Validation of the Pelvic Floor Distress Inventory (PFDI-20) and Pelvic Floor Impact Questionnaire (PFIQ-7) in a Dutch population

Elaine Utomo, Bertil F.M. Blok, Anneke B. Steensma and Ida J. Korfage

Int Urogynecol J. 2014 Apr;25(4):531-44.

ABSTRACT

Aims

The objective of this study was to validate the Pelvic Floor Distress Inventory (PFDI-20) and Pelvic Floor Impact Questionnaire (PFIQ-7) in Dutch women.

Methods

Patients with pelvic floor dysfunction completed the Dutch questionnaires at (1) inclusion to evaluate internal consistency, (2) 1 week later to assess test-retest reliability, and (3) 6 months later to assess responsiveness and interpretability of change. To assess validity, floor and ceiling effects and construct validity were tested. A population-based sample (reference group) completed the questionnaires once.

Results

Data of 111 patients and 283 reference group participants were analysed. Internal consistency of baseline scores in patient and reference groups was moderate (Cronbach's alpha 0.52–0.60) to adequate in the PFDI-20 (Cronbach's alpha 0.71–0.84) and adequate in the PFIQ-7 (Cronbach's alpha 0.88–0.94). Both measures presented adequate test-retest reliability (intraclass correlation coefficient 0.79–0.91) and adequate responsiveness (area under the receiver-operating characteristic curve both 0.77). Interpretability was adequate for PFDI-20 and acceptable for PFIQ-7 with a clinically relevant minimally important change of −23 and −29 points, respectively. At baseline, the scales of the PFIQ-7 showed floor effects (44–55 %) in patients, though the PFIQ-7 summary score did not. No ceiling effects were observed. Construct validity was adequate with all predefined hypotheses confirmed regarding subgroup discrimination using pooled patient and reference group baseline data.

Conclusions

For assessing distress and health-related quality of life of pelvic floor dysfunction, the Dutch PFDI-20 and PFIQ-7 are reliable and valid in the general Dutch population, and also responsive and interpretable among tertiary care seeking women.

INTRODUCTION

Women with complaints of pelvic floor dysfunction may present a variety of symptoms relating to lower urinary tract dysfunction, pelvic organ prolapse, and anorectal dysfunction.[1] Given the coexistence and complex interaction of these symptoms, comprehensive condition-specific health-related quality of life (HRQOL) questionnaires are needed to measure severity and impact of pelvic floor dysfunction in a standardized and reproducible manner.[2]

The Pelvic Floor Distress Inventory (PFDI) and Pelvic Floor Impact Questionnaire (PFIQ)[3] were developed and validated to measure symptom distress and impact on HRQOL, focusing on pelvic organ prolapse (POP), urinary incontinence (UI), and fecal incontinence (FI). Short versions of the PFDI (i.e., PFDI-20) and PFIQ (i.e., PFIQ-7) were developed to reduce respondent's burden.[4,5] The PFDI-20 and PFIQ-7 have been validated in different languages[6-14] though not yet in Dutch. We therefore translated the PFDI-20 and PFIQ-7 into Dutch. Since measures must adequately address measurement properties to be useful in research or in practice[15, 16] we tested their reliability, validity, responsiveness, and interpretability in Dutch women with symptoms of pelvic floor dysfunction.

METHODS

This observational study was conducted at a tertiary urology and gynaecology centre as part of a larger validation study of HRQOL pelvic floor measures. The study was approved by the Institutional Ethics Committee (MEC-2008-376).

Questionnaires

In both PFDI-20 and PFIQ-7, patients reported whether they experienced symptoms of pelvic floor dysfunction, and if so, the extent to which these symptoms were bothersome to them. The PFDI-20 has three scales: Pelvic Organ Prolapse Distress Inventory (POPDI-6), Colorectal-Anal Distress Inventory (CRADI-8), and Urinary Distress Inventory (UDI-6). Response options for rating distress associated with each symptom ranged from 0 ("no" as in no symptoms) to 1 ("not at all" as in symptoms are present but not bothered at all) to 4 ("quite a bit" as in symptoms are present and quite a bit bothered). Per scale, the mean score of answered items is multiplied by 25 to obtain the scale score (range 0–100).[4] Summary scores are calculated by adding up the scale scores (range 0–300). Higher scores indicate more symptom distress.[4] The PFIQ-7 measures impact of bladder, bowel, and vaginal symptoms on daily physical activity, travel, social/relationships, and emotional health. The PFIQ-7 has three scales: the Urinary Impact Questionnaire (UIQ-7), the Colorectal-Anal Impact Questionnaire

(CRAIQ-7), and the Pelvic Organ Prolapse Impact Questionnaire (POPIQ-7). Response options range from 0 ("not at all") to 3 ("quite a bit"). Per scale, the mean score of answered items is multiplied by 33.3 to obtain the scale score (range 0–100).[4] Summary scores are calculated by adding up the scale scores (range 0–300).[4] Higher scores indicate more impact on daily activity.

Linguistic validation

The measures were translated into Dutch following standardized forward-backward procedures,[17] i.e., three independent forward translations, backward translation by a native speaker, and tested face-to-face in ten patients with pelvic floor dysfunction. Several minor problems were identified without the need to adapt the content of the measures. The Dutch versions of PFDI-20 and PFIQ-7 were then finalized (See 'Vragenlijsten').

Study population and study design

For the patient group, inclusion criteria were women aged 18 years or older, with at least one symptom of pelvic floor dysfunction (UI, POP, or FI) existing for at least 3 months, and fluent and literate in the Dutch language. Exclusion criteria were active malignant tumours, dementia, and mental retardation. The definitions used for UI, POP, and FI were according to the standardized International Continence Society terminology.[18, 19] Thus, UI was defined as "the complaint of any involuntary leakage of urine," POP as "the symptomatic descent of one or more of the anterior vaginal wall, the posterior vaginal wall, and the apex of the vagina or vault after hysterectomy," and FI as "any involuntary loss of fecal material".[19] At the initial office visit, the treating physician recruited all consecutive patients potentially eligible for inclusion.

Patients received the same questionnaires at three time points (1) baseline, (2) 1 week later, and (3) 6 months later, and returned them through postal mail. During the test-retest period of 1 week, no treatment was initiated or changed. One week between the repeated administrations was long enough to prevent recall, though short enough to ensure that clinical change had not occurred. Given the observational nature of this study, the treating physician was unrestricted in prescribing therapy after completion of the test-retest for individual patient care. Age and education were documented through the questionnaire. Educational level was classified as "low" (primary school), "intermediate" (high school), or "high" (college or university degree). In the third round the health transition item of the RAND 36-Item Health Survey (RAND 36-HTI)[20, 21] was included through which patients scored the change in their general health compared to 1 year ago. The response options range from 1 ("much better") to 5 ("much worse"). The prescribed treatment between baseline and the third round of the questionnaire was obtained from the patient's medical record. This was categorized into conservative, pharmaceutical, or surgical treatment.

Reference group

We collected baseline data from a representative sample of adult women in the Netherlands through an ISO-certified (ISO 26362) Dutch panel.[22] This subsample was stratified by gender, age, educational level, and residential area and therefore representative of the Dutch population above the age of 18. No medical data were available for this panel, and thus the presence or absence of pelvic floor dysfunction was unknown.

Measurement properties

The psychometric measurement properties addressing the quality domains reliability, validity, and responsiveness were tested, and we assessed interpretability which is an important characteristic of measures.[15]

- Internal consistency (i.e., the degree of interrelatedness among the items) was assessed with Cronbach's alpha. Cronbach's alpha was determined for the summary score as well as for the scale scores. The scales intend to measure a single underlying concept by using multiple items. A high Cronbach's alpha indicates high correlations between the multiple items in a scale, i.e., redundancy of one or more items thus measuring the same concept. Values between 0.70 and 0.95 were considered to reflect adequate internal consistency.[16, 23]
- Test-retest reliability (i.e., the degree to which repeated measurements in stable persons provide similar answers) was assessed through the intraclass correlation coefficient (ICC) for agreement.[24] We used the two-way mixed single measures model in IBM® SPSS software 20.0 to calculate this. Values ≥0.70 were considered to reflect adequate reliability.[16]
- Measurement error (i.e., the systematic and random error of a patient's score not attributed to true changes) was quantified using the limits of agreement (LOA) by Bland and Altman.[25] The LOA were calculated as the absolute mean change in scores ($mean_{change}$) of repeated measurements during the test-retest period ±1.96 SD (SD_{change}). The absolute change in mean scores of the test-retest was presented without a specific direction of change.
- Content validity (i.e., the degree to which the content of a measure is an adequate reflection of the target construct) was subjectively assessed and verified ("face validity") by examining whether the items appeared to be measuring what they are intended to measure. We also assessed floor and ceiling effects. These occur when ≥15 % of scores are at the lower or upper end of the scale.[16, 23]
- Construct validity (i.e., the degree to which measure scores are consistent with hypotheses based on the assumption that the measure validly measures the target construct) was verified by hypotheses testing of the measures using known groups based on patient and reference group baseline scores.[23] Since no medical back-

ground information was available for the reference group (e.g., care-seeking or not, presence or absence of symptoms of pelvic floor dysfunction), we defined the presence of symptoms according to the response options given in the measures. The PFIQ-7 scale scores were used to assess PFDI-20 construct validity, i.e., the presence of POP, FI, or UI was defined as a score above zero on the POPIQ-7, CRAIW-7, or UIQ-7 scale, respectively. Specific items of the PFDI-20 scores were used to assess PFIQ-7 construct validity, i.e., an affirmative response to any of the items of the PFDI-20 indicating a sensation of a "*bulge*" in the pelvic area was defined as presence of POP; an affirmative response to any of the items of the PFDI-20 indicating "*stool loss*" was defined as presence of FI; and an affirmative response to any of the items of the PFDI-20 indicating "*urine leakage*" was defined as presence of UI. We formulated the following hypotheses:

1. The subgroup with POP symptoms will report significantly higher POPDI-6 and POPIQ-7 scale scores than the subgroup without POP symptoms.
2. The subgroup with symptoms of FI will report significantly higher CRADI-8 and CRAIQ-7 scale scores than the subgroup without FI symptoms.
3. The subgroup with symptoms of UI will report significantly higher UDI-6 and UIQ-7 scale scores than the subgroup without UI symptoms.

   Construct validity was considered adequate when at least 75 % of these hypotheses were confirmed.[16]

- The responsiveness (i.e., the ability of a measure to detect changes over time) was assessed in all patients who received any treatment during follow-up. We evaluated the linear relationship of the $mean_{change}$ in measure scores between baseline and 6-month follow-up with the RAND 36-HTI score using the Pearson correlation coefficient (*r*). Additionally, the area under the receiver-operating characteristic (ROC) curve (AUC) for the measures was determined.[23] By determining the AUC, the measures are considered as a diagnostic test, and the RAND 36-HTI functions as the gold standard. The AUC indicates the probability that a measure correctly classifies patients as improved using the RAND 36-HTI as an anchor. The AUC was considered adequate if ≥0.70.[23]
- The interpretability (i.e., the degree to which one can assign qualitative meaning to a measure's quantitative score or change in scores) was tested using the anchor-based ROC method to assess the minimally important change (MIC).[23] As in diagnostic studies, the optimal ROC cut-off point is chosen, i.e., the value for which the sum of the proportions of misclassifications [(1-sensitivity) + (1-specificity)] is smallest. In analogy, the MIC is defined as this optimal ROC cut-off point.[23]

Statistical methods

Statistical analysis was performed using IBM® SPSS software 20.0. Significance was set at a *p*-value <.05. For comparison of two independent groups we used the unpaired *t* test when numerical data were considered and the chi-square test when categorical data were considered. For comparison of more than two independent groups, we used the one-way analysis of variance (ANOVA). General linear models were used to compare measure scores, controlling for variables that differed significantly between patient and reference groups in univariate analysis.

The anchor RAND 36-HTI was dichotomized for ROC analysis: patients reporting to be "a little better" or "much better" were classified as "improved," while "same," "a little worse," or "much worse" were classified as "not improved."

RESULTS

Of 187 consecutive female patients, 161 (86%) were eligible for inclusion of which 117 (73%) consented to participate (Figure 3.1). After three rounds of questionnaires, data were available for 111/117 (95%) patients to test at least one measurement property (i.e., patients completed a questionnaire at baseline and at least at one additional time point) of at least one measure (i.e., patients completed the PFDI-20 and/or PFIQ-7). Regarding the reference group, the measures were sent out to 450 panel participants of which 283 (63 %) responded.

The mean age was higher (p<.001) in the patient group (59±12 years) than in the reference group (47±15 years, Table 3.1). When completing the PFDI-20, 96% of the patients reported symptoms of UI, 57% symptoms of POP, and 47% symptoms of FI. These prevalence rates were lower (p<.001) in the reference group. Of the patients, 74% (n=82) had more than one symptom of pelvic floor dysfunction (i.e., UI, or POP, or FI), compared to 17% (n=49) in the reference group.

Patients reported more symptom distress (PFDI-20) and more impact on daily activity (PFIQ-7) than the reference group (both p<.001). After adjusting for age and educational level, these differences remained significant.

Internal consistency

The PFDI-20 summary score demonstrated adequate internal consistency with Cronbach's alphas of 0.74 and 0.84 in patient and reference groups, respectively (Table 3.2). The scales of the PFDI-20 demonstrated moderate internal consistency for both patient (Cronbach's alpha 0.52–0.71) and reference groups (Cronbach's alpha 0.60–0.74). The PFIQ-7 summary and scale scores demonstrated adequate internal

Table 3.1 Characteristics and measure scores of respondents

Demographics	**Patients** (n = 111)	**Reference group** (n = 283)	***p*-value** univariate	***p*-value** GLM**
Age (years)	58.6 ± 12.2	47.0 ± 15.3	<.001	
Educational level			<.001	
Low	45 (41%)	84 (30%)		
Intermediate	51 (46%)	47 (17%)		
High	15 (14%)	152 (54%)		
Parity	2.3 ± 1.2	NA		
missing	*25 (23%)*	*NA*		
Urinary Incontinence[a] (%)	106 (96%)	128 (45%)	<.001	
Pelvic Organ Prolapse[b] (%)	63 (57%)	59 (21%)	<.001	
Faecal Incontinence[c] (%)	51 (47%)	24 (9%)	<.001	
missing	*2 (2%)*	*0 (0%)*		
Symptoms of pelvic floor dysfunction per patient (n)[d]			<.001	
0	0 (0%)	109 (39%)		
1	29 (26%)	125 (44%)		
2	55 (50%)	34 (12%)		
3	27 (24%)	15 (5%)		
Scores at baseline*				
PFDI-20 (0-300)	93.4 ± 41.5	27.1 ± 31.7	<.001	<.001
missing	*1 (1%)*	*0 (0%)*		
POPDI-6 (0-100)	22.2 ± 16.5	5.6 ± 9.8	<.001	<.001
CRADI-8 (0-100)	26.7 ± 19.5	9.3 ± 12.4	<.001	<.001
missing	*1 (1%)*	*0 (0%)*		
UDI-6 (0-100)	44.7 ± 20.7	12.3 ± 16.1	<.001	<.001
PFIQ-7 (0-300)	69.0 ± 52.5	12.7 ± 27.0	<.001	<.001
missing	*14 (13%)*	*0 (0%)*		
POPIQ-7 (0-100)	14.6 ± 22.4	3.4 ± 10.0	<.001	<.001
missing	*14 (13%)*	*0 (0%)*		
CRAIQ-7 (0-100)	20.9 ± 27.2	4.3 ±11.9	<.001	<.001
missing	*12 (11%)*	*0 (0%)*		
UIQ-7 (0-100)	35.1 ± 26.3	5.0 ± 11.3	<.001	<.001
missing	*6 (5%)*	*0 (0%)*		

Data in mean ± standard deviation or number (%). No data was missing unless otherwise indicated.
Abbreviations: *GLM* general linear models; *NA* not available

*A higher score indicates greater symptom distress or impact on health-related quality of life
***p*-value: corrected for age and educational level with general linear modelling (GLM)
[a]Defined as affirmative response to any of the items based on *"urine leakage"* (i.e. items 15, 16, 17 or 18 of the PFDI-20).

[b]Defined as affirmative response to any of the items based on a sensation of *"a bulge"* in the pelvic area (i.e. items 3, 4, 6 or 14 of the PFDI-20).
[c] Defined as affirmative response to any of the items based on *"stool loss"* (i.e. items 9 or 10 of the PFDI-20).
[d]Pelvic floor dysfunction include urinary incontinence, pelvic organ prolapse, and faecal incontinence

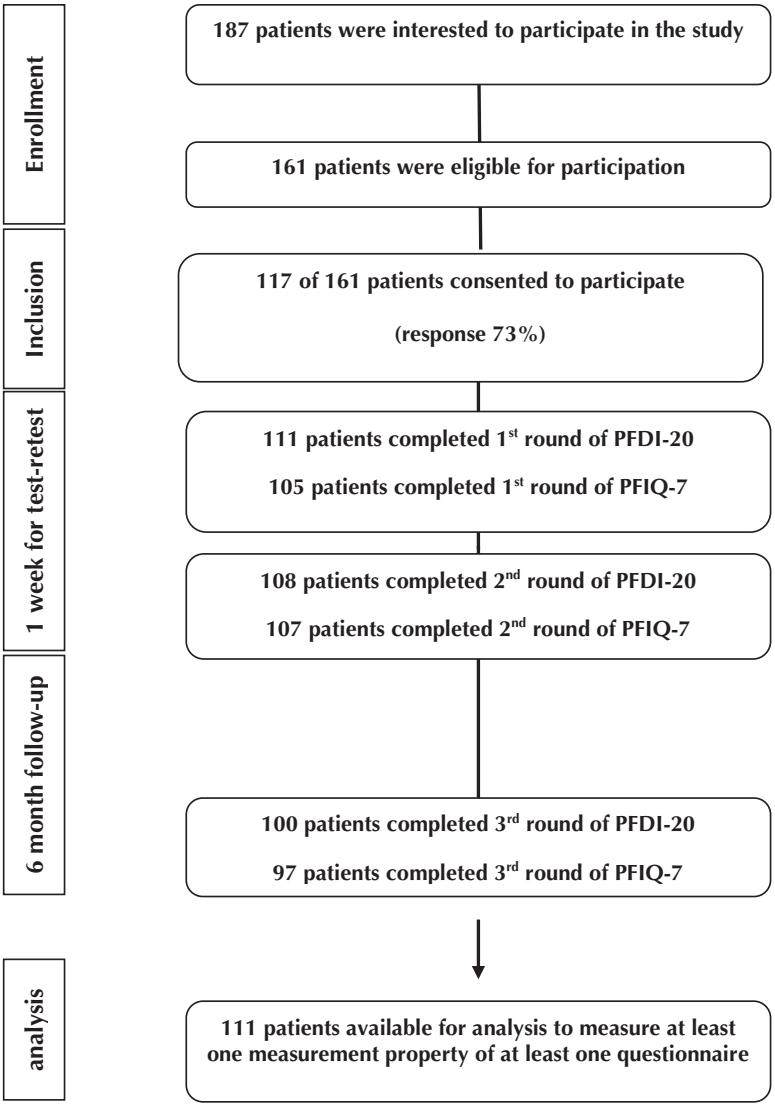

Figure 3.1 Flowchart: inclusion process of patient group

consistency in both patient (Cronbach's alpha 0.89–0.94) and reference groups (Cronbach's alpha 0.88–0.93).

Reliability

The retest assessments were completed on average 8 days after baseline measurement. The $ICC_{agreement}$ was 0.88 for the PFDI-20 summary score and ranged from 0.86 to 0.90 for the PFDI-20 scales (Table 3.2). The $ICC_{agreement}$ of the PFIQ-7 summary score was 0.83 and ranged from 0.79 to 0.91 for the PFIQ-7 scales. All retest assessments indicated adequate reliability.

Measurement error

The absolute $mean_{change}$ of repeated measurements during the test-retest period, the corresponding SD_{change}, and the LOA are presented in Table 3.2. Relating the LOA range to the range of all possible measure scores, the overall magnitude of the measurement error is 10% (59.2 of 600) for the PFDI-20 summary score and 13–14%(27, 27.4, and 26.6

Table 3.2 Internal consistency and reproducibility

Questionnaire	Internal consistency (Cronbach's alpha)		Test-retest reliability Patients		
	Patients	Reference group (n = 283)	Intraclass Correlation Coefficient	$mean_{change} \pm SD_{change}$*	Limits of agreement**
PFDI-20 (0-300)	0.74 (n=96)	0.84	0.88 (n=99)	13.3 ± 15.1	-16.3 – 42.9
POPDI-6 (0-100)	0.52 (n=103)	0.60	0.86 (n=100)	5.6 ± 6.9	-7.9 – 19.1
CRADI-8 (0-100)	0.71 (n=106)	0.71	0.90 (n=99)	5.6 ± 7.0	-8.1 – 19.3
UDI-6 (0-100)	0.60 (n=104)	0.74	0.89 (n=100)	6.8 ± 6.8	-6.5 – 20.1
PFIQ-7 (0-300)	0.89 (n=89)	0.93	0.83 (n=85)	19.2 ± 26.0	-31.8 – 70.1
POPIQ-7 (0-100)	0.92 (n=90)	0.92	0.79 (n=86)	7.3 ± 13.3	-18.8 – 33.4
CRAIQ-7 (0-100)	0.94 (n=94)	0.92	0.91 (n=86)	5.8 ± 10.2	-14.2 – 25.8
UIQ-7 (0-100)	0.90 (n=101)	0.88	0.82 (n=95)	9.9 ± 12.7	-15.0 – 34.8

*Higher scores indicate greater symptom distress or impact on health-related quality of life
** Limits of agreement described by Bland and Altman[25]= $mean_{change} \pm 1.96 * SD_{change}$

of 200 for the POPDI-6, CRADI-8, and UDI-6, respectively) for scale scores, 17% (101.9 of 600) for the PFIQ-7 summary score, and 20–26% (52.2, 40, and 49.8 of 200 for the POPIQ-7, CRAIQ-7, and UIQ-7, respectively) for scale scores.

Floor and ceiling effects

No ceiling effects were observed in the PFDI-20 and PFIQ-7 summary and scale scores, either in the patient or reference group. In patients, floor effects were found in two

Table 3.3 Floor and ceiling effects of baseline scores

Questionnaire	Patients (n=111)				Reference group (n=283)			
	Floor		Ceiling		Floor		Ceiling	
	n	Cumulative %	n	Cumulative %	n	Cumulative %	n	Cumulative %
PFDI-20 (0 – 300)	0	0	0	0	64	23	0	0
POPDI-6 (0 – 100)	14	13	0	0	177	63	0	0
CRADI-8 (0 – 100)	10	9	0	0	127	45	0	0
UDI-6 (0 – 100)	4	4	0	0	117	41	0	0
PFIQ-7 (0 – 300)	5	5	0	0	172	61	0	0
POPIQ-7 (0 – 100)	53	55	1	1	237	84	0	0
CRAIQ-7 (0 – 100)	44	44	1	1	225	80	0	0
UIQ-7 (0 – 100)	14	13	1	1	205	72	0	0

scales of the PFIQ-7, i.e., POPIQ-7 and CRAIQ-7, with 55 and 44% of subjects, respectively, exhibiting the most favourable low score of zero (Table 3.3). No floor effects were observed in the PFIQ-7 summary score. In the reference group, the PFDI-20 and PFIQ-7 summary and scale scores had floor effects ranging from 23 to 84%.

Construct validity

We appropriately confirmed all three predefined hypotheses (Figure 3.2):

1. The subgroup with POP symptoms (symptoms of vaginal bulge) reported higher scores on the POPDI-6 (23.3±17.9) and POIQ-7 (12.8±20.6) scales than the subgroup without symptoms (5.7±9.3 and 3.4±10.7, respectively), (p<.0001).
2. The subgroup with FI symptoms (stool loss) reported higher scores on the CRADI-8 (30.0±18.7) and CRAIQ-7 (30.2±28.4) scales than the subgroup without symptoms (7.0±9.4 and 3.6±10.7, respectively), (p<.0001).
3. The subgroup with UI symptoms (urine leakage) reported higher scores on the UDI-6 (39.1±21.8) and UIQ-7 (19.3±23.7) scales than the subgroup without symptoms (7.2±9.5 and 1.3±6.0, respectively), (p<.0001).

Responsiveness

A total of 100 patients completed the PFDI-20 and 97 patients the PFIQ-7 on average 5.7 months after baseline assessment (third round, Figure 3.1). Of 111 patients, 67 (60%) received treatment during follow-up as determined by the treating physician. Patients received conservative (n=50, 75 %), pharmaceutical (n=11, 16 %), and/or surgical treatment (n =44, 66 %). Table 3.4 presents the $\text{mean}_{\text{change}}$ in measure summary scores between baseline and 6-month follow-up of the measures, stratified by the responses of the RAND 36-HTI. We combined the categories of patients reporting their health

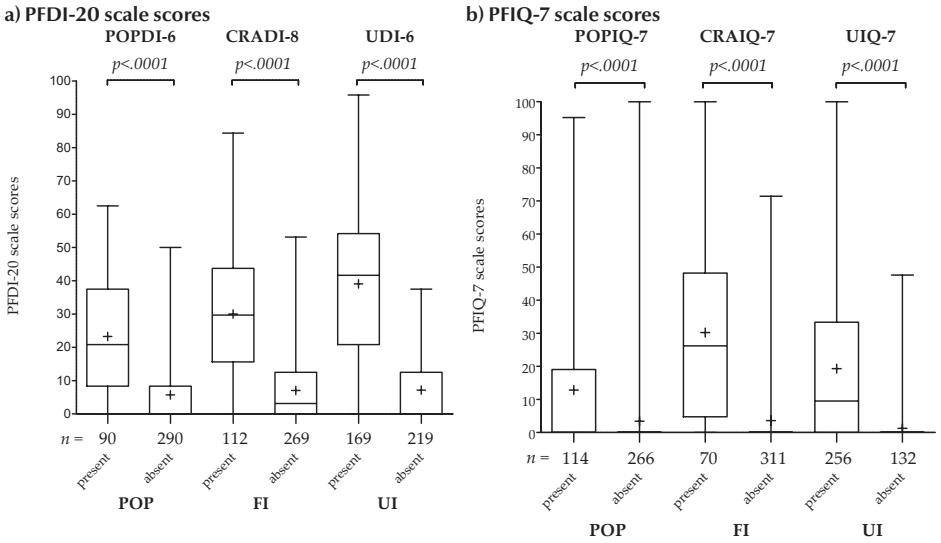

Figure 3.2 Construct validity. Pooled patient and reference baseline scales scores with comparisons between known groups. The bold lines indicates the median scale score and the box the interquartile range, the whiskers presenting the min.-max. The mean is indicated with a plus sign.

compared to 1 year ago as "much worse" (n=4) or "a little worse" (n=9). The relationships between the $mean_{change}$ of the summary scores of the measures and the RAND 36-HTI as anchor were linear (PFDI-20 r =0.49, p<.001; PFIQ-7 r =0.41, p=.002). The AUC for both measures was 0.77 (p<.001 for PFDI-20 and p=.001 for PFIQ-7), demonstrating adequate responsiveness.

Interpretability

Using ROC analysis, the optimum cut-off point was used to determine the MIC. Thus, with 73% correctly identified as improved and 75%as not improved, the MIC for the PFDI-20 summary score was –22.9 (Table 3.4). According to the lower limit of the LOA (Table 3.2), improvements in PFDI-20 summary scores of ≤16.3 cannot be attributed to a true improvement of the respondent, but should be attributed to a measurement error. The MIC of the PFDI-20 (–22.9) lies adequately outside the LOA (–16.3 to 42.9), meaning an improvement of ≥22.9 points on the PFDI-20 summary scale is a true clinically relevant change in score. Treated patients reporting "a little better" on the RAND 36-HTI 6 months after baseline showed mean changes of –32.3 on the PFDI-20 summary score, which indicates a true clinically relevant improvement.

The MIC for the PFIQ-7 summary score was –28.6, with 75% correctly identified as improved and 77 % as not improved (Table 3.4). This MIC (–28.6) lies between the LOA (Table 3.2: –31.8 to 70.1) and therefore cannot be detected as a true clinically relevant improvement, since an improvement of ≤31.8 points should be attributed to

Table 3.4 Responsiveness and Interpretability

Health transition item (RAND-36)*	**Number (%) n = 67**	**PFDI-20 Summary score (0 – 300)**	**PFIQ-7 Summary score (0 – 300)**
Much worse / A little worse	13 (16%)	-1 ± 18.4	-23.2 ± 52.6
Same	31 (39%)	-13.3 ± 25.3	-1.8 ± 40.7
A little better	9 (11%)	-32.3 ± 40.2	-31.3 ± 36.4
Much better	14 (18%)	-47.7 ± 39.6	-73.6 ± 51.7
p-value between RAND-36 items**		.001	.001
Pearson correlation coefficient *r*		0.49	0.41
p-value		<.001	.002
Area under the ROC curve		0.77	0.77
p-value		<.001	.001
Minimal important change		-22.9	-28.6
Sensitivity; specificity		0.73 ; 0.75	0.75 ; 0.77

Data presented in number (%) or $mean_{change} \pm SD_{change}$ between baseline and 6-month follow-up. Negative scores indicate an improvement in pelvic floor distress and/or impact.
Abbreviation: *ROC* receiver-operative characteristic
* *"Compared to one year ago, how would you rate your health in general now?"*
** Analysis of variance (ANOVA) was used

a measurement error. The change values lying between the lower LOA (–31.8) and the MIC (–28.6) are considered important by patients, but cannot be distinguished from measurement error. Treated patients reporting "a little better" on the RAND 36-HTI 6 months after baseline showed mean changes of –31.3 for the PFIQ-7 (Table 3.4), which is close to the lower limit of the LOA.

DISCUSSION

The objective of this study was to validate the PFDI-20 and PFIQ-7 in Dutch to provide an adequate tool to assess symptom distress and impact of pelvic floor dysfunction on HRQOL. Our findings regarding the reliability, construct validity, and responsiveness were generally very positive.

Both summary scales and four of six scales scores showed adequate internal consistency. The POPDI-6 and UDI-6 showed low to moderate internal consistency with Cronbach's alphas of 0.52 and 0.60 in patient groups, respectively, indicating

low homogeneity. Similar alphas were found in the Turkish[11] and Swedish[10] versions. As expected, floor effects were present in the reference group (23–84%) and ceiling effects were absent (0%) in summary and scale scores of both measures. In patients, we observed substantial floor effects in the POPIQ-7 (55%) and CRAIQ-7 scales (44%), indicating a not normal score distribution towards the most favourable low score of zero. The PFIQ-7 summary score did not demonstrate floor effects. Potentially patients typically experience different types of pelvic floor dysfunction (Table 3.1), but not necessarily *all* possible symptoms, e.g., POP and FI without UI.

The MIC of the PFIQ-7 indicated that the measurement error was too wide to detect clinically relevant improvement. However, the LOA are group averages, and just as the normal distribution, the measurement error range is bell-shaped and balances out towards the outer limits and therefore possibly less clinically relevant at the lower limit. Furthermore, the MIC lies close to the lower limit of agreement, with a small difference of 3.2 points on a scale of 0–300 which represents meaningful improvement, but cannot be distinguished from measurement error. We therefore recommend using the lower limit of agreement of −31.8 as MIC for PFIQ-7.

Our study has some limitations. First, it is desirable to have a correlation of 0.5–0.6 between the anchor and the meanchange in scores at the 6-month follow-up[23], while our anchor had a correlation of r =0.49 and r =0.41 for the PFDI-20 and PFIQ-7, respectively. This can be because generic measures, such as our RAND 36-HTI anchor, may not reveal the impact of sexual, urinary, and bowel dysfunctions on patients, because such dysfunctions tend not to be perceived as health problems.[26] Furthermore, it is preferable to measure the change in health prospectively. Second, since medical information about the reference group was not available, we determined the prevalence rates of symptoms of pelvic floor dysfunction using self-reported items of the PFDI-20. Third, this validation study applies to measurements in groups of patients, thus not directly to individual patients for individual measurements. Nevertheless, these questionnaires may create the opportunity to discuss certain important, but possibly less obvious, topics (e.g., emotional health) with the patient. Fourth, our study setting allowed us to determine the MIC of improvement, though not of deterioration. Only patients after treatment were included for analysis of responsiveness and interpretability. These patients were treated with the intention to improve their condition. We therefore did not determine the MIC of deterioration. And lastly, our study lacks the use of objective measures for assessment of symptoms of pelvic floor dysfunction.

The strengths of our study are the extensive assessment of measurement properties, covering the reliability (test-retest), validity, and—in particular—the responsiveness and interpretability of the measures in women after any treatment of pelvic floor dysfunction. An exception is the criterion validity, which requires the availability of a gold standard. Since there are no gold standards of symptoms and bother of pelvic

floor dysfunction, we were not able to test this property. We also included a reference population which is an important addition for the interpretation of the score distribution of the general population. The prevalence rates we found were comparable to the reported rates in the literature. In our study, the prevalence rates of UI and FI in the reference group were 45 and 9%, respectively, and are similar to the reported prevalence of UI in women in the general population (30–60%)[27,28] and FI in men and women in the general population (0–15.2% without significant gender differences).[29] The prevalence of POP in our reference group (21%) was higher than the reported prevalence of 4–15%.[27] Though this may be explained by the definition used by Milsom[27]: the reported rates are based on pelvic heaviness, genital bulge, and digital pressure on the vagina or perineum by defecation, and we included the need for digital pressure in the vaginal area for urination in the definition of POP. Ultimately, we had a high response rate of 73% with a limited dropout rate at 6 months of only 17%.

In conclusion, our findings provide positive evidence for the appropriateness of the Dutch PFDI-20 and PFIQ-7 to measure symptom distress and quality of life changes in women with pelvic floor dysfunction. With this study we validated the Dutch versions of the PFDI-20 and PFIQ-7 allowing use of these short forms for clinical or research purposes.

Acknowledgments

Part of this study was funded by the Urological Research Foundation [Stichting Urologisch Wetenschappelijk Onderzoek (SUWO), Rotterdam, the Netherlands]. SUWO was not involved in the design, conduct, interpretation, and analysis of the study or review or approval of the manuscript.

REFERENCES

1. Haylen BT, de Ridder D, Freeman RM, et al. An International Urogynecological Association (IUGA)/International Continence Society (ICS) joint report on the terminology for female pelvic floor dysfunction. Int Urogynecol J 2010;21:5-26.
2. Barber MD. Questionnaires for women with pelvic floor disorders. Int Urogynecol J Pelvic Floor Dysfunct 2007;18:461-5.
3. Barber MD, Kuchibhatla MN, Pieper CF, Bump RC. Psychometric evaluation of 2 comprehensive condition-specific quality of life instruments for women with pelvic floor disorders. Am J Obstet Gynecol 2001;185:1388-95.
4. 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.
5. Barber MD, Chen Z, Lukacz E, et al. Further validation of the short form versions of the Pelvic Floor Distress Inventory (PFDI) and Pelvic Floor Impact Questionnaire (PFIQ). Neurourol Urodyn 2011;30:541-6.
6. de Tayrac R, Deval B, Fernandez H, Mares P, Mapi Research I. [Development of a linguistically validated French version of two short-form, condition-specific quality of life questionnaires for women with pelvic floor disorders (PFDI-20 and PFIQ-7)] Validation linguistique en francais des versions courtes des questionnaires de symptomes (PFDI-20) et de qualite de vie (PFIQ-7) chez les patientes presentant un trouble de la statique pelvienne. J Gynecol Obstet Biol Reprod (Paris) 2007;36:738-48.
7. El-Azab AS, Abd-Elsayed AA, Imam HM. Patient reported and anatomical outcomes after surgery for pelvic organ prolapse. Neurourol Urodyn 2009;28:219-24.
8. Kaplan PB, Sut N, Sut HK. Validation, cultural adaptation and responsiveness of two pelvic-floor-specific quality-of-life questionnaires, PFDI-20 and PFIQ-7, in a Turkish population. Eur J Obstet Gynecol Reprod Biol 2012;162:229-33.
9. Lowenstein L, Levy G, Chen KO, Ginath S, Condrea A, Padoa A. Validation of Hebrew versions of the Pelvic Floor Distress Inventory, Pelvic Organ Prolapse/Urinary Incontinence Sexual Function Questionnaire, and the Urgency, Severity and Impact Questionnaire. Female Pelvic Med Reconstr Surg 2012;18:329-31.
10. Teleman P, Stenzelius K, Iorizzo L, Jakobsson U. Validation of the Swedish short forms of the Pelvic Floor Impact Questionnaire (PFIQ-7), Pelvic Floor Distress Inventory (PFDI-20) and Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire (PISQ-12). Acta Obstet Gynecol Scand 2011;90:483-7.
11. Toprak Celenay S, Akbayrak T, Kaya S, Ekici G, Beksac S. Validity and reliability of the Turkish version of the Pelvic Floor Distress Inventory-20. Int Urogynecol J 2012;23:1123-7.
12. Treszezamsky AD, Karp D, Dick-Biascoechea M, et al. Spanish translation and validation of four short pelvic floor disorders questionnaires. Int Urogynecol J 2013;24:655-70.
13. Yoshida M, Murayama R, Ota E, Nakata M, Kozuma S, Homma Y. Reliability and validity of the Japanese version of the pelvic floor distress inventory-short form 20. Int Urogynecol J 2012.
14. Zhu L, Yu S, Xu T, et al. Chinese validation of the Pelvic Floor Impact Questionnaire Short Form. Menopause 2011;18:1030-3.
15. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010;63:737-45.

16. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34-42.
17. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1993;46:1417-32.
18. Abrams P, Cardozo L, Fall M, et al. The standardisation of terminology of lower urinary tract function: report from the Standardisation Sub-committee of the International Continence Society. Neurourol Urodyn 2002;21:167-78.
19. Abrams P, Andersson KE, Birder L, et al. Fourth International Consultation on Incontinence Recommendations of the International Scientific Committee: Evaluation and treatment of urinary incontinence, pelvic organ prolapse, and fecal incontinence. Neurourol Urodyn 2010;29:213-40.
20. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. Health Econ 1993;2:217-27.
21. van der Zee KI, Sanderman R. Het meten van de algemene gezondheidstoestand met de RAND-36. Een handleiding. Groningen.
22. Flycatcher Academic Support Centre. 2012, at http://www.flycatcher.eu/nld/academic-support/.
23. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine. 1st ed. Cambridge: Cambridge University Press; 2011.
24. McGraw K, Wong S. Forming inferences about some intraclass correlation coefficients. Psychological Methods 1996;1:30-46.
25. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.
26. Korfage IJ, Hak T, de Koning HJ, Essink-Bot ML. Patients' perceptions of the side effects-of prostate cancer treatment – a qualititative interview study. Soc Sci Med 2006; 63(4):911-19.
27. Milsom I. Epidemiology of Urinary and Faecal Incontinence and Pelvic Organ Prolapse. In: Abrams P, Cardozo L, Khoury S, Wein A, eds. 4th International Consultation on Incontinence. 4th ed. Paris: Health Publication Ltd; 2009.
28. van der Vaart CH, de Leeuw JR, Roovers JP, Heintz AP. [The influence of urinary incontinence on quality of life of community-dwelling, 45-70 year old Dutch women] De invloed van urine-incontinentie op de kwaliteit van leven bij thuiswonende Nederlandse vrouwen van 45-70 jaar. Ned Tijdschr Geneeskd 2000;144:894-7.
29. Pretlove SJ, Radley S, Toozs-Hobson PM, Thompson PJ, Coomarasamy A, Khan KS. Prevalence of anal incontinence according to age and gender: a systematic review and meta-regression analysis. Int Urogynecol J Pelvic Floor Dysfunct 2006;17:407-17.