Comparison of Responsiveness of Validated Outcome Measures After Surgery for Stress Urinary Incontinence

Anna C. Frick,* Beri Ridgeway, Mark Ellerkmann,† Mickey M. Karram,‡ Marie Fidela Paraiso,§ Mark D. Walters‖ and Matthew D. Barber

From the Departments of Obstetrics and Gynecology, Cleveland Clinic (ACF, BR, MFP, MDW, MDB), Cleveland and Good Samaritan Hospital (MMK), Cincinnati, Ohio, and Greater Baltimore Medical Center (ME), Baltimore, Maryland

Purpose: We compared the responsiveness of several validated incontinence, pelvic floor and quality of life outcome measures in women undergoing surgery for stress urinary incontinence to assist investigators in selecting appropriate outcomes in future trials of stress urinary incontinence therapy.

Materials and Methods: This is an ancillary analysis of data from a multicenter, randomized trial comparing tension-free vaginal tape and transobturator slings. All patients were asked to complete outcome measures at baseline and again 1 year postoperatively, including Incontinence Severity Index, Pelvic Floor Distress Inventory-Short Form 20, Pelvic Floor Impact Questionnaire-Short Form 7, Pelvic Organ Prolapse/Urinary Incontinence Sexual Function Questionnaire 12, SF-12® and a 3-day bladder diary. They also completed the Patient Global Index of Improvement at 1 year. We assessed the responsiveness of each outcome measure by calculating a standardized response mean and performing receiver operator characteristics curve analysis.

Results: Incontinence Severity Index, Pelvic Floor Distress Inventory-Short Form 20, Urinary Distress Inventory-Short Form, Pelvic Floor Impact Questionnaire-Short Form 7 and Urinary Impact Questionnaire-Short Form 7 showed excellent responsiveness (standardized response mean $\geq$1.0). Using receiver operator characteristics curve data the bladder diary had the greatest ability to discriminate patients who did vs did not improve (area under the curve 0.97). Incontinence Severity Index, Pelvic Floor Distress Inventory-Short Form 20 and Urinary Distress Inventory-Short Form also showed strong responsiveness according to these data (area under the curve greater than 0.7).

Conclusions: In this study of women undergoing mid urethral sling surgery for stress urinary incontinence the greatest responsiveness was noted on Incontinence Severity Index, Pelvic Floor Distress Inventory-Short Form 20, Urinary Distress Inventory-Short Form and bladder diary. Thus, they may be preferable as primary outcome measures in trials of stress urinary incontinence treatment.

Abbreviations and Acronyms

HRQOL = health related quality of life

IED = incontinence episodes per day

ISI = Incontinence Severity Index

PFDI-20 = Pelvic Floor Distress Inventory (short form)

PFIQ-7 = Pelvic Floor Impact Questionnaire (short form)

PGI-I = Patient Global Index of Improvement

PISQ-12 = Pelvic Organ Prolapse/Urinary Incontinence Sexual Function Questionnaire

SRM = standardized response mean

SUI = stress urinary incontinence

UDI-6 = Urinary Distress Inventory (short form)

UIQ-7 = Urinary Impact Questionnaire (short form)

Key Words: urethra; urinary bladder; urinary incontinence, stress; suburethral slings; questionnaires

Submitted for publication January 28, 2010.

Study received institutional review board approval at each participating center.

Supported by a research grant from American Medical Systems, Minnetonka, Minnesota.

* Correspondence: 9500 Euclid Ave., Desk A-81, Cleveland, Ohio 44195 (telephone: 216-445-6587; FAX: 216-636-0148; e-mail: fricka@ccf.org).

† Financial interest and/or other relationship with Ethicon Women's Health and Urology.

‡ Financial interest and/or other relationship with Astellas, Cooper Surgical, EWH & U, American Medical Systems and Emedsco.

§ Financial interest and/or other relationship with Coloplast.

‖ Financial interest and/or other relationship with American Medical Systems and Boston Scientific.

0022-5347/10/1845-2013/0
THE JOURNAL OF UROLOGY®

Vol. 184, 2013-2017, November 2010
Printed in U.S.A.
DOI:10.1016/j.juro.2010.06.114

To evaluate health related changes with time investigators must use instruments or measures that respond to a clinically significant change in the outcome of interest. Responsiveness or sensitivity to change refers to instrument ability to detect change that occurs as the result of therapy or disease progression. For example, a responsive measure of incontinence symptoms administered before and after surgery would show a large change in score in patients who report relief of incontinence symptoms after surgery. In contrast, use of a measure with poor responsiveness increases the risk of a type II error (failure to find evidence of a difference when in fact there is a difference) and can underestimate the effect of treatment. Adequate responsiveness is an essential property of any questionnaire intended to evaluate the effect of treatment and yet this is the property least often evaluated in the medical literature.[1] While a number of validated, responsive clinical outcome measures exist to assess the efficacy of SUI treatments, to date little data exist on the relative responsiveness of these measures.

We compared the responsiveness of several patient reported outcome measures in women undergoing SUI surgery. Specifically our objectives were to evaluate the responsiveness of 2 incontinence severity measures (ISI[2] and 3-day bladder diary IED), a measure of symptom bother (PFDI-20),[3] a measure of condition specific HRQOL (PFIQ-7),[3] a generic HRQOL measure (SF-12)[4] and a measure of sexual function (PISQ-12).[5] Information from this study should assist investigators in planning future trials of SUI therapy since responsiveness is an important consideration when selecting an appropriate clinical outcome.

MATERIALS AND METHODS

This is an ancillary analysis of data from a multicenter, randomized trial comparing tension-free vaginal tape with transobturator tape for SUI.[6] The study was approved by the institutional review board at each participating center.

Subjects were enrolled from 3 American tertiary care academic medical centers. Subjects were eligible for study if they showed urodynamic SUI on multichannel urodynamic testing, were at least 21 years old and elected surgical correction for incontinence. Subjects requiring concurrent surgery for pelvic organ prolapse were eligible for the study. Exclusion criteria were 1) detrusor overactivity on urodynamic testing, 2) post-void residual urine greater than 100 ml, 3) history of a sling procedure, 4) desire for future childbearing, 5) history of hidradenitis suppurativa, inguinal lymphadenopathy, or an inguinal or vulvar mass, 6) current genitourinary fistula or urethral diverticulum, or 7) another contraindication for surgery. Enrolled subjects were randomized to receive a tension-free vaginal tape (Ethicon, Cincinnati, Ohio) or a Monarc® subfascial hammock system with or without concurrent surgery for pelvic organ prolapse. For analysis purposes the treatment groups were pooled and considered a single group.

Participants completed several outcome measures at baseline and again 12 months after surgery, including ISI, PFDI-20, PFIQ-7, SF-12, PISQ-12 and a 3-day bladder diary. The PFDI-20 and PFIQ-7 each have 3 scales (urinary, colorectal and pelvic organ prolapse) and an overall summary score.[3] For analysis purposes we evaluated the responsiveness of the urinary scales of the PFDI-20 and PFIQ-7, UDI-6 and UIQ-7 as well as the summary scores of each scale. For the SF-12 we evaluated physical and mental summary scores. At 12 months subjects also completed the PGI-I, which asks subjects to rate bladder function on a 7-point scale from very much worse to very much better compared to that before surgery. Details of data collection and trial design were previously reported.[6]

Various methods have been used to assess responsiveness but no single approach has proven to be superior.[7,8] Thus, we evaluated responsiveness using multiple methods. First we calculated the mean change in score between the preoperative and 12-month postoperative visits for each scale by subtracting each participant preoperative score from the score at 1 year and calculating the mean difference in these values. On the PFDI-20, UDI-6, PFIQ-7, UIQ-7, ISI and bladder diary a negative change score indicates improved function, symptoms or quality of life. On the PISQ-12 and SF-12 a positive change score indicates improved sexual function and quality of life, respectively. Pretreatment scores were compared to posttreatment scores in each group using the paired t test. The percent change in score was also calculated by dividing the mean change in score by the mean preoperative score.

We then evaluated the relative responsiveness of each measure by comparing the SRM of each instrument. The SRM is a distribution based measure of responsiveness that is a ratio of individual change to the SD of that change. SRMs were calculated by dividing the absolute value of the mean change in score by the SD of the mean change in score.[9,10] A SRM of 0.2, 0.5 and 0.8 or greater is considered small, moderate and large, respectively.[11] We also measured responsiveness by comparing the ROC curves of each outcome measure. This tested the ability of each measure to discriminate participants with vs without improvement. Improvement was defined as a selection of somewhat better, much better or very much better on PGI-I while no improvement was defined as no different, somewhat worse or much worse on PGI-I. ROC curves for each outcome measure were generated to plot the true positive rate (sensitivity) against the false-positive rate (1 − specificity). The AUC reflects the ability of the measure to discriminate patients with improvement from those who report no improvement. An AUC of 1.0 would represent 100% accuracy while 0.5 would represent an estimate equivalent to chance. Thus, a higher AUC reflects better prediction of clinical improvement. We used Spearman's ρ correlation coefficient to evaluate the correlation between change scores and PGI-I. Data were analyzed using JMP® 8.0.

RESULTS

A total of 170 participants were randomized in this trial. Baseline characteristics and surgical procedure details in the study population were previously reported.[6] Of the participants 135 who completed