Neurourology and Urodynamics 30:541–546 (2011)

Further Validation of the Short Form Versions of the Pelvic Floor Distress Inventory (PFDI) and Pelvic Floor Impact Questionnaire (PFIQ)

Matthew D. Barber,[1]* Zhen Chen,[2] Emily Lukacz,[3] Alayne Markland,[4] Clifford Wai,[5] Linda Brubaker,[6] Ingrid Nygaard,[7] Alison Weidner,[8] Nancy K. Janz,[2] and Cathie Spino[2]

[1]*Obstetrics, Gynecology, and Women's Health Institute, Cleveland Clinic, Cleveland, Ohio*
[2]*Data Coordinating Center, University of Michigan, Ann Arbor, Michigan*
[3]*Department of Reproductive Medicine, University of California-San Diego, San Diego, California*
[4]*Department of Medicine, University of Alabama-Birmingham, Birmingham, Alabama*
[5]*Department of Obstetrics and Gynecology, University of Texas-Southwestern Medical Center, Dallas, Texas*
[6]*Department of Obstetrics & Gynecology and Urology, Loyola University, Chicago, Illinois*
[7]*Department of Obstetrics and Gynecology, University of Utah, Salt Lake City, Utah*
[8]*Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, North Carolina*

Aims: To evaluate validity and responsiveness of PFDI and PFIQ short forms across four multi-center studies and develop conversion formulas between short and long versions. **Methods:** 1,006 participants in four prospective studies of pelvic floor disorders completed long versions of the PFDI, PFIQ, and SF-36 (or SF-12) at baseline and 3 and 12 months after treatment. Responses were used to calculate scores for the short versions. We calculated correlations between scale versions using Pearson's correlation coefficient and compared their relative responsiveness using the standardized response mean. **Results:** PFDI and PFIQ short form scale scores demonstrated excellent correlations with long versions and similar responsiveness. Responsiveness was good to excellent for PFDI-20 urinary and prolapse scales, moderate for PFDI-20 colorectal scale and each of the PFIQ-7 scales, and poor for SF-36 (or SF-12) summary scores. Conversion formulas demonstrated excellent goodness of fit. **Conclusions:** The long and short forms of the PFDI and PFIQ correlate well and have similar overall responsiveness in participants from four different prospective multicenter studies consisting of diverse patient populations with a broad range of pelvic floor disorders. The short forms provide a reliable and valid alternative in situations where reduced response burden is desired. *Neurourol. Urodynam. 30:541–546, 2011.* © 2011 Wiley-Liss, Inc.

Key words: pelvic floor; pelvic organ prolapse; quality of life; questionnaires; responsiveness; urinary incontinence

INTRODUCTION

The Pelvic Floor Distress Inventory (PFDI) and the Pelvic Floor Impact Questionnaire (PFIQ) are two complementary condition-specific health-related quality of life questionnaires for women with pelvic floor disorders.[1] These two instruments are based on the structure and content of two widely-used condition-specific quality of life questionnaires for women with lower urinary tract dysfunction, the Urinary Distress Inventory (UDI) and the Incontinence Impact Questionnaire (IIQ), which were originally described by Shumaker et al.[2] Clinicians and researchers can use the PFDI and PFIQ together to measure how much lower urinary tract, lower gastrointestinal tract and pelvic organ prolapse symptoms affect the quality of life of women with pelvic floor disorders. Each measure has three scales: urinary, colo-rectal anal, and prolapse. The PFDI and PFIQ have each been shown to be psychometrically valid, reliable and responsive to change.[1,3,4] The 46-item PFDI assesses symptom distress in women with pelvic floor disorders and has three scales: the Urinary Distress Inventory (UDI; range 0–300), the Pelvic Organ Prolapse Distress Inventory (POPDI: range 0–300), and the Colorectal-Anal Distress Inventory (CRADI; range 0–400).[1] Similarly, the PFIQ measuring the impact of bladder, bowel, and vaginal symptoms on a woman's daily activities, relationships and emotions is composed of three scales of 31 questions each: the Urinary Impact Questionnaire (UIQ; range 0–400), the Pelvic Organ Prolapse Impact Questionnaire (POPIQ; range 0–400), and the Colorectal-Anal Impact Questionnaire (CRAIQ; range 0–400).[1]

In spite of the strengths of the PFDI and PFIQ, including their comprehensive coverage of symptom distress and impact on quality of life, their relative length may be inefficient or impractical for some clinical or research situations. Table I displays the item reduction used to develop the short forms of the PFDI (PFDI-20) and PFIQ (PFIQ-7).[5] The PFDI-20 and PFIQ-7 demonstrated excellent correlation with their long-form counterparts in the original validation population ($n = 100$) and in a second independent sample of 45 women undergoing pelvic reconstructive surgery ($r = 0.88$–0.94 for scales of PFDI-20; $r = 0.95$–0.96 for scales of PFIQ-7, $P < 0.0001$ for all). The test–retest reliability of each scale in the short forms was good to excellent (ICC 0.70–0.93, $P < 0.001$ for all scales). Moreover, the scales and summary scores of the PFDI-20 and PFIQ-7 demonstrated moderate to excellent responsiveness 3–6 months after surgery.[5]

Conflicts of interest: none.
The included trials registered at clinicaltrials.gov under Registration # NCT00065845 (CARE), NCT00271037 (Colpocleisis), NCT00270998 (ATLAS), and NCT00729144 (ABBI).
Grant sponsor: National Institute of Child Health and Human Development; Grant sponsor: NIH Office of Research on Women's Health; Grant Numbers: U01 HD41249. U10 HD41250. U10 HD41261. U10 HD41267. U10 HD54136. U10 HD54214. U10 HD54215. U10 HD54241.
Heinz Koelbl led the review process.
*Correspondence to: Matthew D. Barber, M.D., MHS, Cleveland Clinic, 9500 Euclid Ave, Desk A81, Cleveland, OH 44195. E-mail: barberm2@ccf.org
Received 27 January 2010; Accepted 24 March 2010
Published online 22 February 2011 in Wiley Online Library (wileyonlinelibrary.com).
DOI 10.1002/nau.20934

TABLE I. Comparison Between Short and Long Versions of the PFDI and PFIQ

Original questionnaire	Original scales	No. of items	Short form	Scales	No. of items
PFDI		46	PFDI-20		20
	Urinary distress inventory (UDI)	28		UDI-6	6
	Pelvic organ prolapse distress inventory (POPDI)	16		POPDI-6	6
	Colorectal-anal distress inventory (CRADI)	17		CRADI-8	8
PFIQ		93	PFIQ-7		21
	Urinary impact questionnaire (UIQ)	31		UIQ-7	7
	Pelvic organ prolapse impact questionnaire (POPIQ)	31		POPIQ-7	7
	Colorectal-anal impact questionnaire (CRAIQ)	31		CRAIQ-7	7

In this analysis, we planned to: (1) further evaluate the validity and responsiveness of the PFDI-20 and PFIQ-7 (short versions of the PFDI and PFIQ) across four multi-center studies using diverse patient samples and treatment approaches for pelvic floor disorders and (2) propose formulas for the conversion of scores between short and long versions of the PFDI and PFIQ.

MATERIALS AND METHODS

We analyzed 1,006 subjects who enrolled in one of four prospective studies (two surgical trials for pelvic organ prolapse, one non-surgical urinary incontinence trial, and one observational cohort study of women with fecal incontinence) conducted by the Pelvic Floor Disorders Network (PFDN) and completed at least one of the scales of the long-form version of PFDI and PFIQ at baseline; 84% and 73% of women also completed these at 3 and 12 months post-treatment follow-up. The designs of each trial, excluding the ongoing observational study of women with fecal incontinence, have been reported previously.[6–10] In brief, participants were: 316 women from the CARE trial, a randomized trial designed to evaluate whether a standardized modified Burch colposuspension, when added to abdominal sacrocolpopexy to treat pelvic organ prolapse, improves urinary stress continence in subjects without preoperative symptoms of stress urinary incontinence;[6,7] 140 from the colpocleisis trial, a cohort study studying the effect of colpocleisis on pelvic organ support, pelvic symptoms, quality of life, report-associated morbidity, and postoperative satisfaction;[8] 435 from the ATLAS trial, a randomized trial comparing behavioral therapy, incontinence pessary, and a combination of the two for treatment of stress urinary incontinence;[9] 115 from ABBI trial, an observational cohort study focusing on describing the use of adaptive behaviors among women undergoing treatment for fecal incontinence. Each clinical site and the data coordinating center in PFDN received institutional review board approval for each of the four trials, and all subjects provided written informed consent.

In each study, the instruments were administered either by telephone or in person at baseline and 3 and 12 months after the intervention. Participant responses to the PFDI and PFIQ individual items that are included in the short form versions were used to calculate the scores for PFDI-20 and PFIQ-7 scales including urinary, prolapse, and colo-rectal/anal subscales. The scales of PFDI-20 and PFIQ-7 (UDI-6, POPDI-6, CRADI-8; UIQ-7, POPIQ-7, CRAIQ-7) all have a range of 0–100, which is different than the scales of their long-form counterparts.[5]

In addition, participants in all the trials, except women with fecal incontinence participating in ABBI, completed the SF-36, a generic health-related quality of life questionnaire.[11] Instead, ABBI participants completed the SF-12, a shortened validated version of the SF-36. The generic health instrument, SF-36 or SF-12, was considered a priori for this ancillary analysis and we selected two summary scales, the mental and physical components. For both the long and short form versions PFDI and the PFIQ, a higher score indicates worse symptom bother or greater impact of symptoms on daily functioning; for the scales of the SF-36 or SF-12, a higher score indicates better health-related quality of life.[1,11]

For all analyses, we used each separate study population, as well as a combined group formed by pooling the 1,006 study participants from the four studies. In particular, the summary scores of the SF-12 were analyzed for ABBI and the SF-36 for each of other three trials first, and then the SF-12 and SF-36 were pooled together for the across-study combined sample.[12] The correlations between the corresponding scales of the long- and short-form versions of the PFDI and PFIQ at baseline were estimated using Pearson's correlation coefficients. In order to evaluate the relative responsiveness of the scales of PFDI, PFIQ, PFDI-20, PFIQ-7, and SF-36 and/or SF-12, the standardized response mean (SRM) of the change in scores from baseline to 3 months and baseline to 12 months after intervention for each scale was assessed; the SRMs were compared between the corresponding scales of the long and short forms of PFDI and PFIQ, between the condition-specific HRQOL (long- and short-form version of PFDI and PFIQ) and the generic HRQOL (SF-36 and/or SF-12) in a descriptive and exploratory fashion. SRM, a commonly used statistic of responsiveness, is equivalent to the change in score over a time period divided by the standard deviation of the change.[13] A higher SRM (in absolute value) indicates better responsiveness. A value of 0.5 is a cutoff for a moderate responsiveness, 0.8 a good responsiveness, and 1.0 an excellent one.[13]

We also used simple linear regression modeling to develop conversion formulas to calculate the scale scores of the PFDI and PFIQ long form from the short versions of the questionnaires. Each subscale of the PFDI or PFIQ short form was regressed on that of the long form and only the statistically significant parameters (i.e., intercept and slope) were retained in the final model. The model assumptions were examined graphically (e.g., linearity, normality, and influence cases) and R^2 values were reported. A split-sample validation method was used to assess the predictive accuracy of the conversion formulas.[14,15] The entire analysis sample ($n = 1,006$) was first stratified by study (condition-specific patient population), then within each study the sample was evenly partitioned into a training set (model development) and a test set (model validation) in a random fashion. The respective stratum-specific sets were recombined into a final training set and a final test set. The same linear models were developed from the training sample and then cross-validated in the test sample.[14,15] Discrimination was assessed by comparing the R^2 in the training set to that achieved in the test set. This was obtained by applying the linear model derived from the training sample to the test sample to yield a predicted long form score and then correlated it with the observed long form scale score in the test sample.[14] The model's calibration was validated graphically in the test set as well.[14] Based on the small shrinkage (the difference between R^2 of the training sample and R^2 of the test sample) and the calibration plot, the training

and test sets were combined for fitting the "final" proposed models.[14,15] The proposed conversion formulas were based on the pooled sample across the four studies.

All reported *P* values were based on the two-sided statistical tests. The analyses were performed in SAS 9.1.3 for Windows (SAS, Inc., Cary, NC).

RESULTS

Table II summarizes the demographics and baseline characteristics of the four study cohorts used in these analyses.[6–10] Variations in prevalence rates for urinary and fecal incontinence are representative of the inclusion criteria for each study. Overall, the rate of missing data was low for all studies at baseline (0.5–2%). All participants provided responses to at least one of the long-version scales of the PFDI, PFIQ, and SF-36 (or SF-12) at baseline, with 84% (849/1,006) available at 3 months and 73% (738/1,006) available at 12 months.

The overall pooled correlation coefficients between the long forms and selected questions representing the short orm version of the PFDI and PFIQ at baseline are shown in Table III. Overall, the correlation coefficients between the short and long version scores pooled across the four populations were excellent, with all subscales having Pearson's correlations of greater than 0.88 (all $P < 0.0001$).

Responsiveness to change of the PFDI and PFIQ short form scales was similar to that of the long versions across all four study populations (Table IV). In the pooled sample, the urinary and prolapse scales of the PFDI-20 demonstrated good responsiveness at 3 and 12 months (SRM −0.71 to −0.85). The highest SRM values were found in the POPDI responses collected from women enrolled in the surgical trials for treatment of prolapse (CARE = −1.35 at 3 months and −1.42 at 12 months; and Colpocleisis = −1.68 at 3 and 12 months) while the lowest SRM values were found in the POPDI responses from women enrolled in the fecal incontinence study (ABBI = −0.44 at 3 months) and women enrolled in the trial seeking conservative therapy for stress incontinence (ATLAS = −0.42 at 12 months) in the ABBI trial (evaluating behavior responses to fecal incontinence) and at 12 months (−0.42) in the POPDI responses from the population of women enrolled in the ATLAS trial. Overall, the colorectal subscale of the PFDI-20 demonstrated fair to moderate responsiveness across the three populations with primarily urinary and prolapse symptoms; however, in the group of women with fecal incontinence (ABBI trial), responsiveness to change was good at 3 months and excellent at 12 months (SRM −0.73 and −1.09, respectively).

Each of the scales of the PFIQ-7 demonstrated a broad range of SRM from −0.21 to −0.90, with higher SRMs in the UIQ and CRAIQ for the women with urinary and fecal incontinence respectively. The lowest values were in the POPIQ scores from the ATLAS group (SRM = 0.23 at 3 months and 0.21 at 12 months), however these values were markedly better in the population of women

TABLE III. Pearson Correlation Between Long and Short Form Versions of PFDI and PFIQ At Baseline

Scale	Subscale	N	Coefficient (range[a])	*P*-Value
PFDI	UDI	1,001	0.88 (0.81, 0.93)	<0.0001
	POPDI	1,000	0.90 (0.86, 0.90)	<0.0001
	CRADI	984	0.93 (0.91, 0.94)	<0.0001
PFIQ	UIQ	995	0.96 (0.95, 0.98)	<0.0001
	POPIQ	994	0.98 (0.97, 0.99)	<0.0001
	CRAIQ	997	0.98 (0.97, 0.98)	<0.0001

N, Pearson coefficient and *P*-value are from pooled sample across four studies.
[a]The range of Pearson correlation across four studies.

TABLE II. Demographic and Baseline Characteristics by Study and Overall

	CARE trial	Colpocleisis cohort	ATLAS trial	ABBI cohort	Total
N	316	140	435	115	1,006
Primary condition studied	Pelvic organ prolapse	Pelvic organ prolapse	Stress urinary incontinence	Fecal incontinence	—
Age (mean ± SD)	61.4 ± 10.3	79.2 ± 5.3	49.7 ± 11.9	57.8 ± 13.9	58.4 ± 14.7
Race (%)					
White/Caucasian	294 (93.0%)	129 (92.1%)	369 (85.0%)	99 (86.1%)	891 (88.7%)
Black/African American	16 (5.1%)	11 (7.9%)	45 (10.4%)	13 (11.3%)	85 (8.5%)
Other	6 (1.9)	0 (0.0%)	20 (4.6%)	3 (2.6%)	29 (2.9%)
Hispanic ethnicity (%)	9 (2.9%)	1 (0.7%)	32 (7.4%)	9 (7.8%)	51 (5.1%)
Educational level (%)					
Less than high school	26 (8.2%)	32 (22.9%)	18 (4.1%)	9 (7.9%)	85 (8.5%)
Completed high school or equivalent	121 (38.3%)	63 (45.0%)	81 (18.6%)	31 (27.2%)	296 (29.4%)
Some college or higher	169 (53.5%)	45 (32.1%)	336 (77.2%)	74 (64.9%)	624 (62.1%)
Body mass index (kg/m^2) (mean ± SD)	27.0 ± 4.5	27.8 ± 5.4	29.5 ± 6.8	29.0 ± 7.0	28.4 ± 6.1
Parity (median, range)	3 (1, 11)	NA	2 (0, 11)	3 (0, 9)	2 (0, 11)
Stress urinary incontinence (%)[a]	60 (19.1%)	105 (75.0%)	430 (99.3%)	77 (67.0%)	672 (67.1%)
Urge urinary incontinence (%)[b]	87 (27.5%)	91 (65.0%)	183 (42.5%)	69 (60.0%)	430 (42.9%)
Pelvic organ prolapse stage (%)[c]					
Stage 0/1	0 (0.0%)	0 (0.0%)	249 (57.9%)	NA	249 (28.3%)
Stage 2	44 (13.9%)	0 (0.0%)	181 (42.1%)		225 (25.5%)
Stage 3/4	272 (86.1%)	135 (100.0%)	0 (0.0%)		407 (46.2%)
Fecal incontinence (%)[d]	64 (20.3%)	44 (31.4%)	104 (24.0%)	104 (90.4%)	316 (31.4%)

NA, data not available.
[a]Defined as any affirmative response to one of the three items in the stress subscale of the Pelvic Floor Distress Inventory (PFDI).
[b]Defined as an affirmative response to item 19 "Do you usually experience urine leakage associated with a feeling of urgency, that is, a strong sensation of needing to go to the bathroom?" or item 28 "Do you usually experience bed-wetting?" of the PFDI.
[c]Pelvic organ prolapse quantitation (POPQ) stage.[26]
[d]Defined as an affirmative response to item 38 "Do you usually lose stool beyond your control if your stool is well formed?" or item 39 "Do you lose stool beyond your control if your stool is loose or liquid?" of the PFDI.

TABLE IV. Standardized Response Mean (SRM) From Baseline to 3 months and to 12 Months

		3 months				12 months			
		Long form		Short form		Long form		Short form	
Scale	Subscale	N	SRM (range[a])	N	SRM (range[a])	N	SRM (range[a])	N	SRM (range[a])
PFDI	UDI	846	−0.816 (−0.490, −1.281)	853	−0.714 (−0.556, −1.091)	737	−0.876 (−0.589, −1.265)	742	−0.749 (−0.535, −1.124)
	POPDI	849	−0.829 (−0.490, −1.450)	853	−0.850 (−0.437, −1.679)	738	−0.878 (−0.433, −1.473)	742	−0.931 (−0.420, −1.682)
	CRADI	826	−0.531 (−0.424, −0.738)	853	−0.464 (−0.361, −0.732)	722	−0.537 (−0.398, −0.947)	742	−0.460 (−0.294, −1.087)
PFIQ	UIQ	835	−0.563 (−0.396, −0.751)	854	−0.582 (−0.393, −0.841)	734	−0.694 (−0.599, −0.807)	743	−0.702 (−0.529, −0.797)
	POPIQ	833	−0.424 (−0.193, −0.663)	852	−0.440 (−0.229, −0.618)	733	−0.472 (−0.213, −0.618)	743	−0.485 (−0.211, −0.645)
	CRAIQ	836	−0.382 (−0.322, −0.693)	855	−0.367 (−0.321, −0.680)	735	−0.378 (−0.292, −0.908)	744	−0.363 (−0.268, −0.895)
SF-36/SF-12[b]	Physical summary	840	0.147 (0.022, 0.277)		NA	726	0.258 (0.084, 0.401)		NA
	Mental summary	840	0.115 (0.041, 0.449)		NA	726	0.137 (0.106, 0.342)		NA

N and standardized response mean (SRM) are from pooled sample across four studies.
NA, not applicable.
For the PFDI and PFIQ (long and short forms), a negative change in score indicates improvement. For the SF-36 or SF-12, a positive change in score indicates improvement.
[a]The range of standardized response mean (SRM) across four studies.
[b]SF36 for CARE, Colpocleisis, and ATLAS studies and SF-12 for ABBI study are pooled together.

undergoing prolapse surgery (greater than 0.60 at 3 and 12 months for both the CARE and Colpocleisis groups). In contrast, the SF-36 summary scores were relatively unresponsive to change (Table IV).

Conversion formulas to estimate long form scale scores were developed from PFDI-20 and PFIQ-7 questions within the baseline questionnaires. Each equation demonstrated goodness of fit ranging from R^2 values of 0.78–0.96 (Table V). For instance, the formula to convert the urinary scale of the PFDI-20 to the long form score is UDI score $= 1.9 \times$ UDI-6 score $+ 11$. To ensure the validity and reliability of the conversion formulas, "one-time data-splitting" was employed. Evidence of a good internal validity was confirmed by the similarity in R^2 values in the development and validation samples (Table V). The calibration (or reliability) and discrimination jointly demonstrated that the derived formulas had good predictive accuracy.[13]

DISCUSSION

We found excellent correlation between the long and short forms of the Pelvic Floor Distress Inventory (PFDI) and the Pelvic Floor Impact Questionnaire (PFIQ), allowing scientifically sound use of these short forms for clinical or research purposes. Barber et al.[1] reported that it took an average (±SD) of 23 ± 11 min to administer both the 46-item PFDI and 93-item PFIQ. The use of short forms instead can reduce participant burden in research settings, as the PFDI and PFIQ are often used in combination with other self-reported measures of interest.

Findings regarding the responsiveness of the short forms over time were generally very positive, with the PFDI-20 that focuses on symptom distress, the urinary and prolapse subscales demonstrating good to excellent responsiveness across populations. Not surprisingly, condition specific subscales of each questionnaire were the most responsive to the respective pelvic floor disorder of primary interest. For example, the colorectal subscale was somewhat more varied with fair to moderate responsiveness across populations of women presenting with prolapse and urinary incontinence, but good to excellent responsiveness in women presenting with a primary complaint of fecal incontinence (SRM −0.73 and −1.09) at 3 and 12 months, respectively. Similarly, the SRM values for symptom distress POPDI were the highest when the study population of interest was prolapse and lowest when it was not the primary outcome. In the PFIQ-7 short form, that focuses on the impact of symptoms on daily activities, the responsiveness of the subscales showed greater variability across patient populations and treatments, with better responsiveness in studies involving prolapse surgery. In the original validation for the short forms, the responsiveness for the CRADI-8 and the CRAIQ-7 were lower than the other subscales, with SRMs of 0.70 and 0.51, respectively.[3] Our findings using a larger cohort of women undergoing specific treatment for pelvic organ prolapse, urinary incontinence, and fecal incontinence suggest a lower responsiveness of the colorectal subscale in comparison to the other subscales (i.e., SRM for CRADI-8 = 0.46 and CRAIQ-7 = 0.37). These lower levels of responsiveness may be accounted for by an overall lower burden of colorectal disease among the majority of the women in the combined groups who presented for treatment for pelvic organ prolapse and/or urinary incontinence. The majority of the women in these multi-center studies were participating in specific intervention studies for pelvic organ prolapse, with and without stress urinary incontinence, with fewer women reporting fecal incontinence. No specific randomized, controlled treatments for colorectal disorders such as fecal incontinence were provided in these studies, although some women with fecal incontinence did receive treatment as part of the clinical observational study. Women in all the

TABLE V. Conversion Formula From Short Form Score to Long Form Score of PFDI and PFIQ At Baseline

Scale	Subscale	N[a]	Conversion formula (long form = slope × short form + intercept)	R^2 development sample[b]	R^2 validation sample[b]	R^2 total population
PFDI	UDI	1,001	UDI score = 1.9 × UDI-6 score + 11	0.77	0.78	0.78
	POPDI	1,000	POPDI score = 2.6 × POPDI-6 score + 13	0.78	0.83	0.80
	CRADI	984	CRADI score = 3.2 × CRADI-8 score + 10	0.86	0.85	0.86
PFIQ	UIQ	995	UIQ score = 3.3 × UIQ-7 score	0.92	0.93	0.92
	POPIQ	994	POPIQ score = 3.3 × POPIQ-7 score	0.95	0.96	0.96
	CRAIQ	997	CRAIQ score = 3.5 × CRAIQ-7 score	0.96	0.97	0.96

N, slope and intercept in conversion formula, and *P*-value are from pooled sample across four multi-center studies.
[a]Based on linear regression models, only statistically significant ($P < 0.0001$) slope and intercept are kept in the formulas.
[b]Sample sizes for development and validation samples are ½ of total N for each row.

studies may have experienced some improvement in colo-rectal symptoms with the treatments under study. Our finding that there is good responsiveness of the CRADI and CRAIQ in the ABBI study that specifically enrolled patients with fecal incontinence supports the use of these scales in this population. Several other quality of life questionnaires for women with fecal incontinence exist including the Fecal Incontinence Quality of Life Scale[16] and the Modified Manchester Questionnaire,[17] however the responsiveness of these instruments have not been evaluated. Future studies should be performed to determine the relative responsiveness of these various instruments.

As anticipated, we found that both the long and short forms of the PFDI and PFIQ were more responsive to change than the SF-36/SF-12 physical and mental component summary scores. In the initial validation of the PFDI-20 and PFIQ-7, low responsiveness was reported for the mental and physical component scores on the SF-36 (SRM range 0.12 and 0.28, respectively).[3] Our findings across four studies found similarly low responsiveness at 3 and 12 months follow-up for the SF-36 mental (SRM 0.12 and 0.14) and physical component scores (SRM 0.15 and 0.26) respectively. These findings are consistent with previous studies showing limited responsiveness of generic QOL measures such as the SF-36 compared to condition-specific measures for women treated for pelvic floor disorders.[18–21] The responsiveness of the SF-36 in studies involving other chronic diseases are somewhat mixed,[22–24] but in large part they are less responsive than condition-specific measures.[25]

The conversion scores reported in our study are intended for use in clinical and research settings for comparing outcomes measured with the PFDI-20 and PFIQ-7 to the more comprehensive instruments. The conversion formulas were developed for use in well-described, clinically relevant patient populations, such as women with specific degrees of pelvic organ prolapse with and without stress urinary incontinence, as well as fecal incontinence, and may not apply to general clinical populations of women without pelvic floor disorders.[7–10] Therefore, investigators should use caution in extrapolating these formulas to other populations.

Our use of four multi-center studies with varying patient demographics, disease characteristics, and treatments is a strength and allows more generalizability of our findings. Another strength of the study is the use of multiple modalities via self-reported and telephone interviews for administration of the PFDI and PFIQ long forms. Additionally, we developed and validated conversion formulas between the short and long forms that should be of benefit to researchers and clinicians. Although we derived scores for the short forms from responses to the long form rather than comparing subject responses from the short form itself, we believe this is a minor limitation. However, since subjects did not complete both questionnaires, we could not perform direct comparisons or evaluate issues of question order, question fatigue, and item grouping. Another limitation is that the measures of responsiveness used in this trial depend in part on the effectiveness of the interventions used. Our analysis of responsiveness included all patients in the four trials considered who completed questionnaires at baseline and follow-up, whether their treatment resulted in symptomatic improvement or not. As such, the responsiveness statistics reported likely represent conservative estimates.

Despite the very positive findings from this study advocating for use of the PFDI and PFIQ short forms in research and clinical settings, there may be some circumstances where the long versions of the PFIQ and PFDI are preferable. For example, the long forms may be preferable when a more comprehensive inventory of symptom distress and impact of pelvic floor disorders on daily activities is a primary study aim. In such cases, the long version could provide better characterization across the full spectrum of the disorder. A generic QOL measure such as the SF-36 or SF-12 may also be desirable if comparability of findings across populations is warranted.

CONCLUSIONS

In conclusion, the PFDI-20 and PFIQ-7 scales are well-correlated with the PFDI and PFIQ long forms and have similar overall responsiveness in four different prospective studies. Our findings provide further evidence that these short forms can be applied to studies that vary in intervention focus and type of pelvic floor disease. These short forms are excellent alternatives to PFDI and PFIQ when decreased response burden is desired in research and clinical settings.

REFERENCES

1. Barber MD, Kuchibhatla MN, Pieper CF, et al. Psychometric evaluation of two comprehensive condition-specific quality of life instruments for women with pelvic floor disorders. Am J Obstet Gynecol 2001;185:1388–95.
2. Shumaker SA, Wyman JF, Uebersax JS, et al. Health-related quality of life measures for women with urinary incontinence: The Incontinence Impact Questionnaire and the Urogenital Distress Inventory. Continence Program in Women (CPW) Research Group. Qual Life Res 1994;3:291–306.
3. Barber MD, Walters MD, Cundiff GW. Responsiveness of the Pelvic Floor Distress Inventory (PFDI) and Pelvic Floor Impact Questionnaire (PFIQ) in women undergoing vaginal surgery and pessary treatment for pelvic organ prolapse. Am J Obstet Gynecol 2006;194:1492–8.
4. Wren PA, Janz NK, Brubaker L, et al. Reliability of health-related quality-of-life measures 1 year after surgical procedures for pelvic floor disorders. Am J Obstet Gynecol 2005;192:780–8.
5. 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;193:103–13.
6. Brubaker L, Cundiff G, Fine P, et al. A randomized trial of colpopexy and urinary reduction efforts (CARE): Design and methods. Control Clin Trials 2003;24:629–42.
7. Brubaker L, Cundiff GW, Fine P, et al. Abdominal sacrocolpopexy with Burch colposuspension to reduce urinary stress incontinence. N Engl J Med 2006;354:1557–66.

8. Fitzgerald MP, Richter HE, Bradley CS, et al. Pelvic support, pelvic symptoms, and patient satisfaction after colpocleisis. Int Urogynecol J Pelvic Floor Dysfunct 2008;19:1603–9.
9. Richter HE, Burgio KL, Goode PS, et al. Non-surgical management of stress urinary incontinence: Ambulatory treatments for leakage associated with stress (ATLAS) trial. Clin Trials 2007;4:92–101.
10. Richter HE, Burgio KL, Brubaker L, et al. Continence pessary compared with behavioral therapy or combined therapy for stress urinary incontinence: A randomized controlled trial. Obstet Gynecol 2010;115:609–17.
11. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83.
12. Ware J, Jr., Kosinski M, Keller SD. A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. Med Care 1996;34:220–33.
13. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 2003;56:395–407.
14. Harrell FE. *Regression modeling strategies*. New York: Springer; 2001.
15. Pedhazur EJ. *Multiple regression in behavioral research*. New York: Holt, Rinehart, & Winston; 1982.
16. Rockwood TH, Church JM, Fleshman JW, et al. Fecal Incontinence Quality of Life Scale: Quality of life instrument for patients with fecal incontinence. Dis Colon Rectum 2000;43:9–16.
17. Kwon S, Visco AG, Fitzgerald MP, et al. Validity and reliability of the Modified Manchester Health Questionnaire in assessing patients with fecal incontinence. Dis Colon Rectum 2005;48:323–31.
18. Barber MD, Amundsen CL, Paraiso MF, et al. Quality of life after surgery for genital prolapse in elderly women: Obliterative and reconstructive surgery. Int Urogynecol J Pelvic Floor Dysfunct 2007;18:799–806.
19. Burgio KL, Kraus SR, Menefee S, et al. Behavioral therapy to enable women with urge incontinence to discontinue drug treatment: A randomized trial. Ann Intern Med 2008;149:161–9.
20. Goode PS, Burgio KL, Locher JL, et al. Effect of behavioral training with or without pelvic floor electrical stimulation on stress incontinence in women: A randomized controlled trial. JAMA 2003;290:345–52.
21. Norton C, Chelvanayagam S, Wilson-Barnett J, et al. Randomized controlled trial of biofeedback for fecal incontinence. Gastroenterology 2003;125:1320–9.
22. Van der Molen T, Postma DS, Schreurs AJ, et al. Discriminative aspects of two generic and two asthma-specific instruments: Relation with symptoms, bronchodilator use and lung function in patients with mild asthma. Qual Life Res 1997;6:353–61.
23. Wolinsky FD, Wyrwich KW, Nienaber NA, et al. Generic versus disease-specific health status measures. An example using coronary artery disease and congestive heart failure patients. Eval Health Prof 1998;21:216–43.
24. Jolly M, Utset TO. Can disease specific measures for systemic lupus erythematosus predict patients health related quality of life? Lupus 2004;13:924–6.
25. Wiebe S, Guyatt G, Weaver B, et al. Comparative responsiveness of generic and specific quality-of-life instruments. J Clin Epidemiol 2003;56:52–60.
26. Bump RC, Mattiasson A, BO K, et al. The standardization of terminology of female pelvic organ prolapse and pelvic floor dysfunction. Am J Obstet Gynecol 1996;175:10–7.