


Intra- and interobserver reproducibility of pelvic ultrasound for the detection and measurement of endometriotic lesions

E. Bean ^{*}, P. Chaggar, N. Thanatsis, W. Dooley, C. Bottomley, and D. Jurkovic

Gynaecology Diagnostic and Outpatient Treatment Unit, Elizabeth Garrett Anderson Wing, University College London Hospital, Lower Ground Floor, 235 Euston Road, London NW1 2BU, UK

*Correspondence address. Tel: +44 (0)20 344 76515; Fax: +44 (0)20 344 76598;
E-mail: elisabeth.bean@nhs.net  <https://orcid.org/0000-0003-0264-2966>

Submitted on September 9, 2019; resubmitted on December 16, 2019; editorial decision on January 3, 2020

STUDY QUESTION: What is the interobserver and intraobserver reproducibility of pelvic ultrasound for the detection of endometriotic lesions?

SUMMARY ANSWER: Pelvic ultrasound is highly reproducible for the detection of pelvic endometriotic lesions.

WHAT IS KNOWN ALREADY: Transvaginal ultrasound (TVS) has been widely adopted as the first-line assessment for the diagnosis and assessment of pelvic endometriosis. Severity of endometriosis as assessed by ultrasound has been shown to have good concordance with laparoscopy (κ 0.79). The reproducibility of TVS for assessment of ovarian mobility and pouch of Douglas obliteration using the 'sliding sign' has already been described in the literature. However, there is no available data in the literature to demonstrate the intraobserver repeatability of measurements for endometriotic cysts and nodules.

STUDY DESIGN, SIZE, DURATION: This was a prospective observational cross-sectional study conducted over a period of 12 months. We included 50 consecutive women who were all examined by two operators (A and B) during their clinic attendance.

PARTICIPANTS/MATERIALS, SETTING, METHODS: The study was carried out in a specialist endometriosis centre. We included all consecutive women who had ultrasound scans performed independently by two experienced operators during the same visit to the clinic. The outcomes of interest were the inter- and intraobserver reproducibility for the detection of endometriotic lesions. We also assessed repeatability of the measurements of lesion size.

MAIN RESULTS AND THE ROLE OF CHANCE: There was a good level of agreement between operator A and operator B in detecting the presence of pelvic endometriotic lesions ($k = 0.72$). There was a very good level of agreement between operators in identifying endometriotic cysts ($k = 0.88$) and a good level of agreement in identifying endometriotic nodules ($k = 0.61$). The inter- and intraobserver repeatability of measuring endometriotic cysts was excellent (intra-class correlation (ICC) ≥ 0.98). There was good interobserver measurement repeatability for bowel nodules (ICC 0.88), but the results for nodules in the posterior compartment were poor (ICC 0.41). The intraobserver repeatability for nodule size measurements was good for both operators (ICC ≥ 0.86).

LIMITATIONS, REASONS FOR CAUTION: Within this cohort, there was insufficient data to perform a separate analysis for nodule size in the anterior compartment. All examinations were performed within a specialised unit with a high prevalence of deep endometriosis. Our findings may not apply to operators without intensive ultrasound training in the diagnosis of pelvic endometriosis.

WIDER IMPLICATIONS OF THE FINDINGS: These findings are important because ultrasound has been widely accepted as the first-line investigation for the diagnosis of pelvic endometriosis, which often determines the need for future investigations and treatment. The detection and measurement of bowel nodules is essential for anticipation of surgical risk and planning surgical excision.

STUDY FUNDING/COMPETING INTEREST(S): The authors have no conflict of interest. No funding was obtained for this work.

Key words: endometriosis / ultrasound / reproducibility / endometrioma / nodule

WHAT DOES THIS MEAN FOR PATIENTS?

Endometriosis is an important medical condition which affects many women. Endometriosis is difficult to detect, and only recently has it been possible to diagnose endometriosis on ultrasound. With every new test, it is important to see whether the same diagnosis is made when the test is carried out by different people. In this study, each woman had two ultrasound scans which were done by two different examiners. We found that in most cases the examiners agreed on the presence of endometriosis and the size of endometriotic cysts and nodules. This is an important finding which showed that ultrasound can be used to diagnose endometriosis in women presenting with pelvic pain and other gynaecological complaints.

Introduction

Ultrasound is a non-invasive diagnostic modality that can be performed in an outpatient setting and provides a detailed assessment of pelvic anatomy. It is relatively quick to perform and inexpensive, usually with minimal discomfort to the woman. Transvaginal ultrasound (TVS) has been widely adopted as the first-line assessment for the diagnosis and assessment of pelvic endometriosis (Abrao *et al.*, 2007; Piketty *et al.*, 2009). There is agreement from an expert panel on how a detailed pelvic ultrasound examination for the detection of pelvic endometriosis should be conducted (Guerriero *et al.*, 2016b). Severity of endometriosis as assessed by ultrasound has been shown to have good concordance with laparoscopy (kappa 0.79) (Holland *et al.*, 2010).

Scan results are used together with clinical signs and symptoms for triaging women for expectant, medical or surgical management. Ultrasound is commonly used to 'map' disease by describing the size and location of lesions (Moore *et al.*, 2002; Guerriero *et al.*, 2016a). This information may be used for surgical planning and counselling women about risks associated with surgery. In women that choose to have expectant or conservative management, TVS may be used at intervals to assess for progression or regression of disease. The reproducibility of TVS for assessment of ovarian mobility and pouch of Douglas obliteration using the 'sliding sign' has already been described in the literature (Holland *et al.*, 2013; Reid *et al.*, 2013). Egekwist *et al.* (2018) reported large inter- and intraobserver variability when 2D or 3D TVS was used for measurement of rectosigmoid deep endometriosis dimensions. Di Giovanni *et al.* (2018) demonstrated 100% specificity and sensitivity of ultrasound for the detection of rectosigmoid endometriosis (<25 cm from the anal verge) prior to surgical segmental resection. In addition, they found no statistically significant difference in the dimensions calculated by pre-operative ultrasound compared to direct measurement at histological examination. There is no single study in the literature that demonstrates the intraobserver repeatability of measurements for endometriotic cysts and nodules.

In this study, we have examined the interobserver and intraobserver reproducibility of ultrasound diagnosis of pelvic endometriosis and we also assessed the repeatability of lesion measurements.

Materials and Methods

This was a prospective observational cross-sectional study of women who were referred to our tertiary endometriosis service at University College London between 5 December 2017 and 13 November 2018. We included 50 consecutive women who underwent TVS by two experienced operators (E.B. and D.J.) during the same visit to the clinic. We included pre-menopausal women more than 18 years of age. We excluded women who had previously visited our clinic and undergone

detailed pelvic ultrasound. All women underwent a detailed pelvic ultrasound assessment using the same model of ultrasound equipment and a 7.5 MHz probe (Voluson E8, GE Medical Systems, Milwaukee, WI, USA). Operator A was a clinical research fellow who has undergone intensive advanced training in gynaecological ultrasound. Prior to commencing this study, Operator A had carried out more than 3300 gynaecological TVS examinations, which included 314 women with pelvic endometriosis. Operator B was a Level III expert with more than 10 years of experience in gynaecological ultrasound (EFSUMB 2006).

All ultrasound examinations were performed using a systematic approach. Each ovary was examined for the presence of endometriotic cysts. Endometriotic cysts were described as well-defined thick-walled cysts with no vascularity on Doppler assessment which contained homogenous low-level internal echoes ('ground glass') (Van Holsbeke *et al.*, 2010). Each endometriotic cyst was measured from the inner cyst wall in three orthogonal planes, and the mean diameter was calculated and recorded. A thorough search for endometriotic nodules was then performed. The anterior compartment of the pelvis was examined for the presence of adhesions and endometriotic nodules in the utero-vesical space, urinary bladder and distal ureters. The posterior compartment of the pelvis was then examined for the presence of endometriotic nodules in all usual sites: utero-sacral ligaments, posterior vaginal wall, recto-vaginal space, anterior rectum and sigmoid. The rectosigmoid colon was also assessed starting from the level of the anorectal verge and advancing cranially to the level of the sigmoid colon. Endometriotic nodules were defined as hypoechoic solid lesions with irregular outer borders that were fixed to neighbouring pelvic structures (Dessole *et al.*, 2003). All nodules were described by their location and size. The size was measured systematically in three orthogonal planes from the opposite outer borders of the lesions. The length (mid-sagittal plane), thickness (anterior-posterior plane) and transverse diameter were measured, and the mean diameter was calculated and recorded for further comparisons (Guerriero *et al.*, 2016b).

It is our routine practice that in all cases the ultrasound findings are checked by an expert operator prior to multi-disciplinary team review and decision about further management. All women were first examined by Operator A who carried out a comprehensive examination of the pelvis and recorded the site and size of all endometriotic lesions.

When endometriotic lesions were detected, the operator advised an independent observer of the location of each lesion. The independent observer, who did not participate in data analysis, recorded the reported location and size of each lesion from dimensions that were displayed on the ultrasound machine monitor. Once the examination was completed, the assessment of the lesions and their sizes was repeated with a delay of 10–15 min to test intraobserver variability. Operator A was blinded to the previous measurements. The patients were then examined in the same way by Operator B who was blinded

to Operator A's findings. All women were assessed by both Operator A and Operator B during the same visit to the clinic.

Intra- and interobserver variability were tested on all 50 women. This sample would provide an estimate of the kappa value that is correct to ± 0.25 for the detection of endometriotic nodules. All recorded measurements were entered into an electronic database (EXCEL™ spreadsheet, Microsoft Corporation, Seattle, WA, USA) by one of the authors (P.C.) who did not participate in data collection.

Statistical analysis

Statistical analysis was performed by an independent statistician using SPSS software (SPSS Inc., Chicago, IL, USA). The agreement between repeat categorical measurements was assessed using the standard (unweighted) kappa and weighted kappa methods. For binary outcomes (e.g. the presence or absence of endometriotic lesions), the standard kappa method was used. Kappa values of 0.81–1.00 indicated very good agreement, 0.61–0.80 indicated good agreement, 0.41–0.60 indicated moderate agreement, 0.21–0.40 indicated fair agreement and kappa values of < 0.20 indicated poor agreement (Brennan *et al.*, 1992). The analysis approach for continuous variables assessed reproducibility using the Bland–Altman limits of agreement method and intra-class correlation (ICC) method (Fleiss *et al.*, 1973; Bland *et al.*, 1986; Bland *et al.*, 1999). The ICC calculations were performed using a one-way random-effects model. Based on the 95% CI of the ICC estimate, values < 0.5 indicated poor reliability, 0.5–0.75 indicated moderate repeatability, 0.75–0.9 indicated good reliability and greater than 0.90 indicated excellent repeatability (Koo *et al.*, 2016).

Ethical approval

Ethical approval was sought from the local research ethics committee, who approved the study but deemed that, as the ultrasound assessments were carried out in accordance with our standard clinical practice, formal application to the local research and ethics committee for full ethical approval was not required.

Results

We examined 50 consecutive women who were referred for pelvic ultrasound scans by the specialist Endometriosis service. Their median age was 35 years (range 21–54). The median gravidity was 1 (range 0–5) and median parity 0 (0–4). Of the 50 women, 31 (62%) had undergone previous surgical treatment of endometriosis. The principal indications for the examinations are presented in Table I.

Reproducibility of diagnosis

There was a good level of interobserver agreement between Operator A and Operator B in detecting the presence of endometriotic lesions. The observers agreed on the presence of lesions in 36/50 (72%, 95% CI 60–84) cases and their absence in 9/50 (18%, 95% CI 7–29) cases (kappa 0.72, 95% CI 0.44–0.99). For all five cases, where there was disagreement between the observers, women had small isolated lesions which were detected by one of the operators: three nodules in the posterior pelvic compartment < 6 mm, an ovarian endometrioma < 12 mm and a bladder nodule < 5 mm.

Table I Principal indications for transvaginal ultrasound examination in the study population of women referred to a tertiary endometriosis centre (N = 50).

Indication for ultrasound scan	N (%)
Pelvic pain*	34 (68)
Subfertility + pelvic pain*	9 (18)
Pelvic pain + menorrhagia	4 (8)
Surveillance of endometriosis	2 (4)
Subfertility	1 (2)

*Included any of the following symptoms: dysmenorrhoea, dyspareunia, dyschezia and dysuria

Table II Agreement between two observers for ultrasound measures of categorical variables.

Agreement	Kappa		Interpretation
	K	95% CI	
Endometriotic cysts	0.88	0.60, > 0.99	Very good
Number of endometriotic cysts*	0.83	0.62, > 0.99	Very good
Endometriotic nodules	0.61	0.33, 0.88	Moderate/good
Number of endometriotic nodules*	0.60	0.42, 0.78	Moderate/good
Bowel nodules	0.82	0.55, > 0.99	Good/very good
Number of bowel nodules*	0.82	0.58, > 0.99	Good/very good
Posterior compartment nodules	0.68	0.41, 0.96	Good
Number of posterior compartment nodules*	0.46	0.27, 0.65	Fair

(*) Analysis performed using weighted kappa method. N = 50 women.

The observers agreed on the presence or absence of endometriotic cysts in 47/50 (94%, 95% CI 87–100) women. In total, 25/50 (50%, 95% CI 36–64) women were diagnosed with endometriotic cysts by Operator A and 26/50 (52%, 95% CI 38–66) women by Operator B. A total of 38 cysts were diagnosed by Operator A and 41 cysts by Operator B. There was agreement in the number of cysts in 41/50 (82%, 95% CI 71–93) cases.

The observers agreed on the presence or absence of endometriotic nodules in 41/50 (82%, 95% CI 71–93) women. In total, 34/50 (68%, 95% CI 55–81) women were diagnosed with endometriotic nodules by Operator A and 38/50 (76%, 95% CI 64–88) women by Operator B. A total of 78 nodules were diagnosed by Operator A and 82 nodules by Operator B. There was agreement in the number of nodules in 25/50 (50%, 95% CI 36–64) cases. There was a very good level of agreement in the identification of bowel nodules and a good level of agreement in identification of nodules in the posterior compartment. The number of nodules in the anterior compartment was too low to carry out the reproducibility and repeatability study. The agreement between Operator A and B for categorical variables is demonstrated in Table II.

Table III Inter- and intraobserver agreement for ultrasound assessment of mean endometriotic cyst diameter.

Agreement	Site	Bland–Altman analysis		ICC (95% CI)
		Mean difference	95% B-A limits	
Interobserver	All	−0.3	−6.4, 5.9	0.99 (0.97, 0.99)
	Left	−2.0	−6.4, 2.4	0.99 (0.97, 1.00)
	Right	0.9	−5.3, 7.1	0.98 (0.96, 0.99)
Intraobserver A	All	−0.3	−5.0, 4.5	0.99 (0.97, 0.99)
	Left	0.7	−3.4, 4.8	0.99 (0.98, 1.00)
	Right	−1.0	−5.7, 3.8	0.98 (0.95, 0.99)
Intraobserver B	All	−0.7	−4.7, 3.3	0.99 (0.98, 1.00)
	Left	−1.3	−6.4, 3.9	0.99 (0.97, 1.00)
	Right	−0.4	−3.6, 2.7	0.99 (0.99, 1.00)

ICC: intra-class correlation. $N = 50$ women.

Table IV Inter- and intraobserver agreement for ultrasound assessment of mean nodule diameter.

Agreement	Site	Bland–Altman analysis		ICC (95% CI)
		Mean diff (*)	95% B-A limits	
Interobserver	All	−1.3	−6.7, 4.2	0.69 (0.53, 0.80)
	Posterior	−1.7	−7.4, 4.0	0.41 (0.12, 0.65)
	Bowel	−0.4	−5.4, 4.6	0.88 (0.72, 0.95)
Intraobserver A	All	0.1	−3.2, 3.5	0.91 (0.86, 0.94)
	Posterior	0.1	−3.3, 3.4	0.86 (0.77, 0.92)
	Bowel	0.5	−2.8, 3.8	0.96 (0.90, 0.98)
Intraobserver B	All	0.0	−2.7, 2.6	0.95 (0.93, 0.97)
	Posterior	0.3	−1.8, 2.5	0.93 (0.88, 0.96)
	Bowel	−0.9	−4.0, 2.2	0.93 (0.82, 0.97)

(*)Differences for interobserver variation calculated as Observer A – Observer B. Differences for intraobserver variation calculated as Measurement 1 – Measurement 2. $N = 50$ women.

Repeatability of measurements

The overall inter- and intraobserver repeatability of endometriotic cyst measurements was excellent, with ICC analysis reported at 0.98 or higher (Table III). The Bland–Altman limits for both inter- and intraobserver variation ranged from ~ -6 to $+6$ mm and were set against the range of mean cysts diameters (10–110 mm) (Fig. 1).

The overall interobserver repeatability of endometriotic nodule measurements was moderate (Table IV). The repeatability of bowel nodule measurements was good, but the result for posterior compartment nodules was poor. The Bland–Altman limits ranged from ~ -6 to $+5$ mm for all nodule sites and were set against the range of mean nodule diameters (5–30 mm) (Fig. 2a). The intraobserver repeatability was good to excellent for both operators (Table IV). The Bland–Altman limits for intraobserver variation ranged from -3 to $+3$ mm. (Fig. 2b and c).

Discussion

Our study has shown a good level of interobserver agreement for the detection of pelvic endometriotic lesions on ultrasound. In particular, the study demonstrates very good TVS performance for the detection of endometriotic cysts and endometriotic bowel nodules. These findings are important because ultrasound has been widely accepted as the first-line investigation for the diagnosis of pelvic endometriosis, which often determines the need for future investigations and treatment (Guerriero et al., 2016b).

Our findings complement previous studies that have shown TVS to be an accurate tool for the diagnosis of endometriosis (Moore et al., 2002; Guerriero et al., 2016a). The interobserver agreement for the detection of endometriotic lesions on ultrasound was similar to results that have been reported comparing the agreement of ultrasound

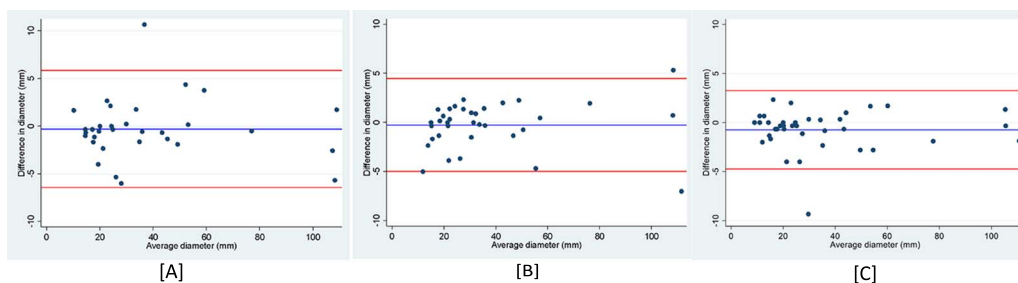


Figure 1 Bland–Altman plots for cyst measurements on transvaginal ultrasound. [A] Interobserver agreement between Operator A and B. [B] Intraobserver agreement for Operator A. [C] Intraobserver agreement for Operator B. In these graphical illustrations for the Bland–Altman analyses, the blue lines represent the mean differences between repeat values, while the red lines represent the limits of agreement.

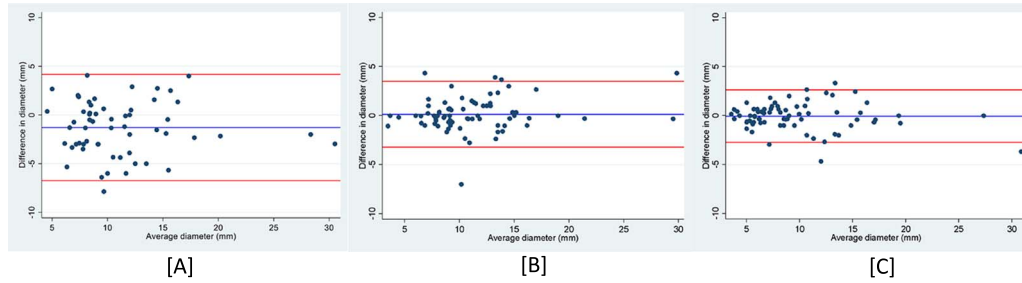


Figure 2 Bland–Altman plots for nodule measurements on transvaginal ultrasound. [A] Interobserver agreement between Operator A and B. [B] Intraobserver agreement for Operator A. [C] Intraobserver agreement for Operator B.

with laparoscopic findings (κ 0.72 versus κ 0.79, respectively) (Holland *et al.*, 2010). Surgical findings and histological diagnosis were not a condition for inclusion of women into our study, because of such previous studies demonstrating that TVS has a high level of agreement with laparoscopy for the assessment of disease severity (Holland *et al.*, 2013; Guerriero *et al.*, 2016a).

Saba *et al.* (2010) reported on the inter- and intra-observer reproducibility of MRI for the diagnosis of endometriosis. The interobserver agreement for the detection of nodules in the posterior compartment was similar to our findings on TVS (k 0.55–0.74 versus 0.68). However, the inter-observer agreement for the presence of endometriotic cysts (k 0.80 versus 0.88) and endometriotic nodules in the rectosigmoid colon (k 0.59 versus 0.82) was superior when using TVS compared to MRI, respectively. There is no available data in the literature on the repeatability of lesion measurements when using MRI.

There were disagreements between the operators in cases of small isolated nodules which were seen by one, but not the other operator. One small echogenic ovarian cyst was classified as an endometrioma by one and as a functional cyst by the other operator. Difficulty differentiating functional haemorrhagic ovarian cysts from endometriomas is a common problem, as they can have similar features on ultrasound. In such cases, a follow up examination is usually arranged to clarify the diagnosis. It is reassuring, however, that in women with multiple endometriotic lesions there was a complete agreement between the operators on the presence of the disease.

A study performed by Mais *et al.* (1993) showed ultrasound to have a sensitivity of 75% and specificity of 99% in detecting ovarian endometrioma. For three women in our study, the operators did not agree on the presence of endometrioma. Where there was disagreement, the cysts all measured less than 2 cm in mean diameter and were classified as either endometriotic cysts or functional haemorrhagic cysts.

The overall results demonstrate that intraobserver agreement is superior to interobserver agreement for both endometriotic cysts and endometriotic nodules. There was excellent agreement for both inter- and intraobserver measurements of cysts, which suggests that monitoring of cyst size can be conducted by different practitioners. However, as the interobserver reproducibility of nodule measurements is poor, we propose that monitoring of deep pelvic disease is likely to be more consistent if performed by the same operator, when possible.

Both the inter- and intraobserver repeatability of endometriotic bowel nodule measurements was excellent. However, the interobserver repeatability for nodules in the posterior compartment, exclud-

ing the bowel, was poor. Bowel nodules appear more hypoechoic on ultrasound than other nodules in the pelvis. They can also be more readily identified by following the anterior muscularis of the bowel from the level of the anal verge to the sigmoid colon in a systematic way, maintaining the anatomical landmarks. Nodules in the posterior compartment are more difficult to outline on ultrasound. They have irregular borders, and precise measurement is impeded by distorted anatomy and adherence to neighbouring structures. This may explain why interobserver agreement was much better for bowel nodules. The detection and measurement of bowel nodules is of utmost clinical importance for anticipation of surgical risk and planning surgical excision. Pre-operative awareness of the extent of bowel involvement also enables appropriate counselling and consenting of women opting for surgery. In terms of clinical practice, the exact size of nodules in the posterior compartment is less relevant for decision-making and precise surgical planning. The interobserver repeatability of measurements for nodules in the posterior compartment was poor, and therefore, reports of nodule size in the posterior compartment should be interpreted with caution.

Within this cohort, there was insufficient data to perform a separate analysis for nodule size in the anterior compartment because these lesions are inherently less common. Maccagnano *et al.* (2012) reported that only 1–2% of women with endometriosis have urinary tract involvement, with 85% of such cases involving the bladder. Further prospective research is required to assess the accuracy of TVS in identifying and measuring anterior compartment involvement and its impact on surgical planning.

A limitation of our study is that all ultrasound examinations were performed within a specialised unit with a high prevalence of the condition and involving clinicians with a high level of expertise. There was a high prevalence of deep endometriosis in this study group, which was expected due to the study being performed within a high-risk population. Our findings may not apply to operators without intensive ultrasound training in the diagnosis of pelvic endometriosis or to those operating within a general gynaecology setting.

We are aware that the time interval used for studying the intraobserver variability is very short. However, the operators were blinded to the measurements and therefore the risk of bias was very low.

In conclusion, this study has shown good reproducibility of TVS for the detection of pelvic endometriotic lesions. Both inter- and intraobserver measurements of endometriotic cysts and bowel endometriotic nodules were shown to be highly reproducible, but interobserver

reproducibility was lower for the nodules in the posterior pelvic compartment.

Acknowledgements

We are grateful to all the women attending and staff working at the Gynaecology Diagnostic Unit at UCLH for their assistance and contribution to this piece of work. We are grateful to Dr Thierry Van den Bosch (University Hospitals, KU Leuven, Leuven, Belgium) for his help with drafting the final manuscript.

Authors' roles

E.B. participated in study design and performed the ultrasound examinations; N.T., W.D. and C.B. participated as independent observers and were not involved in data analysis. P.C. entered recorded measurements into an electronic database and did not participate in data collection. E.B. drafted the manuscript; C.B. redrafted the manuscript; D.J. participated in study design, performed the ultrasound examinations and supervised the drafting and redrafting of the manuscript. All authors approved the final submitted version.

Funding

No funding was obtained for this work.

Conflict of interest

The authors have no conflict of interest.

References

Abrao MS, Gonçalves MO, Dias JA, Podgaec S, Chamie LP, Blasbalg R. Comparison between clinical examination, transvaginal sonography and magnetic resonance imaging for the diagnosis of deep endometriosis. *Hum Reprod* 2007;**22**:3092–3097.

Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**1**:307–310.

Bland M, Altman DG. Measuring agreement in method comparison studies. *Stats Methods Med Res* 1999;**8**:135–160.

Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ* 1992;**304**:1491–1494.

Dessole S, Farina M, Rubattu G, Cosmi E, Ambrosini G, Nardelli GB. Sonovaginography is a new technique for assessing rectovaginal endometriosis. *Fertil Steril* 2003;**79**:1023–1027.

Di Giovanni A, Casarella L, Coppola M, Iuzzolino D, Rasile M, Malzoni M. Combined Transvaginal/Transabdominal Pelvic Ultrasonography Accurately Predicts the 3 Dimensions of Deep Infiltrating Bowel Endometriosis Measured after Surgery: A Prospective Study in a Specialized Center. *J Minim Invasive Gynecol* 2018;**25**:1231–1240.

Education and practical standards committee. European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB). *Ultraschall Med* 2006;**27**:79–105.

Egekvis AG, Forman A, Riiskjær M, Kesmodel US, Mathiasen M, Seyer-Hansen M. Intra- and interobserver variability in nodule size of rectosigmoid endometriosis measured by two- and three-dimensional transvaginal sonography. *Acta Obstet Gynecol Scand* 2018;**97**:734–743.

Fleiss JL, Chen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 1973;**33**:613–619.

Guerriero S, Ajossa S, Orozco R, Perniciano M, Jurado M, Melis GB, Alcazar JL. Accuracy of transvaginal ultrasound for diagnosis of deep endometriosis in the rectosigmoid: systematic review and meta-analysis. *Ultrasound Obstet Gynecol* 2016a;**47**:281–289.

Guerriero S, Condous G, Van den Bosch T, Valentin L, Leone FP, Van Schoubroeck D, Exacoustos C, Installé AJ, Martins WP, Abrao MS et al. Systematic approach to sonographic evaluation of the pelvis in women with suspected endometriosis, including terms, definitions and measurements: a consensus opinion from the International Deep Endometriosis Analysis (IDEA) group. *Ultrasound Obstet Gynecol* 2016b;**48**:318–332.

Holland TK, Hoo WL, Mavrelou D, Saridogan E, Cutner A, Jurkovic D. Reproducibility of assessment of severity of pelvic endometriosis using transvaginal ultrasound. *Ultrasound Obstet Gynecol* 2013;**41**:210–215.

Holland TK, Yazbek J, Cutner A, Saridogan E, Hoo WL, Jurkovic D. Value of transvaginal ultrasound in assessing severity of pelvic endometriosis. *Ultrasound Obstet Gynecol* 2010;**36**:241–248.

Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;**15**:155–163.

Maccagnano C, Pellucchi F, Rocchini L, Ghezzi M, Scattoni V, Montorsi F, Rigatti P, Colombo R. Diagnosis and treatment of bladder endometriosis: state of the art. *Urol Int* 2012;**89**:249–258.

Mais V, Guerriero S, Ajossa S, Angiolucci M, Paoletti AM, Melis GB. The efficiency of transvaginal ultrasonography in the diagnosis of endometrioma. *Fertil Steril* 1993;**60**:776–780.

Moore J, Copley S, Morris J, Lindsell D, Golding S, Kennedy S. A systematic review of the accuracy of ultrasound in the diagnosis of endometriosis. *Ultrasound Obstet Gynecol* 2002;**20**:630–634.

Piketky M, Chopin N, Dousset B, Millischer-Bellaische AE, Roseau G, Leconte M, Borghese B, Chapron C. Preoperative work-up for patients with deeply infiltrating endometriosis: transvaginal ultrasonography must definitely be the first-line imaging examination. *Hum Reprod* 2009;**24**:602–607.

Reid S, Lu C, Casikar I, Mein B, Magotti R, Ludlow J, Benzie R, Condous G. The prediction of pouch of Douglas obliteration using offline analysis of the transvaginal ultrasound 'sliding sign' technique: inter- and intra-observer reproducibility. *Hum Reprod* 2013;**28**:1237–1246.

Saba S, Guerriero S, Sulcis R, Aiossa S, Melis G, Mallarini G. Agreement and reproducibility in identification of endometriosis using magnetic resonance imaging. *Acta Radiol* 2010;**51**:573–580.

Van Holsbeke C, Van Calster B, Guerriero S, Savelli L, Paladini D, Lissoni AA, Czekierdowski A, Fischerova D, Zhang J, Mestdagh G et al. Endometriomas: their ultrasound characteristics. *Ultrasound Obstet Gynecol* 2010;**35**:730–740.