Transvaginal ultrasound (TVS) versus Magnetic Resonance (MR) for diagnosing deep infiltrating endometriosis: a systematic review and meta-analysis

S Guerriero[1],L Saba[2], MA Pascual[3], S Ajossa[1], I Rodriguez[3], V Mais[1] , JL Alcazar[4].

[1]Department of Obstetrics and Gynecology, University of Cagliari, Policlinico Universitario Duilio Casula, Monserrato, Cagliari, Italy.

[2]Department of Radiology, Azienda Ospedaliero Universitaria (A.O.U.), Monserrato, Italy

[3]Department of Obstetrics, Gynecology, and Reproduction, Institut Universitari Dexeus, Barcelona, Spain

[4]Department of Obstetrics and Gynecology, Clínica Universidad de Navarra, School of Medicine, University of Navarra, Pamplona, Spain.

Corresponding author: Professor Stefano Guerriero, MD, Department of Obstetrics and Gynecology, Blocco Q, Azienda Ospedaliero Universitaria- Policlinico Duilio Casula Monserrato s.s. 554, Monserrato, 09045, Italy, University of Cagliari, Cagliari, Italy.

Running title: TVS and MR in DIE

Key words: Transvaginal ultrasound (TVS), Magnetic Resonance (MR), deep infiltrating endometriosis

Abstract

Objectives

To perform a systematic review of studies comparing the diagnostic accuracy of TVS and MRI in Deep Infiltrating Endometriosis (DIE) including only studies in which patients have been underwent both techniques.

Methods

An extensive search of papers comparing TVS and MRI in DIE was performed in Medline (Pubmed) and Web of Sciences from January 1989 to January 2016. Studies were considered eligible if they

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/uog.18961

reported on the use of TVS and MRI in the same set of patients for the preoperative detection of endometriosis in pelvic locations in women with clinical suspicion of DIE using the surgical data as a reference standard. Quality was assessed using QUADAS-2 tool. A random-effects model was used to determine overall pooled sensitivity, specificity, positive likelihood ratio (LR+) and negative likelihood ratio (LR–) and the diagnostic odds ratio (DOR).

Results

Of the 375 citations identified, 6 studies (n=424) were considered eligible. Pooled sensitivity, specificity, LR+ and LR– of MRI in detecting DIE in the recto-sigmoid for MRI were 0.85 (95% CI, 0.78–0.90), 0.95 (95% CI, 0.83–0.99), 18.4 (95% CI, 4.7–72.4) and 0.16 (95% CI, 0.11–0.24), respectively. Pooled sensitivity, specificity, LR+ and LR– of TVS in detecting DIE in the recto-sigmoid for TVS were 0.85 (95% CI, 0.68–0.94), 0.96 (95% CI, 0.85–0.99), 20.4 (95% CI, 4.7–88.5) and 0.16 (95% CI, 0.07–0.38), respectively. DOR was 116 (95% CI, 23-585) and 127 (95% CI, 14 - 1126), respectively.

Pooled sensitivity, specificity, LR+ and LR– of MRI in detecting DIE in the rectovaginal septum for MRI were 0.66 (95% CI, 0.51–0.79), 0.97 (95% CI, 0.89–0.99), 22.5 (95% CI, 6.7–76.2) and 0.38 (95% CI, 0.23–0.52), respectively. Pooled sensitivity, specificity, LR+ and LR– of TVS in detecting DIE in the rectovaginal septum for TVS were 0.59 (95% CI, 0.26–0.86), 0.97 (95% CI, 0.94–0.99), 23.5 (95% CI, 9.1–60.5) and 0.42 (95% CI, 0.18–0.97), respectively. DOR was 65 (95% CI, 21- 204) and 56 (95% CI, 11 - 275), respectively.

Pooled sensitivity, specificity, LR+ and LR– of MRI in detecting DIE in the uterosacral ligaments for MRI were 0.70 (95% CI, 0.55–0.82), 0.93 (95% CI, 0.87–0.97), 10.4 (95% CI, 5.1–21.2) and 0.32 (95% CI, 0.20–0.51), respectively. Pooled sensitivity, specificity, LR+ and LR– of TVS in detecting DIE in the uterosacral ligaments for TVS were 0.67 (95% CI, 0.55–0.77), 0.86 (95% CI, 0.73–0.93),

4.8 (95% CI, 2.6–9.0) and 0.38 (95% CI, 0.29–0.50), respectively. DOR was 32 (95% CI, 12- 85) and 12 (95% CI, 7- 24), respectively.

Wide confidence intervals of pooled sensitivities, specificities and DOR were present for both techniques in all the considered locations. Heterogeneity was moderate or high for sensitivity and specificity for TVS and MRI in most locations assessed. According to QUADAS2, the quality of the studies was considered good for most domains of the included studies.

Conclusions

Overall diagnostic performance of TVS and MRI for detecting DIE involving recto-sigmoid, uterosacral ligaments and rectovaginal septum is similar.

INTRODUCTION

Endometriosis is a pathological condition where the endometrium is localized outside the uterine cavity[1,2] . This condition mainly affects young and childbearing women with a prevalence of 4.7%[3-5] . The medical impact of endometriosis is critical because it could determine several problems and in particular pain and sub-fertility[6-8] .

In the last years it was been well demonstrated that it is possible to identify a sub-class of endometriosis that is known as deep-infiltrating endometriosis (DIE), defined arbitrarily as endometriosis infiltrating the peritoneum by >5 mm [9] and characterized by nodules infiltrating recto-sigmoid, utero-sacral ligaments (USL), vaginal fornix, recto-vaginal septum, and bladder[10-13]. Several imaging techniques have been suggested for the detection of the DIE: ultrasound (US), Computed Tomography (CT), Magnetic Resonance Imaging (MRI) but currently the most used modalities are US (with transvaginal approach: TVUS) and MRI[14-17] . Although transvaginal US is considered as first line technique[18] , there is not a clear evidence of sequential use of imaging methods. The results of meta-analyses are contradictory for several locations ranging from 88%[19] to 58%[20] regarding sensitivity of MR for rectovaginal septum location or from 83%[21] to 92%[20] regarding sensitivity of US for rectosigmoid location. The effect of this disagreement should reduced considering only studies in the same set of patients.

The primary objectives of the present meta-analysis are two: to determine diagnostic performance of magnetic resonance imaging (MRI) and transvaginal ultrasound (TVS), for evaluating the presence of deep infiltrating endometriosis (DIE) in studies evaluating the same set of patients and to compare the diagnostic performance of both techniques.

MATERIALS AND METHODS

Protocol and registration

This systematic review and meta-analysis was performed according to the PRISMA statement (http://www.prisma-statement.org/) and the Synthesizing Evidence from Diagnostic Accuracy Tests (SEDATE) guidelines[22] . All methods for inclusion/exclusion criteria, data extraction and quality assessment were specified in advance. The protocol was registered in PROSPERO (acceptance pending).

Data sources and searches

Studies published between 1989 and October 2016 were identified by one of the authors (S.G.) using two electronic databases (PubMed/MEDLINE and Web of Science), to identify potentially eligible studies. We did not use methodological filters in database searches to avoid possible omission of relevant studies, according to the recommendations of *Leeflang et al*[23,24] . The search terms included and captured the concepts of 'endometriosis', 'ultrasonography', 'sonography', and 'magnetic resonance imaging' (Table S1). There were no language restrictions in the search. Additionally we review the reference list of selected papers.

Study selection and data collection

One author (S.G.) screened the titles and abstracts identified by the searches to exclude obviously irrelevant article, i.e. those not strictly related to the topic under review. Full-text articles were obtained to identify potentially eligible studies, and three reviewers (S.G., S.A. and J.L.A.) applied independently the following inclusion criteria (Table S2):

1) Prospective or retrospective cohort study including patients who underwent both techniques, magnetic resonance imaging (MRI) and transvaginal ultrasound (TVS), for evaluating the presence of deep infiltrating endometriosis (DIE) as index tests. To increase the consistency of the present meta-analysis only head-to-head studies were included with the aim to improve the internal validity reducing the bias due to patients selection and different

included lesions . We selected retrospective studies when data were collected prospectively during real time examination not when data were retrieved by authors by reviewing charts.

2) Series with at least 25 patients.
3) Participants aged over 18 years with clinical suspicion of DIE based on clinical complaints and/or physical examination.
4) Presurgical detection of DIE affecting the recto-sigmoid, vaginal, uterosacral ligaments, rectovaginal septum and bladder.or uterosacral ligaments as reference standard either by histological analysis or laparoscopic findings.
5) Presence of results sufficient to construct the 2×2 table of diagnostic performance as minimum data requirement.

In case of less than four studies were found for a specific location meta-analysis was not done for this locations. To avoid inclusion of duplicate cohorts in the meta-analyses in the case of two studies from the same authors, the study period of each study was examined; if dates overlapped, we chose the latest study according to the publication date, considering that patients from the first study were also included in the latest one. We excluded these types of studies: narrative or systematic reviews, retrospective design studies in which the index test was performed after execution of the reference test and/or participants were selected from a retrospective review of case notes, case reports or case series, and conference proceedings.

In case of missing data we try to contact the authors. The PICOS (Patients, Intervention, Comparator, Outcomes, Study design) criteria used for inclusion and exclusion of studies are shown in Table 1.

Diagnostic accuracy results and additional useful information on patients and procedures were retrieved from selected primary studies independently by the same authors (S.G., S.A. and J.L.A.). Disagreements arising during the process of study selection and data collection were resolved by consensus among three of the authors (S.G, S.A and J.L.A).

Risk of bias in individual studies

Quality assessment was conducted, adapting to this particular review the tool provided by the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) [25] . The QUADAS-2 format includes four domains: (1) patient selection, (2) index test, (3) reference standard, (4) flow and timing. For each domain, the risk of bias and concerns about applicability (the latter not applying to the domain of flow and timing) were analyzed and rated as low, high or unclear risk (Table S3). The results of quality assessment were used for descriptive purposes to provide an evaluation of the overall quality of the included studies and to investigate potential sources of heterogeneity. Three authors (S.G., S.A. and J.L.A.) evaluated independently the methodological quality, using a standard form with quality assessment criteria and a flow diagram; they resolved disagreements by discussion among three of the authors (S.G, S.A. and J.L.A.).

Statistical analysis

We extracted or derived information on diagnostic performance of TVS and MRI. A random-effects model was used to determine overall pooled sensitivity, specificity, positive likelihood ratio (LR+) and negative likelihood ratio (LR–) and the diagnostic odds ratio (DOR). Positive and negative likelihood ratios (LRs) were used to characterize the clinical utility of a test and to estimate the post-test probability of disease. A LR of 0.2–5.0 provides weak evidence for either ruling out or confirming the disease. A LR of 5.0–10.0 and 0.1–0.2 provides moderate evidence to either confirm or rule out the disease. A LR > 10 or < 0.1 provides strong evidence to either confirm or rule out the disease[21] . Using the mean prevalence for each location (pretest probability) in each subset, depending upon the method and LRs, post-test probabilities were calculated and plotted on Fagan nomograms.

We assessed the presence of heterogeneity for sensitivity and specificity using Cochran's Q statistic and the I^2 index[27] . A P-value < 0.1 indicates heterogeneity. The I^2 index describes the percentage of total variation across studies that is due to heterogeneity rather than chance. According to Higgins *et al*., I^2 values of 25%, 50%, and 75% would be considered to indicate low, moderate and high

heterogeneity, respectively[27] . Forest plots of sensitivity and specificity of all studies were plotted. Summary receiver–operating characteristics (sROC) curves were plotted to illustrate the relationship between sensitivity and specificity. Meta-regression was used if heterogeneity existed to assess covariates that could explain this heterogeneity. The covariates analyzed were sample size, prevalence, median patient age and number of observers (single/multiple), index test description and reference standard description.

Comparison of diagnostic performance between TVS and MRI for detecting DIE involvement of the recto-sigmoid, uterosacral ligaments and rectovaginal septum was done using the bivariate method[26] .

All analyses were performed using MIDAS (Meta-analytical Integration of Diagnostic Accuracy Studies) and METANDI commands in STATA version 12.0 for Windows (Stata Corporation, College Station, TX, USA). A *P*-value < 0.05 was considered as statistically significant.

RESULTS

Search results

The electronic search provided a total of 375 citations but after removal of 28 duplicate records, 343 citations remained. Of these, 336 were excluded because it was clear from the title or abstract that they were not relevant to the review (papers not assessing DIE, papers assessing TVS but not MRI, papers assessing MRI but not TVS, papers not assessing diagnostic performance, cases reports, reviews, letters to the editor). We examined the full text of the remaining 11 articles. Finally, 4 studies were discharged because they did not meet inclusion criteria and the remaining 6 studies were included in the review and meta-analysis[29-34]. No additional relevant studies were found from references cited in the papers included in the review. A flowchart summarizing literature identification and selection is given in Figure 1.

Characteristics of included studies

We found less than four studies for bladder[31] and vagina location[30,31,32] , so we didn't performed meta-analysis for those locations. We have no studies with missing data.

A total of 6 studies[29-34] reporting on 424 patients between December 2007 to November 2012 were included in the final analyses.

For detection of recto-sigmoid endometriosis, 424 patients were included in the final analyses. Among these women, 190 had DIE affecting the recto-sigmoid. Mean prevalence was 45%, ranging from 6% to 76%[29-34].

For the detection of USL endometriosis, 261 patients were included in the final analyses. The study of Vimercati et al. reported separately the findings in each USL[31]. Among these women, DIE was detected in 389 USLs. Mean prevalence was 63%, ranging from 15% to 90%[30-34].

For detection of RVS endometriosis, 365 patients were included in the final analyses. Among these women, 125 had DIE affecting the RVS. Mean prevalence was 34%, ranging from 12% to 76%[29-34]. Mean prevalence was considered as the pretest probability (Table 1).

All studies reported the clinical characteristics of the cohort to some extent. Mean patient age was reported in all studies and ranged from 32 to 35 years. Symptoms were reported in all studies, with dysmenorrhea being the most frequent.

Methodological quality of included studies

A graphical display of the evaluation of the risk of bias and concerns regarding applicability of the selected studies is shown in Figure 2. Regarding risk of bias and the domain patient selection, two studies did not report explicitly or were not clear regarding patient inclusion criteria[31-32]. Concerning the domain index test, 4 of the 6 studies adequately described the method of index text as well as how it was performed and interpreted[29-30, 33-34]. Concerning the domain flow and timing, the time elapsed between the index test and reference standard was unclear in three studies [30, 32, 33]. For the domain reference standard, all studies were likely to correctly classify the target condition by the reference standard. However, in most studies it was not specified if the results of the reference standard were

interpreted without knowledge of the results of the index test. In four studies the reference standard was laparoscopic findings instead of histological data[29-32].

Regarding applicability, for the domain patient selection, all but one study[30] were deemed to include patients that matched the review question. For the domain index test, most studies were considered as having low concerns for applicability as the index test was described well enough for study replication, as was the reference standard domain.

Diagnostic performance of TVS and MRI for detection of DIE involving the recto-sigmoid

Overall, pooled sensitivity, specificity, LR+ and LR– of MRI in detecting DIE in the recto-sigmoid for MRI were 0.85 (95% CI, 0.78–0.90), 95 (95% CI, 0.83–0.99), 18.4 (95% CI, 4.7–72.4) and 0.16 (95% CI, 0.11–0.24), respectively. Moderate heterogeneity was found for sensitivity (I^2, 46.8%; Cochran Q, 9.4; $P = 0.09$). Significant heterogeneity was found for specificity (I^2, 85.0%; Cochran Q, 33.3; $P < 0.001$)

Overall, pooled sensitivity, specificity, LR+ and LR– of TVS in detecting DIE in the recto-sigmoid for TVS were 0.85% (95% CI, 0.68–0.94), 0.96% (95% CI, 0.85–0.99), 20.4 (95% CI, 4.7–88.5) and 0.16 (95% CI, 0.07–0.38), respectively. DOR was 116 (95% CI 23 - 585) and 127 (95% CI 14 - 1126), respectively.

Significant heterogeneity was found for sensitivity (I^2, 82.7%; Cochran Q, 28.9; $P < 0.001$) and specificity (I^2, 78.3%; Cochran Q, 23.1; $P < 0.001$)

No statistical differences were found when comparing both methods (p= 0.845)

Figure 3 shows forest plots for both methods. ROC curves are shown in Figure 4.

Fagan nomograms show that a positive test for TVS and MRI increases significantly the pretest probability of DIE involving the rectosigmoid, from 45% to 95%, while a negative test decreases significantly the pretest probability, from 45% to 12% (Figure S1).

Meta-regression showed that sample size, prevalence, median patient age and number of observers (single/multiple), index test description and reference standard description did not explain heterogeneity.

Diagnostic performance of TVS and MRI for detection of DIE involving the rectovaginal septum

Overall, pooled sensitivity, specificity, LR+ and LR– of MRI in detecting DIE in the rectovaginal septum for MRI were 0.66% (95% CI, 0.51–0.79), 97 (95% CI, 0.89–0.99), 22.5 (95% CI, 6.7–76.2) and 0.38 (95% CI, 0.23–0.52), respectively. Significant heterogeneity was found for sensitivity (I^2, 63.7%; Cochran Q, 11.0; P = 0.03) and specificity (I^2, 79.7%; Cochran Q, 19.7; P < 0.001)

Overall, pooled sensitivity, specificity, LR+ and LR– of TVS in detecting DIE in the rectovaginal septum for TVS were 0.59% (95% CI, 0.26–0.86), 0.97 (95% CI, 0.94–0.99), 23.5 (95% CI, 9.1–60.5) and 0.42 (95% CI, 0.18–0.97), respectively. DOR was 65 (95% CI 21- 204) and 56 (95% CI 11 - 275), respectively.

Significant heterogeneity was found for sensitivity (I^2, 88.2%; Cochran Q, 34.5; $P < 0.001$) and specificity (I^2, 67.2%; Cochran Q, 12.2; $P = 0.02$)

No statistical differences were found when comparing both methods (p= 0.855)

Figure 5 shows forest plots for both methods.

ROC curves are shown in Figure 6.

Fagan nomograms show that a positive test for MRI increases significantly the pretest probability of DIE involving the rectovaginal septum, from 34% to 92%, while a negative test decreases significantly the pretest probability, from 34% to 15%. Whereas, a positive test for TVS increases significantly the pretest probability of DIE involving the rectovaginal septum, from 34% to 92%, while a negative test decreases significantly the pretest probability, from 34% to 18% (Figure S2).

Meta-regression showed that sample size, prevalence, median patient age and number of observers (single/multiple), index test description and reference standard description did not explain

heterogeneity.

Diagnostic performance of TVS and MRI for detection of DIE involving the uterosacral ligaments

Overall, pooled sensitivity, specificity, LR+ and LR– of MRI in detecting DIE in the uterosacral ligaments for MRI were 0.70 (95% CI, 0.55–0.82), 0.93 (95% CI, 0.87–0.97), 10.4 (95% CI, 5.1–21.2) and 0.32 (95% CI, 0.20–0.51), respectively. Significant heterogeneity was found for sensitivity (I^2, 83.7%; Cochran Q, 17.7; P = 0.01) and low for specificity (I^2, 14.2%; Cochran Q, 3.5; P = 0.32) .

Overall, pooled sensitivity, specificity, LR+ and LR– of TVS in detecting DIE in the uterosacral ligaments for TVS were 0.67 (95% CI, 0.55–0.77), 0.86 (95% CI, 0.73–0.93), 4.8 (95% CI, 2.6–9.0) and 0.38 (95% CI, 0.29–0.50), respectively. . DOR was 32 (95% CI 12- 85..) and 12 (95% CI 7- 24), respectively.

Moderate heterogeneity was found for sensitivity (I^2, 41.5%; Cochran Q, 5.1; $P = 0.16$) and specificity (I^2, 54.8%; Cochran Q, 6.6; $P = 0.08$) .

No statistical differences were found when comparing both methods (p= 0.250).

Figure 7 shows forest plots for both methods.

ROC curves are shown in Figure 8.

Fagan nomograms show that a positive test for TVS increases significantly the pretest probability of DIE involving the uterosacral ligaments, from 63% to 89%, while a negative test decreases significantly the pretest probability, from 63% to 40%. Whereas, a positive test for MRI increases significantly the pretest probability of DIE involving the uterosacral ligaments, from 63% to 95%, while a negative test decreases significantly the pretest probability, from 63% to 32% (Figure S3).

Meta-regression showed that sample size, prevalence, median patient age and number of observers (single/multiple), index test description and reference standard description did not explain heterogeneity.

DISCUSSION

The purpose of this study was to perform a meta-analysis that compared TVUS and MR for the DIE detection in the same set of patients in order to reach stronger evidences about the performance of these imaging techniques increasing the internal validity. This is the main strength of the present study and may explain the different results obtained in the other reviews present in the literature[19-21, 35, 36] (TABLE 2). The limitations are related to the small number of cases included using only head-to-head studies and also to the locations excluded due to the less than 4 studies for some locations, in particular vaginal and anterior compartment.

The total number of the analyzed patients in the 6 studies was 424 with examinations performed in a 5-years timeline (from December 2007 to November 2012) although the search of databases started from 1989. This is an important point because it allows to include in the present review only papers including more homogeneous technology and similar methodological approach impossible to obtain if older papers were be included. In particular, in the all 6 studies included a 1.5 Tesla scanner was used.

Unfortunately the prevalence of DIE ranged from 6% to 76% introducing a significant heterogeneity among the studies included related to the presence of difference reference standard or surgical methodology (in 4 paper the reference standard was laparoscopic findings instead of histological data[29-32]). Moreover, two studies were not fully clear regarding patient inclusion criteria[31-32]. The presence of a selection bias that we suggest may affect the diagnostic performance because in studies with low prevalence a false negative case reduce dramatically the sensitivity of the methods (for example see Saccardi et al [32]for RS involvement).

The main difference with the only meta-analysis present in the literature[19] including both techniques is the authors did not performed a direct head-to-head comparison of TVS and MRI. This is because for RS, RVS and USL sites they only included 2, 2 and 1 studies, respectively. So, they could not make such a comparison. We have included 6, 5 and 4 studies, respectively. This allow us

to perform a direct comparison, although certainly with a small number of cases.

Regarding in detail the diagnostic performance of TVS and MRI for detection of DIE involving the recto-sigmoid the overall, pooled sensitivity, specificity, LR+ and LR– were similar with 85% of sensitivity for MRI and TVS and with a specificity of 95% and 96% respectively. Performing a comparison with other published meta-analyses[19-21, 35-36] in the present review we observed lower sensitivities in comparison with Nisenblat et al[19] for both methods (90% for US and 92% for MR). On the contrary our overall sensitivities are similar to Noventa et al[20] for TVS and similar to Medeiros et al for MR[21] (Table 2). The level of heterogeneity of the sensitivity that was significant for TVS and only moderate for MRI. Meta-regression showed that sample size, prevalence, median patient age and number of observers (single/multiple), index test description and reference standard description did not explain heterogeneity. These data confirm that the identification of endometriosis involving the recto-sigmoid is very good with both the methodologies.

Also diagnostic performance of TVS and MRI for detection of DIE involving RVS location was similar. The comparison with other meta-analyses showed lower sensitivity in comparison with Nisenblat et al[19] for both methods (88% for US and 812% for MR). On the contrary our overall sensitivies are similar to Noventa et al[20] for TVS and lower to Medeiros et al[21] for MR (Table 2). For RVS for both sensitivity and specificity, the heterogeneity was significant for both techniques. Also in this case, meta-regression did not explain heterogeneity.

The diagnostic performance of TVS and MRI for detection of DIE involving the USL was similar. The comparison with other meta-analyses showed similar sensitivity in comparison with Nisenblat et al[19] for US but higher in comparison with Noventa et al[20] (49.7%). On the contrary our overall sensitivies are lower to Nisenblat et al[19] and Medeiros et al[21] for MR (Table 2). For RVS a moderate heterogeneity was found for sensitivity and specificity using TVS whereas the MRI showed significant heterogeneity for sensitivity (I^2, 83.7%). Again, also in this case, meta-regression did not explain heterogeneity.

In the future studies with better quality design are suggested using a standardized approach, as proposed by IDEA consensus[37] to decrease the heterogeneity reported in the present review. However the results of this meta-analysis demonstrate that diagnostic performance are similar using both imaging methods confirming the role of first line technique of transvaginal ultrasonography as main cost-effective clinical recommendation.

REFERENCES

1) Bulun SE. Endometriosis. N Engl J Med. 2009 15;360(3):268-79.

2) Bondza PK, Maheux R, Akoum A. Insights into endometriosis-associated endometrial dysfunctions: a review. Front Biosci (Elite Ed). 2009 Jun 1;1:415-28.

3) Mehedintu C, Plotogea MN, Ionescu S, Antonovici M. Endometriosis still a challenge. J Med Life. 2014 15;7:349-57.

4) Saba L, Sulcis R, Melis GB, de Cecco CN, Laghi A, Piga M, Guerriero S. Endometriosis: the role of magnetic resonance imaging. Acta Radiol. 2015;56:355-67.

5) Guo SW. An overview of the current status of clinical trials on endometriosis: issues and concerns. Fertil Steril. 2014;101:183-190.e4.

6) Singh SS, Suen MW. Surgery for endometriosis: beyond medical therapies. Fertil Steril. 2017;107:549-554.

7) Speer LM, Mushkbar S, Erbele T. Chronic Pelvic Pain in Women. Am Fam Physician. 2016 1;93:380-7.

8) Haydardedeoglu B, Zeyneloglu HB. The impact of endometriosis on fertility. Womens Health (Lond). 2015;11:619-23.

9) Koninckx PR and Martin DC Deep endometriosis: a consequence of infiltration or retraction or possibly adenomyosis externa?. *Fertil Steril*. 1992; 58: 924–928

10) Ferrero S, Alessandri F, Racca A, Leone Roberti Maggiore U. Treatment of pain associated with deep endometriosis: alternatives and evidence. Fertil Steril. 2015;104:771-92.

11) Gui B, Valentini AL, Ninivaggi V, Marino M, Iacobucci M, Bonomo L. Deep pelvic endometriosis: don't forget round ligaments. Review of anatomy, clinical characteristics, and MR imaging features. Abdom Imaging. 2014;39:622-32.

12) Schindler AE. Pathophysiology, diagnosis and treatment of endometriosis. Minerva Ginecol. 2004;56:419-35.

13) Roman H, Darwish B, Bridoux V, Huet E, Coget J, Chati R, Tuech JJ, Abo C. Multiple nodule removal in multifocal colorectal endometriosis instead of "en bloc" large colorectal resection. Gynecol Obstet Fertil. 2016;44:121-4.

14) Saba L, Guerriero S, Sulis R, Pilloni M, Ajossa S, Melis G, Mallarini G. Learning curve in the detection of ovarian and deep endometriosis by using Magnetic Resonance: comparison with surgical results. Eur J Radiol. 2011;79:237-44.

15) Saba L, Sulcis R, Melis GB, Ibba G, Alcazar JL, Piga M, Guerriero S. Diagnostic confidence analysis in the magnetic resonance imaging of ovarian and deep endometriosis: comparison with surgical results. Eur Radiol. 2014;24:335-43.

16) Guerriero S, Saba L, Ajossa S, Peddes C, Angiolucci M, Perniciano M, Melis GB, Alcázar JL. Three-dimensional ultrasonography in the diagnosis of deep endometriosis. Hum Reprod. 2014 ;29:1189-98.

17) Guerriero S, Spiga S, Ajossa S, Peddes C, Perniciano M, Soggiu B, De Cecco CN, Laghi A, Melis GB, Saba L. Role of imaging in the management of endometriosis. Minerva Ginecol. 2013;65:143-66.

18) Piketty 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.

19) Nisenblat V, Bossuyt PM, Farquhar C, Johnson N, Hull ML. Imaging modalities for the non-invasive diagnosis of endometriosis. Cochrane Database Syst Rev. 2016 ;2:CD009591.

20) Noventa M, Saccardi C, Litta P, Vitagliano A, D'Antona D, Abdulrahim B, Duncan A, Alexander-Sefre F, Aldrich CJ, Quaranta M, Gizzo S. Ultrasound techniques in the diagnosis of deep pelvic endometriosis: algorithm based on a systematic review and meta-analysis. Fertil Steril. 2015 ;104:366-83.

21) Medeiros LR, Rosa MI, Silva BR, Reis ME, Simon CS, Dondossola ER, da Cunha Filho JS. Accuracy of magnetic resonance in deeply infiltrating endometriosis: a systematic review and meta-analysis. Arch Gynecol Obstet. 2015 ;291:611-21.

22) Sotiriadis A, Papatheodorou SI, Martins WP. Synthesizing Evidence from Diagnostic Accuracy TEsts: the SEDATE guideline. Ultrasound Obstet Gynecol. 2016 ;47:386-95

23) Leeflang MM. Reporting diagnostic accuracy studies: where are we now? Biomark Med. 2015;9:897-9.

24) Leeflang MM. Systematic reviews and meta-analyses of diagnostic test accuracy. Clin Microbiol Infect. 2014;20:105-13.

25) Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155: 529–536.

26) EUnetHTA Guideline. Meta-analysis of diagnostic test accuracy studies. Accessed January 2016 (http://www.eunethta.eu/eunethta-guidelines.

27) Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta analyses. *BMJ*2003; 327: 557–560.

28) Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005; 58: 882–893.

29) Abrao MS, Gonçalves MO, Dias JA Jr, 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 ;22:3092-7.

30) Bazot M, Lafont C, Rouzier R, Roseau G, Thomassin-Naggara I, Daraï E. Diagnostic accuracy of physical examination, transvaginal sonography, rectal endoscopic sonography, and magnetic resonance imaging to diagnose deep infiltrating endometriosis.
Fertil Steril. 2009;92:1825-33.

31) Vimercati A, Achilarre MT, Scardapane A, Lorusso F, Ceci O, Mangiatordi G, Angelelli G, Van Herendael B, Selvaggi L, Bettocchi S. Accuracy of transvaginal sonography and contrast-enhanced magnetic resonance-colonography for the presurgical staging of deep infiltrating endometriosis. Ultrasound Obstet Gynecol. 2012 ;40:592-603.

32) Saccardi C, Cosmi E, Borghero A, Tregnaghi A, Dessole S, Litta P. Comparison between transvaginal sonography, saline contrast sonovaginography and magnetic resonance imaging in the diagnosis of posterior deep infiltrating endometriosis. Ultrasound Obstet Gynecol. 2012 ;40:464-9.

33) Saba L, Guerriero S, Sulcis R, Pilloni M, Ajossa S, Melis G, Mallarini G. MRI and "tenderness guided" transvaginal ultrasonography in the diagnosis of recto-sigmoid endometriosis. J Magn Reson Imaging. 2012 ;35:352-60.

34) Gauche Cazalis C, Koskas M, Martin B, Palazzo L, Madelenat P, Yazbeck C. Preoperative imaging of deeply infiltrating endometriosis in: Transvaginal sonography, rectal endoscopic sonography and magnetic resonance imaging.
Gynecol Obstet Fertil. 2012 ;40:634-41.

35) Guerriero S, Ajossa S, Minguez JA, Jurado M, Mais V, Melis GB, Alcazar JL. Accuracy of transvaginal ultrasound for diagnosis of deep endometriosis in uterosacral ligaments, rectovaginal septum, vagina and bladder: systematic review and meta-analysis.
Ultrasound Obstet Gynecol. 2015 ;46:534-45

36) 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. 2016;47:281-9.

37) Guerriero S, Condous G, van den Bosch T, Valentin L, Leone FP, Van Schoubroeck D, Exacoustos C, Installé AJ, Martins WP, Abrao MS, Hudelist G, Bazot M, Alcazar JL, Gonçalves MO, Pascual MA, Ajossa S, Savelli L, Dunham R, Reid S, Menakaya U, Bourne T, Ferrero S, Leon M, Bignardi T, Holland T, Jurkovic D, Benacerraf B, Osuga Y, Somigliana E, Timmerman D. 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. 2016 ;48:318-32.

Legends of Figures

Figure 1.Flowchart showing literature identification and selection.
Figure 2. Quality evaluation of all 11 studies included in the meta-analysis, according to QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies-2) criteria[25], with respect to risk of bias (a) and concerns regarding applicability (b). , low; □, high; ▩, unclear.
Figure 3. Forest plots of studies evaluated for detection of deep infiltrating endometriosis involving recto-sigmoid using Magnetic Resonance (a) and transvaginal ultrasound (b). Summary sensitivity and specificity as well as heterogeneity statistics (Cochran's Q and I^2) are shown.
Figure 4. Summary receiver–operating characteristics (sROC) curves (■) for detection of deep infiltrating endometriosis involving rectosigmoid using Magnetic Resonance (a) and transvaginal ultrasound (b).
Figure 5. Forest plots of studies evaluated for detection of deep infiltrating endometriosis involving rectovaginal septum using Magnetic Resonance (a) and transvaginal ultrasound (b). Summary sensitivity and specificity as well as heterogeneity statistics (Cochran's Q and I^2) are shown.
Figure 6 Summary receiver–operating characteristics (sROC) curves (■) for detection of deep infiltrating endometriosis rectovaginal septum using Magnetic Resonance (a) and transvaginal ultrasound (b).
Figure 7. Forest plots of studies evaluated for detection of deep infiltrating endometriosis involving uterosacral ligaments using Magnetic Resonance (a) and transvaginal ultrasound (b). Summary sensitivity and specificity as well as heterogeneity statistics (Cochran's Q and I^2) are shown.
Figure 8. Summary receiver–operating characteristics (sROC) curves (■) for detection of deep infiltrating endometriosis involving the uterosacral ligaments using Magnetic Resonance (a) and transvaginal ultrasound (b).

Supplemental material

Table S1 The full electronic search strategy.
Table S2 The data collection form.
Table S3 The questions and the criteria for low/high risk assessment.

Figure S1. Fagan nomograms for detection of deep infiltrating endometriosis involving the rectosigmoid using Magnetic Resonance (a) and transvaginal ultrasound (b). Pretest probability (◆) and effect of a positive test result (⊣) and a negative test result (⊣) are indicated.
Figure S2. Fagan nomograms for detection of deep infiltrating endometriosis involving the rectovaginal septum using Magnetic Resonance (a) and transvaginal ultrasound (b). Pretest probability (◆) and effect of a positive test result (⊣) and a negative test result (⊣) are indicated.
Figure S3. Fagan nomograms for detection of deep infiltrating endometriosis involving the uterosacral ligaments using Magnetic Resonance (a) and transvaginal ultrasound (b). Pretest probability (◆) and effect of a positive test result (⊣) and a negative test result (⊣) are indicated.

Table 1. Characteristics of studies included according to PICOS (Patients, Intervention, Comparator, Outcomes, Study design) criteria

Reference	Geographical area	Setting	*n*	Study design	TVS technique	MRI technique	Observers	Reference standard	Cases with DIE (*n*)
Abrao (2007)	South America	Single center	104	Prospective	Non-enhanced	1.5 T + contrast	Single	Surgery and histopathology	RS (54), RVS (41)
Bazot (2009)	Europe	Single center	92	Prospective	Non-enhanced	1.5 T+ contrast	Single	Surgery and histopathology following Bazot's criteria	RS (63), USL (83), RVS (11)
Vimercati (2012)	Europe	Single center	90	Prospective	Non-enhanced	1.5 T+ contrast	Single	Surgery and histopathology	RS (18), USL (112)*, RVS (18),
Saccardi (2012)	Europe	Single center	54	Prospective	Enhanced (saline contrast SVG)	1.5 T+ contrast	Single	Surgery and histopathology	RS (6), USL (9), RVS (36)
Saba (2012)	Europe	Single center	59	Prospective	Non-enhanced	1.5 T+ contrast	Single	Surgery and histopathology following Bazot's criteria	RS (30)
Gauche (2012)	Europe	Single center	25	Retrospective		1.5 T+ contrast	Single	Surgery and histopathology	RS (19), USL (18), RVS (19)

Table 2 Comparison of different published meta-analyses: A Rectosigmoid; B Rectovaginal septum; C, Uterosacral ligaments;

A. Rectosigmoid

Authors	Test methods	Pooled Sensitivity	Pooled Specificity
Nisenblat et al [19]	TVS	90 (95% CI 0.82 -0.97)	96 (95%CI 0.94- 0.99)
Guerriero et al[36]	TVS	91 (95%CI, 85-94%)	97 (95%CI, 95-98%)
Present meta-analysis	TVS	85 (95% CI, 68–94%)	96% (95% CI, 85–99%)
Noventa et al [20]	TVS	85.2% (95% CI, 82.9%–87.2%)	88.9% (95% CI, 86.5%–90.9%)
Medeiros et al[21]	MR	83 (95 % CI 0.78–0.87)	(95 % CI 0.84–0.92)
Nisenblat et al[39]	MR	92 (95% CI 0.86 to 0.99)	96 (95% CI 0.93 to 0.98)
Present meta-analysis	MR	85 (95% CI, 78–90%)	95 (95% CI, 83–99%)

B. RVS

Author	Test methods	Pooled Sensitivity	Pooled Specificity
Nisenblat et al [19]	TVS	88 (95%CI 0.82 -0.94)	100 (95%CI 0.98 - 1.00)
Guerriero et al [35] 2015	TVS	49 (95%CI, 36-62%)	98 (95%CI, 95-99%)
Present meta-analysis	TVS	59 (95% CI, 26–86%)	97 (95% CI, 94–99%)
Noventa et al[20]	TVS	59.8% (95% CI, 55.4%–64%)	87.5% (95% CI, 83.6%–90.5%)
Medeiros et al[21]	MR	77 (95 % CI 0.69–0.83)	95 (95 % CI 0.92–0.96)
Nisenblat et al [19]	MR	81 (95% CI 0.70 - 0.93)	86 (95% CI 0.78 to 0.95)
Present meta-analysis	MR	66 (95% CI, 51–79%)	97 (95% CI, 89–99%)

C. USL

Author	Test methods	Pooled Sensitivity	Pooled Specificity
Nisenblat et al [19]	TVS	64 (95% CI 0.50-to 0.79)	97 (95% CI 0.93 to 1.00)
Guerriero et al[35] 2015	TVS	53 (95%CI, 35-70%)	93 (95%CI, 83-97%)
Present meta-analysis	TVS	67 (95% CI, 55–77%)	86 (95% CI, 73–93%)
Noventa et al [19]	TVS	49.7% (95% CI, 46.1%–53.3%)	68.9% (95% CI, 65.1%–72.5%)
Medeiros et al 21[39]	MR	85 (95 % CI 0.82–0.88)	80 (95 % CI 0.76–0.84)
Nisenblat et al [19]	MR	86 (95% CI 0.80-0.92)	84 (95% CI 0.68 to 1.00)
Present meta-analysis	MR	70 (95% CI, 55–82%)	93 (95% CI, 87–97%)