Pelvic Floor Ultrasound Imaging: Are Physiotherapists Interchangeable in the Assessment of Levator Hiatal Biometry?

Stéphanie Thibault-Gagnon, BSc(PT);* Evelyne Gentilcore-Saulnier, BSc(PT), MSc;†
Cindy Auchincloss, BSc(PT), MSc;* Linda McLean, BSc(PT), MSc(EE), PhD*

ABSTRACT

Purpose: To evaluate inter-examiner reliability in the ultrasound (US) assessment of levator hiatal dimensions when different physiotherapists perform independent data acquisition and analysis. **Methods:** In this cross-sectional observational study, 14 asymptomatic nulliparous women were imaged at rest, during pelvic floor muscle contraction, and during Valsalva manoeuvre by two physiotherapists using three-dimensional (3D) and four-dimensional (4D) transperineal US. Examiners each measured the dimensions of the levator hiatus (area and antero-posterior and transverse diameters) from the US volumes they respectively acquired. Inter-examiner reliability was determined using intra-class correlation coefficients (ICCs), and inter-examiner agreement was determined using Bland–Altman analyses. **Results:** The ICC results demonstrated very good inter-examiner reliability ($ICC = 0.84–0.98$); Bland–Altman results showed high inter-examiner agreement across all measurements. **Conclusions:** Trained examiners may be considered interchangeable in the US assessment of levator hiatal biometry. Overall, trained physiotherapists using transperineal US imaging to assess levator hiatal biometry can be confident when comparing their own clinical findings to those of their colleagues and to findings published in the literature.

Key Words: pelvic floor; reproducibility of results; three-dimensional imaging; ultrasonography.

RÉSUMÉ

Objectif : Examiner la fiabilité entre examinateurs de l'évaluation échographique des dimensions hiatales du muscle releveur de l'anus lorsque des physiothérapeutes différents procèdent à l'acquisition et à l'analyse indépendante de données. **Méthodes :** Au cours de cette étude par observation transversale, 14 femmes nullipares asymptomatiques ont subi un examen d'imagerie au repos, au cours d'une contraction des muscles du plancher pelvien et durant une manœuvre de Valsalva, effectué par deux physiothérapeutes utilisant l'échographie transpérinéale tridimensionnelle (3D) et quadridimensionnelle (4D). Chaque examinateur a mesuré les dimensions du hiatus du muscle releveur (superficie et diamètre antéropostérieur et transversal) à partir des volumes échographiques qu'ils ont acquis respectivement. On a déterminé la fiabilité entre examinateurs à partir des coefficients de corrélation intra-classe (CCI) et déterminé l'uniformité entre examinateur au moyen des analyses de Bland et d'Altman. **Résultats :** Les résultats de l'application des CCI ont révélé une très bonne fiabilité entre examinateurs pour toutes les mesures. ($CCI = 0.84–0.98$); les résultats de Bland et Altman ont révélé une grande uniformité entre examinateurs pour toutes les mesures. **Conclusions :** Les examinateurs qui ont reçu une formation peuvent être considérés comme interchangeables dans l'évaluation échographique de la biométrie hiatale du muscle releveur. Dans l'ensemble, les physiothérapeutes qui ont suivi une formation et utilisent l'imagerie échographique transpérinéale pour évaluer la biométrie hiatale du muscle releveur peuvent comparer en toute confiance leurs constatations cliniques à celles de leurs collègues et à celles des publications.

Over the past decade, three-dimensional (3D) and four-dimensional (4D) transperineal ultrasound (US) imaging has become a popular tool for pelvic floor assessment in women.^{1,2} By providing access to the axial plane, 3D US imaging has made it possible to image the levator ani muscles, a task previously only possible using magnetic resonance imaging (MRI). Furthermore, 4D US imaging, which involves real-time acquisition of 3D US data, can provide important information on the func-

tional integrity of the levator ani and other pelvic structures during manoeuvres such as pelvic floor muscle contraction and Valsalva. Mainly because of its high costs and limited accessibility, MRI has not been adopted in clinical practice for pelvic floor assessment; moreover, although MRI provides superior spatial resolution and remains the gold standard for evaluating muscle morphology in static states, most MRI systems have poor dynamic assessment capabilities and require that the

From the: *School of Rehabilitation Therapy, Queen's University, Kingston, Ont.; †Faculté de Médecine, Université Laval, Québec, Que.

Correspondence to: Stéphanie Thibault-Gagnon, School of Rehabilitation Therapy, Queen's University, 31 George St., Kingston, ON K7L 3N6; 8st31@queensu.ca.

Contributors: All authors designed the study; collected, analyzed, and interpreted the data; drafted or critically revised the article; and approved the final draft.

Competing Interests: None declared.

Physiotherapy Canada 2014; 66(4):340–347; doi:10.3138/ptc.2013-50

person being assessed remain in supine position. Clinically, therefore, transperineal US is a much more feasible method than MRI for evaluating pelvic floor anatomy and function: it is more easily accessible, is less costly, and has better dynamic assessment capabilities than MRI, and many US systems are portable.

The levator ani muscles are thought to play an important role in pelvic organ support,³ and the levator hiatus, the area bordered by the levator ani and the pubic arch, is considered the largest potential hernial portal in the human body.⁴ Evaluating the size of the levator hiatus has been shown to be of particular importance in assessing female pelvic floor dysfunction, particularly pelvic organ prolapse (POP)^{5,6} and childbirth-related levator trauma.^{7–10} Using transperineal US imaging, studies have shown that the levator hiatal dimensions are strongly⁶ and independently¹¹ associated with symptoms and signs of POP. Furthermore, trauma to the levator ani muscles in the form of an avulsion from their insertions onto the pubic bone, which is also strongly¹² and independently¹¹ associated with POP, is significantly associated with increased hiatal dimensions on Valsalva manoeuvre and with reduced levator contractile ability on pelvic floor muscle contraction, as measured by hiatal constriction.^{4,7} Recent research has used 3D/4D transperineal US imaging to assess outcomes in an empirical investigation of the effectiveness of pelvic floor muscle training for POP.¹³ The use of transperineal US imaging may also be of particular interest for therapists assessing women with urinary and fecal incontinence, given that the first line of treatment for these conditions is pelvic floor muscle training.¹⁴

Compared to electromyography biofeedback, which is widely used clinically in rehabilitating pelvic floor muscle function, US offers a unique level of visual feedback to patients and therapists; it allows them to track changes in the morphology of the levator ani and the biometry of the levator hiatus during pelvic floor muscle contraction and relaxation.¹⁵

In a study using transperineal US imaging to evaluate pelvic floor muscle function and teach pelvic floor muscle exercises, Dietz and colleagues¹⁶ reported that of 212 women examined, 56 (26%) were unable to effectively contract their pelvic floor muscles; however, within 5 minutes of biofeedback training using the US system, 57% (32) of those women who had experienced difficulty contracting their pelvic floor muscles were able to perform an elevating contraction.

Due to these recent advances, pelvic floor US imaging has captured the interest of physiotherapists, who are beginning to appreciate the usefulness of this method for assessing and treating pelvic floor dysfunctions. Because pelvic floor US imaging is gaining popularity as a clinical tool, its psychometric properties must be ascertained.

In the assessment of levator hiatal dimensions, US

measures have been validated against MRI results,¹⁷ and test–retest reliability has been established.^{18–20} However, inter-examiner reliability has been evaluated only in the context of two examiners analyzing stored US data acquired by a single examiner,^{21–24} which cannot account for the effect of operator influence during data acquisition. Optimal investigation of inter-examiner reliability for levator hiatal biometry requires that at least two examiners each perform independent US examinations and thereafter analyze the results.^{22,25}

The purpose of our study, therefore, was to investigate inter-examiner reliability and agreement in the measurement of levator hiatal dimensions when two physiotherapists perform independent US assessments and analyses using the same standardized methods.

METHODS

Participants

The study was approved by Queen's University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board, and all participants provided written informed consent before participating in the study.

Healthy female volunteers were recruited through convenience sampling from the Kingston, Ontario, community. Women were included if they were at least 18 years old and were able to follow instructions in English; they were excluded if they were pregnant or had given birth, or if they reported a history of POP, incontinence, urological or gynecological surgery, physical signs and/or symptoms of menopause (e.g., irregular menstruation, hot flashes), pelvic pain in the past 6 months, or any neurological, neuromuscular, or metabolic condition known to affect muscle contractility. We used telephone screening to determine participant eligibility.

Ultrasound procedure

During a single session, two physiotherapists, Examiner 1 (EGS) and Examiner 2 (STG), performed 3D and 4D transperineal US imaging in a randomized order with each participant. Each examiner had at least 6 months of experience using these methods at the time of the study. Imaging was performed using a GE Voluson i ultrasound system (GE Healthcare Austria GmbH & Co. OG, Zipf, Austria) with an 8–4MHz curved array 3D transducer used in 4D Real Time acquisition mode. The transducer was placed transperineally with the main transducer axis oriented in the mid-sagittal plane, and the volume acquisition angle was set to its maximum of 85 degrees. After the bladder was emptied, imaging was performed with the participant in supine position with her hips slightly abducted and flexed approximately 30 degrees, knees flexed approximately 30 degrees, and feet supported in stirrups. Her trunk was lifted and supported by pillows in a way that she could comfortably see and communicate with the investigator. Prior to data collection, participants were trained on how to perform a

Figure 1 (i) Mid-sagittal plane image. The plane of minimal antero-posterior (AP) diameter is indicated by the dashed line joining the posterior-inferior aspect of the symphysis (SP) and the inner border of the *levator ani* muscle at the anorectal junction. (ii) Axial plane image. U urethra, V vagina, A anus. (iii) Measurements in the axial plane of minimal AP diameter. AP diameter of the levator hiatus (LH) (dashed line), transverse diameter of the LH (dotted line), area of the LH (area bordered by the solid tracing).

Figure 2 Mid-sagittal plane image and axial plane image in the plane of minimal antero-posterior diameter taken (i) at rest, (ii) on maximal pelvic floor muscle contraction, (iii) on maximal Valsalva.

correct pelvic floor muscle contraction and a correct Valsalva manoeuvre using US biofeedback teaching when necessary. A correct pelvic floor muscle contraction was defined as producing an antero-cranial displacement

of the anorectal angle (ARA), while a correct Valsalva manoeuvre produced a postero-caudal displacement of the ARA, which could be ascertained by inspection of the mid-sagittal plane image (see Figure 1).

For data acquisition, participants were instructed to keep their pelvic floor muscles relaxed while the examiner acquired three consecutive US volumes at rest. Volumes were then acquired while participants performed three repetitions of a maximal-effort pelvic floor muscle contraction and three repetitions of a maximal-effort Valsalva manoeuvre. Both examiners gave the same standardized instructions and encouragement. Data acquisition continued throughout each task until maximal cranial and maximal caudal displacement of the ARA was attained for contraction and Valsalva respectively, as shown by visual inspection of the mid-sagittal plane image (see Figure 2). Participants generally attained maximal contraction within 5 seconds and maximal Valsalva within 10–15 seconds. Examiners made sure to maintain the symphysis pubis and the ARA in view in the mid-sagittal plane image, as these landmarks are required to be observed during analysis.

Data analysis

Analyses of volume data sets were conducted offline using 4D View version 7.0 software (GE Healthcare Austria GmbH & Co. OG, Zipf, Austria). Each examiner analyzed the volumes she had acquired for each participant and was blinded to the other examiner's measurements. Examiners obtained the following measurements in the axial plane image corresponding to the plane of minimal antero-posterior (AP) diameter of the levator hiatus: the AP diameter of the levator hiatus (LH_{AP}), the right-left transverse diameter of the levator hiatus (LH_T), and the area of the levator hiatus (LH_{area}) (see Figure 1) according to the protocol first described by

Dietz and colleagues¹⁸ and cited in several subsequent studies.^{4–13,15–17,19–24,26–30} Examiners measured each of the three volumes acquired during each task (rest, contraction, Valsalva) with each participant and calculated average measurements for each measure according to each task. Statistical analyses were performed using Excel version 12.2.7 (Microsoft Corp., Redmond, WA) and SPSS version 18.0 (SPSS Inc., Chicago, IL). Measurements and inter-examiner paired differences in measurement were tested for normality using the Shapiro-Wilk test.

The intra-class correlation coefficient (ICC) was used to determine inter-examiner reliability separately for each US measure according to each task. ICC values were computed according to Shrout and Fleiss's ICC model (2,k)³¹ (two-way random model), since examiners were considered to represent a random sample from a larger population of examiners, and ICC estimates were computed for the mean of k measurements. In addition, because examiner variability was considered relevant, we chose "absolute agreement" rather than "consistency" as the level of agreement in SPSS. To determine the level of reliability, we used Altman's scale for the classification of ICC values,³² whereby ICC values <0.20 are considered poor; 0.21–0.41, fair; 0.41–0.60, moderate; 0.61–0.80, good; and 0.81–1.00, very good.

Since a limitation of the ICC is its clinical interpretability, as it does not indicate the magnitude of agreement or disagreement between measurements, we also used the Bland–Altman methods to investigate inter-examiner agreement.³³ Whereas *inter-examiner reliability* can be defined as the degree to which two or more examiners are able to differentiate among subjects under similar assessment conditions, *inter-examiner agreement* pertains to the degree to which two or more examiners achieve identical results under similar assessment conditions.³⁴ Calculations included the mean of differences in the measurements obtained by each examiner ($\bar{d}$), the 95% CI (CI) and the standard error (SE) for $\bar{d}$, the standard deviation of the differences (SD_{diff}), and the 95% limits of agreement for each measure across all tasks. A paired *t*-test of the measurements obtained by each examiner was used to test for the presence of significant bias between examiners ($\alpha = 0.05$ for all tests).

RESULTS

A total of 14 asymptomatic nulliparous female volunteers with a mean age of 27 (SD 6; range 22–39) years and a mean body mass index of 22.4 (SD 3.0) kg/m² participated in the study. All participants were able to correctly perform a pelvic floor muscle contraction. Despite training using pelvic floor US as biofeedback, five participants were unable to perform a Valsalva manoeuvre without concurrent levator contraction, as evidenced by the antero-cranial displacement of the ARA during the task; both examiners independently recognized this in-

Table 1 Dimensions of the Levator Hiatus as Measured by Each Examiner

Measure	Mean (SD)	
	Examiner 1	Examiner 2
Rest (<i>n</i> = 14)		
LH _{AP} , cm	5.57 (0.72)	5.55 (0.70)
LH _T , cm	3.82 (0.39)	3.91 (0.49)
LH _{area} , cm ²	14.76 (3.21)	15.06 (3.24)
Contraction (<i>n</i> = 14)		
LH _{AP} , cm	4.66 (0.50)	4.53 (0.42)
LH _T , cm	3.56 (0.47)	3.61 (0.46)
LH _{area} , cm ²	12.22 (2.38)	11.87 (2.15)
Valsalva (<i>n</i> = 9)		
LH _{AP} , cm	6.15 (1.11)	5.95 (1.05)
LH _T , cm	4.21 (0.53)	4.42 (0.70)
LH _{area} , cm ²	19.64 (5.39)	19.57 (5.73)

LH = levator hiatus; LH_{AP} = anteroposterior diameter of the LH;

LH_T = right-left transverse diameter of the LH; LH_{area} = area of the LH.

ability in all five participants. Valsalva data for these participants were therefore excluded from our analyses.

Table 1 reports the measurements obtained by each examiner across all participants. All measurements, as well as inter-examiner paired differences in measurement, were normally distributed. Results from the inter-examiner reliability and inter-examiner agreement analyses are presented in Table 2. ICC values demonstrated very good inter-examiner reliability (ICCs = 0.84–0.98) across all measures and tasks. Paired *t*-test results for the measurements obtained by each examiner demonstrated no evidence of a significant inter-examiner bias for any of the measurements ($p \geq 0.05$); moreover, for all measurements, the 95% CI for $\bar{d}$ included zero, also which indicates that there was no systematic bias between the two examiners for any of the measures or tasks.^{35,36}

DISCUSSION

Our findings demonstrate that US measures of levator hiatal biometry are highly reproducible when two physiotherapists independently acquire and analyze volume data sets, indicating that operator influence during US data acquisition may be negligible when examiners use the same standardized methods. The overall ICC results indicated very good inter-examiner reliability in the measurement of levator hiatal dimensions at rest, at maximal pelvic floor muscle contraction, and at maximal Valsalva. Analysis of inter-examiner agreement using Bland–Altman methods showed no significant systematic bias between examiners in the measurement of levator hiatal dimensions and provided estimates for expected disagreements in measurements obtained by each examiner.

Inter-examiner reliability and agreement analyses

The ICC is a reliability index calculated using the variance estimates obtained through an ANOVA. The appropriate type of ICC model to use depends on the

Table 2 Inter-examiner Reliability and Agreement—Intra-class Correlation Coefficients (2,k) and Bland–Altman Test Results

Measure	ICC		Bland–Altman tests			
	ICC (95% CI)	$\bar{d}$	SD _{diff}	95% CI of $\bar{d}$	SE of $\bar{d}$	95% limits of agreement
Rest (<i>n</i> = 14)						
LH _{AP} , cm	0.98 (0.92–0.99)	0.02	0.23	–0.11–0.15	0.06	–0.47–0.52
LH _T , cm	0.94 (0.82–0.98)	–0.09	0.19	–0.20–0.02	0.05	–0.51–0.33
LH _{area} , cm ²	0.92 (0.67–0.98)	–0.30	1.09	–0.93–0.33	0.29	–2.65–2.06
Contraction (<i>n</i> = 14)						
LH _{AP} , cm	0.84 (0.52–0.95)	0.13	0.34	–0.07–0.32	0.09	–0.60–0.85
LH _T , cm	0.93 (0.80–0.98)	–0.05	0.24	–0.19–0.09	0.06	–0.56–0.46
LH _{area} , cm ²	0.90 (0.69–0.97)	0.35	1.38	–0.45–1.15	0.37	–2.63–3.33
Valsalva (<i>n</i> = 9)						
LH _{AP} , cm	0.97 (0.85–0.99)	0.21	0.30	–0.02–0.44	0.10	–0.49–0.90
LH _T , cm	0.88 (0.48–0.97)	–0.22	0.38	–0.51–0.08	0.13	–1.09–0.66
LH _{area} , cm ²	0.98 (0.89–0.99)	0.07	1.83	–1.34–1.48	0.61	–4.16–4.30

ICC = intra-class correlation coefficient; $\bar{d}$ = mean difference in the measurements obtained by each examiner (inter-examiner bias); SD_{diff} = standard deviation of the differences; SE = standard error; LH = levator hiatus; LH_{AP} = anteroposterior diameter of the LH; LH_T = right-left transverse diameter of the LH; LH_{area} = area of the LH.

study design and the conceptual intent of the study.³¹ When ICCs are used, the type of ICC chosen and the rationale for using it must be clearly defined. However, many authors fail to report the form of ICC used, as is the case in the inter-examiner reliability studies mentioned previously, which limits the ability to interpret results and makes comparisons between studies difficult. In our study, inter-examiner reliability was investigated using Shrout and Fleiss's ICC model (2,k),³¹ with the assumption that the examiners were randomly chosen and expected to represent the population of examiners using the US methods described above.³⁷ Our aim was to be able to generalize the results to other examiners with similar characteristics. The ICC is easy to interpret and allows us to compare our results to those of previous studies. An important characteristic and limitation of the ICC is that it is highly influenced by the heterogeneity of the population investigated; the more variable the population, the higher the ICC values.²⁵ The high ICC values found in our study, therefore, are likely due in part to the substantial variability within our study population. Lastly, ICC results cannot be interpreted clinically, as they give no indication of the magnitude of agreement or disagreement between measurements. For these reasons, we used Bland–Altman methods to investigate inter-examiner agreement and bias and to obtain limits of agreement that can be used as the 95% CI for expected disagreements between examiners. When two different examiners assess the same person and the difference in their measurements lies within the limits of agreement, this difference can be attributed to the expected disagreement between examiners rather than indicating a change in the person between the two examinations.

The ICC results obtained in our study are comparable to those reported in four previous studies that found

good to very good inter-examiner reliability (ICC range 0.72–0.96) in the measurement of levator hiatus dimensions when two examiners analyzed volume data sets acquired by a single examiner.^{21–24} In two of these studies,^{21,22} Bland–Altman methods were also used to investigate inter-examiner reliability and agreement. Only Majida and colleagues²² reported finding a significant inter-examiner bias and only in the measurement of LH_{AP} at rest and on contraction. Thyer and colleagues²¹ did not report on whether or not they found a significant bias for any of the measures. When the same measures are compared, the limits of agreement reported by both Majida and colleagues²² and Thyer and colleagues²¹ were wider than those we observed; this finding suggests better agreement between examiners in our study despite the fact that our examiners acquired their own images for analysis. The high level of reliability and agreement between examiners in our study may reflect the fact that both examiners had similar experience and were trained to use the standardized protocol for data acquisition and analysis together. A lower level of agreement between examiners might result if examiners with different experience and/or training were compared.

Researchers sometimes report having had more than one examiner involved in data acquisition and/or analysis.^{17,21} The results of our study suggest that if examiners have similar training and experience in acquiring and analyzing US data, using multiple examiners may not add significant bias or variance to clinical or experimental results. That said, when more than one examiner acquires US images or measures the biometry of the levator hiatus, inter-examiner reliability should first be established. When making decisions as to whether examiners' measurements agree sufficiently, one must take into account the context in which the measurements are

to be used.³⁸ Provided that differences between examiners' measurements within the limits of agreement are not clinically important, the examiners can be considered interchangeable. At this time, we are unable to define the maximum acceptable interval for the limits of agreement for measures of levator hiatal biometry, since the existing literature provides scant data regarding what magnitudes of differences constitute clinically important changes. The limits of agreement obtained in our study indicate the magnitude of disagreements that can be expected between the two examiners' measurements on 95% of occasions; hypothetically, in a given study comparing pre- and post-intervention measures in the same sample of participants where the same two examiners acquire and analyze the data, measurement differences from pre- to post-intervention would need to lie outside of the limits of agreement in order for investigators to judge whether differences are genuine or due to measurement error.

Exclusion of Valsalva data

Valsalva data were excluded from our analysis for five participants who, despite training, persisted in demonstrating levator contraction during task performance. This finding is consistent with previous research showing that levator activation occurs during Valsalva in almost half of nulliparous women before training and persists in half of these women even after US biofeedback training.³⁹ Others have also reported excluding Valsalva data from analysis because of confounding levator activation during task performance.²¹

The main strength of our study is its design: to our knowledge, this is the first study to investigate inter-examiner reliability when two examiners perform independent US examinations and thereafter analyze volume data sets to obtain measures of levator hiatal biometry. However, certain limitations must be recognized. The short period (5 minutes) allotted between the different examiners' assessments may be considered a limitation of the study. Given the high reproducibility in measurements found between examiners, it is plausible that participants were able to perform more consistently because they recalled the first examiner's instructions when performing the tasks for the second examiner. This training effect may have reduced within-subject error and thus positively affected inter-examiner reliability. We tried to minimize this effect by randomizing the order in which examiners performed the imaging assessments.

Another limitation of our study is that it included only a small sample of asymptomatic women. On the one hand, it is possible that the small sample size may have prevented us from identifying a significant systematic bias between examiners due to a lack of statistical power; given the results, however, we would expect that if a systematic bias between examiners became apparent

with a larger sample size, it would not be a large bias. Bland and Altman have previously demonstrated that excellent agreement is possible even where there is significant, but small, bias.³³ On the other hand, although our results demonstrated very good inter-examiner reliability and agreement across measurements, we would anticipate that with a larger sample size, the limits of agreement obtained for each measure would in fact be narrower, demonstrating better agreement between examiners.^{33,40}

Lastly, our results do not reflect inter-examiner reliability in the assessment of levator hiatal biometry in women with pelvic floor dysfunction (e.g., POP, incontinence).

CONCLUSION

Pelvic floor US imaging will likely continue to gain clinical popularity in physiotherapy practice as an assessment and rehabilitation (biofeedback) tool. Canadian physiotherapists should have access to training courses that provide clinicians with technical knowledge and expertise as well as evidence-based protocols to standardize practices across the country.

Overall, this study demonstrates that trained physiotherapists using 3D/4D transperineal US imaging to assess measures of levator hiatal biometry can be confident when comparing their clinical findings to their colleagues' findings and to those published in the literature. Given the growing use of pelvic floor US imaging in both research and clinical settings, future studies should aim to investigate the reliability of US measures of levator ani morphology and function in female populations with pelvic floor dysfunction and to define clinically important changes in measures following interventions such as pelvic floor muscle training.

KEY MESSAGES

What is already known on this topic

In the assessment of levator hiatal dimensions using transperineal ultrasound (US) imaging, test-retest reliability has been demonstrated, and inter-examiner reliability has been shown, but these dimensions have been evaluated only when two examiners analyzed stored US data acquired by a single examiner.

What this study adds

In the assessment of levator hiatal dimensions using transperineal US imaging, this study is the first to report on and demonstrate inter-examiner reliability when two examiners perform separate US examinations and thereafter independently analyze the resultant data.

REFERENCES

1. Santoro GA, Wiczorek AP, Dietz HP, et al. State of the art: an integrated approach to pelvic floor ultrasonography. *Ultrasound Obstet Gynecol.* 2011;37(4):381–96. <http://dx.doi.org/10.1002/uog.8816>. Medline:20814874

2. Dietz HP. Pelvic floor ultrasound: a review. *Am J Obstet Gynecol.* 2010;202(4):321–34. <http://dx.doi.org/10.1016/j.ajog.2009.08.018>. Medline:20350640
3. Delancey JOL. Anatomy. In: Cardozo L, Staskin D, editors. *Textbook of female urology and urogynecology*. London: Isis Medical Media; 2001. p. 112–24.
4. Shek KL, Dietz HP. The effect of childbirth on hiatal dimensions. *Obstet Gynecol.* 2009;113(6):1272–8. <http://dx.doi.org/10.1097/AOG.0b013e3181a5ef23>. Medline:19461422
5. Dietz HP. Pelvic floor ultrasound in prolapse: what's in it for the surgeon? *Int Urogynecol J.* 2011;22(10):1221–32. <http://dx.doi.org/10.1007/s00192-011-1459-3>. Medline:21656320
6. Dietz HP, Shek C, De Leon J, et al. Ballooning of the levator hiatus. *Ultrasound Obstet Gynecol.* 2008;31(6):676–80. <http://dx.doi.org/10.1002/uog.5355>. Medline:18470963
7. Abdool Z, Shek KL, Dietz HP. The effect of levator avulsion on hiatal dimension and function. *Am J Obstet Gynecol.* 2009;201(1):e1–5. <http://dx.doi.org/10.1016/j.ajog.2009.02.005>. Medline:19426956
8. Shek KL, Dietz HP. Intrapartum risk factors for levator trauma. *BJOG.* 2010;117(12):1485–92. <http://dx.doi.org/10.1111/j.1471-0528.2010.02704.x>. Medline:20735379
9. Shek KL, Dietz HP. Can levator avulsion be predicted antenatally? *Am J Obstet Gynecol.* 2010;202(6):e1–6. <http://dx.doi.org/10.1016/j.ajog.2009.11.038>. Medline:20079479
10. Albrich SB, Laterza RM, Skala C, et al. Impact of mode of delivery on levator morphology: a prospective observational study with three-dimensional ultrasound early in the postpartum period. *BJOG.* 2012;119(1):51–61. <http://dx.doi.org/10.1111/j.1471-0528.2011.03152.x>. Medline:21985531
11. Dietz HP, Franco AVM, Shek KL, et al. Avulsion injury and levator hiatal ballooning: two independent risk factors for prolapse? An observational study. *Acta Obstet Gynecol Scand.* 2012;91(2):211–4. <http://dx.doi.org/10.1111/j.1600-0412.2011.01315.x>. Medline:22050558
12. Dietz HP, Simpson JM. Levator trauma is associated with pelvic organ prolapse. *BJOG.* 2008;115(8):979–84. <http://dx.doi.org/10.1111/j.1471-0528.2008.01751.x>. Medline:18503571
13. Braekken IH, Majida M, Engh ME, et al. Morphological changes after pelvic floor muscle training measured by 3-dimensional ultrasonography: a randomized controlled trial. *Obstet Gynecol.* 2010;115(2 Pt 1):317–24. <http://dx.doi.org/10.1097/AOG.0b013e3181cbdb35f>. Medline:20093905
14. Hay Smith J, Berghmans B, Burgio K, et al. Adult conservative management. In: Abrams P, Cardozo L, Khoury S, et al, editors. *Incontinence*. 4th ed. Paris: Health Publications Ltd; 2009. p. 1025–120.
15. Thompson JA, O'Sullivan PB, Briffa NK, et al. Assessment of voluntary pelvic floor muscle contraction in continent and incontinent women using transperineal ultrasound, manual muscle testing and vaginal squeeze pressure measurements. *Int Urogynecol J Pelvic Floor Dysfunct.* 2006;17(6):624–30. <http://dx.doi.org/10.1007/s00192-006-0081-2>. Medline:16532264
16. Dietz HP, Wilson PD, Clarke B. The use of perineal ultrasound to quantify levator activity and teach pelvic floor muscle exercises. *Int Urogynecol J Pelvic Floor Dysfunct.* 2001;12(3):166–9. <http://dx.doi.org/10.1007/s001920170059>. Medline:11451004
17. Kruger JA, Heap SW, Murphy BA, et al. Pelvic floor function in nulliparous women using three-dimensional ultrasound and magnetic resonance imaging. *Obstet Gynecol.* 2008;111(3):631–8. <http://dx.doi.org/10.1097/AOG.0b013e3181655dc2>. Medline:18310365
18. Dietz HP, Shek C, Clarke B. Biometry of the pubovisceral muscle and levator hiatus by three-dimensional pelvic floor ultrasound. *Ultrasound Obstet Gynecol.* 2005;25(6):580–5. <http://dx.doi.org/10.1002/uog.1899>. Medline:15883982
19. Braekken IH, Majida M, Engh ME, et al. Test-retest reliability of pelvic floor muscle contraction measured by 4D ultrasound. *Neuro-urol Urodyn.* 2009;28(1):68–73. <http://dx.doi.org/10.1002/nau.20618>. Medline:18932174
20. Braekken IH, Majida M, Ellström-Engh M, et al. Test-retest and intra-observer repeatability of two-, three- and four-dimensional perineal ultrasound of pelvic floor muscle anatomy and function. *Int Urogynecol J.* 2008;19(2):227–35. <http://dx.doi.org/10.1007/s00192-007-0408-7>.
21. Thyer I, Shek C, Dietz HP. New imaging method for assessing pelvic floor biomechanics. *Ultrasound Obstet Gynecol.* 2008;31(2):201–5. <http://dx.doi.org/10.1002/uog.5219>. Medline:18254157
22. Majida M, Braekken IH, Umek W, et al. Interobserver repeatability of three- and four-dimensional transperineal ultrasound assessment of pelvic floor muscle anatomy and function. *Ultrasound Obstet Gynecol.* 2009;33(5):567–73. <http://dx.doi.org/10.1002/uog.6351>. Medline:19402120
23. Santoro GA, Wieczorek AP, Shobeiri SA, et al. Interobserver and interdisciplinary reproducibility of 3D endovaginal ultrasound assessment of pelvic floor anatomy. *Int Urogynecol J.* 2011;22(1):53–9. <http://dx.doi.org/10.1007/s00192-010-1233-y>. Medline:20700728
24. Weinstein MM, Jung SA, Pretorius DH, et al. The reliability of puborectalis muscle measurements with 3-dimensional ultrasound imaging. *Am J Obstet Gynecol.* 2007;197(1):e1–6. <http://dx.doi.org/10.1016/j.ajog.2007.02.041>. Medline:17618762
25. Rovas L, Sladkevicius P, Strobel E, et al. Intraobserver and inter-observer reproducibility of three-dimensional gray-scale and power Doppler ultrasound examinations of the cervix in pregnant women. *Ultrasound Obstet Gynecol.* 2005;26(2):132–7. <http://dx.doi.org/10.1002/uog.1884>. Medline:15959922
26. Yang JM, Yang SH, Huang WC. Biometry of the pubovisceral muscle and levator hiatus in nulliparous Chinese women. *Ultrasound Obstet Gynecol.* 2006;28(5):710–6. <http://dx.doi.org/10.1002/uog.3825>. Medline:17001756
27. Dietz HP, Steensma AB, Vancanillie TG. Levator function in nulliparous women. *Int Urogynecol J Pelvic Floor Dysfunct.* 2003;14(1):24–6, discussion 26. <http://dx.doi.org/10.1007/s00192-002-1001-8>. Medline:12601512
28. Dietz HP. Levator function before and after childbirth. *Aust N Z J Obstet Gynaecol.* 2004;44(1):19–23. <http://dx.doi.org/10.1111/j.1479-828X.2004.00140.x>. Medline:15089863
29. Kruger JA, Dietz HP, Murphy BA. Pelvic floor function in elite nulliparous athletes. *Ultrasound Obstet Gynecol.* 2007;30(1):81–5. <http://dx.doi.org/10.1002/uog.4027>. Medline:17497753
30. Dietz HP, Hyland G, Hay-Smith J. The assessment of levator trauma: a comparison between palpation and 4D pelvic floor ultrasound. *Neuro-urol Urodyn.* 2006;25(5):424–7. <http://dx.doi.org/10.1002/nau.20250>. Medline:16721815
31. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86(2):420–8. <http://dx.doi.org/10.1037/0033-2909.86.2.420>. Medline:18839484
32. Altman DG. *Practical statistics for medical research*. London: Chapman and Hall; 1991.
33. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;327(8476):307–10. [http://dx.doi.org/10.1016/S0140-6736\(86\)90837-8](http://dx.doi.org/10.1016/S0140-6736(86)90837-8). Medline:2868172
34. Kottner J, Audigé L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011;64(1):96–106. <http://dx.doi.org/10.1016/j.jclinepi.2010.03.002>. Medline:21130355
35. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ.* 1992;304(6840):1491–4. <http://dx.doi.org/10.1136/bmj.304.6840.1491>. Medline:1611375

36. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil*. 1998;12(3):187–99. <http://dx.doi.org/10.1191/026921598672178340>. Medline:9688034
37. Portney LG, Watkins MP. Statistical measures of reliability. In: Portney LG, Watkins MP, editors. *Foundation of clinical research: application to practice*. 3rd ed. Upper Saddle River (NJ): Pearson Prentice Hall; 2009. p. 585–618.
38. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstet Gynecol*. 2008;31(4):466–75. <http://dx.doi.org/10.1002/uog.5256>. Medline:18306169
39. Örnö AK, Dietz HP. Levator co-activation is a significant confounder of pelvic organ descent on Valsalva maneuver. *Ultrasound Obstet Gynecol*. 2007;30(3):346–50. <http://dx.doi.org/10.1002/uog.4082>. Medline:17702054
40. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput Biol Med*. 1990;20(5):337–40. [http://dx.doi.org/10.1016/0010-4825\(90\)90013-F](http://dx.doi.org/10.1016/0010-4825(90)90013-F). Medline:2257734