Meta-analysis and systematic review to determine the optimal imaging modality for the detection of rectosigmoid deep endometriosis

B. Gerges¹, W.Li², M. Leonardi¹,⁵, B.W.Mol²,³, G. Condous¹

¹ Sydney Medical School Nepean, University of Sydney Nepean Hospital, Kingswood NSW Australia 2750
² Department of Obstetrics and Gynaecology Monash University, Clayton VIC 3168 Australia
³ Aberdeen Centre for Women’s Health Research University of Aberdeen, Aberdeen, UK
⁴ Sydney West Advanced Pelvic Surgery (SWAPS) Blacktown Hospital, Blacktown NSW Australia 2148
⁵ Department of Obstetrics and Gynecology, McMaster University, Hamilton, L8S 4L8, Canada

*Address for correspondence:
Bassem Gerges
Acute Gynaecology, Early Pregnancy and Advanced Endosurgery Unit
Nepean Hospital
Penrith NSW 2750 Australia
Email: bgerges@gmail.com

Short Title: Imaging and Rectosigmoid Deep Endometriosis
Keywords: endometriosis, imaging, pre-operative, diagnosis, laparoscopy, rectal, rectosigmoid, bowel
Contribution:

What are the novel findings of this work?

Transvaginal ultrasound (TVS) was marginally superior to magnetic resonance imaging and whilst computed tomography also performed well, transrectal endoscopic sonography outperformed all the imaging techniques for the diagnosis of rectal/rectosigmoid deep endometriosis (DE).

What are the clinical implications of this work?

This confirms the importance for the use of imaging for the pre-operative diagnosis of rectal/rectosigmoid DE, with all imaging modalities performing well, although there are key differences for their applicability. As TVS is the simpler, faster, and more readily available, we believe it should be the first line diagnostic tool.
Abstract

Objective: To review the diagnostic accuracy and determine the optimum imaging modality for the detection of rectosigmoid deep endometriosis (DE) in women with a clinical history of endometriosis.

Methods: A systematic review was conducted using PubMed, Medline, Scopus, Embase and Google Scholar to identify studies published between January 1990 and May 2020. Studies were considered eligible if they were prospective and used any imaging modality pre-operatively to assess for the presence of DE in the rectum/rectosigmoid, which was then correlated with the surgical data as the reference diagnosis. The eligibility of studies was restricted to those having at least 10 affected and 10 unaffected women. The QUADAS-2 tool was used to assess quality. This study was prospectively registered with PROSPERO (CRD42017059872).

Results: Of the 1,977 references identified, 30 studies (n = 3,374) were included in the analysis. The overall pooled sensitivity and specificity, from which the likelihood ratio of a positive test (LR+), likelihood ratio of a negative test (LR-) and diagnostic odds ratio (DOR) were calculated, were as follows for transvaginal ultrasound (TVS) 89% (95% CI 83 – 93%), 97% (95% CI 95 – 98%), 28.8 (95% CI 16.2 – 51.0), 0.12 (95% CI 0.08 – 0.18) and 248 (95% CI 104 – 594), for magnetic resonance imaging (MRI) 86% (95% CI 79 – 81%), 97% (96% CI 94 – 97%), 21.0 (95% CI 13.4 – 33.1), 0.15 (95% CI 0.09 – 0.23), and 144 (95% CI 70 – 297), for computed tomography (CT) 93% (95% CI 84 – 97%), 95% (95% CI 81 – 99%), 20.3 (95% CI 4.3 – 94.9), 0.07 (95% CI 0.03 – 0.19), and 280 (95% CI 28 – 2826), and for transrectal endoscopic sonography (RES) 92% (95% CI 87 – 95%), 98% (95% CI 96 – 99%), 37.1 (95% CI 21.1 – 65.4), 0.08 (95% CI 0.05 – 0.14), and 455 (95% CI 196 – 1054),
respectively. There was significant heterogeneity and the studies were considered poor methodologically according to the QUADAS-2 tool.

**Conclusions:** The sensitivity of transvaginal sonography (TVS) for the detection of DE seems to be slightly better than magnetic resonance imaging (MRI), although RES was superior to both. Specificity of both TVS and MRI were excellent. As TVS is the simpler, faster, and more readily available, we believe it should be the first line diagnostic tool for the women with suspected DE.
Introduction

Endometriosis is typically characterised by the deposition of endometrial-like cells outside the uterine cavity resulting in lesions ranging from peritoneal implants to deep endometriosis (DE). DE is defined as endometrial implants or nodules of ≥ 5mm sub-peritoneal infiltration which can involve multiple organs in the pelvis including the bladder, vagina, uterosacral ligaments, bowel and peritoneum. A resultant distortion in anatomy can occur, including obliteration of the pouch of Douglas (POD) or cul-de-sac. Endometriosis can cause symptoms of chronic pelvic pain, dyspareunia, urinary and/or bowel symptoms and infertility.

Multiple imaging modalities have been used in an attempt to pre-operatively accurately diagnose the presence and severity of endometriosis, specifically DE. These have included, but are not limited to, transvaginal sonography (TVS), rectal endoscopy-sonography (RES), magnetic resonance imaging (MRI) and computed tomography (CT). Although, TVS and MRI have been shown to be comparable with regards to diagnostic accuracy, with the differences being in cost and expertise, laparoscopy remains to be the gold standard for the diagnosis of endometriosis. In 2016, the International Deep Endometriosis Analysis (IDEA) Group published a consensus statement to standardise the terms, definitions and sonographic evaluation of the pelvis, and in particular, identified five locations for examination, namely, the bladder, rectovaginal septum, posterior vaginal fornix, uterosacral ligaments/torus uterinus and rectum/rectosigmoid/sigmoid.

There have been a number of systematic reviews separately assessing the accuracy of TVS and MRI for the detection of DE, with Guerriero, et al. and Pereira, et al. performing head-to-head analyses of MRI and TVS to assess the accuracy of TVS for the
detection of rectosigmoid DE and the most recent Cochrane systematic review by Nisenblat, et al. in 2016\textsuperscript{13} which compared multiple imaging technologies. The purpose of this systematic review was to assess the diagnostic accuracy of all imaging modalities for the pre-operative detection of DE in the rectum/rectosigmoid as defined by the IDEA group compared with surgical data in women of reproductive age.
Methods

Protocol and registration

This systematic review and meta-analysis was designed according the PRISMA statement\textsuperscript{14} and the Synthesizing Evidence from Diagnostic Accuracy Tests (SEDATE) guidelines on how to conduct a systematic review on diagnostic test accuracy\textsuperscript{15}. All inclusion/exclusion criteria, data extraction and quality assessment were specified prior to commencement and the protocol was prospectively registered with PROSPERO (CRD42017059872).

Eligibility criteria

Published, peer-reviewed studies that compared one or more imaging modality, considered the “index test”, to pre-operatively evaluate the presence of DE and compared with a surgical/histological diagnosis as the reference standard were included. The gold standard surgical diagnosis of endometriosis was made if any of the following criteria was satisfied: (1) histological confirmation of endometriosis in at least one resected sub-peritoneal nodule; (2) visualization and palpation of a sub-peritoneal nodule without biopsy and another histologically proven location of endometriosis; (3) visualization of complete obliteration of the POD\textsuperscript{16}, although it is important to note that the latter is not always due to rectosigmoid DE, as adhesions may be secondary to adhesions from uterosacral ligaments DE to the rectum, superficial bowel endometriosis to the uterus and previous surgeries, amongst other causes. The studies were included if they were prospective cohort studies including women of reproductive age presenting with a clinical suspicion of DE based on symptoms and/or physical examination from any healthcare center setting.
All types of imaging modalities for deep endometriosis of the rectum/rectosigmoid were included, including variations of conventional techniques, such as bowel preparation, rectal water or gel contrast mediums, with the outcome being the presence and position of DE. Imaging modalities were then assessed separately and as a group. Studies were only included if there was sufficient reported data to construct a 2 x 2 contingency tables. Only studies including at least 10 affected and 10 unaffected women by the reference standard were considered eligible to reduce the risk of selection bias. There was no imposition on any language restriction.

Information sources
Searches were conducted using PubMed, Medline, Scopus and Embase to identify studies published from inception until May 2020 and screened by one author (B.G.) to identify potentially eligible studies. No filters were applied to the databases to reduce any possible omissions of relevant studies. Additionally, authors hand-searched the references from included studies and related reviews. Authors of primary studies were contacted, when necessary.

Search
The above databases were searched using the following terms: (deep AND endometriosis) AND (imaging OR ultrasound OR ultrasonography OR sonography OR (magnetic resonance) OR (shift imaging) OR (proton spin) OR (spin echo) OR MRI OR NMR OR MR OR (computed tomography) OR (computer assisted tomography) OR (beam tomography) OR (Computerized Axial Tomography) OR CT OR CAT) (Appendix 1). These were then
screened for studies that specifically assessed rectal/rectosigmoid DE to ensure that studies that may have adopted outdated or inaccurate DE descriptions were not excluded.

Study selection

The records were screened based on titles and abstracts by one author (B.G.), who obtained the full texts of the records considered to be potentially eligible. Two authors (B.G. and G.C.) independently and blindly examined these full text texts for compliance with the inclusion criteria and select eligible studies. Correspondence with study investigators was made as required, to clarify study eligibility. When papers updated previous publications, which were performed on either part thereof or all of the same study population, the most recent and complete study was used to avoid duplicating the participants or studies, assuming that some if not all of the previous study participants would be included. Similarly, in inter-observer diagnostic studies where two independent reviewers were used, only the most senior and accurate reviewer’s results were included. Studies were excluded if they were conference abstracts, case reports/series, descriptive, systematic reviews and those that were retrospective in nature whether this applied to the study design, review of the case notes or the index test(s) being performed following the reference test. Disagreements were solved by consulting the other author (M.L.). The selection process was documented with a “PRISMA” flow chart\textsuperscript{14}.

Data items

Data were extracted by one author (B.G), including the name of the first author and year of publication, which will be used to identify the study, country, period of enrolment, setting, age (mean and SD), body mass index (mean and SD), presenting symptoms
(dysmenorrhea, chronic pelvic pain, dyspareunia, bowel symptoms, urinary symptoms and infertility), imaging method(s), location of DE, and the results of the 2 x 2 table (true positives, false negatives, true negatives, false positives).

Risk of bias in the individual studies and Quality Assessment

The risk of bias, applicability and methodological quality of each study were independently evaluated by two authors (B.G. and M.L.) as suggested by QUADAS-2, evaluating the following four domains: 1) patient selection, 2) index text, 3) reference standard, and 4) flow and timing (only risk of bias) (Appendix 2). A summary score to calculate the overall quality for each study was not performed.

Risk of bias across studies

In view of the difficulty of detecting and correcting for publication and other reporting biases, we tried to minimize their potential impact by ensuring a comprehensive search for eligible studies, and by being cognisant for duplication of data.

Synthesis of results

We performed mixed-effects diagnostic meta-analysis to determine overall pooled sensitivity and specificity, from which the likelihood ratio of positive and negative tests (LR+, LR−), diagnostic odds ratios (DOR), and area under the curve (AUC) of summary receiver-operating characteristics (sROC) with their respective 95% confidence intervals (CIs) for all
diagnostic modules were calculated. This method requires at least four studies for each meta-analysis\textsuperscript{15}. Likelihood ratios of positive and negative tests (LRs) were parameters that characterizes the utility of a diagnostic module based on pre-test probability of disease. The DOR measures test accuracy that combined both LRs (DOR = LR+ /LR-) and represents how much greater the odds are of having rectosigmoid deep endometriosis for subjects with an abnormal diagnostic test compared with those subjects with a normal test. The AUC is the ability of the imaging technique to correctly classify those with and without rectosigmoid deep endometriosis. We produced forest plots of sensitivity and specificity for diagnostic modules that have adequate studies to be assessed. We also plotted sROC to illustrate AUC and the relationship between sensitivity and specificity. Sub-group analyses, where possible, were performed using the same methods.

We evaluated the presence and magnitude of heterogeneity for sensitivity and specificity using the Cochran's Q test and the I\textsuperscript{2} index. A P-value of Cochran's Q test <0.1 suggests the presence of heterogeneity. The I\textsuperscript{2} index describes the percentage of total variation across studies that can be explained by heterogeneity but not chance. I\textsuperscript{2} values of 25\%, 50\% and 75\% would be considered to indicate low, moderate and high heterogeneity, respectively\textsuperscript{21}. Heterogeneity was also visualized with 95\% prediction interval in the sROC.
Subgroup analyses were sought for 2-D TVS and 2-D MRI where enough studies capacitated the analysis.

We assessed potential publication bias using the Deeks Funnel Plot asymmetry test which computed a regression of diagnostic log odds ratio against $1/\sqrt{\text{effective sample size}}$, weighted by effective sample size. A $p$-value $< 0.10$ for the slope coefficient suggests significant asymmetry and possible publication bias. All analyses were performed using STATA version 16.1 for Windows (Stata Corporation, College Station, TX, USA).
Results

Search results

The literature search of the PubMed, Medline, Scopus, Embase and Google Scholar databases identified 1,977 references, from inception until 01 May 2020. The study selection as per the PRISMA flow diagram are represented in Figure 1. There were 1,061 duplicates, following which a further 591 studies were excluded as they were obviously irrelevant by screening the titles. A further 194 studies were eliminated after reading the abstracts as they were either retrospective or did not answer the research question. Full texts of the remaining 133 articles were reviewed, of which 88 articles were excluded as they did not meet the inclusion criteria or had overlapping study populations. Finally, there were 45 studies considered eligible studies of which 30 studies specifically assessed the rectum and rectosigmoid that were included in the analysis.

The 30 studies included a total of 3,374 patients with a median of 92 women per study (range 23 to 376; interquartile range 56.5 – 138.25). Of the 30 studies, 21 were conducted in Europe, 5 were in South America, 1 in Australia, 1 in Asia and 1 in the Middle East. A total of 22 studies assessed TVS (2,738 patients) of which 12 used 2-D TVS (1,684 patients), 5 assessed 3-D TVS with rectal water contrast (RWC) (502 patients), 1 used sonovaginography, 1 assessed three dimensional (3-D) TVS and 3 assessed TVS with bowel preparation (BP). A total of 9 studies assessed MRI (852 patients), of which one compared two dimensional (2-D) and 3-D MRI, one compared MRI with and without rectal ultrasound gel and the remaining assessed 2-D MRI (817 patients). Six studies assessed CT (402 patients), with three
assessing CT\textsuperscript{123, 145, 152} and three assessing CT colonography (CTC)\textsuperscript{110, 118, 120}. Eight studies assessed RES (850 patients)\textsuperscript{115, 135, 136, 139, 141, 143, 147, 151} and three used double-contrast barium enemas (DCBE)\textsuperscript{133, 141, 153}. The pre-test probabilities of disease for TVS, MRI, CT and RES were 39\%, 40\%, 62\% and 36\%, respectively. The study characteristics are shown in Table 1 and the summary findings are shown in Table 2.

Methodological quality of included studies

The quality of the studies as per QUADAS-2 are presented in figure 2 and figure 3. In summary, the majority of studies were of poor methodological quality.

Nineteen studies were assessed to be low risk of patient selection bias\textsuperscript{112, 115, 116, 121, 125, 126, 130, 132, 133, 135, 137, 138, 141, 142, 146, 147, 151, 154}, eight were high risk\textsuperscript{118, 120, 122, 127, 136, 139, 143, 145} and three were unclear\textsuperscript{110, 152, 153}. Regarding the index test domain, twenty studies were deemed low risk\textsuperscript{110, 118, 121, 125, 127, 130, 132, 133, 136-139, 141-143, 145-147, 151, 154}, eight were high risk\textsuperscript{112, 115, 116, 120, 126, 135, 152, 153} and two were unclear\textsuperscript{122, 123} with how the index test was conducted and interpreted. All studies were considered high risk of bias for the reference standard domain as surgeons were not blinded to the results of the index test, that is the pre-operative imaging performed. With regards to the flow and timing domain, three were deemed high risk\textsuperscript{132, 135, 147}, eight were unclear\textsuperscript{115, 121, 122, 127, 130, 139, 143, 151} and the remaining nineteen were low risk\textsuperscript{110, 112, 116, 118, 120, 123, 125, 126, 133, 136-138, 141, 142, 145, 146, 152-154}. All the studies were considered low risk for the biases concerning applicability given that studies were only included if they (a) had a clinically relevant population that would have undergone index test in real practice, (b) had sufficient information in the index test, of which all imaging modalities were included, and (c) had laparoscopy/laparotomy as a reference test (Appendix 2).
Diagnostic performance of TVS for detection of rectal/rectosigmoid DE

The overall pooled sensitivity and specificity, from which LR+, LR- and DOR were calculated, of detecting rectal/rectosigmoid DE with TVS were 89% (95% CI 83 – 93%), 97% (95% CI 95 – 98%), 28.8 (95% CI 16.2 – 51.0), 0.12 (95% CI 0.08 – 0.18) and 248 (95% CI 104 – 594), respectively. There was significant heterogeneity for sensitivity (I², 86.7%; Cochran Q, 157.9; p = 0.00) and specificity (I², 81.7%; Cochran Q, 114.5; p = 0.00) (Figure 4). The summary ROC is displayed in figure 5 with an AUC of 98% (95% CI 96 – 99%).

Sub-group analyses of TVS

The pooled sensitivity and specificity, from which LR+, LR- and DOR were calculated, of detecting rectal/rectosigmoid DE with 2-D TVS were 84% (95% CI 74 – 90%), 97% (95% CI 93 – 98%), 24.2 (95% CI 11.0 – 53.2), 0.17 (95% CI 0.10 – 0.28), and 144 (95% CI 46 – 453), respectively. There was significant heterogeneity for sensitivity (I², 86.8%; Cochran Q, 83.3; p = 0.00) and specificity (I², 84.4%; Cochran Q, 70.7 p = 0.00) (Figure 4). The ROC is displayed in Figure 5 with an AUC of 97% (95% CI 95 – 98%).
The pooled sensitivity and specificity, from which LR+, LR- and DOR were calculated, of detecting rectal/rectosigmoid DE with TVS-RWC were 88% (95% CI 80 – 93%), 97% (95% CI 93 – 98%), 27.0 (95% CI 12.8 – 56.7), 0.13 (95% CI 0.08 – 0.21), and 214 (95% CI 80 – 572), respectively. There was significant heterogeneity for sensitivity (I², 68.7%; Cochran Q, 12.8; p = 0.01) and specificity (I², 58.3%; Cochran Q, 9.6 p = 0.05) (Figure 4). The ROC is displayed in Figure 5 with an AUC of 98% (95% CI 96 – 99%).

There was no evidence for publication bias for any of these analyses (p = 0.94, p = 0.53 and 0.23, respectively) (Figure 6).

Given the low number of studies, it was not possible to perform sub-analyses for TVS-BP or SVG. However, the results of TVS were improved with BP with sensitivities and specificities of 97 – 100% / 96% - 100% 116, 132, 146, respectively. SVG outperformed 2-D TVS for the rectosigmoid with a sensitivity and specificity of 100% and 93%, retrospectively, however it was similar with regards to the rectum, which appears to be included in their definition of the retrocervix, with a sensitivity of 84% and specificity of 95%122.

Diagnostic performance of MRI for detection of rectal/rectosigmoid DE

The overall pooled sensitivity and specificity, from which LR+, LR- and DOR were calculated, of detecting rectal/rectosigmoid DE with MRI were 86% (95% CI 79 – 81%), 97% (95% CI 94 – 97%), 21.0 (95% CI 13.4 – 33.1), 0.15 (95% CI 0.09 – 0.23), and 144 (95% CI 70 – 297), respectively. There was significant heterogeneity for sensitivity (I², 61.0%;
Cochran Q, 20.5; p = 0.01) and specificity (I2, 34.0%; Cochran Q, 12.1; p = 0.15) (Figure 7). The ROC is displayed in Figure 8 with an AUC of 97% (95% CI 95 – 98%).

Sub-group analyses of MRI

The pooled sensitivity and specificity, from which LR+, LR- and DOR were calculated, of detecting rectal/rectosigmoid DE with 2-D MRI were 85% (95% CI 77 – 90%), 96% (95% CI 93 – 97%), 19.9 (95% CI 12.9 – 30.7), 0.16 (95% CI 0.11 – 0.24), and 125 (95% CI 64 – 243), respectively. There was significant heterogeneity for sensitivity (I2, 62.7%; Cochran Q, 16.1; p = 0.01) and specificity (I2, 35.8%; Cochran Q, 9.3; p = 0.16) (Figure 7). The ROC is displayed in Figure 8 with an AUC of 96% (95% CI 94 – 98%).

There was no evidence for publication bias for any of these analyses (p = 0.98 and p = 0.74, respectively) (Figure 9).

Given the low number of studies, it was not possible to perform sub-analyses for 3-D MRI or MRI with rectal ultrasound gel. 3-D MRI performed similarly to 2-D MRI with sensitivities and specificities between radiologists ranging from 89-100% and 94-100%, respectively127. MRI with rectal ultrasound gel outperformed 2-D MRI with a sensitivity of 99% and specificity of 96%138.

Diagnostic performance of CT for detection of rectal/rectosigmoid DE
The overall pooled sensitivity and specificity, from which LR+, LR- and DOR were calculated, of detecting rectal/rectosigmoid DE with CT were 93% (95% CI 84 – 97%), 95% (95% CI 81 – 99%), 20.3 (95% CI 4.3 – 94.9), 0.07 (95% CI 0.03 – 0.19), and 280 (95% CI 28 – 2826), respectively. There was significant heterogeneity for sensitivity (I2, 84.3%; Cochran Q, 31.9; p = 0.00) and specificity (I2, 80.2%; Cochran Q, 25.2; p = 0.00) (Figure 10). The ROC is displayed in Figure 8 with an AUC of 98% (95% CI 96 – 98%). There was no evidence for publication bias (p = 0.10) (Figure 9). It was not possible to perform sub-analyses of CT colonography, although the results ranged widely with one study performing poorly with a sensitivity and specificity of 68% and 67%, respectively compared with the other two, ranging from 93 – 95% and 87 – 93%.

Diagnostic performance of RES for detection of rectal/rectosigmoid DE

The overall pooled sensitivity and specificity, from which LR+, LR- and DOR were calculated, of detecting rectal/rectosigmoid DE with RES were 92% (95% CI 87 – 95%), 98% (95% CI 96 – 99%), 37.1 (95% CI 21.1 – 65.4), 0.08 (95% CI 0.05 – 0.14), and 455 (95% CI 196 – 1054), respectively. There was significant heterogeneity for sensitivity (I2, 42.8%; Cochran Q, 12.2; p = 0.09) and specificity (I2, 13.0%; Cochran Q, 8.0; p = 0.33) (Figure 10). The ROC is displayed in Figure 8 with an AUC of 99% (95% CI 100 – 0%). There was no evidence of publication bias (p = 0.37) (Figure 9).
Diagnostic performance of double-contrast barium enema for detection of rectal/rectosigmoid DE

As there were only three studies assessing DCBE, it was not possible to perform a meta-analysis, although the reported sensitivities, specificities, positive and negative predictive values for the studies were 84.7% / 93.7% / 98.0% / 62.5%133, 88% / 54% / 78% / 70%141, and 96.4% / 100% / 100% / 98%153, respectively.
Discussion

Summary of evidence

In 2004, Bazot, et al. correlated the ultrasound and surgical findings of deep pelvic endometriosis and, since then, there has been a considerable number of studies published pre-operatively assessing the presence and various characteristics of deep endometriosis. Of all the imaging modalities used for the pre-operative diagnosis of DE, TVS is the most widely studied, and is often used as the first-line imaging technique given its accessibility, relative low cost and invasiveness, although adequate training is required. MRI, whilst more expensive, has the advantage of being able to assess peritoneal and extra-pelvic lesions, as well as adhesions. The results of this meta-analysis for the detection of rectal/rectosigmoid DE revealed a superiority of TVS, albeit marginal, with an overall pooled sensitivity of 89%, specificity of 97%, DOR of 248 and AUC of 98% compared to MRI with an overall pooled sensitivity of 86%, specificity of 97%, DOR of 144 and AUC of 97%. The sub-group analyses, particularly of TVS-WRC, was comparable to 2-D TVS. Interestingly, whilst CT also performed well, with a sensitivity, specificity, DOR and AUC of 93%, 95%, 380 and 98%, respectively, RES outperformed all the techniques with a sensitivity, specificity, DOR and AUC of 92%, 98%, 455 and 99%, respectively. However, whilst RES might be the preferred imaging modality for detecting rectal DE, it is more invasive and not as widely available or used in this context. Furthermore, it is not possible to assess the remainder of the pelvis for DE with RES, and therefore women would require multiple invasive investigations for a complete assessment.
Interpretation of results

Our results differ slightly when compared to other meta-analyses. Nisenblat, et al.13, which assessed a total of 21 studies (2,222 participants), of which the pooled overall sensitivity and specificity for the TVS (14 studies), MRI (6 studies) and RES (4 studies) were 90% / 96%, 92% / 96% and 91% / 96%, respectively. Although similar, the differences are likely due our current meta-analysis including more studies with the limitation of only those with at least 10 affected/unaffected patients. More recently, in 2019 Noventa, et al.155 performed meta-analysis of only head-to-head studies and found TVS to be superior to MRI with sensitivities of 85% and 83%, respectively, and again TVS was superior to RES with sensitivities of 89% and 88%, respectively, and finally RES was superior to MRI with sensitivities of 91% and 84%, respectively. However, they included retrospective studies, which were limited to those determined to be well-conducted. Moura, et al.156 performed a meta-analysis comparing MRI and TVS in the diagnosis of rectosigmoid DE in the same population, of which both modalities were found to have similar a similar sensitivity, specificity and AUC of 90% / 96% / 95% and 90% / 96% / 90%, respectively. This was similar to the more recent comparative study of MRI and TVS, including comparisons between enhanced techniques, by Pereira, et al.8 with a sensitivity and specificity of 80% / 94% and 82% / 94%, respectively. The limitation of the latter two studies are the small number of included studies, being 8, and 11, respectively, although they are both well conducted and confirm the diagnostic accuracy of both TVS and MRI.
The strengths of this study include the inclusion of prospective studies thereby reducing any potential selection bias that may be associated with retrospective studies, regardless of their quality. Also, we believe that the rationale to ensuring, in advance, that all studies had to have at least 10 affected and non-affected participants would further reduce the risk of selection bias, making the results more applicable to the general population. Furthermore, the initial search was purposely broad to ensure all potential studies are included, particularly given the inconsistencies with the descriptions of rectal and rectosigmoid DE. Prior to the IDEA consensus, rectal/rectosigmoid DE had also been referred to as retrocervical, POD, posterior cul-de-sac and bowel DE. Therefore, by including all studies with any mention of “deep” and “endometriosis”, the risk of any of these studies were minimised, although the fact that studies may have still have been missed is also a limitation of this and any other similar systematic-analyses. Moving forward, the introduction of the IDEA terms and definitions into ultrasound practice may help to unify the anatomical classification of rectal and sigmoid endometriosis internationally which in turn may result in more meaningful meta-analyses of pooled data.

Limitations

Whilst there was an attempt to obtain answers with regards to whether different contrasts or bowel preparations improved the performance of the imaging modalities, the sub-analyses were not possible due to the limited number of studies. There was also substantial heterogeneity and high risk of bias in all of the included studies, which decreases the quality of the evidence. Surgeons were not blinded to the pre-operative imaging, and the risk of misdiagnosis due to lack of expertise or histopathological findings are all potential causes of
bias. Furthermore, there is the risk of patient selection bias given that the pre-test probabilities of disease in this analysis ranged from 39% to 66%, which is significantly higher than the reported incidence of rectal/rectosigmoid DE, ranging from 5.3% to 12%\textsuperscript{157, 158}. This is likely due to the inevitable self-selection of women referred to specialised units where they are examined by experienced clinicians, whom would be more attune and subsequently likely to make the correct diagnosis of rectal/rectosigmoid DE. All of these are potential biases which need to be considered. Also, the studies included span from 2007 to 2020, during which time there have been significant advancements in technology as well as improvement in techniques and expert experience, thus it is possible that the accuracy of the imaging modalities may be higher than those obtained.

**Conclusions**

All imaging modalities perform well for the pre-operative diagnosis of rectal/rectosigmoid DE, although there are key differences for their applicability. The sensitivity of TVS for the detection of rectal/rectosigmoid DE seems to be slightly better than MRI, and inferior to RES, however the remainder of the pelvis is unable to be assessed with the latter. Specificity of both MRI and TVS were excellent. As TVS is simpler, faster, more readily available, with a very robust guideline (IDEA) to homogenize test performance that increases its reproducibility all around the globe, we believe it should be the first line diagnostic tool for the women with suspected DE. More studies are required using standardised definitions as those proposed by IDEA, as well as the various contrast mediums that may improve the accuracy of the imaging modalities.
Acknowledgements: None

Conflicts of Interest: None

Funding: None
References


8. Pereira AMG, Brizon VSC, Carvas Junior N, de Carvalho MSR, Cuvero EM, Buehler AM, Pereira TDV and Lopes RGC. Can Enhanced Techniques Improve the Diagnostic Accuracy of Transvaginal Sonography and Magnetic Resonance Imaging for Rectosigmoid


33. Cosma S, Salgarello M, Ceccaroni M, Gorgoni G, Riboni F, La Paglia E, Danese S and Benedetto C. Accuracy of a new diagnostic tool in deep infiltrating endometriosis:


39. Thomeer MG, Steensma AB, Van Santbrink EJ, Willemssen FE, Wielopolski PA, Hunink MG, Sprok S, Laven JS and Krestin GP. Can magnetic resonance imaging at 3.0-


146. Abrao MS, Goncalves MO, Dias JA, Jr., Podgaec S, Chamie LP and Blasbalg R. Comparison between clinical examination, transvaginal sonography and magnetic


Legends

**Figure 1.** Flow of studies identified in literature for systematic review on imaging modalities for the pre-operative diagnosis of rectal/rectosigmoid deep endometriosis.

**Figure 2.** QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies-2) quality evaluation of all 30 included studies.

**Figure 3.** Traffic-light plot summarising the authors’ review of the QUADAS-2 risk of bias and applicability concerns.

**Figure 4.** Forest plots of studies using transvaginal ultrasound for the evaluation of rectal/rectosigmoid deep endometriosis. Imaging modalities analysed are (a) ALL transvaginal ultrasound and (b) sub-analysis of 2-D transvaginal ultrasound and (c) transvaginal ultrasound using rectal water contrast, displaying the pooled sensitivity, specificity and heterogeneity statistics (Cochran’s Q and I²).

**Figure 5.** Summary receiver-operating characteristics (ROC) of studies using transvaginal ultrasound for the evaluation of rectal/rectosigmoid deep endometriosis. Imaging modalities analysed are (a) ALL transvaginal ultrasound (b) sub-analysis of 2-D transvaginal ultrasound and (c) sub-analysis of transvaginal ultrasound with rectal water contrast.

**Figure 6.** Deeks funnel plots of studies using transvaginal ultrasound for the evaluation of rectal/rectosigmoid deep endometriosis, showing no publication bias. Imaging modalities analysed include (a) ALL transvaginal ultrasound (b) sub-analysis of 2-D transvaginal ultrasound and (c) sub-analysis of transvaginal ultrasound with rectal water contrast. ESS, effective sample size.
Figure 7. Forest plots of studies using magnetic resonance imaging for the evaluation of rectal/rectosigmoid deep endometriosis. Imaging modalities analysed are (a) ALL magnetic resonance imaging and (b) sub-analysis of 2-D magnetic resonance, displaying the pooled sensitivity, specificity and heterogeneity statistics (Cochran’s Q and I²).

Figure 8. Summary receiver-operating characteristics (ROC) of studies included for the evaluation of rectal/rectosigmoid deep endometriosis. Imaging modalities analysed are (a) ALL magnetic resonance imaging (b) sub-analysis of 2-D magnetic resonance imaging (c) computed tomography and (d) rectal endoscopic sonography.

Figure 9. Deeks funnel plots of studies included for the evaluation of rectal/rectosigmoid deep endometriosis, showing no publication bias. Imaging modalities analysed include (a) ALL magnetic resonance imaging (b) sub-analysis of 2-D magnetic resonance imaging (c) computed tomography and (d) rectal endoscopic sonography. ESS, effective sample size.

Figure 10. Forest plots of studies included for the evaluation of rectal/rectosigmoid deep endometriosis. Imaging modalities analysed are (a) computed tomography and (b) rectal endoscopic sonography, displaying the pooled sensitivity, specificity and heterogeneity statistics (Cochran’s Q and I²).

Appendix 1. Search strategy example using Medline via OVID (1946 – present) database.

Appendix 2. QUADAS-2 tool used for the assessment of the methodological assessment of included studies.
1,977 records identified by electronic searches

1,061 duplicates removed

918 records screened through titles

591 records excluded as obviously irrelevant by reading titles

327 records screened through titles and abstracts

194 records excluded as retrospective or irrelevant by reading titles and abstracts

133 full-text articles assessed for eligibility

88 full-text articles excluded:
- 19 for differing study design
- 6 for having overlapping populations
- 3 as retrospective
- 34 for not meeting inclusion criteria
- 7 for different standard references
- 19 for different outcomes

45 studies eligible studies assessing Deep Endometriosis (DE)

15 studies excluded as did not assess rectal/rectosigmoid DE

30 studies included in review
<table>
<thead>
<tr>
<th>First Author</th>
<th>Year</th>
<th>Country</th>
<th>Setting</th>
<th>n</th>
<th>Mean Age</th>
<th>Dysmenorrhea (%)</th>
<th>Chronic Pelvic Pain (%)</th>
<th>Dyspareunia (%)</th>
<th>Bowel Symptoms (%)</th>
<th>Urinary Symptoms (%)</th>
<th>Infertility (%)</th>
<th>Index Test(s)</th>
<th>Observers</th>
<th>Reference Standard</th>
<th>Cases with rectosigmoid DE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abrao</td>
<td>2007</td>
<td>Brazil</td>
<td>Single centre</td>
<td>104</td>
<td>33.8</td>
<td>51</td>
<td>17</td>
<td>65</td>
<td>61</td>
<td>14</td>
<td>53</td>
<td>TVS-BP and MRI</td>
<td>Single</td>
<td>Surgery and histopathology</td>
<td>54</td>
</tr>
<tr>
<td>Alborzi</td>
<td>2018</td>
<td>Iran</td>
<td>Multicentre</td>
<td>317</td>
<td>31</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>TVS, RES and MRI</td>
<td>Single</td>
<td>Surgery and histopathology</td>
<td>52</td>
</tr>
<tr>
<td>Baggio</td>
<td>2016</td>
<td>Italy</td>
<td>Single centre</td>
<td>92</td>
<td>32</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>TVS and CTC</td>
<td>Two</td>
<td>Surgery and histopathology</td>
<td>49</td>
</tr>
<tr>
<td>Bahri</td>
<td>2006</td>
<td>France</td>
<td>Single centre</td>
<td>37</td>
<td>35.8</td>
<td>24</td>
<td>100</td>
<td>49</td>
<td>37</td>
<td>22</td>
<td>-</td>
<td>RES</td>
<td>Single</td>
<td>Surgery and histopathology</td>
<td>8</td>
</tr>
<tr>
<td>Barra</td>
<td>2020</td>
<td>Italy</td>
<td>Single centre</td>
<td>68</td>
<td>35.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3D-TVS-RWC and CTC</td>
<td>Single</td>
<td>Surgery and histopathology</td>
<td>37</td>
</tr>
<tr>
<td>Bazot</td>
<td>2009</td>
<td>France</td>
<td>Single centre</td>
<td>92</td>
<td>31.8</td>
<td>85.9</td>
<td>68.5</td>
<td>34.8</td>
<td>3.2</td>
<td>22.8</td>
<td>-</td>
<td>TVS, RES and MRI</td>
<td>Single</td>
<td>Surgery and histopathology</td>
<td>63</td>
</tr>
<tr>
<td>Bazot</td>
<td>2013</td>
<td>France</td>
<td>Single centre</td>
<td>23</td>
<td>34</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>MRI and 3-D MRI</td>
<td>Two</td>
<td>Surgery and histopathology</td>
<td>13</td>
</tr>
<tr>
<td>Bergamini</td>
<td>2010</td>
<td>Italy</td>
<td>Multicentre</td>
<td>61</td>
<td>33.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>TVS-RWC, RES and DCBE</td>
<td>Single</td>
<td>Surgery and histopathology</td>
<td>51</td>
</tr>
<tr>
<td>Biscaldi</td>
<td>2007</td>
<td>Italy</td>
<td>Single centre</td>
<td>98</td>
<td>34</td>
<td>88.8</td>
<td>49</td>
<td>74.5</td>
<td>27.6</td>
<td>-</td>
<td>23.5</td>
<td>CT</td>
<td>Two</td>
<td>Surgery and histopathology</td>
<td>76</td>
</tr>
<tr>
<td>Faccioli</td>
<td>2009</td>
<td>Italy</td>
<td>Single centre</td>
<td>77</td>
<td>31.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>MRI and BE</td>
<td>Two</td>
<td>Surgery and histopathology</td>
<td>59</td>
</tr>
<tr>
<td>Fedele</td>
<td>1998</td>
<td>Italy</td>
<td>Single centre</td>
<td>140</td>
<td>30.2</td>
<td>-</td>
<td>37.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>47.9</td>
<td>RES</td>
<td>Single</td>
<td>Surgery and histopathology</td>
<td>9</td>
</tr>
<tr>
<td>Ferrero</td>
<td>2011</td>
<td>Italy</td>
<td>Single centre</td>
<td>96</td>
<td>33.4</td>
<td>90</td>
<td>63.5</td>
<td>55.1</td>
<td>40.6</td>
<td>-</td>
<td>33.3</td>
<td>TVS-RWC and CT</td>
<td>Four</td>
<td>Surgery and histopathology</td>
<td>48</td>
</tr>
<tr>
<td>Ferrero</td>
<td>2017</td>
<td>Italy</td>
<td>Single centre</td>
<td>70</td>
<td>35.7</td>
<td>91.4</td>
<td>78.6</td>
<td>74.3</td>
<td>62.9</td>
<td>-</td>
<td>-</td>
<td>TVS-RWC and Single</td>
<td>Single</td>
<td>Surgery and histopathology</td>
<td>40</td>
</tr>
</tbody>
</table>

This article is protected by copyright. All rights reserved.
<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Country</th>
<th>Setting</th>
<th>Cases</th>
<th>30-day mortality</th>
<th>90-day mortality</th>
<th>Incidence IA</th>
<th>Incidence IB</th>
<th>Incidence IC</th>
<th>Follow-up</th>
<th>Biopsy Method</th>
<th>Ref.</th>
<th>Surgery and histopathology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goncalves</td>
<td>2009</td>
<td>Brazil</td>
<td>Single centre</td>
<td>194</td>
<td>34.2</td>
<td>56</td>
<td>20</td>
<td>62</td>
<td>58</td>
<td>9</td>
<td>50</td>
<td>TVS-BP</td>
<td>Single Surgery and histopathology</td>
</tr>
<tr>
<td>Griffiths</td>
<td>2008</td>
<td>UK</td>
<td>Single centre</td>
<td>32</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>RES</td>
<td>Single Surgery and histopathology</td>
</tr>
<tr>
<td>Guerriero</td>
<td>2008</td>
<td>Italy</td>
<td>Single centre</td>
<td>88</td>
<td>33</td>
<td>81</td>
<td>-</td>
<td>45</td>
<td>-</td>
<td>-</td>
<td>11</td>
<td>TVS</td>
<td>Single Surgery and histopathology</td>
</tr>
<tr>
<td>Guerriero</td>
<td>2017</td>
<td>Italy</td>
<td>Single centre</td>
<td>159</td>
<td>33</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>TVS, 3D-TVS and MRI</td>
<td>Single Surgery</td>
</tr>
<tr>
<td>Holland</td>
<td>2013</td>
<td>England</td>
<td>Multicentre</td>
<td>198</td>
<td>35</td>
<td>72.2</td>
<td>49.5</td>
<td>45.9</td>
<td>9.6</td>
<td>-</td>
<td>-</td>
<td>TVS</td>
<td>Two Surgery and histopathology</td>
</tr>
<tr>
<td>Hottat</td>
<td>2009</td>
<td>Belgium</td>
<td>Single centre</td>
<td>41</td>
<td>33</td>
<td>46</td>
<td>63</td>
<td>12</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>MRI</td>
<td>Two Surgery and histopathology</td>
</tr>
<tr>
<td>Hudelist</td>
<td>2009</td>
<td>Austria</td>
<td>Single centre</td>
<td>195</td>
<td>33</td>
<td>77.5</td>
<td>22</td>
<td>34.5</td>
<td>14.5</td>
<td>2</td>
<td>17</td>
<td>TVS</td>
<td>Two Surgery and histopathology</td>
</tr>
<tr>
<td>Hudelist</td>
<td>2011</td>
<td>Austria</td>
<td>Multicentre</td>
<td>129</td>
<td>32.2</td>
<td>86</td>
<td>34.8</td>
<td>55.8</td>
<td>30.2</td>
<td>4.6</td>
<td>15.5</td>
<td>TVS</td>
<td>Single Surgery and histopathology</td>
</tr>
<tr>
<td>Hudelist</td>
<td>2013</td>
<td>Austria</td>
<td>Single centre</td>
<td>117</td>
<td>31.6</td>
<td>99</td>
<td>27</td>
<td>63</td>
<td>26</td>
<td>8</td>
<td>19</td>
<td>TVS-RWC and DCBE</td>
<td>Two Surgery and histopathology</td>
</tr>
<tr>
<td>Jiang</td>
<td>2017</td>
<td>China</td>
<td>Single centre</td>
<td>207</td>
<td>32.7</td>
<td>86.4</td>
<td>-</td>
<td>64.1</td>
<td>47</td>
<td>-</td>
<td>-</td>
<td>TVS and RES</td>
<td>Single Surgery and histopathology</td>
</tr>
<tr>
<td>Leon</td>
<td>2014</td>
<td>Chile</td>
<td>Single centre</td>
<td>55</td>
<td>32.9</td>
<td>100</td>
<td>90</td>
<td>76</td>
<td>67</td>
<td>-</td>
<td>-</td>
<td>SVG</td>
<td>Single Surgery and histopathology</td>
</tr>
<tr>
<td>Piketty</td>
<td>2009</td>
<td>France</td>
<td>Single centre</td>
<td>133</td>
<td>32.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>TVS and RES</td>
<td>Single Surgery and histopathology</td>
</tr>
<tr>
<td>Reid</td>
<td>2018</td>
<td>Australia</td>
<td>Multicentre</td>
<td>376</td>
<td>-</td>
<td>64</td>
<td>-</td>
<td>44</td>
<td>36</td>
<td>-</td>
<td>21</td>
<td>TVS</td>
<td>Single Surgery and histopathology</td>
</tr>
<tr>
<td>Ribeiro</td>
<td>2007</td>
<td>Brazil</td>
<td>Single centre</td>
<td>37</td>
<td>35.8</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>RES and DCBE</td>
<td>Single Surgery and histopathology</td>
</tr>
<tr>
<td>Ros</td>
<td>2017</td>
<td>Brazil</td>
<td>Single centre</td>
<td>40</td>
<td>36.8</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>22.5</td>
<td>TVS-BP</td>
<td>Single Surgery and histopathology</td>
</tr>
</tbody>
</table>

This article is protected by copyright. All rights reserved.
Table 1. Characteristics of included studies. Only the first author of each study is given. All studies were prospective and included women with clinical suspicion of rectal/rectosigmoid deep endometriosis (DE). Observers refer to the number of observers involved with each imaging modality. 3-D, 3-dimensional; BP, bowel preparation; DCBE, double contrast barium enema; CT, computed tomography; CTC, CT colonography; MRI, magnetic resonance imaging; RES, transrectal endoscopic sonography; RWC, rectal water contrast; SVG, sonovaginography; TVS, transvaginal ultrasound.

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Country</th>
<th>Setting</th>
<th>Age (years)</th>
<th>BMI (kg/m²)</th>
<th>Size of lesion (cm)</th>
<th>Ulceration</th>
<th>Mass</th>
<th>FISTULA</th>
<th>Surgery and histopathology</th>
<th>Modality</th>
<th>Observers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stabile Ianora</td>
<td>2013</td>
<td>Italy</td>
<td>Single centre</td>
<td>33</td>
<td>33.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>CT</td>
<td>Two</td>
</tr>
<tr>
<td>Tammaa</td>
<td>2015</td>
<td>Austria</td>
<td>Single centre</td>
<td>65</td>
<td>30.2</td>
<td>95</td>
<td>32</td>
<td>63</td>
<td>15</td>
<td>23</td>
<td>TVS</td>
<td>Two</td>
</tr>
</tbody>
</table>

This article is protected by copyright. All rights reserved.
<table>
<thead>
<tr>
<th>Imaging Technique</th>
<th>Number of studies</th>
<th>Number of patients</th>
<th>Sensitivity (CI)</th>
<th>Specificity (CI)</th>
<th>DOR (CI)</th>
<th>AUC (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TVS - Overall</td>
<td>22</td>
<td>2,738</td>
<td>89% (83 – 93%)</td>
<td>97% (95 – 98%)</td>
<td>28.8 (16.2 – 51.0)</td>
<td>98% (96 – 99%)</td>
</tr>
<tr>
<td>2D-TVS</td>
<td>12</td>
<td>1,684</td>
<td>84% (74 – 90%)</td>
<td>97% (93 – 98%)</td>
<td>144 (46 – 453)</td>
<td>97% (95 – 98%)</td>
</tr>
<tr>
<td>TVS-RWC</td>
<td>5</td>
<td>502</td>
<td>88% (80 – 93%)</td>
<td>97% (93 – 98%)</td>
<td>214 (80 – 572)</td>
<td>98% (96 – 99%)</td>
</tr>
<tr>
<td>MRI - Overall</td>
<td>9</td>
<td>852</td>
<td>86% (79 – 81%)</td>
<td>97% (94 – 97%)</td>
<td>144 (70 – 297)</td>
<td>97% (95 – 98%)</td>
</tr>
<tr>
<td>2D-MRI</td>
<td>7</td>
<td>817</td>
<td>85% (77 – 90%)</td>
<td>96% (93 – 97%)</td>
<td>125 (64 – 243)</td>
<td>96% (94 – 98%)</td>
</tr>
<tr>
<td>CT</td>
<td>6</td>
<td>402</td>
<td>93% (84 – 97%)</td>
<td>95% (81 – 99%)</td>
<td>280 (28 – 2826)</td>
<td>98% (96 – 98%)</td>
</tr>
<tr>
<td>RES</td>
<td>8</td>
<td>850</td>
<td>92% (87 – 95%)</td>
<td>98% (96 – 99%)</td>
<td>455 (196 – 1054)</td>
<td>99% (100 – 0%)</td>
</tr>
</tbody>
</table>

Table 2. Summary of findings of the pooled results of the pre-operative diagnostic accuracy of imaging modalities for the detection of rectosigmoid deep endometriosis. 2-D, 2-dimensional; 3-D, 3-dimensional; CI, 95% confidence interval; CT, computed tomography; MRI, magnetic resonance imaging; RES, transrectal endoscopic sonography; RWC, rectal water contrast; TVS, transvaginal ultrasound.