CrossMark

ORIGINAL ARTICLE

Validation of the PFDI-20 and PFIQ-7 quality of life questionnaires in two African languages

Etienne W. Henn[1] · Barry W. Richter[1] · Masekhoko M. P. Marokane[1]

Abstract

Introduction and hypothesis Self-administered quality of life (QOL) questionnaires provide objective evaluation of an individual's symptoms. The Pelvic Floor Distress Inventory-20 (PFDI-20) and the Pelvic Floor Impact Questionnaire-7 (PFIQ-7) are condition-specific short form questionnaires. There are very few validated QOL questionnaires for women in Africa. The aim of this study was to validate these questionnaires in African women for the Afrikaans and Sesotho languages.

Methods Patients with pelvic floor disorders completed the questionnaires at baseline, 1 week later and after 6 months. A control group of women not known to have pelvic floor disorders completed the questionnaires at baseline and 1 week later. Psychometric properties tested were internal consistency, reliability, construct validity and responsiveness.

Results In each language group, 100 control and 100 study participants completed the scheduled rounds. Internal consistency, as measured by the Cronbach's alpha value, was good for the PFDI-20 (0.71–0.89) and the PFIQ-7 (0.81–0.89) for both the Afrikaans-speaking and the Sesotho-speaking patients. The test–retest reliability showed very good intraclass correlation coefficients of 0.89–0.99 across all scales of both

questionnaires and in both language groups. The construct validity was confirmed as was the responsiveness to treatment for both questionnaires.

Conclusions The Afrikaans and Sesotho versions of the PFDI-20 and PFIQ-7 are reliable and valid instruments that can be used in women with pelvic floor disorders speaking these languages.

Introduction

Pelvic floor dysfunction (PFD) manifests as a variety of symptoms which can have a potentially detrimental effect on a woman's quality of life (QOL). The aim of any treatment plan for such an individual is the restoration of function. The objective and standardized assessment of symptoms are of vital importance to jointly guide the planning of treatment and the evaluation of functional outcomes. This assessment can be strengthened through the use of condition-specific health-related QOL questionnaires [1]. The Pelvic Floor Distress Inventory-20 (PFDI-20) and the Pelvic Floor Impact Questionnaire-7 (PFIQ-7) are validated short form questionnaires that measure symptom severity and the impact on QOL [2]. These questionnaires have been linguistically and culturally validated in diverse languages and cultures [3–12]. This process additionally allows international comparison of symptom severity and responses to treatment. They have however never been validated in any African language or culture. The primary objective was the cultural and linguistic validation of these questionnaires for the Afrikaans and Sesotho languages spoken in Southern Africa.

✉ Etienne W. Henn
hennew@ufs.ac.za

1 Department Obstetrics and Gynaecology G71, University of the Free State, PO Box 339, Bloemfontein 9301, South Africa

Materials and methods

This research was performed at a referral urogynecology unit in Universitas Academic Hospital as well as at a general gynecology outpatient clinic in Pelonomi Hospital, Bloemfontein, South Africa. The research was approved by the local ethics committee (HSREC 51/2016). The PFDI-20 measures symptom severity and consists of three scales: the Pelvic Organ Prolapse Distress Inventory-6 (POPDI-6), the Colorectal–Anal Distress Inventory-8 (CRADI-8) and the Urinary Distress Inventory-6 (UDI-6). Each item in a specific scale can be scored from 0 to 4, and the scale score is calculated as a percentage. The PFDI-20 score is the sum of the scale scores out of 300. The PFIQ-7 measures symptom impact on QOL and also consists of three scales: the Urinary Impact Questionnaire-7 (UIQ-7), the Colorectal–Anal Impact Questionnaire-7 (CRAIQ-7) and the Pelvic Organ Prolapse Impact Questionnaire-7 (POPIQ-7). Each item in these scales can be score from 0 to 3, and the scale score is calculated as a percentage. The PFIQ-7 score is the sum of the scale scores out of 300. A higher score in each of these questionnaires indicates more symptom bother.

Linguistic and cultural validation

The questionnaires were translated into Afrikaans and Sesotho by means of three forward translations by clinicians and backward translation by a native speaker not involved in the forward translation [13, 14]. They were then tested on a sample of ten women with PFD who completed the questionnaires and were interviewed afterwards. The interviews revealed that the public awareness of PFD was very low and the phrasing of the questions required simplification to reflect cultural concepts with regard to pelvic floor function. The questionnaire layout was designed to enhance accurate questionnaire completion. They were again tested on a further sample of ten women which confirmed comprehension and ease of completion of the questionnaires, and thereafter it was rolled out in full.

Study design and population

Women were eligible for inclusion if they were over the age of 18 years, did not suffer from any chronic pain syndrome and were literate in the language evaluated. The study group consisted of women referred to the urogynecology unit with complaints of PFD. The control group consisted of women seen at a general gynecology outpatient clinic who were not referred due to PFD. The sample size was determined by the number of items in the questionnaire with a minimum of five participants per item required [15]. The PFDI-20 had the most items and the aim was thus to recruit 100 study and 100 control participants per language group. Definitions used for

pelvic organ prolapse (POP), urinary incontinence (UI) and anal incontinence (AI) were according to the International Urogynecological Association (IUGA)/International Continence Society (ICS) report on terminology for POP [16]. AI was marked as present if the symptom of either fecal and/or flatus incontinence was asserted at baseline data completion.

Participants completed both questionnaires at the baseline visit and 1 week later. The study group also completed the questionnaires 6 months later to allow the evaluation of intervention. The questionnaires were completed after 1 week by telephone [17]. The desired modality is self-administration. This was, however, the only feasible option to permit test–retest assessment in this population due to the vast referral area, cost of transport, lack of internet facilities and lack of an effective postal service. All of these factors are frequently present in most African countries. Only the first 20 participants of the study groups and the first 10 participants of the control groups were used for this due to financial constraints in this study. Any treatment was deferred during this week to avoid test–retest changes related to clinical improvement. The study group completed the patient global impression of improvement (PGI-I) in the third round [18]. The PGI-I was dichotomized into "improved" for scores 1–3 and "not improved" for scores 4–7.

Measurement properties

The psychometric domains of reliability, validity and responsiveness were assessed for each questionnaire as recommended in the consensus-based standards for the selection of health measurement instruments (COSMIN) checklist [19]. Internal consistency was assessed by Cronbach's alpha value for the extent to which the individual scale items measured the same concept. Acceptable values ranged from 0.70 to 0.95 [20]. Test–retest reliability was assessed by completing the questionnaire twice with 1 week between completion rounds. This time period was considered long enough to prevent recall bias, but short enough for any relevant clinical changes to occur. It was calculated in terms of the intraclass correlation coefficient (ICC) between the baseline and second round of completion and was performed for all individual scales as well as for the summary scores. Values of ≥0.70 were considered to indicate adequate reliability [21]. Measurement error was expressed as the calculated limits of agreement (LOA) and are summarized as the mean change of scores during the test–retest period and was a reflection of the random error of scores not attributable to true clinical changes [22].

Construct validity assesses the validity of the questionnaires to measure the given construct, i.e. PFD [21]. We hypothesized that: (1) women with symptoms of POP will score higher in the POPDI-6 and POPIQ-7 scales than those without (2) women with symptoms of UI will score higher in the UDI-

Table 1 Baseline characteristics and questionnaire scores of participants

Metric	Afrikaans			Sesotho		
	Study group (n = 100)	Control group (n = 100)	p value	Study group (n = 100)	Control group (n = 100)	p value
Age (years)	59.94 ± 9.75	48.76 ± 13.38	<0.0001*	56.42 ± 11.77	46.45 ± 11.75	<0.0001*
Education						
Primary	26 (26)	6 (6)	0.0002*	44 (44)	47 (47)	0.3089
Secondary	55 (55)	48 (48)	0.3960	48 (48)	29 (29)	0.0087*
Grade 12	15 (15)	40 (40)	0.0001*	7 (7)	21 (21)	0.0072*
Higher education	4 (4)	6 (6)	0.7475	1 (1)	3 (3)	0.6212
Parity	3 (0–6)	3 (1–5)	0.1615	4 (0–8)	3 (1–7)	0.0009*
Pelvic organ prolapse	88 (88)	13 (13)	<0.0001*	92 (92)	19 (19)	<0.0001*
Urinary incontinence	68 (68)	12 (12)	<0.0001*	53 (53)	9 (9)	<0.0001*
Anal incontinence	34 (34)	6 (6)	<0.0001*	46 (46)	8 (8)	<0.0001*
Symptoms per patient						
0	0 (0)	81 (81)	<0.0001*	0 (0)	74 (74)	<0.0001*
1	35 (35)	9 (9)	<0.0001*	45 (45)	17 (17)	<0.0001*
2	40 (40)	8 (8)	<0.0001*	19 (19)	8 (8)	<0.0001*
3	25 (25)	2 (2)	<0.0001*	36 (36)	1 (1)	<0.0001*
PFDI-20 score (0–300)	131.8 ± 59.2	16.1 ± 22.3	<0.0001*	124.9 ± 55.2	26.8 ± 26.7	<0.0001*
POPDI-6 score (0–100)	52.1 ± 23.2	6.8 ± 12.7	<0.0001*	54.7 ± 21.8	13.2 ± 14.3	<0.0001*
Missing values	7 (1.16)	12 (2.00)	0.3554	6 (1.00)	2 (0.33)	0.2874
CRADI-8 score (0–100)	32.2 ± 25.2	4.4 ± 6.9	<0.0001*	31.7 ± 21.8	6.4 ± 10.2	1.0000
Missing values	8 (1.00)	8 (1.00)	1.0000	7 (0.87)	7 (0.87)	1.0000
UDI-6 score (0–100)	45.2 ± 27.8	4.8 ± 7.3	<0.0001*	36.0 ± 24.9	7.1 ± 9.7	<0.0001*
Missing values	9 (1.50)	4 (0.67)	0.2642	9 (1.50)	5 (0.83)	0.6608
PFIQ-7 score (0–300)	115.5 ± 61.6	13.2 ± 15.1	<0.0001*	116.1 ± 62.4	13.8 ± 16.5	<0.0001*
UIQ-7 score (0–100)	44.2 ± 29.1	4.7 ± 5.8	<0.0001*	39.8 ± 27.8	5.7 ± 6.6	<0.0001*
Missing values	10 (1.43)	13 (1.86)	0.6746	13 (1.86)	22 (3.14)	0.1691
CRAIQ-7 score (0–100)	25.9 ± 26.2	2.9 ± 4.9	<0.0001*	30.2 ± 26.5	3.3 ± 5.5	<0.0001*
Missing values	11 (1.57)	10 (1.43)	0.8261	11 (1.57)	18 (2.57)	0.2608
POPIQ-7 score (0–100)	45.4 ± 27.9	5.1 ± 7.9	<0.0001*	46.1 ± 25.6	4.8 ± 8.0	<0.0001*
Missing values	7 (1.00)	10 (1.43)	0.6266	9 (1.28)	12 (1.71)	0.6608

The data presented are means ± standard deviation, number (%), or medians (range)

Unpaired t-test for means, Mann–Whitney U test for medians and Fisher's and Chi-square tests for numerical data

*$p \leq 0.05$

6 and UIQ-7 scales than those without; and (3) women with symptoms of AI will score higher in the CRADI-8 and CRAIQ-7 scales than those without. Construct validity was considered adequate when at least 75% of these hypotheses were confirmed [21, 23].

Responsiveness in the context of a questionnaire can be defined as the outcome that can be achieved when the instrument is designed in such a way that it is cognizant of and responds appropriately to the clinical result experienced by the individual. Responsiveness therefore measures the ability of the questionnaires to detect change that occurs as a result of treatment. This was assessed during the third round in the study population. The relationship between the mean change in scores and the PGI-I dichotomized classification was summarized. The baseline scores were also compared with the third round scores using the paired t test, and in this case a p value <0.05 was considered to indicate a significant improvement. The standardized response mean (SRM) was used to evaluate whether the questionnaires were responsive to change at the group level in each of the languages [24, 25]. A number of statistical tests have been used to evaluate responsiveness. The SRM is however considered appropriate when evaluating responsiveness in a single group before and after an intervention, as was the case in this population [26, 27]. The difference in mean PFDI-20 and PFIQ-7 scores between those classifying themselves as "improved" compared to "not improved" according to the PGI-I were used to define a cut-off value indicative of improvement.

Table 2 Internal consistency (Cronbach's alpha)

Questionnaire	Afrikaans		Sesotho	
	Study group (n = 20)	Control group (n = 10)	Study group (n = 20)	Control group (n = 10)
PFDI-20	0.89	0.83	0.71	0.75
POPDI-6	0.80	0.91	0.43	0.43
CRADI-8	0.89	0.59	0.85	0.61
UDI-6	0.79	0.78	0.71	0.61
PFIQ-7	0.88	0.54	0.81	0.64
POPIQ-7	0.86	0.27	0.70	0.45
CRAIQ-7	0.97	0.40	0.79	0.31
UIQ-7	0.93	0.47	0.81	0.52

The remainder of the results are summarized categorically as frequencies and percentages. The chi-squared test and Fisher's exact test were used to evaluate univariate associations, and the unpaired t test for continuous variables. Statistical significance was set at a p value of <0.05.

Results

Out of a possible 330 women. the eligible study population consisted of 213 women (64.5%) for the Afrikaans and Sesotho questionnaire validation. Of these 213 women, 208 (97.6%; 104 in each language group) consented to participate in the three rounds and 200 (96.1%) completed the required questionnaires at the specified time points. Of 231 women approached for the control group, 206 (89.2%; 103 in each language group) consented to participate and complete data were available for 200 (97.1%) of these. The baseline characteristics are summarized in Table 1. The mean age of the study groups was higher ($p < 0.0001$) for both languages. There was a significant difference ($p < 0.0001$) in pelvic floor symptoms between the study and control groups for both languages and

the majority of participants in the study groups experienced more than one symptom.

Internal consistency

The PFDI-20 demonstrated good internal consistency with Cronbach's alpha values of 0.89 and 0.84 in the Afrikaans study and control groups and acceptable consistency in the Sesotho study group (0.71) and control group (0.75). The PFIQ-7 demonstrated good consistency (0.88) in the Afrikaans study group, but poor consistency (0.54) in the control group. A similar pattern was found among the Sesotho study (0.81) and control (0.64) participants. These results are presented in Table 2.

Reliability

The test–retest intraclass correlation for both participant groups in both languages ranged from 0.89 to 0.99, confirming very good reliability. These results are presented in Table 3. The correlation additionally confirmed that a telephone follow-up for these questionnaires was feasible and

Table 3 Test–retest reliability (intraclass correlation coefficient)

Questionnaire	Afrikaans		Sesotho	
	Study group (n = 100)	Control group (n = 100)	Study group (n = 100)	Control group (n = 100)
PFDI-20	0.97	0.99	0.97	0.98
POPDI-6	0.95	0.99	0.99	0.98
CRADI-8	0.99	0.99	0.98	0.97
UDI-6	0.99	0.98	0.98	0.95
PFIQ-7	0.99	0.93	0.98	0.97
POPIQ-7	0.99	0.97	0.98	0.96
CRAIQ-7	0.99	0.90	0.96	0.95
UIQ-7	0.99	0.89	0.98	0.92

Table 4 Limits of agreement

Questionnaire	Afrikaans				Sesotho			
	Study group (n = 100)		Control group (n = 100)		Study group (n = 100)		Control group (n = 100)	
	LOA	MME (%)	LOA	MME (%)	LOA	MME (%)	LOA	MME (%)
PFDI-20	−29.6 to 23.5	8.8	−4.1 to 3.2	1.2	−28.9 to 24.2	8.8	−8.6 to 6.6	5.1
POPDI-6	−14.5 to 14.0	14.2	−3.4 to 2.8	3.1	−5.7 to 5.4	5.5	−5.6 to 4.3	4.9
CRADI-8	−3.3 to 3.3	3.3	−1.8 to 1.5	1.6	−9.1 to 8.2	8.6	−4.8 to 4.3	4.5
UDI-6	−6.4 to 5.1	6.4	−2.7 to 2.7	2.7	−8.1 to 9.6	8.8	−5.9 to 5.8	5.8
PFIQ-7	−10.5 to 8.7	3.2	−12.9 to 8.2	3.5	−27.5 to 22.9	8.4	−7.6 to 8.3	2.6
POPIQ-7	−6.1 to 5.3	5.7	−4.2 to 3.4	3.8	−11.1 to 8.8	9.9	−3.9 to 4.3	4.1
CRAIQ-7	−5.2 to 4.8	5.0	−4.8 to 3.1	3.9	−11.6 to 8.9	10.2	−3.7 to 3.1	3.4
UIQ-7	−4.9 to 4.3	4.6	−5.4 to 4.2	4.8	−13.2 to 13.8	13.5	−4.7 to 5.6	5.1

LOA Limits of agreement, *MME* Overall magnitude of measurement error

reliable in the population studied. This represents a potentially significant cost saving in an environment with limited resources.

Measurement error

The LOA are summarized in Table 4 for the different language groups. The overall magnitude of the measurement error was calculated by dividing the LOA by the range of all measures. The measurement error was 8.8% for the PFDI-20 and 3.2–8.4% for the PFIQ-7 among the study participants in both groups. This confirmed low measurement error in both language groups.

Construct validity

Figures 1 and 2 illustrate the construct validity. The Afrikaans subgroup with POP reported higher mean scores for the POPDI-6 (56.7 ± 20.2) and POPIQ-7 scales (50.2 ± 26.0) than those without POP (18.7 ± 15.4 and 10.4 ± 13.7, respectively;

$p < 0001$). The Sesotho subgroup with POP reported higher mean scores for the POPDI-6 (57.8 ± 18.8) and POPIQ-7 scales (48.4 ± 24.9) than those without POP (18.7 ± 22.3 and 20.2 ± 18.4, respectively; $p < 0.0001$). The Afrikaans subgroup with UI reported higher mean scores for the UDI-6 (57.1 ± 24.0) and UIQ-7 scales (52.6 ± 27.7) than those without UI (19.8 ± 15.6 and 26.3 ± 23.8, respectively; $p < 0.0001$). The Sesotho subgroup with UI reported higher mean scores for the UDI-6 (54.6 ± 18.2) and UIQ-7 scales (51.7 ± 25.0) than those without UI (15.0 ± 10.7 and 26.3 ± 24.7, respectively; $p < 0.0001$). The Afrikaans subgroup with AI reported higher mean scores for the CRADI-6 (53.7 ± 24.5) and CRAIQ-7 scales (41.1 ± 29.9) than those without AI (21.1 ± 17.2 and 18.2 ± 20.2, respectively; $p < 0.0001$). The Sesotho subgroup with AI reported higher mean scores for the CRADI-6 (47.1 ± 18.7) and CRAIQ-7 scales (40.9 ± 25.8) than those without AI (18.5 ± 14.2 and 21.0 ± 23.7, respectively; $p < 0.0001$). These findings support the prespecified hypotheses for POP, UI and AI.

Fig. 1 PFDI-20 scale scores. *Afr* Afrikaans, *So* Sesotho, *POP* pelvic organ prolapse, *UI* urinary incontinence, *AI* anal incontinence; *boxes* mean and standard deviation, *whiskers* minimum and maximum scores (n = 100 for each variable)
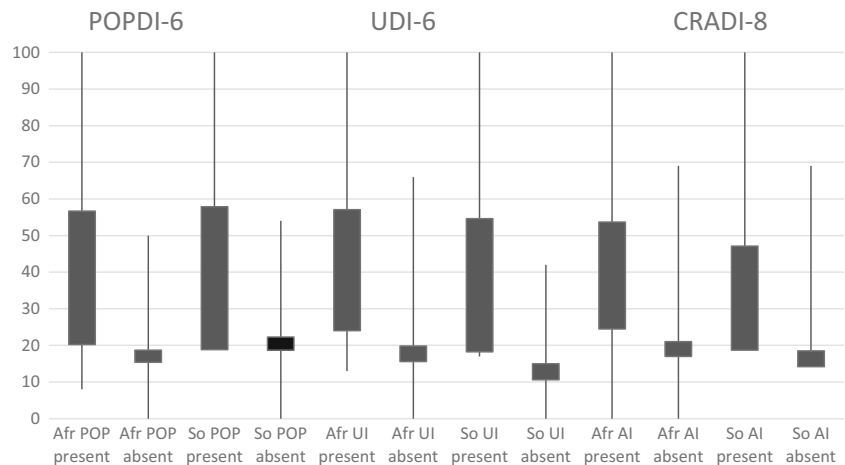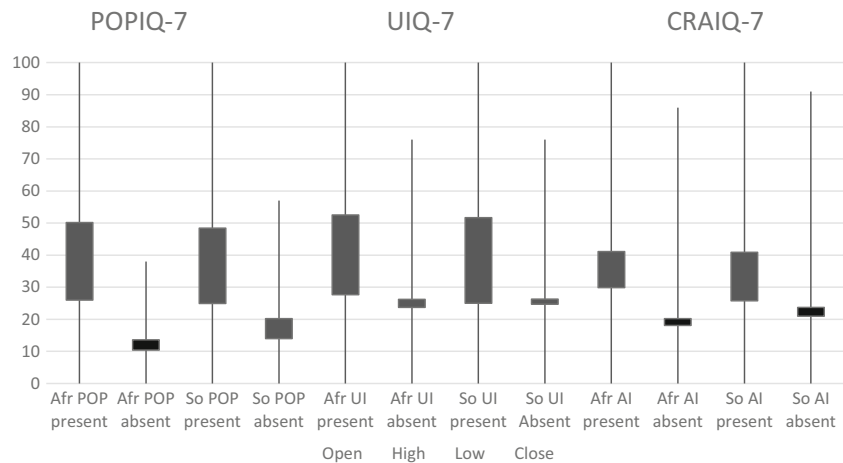
Fig. 2 PFIQ-7 scale scores. *Afr* Afrikaans, *So* Sesotho, *POP* pelvic organ prolapse, *UI* urinary incontinence, *AI* anal incontinence; *boxes* mean and standard deviation, *whiskers* minimum and maximum scores (*n* = 100 for each variable)

Responsiveness

The third completion round was after an average period of 5.2 months. In each language group, 100 participants completed the third round of questionnaires (Table 5). There was a significant improvement in scores across all domains, and this was evident in both language groups. The PGI-I results showed improvement in 79% of the Afrikaans group and in 81% of the Sesotho group. The cut-off score to distinguish women with improvement in symptoms form those without improvement could be determined by comparing the summary scores of the questionnaires. The cut-off scores in the Afrikaans group were <35 for the PFDI-20 and <24 for the PFIQ-7, and in the Sesotho group were <31 and <30, respectively.

Discussion

The objective of this study was to validate the PFDI-20 and PFIQ-7 in African women for the Afrikaans and Sesotho languages. The psychometric properties of

reliability, validity and responsiveness confirmed the validity of these questionnaires in this population and in this healthcare setting. The original PFDI and PFIQ showed very good internal consistency with alpha values of 0.88 and 0.97, respectively. This agreement was confirmed with the development of the short form versions of these questionnaires [2]. Overall Cronbach's alpha values between 0.71 and 0.89 for the PFDI-20 and between 0.81 and 0.88 for the PFIQ-7 were calculated in the two language groups. Similar alpha values have been reported for the Spanish, Turkish, Japanese, Chinese, Hebrew, Dutch and Swedish versions of the PFDI-20 and PFIQ-7 questionnaires [3–5, 7–10].

The questionnaires had to be altered after the initial pilot phase due to challenges that the participants described. It required a more illustrative layout and specific directives on completion in the first round and this was likely a reflection of the low level of education and awareness of PFD. Secondary school had not been completed by 81% of the Afrikaans and 92% of the Sesotho study participants, which is not unusual among African women [28]. This observation

Table 5 Responsiveness

Questionnaire	Afrikaans (*n* = 100)			Sesotho (*n* = 100)		
	Mean ± standard deviation	95% confidence interval	*p* value	Mean ± standard deviation	95% confidence interval	*p* value
PFDI-20	−94.5 ± 63.6	−107.1, −81.9	<0.0001*	−94.7 ± 63.4	−107.3, −82.2	<0.0001*
POPDI-6	−37.4 ± 25.3	−42.4, −32.4	<0.0001*	−44.2 ± 26.8	−49.5, −38.9	<0.0001*
CRADI-8	−22.3 ± 23.4	−26.9, −17.6	<0.0001*	−22.3 ± 19.8	−26.2, −18.3	<0.0001*
UDI-6	−32.3 ± 27.4	−37.7, −26.8	<0.0001*	−25.8 ± 26.3	−31.0, −20.6	<0.0001*
PFIQ-7	−87.6 ± 64.5	−100.4, −74.8	<0.0001*	−87.1 ± 66.6	−100.3, −73.9	<0.0001*
POPIQ-7	−33.6 ± 31.8	−39.9, −27.3	<0.0001*	−37.5 ± 28.6	−43.1, −31.8	<0.0001*
CRAIQ-7	−19.6 ± 24.0	−24.3, −14.8	<0.0001*	−21.2 ± 25.1	−26.2, −16.2	<0.0001*
UIQ-7	−32.6 ± 29.1	−38.4, −26.8	<0.0001*	−28.3 ± 29.2	−34.0, −22.5	<0.0001*

*$p < 0.05$

emphasizes the importance of public education on PFD among African women, something that is still greatly neglected.

Telephonic interviews as discussed by Geller et al. was used for the second round [17]. This was a potential concern with regard to its effect on reliability, but it was the only feasible method to evaluate test–retest reliability in this type of healthcare environment. The observed ICC values of 0.98–0.99 were not inferior to those documented in the validation of these questionnaires by self-administration [3, 5, 8]. The high correlation could be explained by the fact that the women had completed the questionnaire themselves the week before and were thus familiar with the questions being asked over the telephone. This might not have been the case had the questionnaires initially been administered by telephone and hence requires further research.

Responsiveness is an essential element in the assessment of any intervention. This is of particular importance in urogynecology where the primary aim of an intervention is mostly to improve the individual's QOL [29]. The PGI-I was utilized as the gold standard for evaluating responses to an intervention. The responsiveness was statistically significant across all scales ($p < 0.0001$) in this population and for both languages. A similar degree of responsiveness has been shown in the validation of these questionnaires in Turkish and Dutch populations [3, 9]. Responsiveness was however not evaluated in the majority of PFDI-20 and PFIQ-7 validation studies [4–6, 8, 10, 30]. The initial validation of these short-form questionnaires showed that the PFDI-20 is more responsive than the PFIQ-7 [2]. This was not observed in this population, and there is no apparent explanation for this finding at present.

There were some limitations to this study. The internal consistency was evaluated on a smaller sample of women and the Cronbach's alpha value might have been different if a larger sample had been analyzed. The second completion round was conducted by telephone and although the format was structured, it was not possible to ensure that the questionnaires were administered consistently in all interviews. Furthermore, the use of the telephone as a modality for administering the questionnaires was not specifically validated in this population prior to this study. The statistical results for test–retest reliability are, however, reassuring. The education levels of both language groups were generally limited and we had to verbally explain and demonstrate to the majority of patients the process of filling in and scoring symptom severity for the questionnaires in the first round. This was in addition to having redesigned the questionnaires after the pilot phase to permit ease of completion. This resulted in a more labor-intensive process and also emphasizes the importance of short-form questionnaires in an environment with limited human resources. Longer questionnaires would very likely not have been practically feasible in our environment and their use could have resulted in high rates of incomplete questionnaires,

which would have limited clinical value of the study. Pelvic floor symptom screening was not repeated in the third round and only the PGI-I was used to calculate responsiveness of the questionnaires. This lack of additional clinical information could have influenced the evaluation of responsiveness and consequently the determination of improvement or deterioration of function.

The strengths of this study were the use of adequate sample sizes based on the items evaluated in the questionnaires for both the study and the control groups. The methodological design of this study was according to the recommendations of the COSMIN initiative so that adequate evaluation of the different psychometric measurement properties of these questionnaires was ensured. This is particularly important for the validation of patient-reported health outcomes. The control group allowed insight into the score distribution in women who have not been identified with PFD. The results of the study reveal the importance of public education with regard to symptoms of PFD. This is particularly relevant to the Sesotho speaking women, among whom an unexpectedly high number (26%) were identified with symptoms of PFD. This is in contrast to an expected low prevalence of this condition amongst ethnic black African women based on observations in the referral urogynecology unit. The role of the questionnaires as a routine screening tool should therefore be further explored in this population to allow the identification of PFD and appropriate referral and management. It also points out the lack of robust epidemiological data that exists for PFD among African women.

Compliance with ethical standards

Conflicts of interest None.

References

1. Barber MD. Questionnaires for women with pelvic floor disorders. Int Urogynecol J. 2007;18(4):461–465.
2. Barber MD, Walters MD, Bump RC. Short forms of two condition-specific quality-of-life questionnaires for women with pelvic floor disorders (PFDI-20 and PFIQ-7). Am J Obstet Gynecol. 2005;193(1):103–113.
3. Utomo E, Blok BF, Steensma AB, Korfage IJ. Validation of the Pelvic Floor Distress Inventory (PFDI-20) and Pelvic Floor Impact Questionnaire (PFIQ-7) in a Dutch population. Int Urogynecol J. 2014;25(4):531–544.
4. Treszezamsky AD, Karp D, Dick-Biascoechea M, Ehsani N, Dancz C, Montoya TI, et al. Spanish translation and validation of four short pelvic floor disorders questionnaires. Int Urogynecol J. 2013;24(4):655–670.

5. 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. 2013;24(6): 1039–1046.

6. 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(8):1123–1127.

7. Zhu L, Yu S, Xu T, Yang X, Lu Y, Li B, et al. Chinese validation of the Pelvic Floor Impact Questionnaire Short Form. Menopause. 2011;18(9):1030–1033.

8. 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(5):483–487.

9. Kaplan PB, Sut N, Sut HK. Validation, cultural adaptation and responsiveness of two pelvic-floor-specific quality-of-life question-naires, PFDI-20 and PFIQ-7, in a Turkish population. Eur J Obstet Gynecol Reprod Biol. 2012;162(2):229–233.

10. 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(6): 329–331.

11. El-Azab AS, Abd-Elsayed AA, Imam HMK. Patient reported and anatomical outcomes after surgery for pelvic organ prolapse. Neurourol Urodyn. 2009;28(3):219–224.

12. de Tayrac R, Deval B, Fernandez H, Marès P. 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). J Gynecol Obstet Biol Reprod. 2007;36(8):738–748.

13. Eremenco SL, Cella D, Arnold BJ. A comprehensive method for the translation and cross-cultural validation of health status ques-tionnaires. Eval Health Prof. 2005;28(2):212–232.

14. Acquadro C, Conway K, Hareendran A, Aaronson N. Literature review of methods to translate health-related quality of life ques-tionnaires for use in multinational clinical trials. Value Health. 2008;11(3):509–521.

15. Streiner D, Norman G, Cairney J. Health measurement scales: a practical guide to their development and use. New York: Oxford University Press; 2014. p. 7–19.

16. Haylen BT, Maher CF, Barber MD, Camargo S, Dandolu V, Digesu A, et al. An International Urogynecological Association (IUGA)/ International Continence Society (ICS) joint report on the terminol-ogy for female pelvic organ prolapse (POP). Int Urogynecol J. 2016;27(4):655–684.

17. Geller EJ, Barbee ER, Wu JM, Loomis MJ, Visco AG. Validation of telephone administration of 2 condition-specific quality-of-life questionnaires. Am J Obstet Gynecol. 2007;197(6):632.e1–632.e4.

18. Srikrishna S, Robinson D, Cardozo L. Validation of the Patient Global Impression of Improvement (PGI-I) for urogenital prolapse. Int Urogynecol J. 2010;21(5):523–528.

19. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the method-ological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19(4):539–549.

20. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ. 2011;2:53–55.

21. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res. 2013;22(8): 1889–1905.

22. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consen-sus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737–745.

23. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42.

24. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PMM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Qual Life Res. 2003;12(4):349–362.

25. Zou GY. Quantifying responsiveness of quality of life measures without an external criterion. Qual Life Res. 2005;14(6):1545–1552.

26. Eurich DT, Johnson JA, Reid KJ, Spertus JA. Assessing respon-siveness of generic and specific health related quality of life mea-sures in heart failure. Health Qual Life Outcomes. 2006;4:89.

27. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. J Clin Epidemiol. 2001;54(12):1204–1217.

28. Kc S, Lutz W. The human core of the shared socioeconomic path-ways: population scenarios by age, sex and level of education for all countries to 2100. Glob Environ Change. 2017;42:181–192.

29. Aponte MM, Rosenblum N. Repair of pelvic organ prolapse: what is the goal? Curr Urol Rep. 2014;15(2):385.

30. Zhu L, Yu S, Xu T, Yang X, Lu Y, Lang J. Validation of the Chinese version of the Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire short form (PISQ-12). Int J Gynecol Obstet. 2012;116(2):117–119.