerable positive evidence of responsiveness in this population.

The other six instruments have very little evidence to support use in the NHS but to note that some (BP HR-QoL, PPSM and POQ) are less narrow in their content and include dimensions of health not considered by the AUA/IPSS (psychological well-being, satisfaction with treatment).

Recommendations

Despite a lack of good evidence of responsiveness for the EQ-5D in men with BPH undergoing treatment, considerable evidence and support for its use in other conditions make it attractive for the measurement of general health status and to generate a utility value. Therefore, the EQ-5D should be considered.

For the generic measures, there is slightly more evidence to support the NHP but with caution; the evidence presented here in this review is not recent.

Based on appraisal of evidence by the PROM Group, and taking into account ratings and comments from the panel, the AUA/IPSS is clearly recommended as a condition-specific measure for men with BPH. Despite the lack of multi-dimensionality, it has good performance and high patient acceptability.

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				+	-			+	
HUI				-					
Generic measures									
SF-36				+	-			+	
SF-12				+					
NHP				+	++			++	
Condition-specific measures									
AUA/ IPSS	+	+	++	++	+++	+	+	++	+
BPH HR-QoL		+	+	+	+				
BPH II			+	+	++				
ICSmale			+	+	+			+	
NIH CPSI	+		+	+				+	
PPSM		+	+	+	+				
POQ		+	+	+	+			+	

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 4: 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*

Table 5: 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
	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 Guyatt' 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 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 PubMed

Prostat* OR bph OR luts

AND

(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: publication date- 2007, Human, English

Bibliography

Prostat* title and abstract OR keyword ‘urology’

Appendix C

Summary of content and scoring of PROMs included in the review

Instrument (no. items)	*Domains (no. items)*	*Response options*	*Score*	*Administration (completion time, in minutes)*
PREFERENCE-BASED MEASURES				
European Quality of Life instrument, EuroQol/EQ-5D (5+1)	*EQ-5D* Anxiety/depression (1), Mobility (1), Pain/discomfort (1), Self-care (1), Usual activities (1) *EQ-thermometer* Global health (1)	*EQ-5D* Categorical: 3 options *EQ-thermometer* VAS *Health 'today'*	*EQ-5D* Summation: domain profile Utility index (–0.59 to 1.00) *EQ-Thermometer* VAS (0-100)	Interview or self
Health Utilities Index 3 (8)	Vision, Hearing, Speech, Ambulation, Dexterity, Emotion, Cognition, Pain	Three domains have five response options, five have six response options	Global Utility index and single attribute utility scores for the eight separate dimensions	Interview, telephone or self
GENERIC MEASURES				
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	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)	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				
AUA/IPSS	Seven items: frequency, nocturia, weak urinary stream, hesitancy, intermittence, incomplete emptying and urgency. (Barry et al., 1995a).	Symptom responses are obtained on a 6 point Likert scale of frequency of symptoms	Scores are summed for a total score. Scores ranged from 0 to 35 with high scores (20 to 35) representing severe symptoms. Scores between 0 and 7 represent Mild; Moderate= 8 to 19	Self
BPH HR-QoL	49 items in six domains. Domains include Urinary symptoms (12), Bother (12), Psychological Well-being Index (10), BPH interferences with activities (BSIA) (7), Worry and concern (4), Sexual satisfaction (4).	Items are scaled on a 7 point Likert	with responses indicating the intensity of impact of symptoms on activities (none of the time, to all of the time).	Self
BPH II	4-item instrument which assesses the overall impact of BPH on patient's general well-being.	Responses obtained 4 to 5 point Likert..	A total score of all items ranges from 0 to 13 with higher scores denoting greater impact	Self
ICSmale	28 items relating to symptoms	4 point Likert	Prevalence of symptoms followed by a question asking about the degree of bother that problem causes obtained on a 4 point Likert scale of 'no problem' to 'a serious problem'.	Self
NIH CPSI	Nine items in three components: Pain (4) focusing on severity, location and frequency; Urinary function (2 items) concerning irritative and obstructive; Quality of Life (3) about the effect of symptoms on QoL			Self
PPSM	12-item instrument was designed to quantify patients satisfaction with the effects of treatment in 4 areas: control of urinary symptoms (2), strength of urinary stream (2), pain of urination (4), effect on usual activities (2) and two other items relating to overall satisfaction and the other about use of medication.	Responses are obtained on a 7 point Likert satisfaction scale.		Self
POQ	27 items are located in four domains (Urinary symptoms, Complications after surgery, Quality of life and Patient satisfaction with outcome)			Self

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

Instrument	*Instrument domains*								
	Physical function	**Symptoms**	**Global judgement of health**	**Psychological well-being**	**Social well-being**	**Cognitive functioning**	**Role activities**	**Personal constructs**	**Satisfaction with care**
Preference-based measures									
EQ-5D	x	x	x	x	x		x		
HUI3	x	x		x		x			
Generic measures									
NHP	x	x		x	x				
SF-36	x	x	x	x	x		x		
SF-12	x	x	x	x	x		x		
Condition-specific measures									
AUA/IPSS		x	x						
BPH HR-QoL		x	x	x			x	x	
BPH II		x	x						
ICSmale		x							
NIH CPSI		x	x						
PPSM		x	x				x		x
POQ		x	x						x

APPENDIX D: 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 prostatic hypertrophy and interest in Patient-reported Outcome Measures. Five members were recruited and three responded with ratings and comments.

The panel were sent the following documents:

- A structured review of patient-reported outcome measures for elective procedures for benign prostatic hypertrophy:

- Copies of the PROMs included in the review: EQ-5D, NHP and the IPSS

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.

Ratings and comments

Generic measures

Overall the panel agreed that if a generic measure were to be used, the EQ-5D would be the preferred measure in light of thorough validation, population norms existing for a large number of countries, the availability of translated versions and it has been used with a great many health conditions. In the present context while there is really no strong evidence for treatment responsiveness, the review has identified studies providing supportive evidence of discriminant validity. From the patient perspective it is quite straightforward to complete and does not take too long. However, there were several concerns regarding the use of the thermometer. It was thought to pose challenges for patients in terms of completion. Furthermore, over-simplified measurements of one's health should be treated with caution.

NHP

The NHP was considered to have domains and items which captured other aspects of aging and comorbidities which may confound results particularly as large number of men with BPH/LUTS will be older adults.

Overall, there was some concern about using generic measures to evaluate the effectiveness of surgical procedures and quality of services, in particular, the potential to not capturing important aspects of the patient experience related to the health conditions.

One member suggested the SF-12 as a preferred option.

IPSS
All three members scored this questionnaire as 'very suitable' and concluded that it was clinically relevant and acceptable, had good psychometric evidence including high face validity, and was acceptable to patients. It was also reported to be used frequently in clinical practice worldwide.

One member expressed concern about the "bother" element suggesting that there was no straightforward relationship between symptom severity and the subjective experience ("bother") of the symptoms. It was suggested that it may be that some men with clinically significant levels of symptomatology will be relatively undisturbed while others with a lesser degree of symptoms will be clearly distressed. The 'bother' single item was considered to be too narrow to capture the multifaceted aspects of the impact of BPH on patients.

Furthermore, it was pointed out that while there is a need for straightforward, easily administered, quantitative measures, the downside is that such an approach runs the risk of missing out on precisely those issues that are important to the patient and that shape their response to treatment.

Conclusions of scoring

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**
EQ-5D	0	0	0	9	0	**9**
NHP	0	1	0	0	0	**1**
IPSS	0	0	0	0	12	**12**

Recommendations
Based on appraisal of the evidence by the PROM group and taking into account the ratings and comments from the panel, the EQ-5D is recommended for the measurement of general health status and the IPSS for specific quality of life although its lack of multi-dimensionality may pose limitations.

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 prostatic hyperplasia: CONSENSUS GROUP MEMBERS.

Ms Indu Francis
Surgical Nurse Practitioner
Cambridge University Hospitals

Mr Ken Anson
Consultant Urologist
Queen Mary's University Hospital

Dr Kenneth Gannon
Clinical Research Director
University of East London

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.

REFERENCES

Ahmed M, Bell T, Lawrence WT, Ward JP, Watson GM. Transurethral microwave thermotherapy (Prostatron version 2.5) compared with transurethral resection of the prostate for the treatment of benign prostatic hyperplasia: a randomized, controlled, parallel study. *Br J Urol* 1997; 79:181–185

Araki M, Lam PN, Wong C. High-power potassium-titanyl-phosphate laser photoselective vaporization prostatectomy for symptomatic benign prostatic hyperplasia. *J.Endourol.* 2008;**22**:1311-4.

Barkin J, Roehrborn CG, Siami P, Haillot O, Morrill B, Black L et al. Effect of dutasteride, tamsulosin and the combination on patient-reported quality of life and treatment satisfaction in men with moderate-to-severe benign prostatic hyperplasia: 2-year data from the CombAT trial. *BJU Int* 2009; 103(7):919-926.

Barry MJ, Fowler FJJ, O'Leary MP, Bruskewitz RC, Holtgrewe HL, Mebust WK et al. The American Urological Association symptom index for benign prostatic hyperplasia. *Journal of Urology* 1992; 148(1):1549-1557.

Barry MJ, Fowler FJ Jr, O'Leary MP, Bruskewitz RC, Holtgrewe HL, Mebust WK, Cockett AT. The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. 1992 *J Urol* Nov;148(5):1549-57; discussion 1564.

Barry MJ, Williford WO, Chang Y, Machi M, Jones KM, Walker CE et al. Benign prostatic hyperplasia-specific health status measures in clinical research: how much change in the American Urological Association Symptom Index and the Benign Prostatic Hyperplasia Impact Index is perceptible to patients? *Journal of Urology* 1995; 154(5):1770-1774.

Barry MJ, Fowler FJJ, O'Leary MP, Bruskewitz RC, Holtgrewe HL, Mebust WK. Measuring disease-specific health status in men with benign prostatic hyperplasia. *Medical Care* 1995; 33(4 Suppl):AS145-AS155.

Barry MJ, Williford WO, Fowler FJJ, Jones KM, Lepor H. Filling and voiding symptoms in the American Urological Association Symptom Index: the value of their distinction in a veterans affairs randomized trial of medical therapy in men with a clinical diagnosis of benign prostatic hyperplasia. *Journal of Urology* 2000; 164(5):1559-1564.

Black L, Grove A, Morrill B. The psychometric validation of a US English satisfaction measure for patients with benign prostatic hyperplasia and lower urinary tract symptoms. *Health Qual Life Outcomes* 2009; 7:55.

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.
Coyne KS, Wein AJ, Tubaro A , Sexton CC, Thompson CL, Kopp Z , Aiyer LP
The burden of lower urinary tract symptoms: evaluating the effect of LUTS on health-related quality of life, anxiety and depression: EpiLUTS *BJU International* 2009; Volume 103 Issue s3, Pages 4 - 11

Cowles R, Kabalin J, Childs S, Lepor H, Dixon C, Stein B, A. Zabbo. A prospective randomized comparison of transurethral resection to visual laser ablation of the prostate for the treatment of benign prostatic hyperplasia*
Urology, 1995; Volume 46, Issue 2, Pages 155-160

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.

Doll HA, Black NA, Flood AB, McPherson K. Patient-perceived health status before and up to 12 months after transurethral resection of the prostate for benign prostatic hypertrophy. *Br J Urol.* 1993a**71**:297-305.

Doll HA, Black NA, Flood AB, McPherson K. Criterion validation of the Nottingham Health Profile: patient views of surgery for benign prostatic hypertrophy. *Social Science and Medicine* 1993b 37(1):115-122.

Doll HA, Black NA, McPherson K, Williams GB, Smith JC. Differences in outcome of transurethral resection of the prostate for benign prostatic hypertrophy between three diagnostic categories. *Br J Urol.* 1993c 2: 322-30

Donovan JL, Abrams P, Peters TJ, Kay HE, Reynard J, Chapple C et al. The ICS-'BPH' study: the psychometric validity and reliability of the ICSmale questionnaire. *British Journal of Urology* 1996; 77(4):554-562.

Donovan JL, Brookes ST, De la Rosette JJMC, Peters TJ, Porru D, Kondo A et al. The responsiveness of the ICSmale questionnaire to outcome: evidence from the ICS-'BPH' study. *BJU International* 1999; 83(3):243-248.

Donovan JL, Peters TJ, Abrams P, Brookes ST, De la Rosette JJMC, Schafer W. Scoring the short form ICSmaleSF questionnaire. International Continence Society. *Journal of Urology* 2000; 164(6):1948-1955.

Emberton M, Challands A, Styles RA, Wightman JAK, Black NA. Recollected versus contemporary patient reports of pre-operative symptoms in men undergoing transurethral prostatic resection for benign disease. *Journal of Clinical Epidemiology* 1995; 48(6):749-756.

Epstein RS, Deverka PA, Chute CG, Lieber MM, Oesterling JE, Panser L et al. Urinary symptom and quality of life questions indicative of obstructive benign prostatic hyperplasia: results of a pilot study. *Urology* 1991; 38(1 Suppl):20-26.

Epstein RS, Deverka PA, Chute CG, Panser L, Oesterling JE, Lieber MM et al. Validation of a new quality of life questionnaire for benign prostatic hyperplasia. *Journal of Clinical Epidemiology* 1992; 45(12):1431-1445.

EuroQol Group, The. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy* 1990; 16(3):199-208.

Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment* 1998; 2(14).

Fowler C, McAllister W, Plail R, Karim O, Yang Q. Randomised evaluation of alternative electrosurgical modalities to treat bladder outflow obstruction in men with benign prostatic hyperplasia. *Health Technology Assessment* 2005; Vol. 9: No 4

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.

Gittelman M, Ramsdell J, Young J, McNicholas T. Dutasteride Improves Objective and Subjective Disease Measures in Men With Benign Prostatic Hyperplasia and Modest or Severe Prostate Enlargement. *Journal of Urology* 2006; 176(3):1045-1050.

Greco KA, McVary KT. The role of combination medical therapy in benign prostatic hyperplasia. *Int J Impot Res* 2008; 20 Suppl 3:S33-S43.

Hoffman RM, MacDonald R, Wilt T. Laser prostatectomy for benign prostatic obstruction. *Cochrane Database of Systematic Reviews* 2000, Issue 1.

Hoffman RM, Monga M, Elliott SP, MacDonald R, Wilt T. Microwave thermotherapy for benign prostatic hyperplasia. *Cochrane Database of Systematic Reviews* 2007, Issue 4.

Jenkinson CP, Layte R. Development and testing of the UK SF-12. *Journal of Health Services Research and Policy* 1997; 2(1):14-18.

Jenkinson C, Gray A, Doll H, Lawrence K, Keoghane S, Layte R. Evaluation of Index and Profile Measures of Health Status in a Randomized Controlled Trial: Comparison of the Medical Outcomes Study 36-Item Short Form Health Survey, EuroQol, and Disease Specific Measures *Medical Care.* 35(11):1109-1118, November 1997.

Kaplan SA, Roehrborn CG, Chancellor M, Carlsson M, Bavendam T, Guan Z. Extended-release tolterodine with or without tamsulosin in men with lower urinary tract symptoms and overactive bladder: effects on urinary symptoms assessed by the International Prostate Symptom Score. *BJU.Int.* 2008a;**102**:1133-9.

Kaplan SA, Walmsley K, Te AE. Tolterodine extended release attenuates lower urinary tract symptoms in men with benign prostatic hyperplasia. *J.Urol.* 2008b;**179**:S82-S85.

Keoghane SR, Lawrence KC, Jenkinson CP, Doll HA, Chappel DB, Cranston DW. The Oxford laser prostate trial: sensitivity to change of three measures of outcome. *Urology* 1996; 47(1):43-47.

Keoghane SR, Lawrence KC, Gray AM, Doll HA, Hancock AM, Turner K *et al*. A double-blind randomized controlled trial and economic evaluation of transurethral resection vs contact laser vaporization for benign prostatic enlargement: a 3-year follow-up. *BJU Int* 2000;**85**:74-8.

Kirby M & Fitzpatrick J. Facilitating the medical management of benign hyperplasia in primary care. *British Journal of Urology International* 2009; 104, 751-752

Lamping DL, Rowe P, Black NA, Lessof L. Development and validation of an audit instrument: the prostate outcomes questionnaire. *British Journal of Urology* 1998; 82(1):49-62.

Litwin MS, McNaughton CM, Fowler FJJ, Nickel JC, Calhoun EA, Pontari MA et al. The National Institutes of Health Chronic Prostatitis Symptom Index: development and validation of a new outcome measure. *Journal of Urology* 1999; 162(2):369-375.

Lukacs B, Leplege A, McCarthy C, Comet D. Symptom evaluation, quality of life and sexuality. Appendix A. Construction and validation of a BPH specific health related quality of life scale (with special attention to sexuality), for medical outcome research studies. In: Cockett A.T.K et al, Editors. Proceedings of the 2nde International Consultation on Benign Prostatic Hyperplasia (BPH), in Paris. SCI Jersey 1993: 139-143

MacDonagh RP, Cliff AM, Speakman MJ, O'Boyle PJ, Ewings P, Gudex CM. The use of generic measures of health-related quality of life in the assessment of outcome from transurethral resection of the prostate. *British Journal of Urology* 1997; 79(3):401-408.

McAllister WJ, Karim O, Plail RO, Samra DR, Steggall MJ, Yang Q, Fowler CG. Transurethral electrovaporization of the prostate: is it any better than conventional transurethral resection of the prostate *BJU Int.* 2003 Feb;91(3):211-4.

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

McVary KT, Roehrborn CG, Kaminetsky JC, Auerbach SM, Wachs B, Young JM et al. Tadalafil relieves lower urinary tract symptoms secondary to benign prostatic hyperplasia. *J Urol* 2007; 177(4):1401-1407.

Mishriki SF, Grimsley SJ, Nabi G, Martindale A, Cohen NP. Improved quality of life and enhanced satisfaction after TURP: prospective 12-year follow-up study. *Urology* 2008; 72(2):322-326.

Nickel JC, Barkin J, Koch C, Dupont C, Elhilali M. Finasteride monotherapy maintains stable lower urinary tract symptoms in men with benign prostatic hyperplasia following cessation of alpha blockers. *Can.Urol.Assoc.J.* 2008;**2**:16-21.

O'Connor RC, Bales GT, Avila D, Gerber GS. Variability of the International Prostate Symptom Score in men with lower urinary tract symptoms. *Scandinavian Journal of Urology and Nephrology* 2003; 37(1):35-37.

O'Leary MP, Wei JT, Roehrborn CG, Miner M. Correlation of the International Prostate Symptom Score bother question with the Benign Prostatic Hyperplasia Impact Index in a clinical practice setting. *BJU Int* 2008; 101(12):1531-1535.

O'Leary MP, Fowler FJ, Lenderking WR, Barber B, Sagnier PP, Guess HA, Barry MJ.A brief male sexual function inventory for urology. *Urology.* 1995 Nov;46(5):697-706

O'Sullivan MJ, Murphy C, Deasy C, Iohom G, Kiely EA, Shorten G. Effects of transurethral resection of prostate on the quality of life of patients with benign prostatic hyperplasia. *Journa of the American College of Surgeons* 2004; 198(3):394-403.

Roehrborn CG, McVary KT, Elion-Mboussa A, Viktrup L. Tadalafil administered once daily for lower urinary tract symptoms secondary to benign prostatic hyperplasia: a dose finding study. *J Urol* 2008b; 180(4):1228-1234.

Roehrborn CG, Siami P, Barkin J, Damiao R, Major-Walker K, Morrill B et al. The effects of dutasteride, tamsulosin and combination therapy on lower urinary tract symptoms in men with benign prostatic hyperplasia and prostatic enlargement: 2-year results from the CombAT study. *J Urol* 2008a; 179(2):616-621.

Roehrborn CG, Siami P, Barkin J, Damiao R, Becher E, Minana B et al. The influence of baseline parameters on changes in international prostate symptom score with dutasteride, tamsulosin, and combination therapy among men with symptomatic benign prostatic hyperplasia and an enlarged prostate: 2-year data from the CombAT study. *Eur Urol* 2009a; 55(2):461-471.

Roehrborn CG, Siami P, Barkin J, Damiao R, Major-Walker K, Nandy I *et al*. The Effects of Combination Therapy with Dutasteride and Tamsulosin on Clinical Outcomes in Men with Symptomatic Benign Prostatic Hyperplasia: 4-Year Results from the CombAT Study. *Eur.Urol.* 2009b.

Roehrborn CG, Kaminetsky JC, Auerbach SM, Montelongo RM, Elion-Mboussa A, Viktrup L. Changes in peak urinary flow and voiding efficiency in men with signs and symptoms of benign prostatic hyperplasia during once daily tadalafil treatment. *BJU.Int.* 2009c.

Rosen RC, Wei JT, Althof SE, Seftel AD, Miner M, Perelman MA. Association of sexual dysfunction with lower urinary tract symptoms of BPH and BPH medical therapies: results from the BPH Registry. *Urology* 2009;**73**:562-6.

Rosen R, Catania J, Pollack L, Althof S, Leary M, Seftel A. Male Sexual Health Questionnaire (MSHQ): Scale development and psychometric validation *Urology*, 2003a Volume 64, Issue 4, Pages 777-782

Rosen RC, Altwein J, Boyle P, Kirby RS, Lukacs B, Meuleman E et al. Lower urinary tract symptoms and male sexual dysfunction: the multinational survey of the aging male (MSAM-7). *European Urology* 2003b; 44(6):637-649.

Rosen RC, Catania J, Pollack L, Althof S, O'Leary M, Seftel AD. Male Sexual Health Questionnaire (MSHQ): scale development and psychometric validation. *Urology*. 2004 Oct;64(4):777-82.

Rosen RC, Riley A, Wagner G, Osterloh IH, Kirkpatrick J, Mishra A. The international index of erectile function (IIEF): a multidimensional scale for assessment of erectile dysfunction. Urology. 1997 Jun;49(6):822-30.

Schulz MW, Chen J, Woo HH, Keech M, Watson ME, Davey PJ. A comparison of techniques for eliciting patient preferences in patients with benign prostatic hyperplasia. *J Urol.* 2002 Jul;168(1):155-9.

Streiner D, Norman G. Health Measurement Scales: a practical guide to their development and use. 3rd ed. Oxford, UK & New York, USA: Oxford University Press, 2003.

Te AE, Malloy TR, Stein BS, Ulchaker JC, Nseyo UO, Hai MA et al. Photoselective vaporization of the prostate for the treatment of benign prostatic hyperplasia: 12-month results from the first United States multicenter prospective trial. *Journal of Urology* 2004; 172(4, Part 1):1404-1408.

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