ORIGINAL ARTICLE

The minimum important difference for the Pelvic Organ Prolapse-Urinary Incontinence Sexual Function Questionnaire

M. M. Mamik · R. G. Rogers · C. R. Qualls · J. D. Morrow

Abstract

Introduction and hypothesis Although the Pelvic Organ Prolapse-Urinary Incontinence Sexual Function Questionnaire (PISQ) is widely used to assess sexual function in women, the minimum important difference (MID) (defined as the smallest difference in scores of a patient-reported outcome measure that is perceived by patients as beneficial or harmful and which would lead the clinician to consider a change in treatment) is not known. The objective was to estimate the MID for the PISQ.

Methods Two study populations, one of women with overactive bladder (OAB) and urgency UI (UUI) treated with tolterodine in a placebo-controlled trial (cohort I), and one of women treated surgically for prolapse and/or UI (cohort II) were used. Cohort I anchors were the Overactive Bladder Questionnaire (OAB-q), the Patient Perception of Bladder Condition (PPBC), the Patient Perception of Treatment Benefit Questionnaire (PPTBQ), and the change in number of UUI episodes in bladder diaries. Distribution MIDs were also calculated.

Results In the anchor-based analysis, the MID values for changes in PISQ total scores at 3 months in cohort I were 5 points using the UUI anchor (diary-dry women), 5 points using the PPBC anchor, 5 points with the PPTBQ, and 9 points with the OAB-q. In cohort II, the MID at week 12 in PISQ total scores was 7 points in women with improved IIQ-7 scores. The distribution-based MID in PISQ total scores was 5.3 points in cohort I and 5.8 points in cohort II.

Conclusion A reasonable estimate of MID for the PISQ total score is 6 points. Improvements that meet these thresholds may be considered clinically important.

Keywords Minimum important difference · Pelvic Organ Prolapse-Urinary Incontinence Sexual Function Questionnaire (PISQ)

M. M. Mamik (✉) · J. D. Morrow
Icahn School of Medicine at Mount Sinai, 1176 5th Avenue,
Klingenstein Pavilion, 9th Floor, Box 1170, New York, NY 10029,
USA
e-mail: mamta.mamik@gmail.com

R. G. Rogers · C. R. Qualls
University of New Mexico Health Sciences Center, Albuquerque,
NM, USA

Introduction

Pelvic floor disorders (PFD), including pelvic organ prolapse (POP) and urinary and fecal incontinence, are common and negatively affect the lives of women. A community-based survey in the USA found that 23.7 % of women suffered from at least one disorder [1]. PFDs affect women's social interactions, including intimacy with sexual partners, and treatments of PFDs have been associated with both beneficial [2] and deleterious effects on sexual health [3]. In 2001, Rogers et al. published a new measure of sexual health in women with urinary incontinence and/or pelvic organ prolapse, the Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire (PISQ), which is accepted as the standard measure of sexual health in women with pelvic floor dysfunction [4].

Although the PISQ is widely used to assess sexual function in women with pelvic floor disorders, clinically important differences in scores have not been determined. The smallest change in score associated with a clinically meaningful change in a questionnaire has been called the minimum important difference (MID) [5]. The MID is the difference in score that is representative of what patients and physicians perceive as beneficial or harmful and that would mandate, in the absence of troublesome side effects or excessive cost [6], a change in the patient's management. MID provides

interpretation of treatment effectiveness, linking clinical indicators with patient-reported outcomes (PROs) in health-related quality of life (QOL) measures. Our objective was to determine the MID of the PISQ using a dual-cohort methodology: one cohort comprising women with overactive bladder treated with tolterodine, and another cohort comprising women treated surgically for prolapse and/or incontinence.

Materials and methods

This is an ancillary analysis of a randomized clinical trial on the efficacy of tolterodine with regard to overactive bladder symptoms and sexual and emotional quality of life in sexually active women (cohort I; Tolterodine Efficacy in Sexually Active Women [TESA trial]) [7] and a trial that reported sexual function changes after surgery for urinary incontinence and/or prolapse (cohort II) [2]. Since the de-identified data were already available from the TESA trial, this ancillary analysis did not require an Institutional Review Board review. In the study of cohort I, sexually active women with urgency urinary incontinence (UUI) were randomized to placebo or tolterodine extended release for treatment of overactive bladder (OAB) symptoms for 12 weeks. Eligible women had OAB for at least 3 months, were sexually active in a stable heterosexual relationship, had a mean of ≥8 micturitions, ≥ 0.6 urgency incontinence episodes and ≥3 OAB micturitions per 24 h, as recorded in 5-day bladder diaries at baseline, and reported at least "some moderate problems" on the Patient Perception of Bladder Condition (PPBC) questionnaire. In the study for cohort II, women who underwent reconstructive surgery with placement of a suprapubic catheter for postoperative bladder management were randomized to nitrofurantoin monohydrate crystals or placebo for urinary tract infection prophylaxis [8]. A subset of randomized women who were sexually active in a heterosexual relationship participated in the sexual function trial, and the data from these women were used in the current analysis. The surgical trial evaluated sexual function changes in women who underwent incontinence and/or prolapse surgery and measured outcomes at baseline and again at 12 weeks following surgery.

Calculation of MID may be performed using anchor-based or distribution-based methods. Anchor-based approaches use an external indicator, either clinical or patient-reported, to assign subjects into groupings reflecting no change, small or large positive changes, or small or large negative changes in clinical or health status. Anchors include global ratings of change or actual changes in patient reported outcome measures that have ideally demonstrated MID values in the target population. For our anchor-based evaluation of the PISQ, we used multiple anchors in cohort I including the Overactive Bladder Questionnaire (OAB-q), the Patient Perception of Bladder Condition (PPBC), the Patient Perception of

Treatment Benefit Questionnaire (PPTBQ), and bladder diaries. The Incontinence Impact Questionnaire-7 (IIQ-7) was our anchor for cohort II. To evaluate whether or not an anchor was appropriately correlated with the PISQ, Pearson's correlation evaluated the associations between total PISQ scores and the three PISQ domains (Behavioral/Emotive, Physical and Partner-Related) with the various anchors. A correlation of $r=0.3$ was considered an acceptable correlation using Cohen's guidelines [9].

Distribution-based methods assess the MID by using the standard deviation (SD) of observed scores in a relevant sample. An estimate of one-half the SD or standard error of the mean (SEM) may be appropriate for approximating the MID for some PRO instruments. We chose to use half the SD, rather than SEM to approximate the MID by the distribution-based method [10]. For both the anchor- and distribution-based methods, we evaluated the change in scores from baseline to 12 weeks.

We employed a triangulation method of estimating the MID that entailed integrating global ratings with clinical benchmarks of change and statistical methods for estimating magnitude. The quantitative approach can be complemented by qualitative data from clinical experts or patients to provide insight into factors that must be considered when recommending guidelines for interpretation [11].

Measures

For both cohorts, demographic information was collected at baseline. For cohort I, 5-day bladder diaries recorded the time of each micturition UUI episode, at both baseline and at the 12-week follow-up. Validated QOL measures were completed in both cohorts at baseline and at follow-up.

The PISQ measures sexual function in women with pelvic floor disorders and consists of 31 questions in three domains: Behavioral/Emotive, Physical, and Partner-Related. Responses for each question range from 0 ("always") to 4 ("never"), with the exception of one question regarding the frequency of climax with masturbation with responses from 0 ("do not masturbate") and 1 ("never") to 5 ("always"). Total scores range from 0 to 125 with higher scores indicating better sexual function. The joint IUGA/ICS definition and terminology has been used [4].

The Patient Perception of Treatment Benefit Questionnaire (PPTBQ) [12] is a global measure of treatment benefit and satisfaction. It consists of two questions. Responses categorize the degree of benefit and satisfaction into "little" or "much" benefit or "little" or "very" satisfied with treatment.

The Patient Perception of Bladder Condition (PPBC) questionnaire is a validated and responsive global measure of bladder condition [13]. The PPBC consists of a single item that assesses the patient's subjective impression of her current urinary problems. Patients are asked to rate their perceived

bladder condition on a six-point scale ranging from 1 ("no problem at all") to 6 ("many severe problems"). Changes in scores are negative if patients improve and positive if their bladder condition deteriorates.

The OAB-q [14] is a validated condition-specific instrument that comprises an eight-item Symptom Bother scale and a 25-item health-related QOL (HRQL) scale divided into Concern, Coping, Sleep, and Social Interaction aspects for assessment. Responses for each item are scored on a six-point Likert scale. Scores for each scale are normalized to a 100-point scale. Higher Symptom Bother scores indicate greater symptom bother; higher HRQL scores indicate better HRQL. The MID of both the Symptom Bother and the HRQL scales of the OAB-q has been determined to be 10 [15].

The IIQ-7 is a seven-item validated, reliable, and responsive questionnaire [16]. Responses are on a four-point Likert scale; higher scores represent worse symptoms. Improved symptoms result in negative changes in scores. Although the MID for the long form version has been determined to be a difference of 16 points [6], the MID for the short form version of the questionnaire has not been determined.

Cohort I anchors included the change from diary wet to dry on week 12 bladder diaries, as well as changes in the PPBC, PPTBQ and OAB-q questionnaires. For the PPTBQ, we analyzed the difference in means of the PISQ scores between patients who had reported "little" benefit or satisfaction (minimum change) and patients who had reported "much" benefit or satisfaction (maximum change). Using the PPBC, we dichotomized the population into women who reported a score change equal to 0 or −1 (minimum change) and those with a score change ≤−2 (maximum change) to determine differences in mean PISQ scores. For the OAB-q, the population was likewise divided into two groups: those with an absolute change of less than or equal to 10 (minimum change) and those with an absolute change of greater than or equal to 10. The anchor in cohort II was the IIQ-7. The IIQ-7 is scored so that lower scores indicate better quality of life. IIQ scores remained the same (a change equal to 0), improved (changes ranging from −1 to −2) or deteriorated (changes in scores ranging from 1 to 2). To define minimal and maximum change for the IIQ, we dichotomized women into those who had a score change of equal to 0 or −1 (minimum change) and a score change less than or equal to −2 (maximum change); women with deterioration in their scores were not included in the analyses. Distribution MIDs for both cohorts were determined by calculating half the standard deviation in change in PISQ total and the domain scores from baseline to week 12.

Statistical analyses

Differences in the PISQ total and domain scores over 12 weeks were assessed using paired t tests and the differences in the change in PISQ scores over 12 weeks due to categorical

improvement in the anchor were assessed by two-sample t tests. MID is considered "not defined" for a given anchor if the two sample t tests indicated no difference in PISQ changes. If a statistically significant difference was found, then the MID for the PISQ change was defined to be the difference between the PISQ score in the maximum and the minimum groups. If there was a statistically significant difference, but the paired t test indicated no change in PISQ for the minimal improvement category of the anchor, then the MID was defined as being the mean difference in PISQ for the maximum improvement category of the anchor.

Results

In cohort I, 163 of the 202 subjects randomized to tolterodine (81 %) gave data through week 12. In cohort II, 75 of the 102 enrolled subjects whose condition improved or remained unchanged (75 %) completed questionnaires at both baseline and at 12 weeks. The mean age of women in the two cohorts was 49 and 47.1 years respectively, and the majority of women were White (Table 1). Thirty-two women in cohort II had prolapse beyond the hymen; women with stage 3 prolapse were excluded from participation in cohort I. More women in cohort II than in cohort I were premenopausal.

The correlation between all QOL measures and the PISQ was equal to or greater than 0.3 with the exception of the PPTBQ and the PPBC. For these two measures, the correlation as "r-coefficient" was not calculated. Both measures have questions that are reported as "yes/no" responses, which does not allow for correlation statistics. In the anchor-based analysis (Table 2), the MID values for changes in PISQ total scores at week 12 in cohort I were 4.7 points using the UUI anchor (diary-dry women at week 12), 5 points using the PPBC anchor, 5 points with PPTBQ, and 8.7 points with OAB-q. In cohort II, the MID at week 12 in PISQ total scores was 7 points in women with improved IIQ-7 scores. The distribution-based MID (Table 3) in PISQ total scores (i.e., half the SD in score change at week 12) was 5.3 points in cohort I and 5.8 points in cohort II. The range of MID values for total PISQ scores was narrow, from 4.7 to 8.7, using a

Table 1 Demographics

	Cohort I (n=163)	Cohort II (n=75)
Mean age (SD), years	49 (12)	47.1 (11.0)
Race, n (%)		
White	139 (69)	47 (46.5)
Black	37 (18)	20 (20)
Asian/other	25 (2)	34 (33.3)
Pre-menopausal (%)	71 (35)	65 (64)

Table 2 Anchor-based assessment of the minimum important difference (*MID*) of the Pelvic Organ Prolapse-Urinary Incontinence Sexual Function Questionnaire (*PISQ*)

Anchor	PISQ change	Maximum change in anchor, mean ± SD (*p* value)	Minimal change in anchor, mean ± SD (*p* value)	Two-sample *t* test (*p* value)	MID
Cohort 1					
Δ UUI episodes (diary-dry at 12 weeks)	Total PISQ	6.7±12.0 (<0.001)	2.0±8.0 (0.03)	3.2 (0.003)	4.7
	Behavioral/emotive	2.3±6.2 (<0.001)	0.5±5.7 (0.44)	2.0 (0.04)	2.3
	Partner-related	0.8±2.3 (0.001)	0.4±2.1 (0.10)	1.2 (0.22)	Not defined
	Physical	3.4±5.8 (<0.001)	1.2±3.4 (0.002)	3.2 (0.002)	2.2
PPBC	Total PISQ	8.35±11.95 (<0.001)	2.88±9.20 (0.02)	3 (0.004)	5
	Behavioral/emotive	2.74±6.15 (<0.001)	0.8±5.92 (0.30)	1.8 (0.07)	Not defined
	Partner-related	1.10±2.37 (<0.001)	0.4±2.07	1.8 (0.07)	Not defined
	Physical	4.52±5.89 (<0.001)	1.68±3.94 (<0.001)	3.3 (0.001)	3
PPTBQ	Total PISQ	7.3±10.9	2.3±10.2	2.7 (0.009)	5
	Behavioral/emotive	2.38±5.52 (<0.001)	0.55±6.45 (0.55)	1.7 (0.09)	Not defined
	Partner-related	1.05±2.01 (<0.001)	0.55±6.45 (0.46)	1.9 (0.07)	Not defined
	Physical	3.88±5.71 (<0.001)	0.55±6.45 (0.11)	3.0 (0.004)	3
Δ OAB-q score	Total PISQ	6±10.8 (<0.001)	−2.7±7.5 (<0.001)	0.6 (<0.0001)	8.7
	Behavioral/emotive	2.0±6.1 (<0.001)	1.6±5.9 (0.001)	0.6 (0.57)	Not defined
	Partner-related	0.8±2.1 (<0.001)	0.7±2.1 (<0.001)	0.4 (0.68)	Not defined
	Physical	3.1±5.2 (<0.001)	3.0±5.1 (<0.001)	0.2 (0.87)	Not defined
Cohort II					
Δ IIQ-7	Total PISQ	7.1±11.5 (<0.001)	−1.3±7.3 (0.56)	3.1 (0.003)	7
	Behavioral/emotive	0.1±8.6 (0.95)	−2.5±4.8 (0.08)	1.4 (0.17)	Not defined
	Partner-related	1.1±3.2 (0.01)	−0.8±4.6 (0.54)	1.3 (0.20)	Not defined
	Physical	5.5±6.7 (<0.001)	0.7±3.0 (0.56)	3.7 (<0.001)	6

variety of anchor- and distribution-based methods. In general, when choosing a MID value from a range, it is recommended that anchor-based methods are given more weight than distribution methods, because anchor-based methods reflect

Table 3 Distribution-based analyses of ΔPISQ for cohorts I and II

	Parameter	*n*	Δ PISQ± SD	*p* value	MID
Cohort I	ΔPISQ total	163	4.7±10.6	< 0.001	5.3
	ΔPISQ behavioral/emotive	163	1.6±6.0	< 0.001	3.0
	ΔPISQ partner-related	163	0.6±2.2	< 0.001	1.1
	ΔPISQ physical	163	2.5±5.1	< 0.001	2.5
Cohort II	ΔPISQ total	75	5.5±11.5	< 0.001	5.8
	ΔPISQ behavioral/emotive	75	0.5±6.9	0.6	3.5
	ΔPISQ partner-related	75	1.0 ±2.9	0.005	1.5
	ΔPISQ physical	75	4.3±6.0	<0.001	3.0

patient-rated and disease-specific variables [10]. MID estimations obtained from a bladder diary represent a more conservative estimate of clinically meaningful change than that of the patient's perspective, which is subject to recall bias [17]. Since the anchor-based values were lower, we propose an MID for total PISQ scores of 6.

The MID for the three PISQ domains were also analyzed, also using anchor- and distribution-based methods. As above, if the major change in an anchor was not significantly different than the minor change in an anchor, then the MID using that anchor was considered undefined. For the behavioral/emotive domain, the MID, using the anchor of resolution of UUI episodes on bladder diary (change from wet to dry in the diary at 12 weeks), was 2.3, and for all other anchors it was undefined. The distribution-based method of determining the MIDs for the behavioral/emotive domain using half a standard deviation of the PISQ behavioral/emotive domain scores was 3 for cohort I and 3.5 for cohort II. Therefore, the MID for the behavioral/emotive domain ranged from 2.3 to 3.5. For the

physical domain, the MID using the IIQ as the anchor was 6, for the bladder diary was 2.2; for all other anchors for the physical domain, the MID was undefined (Table 2). The distribution-based MID for the physical domain was 2.5 for cohort I and 3 for cohort II, giving a range of MID estimates for the physical domain of 2.2–6. For the partner-related domain, anchor-based methods yielded an undefined value in all instances. Distribution-based analysis showed MID of 1 for cohort I and 1.5 for cohort II. We weighted the MID estimates obtained from the bladder diary for each of the domain estimates because we felt that the diary data represented a more conservative estimate of clinically meaningful change than the patient's perspective as measured by questionnaires, which are subject to recall bias [16]. We therefore propose an MID of 2 for the behavioral/emotional domain, 1 for the partner-related domain, and 3 for the physical domain.

Discussion

This study used five anchor-based approaches and a distribution-based approach to establish the MID for the PISQ and the three domains of the PISQ. The anchor-based approaches included four measures from the patient's perspective (the PPBC, PPTBQ, OAB-q, and IIQ-7) and one clinical measure (the bladder diary). Using the anchor-based method, the range of MID for total PISQ scores was narrow, from 5 to 9. These data were supported for both cohorts by distribution-based analysis, with an MID of both cohorts within this range.

When multiple approaches are used to determine the MID of a scale, a range of values rather than a single-point estimate are the result, as was seen in this study. A narrow range of MID estimates adds strength to our final estimate of the MID of 6. We found a very narrow range of MID identified by multiple anchors, which was then confirmed by our distribution analysis.

In MID determination, distribution-based methods can be used in situations where anchor-based approaches are not available. There are increasing data and a growing consensus that an effect size of 0.5 (or a change of 0.5 SD) is a conservative estimate of the MID that is likely to be clinically significant across different patient-reported questionnaires. A weakness of using the 0.5 SD approach is that the estimates may not represent the minimally significant change [10]. In our study, the 0.5 SD represented the lower boundary of the range of estimates, adding strength to our final MID estimates.

The strengths of the study include the use of multiple anchors in both surgical and nonsurgical cohorts, and the inclusion of women with stress urinary incontinence (with or without prolapse) and OAB. Multiple approaches were then used to triangulate MID estimates following the recommended guidelines. All anchors used in the estimates were at least moderately correlated with the PISQ.

The limitations are that the computation of MID is limited by the strength of the validity and the reliability of the instruments involved. Use of anchors is limited by the degree of correlation they may have with the PISQ.

In conclusion, an estimate of MID for the PISQ total score using anchor- and distribution-based methods is 6 points. Improvements that meet these thresholds may be considered clinically important.

References

1. Nygaard I, Barber MD, Burgio KL, Kenton K, Meikle S, Schaffer J, Spino C, Whitehead WE, Wu J, Brody DJ, Pelvic Floor Disorders Network (2008) Prevalence of symptomatic pelvic floor disorders in US women. JAMA 300(11):1311–1316
2. Rogers RG, Kammerer-Doak D, Darrow A, Murray K, Qualls C, Olsen A, Barber M (2006) Does sexual function change after surgery for stress urinary incontinence and/or pelvic organ prolapse? A multicenter prospective study. Am J Obstet Gynecol 195(5):e1–e4
3. Handa VL, Cundiff G, Chang HH, Helzlsouer KJ (2008) Female sexual function and pelvic floor disorders. Obstet Gynecol 111(5):1045–1052
4. Rogers RG, Kammerer-Doak D, Villarreal A, Coates K, Qualls C (2001) A new instrument to measure sexual function in women with urinary incontinence or pelvic organ prolapse. Am J Obstet Gynecol 184(4):552–558
5. Crosby RD, Kolotkin RL, Williams GR (2003) Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 56(5):395–407
6. Barber MD, Spino C, Janz NK, Brubaker L, Nygaard I, Nager CW, Wheeler TL; Pelvic Floor Disorders Network (2009) The minimum important differences for the urinary scales of the Pelvic Floor Distress Inventory and Pelvic Floor Impact Questionnaire. 200(5):580.e1–580.e7
7. Rogers R, Bachmann G, Jumadilova Z, Sun F, Morrow JD, Guan Z, Bavendam T (2008) Efficacy of tolterodine on overactive bladder symptoms and sexual and emotional quality of life in sexually active women. Int J Urogynecol Pelvic Floor Dysfunct 19(11):1551–1557
8. Rogers RG, Kammerer-Doak D, Olsen A, Thompson PK, Walters MD, Lukacz ES, Qualls C (2004) A randomized, double-blind, placebo-controlled comparison of the effect of nitrofurantoin monohydrate macrocrystals on the development of urinary tract infections after surgery for pelvic organ prolapse and/or stress urinary incontinence with suprapubic catheterization. Am J Obstet Gynecol 191(1):182–187
9. Cohen J (1977) Statistical power analysis for the behavioural sciences. Academic Press, New York
10. Revicki D, Hays RD, Cella D, Sloan J (2008) Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 61(2):102–109
11. Leidy NK, Wyrwich KW (2005) Bridging the gap: using triangulation methodology to estimate minimal clinically important differences (MCIDs). COPD 2(1):157–165

12. Pleil AM, Coyne KS, Reese PR, Jumadilova Z, Rovner ES, Kelleher CJ (2005) The validation of patient-rated global assessments of treatment benefit, satisfaction, and willingness to continue—the BSW. Value Health 8 [Suppl 1]:S25–S34

13. Coyne KS, Matza LS, Kopp Z, Abrams P (2006) The validation of the patient perception of bladder condition (PPBC): a single-item global measure for patients with overactive bladder. Eur Urol 49(6): 1079–1086

14. Kelleher CJ, Cardozo LD, Khullar V, Salvatore S (1997) A new questionnaire to assess the quality of life of urinary incontinent women. Br J Obstet Gynaecol 104:1374–1379

15. Coyne KS, Matza LS, Thompson CL, Kopp ZS, Khullar V (2006) Determining the importance of change in the overactive bladder questionnaire. J Urol 176(2):627–632

16. Uebersax JS, Wyman JF, Shumaker SA, McClish DK, Fantl JA (1995) Short forms to assess life quality and symptom distress for urinary incontinence in women: the Incontinence Impact Questionnaire and the Urogenital Distress Inventory. Continence Program for Women Research Group. Neurourol Urodyn 14(2): 131–139

17. Jaeschke R, Singer J, Guyatt GH (1989) Ascertaining the minimal clinically important difference. Control Clin Trials 10:407–415