Published in final edited form as:

Int Urogynecol J. 2013 January ; 24(1): 15–25. doi:10.1007/s00192-012-1908-7.

Measuring Outcomes in Urogynecologic Surgery: “Perspective is Everything” – A Review

Candace Y. Parker-Autry, MD¹, Matthew D. Barber, MD, MSH², Kimberly Kenton, MD, MS³, and Holly E. Richter, PhD, MD¹

¹Division of Urogynecology and Pelvic Reconstructive Surgery, Department of Obstetrics and Gynecology, University of Alabama at Birmingham, Birmingham, AL

²Center for Urogynecology and Pelvic Reconstructive Surgery, Obstetrics and Gynecology and Women's Health Institute, Cleveland Clinic, Cleveland, OH

³Division of Female Pelvic Medicine & Reconstructive Surgery, Department of Obstetrics and Gynecology & Urology, Loyola University, Maywood, IL

Abstract

Over 10 years have passed since the first US National Institutes of Health consensus panel considered standardization of definitions of pelvic floor conditions and criteria utilized for reporting pelvic floor research study outcomes. The literature is replete with pelvic floor outcome studies; however, a consistent standardized approach to the evaluation of patients and characterization of outcomes is still needed. The purpose of this overview is to describe how the use of outcome measures have evolved over time and attempt to help readers utilize the best measures for their clinical and research needs.

Keywords

pelvic floor disorder research; urogynecology surgical outcomes

INTRODUCTION

Pelvic floor disorders (PFD) collectively refer to the conditions of urinary and fecal incontinence, pelvic organ prolapse, and other lower urinary and bowel tract dysfunction. These multidimensional conditions are prevalent and stigmatizing with wide-reaching consequences on quality of life, intimate and social relationships, and socioeconomic of affected women. To evaluate the effectiveness of non-surgical and surgical interventions for PFDs, we should consider measuring symptom severity from the patient's perspective, measuring symptom severity independent of the patient's viewpoint, and evaluating the multifaceted impact that the PFD may have on a women's life. When symptom severity is measured from the patient's perspective, her perception of the condition is assessed; this is

Corresponding Author: Holly E. Richter, PhD, MD, University of Alabama at Birmingham, 1700 6th Ave South, Suite 10382, Birmingham, AL 35233, hrichter@uabmc.edu, Phone: 205-934-1704, Fax: 205-975-8893.

Drs Parker-Autry, Barber, Kenton and Richter have no Conflicts of Interest.

Author contributions:

Candace Parker-Autry, MD – manuscript writing and editing

Matthew D. Barber, MD, MSH – manuscript writing and editing

Kimberly Kenton, MD, MS – manuscript writing and editing

Holly E. Richter, PhD, MD – manuscript writing and editing

Presented at the Society of Gynecologic Surgeons Annual Clinical Meeting, Baltimore, Maryland, April 15, 2012

known as a subjective outcome measure. Measures of symptom severity using tools that do not account for the individual's perception of their condition are known as objective outcome measures. [1] More recently, experts proposed that multifunctional outcome measures be considered together to create a composite assessment of urogynecologic surgical interventions.

To date, the measures utilized to evaluate and measure PFDs are well summarized in the literature. In this paper, we target clinicians and researchers to provide insight into how to apply these tools to assess the effect of a surgical intervention in practice and in research design. We hope to provide information with which clinician/researchers are able to choose outcome measures that best reflect their clinical or project goals.

The evolution of outcome measures in urogynecologic surgical literature

Surgical treatment of prolapse and urinary incontinence (UI) began to develop in the 18th century concomitantly with the formation of surgical societies and journals.[2, 3] During the 19th century, gynecologic surgeons began to understand the importance of formally reporting surgical outcomes to improve patient care. Early reports assessing urogynecologic surgical outcomes focused on subjective (patient or surgeon) impression of improvement for cure. For example, when Howard Kelly published his case series of 20 women who had undergone a "Kelly plication" procedure for stress urinary incontinence (SUI), he defined success if patient reported 'complete control' of her UI. This innovative, simple surgical technique was widely accepted and performed for over 60 years because it was shown to improve UI over the short and long term (up to 12 years after surgery). [4] While many other techniques existed simultaneously, Kelly's innovation in development and charisma in reporting his technique and its outcomes resulted in its increased popularity.

Concurrently, GR White pioneered the paravaginal repair to correct cystocele after he studied and performed cadaveric dissections to disclose the etiology of the cystocele. He also performed a case series of 19 surgeries, where he defined success as anatomic recurrence on examination up to 3 years after surgery. [5, 6] In his series, there were no recurrences and thus this procedure was deemed superior to the traditional anterior repair of that time. However, in a 1912 report, he argued that surgical failures of anterior vaginal wall prolapse were occurring in high number because other gynecologic surgeons' 'misunderstood the etiology of prolapse'. [2, 6] This report was not widely accepted by his peers, and thus his paravaginal repair procedure was ignored until Richardson popularized White's concepts in the 1970s. [7]

The examples of the Kelly plication and White's paravaginal repair demonstrate how surgical interventions were commonly introduced in the 19th century using surgeon based measures without rigorous testing. Also, the contrast between the popularity of the two techniques demonstrate how popular-opinion of the surgeon was prioritized over the reported success of the procedure as Kelly's technique was utilized by many, while White's technique was ignored despite evidence for superior anatomic outcome. Today in the 21st century, we continue with some of the same concerns.

Assessment methods of urogynecologic surgical interventions continued to evolve after the time of Kelly and White. By the 1970s most surgical outcome studies focused on evaluating the patients' report of symptoms after an intervention. A decade later, subjective measures were negatively viewed as being unsatisfactory for scientific assessment and were replaced with more objective measures, such as urodynamics to define cure of SUI. [8] However, patient-oriented subjective measures and patient-centered goals for treatment returned to the forefront of outcome assessment in the 1990s with the introduction of many diverse and widely available outcome measures. Yet, these measures were seldom used in the clinical

setting and were not applied in a standardized fashion in clinical research. The Agency for Health Care Policy and Research highlighted this lack of standardization in its 1996 update of the Urinary Incontinence Clinical Practice Guidelines. They concluded that urogynecologic surgical literature was deficient in creating “appropriate descriptions of the patient population, in describing the type of incontinence, the methods for accurate diagnosis, the techniques of the surgical procedure, or the outcome in different domains”. [9] A subsequent systematic review on UI treatment revealed that the majority of studies had methodological flaws making “recommendations as to the best clinical practice” impossible due to the lack of good evidence. [10]

In 2001, the National Institutes of Health (NIH) recognized that the majority of urogynecologic literature suffered from a lack of standardized terminology and convened a group of researchers in female pelvic floor disorders to standardize definitions of pelvic floor conditions and to define cure and failure for each condition after an intervention. [11] The application of a standard definition for diagnosis of pelvic floor conditions was an important first step toward improving the quality of urogynecologic surgical literature. Many studies demonstrate that differences in the reported prevalence for any particular condition changes based on how the condition is defined. As we strive to improve the care for women with PFDs, standard definitions in diagnosis of pelvic floor conditions should be utilized at all times (in clinical practice and research).

Importance of outcome measures in clinical practice and research

An outcome measure is an instrument or procedure used to characterize the presence and/or severity of a condition as well as the efficacy, safety, side-effects, and impact of an intervention on the individual with the condition. In the past, subjective physician- or objective test-based measures were used as the primary outcome variable in research studies of pelvic floor disorders (ex. urodynamics for lower urinary tract dysfunction and physical examination parameters for studies of pelvic organ prolapse). Recent studies have increasingly emphasized patient-based outcomes such as symptom questionnaires and quality of life assessments. However, subjective and objective outcomes alone are unable to accurately characterize PFDs or the effect of an intervention on the individual or the condition. Many organizations including the International Continence Society (ICS), NIH, and the World Health Organization’s International Consultation on Incontinence (ICI) developed recommendations to standardize the use of outcome measures in studies of pelvic floor disorders. In general, all agree on several basic principles:

1. outcome assessments should be made using the same measures before and after the intervention
2. both subjective and objective measures should be included, incorporating improvements and deterioration in function as well as complications of the intervention
3. pelvic floor disorders should be assessed from multiple domains to include: [1]
 - Patient’s description and quantification of her symptoms
 - Clinician’s anatomic and functional observations
 - Measurement of the impact on quality of life
 - Socio-economic impact

Using multiple measures collectively refers to a “composite outcome measure” which can include a combination of objective, subjective, patient assessment of improvement, and goal

oriented instruments to provide a more accurate assessment of the true effect of an intervention.

Choosing an outcome measure for clinical practice or study

When choosing an outcome measure to use in clinical practice or when planning a study, a good first step is to identify the primary outcome to be assessed. Identification of valid and reliable tools to measure the desired outcome from all important domains should be the next step. Reliability refers to the repeatability of an instrument. A reliable instrument is able to identify the same diagnosis or severity of a condition with each repeat administration. Once an instrument has been tested and is reliable, the validity is determined. Validity is the degree to which the instrument is able to detect or measure what it intends and is determined by the way the instrument is applied. The process of validation is beyond the scope of this article; however it is important to understand that validity is tested in a target population (ex. English speaking women with a pelvic floor condition) and thus the instrument should only be used in populations similar to that in which it was tested. Administration of the outcome measure should be performed according to how it was validated and should be identically administered before and after the intervention. The remainder of this article will review many types of outcome measures available to clinicians and researchers to assess the effectiveness of treatments for PFDs and will provide insight into how these tools can be applied to improve patient care.

Objective Outcome Instruments

Objective outcomes refer to measures designed to assess the presence and severity of a pelvic floor condition independent from the patient or physician perception of the condition. The ‘objectivity’ of some measures has been questioned because they were designed by physicians to elicit a desired response and therefore may have a subjective component.

Urinary Incontinence—Urinary incontinence (UI) is clinically diagnosed based on patient’s history and depending on the type of UI, treatment may often be started based on symptoms alone. Demonstration of the sign of SUI (transurethral urine loss with increased in abdominal pressure) is necessary prior to surgical therapies; however, in general, the value of other objective measures for assessment of UI in clinical practice is unknown. In the realm of research studies, objective assessment of UI has a role in 1) standardizing the diagnosis of UI as a part of inclusion criteria, and 2) showing a quantifiable difference after an intervention.

Cough stress test: In 1923, Bonney described one of the earliest objective measures of SUI [12] in which one observes the urethra for urine loss during a cough or Valsalva maneuver while the patient has a full bladder and is either in the lithotomy and/or standing positions. The cough stress test is a reliable and accurate method of assessing SUI in women. [13] It is a simple test easily incorporated into the office visit and is the test of choice for diagnosis of SUI in women with uncomplicated SUI symptoms in clinical research studies. [14, 15]

Q-tip test: Stress urinary incontinence may occur in conjunction with urethro-vesical junction (bladder neck) mobility. Thus, assessment of urethral mobility should be considered in the examination of women with UI as some treatments efficacies vary with urethral hypermobility. A cotton-swab Q-tip test is used to diagnosis urethral hypermobility in women. This classic test has been a component of the pelvic examination in women with pelvic floor disorders since its publication in 1971. [16] Using a cut-off of $>130^\circ$ with strain, the test has good test-retest reliability [17], but has lost favor because of patient discomfort and may be replaced by other techniques such as the visual urethral mobility examination. [18]

Pad test: The ‘pad test’ is an objective test that aims to quantify the volume of urine loss by weighing a perineal pad before and after a specified time and/or group of activities. It is currently the only incontinence severity measure that captures the actual volume of leakage. Pad testing has also been used to attempt to distinguish continent from incontinent women. Numerous pad test protocols have been described but in general they can be divided into short- and longer-term tests. The short-term tests may take place in the office and last from 10 minutes to 2 hours, while the long-term tests take place in the home and may last from 24-28 hours. In comparison to long-term tests, short-term pad tests are easy and quick, with excellent direct-observation of patient compliance making them the preferred test in clinical research trials. However, the short-term test lacks authenticity since the subject’s natural environment and daily provocation activities cannot be recreated in the office, therefore the test-retest reliability and validity is low. [19]

Longer-term pad tests reflect everyday life and activities of the patient, which increases its reproducibility and validity. Normative data is also better defined with the longer-term tests. A change in pad weights of up to 4 grams/24 hours can be seen in continent women, so values less than this should be considered insignificant.[20] The disadvantages of longer-term pad tests are decreased compliance and theoretical evaporatory loss, which can be minimized with proper packaging of pads.

Pelvic organ prolapse—Outcome measures that assess changes in anatomy obviously have an essential role in studies of pelvic floor disorders, particularly pelvic organ prolapse. Anatomic outcomes can be assessed by physical examination or radiographic studies. Imaging techniques for PFD include evacuation proctography or defecography, static and dynamic MRI, and ultrasound. All techniques have been used to assess anatomic outcomes in studies of women with pelvic floor disorders. They may provide a better picture of the location, support, and integrity of the pelvic visceral structures than does a simple physical examination. However, the relative cost and lack of standardization of these techniques currently limit their use in many research studies. [21]

Pelvic Organ Prolapse Quantification System: The historical Baden-Walker half-way system was replaced by the Pelvic Organ Prolapse Quantification (POPQ) system in 1996. This was jointly supported by the Society of Gynecologic Surgeons, the American Urogynecologic Society, and the International Continence Society as the accepted method for describing pelvic support and comparing exams over time and after interventions. [22] This ordinal staging system categorizes POP as stage 0, 1, 2, 3, or 4 and quantifies the anatomic descent of the vaginal wall relative to the hymenal ring. The good inter- and intra-examiner reproducibility made this instrument the most commonly used system in peer-reviewed literature. [22] The precision of this tool has allowed improved understanding of the relationship between the anatomic characteristics of pelvic organ prolapse and the development of specific pelvic floor symptoms. The exclusion of features such as vaginal caliber, status of paravaginal support, pelvic floor descent, and urethral mobility are some limitations. Further criticism comes because the ordinal organization limits differentiation of clinically important subgroups. For example, stage 2 prolapse includes a diverse group of women: those with and without prolapse symptoms, and those with and without prolapse beyond the hymen.

Many critiques of the POPQ system have proposed other indices to measure prolapse severity for research purposes. The Pelvic Floor Disorders Network (NICHD) proposed a system of support loss equations to cumulatively account for the diversity in the ordinal staging system and to create a continuous variable that may be more representative of individual patient features. They re-analyzed POPQ staging data from three large, multicenter randomized comparative effectiveness trials of women with PFD, including the

CARE [23], OPUS [24], and ATLAS [25] trials. The authors created 3 new variables to represent loss of vaginal support on a continuum, including adding measures for all the vaginal wall segments. The new scores representing vaginal support loss on a continuum were compared to validated questionnaires and POPQ stage to determine their relationship to symptom and anatomic severity, respectively. The responsiveness of this score to change was also measured. Each support loss score strongly correlated with the POPQ stage at baseline; however, the most distal or prolapsed POPQ point (SLmax) demonstrated the greatest responsiveness to change with anatomic support and thus may be preferable to POPQ stage to compare group data. When considering prolapse symptoms (determined using PFDI questions 4 and 5), each support loss measure was moderately correlated with severity. Similar to POPQ stage, improvements in anatomic support only weakly correlated with prolapse symptoms and had no correlation with the impact of prolapse on quality of life. [26]

Physiologic Tests—Physiologic testing is frequently used in the clinical evaluation of PFD. These tests attempt to describe or quantify the underlying function of the pelvic viscera and pelvic floor, often including an assessment of whether such function is normal or pathologic. Physiologic testing serves two principal roles in pelvic floor research: to characterize subjects' symptoms at study entry and to help define or understand how symptoms may change after treatment. The most common physiologic test of the lower urinary tract is urodynamics. Physiologic tests of the lower gastrointestinal tract include anal manometer and colonic transit studies. Other physiologic tests of the pelvic floor include such things as vaginal pressure transducers to measure levator ani contraction strength and electromyography (EMG) of the pelvic floor muscles, urinary and anal sphincter muscles to evaluate neuromuscular function.

In addition to defining a subject's symptoms at baseline and evaluating treatment outcomes, the use of physiologic testing in pelvic floor research can provide valuable insight into the underlying pathophysiology of the condition and into the mechanisms of treatment success or failure. Additionally, physiologic testing allows correlation between changes in symptoms and changes in physiology. The primary disadvantages of using physiologic tests in research are that they are often costly, time-consuming, and many can be uncomfortable for the patient which directly impacts recruitment. Similarly, reproducibility and validity of most of these tests is not established limiting meaningful interpretation of the results.

Subjective Outcome Measures

Clinically, we determine the presence and severity of patient's symptoms from their history. The information obtained relies upon the clinician's knowledge of important clinical features and interviewing skills, the clinician's documentation practices, and the patient's ability to accurately recall and report their recent health behavior. Unfortunately, history alone is often unreliable and may result in a misunderstanding of the subject's condition clinically as well as introducing recall bias into research studies. Many validated subjective outcome measures are currently available; attention needs to be paid to choosing one that meets the intended goal.

Symptom Diaries—When a patient records the details of their lower urinary and bowel tract symptoms in real-time, this represents a symptom diary. Symptom diaries have good test-retest reliability and provide concrete, objective data allowing the clinician to see the frequency, severity, and provocation of the targeted symptom. [27] This data can be used in clinical practice to add validity to the history interview as well as educate patient's about behavioral patterns contributing to their symptoms. Symptom diaries are used in research studies to aid in diagnosis, symptom severity, and as a tool standardizing inclusion criteria.

Symptom diaries are responsive to change, and thus can be used to measure the effectiveness of an intervention.

In their simplest forms, patients are asked to prospectively record the time and episode of voiding/defecation with a description of any dysfunction (for example, UI, and incomplete defecation with need for splinting). In more complex forms, patients are asked to also record behavioral factors used to deal with their condition. This would include pad usage in women with urinary/bowel incontinence, type and amount of fluid/food intake, voided volume, activities that occur simultaneously with the lower urinary/bowel symptom. The bedtime and awakening times may be recorded to distinguish diurnal and nocturnal symptoms. Bladder diaries are used to record daily fluid intake and volumes of each void, called frequency-volume charts. These charts provide additional data such as average daily fluid intake, voided volume, largest single void (estimates functional bladder capacity), and temporality of the symptoms in question. The duration of a symptom diary is debatable. The most commonly published data have used a 7-day diary. However, the reliability of patient compliance decreases as the duration of the diary increases. The NIH recommends a 3-day diary, which is equally accurate and reproducible with increased patient convenience in comparison to the 7-day diary. [28]

Questionnaires—While symptom specific questionnaires also rely on patients' ability to accurately recall health behaviors, questionnaires' reliability and validity empower them to be able to detect the severity and impact of the condition. Psychometrically robust self-administered questionnaires are designed to measure the presence, severity and/or impact of pelvic floor symptoms on a patient's activities and well-being. In general, pelvic floor questionnaires can be separated into four categories: 1) Symptom-specific questionnaires; 2) Health-related quality of life (HRQoL) questionnaires; 3) Sexual function questionnaires; and 4) Global Impression of Improvement. [29]

As a reminder, when choosing a questionnaire for use in clinical practice or in a research study the outcome of interest should be known and the questionnaire chosen should measure the desired outcome. A detailed review of the questionnaire's content will provide this information. It is important to keep in mind the purpose for which the questionnaire was originally designed and the population in which it was validated. Before questionnaires are used in unique populations or contexts further validation is necessary. The reliability and validity of the questionnaire should then be researched. [29, 30] Use of non-validated questionnaires may provide non-generalizable data or may fail to detect important clinical changes. It is generally preferable to use a validated questionnaire that is widely accepted and has been used in the targeted population. The feasibility of questionnaire use should be determined prior to its use by examining the length and construct. Long questionnaires may be desirable for research studies where detail is needed, however these will likely be too time-consuming and burdensome for effect use in the clinical practice setting.

Symptom Questionnaires: Symptom-specific questionnaires assess the presence or absence of specific pelvic floor symptoms as well as symptom severity and degree of bother. A number of symptom questionnaires are available for women with UI; more recently symptom questionnaires for fecal incontinence and other pelvic floor disorders have been developed. (Table 1) The visual analog scale (VAS), most commonly used to assess pain, is validated for use in PFD research to measure symptom severity. Responders are asked to indicate the severity of their disease on a 10cm VAS with 0=no complaints, and 10=unbearable symptoms.

Health-Related Quality of Life (HRQoL) Questionnaires: Quality of life is a term that globally refers to a person's physical, social, and emotional function as well as a person's

satisfaction with their life. [31] Measuring quality of life in relation to health refers to asking questions about specific quality of life endpoints and how they relate to a women's general health or a specific health condition. [32] Pelvic floor disorders rarely cause mortality or severe morbidity, however they can diminish a women's quality of life. Pelvic floor disorders are associated with decreased activities of daily living, social isolation, and decreased self-esteem, increased depression with avoidance of out-of-home activity, helplessness, and sexual dysfunction. For this reason, assessment of the impact of pelvic floor dysfunction on quality of life has been more recently emphasized as a component in assessment of an intervention both in research and in clinical practice. Health-related quality of life questionnaires are multi-dimensional tools that assess the physical, emotional, social, spiritual, sexual, and occupational impact of a symptom. A good HRQOL questionnaire makes the assessment from the patient's perspective with good reliability, is responsive to change (after an intervention), and is easy to administer.

Measures of HRQOL can be classified into two types: Generic and Condition-specific. Generic instruments have the advantage of allowing comparisons across different groups or illnesses, but may lack sensitivity to the unique aspects of a specific health condition and how it impacts the life of the affected patient. On the other hand, condition-specific HRQOL instruments are designed to measure the impact of a specific health condition on HRQOL. These instruments provide a more in-depth assessment of specific issues and concerns critical to the disease process for which they were designed. They also tend to be more responsive to change than generic instruments. Their primary disadvantage is that they can only be used in the particular patient group for which they were designed and data cannot be compared to normative data from the general population or other groups. Table 2 summarizes by pelvic floor symptom, the more commonly utilized questionnaires. The validation and reliability can be found in the referenced articles.

Sexual Function: Sexual function is an important outcome to consider when evaluating the treatment of pelvic floor disorders. While many reliable and valid sexual function questionnaires exist, their use in women with pelvic floor disorders is limited. One systematic review identified 14 valid and reliable self-administered sexual function measures for men and women, however only two met the highest standards and thus are recommended for use in the general population: the McCoy Female Sexuality Questionnaire (MFSQ) and the Female Sexual Function Index (FSFI). [33] The major difference between these questionnaires is that the MFSQ includes a social and relationship factors, whereas the FSFI is more focused on individual function. Both measures contain questions that are only applicable for women with a current sexual partner. Sexual function is a part of a women's life affected by pelvic floor disorders, thus condition-specific questionnaires to measure this effect are very useful. The only condition-specific questionnaire as it relates to pelvic floor disorders is the Pelvic Organ Prolapse and Incontinence Sexual Function Questionnaire (PISQ). [34] This valid and reliable 31-item instrument was designed for use in sexually active women with pelvic organ prolapse and/or UI and assesses the impact of these conditions on the woman's sexual function. A short form (PISQ-12) makes administration more feasible. [35] An updated version, the PISQ-IR will be published shortly. This version will differ from the previous by: including a measure of recent activity in women who choose not to be active, validated in women with anal incontinence, slightly longer, and no total score. (R. Rogers, personal communication) The PISQ questionnaire has been translated and validated in Portuguese, Swedish, and Turkish, in addition for internet administration. [36]

Measuring the Global Index of Improvement (GPI): A global index is a simple instrument that asks the patient to rate the severity of a specific condition or to rate the response of her condition to treatment. As the name implies, the goal of a global index is to

get an overall appraisal of a complex phenomenon, not to evaluate individual components of the phenomenon. For example, the Patient Global Impression of Improvement (PGI-I), is a single item question that asks patients to rate their improvement after treatment on a 7-point scale from “very much worse” to “very much better”. Responses demonstrate significant correlation with changes in objective testing (such as pad test in women with UI), UI frequency, and I-QOL scores after an intervention for UI. [37] The main disadvantage of global indices is their lack of information and specificity regarding the precise aspects or manifestations of disease severity or improvement that results in an individual patient selecting a specific rating. However, global indices provide the single best measure of the significance of change from the individual perspective. They also provide information that may affect HRQOL more so than other methods for assessing a clinically meaningful change.

The importance of choosing a responsive questionnaire: Most validated questionnaires are reliable; however the responsiveness (the ability to detect a change) of measures may vary. This becomes more important when choosing an instrument to be used in a research study or clinical practice where the effect of an intervention will be measured. For lower urinary tract symptoms, the Incontinence Impact Questionnaire (IIQ) and its short form (IIQ-7) have very good responsiveness. [38] The King’s Health Questionnaire (KHQ) [38] and the Urge-Incontinence Impact Questionnaire (Urge IIQ) questionnaires also have good responsiveness [39], while the Incontinence Quality of Life (I-QOL) questionnaire has poor/moderate responsiveness. [40] Similarly, the Fecal Incontinence Quality of Life (FIQOL) Scale, Modified Manchester Health Questionnaire (MMHQ), and Prolapse Quality of Life Questionnaire (P-QOL) instruments lack any established data on their responsiveness. The Pelvic Floor Distress Inventory (PFDI) and the Pelvic Floor Impact Questionnaire (PFIQ), both available in more feasible short forms, are valid, reliable, and responsive.[41] They are also the most commonly used instruments because of the diversity of administration routes including self-administration on paper, and on the telephone. Several questionnaires, including the PFDI and PFIQ, have also been validated for internet administration.[36, 42]

International Validation of Questionnaires: Pelvic floor disorders are common conditions that affect women of all nationalities. While their use in different populations is necessary, the cultural and psychometric properties of the original country development and validation may not be transferable. Most often the linguistic and cultural adaptation of a questionnaire is done separately in a subsequent study. The principals of adapting a questionnaire from one culture and/or language to another are outlined in the 4th edition of the International Consultation on Incontinence.[29] The available translations for the most commonly used questionnaires are noted in tables 1 and 2.

The ‘composite outcome’ and defining cure

Treatment for pelvic floor disorders can involve medications, behavioral therapy including pelvic floor muscle exercises and strategies, and surgical/procedural interventions. Regardless of the modality, the definition of ‘cure’ used in research investigating the effectiveness of an intervention is variable. While a consensus definition of cure for each pelvic floor disorder is non-existent, the use of a combination of outcome measures to create a ‘composite outcome’ transcends pelvic floor disorder research and provides a more real-life or ‘broader context’ definition of cure.

Composite Outcome of Stress Urinary Incontinence Surgery—The Trial of Midurethral Sling (TOMUS) a multicenter, randomized equivalence trial comparing the effect of the retropubic (RMUS) and transobturator midurethral slings (TMUS) for SUI treatment provide an example of a composite outcome measure. The composite primary

outcome and definition of success included objective measures (negative 300ml retrograde fill stress test, negative 24 hour pad test, and no retreatment) and subjective measures (no stress symptoms on the Medical Epidemiologic and Social Aspects of Aging (MESA) questionnaire, no leakage on a 3-day bladder diary, and no retreatment) assessed at 12 months. The unadjusted composite objective and subjective cure rates were similar between RMUS and TMUS [(80.8% and 77.7%, respectively) and (62.2% and 55.8%, respectively)]. [43] If any single objective or subjective outcome measure rather than a composite measure was used, UI cure rates varied between 65-95% demonstrating the variability of individual measures.[43] (Figure 1)

Composite Outcome of Pelvic Organ Prolapse Surgery—Without a generally accepted definition for success after prolapse surgery, in 2001 the NIH established standardized definitions for prolapse outcomes based on expert opinion. An ‘optimal’ anatomic outcome was defined as Stage 0 prolapse and ‘satisfactory’ anatomic outcome as Stage 1 or less utilizing the POPQ system. [11] Subsequent data suggest these definitions are too strict as only half asymptomatic women presenting for annual gynecologic examinations have stage 0 or 1 prolapse [44]; in other words, over half of asymptomatic women seeking routine gynecologic care do not meet the NIH definition of optimal or satisfactory support. A randomized controlled trial comparing three approaches to the correction of anterior vaginal wall prolapse defined anatomic cure using NIH recommendations. Anatomic success was 30% for traditional anterior colporrhaphy, 46% for ultralateral anterior colporrhaphy, and 42% with absorbable mesh augmentation anterior colporrhaphy, although the majority of patients were asymptomatic post-operatively. [45] Defining success using only anatomic outcome measures and without consideration of patient’s perception of symptom change after surgery reflected low rates of success, that may not truly reflect the patient’s overall experience with surgery.

More recently, data from a large, multicenter trial of women undergoing open sacrocolpopexy with or without Burch procedure was reanalyzed to determine the optimal definition of prolapse surgical treatment success. Eighteen different definitions of ‘success’ were created using variable subjective and objective outcomes and compared to the patient’s overall impression of improvement. Success ranged from 19 – 97% depending on the definition. [46] Similar to prior data comparing techniques for anterior repair [45] when success was defined by anatomic support proximal to the hymen, success rates were low and varied from 19.2-57.6%. The highest cure rates were found when success was defined by absence of bulge symptoms and absence of re-treatment, 92.1% and 97.2%, respectively. These data suggest that subjective outcomes are superior to anatomic outcomes in accessing patients’ overall impression of improvement.[46] The authors concluded that the definition of success should collectively include the absence of bulge symptoms (subjective outcome), anatomic criteria of no leading edge beyond the hymen (objective outcome), and the absence of re-treatment to create a composite outcome. These recommendations inspired many to consider assessment of these multiple domains to measure success after an intervention for a pelvic floor disorder and to apply a ‘composite outcome’ to define assessment of success.

Chmielewski and colleagues recently reanalyzed the trial performed by Weber to apply this recently defined composite outcome of success for pelvic organ prolapse to the data collected by Weber and colleagues. Their composite outcome was defined by: 1) POP-Q measurements Ba, Bp, and most dependent part of vagina 0cm; 2) Visual analog scale (range 0-100) to question “How much are you bothered by symptoms related to vaginal prolapse” with score 20 mm defining success; and 3) no retreatment. Application of this composite outcome resulted in 88% (66/75) of participants being successfully cured after surgery at 1 year with no differences among the 3 groups compared (women receiving

traditional anterior colporrhaphy, “ultralateral” anterior colporrhaphy, and mesh-augmented colporrhaphy).[47]

CONCLUSION

As many of us are both busy clinicians and researchers caring for women with pelvic floor disorders, it is important to prospectively evaluate patients with use of appropriate instruments to help optimize care. The definition of the desired outcome must be clear. In choosing an outcome measure one should consider the definition of the outcome and the validation of the instrument. Research outcomes should be measured as a composite accounting for anatomic changes, patients’ symptoms, need for re-treatment, and their overall impression of the condition. Outcome measures should be administered both prior to and after the intervention. Perspective regarding cure rates or improvement after treatment is highly reflective of the outcome measures which are utilized therefore they need to be chosen in a rigorous and thoughtful manner.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Partially supported by the National Institute of Diabetes and Digestive and Kidney Diseases, 2K24-DK068389 to HER

Abbreviations

PFDs	Pelvic floor disorders
UI	urinary incontinence
SUI	stress urinary incontinence

References

1. Hilton P, Robinson D. Defining cure. *Neurourol Urodyn*. 2011; 30(5):741–5. [PubMed: 21661023]
2. Cooke JN, Gousse AE. A historical perspective on cystocele repair - from honey to pessaries to anterior colporrhaphy: Lessons from the past. *The Journ of Urol*. 2008; 179:2126–2130.
3. Deger RB, Menzin AW, Mikuta JJ. The vaginal pessary: past and present. *Postgrad Obstet Gynecol*. 1993; 13(1)
4. Kelly HA, Dumm WM. Urinary incontinence in women, without manifest injury to the bladder. 1914. *Int Urogynecol J Pelvic Floor Dysfunct*. 1998; 9(3):158–64. [PubMed: 9745976]
5. White GR. A radical cure by suturing lateral sulci of vagina to white line of pelvic fascia. *JAMA*. 1909; 53:1707.
6. White GR. An anatomical operation for the cure of cystocele. *Am J Obstet Dis Women Child*. 1912; 65:286.
7. Richardson AC, Lyon JB, Williams NL. A new look at pelvic relaxation. *Am J Obstet Gynecol*. 1976; 126(5):568–73. [PubMed: 984127]
8. Jarvis GJ. Surgery for genuine stress incontinence. *Br J Obstet Gynaecol*. 1994; 101(5):371–4. [PubMed: 8018606]
9. Fantl, JA.; Colling, ND., et al. Urinary incontinence in adults: acute and chronic management. Mar. 1996
10. Black NA, Downs SH. The effectiveness of surgery for stress incontinence in women: a systematic review. *Br J Urol*. 1996; 78(4):497–510. [PubMed: 8944504]

11. Weber AM, Abrams P, Brubaker L, Cundiff G, Davis G, Dmochowski RR, et al. The standardization of terminology for researchers in female pelvic floor disorders. *Int Urogynecol J Pelvic Floor Dysfunct.* 2001; 12(3):178–86. [PubMed: 11451006]
12. Bonney V. On diurnal incontinence of urine in women. *J Obstet Gynaecol Brit Emp.* 1923; 30:358–365.
13. Swift SE, Yoon EA. Test-retest reliability of the cough stress test in the evaluation of urinary incontinence. *Obstet Gynecol.* 1999; 94(1):99–102. [PubMed: 10389726]
14. Scotti RJ, Myers DL. A comparison of the cough stress test and single-channel cystometry with multichannel urodynamic evaluation in genuine stress incontinence. *Obstet Gynecol.* 1993; 81(3): 430–3. [PubMed: 8437800]
15. Nager CW, Brubaker L, Litman HJ, Aycsynski HM, Varner RE, Amundsen C, et al. A randomized trial of urodynamic testing before stress-incontinence surgery. *N Engl J Med.* 2012; 366(21):1987–97. [PubMed: 22551104]
16. Crystle CD, Charme LS, Copeland WE. Q-tip test in stress urinary incontinence. *Obstet Gynecol.* 1971; 38(2):313–5. [PubMed: 5105346]
17. Swift SE, Barnes D, Herron A, Goodnight W. Test-retest reliability of the cotton swab (Q-tip) test in the evaluation of the incontinent female. *Int Urogynecol J.* 2010; 21(8):963–7. [PubMed: 20379698]
18. Robinson BL, Geller EJ, Parnel BA, Crane AK, Jannelli ML, Wells EC, et al. Diagnostic accuracy of visual urethral mobility exam versus Q-Tip test: a randomized crossover trial. *Am J Obstet Gynecol.* 2012 DOI 10.1016/j.ajog.2012.02.015.
19. Ryhammer AM, Djurhuus JC, Laurberg S. Pad testing in incontinent women: a review. *Int Urogynecol J Pelvic Floor Dysfunct.* 1999; 10(2):111–5. [PubMed: 10384973]
20. Abrams PH, Skidmore R, Poole AC, Follet D. The standardisation of terminology of lower urinary tract function. The International Continence Society Committee on Standardisation of Terminology. *Scand J Urol Nephrol Suppl.* 1998; 114:5–19. [PubMed: 3201169]
21. Derpapas A, Digesu GA, Fernando R, Khullar V. Imaging in urogynaecology. *Int Urogynecol J.* 2011; 22(11):1345–56. [PubMed: 21626038]
22. Hall AF, Theofrastous JP, Cundiff GW, Harris RL, Hamilton LF, Swift SE, et al. Interobserver and intraobserver reliability of the proposed International Continence Society, Society of Gynecologic Surgeons, and American Urogynecologic Society pelvic organ prolapse classification system. *Am J Obstet Gynecol.* 1996; 175(6):1467–70. discussion 1470-1. [PubMed: 8987926]
23. Visco AG, Brubaker L, Nygaard I, Richter HE, Cundiff G, Fine P, et al. The role of preoperative urodynamic testing in stress-continent women undergoing sacrocolpopexy: the Colpopexy and Urinary Reduction Efforts (CARE) randomized surgical trial. *Int Urogynecol J Pelvic Floor Dysfunct.* 2008; 19(5):607–14. [PubMed: 18185903]
24. Wei J, Nygaard I, Richter H, Brown M, Barber M, Xiao X, et al. Outcomes following vaginal prolapse repair and mid urethral sling (OPUS) trial--design and methods. *Clin Trials.* 2009; 6(2): 162–71. [PubMed: 19342469]
25. Richter HE, Burgio KL, Goode PS, Borello-France D, Bradley CS, Brubaker L, et al. Non-surgical management of stress urinary incontinence: ambulatory treatments for leakage associated with stress (ATLAS) trial. *Clin Trials.* 2007; 4(1):92–101. [PubMed: 17327249]
26. Brubaker L, Barber MD, Nygaard I, Nager CW, Varner E, Schaffer J, et al. Quantification of vaginal support: are continuous summary scores better than POPQ stage? *Am J Obstet Gynecol.* 2010; 203(5):512e1–6. [PubMed: 20728072]
27. Abrams PH, Blaivas JG, Stanton SL, Anderson JT. The concept and measurement of bladder work. A review after 6 years. *Br J Urol.* 1983; 55(2):199. [PubMed: 6839095]
28. Dmochowski RR, Sanders SW, Appell RA, Nitti VW, Davila GW. Bladder-health diaries: an assessment of 3-day vs 7-day entries. *BJU Int.* 2005; 96(7):1049–54. [PubMed: 16225527]
29. Staskin, D.; Kelleher, C.; Avery, K.; Bosch, R.; Cotterill, N.; Coyne, K., et al. Committee 5 B: Patient-Reported Outcome Assessment. In: Abrams, P.; Cardozo, L.; Khoury, S.; Wein, A., editors. *Incontinence.* 4th Ed. Health Publication Ltd; Paris, France: 2009. p. 363-412.
30. Uebersax JS, Wyman JF, Shumaker SA, McClish DK, Fantle JA, Continence Program for Women Research Group. 1995) 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*. 14(2):131–9. [PubMed: 7780440]
31. The World Health Organization Quality of Life Assessment (WHOQOL). development and general psychometric properties. *Soc Sci Med*. 1998; 46(12):1569–85. [PubMed: 9672396]
 32. Centers for Disease Control. Measuring healthy days: Population assessment of health-related quality of life. 2000.
 33. Daker-White G. Reliable and valid self-report outcome measures in sexual (dys)function: a systematic review. *Arch Sex Behav*. 2002; 31(2):197–209. [PubMed: 11974645]
 34. Rogers RG, Kammerer-Doak D, Villarreal A, Coates KW, Qualls C. A new instrument to measure sexual function in women with urinary incontinence or pelvic organ prolapse. *Am J Obstet Gynecol*. 2001; 184(4):552–8. [PubMed: 11262452]
 35. Rogers RG, Coates KW, Kammerer-Doak D, Khalsa S, Qualls C. A short form of the Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire (PISQ-12). *Int Urogynecol J Pelvic Floor Dysfunct*. 2003; 14(3):164–8. discussion 168. [PubMed: 12955337]
 36. Parnell BA, Dunivan GC, Connolly A, Jannelli ML, Wells C, Geller EJ. Validation of web-based administration of the Pelvic Organ Prolapse/Urinary Incontinence Sexual Function Questionnaire (PISQ-12). *Int Urogynecol J*. 2011; 22(3):357–61. [PubMed: 20972537]
 37. Yalcin I, Bump RC. Validation of two global impression questionnaires for incontinence. *Am J Obstet Gynecol*. 2003; 189(1):98–101. [PubMed: 12861145]
 38. Geller EJ, Barbee ER, Wu JM, Loomis MJ, Visco AG. Validation of telephone administration of 2 condition-specific quality-of-life questionnaires. *Am J Obstet Gynecol*. 2007; 197(6):632e1–4. [PubMed: 18060958]
 39. Brown JS, Posner SF, Stewart AL. Urge incontinence: new health-related quality of life measures. *J Am Geriatr Soc*. 1999; 47(8):980–8. [PubMed: 10443860]
 40. 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(1):71–6. [PubMed: 9886591]
 41. Barber MD, Walters MD, Bump RC. Short forms of two condition-specific quality-of-life questionnaires for women with pelvic floor disorders (PFDI-20 and PFIQ-7). *Am J Obstet Gynecol*. 2005; 193(1):103–13.
 42. Handa VL, Barber MD, Young SB, Aronson MP, Morse A, Cundiff GW. Paper versus web-based administration of the Pelvic Floor Distress Inventory 20 and Pelvic Floor Impact Questionnaire 7. *Int Urogynecol J Pelvic Floor Dysfunct*. 2008; 19(10):1331–5. [PubMed: 18488134]
 43. Richter HE, Albo ME, Zyczynski HM, Kenton K, Norton PA, Sirls L, et al. Retropubic versus transobturator midurethral slings for stress incontinence. *N Engl J Med*. 2010; 362(22):2066–76. [PubMed: 20479459]
 44. Swift SE, Woodman P, O'Boyle A, Khan M, Valley M, Bland D, et al. Pelvic Organ Support Study (POSST): the distribution, clinical definition, and epidemiologic condition of pelvic organ support defects. *Am J Obstet Gynecol*. 2005; 192(3):795–806. [PubMed: 15746674]
 45. Weber AM, Walters MD, Piedmonte MR, Ballard LA. Anterior colporrhaphy: a randomized trial of three surgical techniques. *Am J Obstet Gynecol*. 2001; 185(6):1299–304. discussion 1304–6. [PubMed: 11744900]
 46. Barber MD, Brubaker L, Nygaard I, Wheeler TL, Schaffer J, Chen Z, et al. Defining success after surgery for pelvic organ prolapse. *Obstet Gynecol*. 2009; 114(3):600–9. [PubMed: 19701041]
 47. Chmielewski L, Walters MD, Weber AM, Barber MD. Reanalysis of a randomized trial of 3 techniques of anterior colporrhaphy using clinically relevant definitions of success. *Am J Obstet Gynecol*. 2011; 205(1):69–e1. [PubMed: 21545996]
 48. 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*. 1993; 47(6):497–9. [PubMed: 8120507]
 49. Avery K, Donovan J, Peters TJ, Shaw C, Gotoh M, Abrams P. ICIQ: a brief and robust measure for evaluating the symptoms and impact of urinary incontinence. *Neurourol Urodyn*. 2004; 23(4):322–30. [PubMed: 15227649]

50. Brookes ST, Donovan JL, Wright M, Jackson S, Abrams P. A scored form of the Bristol Female Lower Urinary Tract Symptoms questionnaire: data from a randomized controlled trial of surgery for women with stress incontinence. *Am J Obstet Gynecol*. 2004; 191(1):73–82. [PubMed: 15295345]
51. Bradley CS, Rahn DD, Nygaard IE, Barber MD, Nager CW, Kenton K, et al. The questionnaire for urinary incontinence diagnosis (QUID): validity and responsiveness to change in women undergoing non-surgical therapies for treatment of stress predominant urinary incontinence. *Neurourol Urodyn*. 2010; 29(5):727–34. [PubMed: 19787711]
52. Rockwood TH, Church JM, Fleshman JW, Kane RL, Mavrantonis C, Thorson AG, et al. Patient and surgeon ranking of the severity of symptoms associated with fecal incontinence: the fecal incontinence severity index. *Dis Colon Rectum*. 1999; 42(12):1525–32. [PubMed: 10613469]
53. Jorge JM, Wexner SD. Etiology and management of fecal incontinence. *Dis Colon Rectum*. 1993; 36(1):77–97. [PubMed: 8416784]
54. Price N, Jackson SR, Avery K, Brookers ST, Abrams P. Development and psychometric evaluation of the ICIQ Vaginal Symptoms Questionnaire: the ICIQ-VS. *BJOG*. 2006; 113(6):700–12. [PubMed: 16709214]
55. Kelleher CJ, Cardozo LD, Khullar V, Salvatore S. A new questionnaire to assess the quality of life of urinary incontinent women. *Br J Obstet Gynaecol*. 1997; 104(12):1374–9. [PubMed: 9422015]
56. Rockwood TH, Church JM, Fleshman JW, Kane RL, Mavrantonis C, Thorson AF, et al. Fecal Incontinence Quality of Life Scale: quality of life instrument for patients with fecal incontinence. *Dis Colon Rectum*. 2000; 43(1):9–16. discussion 16–7. [PubMed: 10813117]
57. Kwon S, Visco AG, Fitzgerald MP, Ye W, Whitehead WE, Pelvic Floor Disorders Network. Validity and reliability of the Modified Manchester Health Questionnaire in assessing patients with fecal incontinence. *Dis Colon Rectum*. 2005; 48(2):323–31. discussion 331–4. [PubMed: 15616750]
58. Digesu GA, Khullar V, Cardozo L, Robinson D, Salvatore S. P-QOL: a validated questionnaire to assess the symptoms and quality of life of women with urogenital prolapse. *Int Urogynecol J Pelvic Floor Dysfunct*. 2005; 16(3):176–81. discussion 181. [PubMed: 15875234]

Figure 1. Proportion of Patients with Treatment Success at 12 Months, According to Objective and Subjective Criteria

* Medical, Epidemiological and Social Aspects of Aging questionnaire

Adapted from ref 43

Table 1
Validated Symptom Specific Questionnaires

Targeted symptom	Questionnaires	Available Translations
Lower urinary tract symptoms	Urinary Distress Inventory -6 [30]	6
	Incontinence Severity Index [48]	2
	International Consultation on Incontinence Modular Questionnaire (ICIQ-UI SF) [49]	38
	Bristol Female Lower Urinary Tract Symptom Questionnaire (BFLUTS) [50]	10
	Questionnaire for Urinary Incontinence Diagnosis (QUID) [51]	English only
Lower gastrointestinal tract dysfunction	Fecal Incontinence Severity Index [52]	English only
	Cleveland Clinic Florida Fecal Incontinence Score/Wexner Index [53]	English only
Pelvic Organ Prolapse symptoms	International Consultation on Incontinence Modular Questionnaire – Vaginal symptoms subscale [54]	2
	Pelvic Organ Prolapse Distress Inventory – 6 [41]	6
Combination of all pelvic floor symptoms	Pelvic Floor Distress Inventory (PFDI) short-form (PFDI-20) * [41]	6

* Pelvic floor distress inventory has been validated for telephone administration. [38]