Received: 4 March 2023 | Revised: 21 June 2023 | Accepted: 18 July 2023

DOI: 10.1111/aogs.14660

ORIGINAL RESEARCH ARTICLE

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

Prubpreet Chaggar[1] | **Tina Tellum**[1,2] | **Abigail Cohen**[1] | **Davor Jurkovic**[1]

[1]Gynecology Diagnostic and Outpatient Treatment Unit, Lower Ground Floor, Elizabeth Garrett Anderson Wing, University College London Hospital, London, UK

[2]Department of Gynecology, Oslo University Hospital, Oslo, Norway

Correspondence
Tina Tellum, Department of Gynecology, Oslo University Hospital, PB 4292 Nydalen, 0450 Oslo, Norway.
Email: tina.tellum@gmail.com

Abstract

Introduction: The number and invasion depth of endometriotic bowel lesions, total length of bowel affected by endometriosis, lesion-to-anal verge distance, and extent of pouch of Douglas obliteration are important factors in preoperatively determining risk and complexity of endometriosis surgery. The intra- and interobserver reproducibility of transvaginal ultrasound in the evaluation of many of these parameters has not yet been investigated. Our study aimed to assess the intra- and interobserver reproducibility of transvaginal ultrasound between an experienced and less experienced examiner for all of these parameters.

Material and methods: This prospective observational cross-sectional study was conducted between July 2019 and November 2020. Fifty consecutive premenopausal women who underwent transvaginal ultrasound examination in our clinic for the first time, were examined by the same two operators during the same attendance. Outcomes of interest were the inter-rater reproducibility of transvaginal ultrasound for detecting the presence, number, depth and size of bowel endometriotic nodules, lesion-to-anal-verge distance, total length of bowel affected, and pouch of Douglas obliteration. The intraobserver reproducibility was assessed for the continuous parameters. Cohen's kappa (κ) statistic, Cohen's weighted kappa (κ), proportions of agreement, intraclass correlation coefficient (ICC) and Bland–Altman limits of agreement were used to assess the reproducibility of the parameters.

Results: The inter-rater agreement and reliability were very good for identifying bowel endometriosis, the number and invasion depth of bowel nodules, determining whether the maximum nodule length was <3 cm, and lesion-to-anal-verge distance <8 cm (proportion of agreement 0.92, 0.94, 0.97, 0.94, 0.96; κ 0.92, 0.91, 0.92, 0.82, 0.89). The inter-rater agreement and reliability were good for assessing pouch of Douglas obliteration (proportion of agreement 0.86, κ 0.80). The intra-rater reliability for the mean nodule diameter (ICC 0.93 and 0.97) and total length of bowel affected

Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient; LAVD, lesion-to-anal verge distance; POD, pouch of Douglas; TVUS, transvaginal ultrasound scan.

(ICC 0.94 and 0.91) were excellent for operators A and B, respectively. The inter-rater reliability for the mean nodule diameter was good (ICC 0.80), and moderate for the total length of bowel affected (ICC 0.70). The Bland–Altman limits of agreement demonstrated clinically acceptable ranges for these two parameters.

Conclusions: This study demonstrated a high intra- and inter-rater reproducibility of transvaginal ultrasound in the diagnosis of bowel endometriosis and measurement of its various components.

KEYWORDS

bowel endometriosis, deep endometriosis, rectosigmoid colon, reliability, reproducibility, surgery planning, transvaginal ultrasound

1 | INTRODUCTION

Deep endometriosis commonly affects the anterior wall of the rectosigmoid colon and can cause severe bowel symptoms.[1,2] Symptomatic rectosigmoid endometriosis is often treated surgically. Surgical treatment can, however, be challenging and associated with severe complications, including unexpected bowel injury, anastomotic leak and rectovaginal fistula.[1]

The size, number and depth of bowel lesions and lesion-to-anal-verge distance (LAVD) are all considered important parameters in predicting surgical complexity.[3] Superficial lesions can be treated by shaving and might not require bowel resection. When deep lesions are present, excision is necessary, which often requires the support of colorectal surgeons. Lesions measuring <30 mm in length are considered amenable to discoid resection, which has been shown to have a lower complication rate than segmental resection.[3] However, segmental resection is still recommended for larger lesions[3] and multifocal lesions,[4] making the total length of the affected bowel segment a relevant factor.[5] Deeper lesions affecting the submucosal or mucosal layers of the bowel have been shown to be associated with >40% involvement of the bowel circumference.[6] Removing a disc of this depth could increase the risk of bowel stenosis.[3,7] Resection of lesions <5–8 cm from the anal verge is associated with an increased risk of postoperative anastomotic leaks,[8] fistula formation[9] and transient neurogenic bladder dysfunction.[10]

The extent of pouch of Douglas (POD) obliteration is also a factor which can increase the complexity of a surgical procedure. Therefore, it is important to presurgically determine the degree of bowel and POD involvement to allow for safe and effective preoperative planning and counseling of patients. This can help to manage patient expectations and minimize the risks of intraoperative complications, residual disease, persistent symptoms, and need for repeat surgery.

Transvaginal ultrasound (TVUS) is considered comparable to laparoscopy for the diagnosis of pelvic endometriosis.[11] Recent studies have shown that ultrasound measurements of the size of rectosigmoid endometriotic nodules,[12,13] LAVD[14] and assessment of POD obliteration[12,15] are accurate compared to findings at laparoscopy as a reference standard. Previous studies have assessed the intra- and interobserver reproducibility of TVUS in the detection and measurement of endometriotic bowel lesions[16–20] and POD obliteration,[15] but not for the number and depth of bowel nodules, the total length of bowel affected by endometriosis when multifocal lesions are present, nor the LAVD.

Key message

Transvaginal ultrasound scan has a high reproducibility for the diagnosis of bowel endometriosis and assessment of lesion features and pouch of Douglas obliteration, associated with surgical complexity. Transvaginal ultrasound can be used for assessment and surgery planning in patients with deep endometriosis.

The aim of our study was to examine the intra- and inter-rater reproducibility of TVUS in assessing the presence, number, size and invasion depth of bowel nodules, the total length of bowel affected by endometriosis when >1 lesion is present, LAVD and POD obliteration, between an experienced and less experienced examiner.

2 | MATERIAL AND METHODS

2.1 | Study setting and patient population

This was a prospective observational cross-sectional study carried out at the Department of Obstetrics and Gynecology at University College London Hospital between July 2019 and November 2020. Our study was conducted and reported according to 'The Guidelines for Reporting Reliability and Agreement Studies' (GRRAS statement)[21] and guidance from existing literature.[22]

Women who were referred to our unit for a detailed TVUS by our specialist endometriosis team were eligible for inclusion. We consecutively included all women that attended the clinic when both operators were present. All participants were examined by the same two operators during a single clinic attendance, using the same ultrasound machine and probe. The patients had not previously had any TVUS

examinations in our clinic and were aged over 18 years. Exclusion criteria were being pregnant or postmenopausal (defined as at least 12 months of amenorrhea, which could not be attributed to hormonal treatment, breastfeeding or endocrine disorders). Operator A was a clinical research fellow with approximately 2 years of prior experience in gynecological ultrasound, having performed around 1500 examinations before the commencement of the study. The fellow had received intensive training in advanced gynecological ultrasound, including the diagnosis of pelvic endometriosis and assessment of its severity. Operator B was a senior consultant gynecologist and a level III expert in this field, with over 30 years of gynecological ultrasound experience.

2.2 | Primary and secondary outcomes

The primary outcome of this study was the inter-rater reproducibility of TVUS for detecting the presence or absence of endometriotic lesions of the bowel.

Secondary outcomes included the inter-rater reproducibility of TVUS for identification of the number, depth and size of bowel endometriotic nodules, LAVD, the lesion length (maximum diameter of lesion measured in the sagittal section along the axis of the bowel), the total length of bowel segment affected by endometriosis in cases of multifocal lesions (sum of the length of all lesions) and POD obliteration. We also studied the intraobserver repeatability of TVUS for the size of nodules and LAVD, the lesion length and the total length of bowel segment affected by endometriosis in cases of multifocal lesions. The latter are deemed the most important measurements in predicting complexity and type of surgical intervention.[3,5]

2.3 | Data collection and image acquisition

A demographic and clinical history was taken prior to the ultrasound examination, and the information was entered into a secure hospital database (Viewpoint Bildverabeitung GmbH), as per routine practice in our clinic.

A two-dimensional (2D) and three-dimensional (3D) TVUS examination of the pelvis was systematically performed by operator A for each patient using a 4–9-MhZ probe (Voluson E8, GE Medical Systems), as described below. In addition to the routine pelvic scan findings of the uterus and adnexa, the presence, number, depth and size of endometriotic lesions of the bowel, LAVD and obliteration of the POD were recorded. When all endometriotic lesions of the bowel had been measured and recorded, all lesions were then remeasured and recorded by the same operator. Upon completion of this process, operator B repeated the procedure in the same manner. The measurements were obtained in real-time and recorded in a concealed fashion by an independent clinical research nurse who was present during the examinations. The research nurse was not involved in the data analysis, and both operators were blinded to their own and each other's results.

Endometriosis of the bowel was diagnosed according to the systematic approach described by the International Deep Endometriosis Analysis (IDEA) group[23] (Figure 1). The rectosigmoid colon was followed from the anal verge, proximally towards the sigmoid colon. Five layers of the normal bowel wall can be differentiated from each other by their distinct appearance on TVUS.[24] (Figure 2). Going from outwards to inwards, they can be described as follows: serosal layer (thin hyperechoic line), muscularis layer (hypoechoic), submucosal layer (hyperechoic), mucosal layer (hypoechoic strip), and interface between the lumen and the mucosal layer (hyperechoic area).

The rectosigmoid colon was assessed for the presence of endometriotic nodules, which were diagnosed when hypoechoic, avascular, solid lesions with irregular outer margins were seen (Figure 2B,C). They were usually tender on palpation with the ultrasound probe, and often adherent to the neighboring pelvic structures, including the uterine serosa, uterosacral ligaments and pelvic side wall. Endometriotic nodules were measured in three orthogonal planes. The average size of each lesion was calculated by taking the mean of all three of these readings.

The depth of invasion of the bowel by endometriosis was determined by assessing which layer of the bowel wall was affected (Figure 2). Submucosal involvement was diagnosed when nodules breached the muscularis-submucosal junction with a partial or complete loss of the definition of the anterior submucosal layer (Figure 2C).

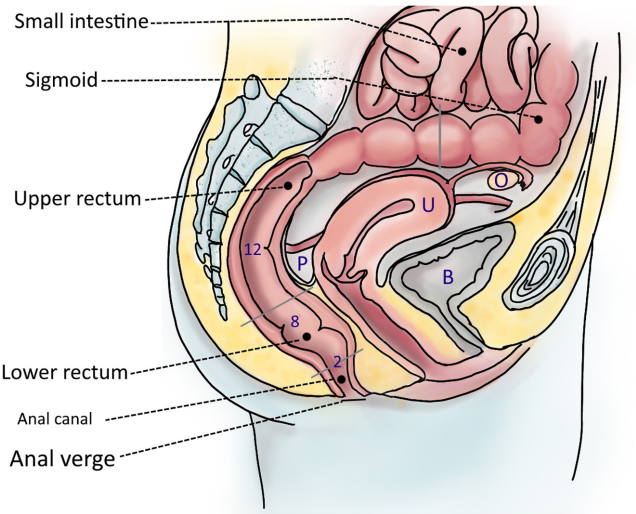

FIGURE 1 Illustration of the female pelvis and the anatomical relation of bowel to internal genitals. Definition of bowel segments according to the International Deep Endometriosis Analysis (IDEA) consensus. The anal canal ends at the dental line, followed by the lower rectum, which is retroperitoneal, thus not visible during diagnostic laparoscopy. The beginning of the upper rectum corresponds approximately to the pouch of Douglas/retrocervical area, where this bowel segment is first only anteriorly intraperitoneal while the posterior wall is still in the retroperitoneal space and immobile. When bending to the left, the rectosigmoid enters the peritoneal cavity also posteriorly, becoming completely intraperitoneally both on the anterior and posterior surface. The rectosigmoid junction is defined to be approximately at the uterine fundus, from where the sigmoid colon continues. The numbers (2, 8, 12) indicate the distance from the anal verge and their corresponding landmarks in centimeters (cm). The anatomy is consistent in-between people. The dental line is at 2 cm distance from the anal verge; the second anal valve is found at 8 cm and the third anal valve at 12 cm. B, bladder; O, ovary; P, pouch of Douglas; U, uterus.

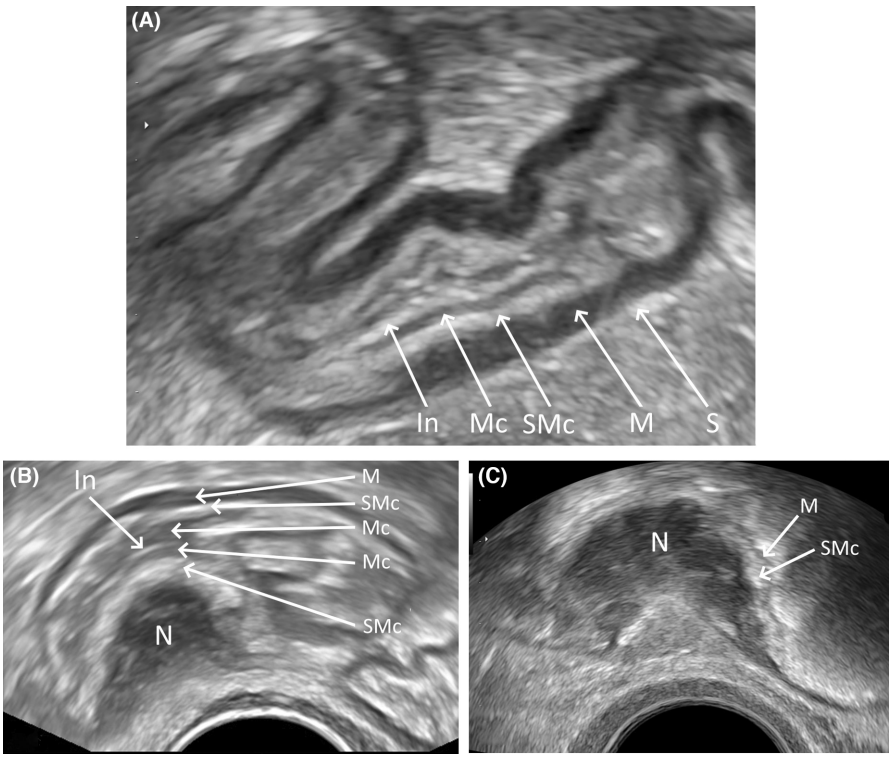

FIGURE 2 Transvaginal ultrasound, B-mode, illustrating the different layers of the bowel wall and endometriotic nodules of the bowel. (A) Normal bowel anatomy without endometriosis. (B) Endometriotic nodule (N) confined to the anterior bowel muscularis, with no involvement of other parts of the bowel wall. (C) Endometriotic nodule (N) invading the anterior submucosal layer of the bowel. Note the loss of normal anatomy of the anterior bowel which is replaced by deep endometriosis nodule. The submucosal (SMc) and muscularis (M) layers are only discernible within the posterior bowel wall. In, layer interface; Mc, mucosal layer; M, muscularis layer; N, endometriotic nodule; S, serosal layer; SMc, submucosal layer.

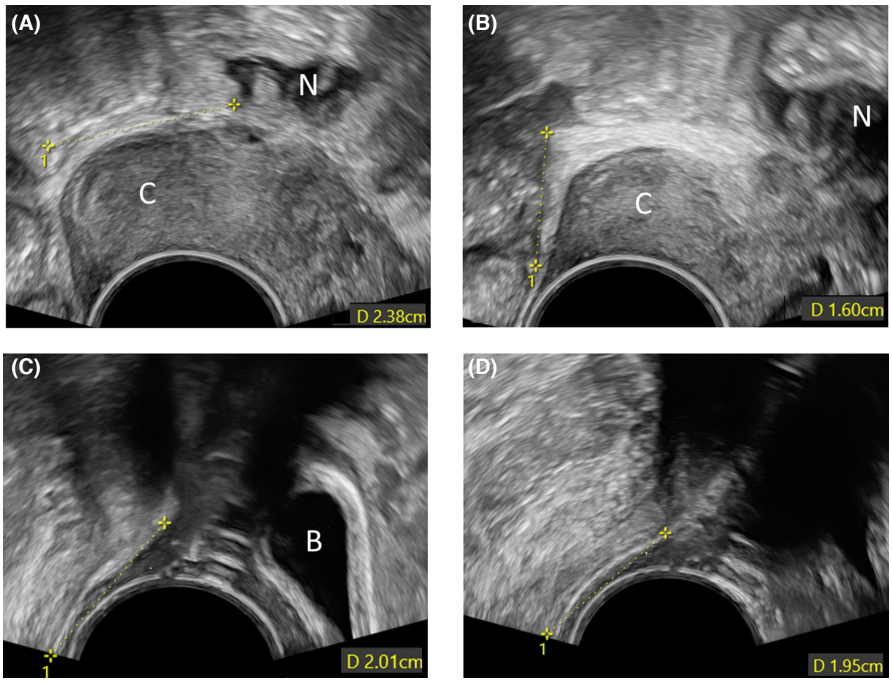

FIGURE 3 Illustration of the approach to measure the lesion-to-anal verge distance (LAVD). LAVD was measured in a stepwise fashion (A–D), starting at the most distal point of the lesion (A) and following the hypoechoic line of the muscularis layer of the bowel down to the anal verge, taking measurements successively between identifiable anatomical landmarks and adding them together to calculate the total distance (D) at the end. B, bladder; C, cervix; N, endometriotic nodule; V, vaginal wall.

The LAVD was defined as the distance between the most distal aspect of the lowermost endometriotic lesion of the bowel and the anal verge (Figure 1). It was measured in a stepwise fashion by starting at the most distal point of the lesion and following the hypoechoic line of the muscularis layer of the bowel down to the anal verge, taking measurements successively between identifiable anatomical landmarks (Figure 3), and adding them together to calculate the total distance at the end.

Obliteration of the POD was assessed using the well-described "sliding organs sign", and it was classified as none, partial or complete.[23] The "flapping sail sign"[25] was used to confirm the presence of thin, filmy adhesions. When filmy adhesions were present, the POD was described as partially obliterated.

2.4 | Statistical analyses and sample size calculations

The inter-rater reproducibility of TVUS for categorical variables (presence/absence of endometriotic nodules of the bowel, depth of

invasion, maximum length of nodules </> 3 cm, and LAVD </> 8 cm) was assessed by calculating reliability and agreement using Cohen's kappa (κ) statistic and proportions of agreement respectively. Cohen's weighted kappa (κ) statistic and proportions of agreement were used when measuring the inter-rater reproducibility of TVUS for identification of the number of nodules present and assessment of obliteration of the POD. We considered a κ value of ≤0.2 to be very poor, 0.21–0.4 poor, 0.41–0.6 moderate, 0.61–0.8 good, and 0.81–1.0 very good.[26] Confidence intervals (CI) for proportions were calculated using the Wilson efficient-score method, corrected for continuity.

The inter- and intra-rater reproducibility of TVUS for measurements of the average size and maximum length of bowel endometriotic nodules, and total length of the bowel segment involved in cases of multifocal lesions, and LAVD were calculated by assessing the reliability and agreement using the intraclass correlation coefficient (ICC) and Bland–Altman limits of agreement method, respectively. The ICC was calculated using a two-way random-effects model, with a corresponding 95% CI. Values of <0.5 signified poor reliability, 0.5–0.75 moderate, 0.75–0.9 good, and >0.9 excellent.[27] The Bland–Altman analysis gave the intervals into which 95% of all differences between measurements should lie. Its interpretation was based on clinical judgment to determine whether the interval demonstrated an acceptable degree of variation between measurements.

A previous comparable study reported a κ value of 0.82 for the intraobserver agreement in detecting endometriotic nodules of the bowel.[19] In our study, the κ value in the null hypothesis was set at 0.3, as agreement between the observers was likely to be better than expected by chance and using a value of zero would therefore be inappropriate.[28] Assuming an expected κ value of 0.8, power of 80% and alpha of 5%, a sample size of 23 was required. However, because the proportion of women with either presence or absence of bowel endometriosis was not expected to be equal, we multiplied this minimum sample size by two to accommodate this variation. This resulted in a minimum required sample size of 46.[29]

2.5 | Ethics statement

This study was approved by the Joint Research Office (JRO) at University College London Hospital (decision date June 6, 2019). All women gave their consent to be included in the study. All women who are referred to our tertiary endometriosis center have detailed transvaginal ultrasound scans by clinical research fellows, to map the extent of their disease. In all cases, these findings are routinely checked by a consultant with expert skills in the sonographic diagnosis of endometriosis during the same clinic visit. Given this, the JRO assessed that the scans being undertaken for the purpose of this study were part of routine clinical care and further approval from the Health Research Authority, the Research Ethics Committee, or Research and Development, was waived.

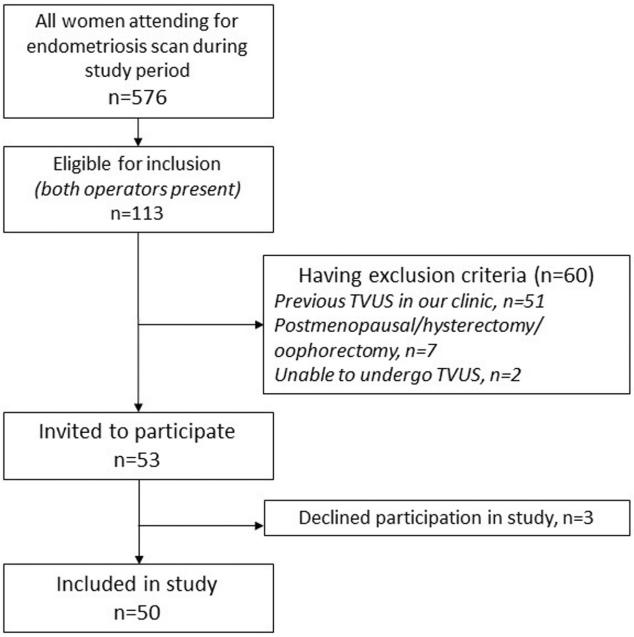

FIGURE 4 Flow chart summarizing the inclusion and exclusion of eligible women in the study. TVUS, transvaginal ultrasound.

3 | RESULTS

Out of 576 patients attending the unit for detailed endometriosis scans during our study period, 50 participants formed the final study sample (Figure 4). The repeat ultrasound was well-tolerated, and no examination had to be aborted due to patient discomfort. Their demographic and clinical characteristics are shown in Table 1, and indications for attendance in Table 2.

Both the inter-rater reliability and agreement for diagnosing bowel endometriosis on TVUS were very good (Table 3). There was disagreement in only two patients, in whom both operators had identified the same nodules but classified them differently. Operator A had reported infiltration of the bowel muscularis, whereas operator B had described a nodule in the retrocervical area tethered superficially to the bowel serosa.

When assessing the number of bowel nodules, both inter-rater agreement and reliability reached a comparably high level (Table 3), and there was disagreement in only three cases, including the two patients discussed above. In the third case, the more experienced operator diagnosed a second nodule in the sigmoid colon, which operator A did not identify.

TVUS also showed very good inter-rater agreement and reliability for assessing the depth of bowel infiltration, as demonstrated in Table 3, with only one case of disagreement, where the nodule in question was also located in the sigmoid colon.

When assessing the extent of obliteration of the POD, both inter-rater agreement and reliability were good, even if there was disagreement in five cases (Table 3). In three women, operator A described unilateral partial obliteration of the POD, whereas operator B reported no obliteration, and in two cases, operator A reported complete obliteration and operator B bilateral partial obliteration.

TABLE 1 Demographic characteristics of the study cohort ($N = 50$).

	Value
Age, years, median (IQR)	35 (30–41)
Body mass index, kg/m^2, median (IQR)	24.3 (20.9–29.8)
Ethnicity, *n* (%; 95% CI)	
Caucasian	26 (52; 37.4–66.3)
Asian	10 (20; 10.0–33.7)
Afro-Caribbean	8 (16; 7.2–29.1)
Mixed/other	6 (12; 4.5–24.3)
Smoking status, *n* (%; 95% CI)	
Nonsmoker	35 (70; 55.4–82.1)
Ex-smoker	9 (18; 8.6–31.4)
Current smoker	6 (12; 4.5–24.3)
Gravidity, *n* (%; 95% CI)	
0	24 (48; 33.7–62.6)
1	10 (20; 10.0–33.7)
2+	16 (32; 19.5–46.7)
Parity, *n* (%:95% CI)	
0	32 (64; 49.2–77.1)
1	5 (10; 3.3–21.8)
2+	13 (26; 14.6–40.4)

Abbreviations: CI, confidence interval; IQR, interquartile range.

TABLE 2 Indications for visit to clinic.

Indication for attendance	Number of women, (%; 95% CI)
Pelvic pain[a]	30 (60; 45.2–73.6)
Surveillance of endometriosis	8 (16; 7.2–29.1)
Subfertility and pelvic pain[a]	6 (12; 4.5–24.3)
Menorrhagia and pelvic pain[a]	4 (8; 2.2–19.2)
Intermenstrual bleeding and pelvic pain[a]	2 (4; 0.5–13.7)

Abbreviations: CI, confidence interval.

[a]Pelvic pain included any of the following: menstrual and nonmenstrual pelvic pain, deep dyspareunia, dyschezia, dysuria.

The intra-rater reliability for the mean nodule diameter, the maximum length of the nodule and the total length of bowel affected were excellent for both operators and good for LAVD (Table 4). The inter-rater reliability was not as high but was still good for the mean nodule diameter and LAVD, and moderate for the maximum length of the nodule and total length of bowel affected (Table 4). Clinically, the Bland Altman limits of agreement demonstrated a narrow range, suggesting good intra-rater agreement for mean nodule diameter, maximum nodule length and total length of bowel affected, but a wider range for LAVD (Figure 5). The intra-rater agreement for these continuous parameters was overall better than the inter-rater agreement, which was still good for mean nodule diameter and total length of bowel affected, but again not as good for LAVD, or maximum nodule length (Table 4). The inter-rater agreement and reliability were significantly higher for maximum length of bowel nodule and LAVD when operators were compared in their judgment of whether these measurements were <3 and <8 cm, respectively (Table 3).

4 | DISCUSSION

Our study showed a high intra- and inter-rater reproducibility of TVUS in the detection of the presence, number, size and depth of endometriotic nodules in the bowel, whether LAVD was <8 cm, and extent of POD obliteration. These parameters are crucial for planning endometriosis surgery.

Only two disagreements occurred between observers regarding the presence of bowel nodules. In one of these two cases, a nodule was located in the sigmoid colon. The literature reports that nodules >25 cm from the anal verge are difficult to visualize using TVUS,[14] requiring a very experienced operator when there is suspicion of endometriotic lesions in the sigmoid. The disagreements in our study were not clinically significant, as they would not affect the surgical approach or assessment of surgical risks.

When assessing continuous parameters, such as the maximum nodule length, LAVD and total length of bowel affected, the intra-rater reliability and agreement were overall higher than the inter-rater measurements, which is often the case with reproducibility studies. The inter-rater reliability was only moderate for "total length of bowel involved;" however, there were only four cases with more than one bowel nodule identified by both operators.

When operators were assessing whether LAVD was <8 cm or not, inter- and intra-rater reliability and agreement were much higher than when individual LAVD measurements were being compared. Given that the surgical approach and risks of the procedure considerably change if the LAVD is <8 cm vs ≥8 cm, this analysis for LAVD is clinically more relevant than the latter.

Although measurement of the LAVD using TVUS is recommended by the IDEA group, there is no standardized guidance on the methodology.[23] A recent study by Aas-Eng et al. demonstrated good reproducibility of TVUS in measuring LAVD, when compared to intraoperative measurement (IOM) using a rectal probe (ICC 0.81 and good agreement on Bland–Altman plot).[14] A second study by the same group also found TVUS and MRI to have overall similar performance in their measurement of LAVD when compared to IOM.[30] However, their TVUS and IOM measurements do not account for the curvature of the bowel, leading to an underestimation of LAVD in lesions located further from the anal verge, compared to MRI technique which allows for the bowel curvature. Furthermore, they have not assessed the reproducibility of their method.

In clinical practice, the nodule length is a significant factor affecting the management when it reaches a cutoff of 30 mm, as the surgical approach becomes more invasive (segmental resection rather than discoid).[3] The inter-rater agreement and reliability significantly improved when operators assessed whether the length was <30 mm or not, rather than interpreting the exact measurements.

TABLE 3 Inter-rater agreement and reliability for categorical outcomes.

Outcome	Category	Operator A *n* (%)	Operator B *n* (%)	Agreement Proportions of agreement (95% CI)	Reliability Kappa (95% CI)	Interpretation of reliability
Bowel endometriosis	No	20 (40)	22 (44)	0.92 (0.64–1.00)	0.92 (0.64–1.00)	Very good
	Yes	30 (60)	28 (56)			
Number of nodules	0	20 (40)	22 (44)	0.94 (0.83–0.98)	0.91 (0.69–1.00)[a]	Very good
	1	26 (52)	23 (46)			
	2	4 (8)	5 (10)			
POD obliteration	No	8 (16)	11 (22)	0.86 (0.73–0.93)	0.80 (0.60–1.00)[a]	Good
	Partial	36 (72)	35 (70)			
	Complete	6 (12)	4 (8)			
Depth nodules	Muscularis	23 (72)	24 (75)	0.97 (0.82–1.00)	0.92 (0.58–1.00)	Very good
	Submucosa	9 (28)	8 (25)			
Maximum nodule length <3 cm	No	24 (75)	26 (81)	0.94 (0.78–0.99)	0.82 (0.57–1.00)	Very good
	Yes	8 (25)	6 (19)			
LAVD <8 cm	No	23 (82)	22 (79)	0.96 (0.80–1.00)	0.89 (0.67–1.00)	Very good
	Yes	5 (18)	6 (21)			

Abbreviations: CI, confidence interval; LAVD, lesion to anal verge distance; POD, pouch of Douglas.

[a]Analysis using weighted kappa.

TABLE 4 Intra and inter-rater reliability and agreement for continuous outcomes – ICC and Bland Altman analysis.

Outcome	Assessment type	Reliability ICC (95% CI)	Interpretation of reliability	Mean difference	SD difference	Agreement 95% BA limits
Mean nodule diameter	intra-rater[a]					
	Operator A	0.93 (0.87–0.97)	Excellent	0.2	1.2	−2.2 – 2.7
	Operator B	0.97 (0.92–0.99)	Excellent	0.3	0.9	−1.5 – 2.1
	Inter-rater[b]	0.80 (0.61–0.90)	Good	−1.0	3.1	−7.2 – 5.1
Maximum nodule length	Intra-rater[a]					
	Operator A	0.93 (0.87–0.96)	Excellent	0.1	3.5	−6.8 – 7.1
	Operator B	0.98 (0.88–0.99)	Excellent	1.2	2.1	−2.9 – 5.4
	Inter-rater[b]	0.65 (0.37–0.82)	Moderate	−0.6	10.2	−20.5 – 19.4
Total length of bowel affected	Intra-rater[a]					
	Operator A	0.94 (0.51–1.00)	Excellent	2.5	1.2	0.2–4.8
	Operator B	0.91 (0.37–0.99)	Excellent	0.5	3.5	−6.4 – 7.4
	Inter-rater[b]	0.70 (0.00–0.98)	Moderate	4.4	5.9	−7.2 – 16.0
LAVD	Intra-rater[a]					
	Operator A	0.85 (0.70–0.92)	Good	−1.0	15.0	−30 – 28
	Operator B	0.85 (0.88–0.93)	Good	3.0	14.0	−24 – 30
	Inter-rater[b]	0.79 (0.60–0.90)	Good	−3.0	20.0	−42 – 36
	Inter-rater[b,c]	0.73 (0.48–0.87)	Good	−1.0	15.0	−30 – 29

Abbreviations: BA, Bland Altman; CI, confidence interval; ICC, intraclass correlation coefficient; LAVD, lesion to anal verge distance; SD, standard deviation.

[a]Differences calculated as measurement 2 minus measurement 1.

[b]Differences calculated as operator B minus operator A.

[c]Bland Altman analysis omitting one outlying value with large inter-rater difference.

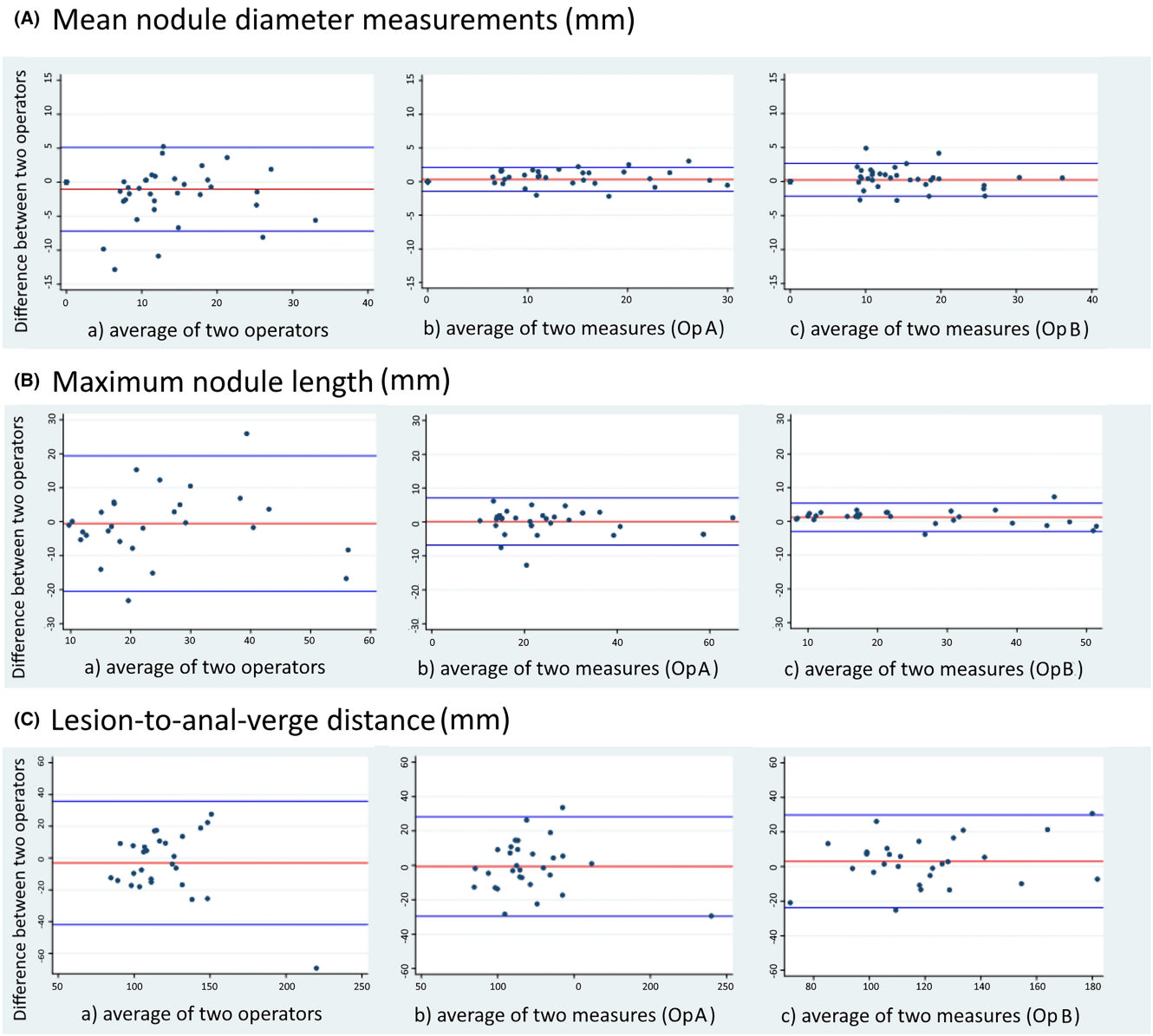

FIGURE 5 Bland–Altman plots for measurements on transvaginal ultrasound. The red lines represent the mean differences between the measurements, the blue lines represent the upper and lower limits of agreement (95% confidence interval). All measurements are indicated in millimeters (mm). Images (a) represent interobserver agreement between operators (Op A; Op B); Images (b) represent the intraobserver agreement for Operator A (Op A); Images (c) represent intraobserver agreement for Operator B (Op B) for the respective measurements.

Of six existing studies evaluating the intra- and interobserver reproducibility of TVUS for the identification +/− measurement of rectosigmoid endometriotic nodules,[15–20] ours was comparable to only three of these,[15,19,20] due to significant differences in the study design and statistical analysis.

Two of these three studies were conducted similarly to ours, other than both their examiners being highly experienced in the TVUS diagnosis of endometriosis.[15,19] While one of the studies is more recent, from 2020,[19] the other is significantly older, being published in 2013.[15] This study reported a lower interobserver reliability in the detection of bowel endometriosis compared to our study (κ 0.56 vs. κ 0.96), which could be explained by technological advances in ultrasound and the development of standardized criteria for diagnosing endometriosis since.[23] The more recent study[19] demonstrated results comparable to ours, reporting very good inter-rater reliability of TVUS for identification of bowel nodules and inter- and intra-rater reliability of measurements of bowel nodules, with κ 0.82 and ICC 0.88, 0.96 and 0.93, respectively.[19] Their results for intra- and interobserver agreement of measurement of these nodules were also comparable to ours (Bland Altman limits of agreement for observers A and B and interobserver agreement: −2.8–3.8, −4.0–2.2 and −5.4–4.6, respectively).

The third study, conducted by Guerriero et al., compared three highly experienced examiners to three with less than two

years' experience, which is similar to the difference in the level of expertise between observer A and B in our study. However, they observed lower levels of reliability for TVUS detection of rectosigmoid endometriosis than in our study, with moderate to good intraobserver reliability (κ values between 0.49–0.96) and fair to very good interobserver reliability (κ values between 0.21–0.87).[20] Different teaching setups for skill development in less experienced investigators, individual differences, and diagnoses being made from stored 3D volumes might be the reason for this discrepancy.

That specific training in the ultrasound assessment of deep endometriosis and real-time analysis appears to be relevant, is supported by the comparison of our results to those of Reid et al.[12] They reported a lower intra- and inter-rater reliability of TVUS in determining POD obliteration compared to our study (κ 0.60–0.95 and 0.46–1.0 for different observers, and κ 0.646, respectively). Their study included four observers, two of whom were fetal medicine experts with no experience in gynecological scanning. Furthermore, the study was based on off-line prerecorded video sets rather than real-time scans. Similar issues existed for the study by Menakaya et al. who also reported a lower inter-rater reliability than our study.[31]

Few studies have assessed the inter-rater reproducibility of TVUS in determining POD obliteration.[12,15,31,32] Holland et al., who reported a higher inter-rater reliability than in our study (κ 0.947),[15] only classed obliteration into "yes" or "no". However, reporting partial or complete POD obliteration, as done in this study, is clinically important as it considerably influences the complexity of a surgical case. Leonardi et al. assessed the accuracy of TVUS assessment of POD obliteration between three gynecology trainees compared to an expert in gynecological scanning, but not the intraobserver reproducibility.[32]

A strength of our study is that it is the first to assess the invasion depth of bowel nodules, measurement of total length of affected bowel segment and LAVD. The other strengths include the prospective approach with predefined diagnostic parameters, blinding examiners to their own and each other's findings, and that all assessments were performed using real-time ultrasound examinations. Off-line image and video analysis for re-examination are a source of bias in assessing the relevance of operator experience. The reliability of our study was further strengthened by adhering to recommended guidelines on reproducibility studies.[21]

One of the limitations of our study was that examiners were not blinded to the fact that they were participating in a research study. Second, both examiners worked in the same institution, a tertiary endometriosis center, and conducted examinations using the same routine, technique and high-quality ultrasound machines. Therefore, our findings may not be transferable to operators working in a different setting, using different techniques and equipment. Also, there was no corroboration of the imaging finding with surgical findings. However, the accuracy of TVUS in diagnosing deep endometriosis is now considered comparable to that of laparoscopy[11] and has been demonstrated in multiple studies.[13,14,33,34,35] When the intraobserver reproducibility was assessed, there was a very short interval between examinations, while in clinical practice several weeks may pass between examinations. It is important to note this limitation when considering our results of intraobserver variability.

5 | CONCLUSION

Reproducibility studies were important before the introduction of diagnostic tests into clinical practice.[36] Our study demonstrated that both less and more experienced operators can achieve similar results when performing a preoperative sonographic assessment of various parameters of bowel endometriosis and degree of POD obliteration. Further research is required to investigate an optimal sonographic training and quality markers that define proficiency in bowel endometriosis scanning. The use of TVUS in predicting the complexity of surgical cases and assessing the risk of complications should be addressed in future studies.

AUTHOR CONTRIBUTIONS

P.C. and D.J. designed the study, performed the ultrasound examinations and wrote the manuscript. A.C. contributed to the data collection and contributed to writing the manuscript. T.T. contributed with data interpretation, writing the manuscript and Figures. All authors approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank Sarah Ekladios for her support during data collection and Paul Basset for the statistical analysis.

FUNDING INFORMATION

No specific funding was received for this study. During the conduction of the study, TT received a research grant from the South-Eastern Norwegian Health Authority (grant no. 2020083).

CONFLICT OF INTEREST STATEMENT

TT reports receiving personal fees for lectures on ultrasound from GE Healthcare, Samsung, Medtronic and Merck, outside of this study. The other authors report no conflict of interest.

ORCID

Prubpreet Chaggar https://orcid.org/0000-0002-8432-6117
Tina Tellum https://orcid.org/0000-0003-2635-4504
Davor Jurkovic https://orcid.org/0000-0001-6487-5736

REFERENCES

1. Byrne D, Curnow T, Smith P, Cutner A, Saridogan E, Clark TJ. Laparoscopic excision of deep rectovaginal endometriosis in BSGE endometriosis centres: a multicentre prospective cohort study. *BMJ Open*. 2018;8:e018924.
2. Bean E, Naftalin J, Horne A, Saridogan E, Cutner A, Jurkovic D. Prevalence of deep and ovarian endometriosis in early pregnancy: ultrasound diagnostic study. *Ultrasound Obstet Gynecol*. 2021;59:107-113.

3. Abrao MS, Petraglia F, Falcone T, Keckstein J, Osuga Y, Chapron C. Deep endometriosis infiltrating the recto-sigmoid:critical factors to consider before management. *Hum Reprod Open*. 2015;21:329-339.
4. Nezhat C, Li A, Falik R, et al. Bowel endometriosis: diagnosis and management. *AJOG*. 2017;218:549-562.
5. Keckstein J, Saridogan E, Ulrich UA, et al. The #Enzian classification: a comprehensive non-invasive and surgical description system for endometriosis. *Acta Obstet Gynecol Scan*. 2021;100:1165-1175.
6. Abrao MS, Podgaec S, Dias JAJ, et al. Endometriosis lesions that compromise the rectum deeper than the inner muscularis layer have more than 40% of the circumference of the rectum affected by the disease. *J Minim Invasive Gynecol*. 2008;15:280-285.
7. Bendifallah S, Puchar A, Vesale E, Moawad G, Daraï E, Roman H. Surgical outcomes after colorectal surgery for endometriosis: a systematic review and meta-analysis. *J Minim Invasive Gynecol*. 2021;28:453-466.
8. Ruffo G, Scopelliti F, Scioscia M, Ceccaroni M, Mainardi P, Minelli L. Laparoscopic colorectal resection for deep infiltrating endometriosis: analysis of 436 cases. *Surg Endosc*. 2010;24:63-67.
9. Ruffo G, Scopelliti F, Manzoni A, et al. Long-term outcome after laparoscopic bowel resection or deep infiltrating endometriosis: a single-center experience after 900 cases. *Biomed Res Int*. 2014;2014:463058.
10. Dousset B, Leconte M, Borghese B, et al. Complete surgery for low rectal endometriosis: long-term results of a 100-case prospective study. *Ann Surg*. 2010;251:887-895.
11. Becker CM, Bokor A, Heikinheimo O, et al. ESHRE guideline: endometriosis. *Hum Reprod Open*. 2022;2022:hoac009.
12. Reid S, Lu C, Casikar I, et al. Prediction of pouch of Douglas obliteration in women with suspected endometriosis using offline analysis of the transvaginal ultrasound 'sliding sign' technique: inter- and intra-observer reproducibility. *Ultrasound Obstet Gynecol*. 2013;28:1237-1246.
13. Aas-Eng MK, Lieng M, Dauser B, et al. Transvaginal sonography determines accurately extent of infiltration of rectosigmoid deep endometriosis. *Ultrasound Obstet Gynecol*. 2021;58:933-939.
14. Aas-Eng MK, Dauser B, Lieng M, et al. Transvaginal sonography accurately measures lesion-to-anal-verge distance in women with deep endometriosis of the rectosigmoid. *Ultrasound Obstet Gynecol*. 2020;56:766-772.
15. Holland T, Hoo WL, Mavrelos D, Saridogan E, Cutner A, Jurkovic D. Reproducibility of severity of pelvic endometriosis using transvaginal ultrasound. *Ultrasound Obstet Gynecol*. 2013;41:210-215.
16. Egekvist AG, Forman A, Seyer-Hansen M. Transvaginal ultrasonography of rectosigmoid endometriosis: interobserver variation of lesion size. *Acta Obstet Gynecol Scand*. 2011;91:264-268.
17. Tammaa A, Fritzer N, Lozano P, et al. Interobserver agreement and accuracy of non-invasive diagnosis of endometriosis by transvaginal sonography. *Ultrasound Obstet Gynecol*. 2015;46:737-740.
18. Egekvist AG, Forman A, Riiskjaer M, Kesmodel US, Mathiasen M, Seyer-Hansen M. Intra- and interobserver variability in nodule size of rectosigmoid endometriosis measured by 2- and 3-dimensional transvaginal sonography. *Acta Obstet Gynecol Scand*. 2018;97:734-743.
19. Bean E, Chaggar P, Thanatsis N, Dooley W, Bottomley C, Jurkovic D. Intra- and inter-observer reproducibility for the detection and measurement of endometriotic lesions. *Hum Reprod Open*. 2020;2020:hoaa001.
20. Guerriero S, Pascual M, Ajossa S, et al. The reproducibility of ultrasonographic findings of rectosigmoid endometriosis among examiners with different level of expertise. *J Ultrasound Med*. 2021;41:403-408.
21. Kottner J, Audige L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol*. 2011;64:96-106.
22. Bartett J, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstet Gynecol*. 2008;31:466-475.
23. Guerriero S, Condous G, Van den Bosch T, 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. *Utrasound Obstet Gynecol*. 2016;48:318-332.
24. Abrao MS, Goncalves MO, Dias JA Jr, Podgaec S, Chamie LP, Blasbalg R. Comparison between clinical examination, transvaginal sonography and magentic resonanceimaging for the diagnosis of deep endometriosis. *Hum Reprod*. 2007 Dec;22:3092-3097.
25. Savelli L, de Laco P, Ghi T, Bovicelli L, Rosati F, Cacciatore B. Transvaginal sonographic appearance of peritoneal pseudocysts. *Ultrasound Obstet Gynecol*. 2004;23:284-288.
26. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ*. 1992;304:1491-1494.
27. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155-163.
28. Sim J, Wright C. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005;85:257-268.
29. Bujang M, Baharum N. Guidelines of the minimum sample size requirements for Cohen's kappa. Epidemiol biostat. *Public Health*. 2017;14:e12267-1-10.
30. Aas-Eng MK, Young VS, Dormagen B, Pripp AH, Hudelist G, Lieng M. Lesion-to-anal-verge distance in rectosigmoid endometriosis on transvaginal sonography vs magnetic resonance imaging: prospective study. *Ultrasound in Obstet Gynecol*. 2023;61:243-250.
31. Menakaya U, Infante F, Lu C, Phua C, Model A, Messyne F. Interpreting the real-time dynamic 'sliding sign' and predicting pouch of Douglas obliteration: an interobserver, intraobserver, diagnostic-accuracy and learning-curve study. *Ultrasound Obstet Gynecol*. 2015;48:113-120.
32. Leonardi M, Ong J, Espada M, et al. One-size-Fits_All approach does not work for gynecology trainees learning endometriosis ultrasound skills. *J Ultrasound Med*. 2020;39:2292-2303.
33. Hudelist G, Tuttlies F, Rauter G, Pucher S, Keckstein J. Can transvaginal sonography predict infiltration depth in patients with deep infiltrating endometriosis of the rectum? *Hum Reprod*. 2009;24:1012-1017.
34. 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.
35. Hudelist G, English J, Thomas AE, Tinelli A, Singer CF, Keckstein J. Diagnostic accuracy of transvaginal ultrasound for non-invasive diagnosis of bowel endometriosis: systematic review and meta-analysis. *Ultrasound Obstet Gynecol*. 2011;37:257-263.
36. Coelho NM, Roncato P, Nastri CO, Martins WP. True reproducibility of ultrasound techniques (TRUST): systematic review of reliability studies in obstetrics and gynecology. *Ultrasound Obstet Gynecol*. 2015;46:14-20.

How to cite this article: Chaggar P, Tellum T, Cohen A, Jurkovic D. Intra- and interobserver reproducibility of transvaginal ultrasound for the detection and measurement of endometriotic lesions of the bowel. *Acta Obstet Gynecol Scand*. 2023;102:1306-1315. doi:10.1111/aogs.14660