Neurourology and Urodynamics 17:255–262 (1998)

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

Gunnar Lose,* J. Andrew Fantl, Arne Victor, Steen Walter, Thelma L. Wells, Jean Wyman, and Anders Mattiasson

INTRODUCTION

The purpose of this communication is to offer the clinical and research communities an initial attempt at incorporating outcome measurements within identifiable domains, as well as providing initial information as to the reliability of those measurements most commonly used.

Scientific evaluation of the outcome of therapeutic interventions is based on assessment before and after treatment. However, the reliability of methods and measurements used in the evaluation is often poor or unclear, which makes interpretation of outcome difficult. The reliability of a test depends on the accuracy and reproducibility of the method. The diagnostic accuracy is determined by verifying test results against a reference (''gold'') standard that defines true disease status. The predictive value of a measure is considered the most important. Since it may be impossible to establish a final true diagnosis, reliability must in some cases be measured by reproducibility, which is determined by comparing results of repeated examinations of the same patient.[1] Reproducibility for tests where the result is given on a continuous scale may be expressed as bias with the 95% confidence limit [Bland and Altman, 1986], while for binary tests the kappa coefficient, which adjusts the observed agreement for expected chance agreement, is used [Gjørup, 1997].

Urinary incontinence is defined as ''a condition in which involuntary urine loss is a social or hygienic problem and is objectively demonstrable.'' It represents a multidimensional phenomenon with wide-reaching effects which may be quantified within various areas or domains. These areas or domains include:

Produced by the Standardisation Committee of the International Continence Society, A. Mattiasson, Chairman. Subcommittee on Outcome Research in Women, G. Lose, Chairman.

*Correspondence to: Gunnar Lose, Department of Obstetrics/Gynecology, Glostrup County Hospital, Copenhagen University, Glostrup, Denmark

[1]Other terms might have been chosen to denote accuracy and reproducibility. Terms such as precision, validity, observer or interrater variability, observer error, and efficacy have been used. At the moment it does not seem possible to establish a commonly accepted terminology. Readers of papers about reliability must therefore in each case examine the author's use of the terms.

1. The patient's observations (symptoms)
2. Quantification of symptoms (e.g., urine loss)
3. The clinician's observations (anatomical and functional)
4. Quality of life
5. Socioeconomic measures.

Outcome measures should be selected within the context of a specific study.

The ideal combination or measurement ''in aggregate,'' particularly if belonging to different domains, will enhance the overall significance and value of the study results. A multidimensional approach is important, since the effect of intervention depends on the outcome measure chosen and may vary significantly between various domains and even within a certain domain.

DOMAIN OF SYMPTOMATOLOGY

Symptoms

General and condition-specific. Here are included a respondent's overall opinion of the condition (incontinence) and/or one or more characteristics of the condition, e.g., frequency, quantity, or magnitude. Different methods to obtain this measure include: a question with a forced choice, a graded response, a statement with a Likert scale agree-disagree response, and a statement with a visual analog graded scale response.

Recommendations/Comments

There is no general symptom (opinion) measure with established methodologic reliability. Therefore, researchers should clearly describe their instrument and procedure and provide reliability data or indicate their absence.

DOMAIN OF SYMPTOM QUANTIFICATION

Diary (Frequency/Volume Chart) [Larsson and Victor, 1988; Larsson et al., 1991; Wyman et al., 1988]

This entails a self-monitored record of selected lower urinary functions kept for specific time periods. Selected variables include episodes of incontinence/pad use frequency (diurnal and nocturnal), voiding/toileting frequency (diurnal and nocturnal), total voided volume, mean voided volume, and largest single void. Accuracy depends on the patient's ability to follow instructions but is still difficult to assess because there is no gold standard against which the test result can be compared. Reproducibility depends on the parameter used and improves with an increase in the number of days that self-recording is obtained.

The highest reproducibility has been found for mean voided volume.

The circumstances under which a diary is kept should be approximate to those of everyday life, and should be similar before and after intervention to allow for meaningful comparison.

Recommendations/Comments

A diary kept for a minimum of three days is recommended as an outcome measure.

Pad-Weighing Tests [Lose and Versi, 1992; Victor, 1990]

Pad tests can be divided into short-term tests, generally performed under standardized conditions as office tests, and long-term tests, generally performed by the patient at home during 24–48 hr.

Accuracy is difficult to assess. Reproducibility is generally poor, but improves if the circumstances are standardized as much as possible. For long-term tests, the test period should be sufficiently long.

Recommendations/Comments

Pad-weighing offers a potential for quantifying the degree of incontinence. For short-term tests, the experimental conditions should be described. Standardized bladder filling volumes are generally recommended. For long-term tests, the time period should be as long as practical.

DOMAIN OF CLINICIAN'S OBSERVATIONS

Pelvic Muscle Activity

This is the force of voluntary pelvic muscle contraction determined in direct (digital or air pressure) or indirect (surface electromyography) measures.

Accuracy is difficult to assess. Test-retest data are mainly based on correlations analysis, which is problematic to interpret.

Recommendations/Comments

At this time there is no conclusive information as to the potential of muscle activity as an outcome measure.

Researchers should clearly describe their instrument and procedure and provide reliability data or indicate their absence.

Cystometry

Cystometry variables include sensation, compliance, capacity, and activity of the bladder. The investigation can be carried out in a stationary or ambulatory setting. Data obtained are method-dependent.

Accuracy is difficult to assess, since variation in various parameters is significant [Lose and Thyssen, 1996; Sørensen et al., 1988; Sørensen, 1988].

Recommendations/Comments

It is recommended that authors provide reproducibility data or indicate their absence.

Uroflowmetry

Normal female voiding produces a bell-shaped curve on uroflowmetry. Voiding dysfunctions may give rise to an intermittent or multiple peaked flow, suggesting either abdominal straining, unsustained bladder contractions, intermittent sphincter contractions, or bladder outlet obstruction.

Maximum and average flow are directly dependent upon voided volume. Consequently, interpretation and comparison of flow rates require that the voided volume is taken into account. Various nomograms are available. Flow rate depends to a lesser extent on age and sex, while parity, weight, menstrual cycle phase, or menopause status seem not to influence flow rate [Fantl et al., 1982].

A certain test-retest variation exists in terms of flow rates and pattern [Sørensen et al., 1988; Sørensen, 1988].

Recommendations/Comments

Uroflowmetry represents a simple initial test to assess the emptying phase of the lower urinary tract. However, researchers should describe their methodology and provide reproducibility data or indicate their absence.

Pressure-Flow Study

This represents the simultaneous study of urine flow and intravesical or detrusor pressure during voluntary voiding. Graphic plotting of pressure-flow data enables classification of individual patients into one of the following groups: obstructed, equivocal, or unobstructed patterns.

In patients with an irregular micturition pattern, the technique should be accomplished with simultaneous EMG recording to allow differentiation between abdominal straining, intermittent detrusor contraction, or detrusor/sphincter dyssynergia. Pressure-flow data are subject to significant test-retest variation [Lose and Thyssen, 1996].

Recommendations/Comments

Researchers should describe their methodology and reproducibility data or indicate their absence.

Electromyography

Electromyographic (EMG) recordings are performed to examine the activity (behavior) of pelvic and sphincter muscles during different maneuvers, particularly during bladder filling and voiding. This type of EMG is also called kinesiologic EMG. Both surface and intramuscular electrodes may be used for recording. Surface electrodes are more nonselective and may be used for assessing the overall behavior of the muscle bulk underlying the electrodes. Intramuscular electrodes (needle or wire)

provide more selective recordings of higher quality, but reflect only the activity of a small muscle volume.

Electromyographic techniques (concentric-needle EMG, single-fiber EMG) allow acquisition of electrophysiologic parameters that define a striated muscle as normal or abnormal. With concentric-needle EMG, muscle denervation and reinnervation can be assessed by recording of spontaneous activity and motor-unit action potentials. Single-fiber EMG reveals muscle changes due to reinnervation by recording muscle-fiber density [Flink, 1997; Vodusêk, 1996].

The reliability of EMG recordings is poorly determined.

Recommendations/Comments

Surface electrodes can be used to identify striated muscle activity during different maneuvers, particularly during bladder filling and voiding, and also as a biofeedback therapeutic technique.

Intramuscular electrodes can be used in the more precise studies to identify the activity patterns of individual striated muscles. Special needle electrodes are required to assess denervation and reinnervation of the pelvic floor muscles and striated sphincters.

Researchers should describe their methodology and reproducibility data or indicate their absence.

Nerve Stimulation

This entails neurophysiological tests involving stimulation of nervous system structures.

Electrical and magnetic stimulators are used to depolarize nervous structures (particularly the pudendal nerve and its branches, sacral roots, and the motor cortex). Recordings are obtained both from pelvic floor muscles and striated sphincters, but also from nerve structures, particularly over the somatosensory cortex. Several tests exist and can be broadly grouped into evoked potentials and reflex responses.

The reproducibility of several tests is fairly acceptable if the tests are performed under standardized conditions, but no overall accepted standards for individual tests as yet exist.

Recommendations/Comments

The clinical usefulness of these tests has not yet been clarified. They seem to be helpful in research. Researchers should describe their methodology and their laboratories' normal values. They should indicate reproducibility data or indicate their absence.

Urethral Pressure Measurement

Conceptually, urethral closure pressure represents the ability of the urethra to prevent urine leakage. Technically, it represents the difference between the pressure at one given point in the urethra and the simultaneously recorded pressure in the bladder.

Measurements of urethral closure pressure may be made at one point in the urethra over a period of time (continuous urethral pressure recording), or at several

points along the urethra consecutively, forming a urethral pressure profile (UPP). These measurements can be made at rest or during exertion.

The results of urethral pressure measurements are highly influenced by methodological and biological factors and have been found to have significant test-retest variability. The overlap between both the static and dynamic urethral closure measurements in continent and incontinent women is significant [Lose, 1997].

Recommendations/Comments

Urethral pressure profile parameters are of limited value in the assessment of urethral sphincter function. Urethral pressure measurement may be useful in evaluating local pathology (e.g., diverticulum and stricture), in assessment of changes with intervention, and in therapeutic selection (e.g., intrinsic sphincter deficiency).

Imaging

Imaging techniques in terms of radiography, ultrasound, and magnetic resonance imaging (MRI) are used to provide morphological and functional information on the lower urinary tract and the pelvic floor. Until now, as with conventional radiology, neither ultrasound nor MRI appeared to provide conclusive and discriminatory diagnostic information. Scant information is available on reproducibility of measurement, while significant inter- and intraobserver variation in the assessment of radiography pictures has been reported.

Recommendations/Comments

At this time, imaging techniques are of limited value as outcome measures in female incontinence. Researchers should clearly describe their instrument and procedure and provide reliability data or indicate their absence.

DOMAIN OF QUALITY-OF-LIFE MEASUREMENTS

Health-related quality of life (HRQOL) is a multidimensional construct that refers to an individual's perceptions of the effect of a health condition or disease and its subsequent treatment. Primary domains of HRQOL include physical, psychological, and social functioning, overall life satisfaction/well-being, and perceptions of health status. Secondary domains include somatic sensations (symptoms), sleep disturbance, intimacy and sexual functioning, and personal productivity (e.g., household, occupational, volunteer, or community activities).

Three measurement approaches are commonly used to assess HRQOL: generic, condition-specific, and dimension-specific. Generic HRQOL instruments are designed to be used across multiple disease states or conditions, and allow for comparisons across groups by having established age and gender norms. Condition-specific instruments are designed to measure the impact of a particular disease or condition. These instruments tend to be more responsive than generic instruments in detecting treatment effects. Symptom scales are considered condition-specific; generally, these scales should include measurement of the presence of a symptom as well as its ''bothersome'' or ''troublesome'' nature. The majority of generic and condition-specific instruments are multidimensional, i.e., they measure more than one aspect of HRQOL. Dimension-specific instruments, in contrast, are designed to assess

a single component of HRQOL, such as emotional distress. The trend in assessing HRQOL outcomes has been toward the use of a multidimensional generic and/or condition-specific instrument, supplemented with dimension-specific instruments, as needed [Grimby et al., 1993; Shumaker et al., 1994; Wagner et al., 1996; Naughton and Wyman, 1997].

Recommendations/Comments

The selection of an HRQOL instrument should be based upon the purpose of the study. Descriptive or epidemiological studies should consider both generic and condition-specific instruments. Intervention studies should include a condition-specific instrument. Dimension-specific instruments should be used when more detail about a specific subdomain of HRQOL is desired. Researchers should define HRQOL as it is conceptualized for their study, clearly describe their instrument(s) and data collection, and provide reliability data if available.

It is recommended that researchers select instruments with reliability and sensitivity. In adopting HRQOL instruments, we recommend comparing results obtained in the study population with published norms. If a new condition-specific instrument will be used in a study, adequate pretesting should be done to establish its clinimetric characteristics (e.g., reliability and sensitivity).

DOMAIN OF SOCIOECONOMIC COSTS

Cost-effectiveness analysis is a method to decide on the relative merits of alternative courses of action by comparing costs to achieve a given health effect. Costs can be divided into direct and indirect costs. Direct costs are all costs incurred in the routine care, diagnosis, treatment, and management of adverse consequences associated with treatment or with being incontinent, which can be assessed from the market value of goods and services. Indirect costs are those incurred as a result of being incontinent or of having to care for someone who is incontinent. These opportunity costs, e.g., time lost enjoying leisure activity or engaging in productive effort, are more difficult to quantify, and thus, are usually not measured. If morbidity and mortality are considered, they should be converted to a monetary value.

Effectiveness is characterized as a specific health outcome, e.g., disease prevented or cured, function restored, or symptoms alleviated.

Direct costs are all costs incurred in routine care, diagnosis, treatment of incontinence, and management of adverse consequences associated with treatment or with being incontinent. Indirect costs are those incurred as a result of being incontinent or having to care for someone who is incontinent. These opportunity costs, e.g., time lost enjoying leisure activity or engaging in productive effort, are more difficult to quantify, and thus, are usually not measured.

Recommendations/Comments

The types of costs and how they are determined should be clearly articulated. Typical resource uses to be collected in a study include:

1. Clinician's time/service
2. Laboratory and imaging studies

3. Treatment expenses (e.g., procedural charges, medication costs)
4. Supplies (disposable and reusable products)
5. Side and adverse effects and their management
6. Travel costs to obtain health care
7. Loss of wages from receiving health care or surgery.

CONCLUSIONS

Research on lower urinary tract symptoms in adult women remains impaired by lack of standardization of outcome variables. The use of primary and secondary outcomes on different areas or domains should help overcome some of these difficulties. However, it is imperative that investigators focus attention on these issues. Studies leading to the development and standardization of outcomes should be entertained. Success in this regard will not only enhance knowledge of actual treatment strategies but stimulate the development of new ones.

REFERENCES

Bland JA, Altman DG (1986): Statistical method for assessing agreement between two methods of clinical measurement. Lancet 1:307–310.

Fantl JA, Smith PS, Schneider V, Hurt WG, Dunn LJ (1982): Fluid weight uroflowmetry in women. Am J Obstet Gynecol 145:1017–1024.

Flink R (1997): Clinical neurophysiological methods for investigating the lower urinary tract in patients with micturition disorders. Acta Obstet Gynecol Scand [Suppl] 166:50–58.

Gjørup T (1997): Reliability of diagnostic tests. Acta Obstet Gynecol Scand [Suppl] 166:9–14.

Grimby A, Milson L, Molander U, Wiklund I, Ekelund P (1993): The influence of urinary incontinence on the quality of life of elderly women. Age Ageing 22:82–89.

Larsson G, Victor A (1988): Micturition patterns in a healthy female population, studied with a frequency/volume chart. Scand J Urol Nephrol [Suppl] 144:53–57.

Larsson G, Abrams P, Victor A (1991): The frequency/volume chart in detrusor instability. Neurourol Urodyn 10:533–543.

Lose G (1997): Urethral pressure testing. Acta Obstet Gynecol Scand [Suppl] 166:39–42.

Lose G, Thyssen H (1996): Reproducibility of cystometry and pressure-flow parameters in female patients. Neurourol Urodyn 15:302–303.

Lose G, Versi E (1992): Pad-weighing tests in the diagnosis and quantification of incontinence. Int Urogynecol J 3:324–328.

Naughton MJ, Wyman JF (1997): Quality of life of geriatric patients with lower urinary tract dysfunctions. Am J Med Sci 314:219–224.

Shumaker SA, Wyman JF, Uebersac J, McClish DK, Fantl JA (1994): Health-related quality of life measures for women with urinary incontinence: The urogenital Distress Inventory and the Incontinence Impact Questionnaire. Quality Life Res 3:291–306.

Sørensen S (1988): Urodynamic investigations and the reproducibility in healthy postmenopausal females. Scand J Urol Nephrol [Suppl] 114:42–47.

Sørensen S, Gregersen H, Sørensen M (1988): Long term reproducibility of urodynamic investigation in healthy females. Scand J Urol Nephrol [Suppl] 114:35–41.

Victor A (1990): Pad weighing tests—A simple method to quantitate urinary incontinence. Ann Med 22:443–447.

Vodusêk DB (1996): Evoked potential testing. Urol Clin North Am 23:427–446.

Wagner TH, Patrick DL, Buesching DP (1996): Quality of life of persons with urinary incontinence. Development of a new measure. Urology 47:67–73.

Wyman JF, Choi SC, Harkins SW, Wilson MS, Fantl JA (1988): The urinary diary in evaluation of incontinent women: A test-retest analysis. Obstet Gynecol 71:812–817.