

The evaluation of pelvic floor muscle strength in women with pelvic floor dysfunction: A reliability and correlation study

Beatriz Navarro Brazález PT, PhD¹  | María Torres Lacomba PT, PhD¹  |
Pedro de la Villa MD, PhD¹ | Beatriz Sánchez Sánchez PT, PhD¹ |
Virginia Prieto Gómez PT, MSc¹ | Ángel Asúnsolo del Barco MD, PhD² |
Linda McLean PT, PhD³

¹ Physical Therapy in Women's Health Research Group, Faculty of Medicine and Health Sciences, Physical Therapy Department, University of Alcalá, Alcalá de Henares, Madrid, Spain

² Preventive Medicine and Health Science Unit, Faculty of Medicine and Health Sciences, Department of Surgery, Medical and Social Sciences, University of Alcalá, Alcalá de Henares, Madrid, Spain

³ Faculty of Health Sciences, School of Rehabilitation Sciences, University of Ottawa, Ottawa, Ontario, Canada

Correspondence

María Torres Lacomba, PT, PhD, Physical Therapy in Women's Health Research Group, Faculty of Medicine and Health Sciences, Physical Therapy Department, University of Alcalá. Campus Externo. Ctra. Madrid-Barcelona km 33.600. 28071. Alcalá de Henares, Madrid, Spain.
Email: maria.torres@uah.es

Aims: The purposes of this study were: (i) to evaluate the reliability of vaginal palpation, vaginal manometry, vaginal dynamometry; and surface (transperineal) electromyography (sEMG), when evaluating pelvic floor muscle (PFM) strength and/or activation; and (ii) to determine the associations among PFM strength measured using these assessments.

Methods: One hundred and fifty women with pelvic floor disorders participated on one occasion, and 20 women returned for the same investigations by two different raters on 3 different days. At each session, PFM strength was assessed using palpation (both the modified Oxford Grading Scale and the *Levator ani* testing), manometry, and dynamometry; and PFM activation was assessed using sEMG.

Results: The interrater reliability of manometry, dynamometry, and sEMG (both root-mean-square [RMS] and integral average) was high (Lin's Concordance Correlation Coefficient [CCC] = 0.95, 0.93, 0.91, 0.86, respectively), whereas the interrater reliability of both palpation grading scales was low (Cohen's Kappa [k] = 0.27-0.38). The intrarater reliability of manometry (CCC = 0.96), and dynamometry (CCC = 0.96) were high, whereas intrarater reliability of both palpation scales (k = 0.78 for both), and of sEMG (CCC = 0.79 vs 0.80 for RMS vs integral average) was moderate. The Bland-Altman plot showed good inter and intrarater agreement, with little random variability for all instruments. The correlations among palpation, manometry, and dynamometry were moderate (coefficient of determination [r^2] ranged from 0.52 to 0.75), however, transperineal sEMG amplitude was only weakly correlated with all measures of strength (r^2 = 0.23-0.30).

Conclusions: Manometry and dynamometry are more reliable tools than vaginal palpation for the assessment of PFM strength in women with pelvic floor disorders,

Mickey Karram led the peer-review process as the Associate Editor responsible for the paper.

especially when different raters are involved. The different PFM strength measures used clinically are moderately correlated; whereas, PFM activation recorded using transperineal sEMG is only weakly correlated with PFM strength. Results from perineal sEMG should not be interpreted in the context of reporting PFM strength.

KEYWORDS

dynamometer, manometer, pelvic floor dysfunction, pelvic floor muscle strength, reliability, surface electromyography, vaginal palpation

1 | INTRODUCTION

The pelvic floor muscles (PFM) appear to act as a functional unit^{1–3} in synergy with the external urethral⁴ and anal² sphincters. Pelvic floor dysfunctions (PFD) refers to a collective of disorders in which the PFM or connective tissues are thought to be implicated, including pelvic organ prolapse (POP), urinary and/or anal incontinence, and dyspareunia.⁵ Despite the difficulty in evaluating PFM function,^{1,6} measuring PFM strength, endurance, and neuromuscular activation are essential to the clinical evaluation of women with PFD in order to direct management.⁵ There is no gold standard for measuring PFM strength nor endurance, yet many tools have been used for this purpose.^{1,5} Vaginal manometry measures maximum vaginal squeeze pressure, and has been shown to be reliable both within^{7,8} and between^{7,9,10} raters, however, it is highly influenced by intra-abdominal pressure.⁵ Vaginal dynamometry is also used to measure PFM strength. Dynamometry is performed through a speculum that measures the anteroposterior peak force generated through contraction of the PFM, and appears to have strong intrarater^{11,12} and interrater agreement,^{13,14} however, the reliability and concurrent validity of a commercially available dynamometry has not been studied.

Vaginal palpation^{1,8,9,14–19} is inexpensive and is used widely in clinical practice to evaluate the quality, power, and endurance of PFM contractions. However, it is subjective and has poor interrater reliability.^{9,10,16} The modified Oxford Grading Scale is a commonly used scale to manually evaluate PFM strength.¹ *Levator ani* testing^{20,21} is a method of vaginal palpation currently used in physical therapy assessment protocols at Spanish, French, and Belgian hospitals, which has not been validated against measures of PFM strength or activation. The *Levator ani* testing uses the same 6-point grading criteria as the modified Oxford Grading Scale to quantify PFM strength, but further considers the quality of the contraction by ensuring that PFM activation can be sustained and repeated (Table 1).

Surface electromyography (sEMG) quantifies the neuromuscular activation of the PFM.^{3,6,22} The between-session reliability of peak sEMG recorded using intravaginal probes is variable,^{23,24} with reports suggesting much better within than between session reliability. Adhesive electrodes located on either side of the anus, or transperineal sEMG is a used approach to the clinical assessment of PFM activation. Such an electrode configuration is not recommended, as the electrodes are large, and located on superficial perineal muscles, thus it lends itself to crosstalk contamination.^{25,26} However, to our knowledge, the reliability, and concurrent validity of sEMG activation of the PFM, when recorded using adhesive electrodes has not been evaluated.²⁶

Some studies have evaluated the reliability and the concurrent validity of different measures of PFM strength.^{7–14,18,19,23,24,27} However, the repeatability and force-generating capacity of PFM contractions may be different between healthy women and those with PFD.²¹ Further, it is not uncommon that several forms of PFD coexist in a single woman.² To our knowledge, no study has tested the reliability and concurrent validity of different methods of assessing PFM strength among women with one or more PFD. The first aim of this study was therefore to determine the intra- and inter-rater reliability of PFM strength measured using vaginal palpation quantified by modified Oxford Grading Scale and *Levator ani* testing, manometry, a commercial dynamometry, and PFM activation amplitude recorded using transperineal sEMG in women who suffer from PFD. Second, we aimed to explore the correlations among measures of PFM strength and activation using these different instruments.

2 | MATERIALS AND METHODS

The study was approved by the Research Committee of the University of Alcala (D2013/003/20130520). All evaluation procedures were conducted at the laboratory of the Physical

TABLE 1 *Levator ani* testing for the vaginal palpation assessment of pelvic floor muscles strength

Grade	Pelvic floor muscles response	Endurance ^a
0	No contraction.	
1	Flicker contraction without displacement.	1 contraction for 1 s.
2	Weak contraction with a slightly displacement.	2 contractions for 2 s.
3	Moderate contraction, inward totally displacement without resistance.	3 contractions for 3 s.
4	Complete contraction against moderate resistance.	4 contractions for 4 s.
5	Strong contraction against strong resistance with complete inward displacement.	5 contractions for 5 s.

^a4-5 s rest between each repetition allowed.

Therapy in Women’s Health Research Group at the University of Alcalá (Madrid, Spain) from May 2013 to September 2015.

2.1 | Participants

Women defined as having PFD based on reporting symptoms of urinary incontinence, anal incontinence, or medically diagnosed stage 1 or 2 POP according to POP Quantification Scheme,²⁸ who were referred for conservative treatment, were invited to participate. Exclusion criteria were aged younger than 18 years, pregnancy, a history of pelvic floor surgery in the previous year, pelvic pain during digital examination, or known or suspected urinary tract infection. The participants who accepted to participate provided written informed consent prior to entering the study.

2.2 | Measurement instruments

The modified Oxford Grading Scale^{9,10,15} and *Levator ani* testing²⁰ (Table 1) were used to quantify PFM strength through vaginal palpation. The physical therapists wore latex or vinyl gloves and inserted the second and third fingers into the vagina. The fingers were abducted slightly (ie, 5 and 8 o’clock position) to palpate the muscles at the lateral vaginal walls. Participants were encouraged to squeeze against the resistance provided by the therapists’ fingers, which resulted in both cranial and anterior forces. The instruction given was “squeeze and lift my fingers as strongly as you can.” For the *Levator ani* testing, participants were asked to perform a strong, sustained, and repeated PFM contraction through the instruction “squeeze and lift my fingers as strongly as you can, and try to maintain the contraction until I say to relax. We will repeat the contraction many times.”

Manometry was assessed using an air filled vaginal probe of 28 mm in diameter, 108 mm total length, and 55 mm active length (Peritron, Melbourne, Australia). The manometry measures pressure in units of cmH₂O with a resolution of 0.1 cmH₂O.

The dynamometer was a commercially available instrumented plastic speculum (Pelvimetre Phenix, Montpellier, France) that incorporates two branches which are inserted into the vagina. The length of the dynamometer arms was 83 mm and each arm had a width of 25 mm, with a total thickness of 24 mm, when the arms were in the closed position. The measurement was performed in the mid-sagittal plane with the dynamometer arms in the closed position. The dynamometer measures force in units of grams (g) with a resolution of 0.1 g.

Both probes were protected by latex or by polyethylene covers. During the manometry and dynamometry assessments, participants were first instructed to remain relaxed, and then to “squeeze and lift the probe as strongly as possible.” The manometer and dynamometer were interfaced with a Phenix USB2 biofeedback system (Vivaltis, Montpellier, France) and an IBM compatible personal computer.

For sEMG evaluations, conductive adhesive hydrogel foam electrodes (Kendall 100 series, Covidien, MA) were placed on the skin overlying both sides of the anus. A reference electrode was located on the skin over the right antero-superior iliac spine.²¹ The amplifier gain was X10 000 (Amplifier AC, CP511, Astr-Med, Inc. Grass Product Group, Warwick) and data were band-pass filtered (10-500 Hz) with an additional notch filter at 50 Hz. Differential sEMG data were digitized using a 16-bit A/D board ([PowerLab 8/30, ADInstruments, Sydney, Australia]) at a sampling rate of 1000 samples per second and a range of ±10 V. All data were processed and analyzed on an IBM compatible personal computer using LabChart 7 software (ADInstruments, Sydney, Australia).

2.3 | Assessment procedure

2.3.1 | Phase I: Reliability study

For the reliability study, a sample size of $n = 20$ was deemed necessary using an alpha of 0.05, a beta of 0.2 (two-sided test), and an expected correlation coefficient of 0.60. The intra and interrater reliability protocol was conducted by two experienced women’s health physical therapists. Physical therapist 1 (PT1) had 20 years of experience, and physical therapist 2 (PT2) had 5 years of experience. Three assessments were performed on each participant. The first (A1) and the second (A2) assessments were performed within a period of 7 days, and were performed by PT1. The third assessment (A3) was performed by PT2, 1-2 days after A2. As such, the A1 and A2 data were used to determine intrarater, between-day reliability; and A2 and A3 data were used to determine interrater reliability. Both raters and participants were blinded to assessment outcomes from previous sessions.

PT1 collected demographic and clinical data through individual interviews, and PFD symptoms were assessed

TABLE 2 Outcomes and reliability metrics for the different Instruments used in this study

Measurement instrument	Assessment 1 (Physical therapist 1)				Assessment 2 (Physical therapist 1)				Assessment 3 (Physical therapist 2)				Interrater reliability A2 vs A3	
	Me	IQR	Min	Max	Me	IQR	Min	Max	Me	IQR	Min	Max	CCC (95%CI) ^a	CCC (95%CI) ^a
Vaginal manometry (cmH ₂ O)	17.40	9.40-39.30	4.60	60.70	20.70	8.20-37.70	6.00	53.40	19.20	10.50-38.60	2.90	54.70	0.95 (0.88-0.98)	0.95 (0.88-0.98)
Vaginal dynamometry (g)	305.50	145.70-580.10	103.70	906.30	276.50	141.30-568.40	98.70	999.30	273.70	142.10-727.70	51.30	974.70	0.93 (0.83-0.97)	0.93 (0.83-0.97)
PFM ^f sEMG ^g RMS ^h (μV)	42.47	27.60-52.37	15.23	71.13	46.88	28.06-73.33	16.40	100.77	52.01	31.67-72.89	19.17	102.70	0.87 (0.70-0.95)	0.87 (0.70-0.95)
PFM ^f sEMG ^g integral (μV/s)	155.45	111.85-220.58	31.80	313.4	157.25	121.30-232.60	42.00	584.50	191.85	138.85-309.52	39.10	400.10	0.92 (0.81-0.97)	0.92 (0.81-0.97)

A1, assessment 1; A2, assessment 2; A3, assessment 3; Me, median; IQR, interquartile range; Min, minimum; Max, maximum; CCC, concordance correlation coefficient; CI, Confidence interval; ^fPFM, Pelvic floor muscles; ^gsEMG, Surface electromyography; ^hRMS, Room mean square.

Measures obtained for manometry, dynamometry, and surface electromyography (both root mean square and integral average) were the mean of three trials for each participant and grouped data are reported as medians.

^aCCC and IC were calculated with logarithmic transformation of variables to adjust them into a normal distribution.

using the Spanish validated version of the Pelvic Floor Distress Inventory Short Form.²⁹ Participants were evaluated in the supine (lithotomy) position. The vaginal palpation assessment was completed first to ensure that women were able to contract their PFM. Following this, the instruments (manometer, dynamometer and sEMG) were used in a randomly assigned order. For each device, three maximum voluntary effort contractions were performed, with 10 s rest between attempts, and with 10 min of rest provided between subsequent instruments. Moreover, an aspect of endurance was tested using sEMG, in which women performed three maximal PFM contractions that were held for 10 s, with a rest of 20 s provided between trials.

2.3.2 | Phase II: Correlation study

For the correlation study, the sample size needed was estimated at $n = 141$, considering a Pearson's correlation coefficient of 0.25, accepting an alpha risk of 0.05, and a beta risk of 0.15 in a two-sided test. Data were acquired exclusively by PT1 using the same procedure described for the reliability study. The primary outcomes included PFM strength assessed through vaginal palpation (modified Oxford grading scale and *Levator ani* testing), manometry, and dynamometry; PFM activation assessed during the strength task (peak root mean square [RMS]), and PFM endurance assessed during the 10 s hold task (integral average). During data collection, observations were noted by PT1 including breath holding and/or contraction of the gluteal, adductor, or abdominal muscles during the PFM contractions, recorded discretely as yes/no in all cases.

2.4 | Data processing

The manometer and dynamometer probes were calibrated after insertion into the vagina and prior to data collection, such that the baseline pressure/force was recorded as zero, when the women attempted to completely relax their PFM, thus eliminating the effect of passive tissue properties or tonic activation. The device software automatically detected peak pressure or force for each contraction attempt. The mean value of the three maximum contractions was entered into the dataset for each instrument. sEMG data were full-wave rectified. To evaluate peak PFM sEMG, the highest RMS value determined over a moving window of 500 ms duration was obtained. To evaluate the endurance, the integral average of the rectified sEMG waveform, recorded over the entire 10 s window was determined. Values were determined using the LabChart software tools. Again, the average of the three trials was used as the outcome for both strength (RMS) and endurance (integral average).

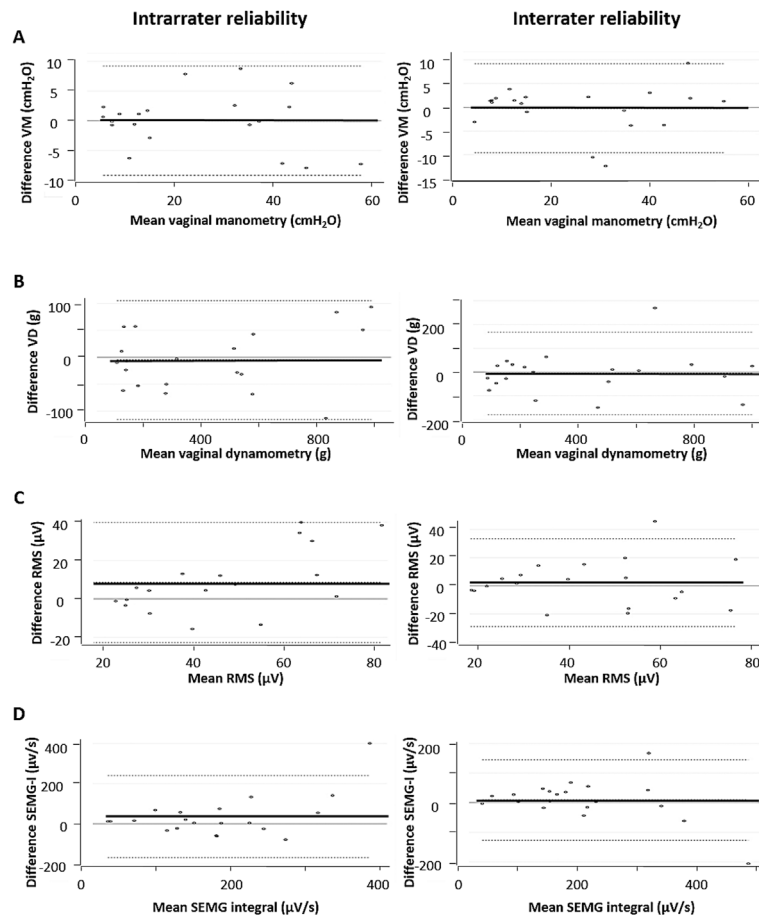


FIGURE 1 Bland-Altman plot of the intrarater and interrater reliability study. Intrarater reliability graphs: The plots show the difference between mean values from the first assessment (A1) and the second assessment (A2), as performed by the first physical therapist (PT1). Interrater reliability graphs: The plots show the differences between mean values obtained by the first physical therapist (PT1-A2) and the second physical therapist (PT2-A3). The dotted horizontal lines represent the mean difference values and the 95% limits of agreement. The continuous horizontal line shows the ideal mean different value (equal to zero), and the thick horizontal line shows the actual mean difference as calculated. (A) Vaginal manometry (VM). (B) Vaginal dynamometry (VD). (C) Surface electromyography root mean square (RMS). (D) Surface electromyography integral average (sEMG-I)

2.5 | Statistical analysis

All statistical analyses were performed using STATA 14.2 SE software. Intra and interrater reliability of the modified Oxford Grading Scale and the *Levator ani* testing were assessed with the Cohen’s Kappa Index (*k*). As the values obtained from the different measurement devices were not normally distributed, Lin’s Concordance Coefficient Correlation (CCC) was calculated from logarithmic transformed variables. Bland-Altman plots were performed using the original data as there is no non-parametric equivalent to the Bland-Altman approach.

Simple linear regression models were used to study the relationships between the different PFM strength/activation measurement methods. The coefficients of determination (r^2) were considered to estimate the goodness-of-fit after checking the residual plots to ensure linearity. Box plots

were performed to compare PFM strength measurements against the palpation ratings (both modified Oxford Grading Scale and *Levator ani* testing). In all cases, 95% confidence intervals (CIs) were computed.

3 | RESULTS

3.1 | Phase I: Reliability study

Twenty-two women were enrolled in the reliability study; however, data from two participants were excluded for technical reasons. The women were aged (mean [SD]) 40 (10) years old, with a mean (SD) body max index of 23.90 (3.50) kg/m². Among the participants, fifteen women (75%) had urinary incontinence, nine (45%) had anal incontinence, and seven (35%) had POP stage 1 or 2. None of the

TABLE 3 Demographic and clinical characteristics of participants' in the correlation study

Participants characteristics	N = 150
Age years, mean (SD)	50 (12)
BMI kg/m ² , mean (SD)	25.70 (4.50)
Menopausal status, n (%)	68 (45)
Pregnancies, mean (SD)	2 (1)
Previous pelvic surgery, n (%)	19 (13)
Pelvic Floor Dysfunction, n (%)	
Urinary incontinence	
SUI	59 (39)
UUI	11 (7)
MUI	39 (26)
Total	109 (72)
Anal incontinence	
Flat I	40 (27)
FI	1 (1)
AI	13 (9)
Total	54 (37)
POP	93 (62)
1PFD	64 (43)
>1PFD	86 (57)
Nearby muscle contraction, n (%)	74 (49)
Abdominal muscles, n (%)	40 (26.5)
Adductor muscles, n (%)	25 (17)
Gluteus muscles, n (%)	58 (38)
Apnea during PFM contraction, n (%)	82 (55)
PFM weekly training, n (%)	95 (63)
Vaginal manometry cmH ₂ O, Me (IQR)	19.60 (12.10-29.60)
Vaginal dynamometry g, Me (IQR)	266.60 (166.70-515.30)
sEMG RMS μV, Me (IQR)	45.55 (26.20-76.91)
sEMG integral μV/s, Me (IQR)	173.85 (104.40-280)

SD, standard deviation; Me, median; IQR, interquartile range N/n, number; BMI, body mass index; PFD, pelvic floor dysfunction; SUI, Stress urinary incontinence; UUI, urgency urinary incontinence; MUI, mixed urinary incontinence; AI, anal incontinence; Flat I, flatulence incontinence; FI, fecal incontinence; POP, pelvic organ prolapse; 1PFD, one pelvic floor dysfunction; >1PFD, more than one pelvic floor dysfunction; PFM, pelvic floor muscle regular training; sEMG, surface electromyography; RMS, root mean square.

participants had POP> stage 2. PFM median (IQR) strength values are shown in Table 2. The Bland-Altman plots showed good agreement, with little random variability for all instruments (Fig. 1). The intrarater reliability of modified Oxford Grading Scale and *Levator ani* testing was high, with $k = 0.78$ for both raters. Between A1(PT1) and A2(PT1), the strength grades were equivalent for 17 out of 20 participants (85%), however, the interrater reliability

(A2[PT1] vs A3[PT2]) of both the modified Oxford Grading Scale and the *Levator ani* testing were low ($k = 0.38$ and 0.27 , respectively). Both physical therapists agreed on 12 out of 20 (60%) assessments in modified Oxford Grading Scale, and 10 out of 20 (50%) assessments in *Levator ani* testing.

3.2 | Phase II: Correlation study

One hundred sixty-eight women were recruited. Three women were unable to participate in the study as they reported vaginal pain during the examination. Three women were excluded for pregnancy, five women were excluded because they were unable to contract their PFM (modified Oxford Grading Scale = 0) and data from seven women were excluded due to technical problems with one of the devices. Data from 150 women were available for the correlation analysis. Participant characteristics are shown in Table 3. Seventy-four (49%) participants were unable to contract PFM without nearby muscle contraction.

Significant correlations were found among modified Oxford Grading Scale, *Levator ani* testing, manometry, dynamometry, and sEMG. The correlation between the two vaginal palpation methods, modified Oxford Grading Scale, and *Levator ani* testing, was moderate ($k = 0.70$) with 77.30% agreement. Modified Oxford Grading Scale scores were moderately correlated with manometry ($r^2 = 0.57$) and with dynamometry ($r^2 = 0.52$) outcomes; whereas the correlation between modified Oxford Grading Scale scores and the RMS of the perineal sEMG was low ($r^2 = 0.25$). The association between *Levator ani* testing and manometry ($r^2 = 0.54$) was moderate, and between *Levator ani* testing and dynamometry ($r^2 = 0.47$) was low. The association between *Levator ani* testing and both endurance (integral average; $r^2 = 0.35$), and activation amplitude (RMS; $r^2 = 0.26$) measured by sEMG was very low. Manometry was moderately correlated with dynamometry ($r^2 = 0.75$), and both were very weakly correlated with sEMG RMS ($r^2 = 0.23-0.29$). Figure 2 shows the relationship between manometry and dynamometry values according to modified Oxford Grading Scale and *Levator ani* testing categories.

4 | DISCUSSION

Our results show that manometry, dynamometry, and transperineal sEMG are reliable instruments for measuring PFM contraction (force or activation) in women with PFD. Vaginal palpation, quantified by modified Oxford Grading Scale or by *Levator ani* testing, showed good intrarater reliability but only fair interrater agreement. Despite the strengths and limitations of each tool, PFM strength measured using the different systems was generally moderately correlated, but correlations between the different devices

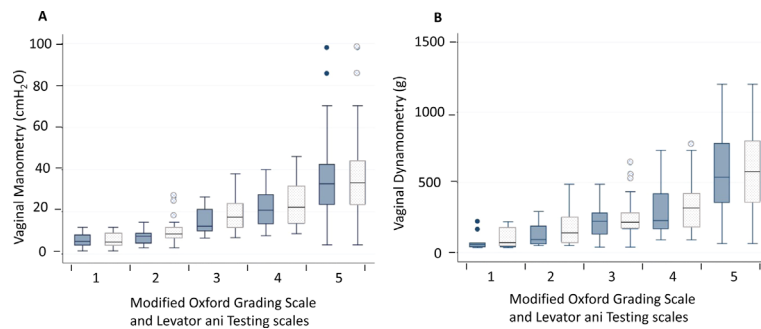


FIGURE 2 Relationship between Modified Oxford Scale and *Levator ani* testing classification and assessments instruments. The box and whisker plots represents vaginal manometry (A), and vaginal dynamometry (B) scores according to the Modified Oxford grading scale and the *Levator ani* testing scale. The plain boxes represent the association with modified Oxford grading scale, and the dotted boxes the associations with *Levator ani* testing. The horizontal thick lines of each box represent the median scores. The isolated points constitute atypical values

used to measure PFM strength and neuromuscular activation of the PFM measured through transperineal sEMG were weak.

Measuring PFM function in women with PFD is challenging as the nature of the disorders themselves may affect the measurement through anatomical changes^{2,30} and/or neuromuscular abnormalities.^{3,22} Previous studies have reported on the reliability of vaginal palpation,^{15,17} manometry,⁸ dynamometry,^{11,14} and sEMG²⁴ in women with urinary incontinence, but not in women with other PFD. The present study included women with different PFD including urinary incontinence, anal incontinence, and POP. Devreese et al explored the interrater reliability of a vaginal palpation assessment of PFM strength in continent and incontinent women and obtained high kappa values for the assessment in both groups ($k = 0.94$). Despite our intrarater reliability also being good ($k = 0.78$), our interrater results differed substantially from those of Devreese. The complexity of the disorders that affected our sample, the quantification scales used, and the differences in rater experience may be reasons for disagreement. Nevertheless, other reliability studies including healthy women, and experienced physical therapists obtained similarly low interrater agreement for the modified Oxford Grading Scale ($k = 0.33^9$ - 0.37^{10}) as obtained in the current study. The implicit subjectivity of measuring PFM strength through vaginal palpation may limit its use,^{9,10} even more so when one considers the higher interrater reliability results obtained using manometry, dynamometry, and sEMG. In any event, the intrarater reliability of palpation was fairly high, suggesting that patient evaluation and re-evaluation by the same physical therapist using the modified Oxford Grading Scale or the *Levator ani* testing may be clinically meaningful in the context of evaluating improvements with therapy.

Methodological variability among different studies may further explain differences in results. The Guidelines for Reporting Reliability and Agreement Studies³¹ suggest that taking an average of several measurements is a better approach to assessing reliability. We calculated the average of

three maximum PFM contractions⁹ for each measure, however, this may have resulted in an underestimation of maximum measurements.¹² Furthermore, there is no specified optimum time between examinations.³¹ Our reliability protocol lasted 7-9 days, with a goal of limiting any functional changes in the strength of the PFM while also ensuring that there was adequate recovery between sessions. Others have performed intrarater assessments the same day,^{7,8,14,24} which may have resulted in higher reliability metrics, especially in sEMG,²³ yet does not reflect the clinical reality that patients are assessed on separate visits.

The comparison between different forms of PFM assessment have previously been reported in incontinent women,^{8,14,19} where manometry has shown good correlation with vaginal palpation results.^{7,8,18} Our correlations between manometry and vaginal palpation were moderate ($r^2 = 0.54$ - 0.57), but lower than reported previously.^{7,18} Both Hundley et al,⁷ and Isherwood and Rane¹⁸ used a discrete scale to quantify manometry (scores ranging from 0 to 12),^{7,18} which may have reduced the responsiveness of manometry while enhancing its reliability and strengthening its relationship with vaginal palpation. In our study, manometry values ranged from 1.40 to 98 cmH₂O, providing a broad range of values over which to calculate correlations. Coupled with the large sample size, our correlation values are likely to be representative of the broader population of women with PFD including urinary incontinence, anal incontinence, and minimal POP.

The correlation between manometry and dynamometry has not previously been reported. Our results showed good ($r^2 = 0.75$) agreement between manometry and dynamometry, when peak values are used, which suggests that both are suitable tools for measuring PFM strength during maximum contraction efforts performed by women with PFD. Non-commercial dynamometry has previously been investigated.^{11-14,19} Consistent with our results, Morin et al¹⁹ reported a Pearson's correlation (r_s) of 0.564 between modified Oxford Grading Scale and dynamometry, and found lower correlation values in incontinent than in continent women ($r_s = 0.450$ vs $r_s = 0.727$).

We hypothesize that, as in Morin et al, our correlation values were lower for women with PFD because they may have larger urogenital hiatuses³⁰ which could affect vaginal closure force recorded at a fixed dynamometer setting. Palpation would presumably be less affected by the size of the urogenital hiatus as the fingers locate the muscles, adjusting for differences in morphology from one participant to the next. Therefore, in women with PFD, defects in the *levator ani*³⁰ and alterations in muscle tone²¹ may impact vaginal force measured by the dynamometer.³⁰ A previous study showed that the best reliability of maximum PFM force generating capacity was achieved for a 24 mm dynamometry opening,¹¹ which is consistent with that used in the current study as the total thickness of our dynamometer was 24 mm, when it was in the closed-position. Nevertheless, in women with PFD, customizing the amount of dynamometer opening based on the size of the urogenital hiatus, or perhaps baseline passive resistance, may improve the validity of PFM strength measurements.

The relationship between intravaginal sEMG and manometry has previously been investigated³²; however, as far as we are aware, the transperineal sEMG approach has not been studied. We found that PFM activation recorded transperineally using sEMG is only weakly correlated with measurements of PFM strength made intravaginally. Others have demonstrated that PFM strength and neuromuscular activation are not linearly related³² and as such, this result is not surprising. Bothelo et al²⁷ explored the correlation between intravaginal sEMG and vaginal palpation in a large group of women with no PFD and found stronger correlations ($\rho = 0.739$). However, depending on the electrode configuration used, measures of PFM activation using intravaginal probes are likely to be less affected by crosstalk,²⁵ and women with PFD may have neurologic impairments,²² which may result in higher sEMG values being recorded at lower force outputs.³³ Such relationships should be explored and compared among groups of women with different PFD. sEMG may be more useful as a screening tool for neuromuscular abnormalities than as a measure of PFM contractile ability.²⁴

Crosstalk contamination⁶ is a large limitation, when measuring PFM sEMG. To mediate this problem, women were taught how to contract their PFM through vaginal palpation, and were given verbal feedback during PFM contractions to correct for unwanted muscle activation, as hip rotators, adductors, or abdominal muscles. Nonetheless, 74 of 150 women (49%) in the current study produced visible contractions of gluteus, hip adductors, and/or abdominal muscles, which may have resulted in crosstalk and underpinned the low correlations between sEMG and the PFM strength outcomes. Madill and McLean³⁴ showed that, women generally co-contract their abdominal muscles to maximize PFM force generating capacity. Based on the lower correlations between transperineal sEMG and PFM force found here compared to previously reported correlations

between intravaginal sEMG and PFM force, it would be useful to compare crosstalk recorded using different sEMG electrode configurations.

Although the PFMs are composed mainly of slow-twitch muscles, presumably needed to support the pelvic organs and promote urethral closure,² PFM endurance is rarely measured.^{15,21} Some authors have measured PFM endurance though recording the duration a sustained maximal contraction can be held^{15–17} however, contraction intensity is difficult to standardize using this approach. The standardization of contraction intensity must be considered in order for contraction time to be meaningful.²¹ In the current study, we measured the integrated EMG over a 10 s maximum effort and this approach produced good intra and interrater reliability (CCC = 0.80–0.92). However, this measure was poorly correlated with *Levator ani* testing ($r^2 = 0.35$), as these techniques measure different aspects of PFM function. The area under the curve measured by manometry or dynamometry may be viable options to measure PFM endurance. The optimal PFM endurance protocol has yet to be established.

5 | CONCLUSION

The results of this study suggest that manometry, dynamometry, and sEMG are reliable instruments for measuring aspects of PFM function in women with early stage POP, anal incontinence, and/or urinary incontinence, and are more reliable than vaginal palpation. The study also demonstrated that manometry and dynamometry are highly correlated, suggesting that both approaches may be valid, when measuring PFM force generating capacity in a supine position. Although vaginal palpation may be useful to verify aspects of morphology and motor control, PFM strength appears to be more accurately measured using manometry and dynamometry than through modified Oxford Grading Scale or *Levator ani* testing. Although it is reliable, transperineal sEMG should not be used to infer information about PFM strength generating capacity in women with PFD.

CONFLICT OF INTEREST

The authors report no conflict of interest.

REFERENCES

1. Bø K, Sherburn M. Evaluation of female pelvic-floor muscle function and strength. *Phys Ther*. 2005;85:269–282.
2. Raizada V, Mittal RK. Pelvic floor anatomy and applied physiology. *Gastroenterol Clin North Am*. 2008;37:493–509, vii.
3. Enck P, Vodusek DB. Electromyography of pelvic floor muscles. *J Electromyogr Kines*. 2006;16:568–577.
4. Bø K, Stein R. Needle EMG registration of striated urethral wall and pelvic floor muscle activity patterns during cough valsalva,

- abdominal, hip adductor, and gluteal muscle contraction in nulliparous healthy females. *Neurol Urodynam*. 1994;13:35–41.
5. Bø K, Frawley HC, Haylen BT, et al. An international urogynecological association (IUGA)/international continence society (ICS) joint report on the terminology for the conservative and nonpharmacological management of female pelvic floor dysfunction. *Neurol Urodynam*. 2016;9999:1–24.
 6. Keshwani N, McLean L. A differential suction electrode for recording electromyographic activity from the pelvic floor muscles: crosstalk evaluation. *J Electromyogr Kinesiol*. 2013;23:311–318.
 7. Hundley AF, Wu JM, Visco AG. A comparison of perineometer to brink score for assessment of pelvic floor muscle strength. *Am J Obstet Gynecol*. 2005;192:1583–1591.
 8. Kerschman-Schind K, Uher E, Wiesinger G, et al. Reliability of pelvic floor muscle strength measurement in elderly incontinent women. *Neurol Urodynam*. 2002;21:42–47.
 9. Ferreira CHJ, Brentegani Barbosa P, de Oliveira Souza F, Ignácio Antônio F, Menezes Franco M, Bø K. Inter-rater reliability study of the modified oxford grading scale and the peritron manometer. *Physiotherapy*. 2011;97:132–138.
 10. Bø K, Finckenhagen HB. Vaginal palpation of pelvic floor muscle strength: inter test reproductibility and comparison between palpation and vaginal squeeze pressure. *Acta Obstet Gynecol Scand*. 2001;80:883–887.
 11. Dumoulin C, Gravel D, Bourbonnais D, Lemieux MC, Morin M. Reliability of dynamometric measurements of the pelvic floor musculature. *Neurol Urodynam*. 2004;23:134–142.
 12. Miller JM, Asthon-Miller J, Perruchicni D, DeLancey JOL. Test-retest reliability of an instrumented speculum for measuring vaginal closure force. *Neurol Urodynam*. 2007;26:858–863.
 13. Martinho NM, Marques J, Silva VR, Silva SLA, Carvalho LC, Bothelo S. Intra and inter-reliability study of pelvic floor muscle dynamometric measurements. *Braz J Phys Ther*. 2015;19:97–104.
 14. Romero-Cullerés G, Peña-Pitarch E, Jané-Feixas C, Arnau A, Montesions J, Abenoza-Guardiola M. Intra-rater reliability and diagnostic accuracy of a new vaginal dynamometer to measure pelvic floor muscle strength in women with urinary incontinence. *Neurol Urodynam*. 2017;36:333–337.
 15. Laycock J, Jerwood D. Pelvic floor muscle assessment: the PERFECT scheme. *Physiotherapy*. 2001;87:631–642.
 16. Sliker-ten Hove MCP, Pool-Goudzwaard AL, Eijkemans MJC, Steegers-Theunissen RPM, Burger CW, Vierhout ME. Face validity and reliability of the first digital assessment scheme of pelvic floor muscle function conform the new standardized terminology of the international continence society. *Neurol Urodynam*. 2009;28:295–300.
 17. Devreese A, Staes F, De Weerd W, et al. Clinical evaluation of pelvic floor muscle function in continent and incontinent women. *Neurol Urodynam*. 2004;23:190–197.
 18. Isherwood PJ, Rane A. Comparative assessment of pelvic floor strength using a perineometer and digital examination. *Br J Obstet Gynaecol*. 2000;107:1007–1011.
 19. Morin M, Dumoulin C, Bourbonnais D, Gravel D, Lemieux MC. Pelvic floor maximal strength using vaginal digital assessment compared to dynamometric measurements. *Neurol Urodynam*. 2004;23:336–341.
 20. Mansoor A, Jacquetin B, Ohana M. Évaluation des facteurs de l'incontinence urinaire féminine et indications thérapeutiques. *Ann Urol (Paris)*. 1993;27:292–305.
 21. Castro-Pardiñas MA, Torres-Lacomba M, Navarro-Brazález B. Pelvic floor muscle function in healthy, postpartum, and pelvic floor dysfunction women. *Actas Urol Esp*. 2017; [Epub ahead of print]. <https://doi.org/10.1016/j.acuro.2017.03.004>
 22. Weidner AC, Barber MD, Visco AG, Bump RC, Sanders DB. Pelvic muscle electromyography of levator ani and external sphincter in nulliparous women and women with pelvic floor dysfunction. *Am J Obstet Gynecol*. 2000;183:1390–1401.
 23. Auchincloss C, McLean L. The reliability of surface EMG recorder from the pelvic floor muscles. *J Neurosci Meth*. 2009;182:85–96.
 24. Glazer H, Romanzi L, Polanezky M. Pelvic floor muscle surface electromyography. reliability and clinical predictive validity. *J Reprod Med*. 1999;44:779–782.
 25. Keshwani N, McLean L. State of art review: intravaginal probes for recording electromyography from the pelvic floor muscles. *Neurol Urodynam*. 2015;34:104–112.
 26. Gentilcore-Saulnier E, Auchincloss CC, McLean L. The Overactive Pelvic floor. In: Padoa A, Rosenbaum T, eds. *Electromyography 1st ed*. Switzerland: Springer; 2016;175–203.
 27. Botelho S, Carvalho-Pereira L, Marques J, et al. Is there correlation between electromyography and digital palpation as means of measuring pelvic floor muscle contractility in nulliparous, pregnant, and postpartum women? *Neurol Urodynam*. 2013;32:420–423.
 28. Haylen BT, Maher CF, Barber MD, et al. An international urogynecological association (IUGA)/international continence society (ICS) joint report on the terminology for female pelvic organ prolapse (POP). *Neurol Urodynam*. 2016;35:137–168.
 29. Sánchez-Sánchez B, Torres-Lacomba M, Yuste-Sánchez M, 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;170:281–285.
 30. DeLancey JOL, Morgan DM, Fenner DE, et al. Comparison of levator ani muscle defects and function in women with and without pelvic organ prolapse. *Obstet Gynecol*. 2007;109:295–302.
 31. Kottner J, Audige L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol*. 2011;64:96–106.
 32. Madill SJ, McLean L. Relationship between abdominal and pelvic floor muscle activation and intravaginal pressure during pelvic floor muscle contractions in healthy continent women. *Neurol Urodynam*. 2006;25:722–730.
 33. Finisterer J. EMG-interference pattern analysis. *J Electromyogr Kines*. 2011;11:231–246.
 34. Madill SJ, McLean L. Quantification of abdominal and pelvic floor muscle synergies in response to voluntary pelvic floor muscle contractions. *J Electromyogr Kines*. 2008;18:955–956.

How to cite this article: Navarro Brazález B, Torres Lacomba M, de la Villa P, et al. The evaluation of pelvic floor muscle strength in women with pelvic floor dysfunction: A reliability and correlation study. *Neurourology and Urodynamics*. 2018;37:269–277. <https://doi.org/10.1002/nau.23287>