Published in final edited form as:

Int Urogynecol J Pelvic Floor Dysfunct. 2009 June ; 20(6): 623–639. doi:10.1007/s00192-008-0792-7.

Validation of Spanish Versions of the Pelvic Floor Distress Inventory and Pelvic Floor Impact Questionnaire : A Multicenter Validation Randomized Study

Tola B. Omotosho, MD,

Division of Urogynecology, Department of Obstetrics and Gynecology, University of New Mexico Health Sciences Center, 1 University of New Mexico, MSC 10 5580, Albuquerque, NM 87131-0001, tomosho@salud.unm.edu, Phone: 505-272-9702, Fax: 505-272-1336

Anne Hardart, MD,

Department of Obstetrics and Gynecology, St. Luke's Roosevelt Hospital Center, 1000 Tenth Avenue, 10th Floor, New York, NY 10019

Rebecca G. Rogers, MD,

Division of Urogynecology, Department of Obstetrics and Gynecology, University of New Mexico Health Sciences Center, 1 University of New Mexico, MSC 10 5580, Albuquerque, NM 87131-0001

Joseph I. Schaffer, MD,

Department of Obstetrics and Gynecology, University of Texas, Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd., Dallas, TX 75390

William H. Kobak, MD, and

Department of Obstetrics and Gynecology, University of Illinois Chicago Medical Center, 820 South Wood St., Chicago, IL 60612

Audrey A. Romero, MD

Rubino Obstetrics and Gynecology, 101 Old Short Hills Road, Suite 410, West Orange, NJ 07052

Abstract

Introduction and Hypothesis—The purpose of this study is to validate Spanish versions of the Pelvic Floor Distress Inventory (PFDI) and Pelvic Floor Impact Questionnaire (PFIQ).

Methods—Spanish versions were developed using back translation. Validation was performed by randomizing bilingual women to complete the Spanish or English versions of the questionnaires first. Weighted kappa statistics assessed agreement for individual questions; interclass correlation coefficients (ICC) compared primary and subscale scores. Cronbach's alpha assessed internal consistency of Spanish versions. To detect a 2.7 point difference in scores with 80% power and alpha of 0.05, 44 bilingual subjects were required.

Results—Individual questions showed good to excellent agreement (kappa .6) for all but 8 questions on the PFIQ. ICC's of primary and subscale scores for both questionnaires showed excellent agreement. (All ICC .79). All Cronbach's alpha's values were excellent (.84) for the primary scales of both questionnaires.

Conclusions—Valid and reliable Spanish versions of the PFIQ and PFDI have been developed.

Keywords

Fecal incontinence; Pelvic organ prolapse; Quality of life; Spanish translation; Urinary incontinence

Introduction

The Pelvic Floor Distress Inventory (PFDI) and Pelvic Floor Impact Questionnaire (PFIQ) are validated and reliable symptom severity and condition-specific quality of life questionnaires[1]. These companion questionnaires are widely used by both researchers and clinicians who treat women with pelvic floor disorders and provide an accurate measure of pelvic organ prolapse, urinary and colorectal-anal symptom severity, bother and quality of life changes, and have proven responsiveness to change[2].

Both questionnaires were developed in English, and therefore limit the participation of other cultural and language groups, including Hispanics. Hispanics comprise one of the fastest growing populations in the United States (U.S.). In 2000, the U.S. Census Bureau estimated that 32.8 million Hispanics reside in the United States and that by the year 2050 one in four women will identify themselves as Latina[3, 4]. Although many Hispanics living in the U.S. are bilingual, Spanish often remains their preferred language and for recent immigrants, Spanish may be their only language. We aimed to create valid, reliable and culturally sensitive Spanish versions of the PFDI and PFIQ.

Materials and Methods

The PFDI is a 46-item condition-specific symptom questionnaire that consists of 3 primary scales; the Urinary Distress Inventory (UDI scores range 0–300), Pelvic Organ Prolapse Distress Inventory (POPDI scores range 0–300) and Colo-rectal-Anal Distress Inventory (CRADI scores range 0–400). Each of the primary scales is further divided into subscales. The UDI has Obstructive, Irritative/Discomfort and Stress subscales, the POPDI has General, Anterior and Posterior subscales, and the CRADI has Obstructive, Incontinence, Pain/Irritation, and Rectal Prolapse subscales[1].

The PFIQ, a condition-specific measure of functional status, has 93 questions and contains three primary scales; Urinary Impact Questionnaire (UIQ), the Pelvic Organ Prolapse Impact (POPIQ) and the Colo-Rectal-Anal Impact (CRAIQ). Each of the primary scales is further subdivided into four subscales including Travel, Social, Emotional and Physical Activity. Each primary scale of the PFIQ has scores which range from 0–400. Higher scores on both the PFIQ and the PFDI reflect greater pelvic floor dysfunction[1].

Validation and reliability testing of the questionnaires was conducted in two phases. In the first phase, a translation-back translation method was utilized to generate a single Spanish translation of each questionnaire reflective of the content and meaning of the original English questionnaires[5,6]. English versions of the PFIQ and PFDI were translated into Spanish by three independent researchers fluent in Spanish, English and medical terminology. Three other independent translators then translated the three Spanish versions back to three English versions for each questionnaire. Translators were bilingual medical personnel. The newly created English versions were then compared to the original English versions to ensure comparability of language and similarly of interpretation. *Comparability of language* refers to the formal similarity of words, phrases and sentences. *Similarity of interpretability* refers to the degree to which the two versions engender the same response even if the exact wording varies. A scoring system was utilized in order to determine which of the three Spanish translations would be used[5]. (Table 1.) Seven investigators compared each English back-translated version of each question from the PFIQ and PFDI to the original English versions with an item-by-item rating using Likert scales ranging from 1 (extremely comparable / extremely similar) to 7 (not at all comparable /not at all similar).

The final back- translated versions of each questionnaire were reviewed by bilingual translators for comprehension and to ensure cultural sensitivity and minor corrections were then made to produce the two final Spanish versions.

In phase two of the study, bilingual women with pelvic floor disorders who could read both English and Spanish were recruited from the general gynecology and urogynecology clinics from the University of New Mexico Health Sciences Center, St. Luke's Roosevelt Hospital Center in New York, University of Texas Southwestern Medical Center and University of Illinois at Chicago. Four sites throughout the country were chosen to reflect differences in dialect that may exist geographically.

Institutional review board approval was obtained for each facility that participated in the study. All women provided written informed consent before completing questionnaires. Inclusion criteria included all women over the age of 18 who stated that they were able to read both English and Spanish, and provided an affirmative answer to the question "Do you have symptoms of urine leaking, pressure in the vagina, or the feeling that things are falling out of the vagina?" All participants provided personal information including age, parity, ethnicity, race, their primary language spoken at home and highest level of education.

Women were randomized to complete either the English or Spanish versions first followed by the alternative language versions. (Figure 1) Randomization was assigned by using a random number table in blocks of four and was assigned by the consenting investigator. Assignments were kept in sealed, sequentially numbered opaque envelopes and were opened after written consent was obtained. In order to establish test-retest reliability, all subjects were asked and 22 women agreed to repeat the Spanish versions within 1–2 weeks.

Agreement between versions (English-Spanish and Spanish-Spanish) was assessed by weighted kappa statistics for individual questions. Kappa scores reflect the following agreement: for scores of < 0.2 poor, 0.21–0.4 fair, 0.41 to 0.6 moderate, 0.61–0.8 good and > 0.81 near perfect[2]. Interclass correlation coefficients (ICC) were calculated to compare primary and subscale scores. Interclass correlation coefficients of 0.41–0.6 were considered moderate agreement, 0.61–0.8 good agreement, and 0.81 or more excellent agreement[7]. Cronbach's alpha coefficient was used to assess the internal consistency of the new Spanish versions with 0.7–0.8 moderate, 0.81–0.9 good, and > 0.9 excellent internal consistency[8]. Differences in mean scores were assessed with Wilcoxin signed rank test. Descriptive statistics were used to report patient characteristics. In order to detect a 2.7 point difference in scores with 80% power and alpha of 0.05, 44 bilingual subjects were required[7].

Results

For Phase 1 of the validation, both the PFDI and PFIQ individual questions had comparability of language scores ranging from 1.00–3.50 (PFDI) and 1.00–2.50 (PFIQ), indicating moderate to excellent comparability. Similarity of interpretability scores for both questionnaires ranged from 1.00– 2.83, indicating excellent interpretability[5].

For Phase 2 of the study, 44 patients were consented and completed both English and Spanish versions of the PFDI and PFIQ. Twenty two patients repeated the Spanish versions of both questionnaires. The mean age of patients was 49 +/- 13 years. Study subjects were representative of a number of different ethnic backgrounds and approximately half considered both Spanish and English their primary language.(Table 2).

The PFDI demonstrated good to nearly perfect agreement between Spanish and English versions with weighted kappas 0.65 for all 46 individual questions. Interclass correlation coefficients for the UDI, POPDI and the CRADI (primary scales of the PFDI) demonstrated

excellent correlation with all ICC scores = 0.95, as illustrated on the scatter plots in Figure 2. Excellent agreement was also noted for the three subscales of the UDI and POPDI where all ICC's were = 0.9. For three of the four subscales of the CRADI, the ICC's were also = 0.9. The rectal prolapse subscale of the CRADI was lower, but still excellent at 0.87. For the UDI, POPDI and CRADI primary scales there were no significant differences between the Spanish and English mean scores (All $P=0.17 - 0.96$). Patients did have significantly higher English than Spanish scores on the incontinence subscale of the CRADI (19 ± 25 English versus 16 ± 24 Spanish, $P=0.04$, Wilcoxin signed rank test).

The PFIQ demonstrated good to nearly perfect agreement between Spanish and English versions with weighted kappas = 0.60 for all but eight individual questions; three questions in the CRAIQ scale and five questions in the POPIQ scale had kappas which ranged from 0.28–0.56. The PFIQ had excellent interclass correlations with all three primary scales (UIQ, POPIQ and CRAIQ) = 0.84. (Figure 3) For all 12 subscale comparisons of the PFIQ, ICC's were good to excellent at = 0.79. No differences were noted in mean scores between the Spanish and English versions of PFIQ.

Equal numbers of women in both groups completed a second Spanish questionnaire. For test-retest reliability of the Spanish versions of both questionnaires, interclass correlations for the primary scales of the PFIQ ranged from 0.63 to 0.84, and the interclass correlations for primary scales of the PFDI ranged from 0.71 to 0.86, indicating good to excellent agreement. Cronbach's alpha for the three primary scales of the PFIQ were all = 0.93 and for the three primary scales of the PFDI were = 0.84. In addition, the individual items all had high correlations within their subscales, indicating that the Spanish version subscales are well supported by the individual questions. This is consistent with the items in the Spanish version supporting the factors defined in the original English version.

We performed a subset analysis of the questions utilized in the short versions of both questionnaires (PFDI-20 and PFIQ-7). Weighted kappas for questions in both the PFDI-20 and PFIQ-7 were all > 0.60 except for one question in the POPIQ scale of the PFIQ-7. PFDI-20 and PFIQ-7 summary and scale scores demonstrated good to excellent agreement between English and Spanish versions with all ICC's = 0.77. Mean Spanish and English summary and scale scores did not differ between questionnaires with the exception of scores for the UDI-6. The mean score of the Spanish version of the UDI-6 was higher than the English version (41.8 ± 25.9 vs 38.4 ± 29.2 $P= 0.04$).

Discussion

Pelvic floor disorders are common and, although not life threatening, affect the quality of life of women who have these conditions. The PFDI and PFIQ are reliable and valid condition-specific questionnaires that measure symptom bother and quality of life in women with pelvic organ prolapse, urinary, and colo-rectal and anal problems [1, 2]. In addition, both questionnaires have proven sensitivity to change[9, 10]. Previous studies utilizing the PFDI and PFIQ have focused on Caucasian women[1, 2, 9]. The limited availability of valid versions of questionnaires in other languages may contribute to the under-representation of different ethnic and racial groups in pelvic floor studies.

Standards for translation of questionnaires do not exist. Back translation methodologies as well as further validation in bilingual women are both accepted means of validating alternate language versions of questionnaires[5, 6]. Others have purported that simultaneous development of questionnaires in multiple languages may be a more valid approach for developing alternative language questionnaires, however, this method can be expensive and laborious[5].

Our results are similar to other Spanish translations of the PFIQ and PFDI with good to excellent ICC correlations for the primary scales and subscales[11]. For the PFDI, unlike previous investigators, we had good to excellent agreement for all individual questions in the questionnaire[11]. We also noted that the incontinence subscale of the CRADI had lower scores for the Spanish when compared to the English version; while others found that CRADI mean total scores had lower Spanish than English scores. Although we have no clear explanation for these findings, perhaps poorer correlation for bowel questions may be secondary to patient's reticence in reporting bowel complaints, while urinary incontinence and pelvic organ prolapse may not hold the same social stigma[11].

Agreement for individual questions on the PFIQ were good to excellent with the exception of 8 questions in the bowel and prolapse primary scales; while the previous study found poor agreement in four questions[11]. We also noted kappas were poorer in questions that had a lack of diversity in responses that is necessary for kappa to provide a stable estimate[11].

A limitation of this study is the use of bilingual patients since the translated questionnaires are intended for use by monolingual patients. Bilingual women may not be fully representative of monolingual patients. Because each subject served as her own control recall bias was inevitable and may have contributed to a greater agreement between English and Spanish versions than what actually exists. We minimized this bias by randomizing women to complete alternate Spanish or English versions. In addition, unlike previous studies, we quantified our back translation methodology to provide rigorous Spanish versions for bilingual testing[11].

We diversified our population by enrolling subjects at 4 different locations to reflect the regional and cultural differences of Spanish speaking women in the U.S. Our Spanish versions are not validated in all social contexts in which Spanish is spoken.

Another limitation of this study was that construct validity was not established because we did not use objective assessments of type or symptom severity of the pelvic floor condition[1]. However, all women who participated in the study presented to our clinics with complaints of pelvic floor dysfunction and were not medical personnel who may have greater knowledge of pelvic floor dysfunction than clinic patients.

Unlike previous studies we also investigated the test-retest reliability of the Spanish versions of the questionnaires by having 22 women complete questionnaires after at least a one week period, with good to excellent reliability.

Although a preliminary power analysis was performed, a post hoc analysis was also performed based on the number of participants in the study. Using the variability obtained in the study for the paired comparison of the English and Spanish versions of the PFDI, the detectable differences for the primary scales of the PFDI (UDI, POPDI, CRADI) are 3.2, 8.6, 8.8 points respectively with 80% power and $\alpha = 0.05$. These differences are 1%, 3%, and 2% of the maximum score for each scale respectively, and illustrate that we were able to detect small differences between Spanish and English versions. In addition a post hoc power analysis of our sample size of 22 women for test retest reliability was adequate to detect a 14% difference in the 6 scales of the two questionnaires with 80% power and α of 0.05.

Short versions of the PFDI and PFIQ are valid, reliable and sensitive to change[12]. We performed a subset analysis of the questions utilized in the short versions of the PFDI and PFIQ and confirmed good to excellent agreement for the individual questions as well as good to excellent interclass correlations for primary scales and summary scores.

Our aim was to create valid and reliable Spanish versions of the PFDI and PFIQ, using a rigorous back translation methodology, followed by further validation in a population of bilingual women. The versions of the questionnaires developed here perform similarly to their English versions, are valid and reliable, and can be used to evaluate the impact of pelvic floor disorders in Spanish speaking women. Use of Spanish questionnaires allows representation of Hispanic women in research and valid assessment of outcomes of their clinical care.

Acknowledgments

This work supported in part by DHHS/NIH/NCRR/GCRC Grant # 5M01 RR 00997 and R01 HD049819-02; Reprints are not available.

References

1. Barber MD, Kuchibhatla MN, Pieper CF, Bump RC. Psychometric evaluation of comprehensive condition-specific quality of life instruments for women with pelvic floor disorders. *Am J Obstet Gynecol.* 2001; 185(6):1388–1395.
2. Wren PA, Janz NA, Brubaker L, Fitzgerald MP, Weber AM, LaPorte FB, Wei JT. Reliability of health-related quality of life measures 1 year after surgical procedures for pelvic floor disorders. *Am J Obstet Gynecol.* 2005; 192:780–788. [PubMed: 15746672]
3. Therrien, M.; Ramirez, RR. Current Population Reports, P20-535. Washington: US Census Bureau; The Hispanic population in the United States: March 200.
4. US Census Bureau. US Interim Projections by, Age, Sex, Race and Hispanic Origin, Internet Release Date: March 18, 2004.
5. Sperber AD, DeVellis RF, Boehlecke B. Cross-cultural translation: Methodology and validation. *J Cross-Cult Psychol.* 1994; 25:501–524.
6. Berry, JW. Introduction to Methodology. In: Triandis, H.; Berry, JW., editors. *Handbook of Cross-Cultural Psychology*. Vol. Volume 2. Boston: Allyn and Bacon; 1980. p. 1-34.
7. Romero AA, Hardart A, Kobak W, Qualls C, Rogers R. Validation of a Spanish Version of the Pelvic Organ Prolapse Incontinence Sexual Questionnaire. *Obstet and Gynecol.* 2003; 102:1000–1005.
8. Fayes, PM.; Machin, D. *Quality of Life: assessment, analysis and interpretation*. Chichester: John Wiley & Sons; 2000.
9. Barber MD, Walters MD, Cundiff GW. PESSRI Trial Group. 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 May; 194(5):1492–1498. [PubMed: 16647933]
10. Barber MD. Questionnaires for Women with Pelvic Floor Disorders. *Int. Urogynecol J.* 2007; 18:461–465.
11. Young AE, Fine PM, McCrery R, Wren PA, Richter HE, Brubaker L, et al. Spanish Language Translation of Pelvic Floor Disorders Instruments. *Int Urogynecol J Pelvic Floor Dysfunct.* 2007 Jun 19. [Epub ahead of print].
12. 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; 193:103–113. [PubMed: 16021067]

Brief Summary

We validated and performed reliability testing for Spanish versions of the Pelvic Floor Distress Inventory and Pelvic Floor Impact Questionnaires using a rigorous back translation methodology, followed by further validation in a population of bilingual women. The Spanish versions of these quality of life questionnaires performed similarly to their English versions, are valid and reliable, and can be used to evaluate the impact of pelvic floor disorders in Spanish speaking women.

Figure 1.
Randomization of Bilingual Patients

Figure 2.

Scatter Plots of the PFDI Primary Scores

2a. UDI (Urinary Distress Inventory)

21b. POPDI (Pelvic Organ Prolapse Distress Inventory),

2c. CRADI (Colo-Rectal-Anal Distress Inventory)

Figure 3.

Scatter Plots of the PFIQ Primary Scores

3a. UDI (Urinary Distress Inventory)

3b. CRAIQ (Colo-Rectal-Anal Impact Questionnaire)

3c. POPIQ (Pelvic Organ Prolapse Impact Questionnaire)

Table 1

Comparability/ Interpretability rating sheet

Please Circle the response which most closely represents how you would rate the following pairs of items in terms of:

- A.** Comparability of language (how comparable is the formal wording?) and
B. Similarity of interpretation (would the paired items be interpreted similarly, even if the wording is different ?) Please circle only one response for A?) and one response for B) for each pair of items.

Original English version	Back translated English version	<u>A). COMPARABILITY OF LANGUAGE</u>						
		Extremely Comparable			Moderately Comparable		Not At All Comparable	
		1	2	3	4	5	6	7
		<u>B). SIMILARITY OF INTERPRETATION</u>						
		Extremely Similar			Moderately Similar		Not At All Similar	
		1	2	3	4	5	6	7

Table 2

Patient Characteristics

Characteristic	N=44
Age	49 ± 13
Ethnicity (%)	
Hispanic/Latina	17 (38)
Mexican	11 (26)
Puerto Rican	9 (21)
Primary Language (%)	
English	13 (30)
Spanish	9 (20)
Both	22 (50)
Education Level (%)	
Grade School	13 (30)
High School	8 (18)
College	19 (44)
Professional/Graduate School	4 (8)