Outcome Measures for Research in Treatment of Adult Males With Symptoms of Lower Urinary Tract Dysfunction


INTRODUCTION

The assessment of outcome may address one or more aspects: lower urinary tract symptoms, impact of symptoms on quality of life, physical findings, and changes in urodynamics following treatment. There has been much confusion with respect to nomenclature, and in this report the following terms are used:

LUTS, lower urinary tract symptoms
QoL, quality of life; QoL can be regarded as covering bothersomeness, impact on quality of life, and sexual functioning
BPH, benign prostatic hyperplasia; this term is reserved for the characteristic histological changes associated with BPH
BPE, benign prostatic enlargement; this term refers to documented enlargement of the gland
BPO, benign prostatic obstruction; this term refers to obstruction documented by urodynamic studies in the presence of an enlarged gland
BOO, bladder outlet obstruction; BOO is the generic term for obstruction.

Outcome measures will differ according to whether the investigation/trial applies to the initial development of a new therapy, or whether the study is at a later phase of evaluation of that therapy. Phase I studies examine efficacy, safety, and tolerability, and are usually open and hence unblinded. Phase II studies are randomized controlled trials in which a ranges of “doses” of the new therapy are tested against a placebo or sham. Randomized controlled trials (RCTs) are the most important method for demonstrating the effectiveness of treatments. Phase III studies are usually RCT in format, using the regime indicated from phase II, in comparison with placebo and/or a comparator, such as an equivalent drug or other therapy.
Outcome measures may be primary or secondary. Primary outcome measures are used in the power calculation needed to determine the size of any study.

During the last decade, many new treatments for BPO have been introduced, including invasive treatments such as superficial or interstitial laser treatment, rectal or urethral heating of the prostate by microwaves, balloon dilatation, temporary or permanent stents, and medical treatments such as 5-alpha-reductase inhibitors and alpha-adrenoceptor-blocking agents. The effectiveness of these treatments has been assessed in many different ways, e.g., by changes in symptoms including symptom score, quality of life, urinary flow rate, residual urine volume, prostatic volume, voiding diary, and pressure-flow parameters, or by complications such as the number of patients going into retention. The side effects and complications of treatments have also been assessed in many different ways, so that neither the consumer (the patient) nor the treating doctor can properly compare therapies. A major reason for this is that there is no consensus on which parameters should be used to measure results of treatment. Furthermore, the tools used to measure changes are often poorly described concerning reproducibility, validity, reliability, sensitivity, specificity, and responsiveness.

The ICS report on “Standardization of Outcome Studies in Patients With Lower Urinary Tract Dysfunction” [Mattiasson et al., this issue] gives the general principles governing the present report. This report analyzes the tools used to assess the various aspects of LUTS and BPO, and on that basis gives recommendations on how to assess the outcome of treatment.

OUTCOME MEASUREMENTS
Measurement of Patient Observations and Symptoms Assessment of Lower Urinary Tract Symptoms

From the patient’s perspective, the assessment of LUTS is the most important outcome measure following therapy. In order to recognize the importance of the patient’s view, the instruments devised over recent years have been designed for patient and not doctor completion. LUTS constitute the main indication for treating BPO. Scoring of severity of symptoms is therefore necessary in order to quantify the symptoms and their influence on the patient’s quality of life (see below) and to register/monitor changes after treatment. Many symptom scores have been used in the evaluation of LUTS suggestive of BPO. None of them have been designed to specifically measure outcomes of treatment. None is BPE- or BPO-specific. Finally, individual symptoms may be considered separately to evaluate changes.

Validated questionnaires. The following instruments have been tested for some aspects of validity and reliability.

The International Prostate Symptom Score (I-PSS) [Barry et al., 1992; Cockett et al., 1991]. The I-PSS was derived from the AUA symptom score with the addition of a condition-specific quality-of-life question. The AUA score contains a symptom score and a bothersomeness score. The AUA symptom score was originally created by selecting seven LUTS from 15 preliminary questions based on their correlation to two global urinary bothersomeness scores and their ability to discriminate between LUTS patients and younger males <55 years old. However, it does not contain
questions related to symptoms which may develop following treatment, such as urinary stress incontinence.

The Danish Prostatic Symptom Score (DAN-PSS-1) [Hald et al., 1991; Hansen et al., 1995]. The DAN-PSS-1 contains a symptom/symptom-bothersomeness part and a sexual questionnaire part (see below). A total of 24 questions, combining symptoms and bothersomeness of symptoms, causes longer page layout and smaller type size and may be confusing, especially to aged patients. Therefore, usability among different populations and different socioeconomic groups might be lower than that of the I-PSS.

Bolognese Symptom Questionnaire for BPH [Bolognese et al., 1992]. This score system contains a symptom score, a symptom bothersomeness score, and a global urination problems question. It is an adaptation of the Boyarsky score. The Bolognese Symptom Questionnaire was constructed to assess the effect of finasteride in men with LUTS suggestive of BPO. It has otherwise not been used.

ICSmale Questionnaire [Donovan et al., 1996]. The ICSmale Questionnaire contains questions on 20 urinary symptoms, 19 of which also have an additional question to ascertain the degree of bother that they cause. This questionnaire is easy to complete. It can differentiate between men in clinical and community populations. It demonstrates reasonable agreement between relevant parts of the questionnaire and frequency volume charts when a ‘relatively flexible approach’ is taken, but there is a very poor relationship between questions assessing strength of stream and the results of uroflowmetry. It has a good internal consistency and good test-retest reliability.

Unvalidated questionnaires. The following symptom questionnaires were the first published. In general they have received no psychometric testing.

Boyarsky Score [Boyarsky et al., 1977]. The first symptom score published to assess LUTS was designed for doctor completion. It has never been validated in its original form, and is normally used in a more or less modified form.

Madsen-Iversen Score [Madsen and Iversen, 1983]. This symptom score has never been validated in its original form. It is designed for doctor completion and is often used in a more or less modified form, including weighting of symptoms by a grading system, individual to each symptom.

The Maine Medical Assessment Program (MMAP) Instrument [Fowler et al., 1988]. Little validation has been done and that only in relation to responsiveness. The MMAP was a simple questionnaire, consisting of five questions, and it was the first to be designed for patient completion.

Bothersomeness of symptoms. Another method of assessing symptoms is to ask the patient about the degree of bother each symptom causes. This approach has been used in the AUA Bother Index, the ICSmale Questionnaire, and the DAN-PSS-1. At present it is not known whether the bother index or the quality of life index is more responsive to therapies, or whether they measure different aspects of the impact of therapy.

Sexual function assessment. It is now recognized that sexual function is important when assessing therapies. There are a number of instruments available, although none have been fully validated in LUTS or after therapies for BPO. The ICS sex and its derivative PROTOsex, O’Leary’s Brief Sexual Function Inventory [O’Leary et al., 1995], and DANPSS assess similar aspects of sexual function.
Recommendations/Comments

A validated, patient-completed symptom questionnaire must always be included in trials of therapies for LUTS and BPO to capture the patient’s perspective of their condition. ample research has illustrated that symptom questionnaires are not sex-, age-, or condition-specific. Therefore, any symptom questionnaire/score may only be used to document the presence and/or severity of its constituent symptoms at any point in time. However, if this limitation is appreciated, these instruments are valuable in trials of therapies for LUTS and BPO. If such a score is used, reliability and validity data should be provided if available or their absence indicated.

Documentation of Voiding Pattern

Urinary diary (frequency volume chart, voiding diary, and bladder diary) [Bødker et al., 1989]. A voiding diary is a self-monitored record of selected lower urinary tract functions over specific time periods. In LUTS and BPO, selected variables are often voiding/toileting frequency (diurnal and nocturnal), voided volume, and fluid intake volume. A voiding diary is essential both in the clinical work and in research to obtain semiobjective data on frequency and volume voided in particular. E.g. provides the average voided volume from the voiding diary important redundant information to judge free uroflowmetry and pressure/flow studies.

Recommendations/Comments

Sparse information is available concerning validation of voiding diaries in men with LUTS. If a voiding diary is used, reliability and validity data should be provided, if available, or their absence indicated. It should be emphasized that the committee considers voiding diaries to be important in the evaluation of LUTS in order to diagnose high diuresis or reversed day and night diuresis as the cause of frequency and/or nocturia.

Anatomical and Functional Measurements: Clinical Outcome Measures

Estimation of prostate size. Prostate size is of little interest in relation to outcome in general, in view of the fact that prostatic size correlates poorly with symptoms, BOO, and treatment outcome. However, where therapies aim at reducing prostate size, either by shrinkage or by tissue destruction, exact measurement of prostate size may be important. A number of methods may be used. Of the methods available, transrectal ultrasound scanning (TRUS) is the most accurate, and digital rectal examination is the easiest test but the most inaccurate. MRI is accurate, but expensive.

From TRUS, planimetric volumetry is regarded as the “gold standard” for prostate volume determination, and excellent correlation between planimetric volume and absolute volume in cadavers has been demonstrated [Jones et al., 1989; Hastak et al., 1982; Hendriks et al., 1991; Peeling, 1989]. Reliability is very good [Peeling, 1989; Nathan et al., 1996]. Different methods and formulas to measure and calculate prostate volume from different axes of the prostate are less accurate [Nathan et al., 1996; Bangma et al., 1996]. Other, simpler formulas are used to calculate prostate volume from axial, sagittal, and coronal diameters. The accuracy of these seem sufficient for daily clinical use [Terris and Stamey, 1991]. Transition zone volume
and transition zone index may be a valuable supplement to whole-gland volume
determination.

It has been reported that the anteroposterior diameter increases more than the
transverse diameter with age, and that the shape of prostate becomes round by trans-
verse section [Peeling, 1989], although controversy exists [Bosch et al., 1994].
Changes in microscopic features of glandular, muscle, and connective tissue content
in the gland might prove to be relevant [Marks et al., 1994], and may be important in
assessing some new therapies.

**Recommendations/Comments**

If treatment intends to change prostate volume, measurements should be done
before and after treatment. Timing of posttreatment testing depends on mechanism of
action of the modality. The method used and its reliability and validity should be
provided if available, or their absence indicated.

**Assessment of bladder outlet obstruction**

*Pressure flow studies (pQS).* The only method of accurately diagnosing and
grading BOO is pressure-flow studies with simultaneous recording of abdominal
pressure, intravesical pressure, and urinary flow rate. As this is an invasive and
time-consuming procedure, it will probably never be adopted as routine in all patients
with LUTS suggestive of BPO. It is, however, the cornerstone of the assessment of
outcome after treatments claiming to influence BOO, or to relieve symptoms because
of reduction in BOO.

*Urine flow rate.* The urinary flow rate reflects both detrusor function and outlet
conditions. Because both muscle mechanics and bladder geometry depend on bladder
volume, urinary flow rate increases with increasing bladder volume until a certain
limit, and thereafter decreases. This underlines the high dependence of urinary flow
rates on detrusor muscle function and explains why urinary flow rate alone is of
limited help in the diagnosis of outlet obstruction.

Test-retest measurements of maximum urinary flow rates have a standard de-
viation of about 3.5 ml/sec, with a wide range [Neal et al., 1987; Grino et al., 1993].
A significant overlap exists between symptomatic and asymptomatic men [Diokno et
al., 1994], as well as between obstructive and nonobstructive voiding [Abrams and
Griffiths, 1979]. Correlation for voided volume does not change this finding.

*Residual urine volume.* Wide intraindividual variation of residual urine volume
has been demonstrated by catheter or ultrasound measurement in patients with LUTS
suggestive of BPO [Bruskewitz et al., 1982; Dunsmuir et al., 1996]. In the latter study,
the individual variation was 42%, with a confidence interval of 55–228 ml. As poor
bladder emptying may be due to either outlet obstruction or detrusor underactivity or
both, the correlation of residual urine volume with outlet obstruction alone is poor
[Neal et al., 1987, Bruskewitz et al., 1982; Poulsen et al., 1994].

*Cystourethroscopy.* Prostatic intraurethral protrusion and bladder trabeculation
are poor indicators of bladder outlet obstruction [Andersen et al., 1980; Rosier et al.,
1996]. As a direct indicator of BOO after treatment, cystourethroscopy is therefore
useless. It might in certain treatments give a clue as to why a therapy may fail, e.g.,
presence of the median lobe in thermotherapy.
Recommendations/Comments

In research studies of outcome, pQS must be included to document the presence and degree of change in BOO.

Results should be presented as stated in the ICS 1997 standardization report on pressure-flow studies of voiding, urethral resistance, and urethral obstruction [Griffiths et al., 1997].

Change in flow rates in response to treatment is sensitive, but the degree of change is meaningless unless pretreatment detrusor voiding pressure is known. A slight decrease in outlet resistance might produce a pronounced increase in maximum urinary flow rate if outlet resistance is low before treatment; and a significant decrease in outlet resistance might produce only a small increase in maximum urinary flow rate if outlet resistance were high before treatment and an element of obstruction persists.

Posttreatment reduction of residual urine volume indicates improvement of outlet conditions and is likely to be more significant in assessing treatment response than for diagnosis. Methods used for the assessment of BOO should be stated, and reliability and validity data should be provided if available or their absence indicated.

Assessment of Impact on Quality of Life

As mentioned in the Introduction, quality of life may also include sexual function and bothersomeness of LUTS. Considerable developmental work is being undertaken to develop measurement tools within this area, but so far a widely accepted quality-of-life scale for LUTS does not exist. Quality-of-life aspects should be addressed in outcome studies.

Quality-of-life assessment. There are numerous instruments in this ever-expanding field. They can be divided into two broad categories, generic quality-of-life instruments and disease-specific quality-of-life instruments.

Generic quality of life instruments. The best-known generic instruments are the Nottingham Health Profile, the SF36, and the EuroQoL [Donovan et al., 1997]. At the time of writing there is very little scientific work on the use of these instruments in men with LUTS and/or BPO. The aim of generic instruments is to get an overview of health status, and to allow comparisons between patient groups. It is unlikely that these instruments will be very sensitive to the changes following therapies, in view of the aged population with coexisting pathologies.

Disease-specific quality-of-life instruments. The most commonly used instrument is the single question attached to the I-PSS. There has been concern that a single question may not adequately capture the impact of LUTS on quality of life, and therefore other instruments are being developed. The ICSQoL is part of the ICS-‘‘BPH’’ Questionnaire [Donovan et al., 1997]. Lukacs et al. [1994] and Girman et al. [1994] have produced disease-specific quality-of-life questionnaires which are currently under evaluation, and the BII (BPH Impact Index) [Barry et al., 1995] is derived from the AUA quality-of-life questionnaire.

Recommendations/Comments

Treatment effects on quality of life are probably among the most important outcome parameters. Further research should therefore be done to identify the best way to measure this after treatment of LUTS suggestive of BPO.
When a QoL assessment is used, the particular method should be stated, and reliability and validity data should be provided if available or their absence indicated.

**Socioeconomic Measurements**

As health care resources are limited and both LUTS and BPO are common conditions, outcome studies must include an assessment of relative costs, at least when new treatments are introduced.

Measurement tools are under development, but are even less complete than the quality-of-life measuring tools. Therefore, this research area deserves much attention and might produce surprising results when properly applied [Keoghane et al., 1996].

**Recommendations/Comments**

Cost are not, strictly speaking, an outcome measure. An economic evaluation (specifically a cost-effectiveness study) should accompany any evaluation of effectiveness. Costs need to be collected from the viewpoint of the patient and society and then combined with effectiveness to give an indication of the cost-effectiveness of different treatments.

**Treatment Complications**

The treatment complications of a specific treatment must also be described and quantified in a complete outcome evaluation of any treatment. Complications such as mortality and morbidity must be addressed, although complications may vary considerably with different treatments. It is therefore not possible to give specific recommendations in this area.

**Durability**

Durability of a treatment is important. It is therefore essential that long-term follow-up be included in the evaluation of any new treatments. The Third International Consultation on Benign Prostatic Hyperplasia defines evaluation time into three categories: short-term, up to and including 3 months; medium-term, up to and including 12 months; and long-term, studies reporting after 12 months [Cockett et al., 1996]. Issues such as retreatment rate and treatment failure rate should be considered.

**CONCLUSIONS**

Measuring outcomes in LUTS and BPO depends on the purpose of the measurement. In the study of new treatments, all domains must be covered, meaning that those domains allegedly affected by that treatment must be investigated properly. Symptoms must be evaluated by a validated symptom score, prostatic obstruction must be evaluated by pressure-flow studies, and quality-of-life and cost issues should be addressed. Other critical issues are the length of follow-up and the long-term effectiveness of treatments.

In a quality assurance program, outcome could be assessed in a less ambitious way, using changes in a validated symptom score and changes in an indirect measure of outlet obstruction such as urinary flow rate and/or residual urine volume. However, the assessment of quality of life and costs should also be considered in such a
program. In daily clinical work, the follow-up that is actually performed will depend on the resources available. A general practitioner will not have the same possibilities as a hospital clinic, and the diagnostic setup will also necessitate different methods of assessment of outcome.

A wide range of methods of evaluating outcomes in LUTS and BOO is available. The precise tools to be used will depend on the aims of the study and the type of design. Outcomes are best assessed through randomized controlled trials, in which a new treatment or type of service is evaluated in comparison with an existing standard therapy (such as TURP), placebo, or conservative management. Within such trials, the assessment of outcome should always include measurements of LUTS (using validated questionnaires or scores), $Q_{\text{max}}$, residual urine and voided volume (uroflowmetry), quality of life (generic or condition-specific questionnaires or questions, sexual function), and treatment complications. All researchers should also consider including pQS, the measurement of prostate size, and an economic evaluation, particularly for trials of new treatments or services.

REFERENCES


