



Pelvic floor distress inventory (PFDI)—systematic review of measurement properties

Guilherme Tavares de Arruda¹ · Thaise dos Santos Henrique¹ · Janeisa Franck Virtuoso¹

Received: 4 December 2020 / Accepted: 22 February 2021 / Published online: 12 March 2021
© The International Urogynecological Association 2021

Abstract

Introduction and hypothesis The Pelvic Floor Distress Inventory (PFDI) and PFDI-20 have been translated and validated into several languages with different measurement property values and are recommended by the International Consultation on Incontinence (ICI) as grade A for assessing pelvic floor dysfunction. Thus, the aim of the current study was to investigate the measurement properties of the PFDI and PFDI-20.

Methods Systematic review conducted in August 2020 through a search performed in PubMed, SCOPUS, WoS, ScienceDirect, CINAHL, and Google Scholar for studies that evaluated the measurement properties of the PFDI and PFDI-20. The data were analyzed according to the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN).

Results Initially, 2857 studies were found, and 7 studies on PFDI and 25 on PFDI-20 were analyzed. The PFDI presented high quality of evidence for hypothesis testing, moderate for test-retest reliability and responsiveness, and very low quality of evidence for content validity. The PFDI-20 presented high quality of evidence for criterion validity, hypothesis testing, and responsiveness, moderate quality for test-retest reliability and measurement errors, and very low quality of evidence for content validity. It was not possible to rate the quality of evidence of the internal consistency of the PFDI and PFDI-20. No studies assessed the cross-cultural validity.

Conclusion Only the hypothesis testing presented high quality of evidence for the PFDI. Criterion validity, hypothesis testing, and responsiveness presented a high quality of evidence for the PFDI-20. Due to the high degree of recommendation of the PFDI and PFDI-20 given by the ICI, further studies are needed to reevaluate all the measurement properties of these instruments.

Keywords Validation study · Patient-reported outcome measures · Pelvic floor disorders · Women

Introduction

Pelvic floor dysfunction (PFD) is a broad term used to describe specific clinical conditions of the pelvic floor (PF), such as urinary incontinence (UI), pelvic organ prolapse (POP), sexual dysfunction, and functional defecation disorders [1]. The prevalence of PFD varies widely in the literature [2, 3]; however, it is known to be more frequent in females [1] and can negatively affect the woman's quality of life, causing bother in the PF, depression, and a sedentary lifestyle [4–6].

For the assessment and quantification of the intensity and severity of symptoms, patient-reported outcome measures (PROMs) are often used [19]. Among the existing PROMs to assess the symptoms of PFD in women, the Pelvic Floor Distress Inventory (PFDI) is frequently used [7, 8] and recommended as grade A by the International Consultation on Incontinence (ICI) [9]. The ICI classifies a PROM as grade A if the Committee found “published data indicating that the questionnaire is valid, reliable, and responsive to change following standard psychometric testing,” without taking into account the details of how the study was conducted and the values for the measurement properties [9]. In addition, the 46 items of the PFDI [10] were abbreviated in a short version with 20 items, the PFDI-20 [6]. The purpose of both forms is to assess the distress of the symptoms of PFD through items divided into three subscales—Pelvic Organ Prolapse Distress Inventory (POPDI), Colorectal-Anal Distress Inventory (CRADI), and Urinary Distress Inventory (UDI)—which assess POP, anorectal, and urinary symptoms, respectively [6, 10].

✉ Guilherme Tavares de Arruda
gui_tavares007@hotmail.com

¹ Programa de Pós-Graduação em Ciências das Reabilitação, Universidade Federal de Santa Catarina, Rodovia Governador Jorge Lacerda, 3201, Jardim das Avenidas, SC-CEP: 88.906-072, Araranguá, Brazil

The PFDI and PFDI-20 have been translated and validated into several languages [11–15]. Both versions have also been used in a large number of studies [7, 8] and in clinical practice to identify the symptoms of PFD and the intensity of distress caused by these symptoms. However, it is necessary to evaluate the measurement properties of the two versions to determine whether they are suitable for the clinical and scientific context. Furthermore, the classification of a PROM proposed by the ICI is simplistic and does not analyze the study methods or values of the measurement properties. To our knowledge, no systematic reviews of the measurement properties of the PFDI or PFDI-20 have been published. Thus, the aim of the current study was to investigate whether the measurement properties of the PFDI and PFDI-20 were confirmed in previous studies and are suitable as criteria for good measurement properties, according to the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN).

Materials and methods

This systematic review (PROSPERO ID CRD42020157083) followed the guidelines of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) [16]. The search for studies was carried out by two independent reviewers in the PubMed, SCOPUS, Web of Science, ScienceDirect, and CINAHL databases. We also conducted a search in Google Scholar to identify studies not indexed in databases. Instrument names combined with their acronyms were used in a search filter for measurement properties adapted for each database (Appendix A) and recommended by COSMIN [17]. The search was carried out in August 2020, with articles evaluated without language restrictions and publication date. The reference lists of all included studies were verified to identify studies not found in the database searches.

In this systematic review, we followed the COSMIN methodology, which is based on the definitions of measurement properties: content validity (degree to which the content of a PROM is an adequate reflection of the construct to be measured); structural validity (degree to which the scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured); criterion validity (degree to which a PROM score is an adequate reflection of a “gold standard”); cross-cultural validity (degree to which performance of items in a translated or culturally adapted PROM is an adequate reflection of the performance of items in the original PROM version); hypothesis testing for construct validity (degree to which the scores of a PROM are consistent with hypotheses based on the assumption that the PROM validly measures the construct to be measured), divided into convergent validity (correlations must be related to other measures), discriminative validity or comparison between known groups (comparison between groups that are different in the construct), and

divergent validity (correlations should not be related to other measures); reliability (degree to which a measure is free from measurement errors: test-retest, intra- and inter-evaluators); internal consistency (degree of interrelation between items); measurement errors (systematic and random error of a patient’s score that is not attributed to real changes in the construct to be measured); responsiveness (PROM ability to detect changes over time in the construct to be measured) [18].

Original studies that reported one or more measurement properties of the PFDI and PFDI-20 and their subscales were included. Studies that were available only as event summaries or proceedings and that assessed the reliability of PFDI and PFDI-20 paper versions versus telephone/computer administration were excluded. According to the COSMIN guideline, these types of reliability studies should be ignored as they do not provide enough information about this PROM measurement property [19]. Titles, abstracts, and full texts were reviewed by two independent researchers (G.T.A., T.S.H.). In case of disagreements between the researchers, a third (J.F.V.) was consulted.

The extraction of data on the measurement properties and the evaluation of the methodological quality of the studies were carried out by two independent researchers (G.T.A., T.S.H.). The type of measurement property, results of the measurement property, objective of the study, sample size, type of PFD included, country where the study was carried out, and version of the PFDI used were extracted. For analysis, recording, and storage of results, the reference manager EndNote Web and a spreadsheet in the Microsoft Excel program were used.

Data analysis was performed in three stages. In the first step, the methodological quality of the included studies was rated using the 4-point scoring system on the COSMIN checklist [19]. The methodological aspects of the studies and the statistical methods used for each measurement property were rated as “very good,” “adequate,” “doubtful,” or “inadequate.” To rate the methodological quality of the studies according to the sample number, the following criteria were used as a standard: the study was rated as “very good” if the sample number was seven times the number of items and ≥ 100 individuals; “adequate” if the sample number was five times the number of items and ≥ 100 individuals, or six times the number of items, but < 100 individuals; “doubtful” if the sample number was equal to five times the number of items and < 100 individuals; “inappropriate” if the sample number was less than five times the number of items. At the end of this stage, the methodological quality of each measured property was summarized by study. For content validity, separate aspects were evaluated regarding relevance, comprehensiveness, and comprehensibility in the three stages [19].

In the second step, each measurement property in each study was rated as “sufficient,” “insufficient,” or “indeterminate,” according to criteria for good measurement properties

of the COSMIN guideline for systematic reviews of PROMs [19]. These ratings were summarized qualitatively to determine the overall rating of the measurement property for the PFDI and PFDI-20. If all studies indicated a “sufficient,” “insufficient,” or “indeterminate” rating for a specific measurement property, the overall rating of that measurement property was rated accordingly. If there were inconsistencies between studies on differences in methodological quality, populations, etc., explanations were provided. The explanations were discussed until a consensus was reached on the overall rating of the measurement property. If no explanation was found, the overall rating was considered “inconsistent.”

In the third step, the overall rating of evidence by measurement property was complemented by a level of quality of evidence, using the modified Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach of the COSMIN methodology [19]. This approach takes into account the quality of the study, the indirectness of the evidence, the inconsistency of the results, and the imprecision of the evidence (number of studies and sample size). The overall quality of the evidence was rated as “high,” “moderate,” “low,” or “very low.” This classification considers “high” quality of evidence when there is a lot of confidence that “the true measurement property lies close to that of the estimate of the measurement property,” “moderate” quality when “the true measurement property is likely to be close to the estimate of the measurement property, but there is a possibility that it is substantially different,” “low” quality when “the true measurement property may be substantially different from the estimate of the measurement property,” and “very low” quality of evidence when “the true measurement property is likely to be substantially different from the estimate of the measurement property” [19]. The measurement properties that were rated as “indeterminate” in the previous step did not receive a rating in this third step, as there was no evidence to rate.

Results

Through the database searches, 2857 studies were found. After removing duplicates ($n = 793$) and reading titles and abstracts, 46 studies were read in full. Of these, 15 were excluded for not evaluating measurement properties ($n = 11$), for analyzing only a subscale of the PFDI-20 ($n = 2$), and for evaluating reliability through paper versus internet or telephone ($n = 2$). Of the 31 studies selected after applying the eligibility criteria, 1 study was added through the reference lists, with 32 studies included at the end (Fig. 1).

Of the 32 studies included, 7 evaluated the measurement properties of the PFDI and 25 evaluated the PFDI-20, totaling the participation of 8217 women. The majority of studies were conducted in the USA ($n = 9$; 28.13%) and with women with

PFD ($n = 16$; 50%). The characteristics of the included studies are shown in Chart 1. Except for cross-cultural validity, all the 32 studies included reported at least one measurement property.

Content validity

Content validity was assessed by 21 studies, of which 2 were on the PFDI [10, 11, 13, 20] and 19 on the PFDI-20 [6, 12, 14, 15, 21–33] (Appendix B). Content validity was not assessed in one study on the PFDI [10] and three studies on the PFDI-20 [6, 30, 32] because it was not clear which aspects were assessed, relevance, comprehensiveness, or comprehensibility.

The overall methodological quality of all studies [11–15, 20–29, 31, 33] was rated as “inadequate” because how the patients were asked about the relevance, comprehensiveness, and comprehensibility and how the professionals assessed the relevance and comprehensiveness are not clear or not presented in enough detail. Because of this, the rating of the content validity of these studies was assessed as “insufficient.”

Structural validity

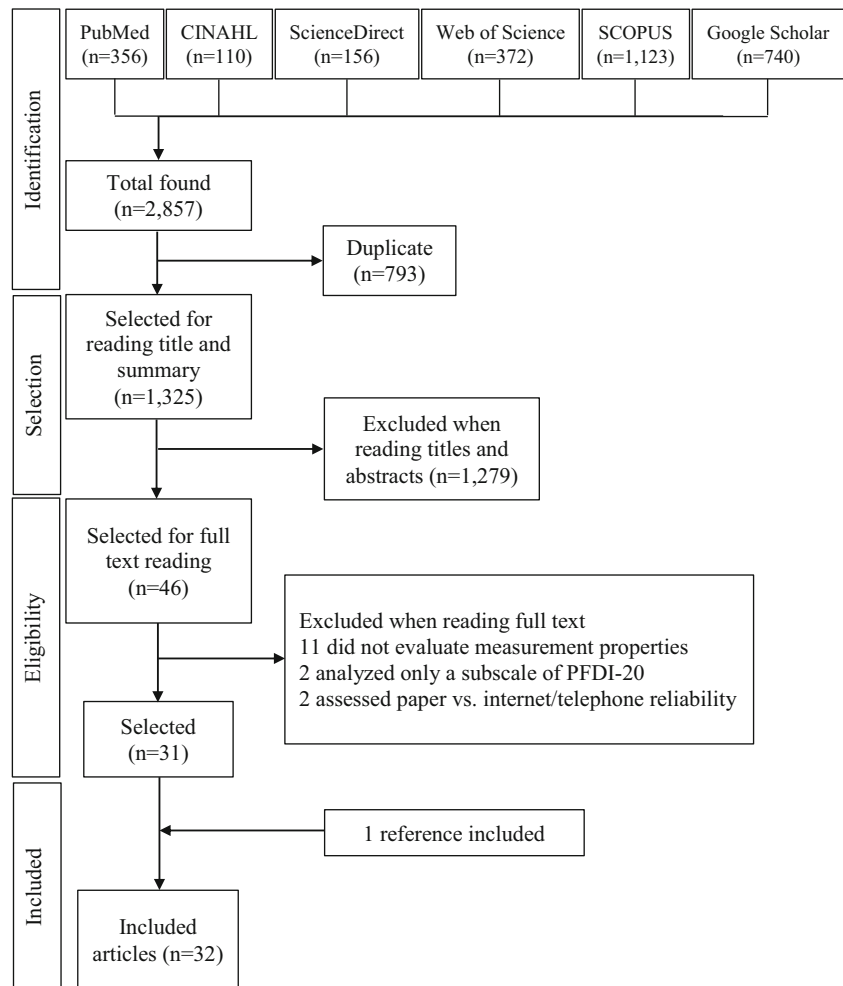
No studies evaluated the structural validity of the PFDI, and only one study reported this measurement property for the PFDI-20 [15]. The study by Ma et al. [15] used confirmatory factor analysis (CFA) and found five factors that explain 69.55% of the cumulative variance: anal and colorectal distress, direct feelings of POP and symptoms of irritation or obstruction of the lower urinary tract, several types of UI, external force to defecate, and symptoms of a rectocele.

The methodological quality of the study by Ma et al. [15] was assessed as “adequate” because of the small sample size (less than seven times the number of items on the PFDI-20). The structural validity rating of the PFDI-20 was considered “indeterminate” since the unidimensionality of each subscale was not confirmed because three factors were not found for the PFDI-20 or a single factor that assesses the distress of the symptoms of PFD.

Internal consistency

Internal consistency was reported by 25 studies, of which 5 were on the PFDI with “very good” [20, 36] and “inadequate” [10, 34, 35] methodological quality and 20 on the PFDI-20 with “very good” [11, 12, 15, 23–25, 28, 32, 37, 38], “doubtful” [26, 30, 31, 33], and “inadequate” [13, 14, 21, 22, 25, 29] methodological quality (Table 1). The “doubtful” methodological quality of the studies occurred because of the low sample number equal to five times the number of items and fewer than 100 individuals. The studies were assessed as “inadequate” quality because of the low sample number less than

Fig. 1 Flowchart for selecting studies based on PRISMA



five times the number of items and because they did not report the Cronbach's α value for each subscale.

The internal consistency of all studies on the PFDI and PFDI-20 was rated as “indeterminate” because of their factorial structure being unclear, contradicting the unidimensionality of each of the three subscales stipulated by the instrument's developers. This is because only one study [15] reported the structural validity of the PFDI-20 and found 5 factors through the CFA: anal and colorectal distress (factor 1); direct POP feelings and symptoms of irritation or obstruction of the lower urinary tract (factor 2); various types of UI (factor 3); external force to defecate (factor 4); symptoms of a rectocele (factor 5). Thus, for COSMIN, the methodological quality of the studies that analyzed the internal consistency must be rated as “doubtful,” and no conclusions can be made.

Test-retest reliability

Of the total studies selected, 4 on the PFDI reported test-retest reliability as “doubtful” [34] and “inadequate” [10, 20, 36] methodological quality and 17 on the PFDI-20 as “doubtful” [13–15, 22, 25–29, 32, 38] and “inadequate” [6, 11, 12, 24,

30, 31] methodological quality (Table 2). The “doubtful” quality occurred because the low sample number, or the time interval between the test and the retest was doubtful, or it was not clear whether the patients were stable between measurements, or they did not calculate the intraclass correlation coefficient (ICC) and because the conditions—environment or administration—of the test are different from the retest. The “inadequate” quality occurred because the time interval between the test and retest was considered inappropriate.

Test-retest reliability of the PFDI was rated in all studies as “sufficient.” This measurement property was rated as “sufficient” in the majority of the studies on the PFDI-20, and only one study [38] was rated as “indeterminate” for not having used the ICC in the analysis.

Measurement errors

Five studies, all on the PFDI-20, reported measurement errors as “doubtful” [14, 15] and “inadequate” [11, 12, 30] methodological quality (Appendix C) because of the time interval between measurements being greater than or less than 2 weeks and the sample size being equal to or less than five times the

Chart 1 Characteristics of the included studies ($n = 32$)

Reference	Country	Population	Sample size	Main objective of the study
PFDI				
Barber et al. [10]	EUA	Women with PFD	100	Psychometrically assess PFDI and PFIQ in a population of women with symptoms of PFD
Barber, Walters, and Cundiff [39]	EUA	Women with POP	106	To assess the responsiveness of PFDI and PFIQ in women with advanced POP undergoing surgical and non-surgical treatment
Omotosho et al. [34]	EUA	Women with PFD	44	Validate Spanish versions of PFDI and PFIQ
Lowder et al. [35]	EUA	Women with PFD	67	Report measures of central tendency for PFDI, PFIQ, and PISQ-12 in the 12 items form in women with POP-Q stage ≤ 1
Chan et al. [20]	China	Women with PFD	597	Investigate the reliability and validity of the Chinese version of PFDI and PFIQ
Chan et al. [41]	China	Women with PFD and/or SUI	156	To assess the responsiveness of Chinese PFDI and PFIQ in women with POP and/or SUI undergoing treatment
Chan et al. [36]	China	Women with PFD	597	To investigate the reliability and validity of a Chinese version of PFDI and PFIQ in Chinese women with PFD
PFDI-20				
Barber, Walters, and Bump [6]	EUA	Women with PFD	145	To develop short forms of PFDI and PFIQ for women with PFD
El-Azab, Abd-Elisayed, and Imam [21]	Egypt	Women with POP	20	Modify PFIQ and introduce it with PFDI-20 in clinical and research practice to assess POP among Arab Muslim women
Frick et al. [42]	EUA	Women with UI	135	Compare the responsiveness of various measures of PFD result and quality of life in women undergoing SUI surgery
Barber et al. [39]	EUA	Women with POP	1006	Evaluate the validity and responsiveness of PFDI-20 and PFIQ-7 in four multicenter studies and develop formulas for conversion between short and long versions
Teleman et al. [22]	Sweden	Women with POP	44	Psychometrically evaluate the Swedish translations of PFIQ-7, PFDI-20, and PISQ-12
Gelhorn et al. [37]	Australia, Germany, UK, EUA, France, Belgium, The Netherlands	Women with POP	275	To evaluate the psychometric characteristics of PFDI-20, PFIQ-7, and PGIC in women undergoing POP repair surgery
Kaplan, Sut, and Sut [23]	Turkey	Women with or without POP	248	Culturally adapt, validate and assess the responsiveness of PFDI-20 and PFIQ-7 in a Turkish population
Lowenstein et al. [38]	Israel	Women with PFD	54	Translate and validate Hebrew versions of three validated questionnaires commonly used in the field of urogynecology
Celenay et al. [24]	Turkey	Women with PFD	128	Investigate the reliability and validity of the Turkish version of PFDI-20
Treszezamsky et al. [25]	EUA	Women with PFD	78	Develop valid Spanish translations of QUIJ, 3IQ, PFDI-20 and PFIQ-7
Due, Broström, and Lose [27]	Denmark	Women with POP	132	Translate PFDI-20 and PFIQ-7 and evaluate their psychometric properties in Danish women with symptomatic POP
Grigoriadis et al. [26]	Greece	Women with PFD	185	Translate and validate the PFDI-20 and PFIQ-7 questionnaires in Greek
Sánchez-Sánchez et al. [28]	Spain	Women with PFD	114	Develop a Spanish version of PFDI-20 and PFIQ-7, linguistically and psychometrically validated, to assess symptoms and quality of life in Spanish women with PFD

Chart 1 (continued)

Reference	Country	Population	Sample size	Main objective of the study
Yoshida et al. [29]	Japan	Women with or without PFD	59	Translate the PFDI-20 into Japanese and test its reliability and validity among Japanese women
Utomo et al. [30]	The Netherlands	Women with PFD	111	Validate PFDI-20 and PFIQ-7 in Dutch women
Sánchez-Sánchez et al. [43]	Spain	Women with PFD	85	Assess the responsiveness of the Spanish versions of PFDI-20 and PFIQ-7 to assess symptoms and quality of life in Spanish women with PFD
Arouca et al. [11]	Brazil	Women with or without PFD	250	Translate and validate PFDI-20 and PFIQ-7 in Brazilian Portuguese
Henn, Richter, and Marokane [12]	South Africa	Women with or without PFD	400	Validate PFDI-20 and PFIQ-7 in African women for Afrikaans and Sesotho languages
Mattsson et al. [13]	Finland	Women with POP	63	Translate PFDI-20, PFIQ-7, and PISQ-12 into Finnish and evaluate their psychometric properties in Finnish women with symptomatic POP
Teig et al. [14]	Norway	Women with PFD	205	Translate into Norwegian and validate PFDI-20 and PFIQ-7 using a sample of women with symptomatic POP and PFD
Goba et al. [31]	Ethiopia	Women with or without PFD	118	Translate and adapt the English versions of PFDI-20 and PFIQ-7 to Tigrigna and test the validity and reliability of translated questionnaires
Grzybowska et al. [32]	Poland	Women with PFD	244	Develop a Polish language version of PFDI-20 and validate it in a sample of Polish-speaking women with PFD
Ma et al. [15]	China	Women with PFD	126	Translate PFDI-20 into Chinese and evaluate its psychometric properties in Chinese women with PFD, according to COSMIN
Pangastuti et al. [33]	Indonesia	Health professionals	77	Obtain the validity and reliability of PFDI-20 in healthcare professionals in Indonesia
Mattsson et al. [44]	Finland	Women with POP	2248	To evaluate the correlation between PFDI-20, 15D and PGI-1 in the change in health-related quality of life after POP surgery

PFDI: Pelvic Floor Distress Inventory. PFDI-20: Pelvic Floor Distress Inventory short form. PFIQ: Pelvic Floor Impact Questionnaire short form. PFD: Pelvic floor dysfunctions. UI: Urinary incontinence. SUI: Stress urinary incontinence. POP: Pelvic organ prolapse. POP-Q: Pelvic Organ Prolapse Quantification. PISQ-12: Pelvic organ prolapse/urinary incontinence Sexual Questionnaire. PGIC: Patient Global Impression of Change. QUID: Questionnaire for Urinary Incontinence Diagnosis. 3IQ: Three Incontinence Questions. 15D: 15-dimensional instrument. PGI-1: Patient Global Impression of Improvement. COSMIN: Consensus-based Standards for the selection of health Measurement Instruments

Table 1 Cronbach's α values for internal consistency of PFDI and PFDI-20 and subscales

Reference	Assessments	Total score	Subscales				Methodological quality	Rating
			PFDI	POPDI	CRADI	UDI		
Barber et al. [10]	Test	0.88	0.82	0.84	0.89	Inadequate	Indeterminate	
Omotosho et al. [34]	Test	NS	≥ 0.84	≥ 0.84	≥ 0.84	Inadequate	Indeterminate	
Lowder et al. [35]	Test	0.93	NS	NS	NS	Inadequate	Indeterminate	
Chan et al. [20]	Test	0.93	0.87	0.86	0.89	Very good	Indeterminate	
Chan et al. [36]	Test	0.93	0.87	0.86	0.89	Very good	Indeterminate	
		PFDI-20	POPDI-6	CRADI-8	UDI-6			
El-Azab, Abd-Elsayed, and Imam [21]	Test	NS	NS	NS	NS	Inadequate	Indeterminate	
Teleman et al. [22]	Test	0.841	0.569	0.787	0.737	Inadequate	Indeterminate	
	Retest	0.842	0.637	0.782	0.769			
Gelhom et al. [37]	Study 1	0.85	0.68	0.78	0.72	Very good	Indeterminate	
	Study 2	0.82	0.68	0.64	0.70			
Kaplan, Sut and Sut [23]	Test	0.908	0.952	0.964	0.965	Very good	Indeterminate	
Lowenstein et al. [38]	Test	NS	0.85	0.65	0.71	Very good	Indeterminate	
Celenay et al. [24]	Test	0.79	0.80	0.73	0.66	Very good	Indeterminate	
Treszezamsky et al. [25]	Test	NS	0.81	0.79	0.83	Inadequate	Indeterminate	
Due, Broström, and Lose [27]	Test	0.804	0.606	0.782	0.721	Very good	Indeterminate	
Grigoriadis et al. [26]	Test	0.852	0.832	0.803	0.758	Doubtful	Indeterminate	
	Retest	0.836	0.796	0.807	0.720			
Sánchez-Sánchez et al. [28]	Test	0.837	0.787	0.630	0.699	Very good	Indeterminate	
Yoshida et al. [29]	Test	0.83	0.82	0.52	0.75	Inadequate	Indeterminate	
Utomo et al. [30]	Group 1	0.74	0.52	0.71	0.60	Doubtful	Indeterminate	
	Group 2	0.84	0.60	0.71	0.74			
Arouca et al. [11]	Group 1	0.816	0.839	0.842	0.841	Very good	Indeterminate	
	Group 2	0.844	0.850	0.855	0.848			
Henn, Richter, and Marokane [12]	Group 1	0.89	0.80	0.89	0.79	Very good	Indeterminate	
	Group 2	0.83	0.91	0.59	0.78			
	Group 3	0.71	0.43	0.85	0.71			
	Group 4	0.75	0.43	0.61	0.61			
Mattsson et al. [13]	Test 1	0.88	0.71	0.80	0.71	Inadequate	Indeterminate	
	Test 2	0.88	0.74	0.80	0.71			
	Retest 1	0.89	0.78	0.83	0.69			
	Retest 2	0.89	0.78	0.83	0.69			
Teig et al. [14]	Test	0.83	0.66	0.72	0.71	Inadequate	Indeterminate	
Goba et al. [31]	Test	0.930	0.897	0.892	0.891	Doubtful	Indeterminate	
Grzybowska et al. [32]	Test	0.89	0.81	0.85	0.80	Very good	Indeterminate	
Ma et al. [15]	Test	0.88	0.77	0.84	0.80	Very good	Indeterminate	
Pangastuti et al. [44]	Test	0.911	0.902–0.913	0.902–0.913	0.902–0.913	Doubtful	Indeterminate	

NS: Not shown. PFDI: Pelvic Floor Distress Inventory. PFDI-20: Pelvic Floor Distress Inventory short form. POPDI: Pelvic Organ Prolapse Distress Inventory long form. POPDI-6: Pelvic Organ Prolapse Distress Inventory short form. CRADI: Colorectal-Anal Distress Inventory long form. CRADI-8: Colorectal-Anal Distress Inventory short form. UDI: Urinary Distress Inventory long form. UDI-6: Urinary Distress Inventory short form

number of items on the PFDI-20. All the studies used appropriate methods to assess measurement errors, such as standard error of measurement (SEM), limits of agreement (LoA), and smallest detectable change (SDC).

Regarding the rating of measurement errors by study, studies were rated with measurement errors “sufficient” [14], “indeterminate” [11, 12], and “insufficient” [15, 30]. Two studies had measurement errors rated as “indeterminate” because they did not present the values of minimal important change

(MIC), and two were rated as “insufficient” because they presented an MIC less than the LoA or SDC.

Criterion validity

As there is no gold standard for assessing the distress of all PFD symptoms assessed by the PFDI-20 (POP, anorectal, and urinary), only two studies [6, 39] reported criterion validity for the PFDI-20 with “very good” methodological quality and

Table 2 Test-retest reliability of PFDI and PFDI-20

Reference	Coefficient	Total score	Subscales			Methodological quality	Rating
			PFDI	POPDI	CRADI		
Barber et al. [10]	ICC	0.87	0.87	0.86	0.87	Inadequate	Sufficient
Omotosho et al. [34]	ICC	0.71–0.86	NS	NS	NS	Doubtful	Sufficient
Chan et al. [20]	ICC	0.77	0.79	0.80	0.83	Inadequate	Sufficient
Chan et al. [36]	ICC	0.77	0.79	0.80	0.83	Inadequate	Sufficient
		PFDI-20	POPDI-6	CRADI-8	UDI-6		
Barber, Walters, and Bump [6]	ICC	0.93	0.91	0.84	0.82	Inadequate	Sufficient
Teleman et al. [22]	ICC	0.932	0.888	0.926	0.924	Doubtful	Sufficient
Lowenstein et al. [38]	Correlation	NS	0.81	0.72	0.73	Doubtful	Indeterminate
Celenay et al. [24]	ICC	0.98	0.98	0.96	0.98	Inadequate	Sufficient
Treszezamsky et al. [25]	ICC	NS	0.91	0.84	0.92	Doubtful	Sufficient
Due, Broström, and Lose [27]	ICC	0.878	0.894	0.847	0.860	Doubtful	Sufficient
Grigoriadis et al. [26]	ICC	0.85	0.87	0.75	0.78	Doubtful	Sufficient
Sánchez-Sánchez et al. [28]	ICC	0.644	0.711	0.771	0.428	Doubtful	Sufficient
Yoshida et al. [29]	ICC	0.87	0.83	0.77	0.90	Doubtful	Sufficient
Utomo et al. [30]	ICC	0.88	0.86	0.90	0.89	Inadequate	Sufficient
Arouca et al. [11]	ICC	0.803	0.764	0.789	0.767	Inadequate	Sufficient
Henn, Richter, and Marokane [12]	ICC group 1	0.97	0.95	0.99	0.99	Inadequate	Sufficient
	ICC group 2	0.99	0.99	0.99	0.98		
	ICC group 3	0.97	0.99	0.98	0.98		
	ICC group 4	0.98	0.98	0.97	0.95		
Mattsson et al. [13]	ICC	0.92	0.83	0.90	0.89	Doubtful	Sufficient
Teig et al. [14]	ICC	0.944	0.895	0.938	0.918	Doubtful	Sufficient
Goba et al. [31]	ICC	0.941	0.969	0.765	0.881	Inadequate	Sufficient
Grzybowska et al. [32]	ICC	0.87	0.80	0.87	0.86	Doubtful	Sufficient
Ma et al. [15]	ICC	0.997	0.994	0.997	0.995	Doubtful	Sufficient

NS: Not shown. ICC: Intraclass correlation coefficient. PFDI: Pelvic Floor Distress Inventory. PFDI-20: Pelvic Floor Distress Inventory short form. POPDI: Pelvic Organ Prolapse Distress Inventory long form. POPDI-6: Pelvic Organ Prolapse Distress Inventory short form. CRADI: Colorectal-Anal Distress Inventory long form. CRADI-8: Colorectal-Anal Distress Inventory short form. UDI: Urinary Distress Inventory long form. UDI-6: Urinary Distress Inventory short form

“sufficient” criterion validity. According to COSMIN [19], the long version of an abbreviated PROM is considered the gold standard when it is not available. Pearson’s correlation coefficient was used to assess the criterion validity of the PFDI-20 subscales in the studies by Barber et al. [6] (POPDI-6 $r = 0.92$; CRADI-8 $r = 0.93$; UDI-6 $r = 0.86$) and Barber et al. [39] (POPDI-6 $r = 0.90$; CRADI-8 $r = 0.93$; UDI-6 $r = 0.88$).

Hypothesis testing (convergent validity)

Convergent validity was reported by 15 studies, including 3 studies on the PFDI with “very good” [20] and “inadequate” [10, 36] methodological quality and 12 studies on PFDI-20 with “very good” [14, 23, 32, 37], “doubtful” [24, 27, 28, 31], and “inadequate” methodological quality [13, 15, 22, 29] (Appendix D). The “doubtful” methodological quality rating was given because of the low sample size (equal to 5 times the

number of items). The studies were evaluated with “inadequate” methodological quality due to the very small sample size (less than 5 times the number of items).

All studies performed correlations as the predominant statistical method. Instruments to assess quality of life, whether or not related to PFD, such as the SF-12 Health Survey (SF-12), the Short Form Health Survey (SF-36), the Visual Analogue Scale (VAS), the Pelvic Floor Impact Questionnaire (PFIQ-7), and the Urinary Incontinence Quality of Life Scale (I-QoL), were predominantly used [13–15, 20, 22, 24, 27–29, 32, 36]. The majority of studies quantified the degree of POP objectively through the Pelvic Organ Prolapse Quantification (POP-Q) by most studies [10, 20, 23, 27, 29, 31, 36, 37].

The convergent validity rating for the PFDI was “sufficient” [20], “indeterminate” [39], and “insufficient” [36]. The “indeterminate” rating occurred because there was no clearly defined hypothesis and “insufficient” because the

results were not in line with the hypotheses. For the PFDI-20, the convergent validity rating for each study was “sufficient” [14, 22–24, 28, 29, 31, 32] and “insufficient” [13, 15, 27, 37]. The rating was “insufficient” as < 75% of the results were in accordance with the hypotheses.

Hypothesis testing (divergent validity)

Divergent validity was reported by all six studies on the PFDI-20 with “very good” [32], “doubtful” [28, 31], and “inadequate” [13, 22, 29] methodological quality (Appendix E). Studies were evaluated as “doubtful” and “inadequate” quality because of the low (equal to 5 times the number of items) and very low (less than 5 times the number of items) sample sizes, respectively.

Regarding the rating of divergent validity, 5 studies were rated as “sufficient” [22, 28, 29, 31, 32] and one as “insufficient” [13]. The study by Mattsson et al. [13] was rated as “insufficient” divergent validity as < 75% of the results were in accordance with the hypotheses.

Hypothesis testing (comparison between known groups)

The comparison between known groups or discriminative validity was reported by one study on the PFDI with “very good” [20] methodological quality and by eight studies on the PFDI-20 with “very good” [23], “doubtful” [11, 12, 15, 30, 31, 37], and “inadequate” [29] methodological quality (Appendix F). The studies were evaluated as “doubtful” quality because of the small sample size, because they used incorrect statistical tests, or because they did not present values for each subscale of the PFDI-20. The study with “inadequate” methodological quality was assessed as the sample size was considered too small and because of the use of an inadequate statistical test. All studies on the PFDI and PFDI-20 included women with and without POP or with different stages of POP.

The rating of the comparison of known groups of the PFDI was “sufficient” and of the PFDI-20 was “sufficient” [11, 12, 15, 23, 29, 31] and “insufficient” [30, 37]. The “insufficient” rating was due to < 75% of the results being in accordance with the hypotheses.

Responsiveness

Seventeen studies, 2 on the PFDI with “very good” [41] and “inadequate” [40] methodological quality and 15 on the PFDI-20 with “very good” [6, 26], “doubtful” [11, 12, 14, 15, 23, 27, 30, 37, 39, 42, 44], and “inadequate” [21, 43] methodological quality evaluated responsiveness (Appendix G). The “doubtful” quality of the studies was due to the small sample size, the use of an inappropriate method to test the hypotheses, and the lack of an adequate description of the intervention.

The studies were rated as “inadequate” because of the very small sample size or the use of inappropriate statistical methods. Surgery was the most commonly used intervention by studies [6, 11, 12, 14, 15, 21, 23, 26, 27, 30, 37, 39–42, 44].

The PFDI responsiveness rating was “sufficient” [40] and “insufficient” [41]. Seven studies on the PFDI-20 received “sufficient” ratings of responsiveness [6, 14, 23, 26, 27, 37, 44], six had an “indeterminate” rating [11, 12, 15, 21, 30, 42], and two were rated as “insufficient” [39, 43]. Responsiveness was rated as “indeterminate” because the studies did not present values for the subscales or presented only values of statistical significance and “insufficient” because the results were not in line with the hypotheses.

Data synthesis

Table 3 presents the overall rating of the measurement properties and the quality of the evidence of the PFDI and PFDI-20. The general content validity rating of the PFDI and PFDI-20 was “insufficient” and the quality of the evidence “very low” due to the existence of only studies with inadequate methodological quality.

The structural validity of the PFDI-20 and the internal consistency of the PFDI and PFDI-20 were rated as “indeterminate” because the PROM subscales do not have proven one-dimensionality. Thus, the quality of the evidence was not assessed.

The test-retest reliability of the PFDI was rated as “sufficient” because all studies were assessed as “sufficient” and with an ICC ≥ 0.7 . However, the quality of the evidence of the PFDI was considered “moderate” due to the fact that all studies were evaluated as having “doubtful” or “inadequate” methodological quality. The test-retest reliability of the PFDI-20 was also rated as “sufficient” with “moderate” quality of evidence due to the existence of several studies of “doubtful” or “inadequate” quality.

For the PFDI-20, measurement errors were rated as “inconsistent” because the majority of the existing studies were rated as “indeterminate” or “insufficient.” The quality of the evidence was assessed as “moderate” because all studies were of “doubtful” or “inadequate” quality and because the majority of studies had an “indeterminate” or “insufficient” rating.

The criterion validity and construct validity for hypothesis testing of the PFDI-20 were rated as “sufficient” and with high evidence. However, the construct validity for PFDI hypothesis testing was rated as “insufficient” and the quality of the evidence was assessed as high. The “inconsistent” rating occurred because 50% of the studies were rated as “indeterminate” or “insufficient” and 50% as “sufficient.” This criterion was applied as if < 75% of the results are rated as “insufficient,” the rating by measurement property is also “insufficient.”

Table 3 Classification and quality of evidence of the measurement properties of PFDI and PFDI-20

Measurement property	Rating of measurement property	Quality of evidence
PFDI		
Content validity	Insufficient	Very low
Internal consistency	Indeterminate	NR
Test-retest reliability	Sufficient	Moderate
Construct validity-hypothesis testing	Insufficient	High
Responsiveness	Insufficient	Moderate
PFDI-20		
Content validity	Insufficient	Very low
Structural validity	Indeterminate	NR
Internal consistency	Indeterminate	NR
Test-retest reliability	Sufficient	Moderate
Measurement errors	Inconsistent	Moderate
Criterion validity	Sufficient	High
Construct validity-hypothesis testing	Sufficient	High
Responsiveness	Inconsistent	High

NR: Not rated. PFDI: Pelvic Floor Distress Inventory long form. PFDI-20: Pelvic Floor Distress Inventory short form

The responsiveness of the PFDI was rated as “insufficient” because 50% of the studies were rated as “sufficient” and 50% as “insufficient.” This is also due to the small number of studies that evaluated this measurement property for the PFDI. The quality of the evidence of responsiveness of the PFDI was assessed as moderate based on the disagreement between the ratings of the studies, as only two studies are available, one rated as “sufficient” and the other as “insufficient.” Regarding the responsiveness of the PFDI-20, this measurement property was rated as “inconsistent” and with a high quality of evidence. The rating was “inconsistent” because of the considerable number of studies with “indeterminate” and “sufficient” ratings.

Discussion

For the PFDI, only the hypothesis testing presented a high quality of evidence, while the PFDI-20 had a high quality of evidence for criterion validity, hypothesis testing, and responsiveness. For the structural validity and internal consistency of both PROMs, it was not possible to determine the quality of the evidence. The content validity of both PROMs had very low quality of evidence. No studies on the PFDI and PFDI-20 assessed cross-cultural validity. This demonstrates a serious problem in the validation process of these PROMs, reflecting the need for a new process to evaluate the measurement properties of the instruments. It is possible that a process of validation and evaluation of the measurement properties of the PROM, mainly evaluating relevance, comprehensiveness, and comprehensibility of the content validity and following the COSMIN methodology, would be most appropriate.

Content validity is considered the most important measurement property of a PROM because, if the PROM construct is unclear, the evidence for the remaining measurement properties is not valuable [18, 19]. This measurement property is evaluated by asking the patient about relevance, comprehensiveness, and comprehensibility, and professionals about relevance and comprehensiveness [18]. Thus, the content validity of the PFDI and PFDI-20 would be an adequate reflection of the construct to be measured if relevance, comprehensiveness, and comprehensibility were evaluated. In addition, whether the items of both PROMs are relevant for the assessment of the distress of the symptoms of PFD, comprehensive for the assessment of this distress, and understandable to the population could be assessed. In this systematic review, 1 study on the PFDI and 16 studies on the PFDI-20 almost exclusively assessed comprehensibility by patients; only 1 study assessed the comprehensiveness of the PFDI-20 and 3 studies assessed the relevance of the PFDI-20 to patients. Regarding the evaluation by professionals, 16 studies assessed the relevance and comprehensiveness of the PFDI-20. Overall, the studies are not clear and do not present details on how patients and professionals are questioned on these aspects of content validity. This represents a major flaw in the validation process for both instruments and reflects the quality of the remaining measurement properties.

In the current systematic review, the methodological quality for structural validity was not assessed by any studies on the PFDI and was assessed by only one study on the PFDI-20. For the PFDI-20, the quality of evidence of internal consistency was not assessed because three factors were not found, one for each subscale, or any factors to assess the distress from PFD symptoms. In addition, in the study by Ma et al. [15], the

method of factor analysis is not very clear, since it is advisable to perform exploratory factor analysis before CFA if there is no evidence about the dimensionality of an instrument [18]. The lack of studies on the structural validity of the PFDI and the failure to assess the quality of evidence on the structural validity of the PFDI-20 compromised the quality of the evidence of internal consistency. This was because the structural validity is a prerequisite for assessing internal consistency, because for each factor identified in the instrument there must be a unidimensionality value: Cronbach's alpha [18].

The cross-cultural validity or invariance of the measure was the only measurement property not evaluated for the PFDI and PFDI-20. For the assessment of cross-cultural validity, data from two different populations are needed, in which one population completes the original version of the instrument and the other completes the culturally adapted version. For the analysis between the versions of the instrument, some statistical methods are generally used, such as the differential item functioning (DIF), which detects items that differ between subgroups of a population; factor analysis; logistic regression analysis; and techniques of item response theory [18, 19].

In addition to cross-cultural validity, criterion validity was not evaluated for the PFDI. However, two studies showed these measurement properties for the PFDI-20 [6, 39]. According to the COSMIN guideline [19], if a measure does not have a gold standard defined in the literature, one way to check the criterion validity is by comparing the long version of the PROM with the short version. In the case of the PFDI and PFDI-20, which assess different symptoms of PFD and their distress in women, there is no established gold standard. Thus, the comparison between the PFDI and PFDI-20 is useful for assessing criterion validity, being rated as “sufficient” and high quality of evidence for the PFDI-20.

Test-retest reliability is analyzed based on two measurements made with the same subject after a period of time. According to COSMIN [18, 19], some requirements are necessary for the “sufficient” rating of reliability, such as the appropriate time interval between the test and the retest, patient stability in the interim period, similar conditions for the measurements, appropriate sample number, and analysis of the ICC with its specifications for continuous results. The reliability of the PFDI and PFDI-20 were rated as “sufficient” because of the use of the ICC and values obtained above 0.70 in more than 70% of the studies, but the evidence was considered moderate because several studies had “doubtful” or “inadequate” quality. This criterion followed the only downgrade of the risk of bias by the modified GRADE. Thus, although some studies showed flaws in relation to the requirements for reliability analysis, what most defined the overall quality of the evidence from the PFDI and PFDI-20 were the results of the ICC.

Similar to reliability, measurement errors are assessed for the same subject based on two time-span measurements. For the evaluation, each study must clearly provide the calculation and value of the MIC, SEM, SDC, or LoA. The MIC is defined by COSMIN as “the smallest change in the score in the construct to be measured that patients perceive as important” and can come from different studies [18]. In addition, for a PROM to have a “sufficient” overall rating and high quality of evidence of measurement errors, it should be assessed, respectively, whether the SDC or LoA is less than the MIC and if the risk of bias, inconsistency, imprecision, and indirectness assessed by the modified GRADE are satisfactory [18]. In the case of the PFDI-20, which had an “inconsistent” overall rating of measurement errors and moderate quality of evidence, these measurement error assessment requirements were not met. The overall rating was considered “inconsistent” for the PFDI-20 measurement errors because > 75% of the studies had SDC or LoA values greater than the MIC value or did not define the MIC.

When a gold standard does not exist or is not used by studies to assess criterion validity, the hypothesis test for construct validity must be performed through the expected relationships with other measures (convergent validity), not expected with other measures (divergent validity), and/or through expected differences between relevant groups (discriminatory or known groups). For this, one of the prerequisites for evaluating the hypothesis testing is the previous formulation of hypotheses with inclusion of the directions and expected magnitudes of the correlations between the instruments or of the average differences between the groups [18]. In this systematic review, the overall rating of the hypothesis testing for construct validity of the PFDI was considered “insufficient” because < 75% of the studies were rated as “sufficient.” For the convergent validity of the PFDI, 66.67% of the studies received the “indeterminate” or “insufficient” rating because no hypothesis was defined or because the result was not in accordance with the hypothesis presented. Divergent validity of the PFDI was not evaluated by any studies, and only one study [20] assessed the PFDI as “sufficient” to compare known groups. In contrast, the PFDI-20 had a “sufficient” rating for hypothesis testing and high quality of evidence.

Responsiveness is an important measure for clinical practice because it has the ability to detect changes in the construct to be measured over time. Thus, the responsiveness of a PROM helps clinical professionals to observe improvement or worsening in a patient's health after an intervention [18]. In addition to being evaluated in longitudinal studies, for this measurement property, four approaches are used to rate the methodological quality of the studies: (1) criterion approach; (2) construct approach—comparison with other PROMs; (3) construct approach—comparison between subgroups; (4) construct approach—pre- and post-intervention. The

evaluation of approaches 1, 2, and 3 can be compared, respectively, to the evaluations of the criterion validity and the hypothesis testing for construct validity-convergent validity and known groups. However, the general responsiveness of the PFDI was rated as “insufficient” and of moderate quality of evidence due to the lack of hypotheses or because there was no agreement between the results and the hypotheses of the two existing studies [40, 41]. For the PFDI-20, responsiveness had an overall “inconsistent” rating and high quality of evidence. The rating was “inconsistent” because most studies on the PFDI-20 received a “doubtful” rating as they did not provide adequate information about the intervention and/or use an inappropriate statistical method to test the hypotheses.

This was the first systematic review study to investigate the measurement properties of two instruments frequently used in scientific research [7, 8] and recommended by the ICI as grade A for the assessment of PFD in women [9]. In clinical practice, the need for PROMs with high-quality evidence measurement properties contributes to the correct assessment of the patient’s health status. In scientific research, high-quality PROMs improves the accuracy of data analysis on the health status of a particular group or population. For researchers and health professionals in the field of urogynecology who use the PFDI or PFDI-20, this systematic review presented the quality of evidence of the measurement properties of both instruments in detail and also identified the measurement properties that need further investigation. Despite this, the failure to assess interpretability, “the degree to which qualitative significance can be attributed to quantitative scores or changes in the scores of an instrument,” can be considered a limitation of this study. Interpretability is not considered a measurement property for COSMIN because it does not assess the quality of studies related to validation and reliability [18]. However, the evaluation of interpretability is recommended in order to understand the score in different populations.

According to the results of this systematic review, the PFDI has a high quality of evidence for construct validity-hypothesis testing, moderate for test-retest reliability and responsiveness, and very low quality of evidence for content validity. The PFDI-20 demonstrates a high quality of evidence for criterion validity, construct validity-hypothesis testing and responsiveness, moderate quality for test-retest reliability and measurement errors, and very low quality of evidence for content validity. The internal consistency of the PFDI and PFDI-20 was not evaluated because of the scarcity of studies that confirmed the unidimensionality of the PFDI, the PFDI-20, or its subscales. Thus, we suggest that future studies assess the factorial structure of both PROMs to determine the dimensionality and internal consistency of their dimensions. This would help in identifying items that assess the same construct, such as POP, anorectal and urinary symptoms, or just distress from PFD symptoms. In addition, it is necessary to evaluate the analysis of the cross-cultural validity as a measure of the

behavior of the translated items in relation to the original version of the instrument and, mainly, the content validity throughout its methodological process in order to improve the quality of evidence of this measurement property.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1007/s00192-021-04748-4>

Funding This study was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (grant number 3.437.754).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

- Haylen BT, de Ridder D, Freeman RM, Swift SE, Berghmans B, 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. <https://doi.org/10.1007/s00192-009-0976-9>.
- Islam RM, Oldroyd J, Rana J, Romero L, Karim MN. Prevalence of symptomatic pelvic floor disorders in community-dwelling women in low and middle-income countries: a systematic review and meta-analysis. *Int Urogynecol J.* 2019;30(12):2001–11.
- Almoussa S, van Loon AB. (2018). The prevalence of urinary incontinence in nulliparous adolescent and middle-aged women and the associated risk factors: a systematic review. *Maturitas* 107:78–83.
- Jerez-Roig J, Booth J, Skelton DA, Giné-Garriga M, Chastin SFM, Hagen S. Is urinary incontinence associated with sedentary behaviour in older women? Analysis of data from the National Health and nutrition examination survey. *PLoS One.* 2020;15(2):e0227195.
- Mazi B, Kaddour O, Al-Badr A. Depression symptoms in women with pelvic floor dysfunction: a case-control study. *Int J Women's Health.* 2019;11:143–8.
- 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):103–13.
- Aguilar VC, White AB, Rogers RG. Updates on the diagnostic tools for evaluation of pelvic floor disorders. *Curr Opin Obstet Gynecol.* 2017;29(6):458–64.
- Zuchelo LTS, Bezerra IMP, Silva ATM, Gomes JM, Júnior JMS, Baracat EC, et al. Questionnaires to evaluate pelvic floor dysfunction in the postpartum period: a systematic review. *Int J Women's Health.* 2018;10:409–24.
- Abrams P, et al. *Incontinence.* 6th ed. Paris: Health Publication; 2017.
- 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(6):1388–95.
- Arouca MAF, Duarte TB, Lott DAM, Magnani PS, Nogueira AA, Rosa-E-Silva JC, et al. Validation and cultural translation for Brazilian Portuguese version of the pelvic floor impact questionnaire (PFIQ-7) and pelvic floor distress inventory (PFDI-20). *Int Urogynecol J.* 2016;27(7):1097–106.

12. Henn EW, Richter BW, Marokane MMP. Validation of the PFDI-20 and PFIQ-7 quality of life questionnaires in two African languages. *Int Urogynecol J*. 2017;28(12):1883–90.
13. Mattsson NK, Nieminen K, Heikkinen A-M, Jalkanen J, Koivurova S, Eloranta M-L, et al. Validation of the short forms of the pelvic floor distress inventory (PFDI-20), pelvic floor impact questionnaire (PFIQ-7), and pelvic organ prolapse/urinary incontinence sexual questionnaire (PISQ-12) in Finnish. *Health Qual Life Outcomes*. 2017;15(1):88.
14. Teig CJ, Grotle M, Bond MJ, Prinsen CAC, Engh MAE, Cvancarova MS, et al. Norwegian translation, and validation, of the pelvic floor distress inventory (PFDI-20) and the pelvic floor impact questionnaire (PFIQ-7). *Int Urogynecol J*. 2017;28(7):1005–17.
15. Ma Y, Xu T, Zhang Y, Mao M, Kang J, Zhu L. Validation of the Chinese version of the pelvic floor distress Inventory-20 (PFDI-20) according to the COSMIN checklist. *Int Urogynecol J*. 2019;30(7):1127–39.
16. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.
17. Terwee CB, Jansma EP, Riphagen II, De Vet HCW. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 18(8):1115–1123.
18. Mokkink LB, Prinsen CAC, Bouter LM, de Vet HCW, Terwee CB. (2016). The COSensus-based standards for the selection of health measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Braz J Phys Ther* 20(2):105–113.
19. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147–57.
20. Chan SSC, et al. Chinese validation of pelvic floor distress inventory and pelvic floor impact questionnaire. *Int Urogynecol J*. 2011;22:1305–12.
21. El-Azab AS, Abd-Elsayed AA, Imam HM. Patient reported and anatomical outcomes after surgery for pelvic organ prolapse. *Neurourol Urodyn*. 2009;28(3):219–24.
22. 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.
23. 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.
24. Celenay ST, 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.
25. Treszezamsky AD, et al. Spanish translation and validation of four short pelvic floor disorders questionnaires. *Int Urogynecol J*. 2012;24(4):655–70.
26. Grigoriadis T, et al. Translation and psychometric evaluation of the Greek short forms of two condition-specific quality of life questionnaires for women with pelvic floor disorders: PFDI-20 and PFIQ-7. *Int Urogynecol J*. 2013;24(12):2131–44.
27. Due U, Broström S, Lose G. Validation of the pelvic floor distress Inventory-20 and the pelvic floor impact Questionnaire-7 in Danish women with pelvic organ prolapse. *Acta Obstet Gynecol Scand*. 2013;92:1041–8.
28. Sánchez-Sánchez B, et al. Cultural adaptation and validation of the pelvic floor distress inventory short form (PFDI-20) and pelvic floor impact questionnaire short form (PFIQ-7) Spanish versions. *Eur J Obstet Gynecol Reprod Biol*. 2013;170:281–5.
29. 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;24:1039–46.
30. 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):531–44.
31. Goba GK, et al. Reliability and validity of the Tigrigna version of the pelvic floor distress inventory-short form 20 (PFDI-20) and pelvic floor impact Questionnaire-7 (PFIQ-7). *Int Urogynecol J*. 2019;30(1):65–70.
32. Grzybowska ME, et al. Validation of the polish version of the pelvic floor distress inventory. *Int Urogynecol J*. 2019;30(1):101–5.
33. Pangastuti N, et al. Validation test of Indonesian pelvic floor distress inventory-20 (Indonesian PFDI-20). *J Med Sci*. 2020;52(2):131–7.
34. Omotosho TB, et al. Validation of Spanish versions of the pelvic floor distress inventory (PFDI) and pelvic floor impact questionnaire (PFIQ): a multicenter validation randomized study. *Int Urogynecol J*. 2009;20:623–39.
35. Lowder JL, et al. Normative data for commonly used validated pelvic floor disorder questionnaires in women. *Female Pelvic Med Reconstr Surg*. 2010;16:296–8.
36. Chan SS, Pang SM, Lai BP, Choy KW. Quality of life and symptom measurement in Chinese women with pelvic floor disorders: validation study of pelvic floor distress inventory and pelvic floor impact questionnaire. *Hong Kong Med J* 23 Suppl. 2017;2(3):38–41.
37. Gelhorn HL, Coyne KS, Sikirica V, Gauld J, Murphy M. Psychometric evaluation of health-related quality-of-life measures after pelvic organ prolapse surgery. *Female Pelvic Med Reconstr Surg*. 2012;18(4):221–6.
38. 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):329–31.
39. Barber MD, et al. Further validation of the short form versions of the pelvic floor distress inventory (PFDI) and pelvic floor impact questionnaire (PFIQ). *Neurourol and Urodyn*. 2011;30:541–6.
40. Barber MD, Walters MD, Cundiff GW. Responsiveness of the pelvic floor distress inventory (PFDI) and pelvic floor impact questionnaire (PFIQ) in women undergoing vaginal surgery and pessary treatment for pelvic organ prolapse. *Am J Obstet Gynecol*. 2006;194:1492–8.
41. Chan SSC, et al. Responsiveness of the pelvic floor distress inventory and pelvic floor impact questionnaire in women undergoing treatment for pelvic floor disorders. *Int Urogynecol J*. 2013;24:213–21.
42. Frick AC, et al. Comparison of responsiveness of validated outcome measures after surgery for stress urinary incontinence. *J Urol*. 2010;184:2013–7.
43. Sánchez-Sánchez B, et al. Responsiveness of the Spanish pelvic floor distress inventory and pelvic floor impact questionnaires short forms (PFDI-20 and PFIQ-7) in women with pelvic floor disorders. *Eur J Obstet Gynecol Reprod Biol*. 2015;190:20–5.
44. Mattsson NK, et al. Agreement between patient global impression scale of improvement, pelvic floor distress inventory and 15D in measuring the outcome of pelvic organ prolapse surgery. In: *Neurourol Urodyn* 1–8; 2020.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.