

The repeatability of sonographic measures of functional pelvic floor anatomy

Li Tan¹ · Ka Lai Shek^{2,3} · Ixora Kamisan Atan³ · Rodrigo Guzman Rojas^{4,5} · Hans Peter Dietz³

Received: 1 February 2015 / Accepted: 1 June 2015 / Published online: 14 June 2015
© The International Urogynecological Association 2015

Abstract

Introduction and hypothesis Translabial 3D/4D ultrasound is increasingly being used in the diagnostic evaluation of pelvic floor dysfunction. The result of the assessment is influenced by a number of confounders that are generally unrecognised. The aim of this study was to determine the short- to medium-term repeatability of translabial ultrasound measures of female pelvic organ support and pelvic floor anatomy.

Methods This is a retrospective study analyzing archived ultrasound volume datasets of 106 patients with pelvic floor dysfunction. Every subject was assessed twice at an average interval of 73 days. Outcome measures including hiatal area on Valsalva, descent of the bladder neck, bladder, uterus and rectal ampulla, rectocele depth, diagnosis of true rectocele, and levator integrity (avulsion) were compared at the first and second appointments.

Results All parameters of organ descent demonstrated good to excellent reliability (ICC 0.73–0.93) except for rectocele descent, which showed moderate reliability (ICC 0.44, CI 0.26–

0.58). The most highly repeatable measure was hiatal area on Valsalva or “ballooning” (ICC 0.93, CI 0.90–0.95). For the diagnosis of levator avulsion and true rectocele, agreement was very high (kappa 0.91 for avulsion (CI 0.77–0.94) and kappa 0.73 (CI 0.56–0.84) for true rectocele).

Conclusions The short- to medium-term repeatability of translabial ultrasound measures of functional pelvic floor anatomy seems to be high. Hiatal area on Valsalva (ballooning) and diagnosis of levator avulsion were the most repeatable measures. The least repeatable measures related to the posterior compartment.

Keywords Repeatability · Female pelvic organ prolapse · Translabial ultrasound · Pelvic floor · Levator avulsion

Introduction

Ultrasound is increasingly used in the diagnostic evaluation of patients with symptoms and signs of pelvic floor dysfunction [1, 2] and is the most appropriate form of imaging in urogynecology for reasons of cost, access and performance, and because it provides information in real time. Translabial (or “perineal”, or “transperineal”) 3D/4D imaging is particularly useful in assessing women with pelvic organ prolapse, since 3D ultrasound gives easy access to the axial plane, which was previously only possible with the use of magnetic resonance imaging (MRI).

To date, a number of studies have demonstrated moderate to almost perfect repeatability for measurements of organ prolapse and pelvic floor anatomy using translabial 3D ultrasound [3–9]. However, these studies commonly focus on the repeatability of the offline analysis of stored volume data sets and of measurements obtained on the same day. Hence, there are few data on the agreement between two independent

✉ Hans Peter Dietz
hpdietz@bigpond.com

¹ Department of Ultrasound, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing 100730, China

² Liverpool Hospital, University of Western Sydney, Sydney, NSW, Australia

³ Sydney Medical School Nepean, University of Sydney, Nepean Hospital, Penrith, NSW 2750, Australia

⁴ Departamento de Ginecología y Obstetricia, Facultad de Medicina, Clínica Alemana, Universidad del Desarrollo, Santiago, Chile

⁵ Departamento de Ginecología y Obstetricia, Hospital Clínico de la Universidad de Chile, Santiago, Chile

measurements obtained at longer intervals. This is relevant because of the potential confounding effect of bowel filling and stool consistency, neuromuscular activation state, and the varying efficacy of patient instruction and investigator performance. It is recognised that the repeatability of measures is influenced not only by the quality of post-processing, but also by the quality of volume acquisition at the time of the examination.

Owing to the nature of clinical practice at our unit, patients undergoing surgical management for pelvic floor disorders are assessed twice, once when they are first seen in our clinic, and a second time when they undergo urodynamic testing prior to surgical consent, or occasionally in reverse order. In this study, we assessed the short- to medium-term repeatability of translabial ultrasound measures of pelvic floor anatomy by comparing archived volume ultrasound datasets obtained at two separate appointments at a maximum interval of 6 months.

Materials and methods

This retrospective study was performed by analysing archived ultrasound volume datasets obtained in the context of routine clinical practice at a tertiary urogynecological unit. Patients undergoing surgical management for urinary incontinence or female pelvic organ prolapse are assessed twice: once at the time of the initial visit to a public hospital, and again after urodynamic testing in a private setting. In some instances the order was reversed. One hundred and fifty-six patients with two separate assessments and two available ultrasound volume datasets prior to surgical management between January 2008 and November 2012 were identified from the institutional databases. At both assessments participants had undergone a physician-administered, standardised interview regarding demographic data. All complained of symptoms of pelvic floor dysfunction. A clinical assessment was performed using the International Continence Society Prolapse Quantification (ICS POP-Q) system [10].

Four-dimensional (4D) translabial ultrasound was performed after voiding and in the supine position, using a Voluson system with a RAB 4- to 8-MHz transducer, as previously described [11]. The probe was placed on the perineum in the sagittal direction. After an initial 2D assessment of bladder volume, detrusor, urethra and organ descent, ultrasound volumes were obtained by a total of approximately 20 individual examiners (medical doctors in gynaecology or urogynaecology training, of widely varying degrees of ultrasound expertise), under the supervision of the senior author or five other staff members trained by him for 3 months or more, at rest, on pelvic floor muscle contraction (PFMC) and on maximal Valsalva. Each manoeuvre was recorded three times. Patients were instructed not to be inhibited by the presence of urinary, faecal, or flatus incontinence in order to achieve

maximum pushing effort, and a Valsalva was required to last at least 6 s [12]. Care was taken to include the dorsal aspect of the symphysis pubis, as it is an essential landmark for all the parameters assessed in this study. During Valsalva, we also tried to control for levator co-activation [13]. To achieve an optimal PFM contraction we provided visual biofeedback to the patient.

Ultrasound data analysis of stored 4D volume data sets was performed offline by LT, who was blinded to all other data, on a desktop PC using the software 4D View v 10.0 (Kretz Medizintechnik, Zipf, Austria). Assessment of the second appointment was undertaken blinded to the results of the assessment of the first appointment. The most effective Valsalva and contraction volume data were utilized, defined as the volume data showing the most organ displacement. Bladder neck and pelvic organ descent were determined relative to a horizontal reference line placed at the level of the inferoposterior margin of the symphysis pubis [14]. Bladder neck descent (Fig. 1a, b) is a dynamic measurement obtained by comparing measurements at rest with those obtained on maximal Valsalva, while measurements of maximum bladder, uterine and rectal descent (Fig. 1c) are determinants of organ position on maximal Valsalva without reference to organ position at rest. Hiatal area on Valsalva was measured in the axial plane at the location of the minimal anteroposterior diameter of the hiatus, as previously described (Fig. 2) [5]. A true rectocele was defined as a sharp discontinuity in the anterior contour of the anorectal muscularis layer and a resulting herniation into the vagina of ≥ 10 mm in depth [15]. Rectocele depth was measured perpendicular to a line projected along the expected contour of the anterior anorectal muscularis [15]. Levator avulsion was determined on maximal PFMC by tomographic ultrasound imaging (TUI), as previously described [16], if all three central slices, i.e. the plane of minimal dimensions plus slices 2.5 mm and 5 mm cranial to this plane, showed an abnormal insertion of the puborectalis muscle on the inferior pubic ramus (Fig. 2). In doubtful cases, the “levator urethra gap” (LUG) was measured, with the insertion regarded as abnormal if the LUG is over 2.5 cm [17].

Prior to this study, after an initial training session of 10–20 cases assessed under direct supervision of one of the authors, an offline test–retest series of 20 ultrasound volume datasets was undertaken between the first author (who had no prior experience in pelvic floor ultrasound) and RGR, a senior trainee with over 1 year’s experience in translabial ultrasound.

Statistical analysis was undertaken using SPSS V 16 (IBM, Armonk, NY, USA). Cohen’s kappa was used to assess qualitative data, such as for levator avulsion and true rectocele. Intraclass correlation coefficient (ICC) statistics (absolute agreement definition) were obtained to test the repeatability of continuous measurements. We did not perform any power calculations because of the absence of pilot data and the retrospective nature of this research.

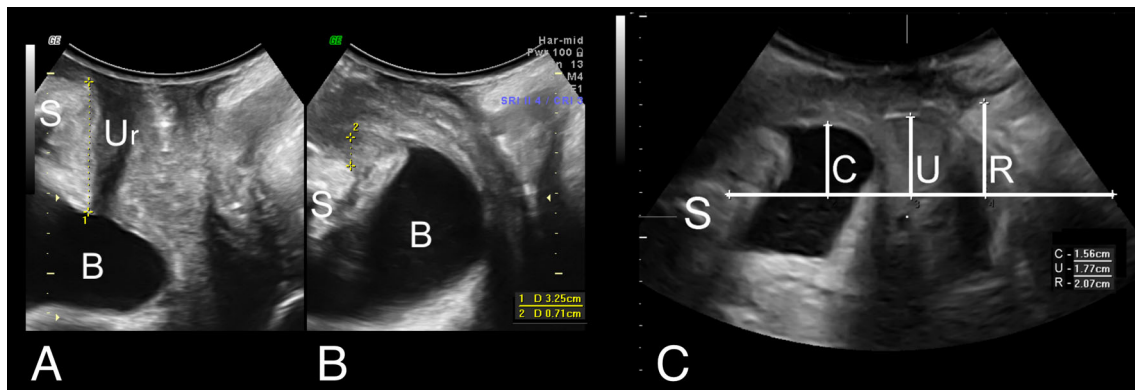


Fig. 1 Assessment of **a, b** bladder neck descent and **c** prolapse by translabial 4D ultrasound. *S* symphysis pubis, *Ur* urethra, *B* bladder, *C* cystocele, *U* uterine cervix, *R* rectal ampulla

ICC values under 0.20 were considered poor, 0.21–0.40 was considered fair, 0.41–0.60 moderate, 0.61–0.80 good and 0.81–1.00 very good or excellent [18]. To provide a wide spectrum of measures of repeatability, we calculated not just intraclass correlations but also systematic bias and the mean absolute difference between measurements.

This study was approved by our local institutional review board (NBMLHD HREC, reference 13–07). Owing to the retrospective nature of the research and the fact that all data collection occurred as part of routine clinical care, the committee waived the requirement for individual informed consent.

Results

Of 156 patients, 40 with intervals greater than 6 months and 10 with missing volume datasets were excluded, leaving 106 for analysis. All reported data pertain to those 106 patients. The interval between assessments was a mean/median of 73 days (range, 1–178). Demographic data and information on symptoms are given in Table 1. On examination, 96 women (90.6 %) were diagnosed with significant prolapse (ICS POP-Q grade 2 or higher), 73 (68.9 %) in the anterior

compartment, 64 (60.4 %) in the posterior compartment, and 13 (12.3 %) in the central compartment.

In the test–retest series undertaken prior to this study, we documented good to excellent inter-observer repeatability for the postprocessing analysis of bladder neck descent (ICC 0.82, 95 % CI 0.60–0.93), cystocele (ICC 0.82, 95 % CI 0.51–0.93), uterine descent (ICC 0.81, 95 % CI 0.53–0.93), hiatal area on Valsalva (ICC 0.92, 95 % CI 0.80–0.97), rectal ampulla position (ICC 0.73, 95 % CI 0.39–0.89), rectocele depth (ICC 0.75, 95 % CI 0.48–0.89), true rectocele (Kappa 0.69, 95 % CI 0.29–1.09) and levator avulsion on tomographic ultrasound (Kappa 0.77, 95 % CI 0.47–1.0).

Table 2 shows a comparison of ultrasound measurements for the first and second appointments, with means, ranges, intraclass correlation and confidence intervals for the same, systematic bias and mean absolute difference between measurements. All the parameters measured demonstrated good to very good repeatability between the two assessments (ICC, 0.73–0.93) with the exception of moderate repeatability for rectocele descent (ICC, 0.44). Among them, hiatal area on Valsalva showed the best repeatability (ICC, 0.93). As regards qualitative findings, such as the diagnosis of levator avulsion and true rectocele, agreement was very high (agreement in 101/106 cases, kappa 0.91 for avulsion [CI 0.77–0.94] and

Fig. 2 Assessment of **a** levator hiatal area on Valsalva and **b** levator avulsion on translabial ultrasound. *SP* symphysis pubis, *LH* levator hiatus, *L* levator ani. In **b** the *asterisk* marks the location of a full avulsion on the patient's right (the left aspect of slices 3–8)

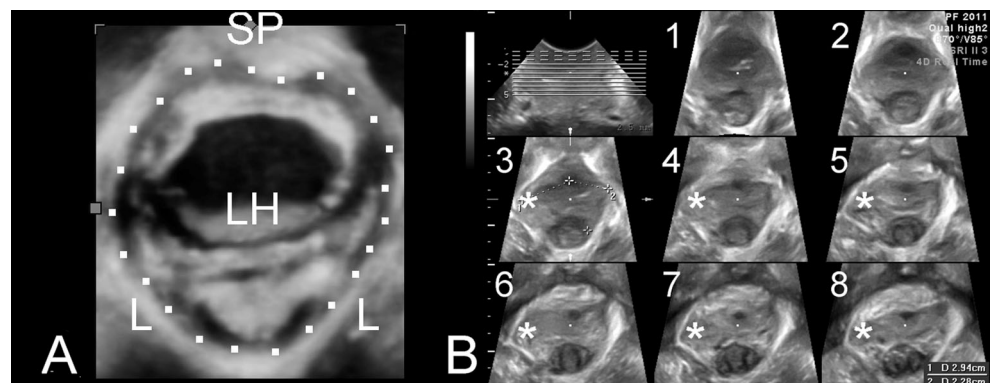


Table 1 Demographic data and prevalence of symptoms at first assessment ($n=106$)

Parameter	Mean/*median/ n	SD/%	Range
Age (years)	58	12.8	27–87
BMI (kg/m^2)	30.4	5.5	18–46
Parity	3*		1–10
Previous hysterectomy	53	50 %	
Previous POP/incontinence surgery	34	32 %	
Prolapse symptoms	71	67 %	
Stress incontinence	83	78 %	
Urgency incontinence	80	75 %	
Voiding dysfunction	46	43 %	
Chronic constipation	36	34 %	
Obstructed defecation	61	58 %	
Faecal incontinence	17	16 %	

agreement in 92/106 cases, kappa 0.73 [CI 0.56–0.84] for true rectocele). There was no appreciable systematic bias, with measurements showing mean differences of 0.9 mm for bladder neck descent and cystocele descent and 1.2 mm for uterine descent.

Discussion

The present study demonstrated good to excellent repeatability of measurements of bladder neck descent, cystocele, uterine descent, rectocele depth and hiatal area on Valsalva (ICC, 0.73–0.93) in two assessments at an average interval of 73 days, with a range of 1–178 days. The qualitative diagnoses of levator avulsion and true rectocele showed very high agreement. Our findings confirm and reinforce previous test–retest, intra- and interobserver studies using translabial ultrasound [5, 6, 8, 9, 19], which suggest that 3D/4D translabial ultrasound is a highly reliable method for the evaluation of pelvic organ prolapse and pelvic floor anatomy.

With regard to the hiatal area on Valsalva, our results demonstrate that this is likely to be amongst the most highly repeatable measures of pelvic organ descent. Recent studies demonstrated that even an inexperienced observer could adequately perform evaluation of the hiatal area for both the offline measurement technique [8, 9, 19] and volume acquisition [8], after a limited amount of teaching. As was the case in those studies, volumes in our study were obtained by a large number of trainees under supervision of the senior author or other staff trained by him for 3 months or more. Postprocessing analysis was performed by a novice after less than 2 weeks of training. These observations suggest that measurement of the hiatal area on Valsalva is a robust parameter and easy to learn with an acceptable level of effort. It is also clearly a measure of high clinical validity as it is strongly associated with prolapse and prolapse symptoms [6].

Regarding the displacement of the rectal ampulla, our observation of moderate repeatability is in contrast to the findings of Dietz and Steensma [15], who reported high interobserver reproducibility in a test–retest series, with an ICC of 0.75. This might be explained by the fact that this earlier study used two experienced observers for the evaluation of volumes that were obtained on the same day, with similar rectal filling, while we used volumes obtained at an average interval of 73 days. Regardless of this, it appears that offline assessment of the posterior compartment is more challenging for trainees and requires more extensive training [20]. This may be related to the effect of bowel filling and stool consistency, muscle resting tone and neuromuscular activation status, which is variable especially for evaluations carried out at longer intervals.

The fact that volume acquisition for this present study was performed by multiple observers is also bound to add variability, since it involves the effectiveness of instruction and investigator performance. The quality of a Valsalva manoeuvre is strongly reliant on instructions given to the patient and the patient's cooperation. The quality of instruction or coaching depends to a large degree on the operator's awareness of confounders such as levator co-activation [13], and bladder and bowel filling. Real-time imaging has the advantage of demonstrating these confounders, allowing correction of suboptimal

Table 2 Repeatability of sonographic measures of pelvic floor functional anatomy ($n=106$)

Parameter	First assessment	Second assessment	Intraclass correlation (95 % CI)	Systematic bias (95 % CI)	Mean absolute difference (SD)
Bladder neck descent (mm)	28.4 (1.3 to –58.4)	29.3 (1.3 to –61)	0.74 (0.57 to 0.83)	–0.9 mm (–3.1 to 1.4)	5.7 (7.9)
Cystocele descent (mm)	–9.4 (25.9 to –59.4)	–8.5 (28.8 to –54.2)	0.81 (0.73 to 0.87)	–0.9 mm (–3.3 to 1.6)	15.8 (8.8)
Uterine descent (mm)	–2.5 (45.8 to –49.5)	–1.3 (41.2 to –52.6)	0.79 (0.66 to 0.87)	–1.2 mm (–5.4 to 2.9)	6.3 (9.6)
Rectocele descent (mm)	–16.3 (16 to –45.6)	–12.1 (26.5 to –42.9)	0.44 (0.26 to 0.58)	–4.2 mm (–6.8 to –1.7)	8.2 (8.3)
Rectocele depth (mm)	18.7 (10 to 35.4)	19.4 (10.1 to 45.1)	0.73 (0.45 to 0.86)	3.4 mm (1.4 to 5.4)	4.1 (3.2)
Hiatal area (cm^2)	33.5 (15.5 to 66.3)	33.4 (12.7 to 70)	0.93 (0.90 to 0.95)	+0.1 cm^2 (–0.7 to –1.0)	4.7 (2.7)

effort and recognition of organ filling and levator activation [1]. In order to achieve maximum pushing effort, we instructed patients to continue the Valsalva manoeuvre for at least 6 s [12] and not be restrained by urinary or anal incontinence, which may be regarded as a form of standardization. To achieve an optimal PFM contraction we provided visual bio-feedback to the patient, as previously described [21]. The excellent ICC values of most measurements suggest that we were successful in controlling for confounders.

There are a number of potential weaknesses of this study that have to be acknowledged. The study design was retrospective, with about 20 individual ultrasound operators with varying experience being involved in volume acquisition. However, this may rather be seen as a strength as it implies that our data might be more widely applicable. We used archived 4D ultrasound data sets rather than 2D cine loops or stills for reasons of convenience as this is required for the hiatus and tomographic imaging of the levator. It is recognised, however, that sonographic assessment for organ descent does not need 3D/4D imaging as it can just as well be performed on 2D images. Another potential weakness is that hiatal area was determined in a simple axial plane of the minimal anteroposterior diameter of the hiatus. In spite of its high reproducibility [2, 21] and comparability to magnetic resonance imaging [22] it has recently been shown that determination of hiatal area in a rendered volume may be more appropriate, although this latter method is not available on all 3D ultrasound systems [22]. Additionally, it has to be recognised that in some instances intercurrent treatment such as pelvic floor muscle exercise teaching or temporary pessary placement may have occurred, confounding the results. The absence of systematic bias between the two assessments argues against any significant effect of intercurrent treatment. Ageing is unlikely to be a significant confounder owing to the short interval of 73 days on average.

Finally, while we analysed the qualitative diagnosis of levator avulsion, other measures such as the numeric value of the levator–urethra gap, and other quantitative parameters, such as detrusor wall thickness were not assessed, leaving room for future work.

The main strength of the study is that, as opposed to other similar work, we assessed the medium-term repeatability of parameters obtained on sonographic assessment of functional pelvic floor anatomy, which includes the variability of operator performance, patient instruction, and recording ultrasound volumes. Our results are reassuring, especially given that multiple different individuals participated in volume acquisition, and given that offline analysis was performed by a novice trainee. Our study suggests that despite high potential operator dependence of sonographic assessments, 3D systems can reduce this operator dependence and help facilitate both performance and analysis of a translabial ultrasound examination.

In conclusion, the repeatability of translabial ultrasound measures of functional pelvic floor anatomy performed with a mean interval of 73 days was moderate to excellent. Based on the results of the present study, we feel that 3D/4D translabial ultrasound is easy to perform and interpret. The technology appears to be suitable for introduction into clinical practice.

Conflicts of interest H.P. Dietz and K.L. Shek have received unrestricted educational grants from GE Medical.

References

1. Dietz HP (2011) Pelvic floor ultrasound in prolapse: What's in it for the surgeon? *Int Urogynecol J* 22:1221–1232
2. Dietz HP (2011) Pelvic floor ultrasound in incontinence: What's in it for the surgeon? *Int Urogynecol J* 22(9):1085–1097
3. Hoff Braekken I, Majida M, Ellstrom-Engh M, Dietz HP, Umek W, Bo K (2008) Test-retest and intra-observer repeatability of two-, three- and four-dimensional perineal ultrasound of pelvic floor muscle anatomy and function. *Int Urogynecol J* 19:227–235
4. Chen R, Song Y, Jiang L, Hong X, Ye P (2011) The assessment of voluntary pelvic floor muscle contraction by three-dimensional transperineal ultrasonography. *Arch Gynecol Obstet* 284:931–936
5. Dietz H, Shek K, Clarke B (2005) Biometry of the pubovisceral muscle and levator hiatus by three-dimensional pelvic floor ultrasound. *Ultrasound Obstet Gynecol* 25:580–585
6. Dietz H, De Leon J, Shek K (2008) Ballooning of the levator hiatus. *Ultrasound Obstet Gynecol* 31:676–680
7. Peschers UM, Fanger G, Schaer GN, Vodusek DB, DeLancey JO, Schuessler B (2001) Bladder neck mobility in continent nulliparous women. *BJOG* 108(3):320–324
8. Siafarikas F, Staer-Jensen J, Braekken I, Bo K, Engh M (2013) Learning process for performing and analyzing 3D/4D transperineal ultrasound imaging and interobserver reliability study. *Ultrasound Obstet Gynecol* 41(3):312–317
9. Zhuang R, Song Y, Chen Q, Ma M, Huang H, Chen J et al (2011) Levator avulsion using a tomographic ultrasound and magnetic resonance-based model. *Am J Obstet Gynecol* 205:232.e1–232.e8
10. Bump RC, Mattiasson A, Bo K, Brubaker LP, DeLancey JO, Klarskov P et al (1996) The standardization of terminology of female pelvic organ prolapse and pelvic floor dysfunction. *Am J Obstet Gynecol* 175(1):10–17
11. Dietz H (2004) Ultrasound imaging of the pelvic floor. II. Three-dimensional or volume imaging. *Ultrasound Obstet Gynecol* 23(6): 615–625
12. Orejuela F, Shek K, Dietz H (2012) The time factor in the assessment of prolapse and levator ballooning. *Int Urogynecol J* 23:175–178
13. Oerno A, Dietz H (2007) Levator co-activation is a significant confounder of pelvic organ descent on Valsalva maneuver. *Ultrasound Obstet Gynecol* 30:346–350
14. Dietz HP, Haylen BT, Broome J (2001) Ultrasound in the quantification of female pelvic organ prolapse. *Ultrasound Obstet Gynecol* 18(5):511–514
15. Dietz HP, Steensma AB (2005) Posterior compartment prolapse on two-dimensional and three-dimensional pelvic floor ultrasound: the

- distinction between true rectocele, perineal hypermobility and enterocele. *Ultrasound Obstet Gynecol* 26:73–77
16. Dietz H, Bernardo M, Kirby A, Shek K (2011) Minimal criteria for the diagnosis of avulsion of the puborectalis muscle by tomographic ultrasound. *Int Urogynecol J* 22(6):699–704
 17. Dietz H, Abbu A, Shek K (2008) The levator urethral gap measurement: a more objective means of determining levator avulsion? *Ultrasound Obstet Gynecol* 32:941–945
 18. Bland J, Altman D (1986) Statistical methods for assessing agreement between measurement. *Biochim Clin* 11:399–404
 19. Van Veelen G, Schweitzer K, Van Der Vaart C (2013) Reliability of pelvic floor measurements on three- and four-dimensional ultrasound during and after first pregnancy: implications for training. *Ultrasound Obstet Gynecol* 42:590–595
 20. Dietz H, Guzman Rojas R, Shek K (2014) Postprocessing of pelvic floor ultrasound data: how repeatable is it? *Aust N Z J Obstet Gynaecol* 54(6):553–557
 21. Dietz HP, Wilson PD, Clarke B (2001) The use of perineal ultrasound to quantify levator activity and teach pelvic floor muscle exercises. *Int Urogynecol J Pelvic Floor Dysfunct* 12(3):166–168, discussion 8–9
 22. Dietz H, Wong V, Shek KL (2011) A simplified method for determining hiatal biometry. *Aust N Z J Obstet Gynaecol* 51:540–543