



## Review – Incontinence

# Outcomes in Urinary Incontinence: Reconciling Clinical Relevance with Scientific Rigour

Cara Tannenbaum<sup>a</sup>, Jacques Corcos<sup>b,\*</sup>

<sup>a</sup> University of Montreal, Montreal, QC, Canada

<sup>b</sup> Urology Department, Jewish General Hospital, McGill University, Montreal, QC, Canada

### Article info

#### Article history:

Accepted February 12, 2008

Published online ahead of  
 print on February 20, 2008

#### Keywords:

Patient outcome measures

Review

Urinary incontinence

### Abstract

**Objectives:** To aid clinicians in selecting an outcome measure for the assessment of urinary incontinence (UI), from the perspective of both scientific rigour and clinical utility.

**Methods:** We conducted a comprehensive review of the literature on outcome measures for the assessment of UI in adults. Tools were classified by instrument type (ie, subjective measures, objective measures, clinical observations, quality of life, and combined instruments) and assessed for scientific rigour based on their psychometric properties (reliability, validity, responsiveness). The clinical relevance of each tool was considered in terms of current usage and practicality.

**Results:** The most rigorous validation processes were identified for quality-of-life questionnaires, including the Incontinence Impact Questionnaire, King's Health Questionnaire, Incontinence Quality of Life questionnaire, and Urogenital Distress Inventory. Bladder diaries, goal-attainment scales, and combined measures such as the International Consultation on Incontinence Questionnaire appear to be more practical for use in clinics. The Clinical Global Impression of Improvement is the outcome most widely used clinically, but least well validated.

**Conclusions:** To elevate the level of outcome assessment for UI to meet that of other urology specialties, it is necessary to reconcile the realities of clinical practice with the scientific rigour of UI research, and to mainstream outcome measures that are reciprocally translatable between the two settings.

© 2008 European Association of Urology. Published by Elsevier B.V. All rights reserved.

\* Corresponding author. Urology Department, Jewish General Hospital, McGill University, 3755 Cote Ste-Catherine, Montreal, QC, Canada H3T 1E2. Tel. +1 (514) 340 8222#5166; Fax: +1 (514) 340 7559.  
 E-mail address: [jcorcos@uro.jgh.mcgill.ca](mailto:jcorcos@uro.jgh.mcgill.ca) (J. Corcos).

## 1. Introduction

One of the greatest challenges in the nonsurgical management of urinary incontinence (UI) is determining when a satisfactory improvement has been achieved. In the absence of complete dryness, clinical trialists and UI specialists have yet to concur on a definition of clinically meaningful change in UI status for their patients. Consider a 70% reduction in the frequency of UI episodes. Is this a genuine marker of efficacy for a new overactive bladder medication if the patient still reports poor UI-related quality-of-life (QOL), is reluctant to socialise, and has to change pads four times per day?

Despite the existence of several subjective and objective outcome measurements for UI, no definitive UI measurement protocol has been established for research or clinical practice to determine the impact of UI on the patient and its response to therapeutic interventions. The 3rd International Consultation on Urinary Incontinence recommends using a combination of both subjective and objective outcome measurements [1]. However, due to time constraints, unfamiliarity with many of the tools, and a lack of widely accepted efficacy criteria, these outcome measures are seldom used by clinicians in routine office practice [2]. Even when they are used, there is rarely consensus among the different measures as to the benefits achieved [3].

In the midst of this confusion, each clinician must decide on the most appropriate outcome to use in his or her practice. In research, selection of UI outcome measures is left to the individuals designing the studies and varies considerably from one study to the next. As a result, clinicians are often unable to compare their findings to the outcomes obtained in research studies, and comparison of improvements across patients and interventions is not possible. Furthermore, the type of UI is of paramount importance in clinical practice and few of the measures commonly used in clinical practice consider the aetiology of UI. This is in contrast to well-conducted studies of other urology specialties such as prostate cancer, in which the same standardised outcomes (biologic response [prostate-specific antigen], survival time, disease-free survival, etc) are used across different studies and settings.

Unlike prostate cancer, UI is not life-threatening and has no biologic markers to characterise the disease or its response to treatment. UI is actually a symptom, not a disease, which can be explained by different aetiologies. The severity and impact of UI is perceived differently by patients according to certain factors, including their sex, the degree of leakage, the activity level of the patient, and the familial and social

environment. It therefore seems logical that the type of outcomes used in studies and practice will vary depending on the goals of the clinician or researcher, the type of UI being evaluated, and the outcomes desired by each patient.

A number of outcome measures are currently being used in research trials for UI. However, even in research studies marked heterogeneity exists with a minority of studies using the most rigorously validated instruments. According to Avery et al, among 130 randomised, controlled trials for treatment of UI published from 2001 to 2004, only 38% used the most scientifically validated UI symptom and QOL questionnaires as outcome measures [4]. These included the Incontinence Impact Questionnaire (IIQ) or its short form (IIQ-7), the King's Health Questionnaire (KHQ), the Incontinence Quality of Life questionnaire (I-QOL), or the Urogenital Distress Inventory (UDI) or its short form (UDI-6). In a separate review, Ross et al confirmed that the IIQ was the most frequently used measure of incontinence-specific QOL used in research studies, followed by the I-QOL [5].

Few data are available detailing clinicians' routine use of outcome measures in practice. A postal survey of 156 UI specialist members of the United Kingdom branch of the International Continence Society indicated that 42% of respondents used subjective improvement (patient observations of the severity of urine loss) alone, and 36% used improvement in QOL as a clinical outcome measure [2]. Pad testing and the results of urodynamic tests to evaluate outcomes were rarely used. A survey of Canadian family physicians revealed that the majority see patients with UI in their practice, but only one third have an organised plan for evaluation and treatment outcome [6]. Many of these physicians use their global impression of change rather than any standardised QOL questionnaire, pad test, or bladder diary to assess patient's improvement.

The aim of this review is to aid clinicians in selecting the most appropriate UI outcome measures for use in their practice. Both scientific rigour and practical features are considered. To elevate the level of outcome assessment for UI to meet that of other urology specialties, measures that can transfer easily between research and practice are emphasised.

## 2. Methods

### 2.1. Search strategy

A comprehensive review of published literature was conducted using the following databases: CINAHL, from 1982 to

2007; Cochrane systematic reviews, from 1980 to 2007; Psychology and Social Tools, from 1985 to 2007; MEDLINE (PubMed), from 1980 to 2007; and ISI Science Web databases, from 2003 to 2007. The Medical Subject Heading (MeSH) term “urinary incontinence” was combined with the following MeSH terms in the literature search strategy: “outcome measures,” “quality of life,” “VAS,” “goal-attainment scaling,” “pad test,” “voiding diary,” and “self-efficacy.” From these queries, 947 articles were retrieved.

The abstracts from these articles were reviewed by three researchers for the following inclusion criteria: adult patients, outcome measures assessing UI symptoms and QOL, validation of instrument, original research and review articles, and English language. In addition, references from these articles were reviewed to identify further information on developed instruments and literature concerning validation findings. The textbook of the 3rd International Consultation on Incontinence was included [1].

After review, 61 articles were selected and classified into the following topics: subjective measures, objective measures, QOL measures, and clinical considerations. In addition, the psychometric properties (reliability, validity, and responsiveness) of each outcome measure were evaluated. Only outcome measures having undergone a rigorous evaluation process or currently used in practice are strategically appraised in this review.

## 2.2. Classification of outcome measures

We adhered to the classification of outcome measures put forth by Lose and colleagues, whose system divided outcome measures for UI into five types [7]: (1) subjective measures, for example, self-reported patient observations of the frequency, quantity, or magnitude of UI symptomatology; (2) objective measures, that is, quantification of urine loss; (3) clinical observations, that is, cystometry, uroflowmetry; (4) QOL questionnaires; and (5) socioeconomic measures. We added the category of combined measures to reflect advancements in the field since Lose’s classification.

## 2.3. Properties of outcome measures

Within each category outcome measures were evaluated based on their psychometric and clinometric properties. Psychometric properties are attributes of the outcome measure that allow it to measure what it is supposed to be measuring [8]. Validity, reliability, and responsiveness are the most scientifically rigorous markers of a psychometrically sound outcome measure (Table 1). Clinometric properties, on the other hand, refer to a measure’s ability to reflect clinically relevant information on the condition being studied, with the potential to guide practice, be acceptable to patients and clinicians, and be easy to use and to interpret [9].

## 3. Results

Table 2 summarises all outcome measures according to their type, scientific rigour, and clinical relevance.

### 3.1. Subjective measures

#### 3.1.1. Bladder diaries

Subjective measures involve patient or physician reports of UI status. Frequency, quantity, or magnitude of UI is typically ascertained through either patient recall, daily self-monitoring bladder diaries, or questionnaires. Recall data are not necessarily unreliable [10,11], but diary recordings are more accurate [11–14]. Diaries permit the assessment of parameters such as incontinence episodes, frequency of voiding, type and amount of fluid intake, total 24-h voided volume, and circumstances associated with leaks. The recommended diary length to obtain reliable estimates of UI frequency depends on the absolute frequency and variability of each

**Table 1 – Psychometric properties of useful outcome measures [8]**

| Criterion                | Definition                                                                                               |
|--------------------------|----------------------------------------------------------------------------------------------------------|
| Validity                 |                                                                                                          |
| •Face validity           | Measure appears to be true to its intended objective                                                     |
| •Content validity        | Measure contains a sufficiently representative collection of elements to allow for a complete evaluation |
| •Criterion validity      | Comparison of a given measure against a gold standard                                                    |
| •Construct validity      | Comparison of measurement results with their anticipated outcome based on preconceived theories          |
| Reliability              |                                                                                                          |
| •Internal consistency    | The extent to which items within the measure are related to each other                                   |
| •Test-retest reliability | Subjects who have not changed evaluated at different points in time                                      |
| •Interrater reliability  | Parallel evaluations conducted by different raters                                                       |
| Responsiveness           | Ability to measure a clinically significant difference over time                                         |

Table 2 – A summary of outcome measures by type

| Outcome measure                                                                                                                                                                       | Description                                                                                                                                                                                                                                                               | Type of UI | Psychometric properties                                                                                                                                   | Pros                                                                                                                                                                                                                                                                                                                                                                      | Cons                                                                                                                                                                                                                                      |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>1. Subjective measures</b>                                                                                                                                                         |                                                                                                                                                                                                                                                                           |            |                                                                                                                                                           |                                                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                           |
| <i>Bladder diary</i>                                                                                                                                                                  | Patients are asked to fill out a chart that records the frequency and time of voiding, incontinence episodes, type and amount of fluid intake, total 24-h voided volume and the circumstances associated with leaks. Usually 1 page/d; 3–7-d assessments are recommended. | All        | Acceptable validity and reliability. The accuracy and responsiveness decrease for patients with fewer than three episodes of incontinence per day.        | <ul style="list-style-type: none"> <li>• Talking about the number of UI episodes per day is relevant to the patient.</li> <li>• Easy to objectively compare status before and after intervention.</li> <li>• The diary may serve as an intervention in and of itself for raising awareness of lifestyle habits.</li> <li>• Used routinely in research studies.</li> </ul> | <ul style="list-style-type: none"> <li>• Difficult for some older patients to complete.</li> <li>• Logistically more challenging; requires effort to send the diary to patients before the visit and have them complete it.</li> </ul>    |
| <i>Symptom questionnaires</i><br>For women: Incontinence Severity Index, Urogenital Distress Inventory, Bristol Female Lower Urinary Tract Symptoms<br>For men: ICSmale questionnaire | Self-administered or physician-administered questionnaires querying the severity of symptoms and the degree to which UI symptoms are troubling. Number of items vary from 2 to 19.                                                                                        | All        | Grade A (highest) recommendation from the Symptoms and Quality of Life assessment committee of the 3rd International Consultation on Incontinence         | <ul style="list-style-type: none"> <li>• Easy for patients to understand</li> <li>• Clinically relevant</li> <li>• Used in many research studies</li> </ul>                                                                                                                                                                                                               | <ul style="list-style-type: none"> <li>• Scores may be complicated to compute and interpret.</li> <li>• Different versions are required for female and male patients.</li> <li>• Symptoms do not always correlate with impact.</li> </ul> |
| <i>Goal-attainment scale</i>                                                                                                                                                          | Evaluates progress towards personal goals. Uses a 5-point scale ranging from –2 to 2, with achievement of the UI-related goal being rated as 0.                                                                                                                           | All        | Very good validity, reliability, and responsiveness.                                                                                                      | <ul style="list-style-type: none"> <li>• Patient-centered</li> <li>• Clinically relevant</li> <li>• Validated in the elderly</li> <li>• Convenient to do during the patient encounter</li> </ul>                                                                                                                                                                          | <ul style="list-style-type: none"> <li>• May be time-consuming</li> <li>• Some individuals are unable to formulate realistic goals.</li> <li>• Used infrequently in research studies</li> </ul>                                           |
| <i>Global impression of improvement</i>                                                                                                                                               | A single-item question that asks how UI status after treatment compares with before treatment. Seven response options (very much worse to very much better) are possible.                                                                                                 | All        | Only the validity has been tested for the Patient's Global Impression of Improvement. No validation studies have been conducted on the physician's index. | <ul style="list-style-type: none"> <li>• Quickly and easily interpretable from a clinical standpoint</li> <li>• Commonly used in practice</li> </ul>                                                                                                                                                                                                                      | <ul style="list-style-type: none"> <li>• Physician's index has a lack of accuracy with respect to patients' perceptions.</li> <li>• Used infrequently in research studies</li> </ul>                                                      |
| <i>Self-efficacy index</i>                                                                                                                                                            | A 12-item questionnaire measuring the patient's degree of confidence in retaining urine under different sets of circumstances.                                                                                                                                            | All        | Valid, reliable, and responsive                                                                                                                           | <ul style="list-style-type: none"> <li>• Clinically meaningful</li> <li>• Can be self-administered or administered by the clinician</li> </ul>                                                                                                                                                                                                                            | <ul style="list-style-type: none"> <li>• Visual analogue scales may be difficult for elderly patients.</li> <li>• More time-consuming than symptom questionnaires</li> </ul>                                                              |

Table 2 (Continued)

| Outcome measure                                                                                                                                                  | Description                                                                                     | Type of UI                                   | Psychometric properties                                                            | Pros                                                                                                                                                                                                                                                                                | Cons                                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|----------------------------------------------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>2. Objective measures</b>                                                                                                                                     |                                                                                                 |                                              |                                                                                    |                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                              |
| <i>Pad weighing</i>                                                                                                                                              | 1-h or 24-h measure of pad wetness                                                              | All                                          | Variable                                                                           | <ul style="list-style-type: none"> <li>Used in research studies</li> </ul>                                                                                                                                                                                                          | <ul style="list-style-type: none"> <li>Can be labour-intensive and time-consuming if done in the clinic.</li> <li>Distasteful to the patient if done at home.</li> <li>Test-retest reliability is considerable, especially in the elderly, limiting its usefulness in evaluating outcomes.</li> </ul>                                                                        |
| <i>Wet checks</i>                                                                                                                                                | Examination of pads/diapers at regular intervals to identify episodes of incontinence           | All                                          | Valid, reliable, and possibly responsive depending on the frequency of the checks  | <ul style="list-style-type: none"> <li>Most practical method in patients who are cognitively impaired</li> <li>Used routinely in the nursing home and in nursing home research</li> </ul>                                                                                           | <ul style="list-style-type: none"> <li>Dependent on the reliability of the caregiver</li> <li>Inappropriate for cognitively intact individuals</li> </ul>                                                                                                                                                                                                                    |
| <b>3. Clinical observations</b>                                                                                                                                  |                                                                                                 |                                              |                                                                                    |                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                              |
| <i>Cystometry, uroflowmetry, multichannel urodynamics, ultrasound</i>                                                                                            | Office procedures to measure bladder sensation, capacity, activity, and urine flow              | All                                          | Test-retest and interrater reliability are variable.                               | <ul style="list-style-type: none"> <li>May be useful for measuring the effect of pharmacologic management</li> </ul>                                                                                                                                                                | <ul style="list-style-type: none"> <li>Time-consuming and may be invasive</li> <li>Rarely repeated after treatment; not useful as outcome measures</li> <li>Less reliable in the elderly</li> </ul>                                                                                                                                                                          |
| <b>4. QOL questionnaires</b>                                                                                                                                     |                                                                                                 |                                              |                                                                                    |                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                              |
| Incontinence Impact Questionnaire (IIQ) or its short form (IIQ-7), the King's Health Questionnaire (KHQ), the Incontinence Quality of Life questionnaire (I-QOL) | Questionnaires of variable length (7–22 items) measuring the impact of UI on the patient's life | All<br>Type-D specific questionnaires exist. | Highest grade A recommendation from the International Consultation on Incontinence | <ul style="list-style-type: none"> <li>Shortened versions exist.</li> <li>May be self-administered or administered by the clinician</li> <li>Appropriate for measuring the impact of UI on the patient's life</li> <li>Starting to be used routinely in research studies</li> </ul> | <ul style="list-style-type: none"> <li>Full versions are lengthy to administer.</li> <li>Complicated scoring system involving subscales and summary scores</li> <li>Correct interpretation may be difficult because the direction of scores vary from one index to the next.</li> <li>No thresholds for guiding treatment</li> <li>Poor correlation with symptoms</li> </ul> |

Table 2 (Continued)

| Outcome measure                                                                                            | Description                                         | Type of UI | Psychometric properties                                                            | Pros                                                                                                                                                                                                                                           | Cons                                                                                      |
|------------------------------------------------------------------------------------------------------------|-----------------------------------------------------|------------|------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|
| <b>5. Combined symptom and QOL questionnaires</b>                                                          |                                                     |            |                                                                                    |                                                                                                                                                                                                                                                |                                                                                           |
| For men and women: International Consultation on Incontinence Questionnaire (ICI-Q)                        | Short-form questionnaires querying symptoms and QOL | All        | Highest grade A recommendation from the International Consultation on Incontinence | <ul style="list-style-type: none"> <li>• Encompasses all aspects of incontinence</li> <li>• Short and easy to administer</li> <li>• Scoring systems are simpler, but not always clinically meaningful, especially during follow-up.</li> </ul> | <ul style="list-style-type: none"> <li>• Some questionnaires are sex-specific.</li> </ul> |
| For women: Bristol Female Lower Urinary Tract Symptoms-short form questionnaire                            |                                                     |            |                                                                                    |                                                                                                                                                                                                                                                |                                                                                           |
| For men: ICSmaleSF                                                                                         |                                                     |            |                                                                                    |                                                                                                                                                                                                                                                |                                                                                           |
| UI = urinary incontinence; ICS = International Continence Society; QOL = quality of life; SF = short form. |                                                     |            |                                                                                    |                                                                                                                                                                                                                                                |                                                                                           |

individual's UI episodes [11]. Nocturnal events are more variable than daytime events [11,13], and incontinence frequency is more variable for urge incontinence than stress incontinence [11]. A 7-d diary provides acceptable accuracy and responsiveness for patients with a mean of three episodes of daytime incontinence [11,14,15], but a diary length of 31 d is required to detect a 50% change in UI severity for subjects with an expected daytime incontinence frequency of 0.5 (once every 2 d) [11]. Two-day and 3-d diaries are associated with decreased reliability [11,16] but higher compliance rates [16]. The impact of UI frequency on physical and psychological function is difficult to interpret [17,18]. Diaries serve as one of the primary end points in most pharmacologic studies of UI, with improvements reported as percentage reductions in UI from baseline or the absolute number of episodes of UI per day or both. For the most part, diaries have not been regularly adopted in practice. This may be due to the logistic difficulty of having patients fill out the diary prior to their first visit/before initiation of treatment and then immediately prior to their follow-up visit. Also, it is unclear whether using the concept of percentage reduction in UI frequency to compare baseline and UI frequency after intervention is practical or meaningful to patients or physicians.

### 3.1.2. Questionnaires on symptoms

Several UI severity scales have rigorous psychometric testing. For women, these include the Incontinence Severity Index [19,20], the UDI short and long forms [21-23], and the Bristol Female Lower Urinary Tract Symptoms (BFLUTS) questionnaire [24]. For men, the International Continence Society

(ICS) male questionnaire is recommended [25]. Some of the questionnaires are brief, but all require calculation of a final score. For instance, the Incontinence Severity Index comprises only two questions: How often do you experience urine leakage? (four levels), and how much urine do you lose? (two levels). The index is then calculated by multiplying the two responses together and is categorised into slight, moderate, severe, and very severe. The main criticism of symptoms questionnaires relates to the known discrepancy between UI symptoms and impact, as well as the clinical interpretation of the ratings. There may be value in knowing that the patient has improved from a very severe rating to a moderate one, but more information must be obtained to decide what this really means and whether treatment should be considered successful or further therapeutic options pursued.

### 3.1.3. Goal-attainment scale

Goal-attainment scaling is a systematic measure of treatment outcome that evaluates progress towards patient-determined goals. It embodies the hallmark of clinical care, requiring that patients formulate specific goals in relation to their UI treatment. For example, a goal for stress UI might be to be able to run for 30 min without needing to use a pad. For urge UI, one would like to be able to take a 3-h outing without experiencing urinary urgency or leakage. Each goal is then put on a 5-point scale ranging from -2 to 2, with achievement of the goal being rated as 0, and higher and lower benchmarks ranked accordingly. Other scales use ratings of worse, slightly worse, same, slightly better, and much better than expected. Goal-attainment scaling has

been used to assess outcomes in surgical studies of UI [26] and has been successfully evaluated for validity, reliability, and responsiveness in geriatric populations [27,28]. Its theoretical advantages are its patient-centeredness and clinical relevance; however, some health professionals find it time-consuming to develop a scale for each individual [29], and some individuals are unable to formulate specific goals. Though not being used routinely in clinical trials or private practice, goal-attainment scaling is a useful outcome measure to consider for spanning the two worlds.

#### 3.1.4. Global impressions of severity or improvement

These indexes provide global assessments of UI severity or response to treatment. For instance, the Patient's Global Impression of Improvement index is a single-item question that asks how UI status after treatment compares with before treatment [30]. Seven responses are possible: very much, much, and a little better, no change, and a little, much, and very much worse. Although it has great clinometric appeal, and is probably the most common outcome measure used in practice, only its validity has been tested [30]. The Clinician's Global Impression of Improvement is another subjective global rating measure, similar to the Patient's Global Impression of Improvement scale, where the clinician is asked to gauge the patient's degree of change in UI following intervention [31]. There have been no validation studies on the physician's measure, except to document its poor concordance with patients' perceptions of improvement [31,32], so it is rarely used in research trials.

### 3.2. Objective measures

#### 3.2.1. Pad weighing

Objective measures quantify the degree of UI in terms of frequency or amount and mostly include pad weighing. Pad weighing is performed either at the patient's home or in the physician's office by noting perineal pad weights before and after specified time intervals where urine loss is measured as weight gain of the pad. Pad weighing has been accepted as a useful outcome in clinical trials, but a 1-h pad-weighing test performed in the clinic following provocative measures that elicit urine loss can be labour-intensive and time-consuming in a busy office practice [33,34]. Test-retest within-patient variability is considerable and limits the usefulness of this test in the analysis of treatment outcomes, for determining the impact of UI on function, or for guiding treatment choices.

#### 3.2.2. Wet checks

Wet checks refer to patient examination at regular intervals to identify episodes of incontinence and are used commonly in nursing homes and hospitals for cognitively or functionally impaired individuals with UI. Clinical studies indicate that this method is reliable in identifying incontinence [35]; however, the method depends on the staff's or caregiver's level of commitment for regularly performing and accurately recording the results of the checks.

### 3.3. Clinical observations

Clinical observations consist of more invasive methods of analysis. Cystometry and uroflowmetry are procedures to measure bladder sensation, capacity, and activity and urine flow, respectively [7]. Other clinical measures include multichannel urodynamic testing, a safe and effective parameter but complicated to perform and unfeasible in the busy office practice, and ultrasound, which can provide valuable information about postvoid residual urine although sensitivity typically decreases with volumes >200 ml [35]. All of these tests have good diagnostic reproducibility for UI and its underlying causes, but hold little or no utility as outcome measures for measuring improvements in UI status. This is especially true of urodynamic testing where the results correlate poorly with urinary symptoms and incontinence [36,37].

### 3.4. QOL questionnaires

UI-specific QOL scales were created to capture the impact of UI on individuals' lives and add an important dimension to the evaluation and assessment of UI [4,5]. Health-related QOL assessments have become an integral part of patient-reported outcomes in research and pharmacologic trials [38]. Vetted grade A (highest level according to their psychometric properties) QOL questionnaires used most commonly in research include the I-QOL [39], the IIQ and its short-form [21–23,40], and the KHQ [41]. However, only one third of UI specialists are currently using these questionnaires to guide treatment decisions in daily clinical practice [2]. The question is why. One possibility is their length, although short forms are now available. A second explanation is confusion regarding their complicated scoring systems. Minimally important changes range from 2% to 5%, but following numbers on questionnaires requires both time and calculation to derive a score that is not as instinctively easily interpretable for clinical treatment as laboratory tests or biomarkers.

Furthermore, there are no thresholds on QOL questionnaires for guiding treatment decisions, and controversy still reigns around the issue of whether UI should be treated at all if it is not bothersome to the patient. Significant discrepancy often exists between UI symptom severity and QOL, for example, a 50% reduction in UI episodes may be a life-changing improvement in one patient's QOL, but may confer no significant perceived benefit to another. For these reasons, though all experts agree that the impact of UI on the patient must be evaluated, the feasibility of using QOL questionnaires in practice remains to be established. Because the scope and content of the questionnaires vary considerably, physicians should review each one before selecting which, if any, are appropriate for their clientele. If the language of administration is other than English, proper validation and translation protocols are required.

### 3.5. Socioeconomic evaluation

Socioeconomic evaluation of UI typically falls under one (or more) of three parameters: cost, cost benefit, and cost effectiveness. These measures are not being used in clinical trials or in office practice at the current time, and except in nursing home settings, it is unlikely that they will become useful on a day-to-day basis in the future to measure outcomes. Consequently, they are not included in this review.

### 3.6. Combined symptom and QOL impact measures

Only one questionnaire, the International Consultation on Incontinence Questionnaire (ICI-Q), captures both symptoms and QOL impact for men and women and encompasses all aspects of incontinence. It has been successfully subjected to considerable psychometric testing, and the final short-form version comprises three scored items (assessment of the frequency, severity, and perceived impact of incontinence) as well as an unscored self-diagnostic item [42]. It is an excellent choice for clinical practice. Other validated combined questionnaires exist, such as the BFLUTS short-form questionnaire (BFLUTS-SF) for women [43] and the ICSmaleSF for men [44].

## 4. Specificity according to UI type

A trend that has become apparent in research is the development of questionnaires for use according to UI type. The IIQ and UDI have been modified for use specifically in patients with urge urinary incon-

tinence [45]; another is the Overactive Bladder Questionnaire (OAB-Q) [46]. The Stress and Urge Incontinence and Quality of Life Questionnaire (SUIQQ) is a combined symptoms and QOL questionnaire that includes separate indices for stress, urge, and QOL [47]. The difficulty with UI type-specific questionnaires in clinical practice is that they can only be administered to the patient once the type of UI has been confirmed. The logistics of administering a baseline type-specific UI questionnaire after the diagnostic evaluations have been completed are more complicated than routine administration of the questionnaire immediately before or during the first visit. Additionally, outcome measures should not be confused with diagnostic measures, which are developed, tested, and validated very differently.

## 5. Special considerations in the elderly

As the population ages and the prevalence of UI increases, clinicians will continue to see growing numbers of elderly patients with UI. Caveats regarding the use of outcome measures in the elderly are therefore worth considering. The importance of using the same outcome measure for all patients cannot be over-emphasised. Familiarity with the outcome measure of choice is required by physicians for cross-patient and cross-intervention comparisons. Selection of an outcome measure that is appropriate across age groups is therefore most practical, though not always feasible. Specifically, factors such as sexual and social activity will certainly vary with age and outcome measures should reflect this.

Bladder diaries have not been validated in frail elderly individuals who are housebound or residents at institutions [35]. Diaries are often not filled in properly by patients and patients may change their routine activities when they are using short-term bladder diaries (ie, for <3 d). Indeed, filling in the diary in itself may alter behaviour. In addition, self-assessment of symptoms may not be practical for elderly patients, who may have difficulties with comprehension, dexterity, vision, and urine measurement. Home pad-weighing tests for measuring urine loss may also not be reliable in older individuals. For QOL assessment, UI appears to have less of an impact on QOL in older individuals than in younger individuals [39,48]. In a study by Robinson et al [48], women aged  $\geq 80$  yr were 70% less likely to report being negatively affected by UI than women aged 60–65 yr. This may be related to the fact that the very elderly have other conditions

impinging on QOL and may have adapted behaviourally and psychologically to urine loss.

In general, to be robust, outcome measures must be validated for all populations for which they are intended. Two outcome measures have been developed and validated specifically for the older population: the Urge Impact Scale (URIS) [49] and the Geriatric Self-Efficacy Index for Urinary Incontinence (GSE-UI) [50]. The URIS-24 comprises 24 items and has been shown to be an internally consistent, highly reproducible tool for the assessment of QOL in men and women older than 60 yr experiencing urge incontinence. The GSE-UI draws on a separate concept, the confidence a person has in the ability to prevent urine loss under different sets of circumstances. The validity, reliability, and responsiveness have recently been evaluated, but its practicality remains to be tested in clinical practice.

## 6. Discussion

Beyond the need for an outcome measure to be scientifically rigorous, a useful measure must provide clear information that has the potential to guide clinical practice, be acceptable to patients and clinicians, and be easy to use and to interpret. Clearly, the broad range of current UI outcome variables needs to be streamlined to simplify the interpretation of clinical trials in UI and to apply more rigorous outcome measurement in everyday practice for UI. When considering the “best” outcome measure, we need to consider whether a single measure can be accurate enough for research yet practical enough for clinical practice.

Time constraints, length of the tool to be administered, its clinical relevance, and the ability to guide treatment decisions are the main elements that physicians consider during clinical encounters. Often, it is the response choices that pose problems, not the questions or tools themselves. The simplest and most direct method of patient interviewing is “yes/no” responses to a series of questions, but the disadvantage is the loss of ability to discern gradations of “yes” or “no,” necessary to conduct sensitive analyses in research. To elicit more detailed information, categorical scaling and cross-modality matching of degree of symptoms to line length (ie, visual analogue scale [VAS]) are used. Categorical scaling entails a choice of categories, ranging from three (eg, mild, moderate, or severe) to as many as nine; the standard range is four to seven categories. The VAS uses a line with the ends as the extreme boundaries; zero at the left end means no symptom experienced,

and the maximum at the right end signifies the worst possible symptom. However, the use of these response options may appear unnatural during a clinical encounter.

For outcome measures to be incorporated routinely into clinical practice, the issue of when to administer them must also be addressed. Should all patients fill out the outcome measure questionnaires in the waiting room prior to each visit? This might work if patients are identified at the reception as waiting to be seen for incontinence, or if their type of UI is known, in the case of UI type-specific questionnaires. For other outcome assessments, such as bladder diaries and pad tests, more structured administration protocols and instructions must be envisioned. Having additional allied health professionals as team members (nurse continence advisors or physiotherapists), greatly facilitates collection of outcome data.

Finally, collection of data is pointless unless outcomes are systematically examined and compared. This is true not only to satisfy patients but as a method of quality assurance. Quarterly or annual reviews of patient outcomes according to type of UI and type of intervention are required for clinicians to know if their practice results are replicating what has been observed in research studies. Internal feedback and audit on the effectiveness of our interventions as UI specialists is necessary for advancement of the field and for credibility in the eyes of the scientific community.

## 7. Conclusion

Each UI specialist should choose one or two of the outcome measures listed in this review for trial use in his or her practice. The alternative options, not to systematically measure UI outcomes or to keep outcome measures for research separate from those for clinical practice, constitute a setback for the specialty as a whole and patients in particular. This approach leads to research results not being applied to clinical practice, as is currently the case. Reconciling scientific rigour with clinical utility is thus the most useful strategy to advance the field of outcome measurement for UI.

*Author contributions:* Jacques Corcos had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

*Study concept and design:* Corcos and Tannenbaum.

*Acquisition of data:* Corcos and Tannenbaum.

*Analysis and interpretation of data:* Corcos and Tannenbaum.

*Drafting of the manuscript:* Tannenbaum.

Critical revision of the manuscript for important intellectual content: Corcos.

Statistical analysis: Corcos and Tannenbaum.

Obtaining funding: Corcos.

Administrative, technical, or material support: NA.

Supervision: NA.

Other (specify): None.

**Financial disclosures:** I certify that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/ affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: Tannenbaum and Corcos are consultants for Astellas, Triton, and Oryx Pharmaceuticals. Corcos also consults for Pfizer and Ortho-McNeil Pharmaceuticals.

**Funding/Support and role of the sponsor:** Collection of data and preparation.

## References

- [1] Abrams P, Cardozo L, Khoury S, Wein A. Incontinence. 3rd International Consultation on Incontinence, Monaco, June 2004, ed. 3. United Kingdom: Health Publications, 2004.
- [2] Robinson D, Anders K, Cardozo L, Bidmead J. Outcome measures in urogynaecology: the clinician's perspective. *Int Urogynecol J* 2007;18:273-9.
- [3] Albo M, Wruck L, Baker J, et al. The relationships among measures of incontinence severity in women undergoing surgery for stress urinary incontinence. *J Urol* 2007;177:1810-4.
- [4] Avery KNL, Bosch JLHR, Gotoh M, et al. Questionnaires to assess urinary and anal incontinence: review and recommendations. *J Urol* 1997;177:39-49.
- [5] Ross S, Soroka D, Karahalios A, Glazener CA, Hay-Smith EJ, Drutz HP. Incontinence-specific quality of life measures used in trials of treatments for female urinary incontinence: a systematic review. *Int Urogynecol J Pelvic Floor Dysfunct* 2006;17:272-85.
- [6] Swanson JG, Hutchison B. Urinary incontinence in Canada. National survey of family physicians' knowledge, attitudes, and practices. *Can Fam Physician* 2002;48:86-92.
- [7] Lose G, Fantl JA, Victor A, et al. Outcome measures for research in adult women with symptoms of lower urinary tract dysfunction. Standardization Committee of the International Continence Society. *Acta Obstet Gynecol Scand* 2001;80:981-5.
- [8] Finch E, Brooks D, Stratford PW, Mayo NE. Physical rehabilitation outcome measures: a guide to enhanced clinical decision making. ed. 2. Hamilton, ON: BC Decker; Canadian Physiotherapy Association; 2002.
- [9] Feinstein AR. *Clinimetrics*. New Haven, CT: Yale University Press; 1987.
- [10] Elser DM, Fantl JA, McClish DK. Comparison of "subjective" and "objective" measures of severity of urinary incontinence in women. Program for Women Research Group. *Neurourol Urodyn* 1995;14:311-6.
- [11] Homma Y, Ando T, Yoshida M, et al. Voiding and incontinence frequencies: variability of diary data and required diary length. *Neurourol Urodyn* 2002;21:204-9.
- [12] Wyman JF, Choi SC, Harkins SW, Wilson NS, Fantl JA. The urinary diary in evaluation of incontinent women: a test-retest analyses. *Obstet Gynecol* 1988;71:812-7.
- [13] Blanker MH, Bohnen AM, Groeneveld FPMJ, Bernsen RMD, Prins A, Bosch JLHR. Normal voiding patterns and determinants of increased diurnal and nocturnal voiding frequency in elderly men. *J Urol* 2000;164:1201-5.
- [14] Locher JL, Goode PS, Roth DL, Worrell RL, Burgio KL. Reliability assessment of the bladder diary for urinary incontinence in older women. *J Gerontol* 2001;56A:M32-5.
- [15] Brown JS, McNaughton KS, Wyman JF, et al. Measurement characteristics of a voiding diary for use by men and women with overactive bladder. *Urology* 2003;61:802-9.
- [16] Groutz A, Blaivas JG, Chaikin DC, et al. Noninvasive outcome measures of urinary incontinence and lower urinary tract symptoms: a multicenter study of micturition diary and pad test. *J Urol* 2000;164:698-701.
- [17] Fultz NH, Herzog AR. Self-reported social and emotional impact of urinary incontinence. *J Am Geriatr Soc* 2001;49:892-9.
- [18] Frazer MI, Haylen BT, Sutherst JR. The severity of urinary incontinence in women: comparison of subjective and objective tests. *BJU Int* 1989;63:14-5.
- [19] Sandvik H, Hunskaar S, Seim A, Hermstad R, Vanvik A, Bratt H. Validation of a severity index in female urinary incontinence and its implementation in an epidemiological survey. *J Epidemiol Community Health Med* 1993;47:497-9.
- [20] Hanley J, Capewell A, Hagen S. Validity study of the severity index, a simple measure of urinary incontinence in women. *BMJ* 2001;322:1096-7.
- [21] Uebersax JS, Wyman JF, Shumaker SA, McClish DK, Fantl JA, and the Continence Program for Women Research Group. Short forms to assess life quality and symptom distress for urinary incontinence in women: the Incontinence Impact Questionnaire and the Urogenital Distress Inventory. *Neurourol Urodyn* 1995;24:131-9.
- [22] Van der Vaart CH, de Leeuw JRJ, Roovers JPWR, Heintz APM. Measuring health-related quality of life in women with urogenital dysfunction: the Urogenital Distress Inventory and Incontinence Impact Questionnaire revisited. *Neurourol Urodyn* 2003;22:97-104.
- [23] Hagen S, Hanley J, Capewell A. Test-retest reliability, validity, and sensitivity to change of the Urogenital Distress Inventory and the Incontinence Impact Questionnaire. *Neurourol Urodyn* 2002;21:534-9.
- [24] Jackson S, Donovan J, Brookes S, Eckford S, Swithinbank L, Abrams P. The Bristol Female Lower Urinary Tract Symptoms questionnaire: development and psychometric testing. *BJU* 1996;77:805-12.

- [25] Donovan JL, Abrams P, et al. The ICS-‘BPH’ Study: the psychometric validity and reliability of the ICSmale questionnaire. *BJU Int* 1996;77:554–62.
- [26] Hullfish KL, Bovbjerg VE, Steers WD. Patient-centered goals for pelvic floor dysfunction surgery: long-term follow-up. *Am J Obstet Gynecol* 2004;191:201–5.
- [27] Rockwood K, Stolee P, Fox RA. Use of goal attainment scaling in measuring clinically important change in the frail elderly. *J Clin Epidemiol* 1993;46:1113–8.
- [28] Rockwood K, Howlett S, Stadnyk K, Carver D, Powell C, Stolee P. Responsiveness of goal attainment scaling in a randomized controlled trial of comprehensive geriatric assessment. *J Clin Epidemiol* 2003;56:736–43.
- [29] Yip AM, Gorman MC, Stadnyk K, et al. A standardized menu for goal attainment scaling in the care of frail elders. *Gerontologist* 1998;38:735–42.
- [30] Yalcin I, Bump RC. Validation of two global impression questionnaires for incontinence. *Am J Obstet Gynecol* 2003;189:98–101.
- [31] Yalcin I, Viktrup L. Comparison of physician and patient assessments of incontinence severity and improvement. *Int Urogynecol J* 2007;18:1291–5.
- [32] Tannenbaum C, Brouillette J, Corcos J. Rating improvements in urinary incontinence: do patients and their physicians agree? *Age Ageing*. In press.
- [33] Jorgensen L, Lose G, Andersen JT. One-hour pad-weighting test for objective assessment of female urinary incontinence. *Obstet Gynecol* 1987;69:39–42.
- [34] Fantl JA, Harkins SW, Wyman JF, Choi CC, Taylor JR. Fluid loss quantitation in women with urinary incontinence: a test-retest analysis. *Obstet Gynecol* 1987;70:739–43.
- [35] Fonda D, Resnick NM, Colling J, et al. Outcome measures for research of lower urinary tract dysfunction in frail older people. *Neurourol Urodynam* 1998;17:273–81.
- [36] Goode PS, Burgio KL, Locher JL, Umlauf MG, Lloyd LK, Roth DL. Urodynamic changes associated with behavioural and drug treatment of urge incontinence in older women. *J Am Geriatr Soc* 2002;50:808–16.
- [37] Elser DM, Wyman JF, McClish DK, Robinson D, Fantl JA, Bump BC. The effect of bladder training, pelvic floor muscle training, or combination training on urodynamic parameters in women with urinary incontinence. Continence Program for Women Research Group. *Neurourol Urodyn* 1999;18:427–36.
- [38] Draft guidance for industry on patient-reported outcome measures: use in medicinal product development to support labeling claims. *Fed Regist* 2006;71:5862–63; Committee for Medicinal Products for Human Use. Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products. July 25, 2005: <http://www.emea.europa.eu/pdfs/human/ewp/1393104en.pdf>. Accessed January 11, 2008.
- [39] Patrick DL, Martin ML, Bushnell DM, Yalcin I, Wagner TH, Buesching DP. Quality of life of women with urinary incontinence: further development of the Incontinence Quality of Life instrument (I-QOL). *Urology* 1999;53:71–6.
- [40] Shumaker SA, Wyman JF, Uebersax JS, McClish D, Fantl JA, for the Continence Program in Women (CPW) Research Group. Health-related quality of life measures for women with urinary incontinence: the Incontinence Impact Questionnaire and the Urogenital Distress Inventory. *Qual Life Res* 1994;3:291–306.
- [41] Kelleher CJ, Cardozo LD, Khullar V, Salvatore S. A new questionnaire to assess the quality of life of urinary incontinent women. *BJOG* 1997;104:1374–9.
- [42] Avery K, Donovan J, Peters TJ, Shaw C, Gotoh M, Abrams P. A brief and robust measure for evaluating the symptoms and impact of urinary incontinence. *Neurourol Urodynam* 2004;23:322–30.
- [43] Brookes ST, Donovan JL, Wright M, Jackson S, Abrams P. A scored form of the Bristol Female Lower Urinary Tract Symptoms. Incontinence Quality of surgery for women with stress incontinence. *Am J Obstet Gynecol* 2004;191:73–82.
- [44] Donovan JL, Peters TJ, Abrams P, Brookes ST, de la Rosette JJ, Shaefer W. Scoring the short form ICSmale SF questionnaire. International Continence Society. *J Urol* 2000;164:1948–55.
- [45] Lubeck D, Prebil L, Peebles P, Brown JS. A health related quality of life measure for use in patients with urge urinary incontinence: a validation study. *Qual Life Res* 1999;8:337–44.
- [46] Coyne K, Revicki D, Hunt T, et al. Psychometric validation of an overactive bladder symptom and health-related quality of life questionnaire: the OAB-q. *Qual Life Res* 2002;11:563–74.
- [47] Kulseng-Hanssen S, Borstad E. The development of a questionnaire to measure the severity of symptoms and the quality of life before and after surgery for stress incontinence. *BJOG* 2003;110:983–8.
- [48] Robinson D, Pearce KF, Preisser JS, Dugan E, Suggs PK, Cohen SJ. Relationship between patient reports of urinary incontinence symptoms and quality of life measures. *Obstet Gynecol* 1998;91:224–8.
- [49] DuBeau CE, Levy B, Mangione CM, Resnick NM. The impact of urge urinary incontinence on quality of life: importance of patients’ perspective and explanatory style. *J Am Geriatr Soc* 1998;46:683–92.
- [50] Tannenbaum C, Brouillette J, Korner-Bitensky N, et al. Creation and testing of the geriatric self-efficacy index for urinary incontinence. *J Am Geriatr Soc*. In press.