Comparison of physical examination, ultrasound techniques and magnetic resonance imaging for the diagnosis of deep infiltrating endometriosis: A systematic review and meta-analysis of diagnostic accuracy studies

XIAO ZHANG, TAO HE and WEN SHEN

Department of Ultrasound, Hanchuan People's Hospital, Hanchuan, Hubei 431600, P.R. China

Received December 10, 2019; Accepted May 15, 2020

DOI: 10.3892/etm.2020.9043

Abstract. Visual inspection via laparoscopy is considered the gold standard for the diagnosis of deep infiltrating endometriosis. Laparoscopy is an invasive procedure; therefore, it would be beneficial to patients if accurate non-invasive modalities were available for the diagnosis of deep infiltrating endometriosis. The purpose of the current review and meta-analysis was to assess the diagnostic accuracy of clinical examination, transvaginal ultrasound (TVUS), transrectal ultrasound (TRUS) and MRI as alternative methods for diagnosis of deep infiltrating endometriosis. A systematic search of the Medline, Scopus, Embase and Cochrane library databases, between their inception and September 2019, was performed. The quality of trials was assessed using the quality assessment of diagnostic accuracy studies-2 tool. Meta-analyses were conducted to obtain the pooled sensitivity, specificity, positive and negative likelihood ratios and diagnostic odds ratio for each of the three imaging modalities and clinical examination. A total of 30 studies with 4,565 participants were included in the review. Physical examination had a pooled sensitivity of 71% and a specificity of 69%, with an average diagnostic accuracy [area under the curve (AUC) =0.76]. TVUS had a pooled sensitivity of 76% and a specificity of 94%, with higher diagnostic accuracy than physical examination (AUC =0.92). TRUS had a pooled sensitivity of 91% and a specificity of 80% with an AUC of 0.93. MRI had a pooled sensitivity of 82% and a specificity of 87% with higher diagnostic accuracy than physical examination (AUC =0.91). All the imaging modalities had good clinical utility, as indicated by the Fagan plot. The present analysis demonstrates that the imaging modalities TVUS, TRUS and MRI may be highly useful alternatives to laparoscopy for diagnosis of deep infiltrating endometriosis and that these techniques have a high sensitivity and specificity.

Introduction

Endometriosis is one of the most common benign conditions affecting women. The prevalence of the disease is ~10% in females of reproductive age (15-49 years) and 20-50% in those diagnosed with infertility (1,2). Dysmenorrhea, chronic pelvic pain and infertility are commonly associated with the disorder (3). Endometriosis can be classified as either superficial or deep. Peritoneal infiltration of <5 mm is defined as superficial endometriosis. The presence of endometrial tissue, fibrosis and hyperplasia with >5 mm peritoneal infiltration is defined as deep infiltrating endometriosis (DIE) (3). DIE accounts for 15-30% of all diagnosed endometriosis cases (4,5). While there is no known cure for the disease, available treatment options include expectant management, hormonal therapy and surgical management (6). Treatment is usually personalized, considering the therapeutic goal (subfertility or pain relief) and the site of disease (6).

Pharmacological therapies for endometriosis include progestogens, the combined oral contraceptive pill, gonadotropin-releasing hormone agonists and weak androgens (7). These drugs alleviate the symptoms of chronic pelvic pain and dysmenorrhea; however, they are associated with adverse events, including irritability, breast discomfort, bone loss and androgenic symptoms (7). Surgical treatment of endometriosis can alleviate pain and improve fertility. However, 5-year recurrence rates following surgery are high (40-50%) (8,9). The potential success and impact of these strategies is dependent on early and accurate diagnosis (10). Laparoscopic observation and biopsy are considered to be the gold standard for diagnosis of DIE (9). However, laparoscopy is invasive and does not allow any time for preoperative planning. Several non-invasive imaging modalities, such as physical examination, transvaginal ultrasonography (TVUS), transrectal sonography (TRUS) and MRI, are also available for the diagnosis of DIE (11). Accuracy of such non-invasive modalities would ensure their use as standard, significantly reducing the risk associated with surgery while facilitating early diagnosis and treatment (11). A number of studies

Correspondence to: Dr Tao He, Department of Ultrasound, Hanchuan People's Hospital, 1 People's Avenue, Hanchuan, Hubei 431600, P.R. China
E-mail: 17771215810@163.com

Key words: deep infiltrating endometriosis, MRI, ultrasonography, validation
have attempted to assess the accuracy of such non-invasive diagnostic modalities for DIE in various settings with different results. However, to the best of our knowledge, there have been no systematic efforts to collate evidence comparing these different diagnostic techniques. Therefore, the purpose of the present study was to systematically search the literature and perform a meta-analysis of diagnostic data to compare the accuracy of physical examination, ultrasound techniques (TVUS & TRUS) and MRI in the diagnosis of DIE.

Materials and methods

Inclusion criteria

Studies. All studies examining the diagnostic accuracy of physical examination, ultrasound techniques (TVUS or TRUS) and MRI for diagnosis of DIE were included where they reported sensitivity and specificity values of any of the aforementioned diagnostic techniques or provided the data required to calculate these rates. Studies published as full-text manuscripts were included while abstracts and case reports were excluded.

Participants. Studies conducted on patients with suspected DIE were included irrespective of the medical co-morbidities suffered by the participants and the setting in which the study was conducted.

Index test. Studies utilizing physical examination, TVUS, TRUS or MRI to diagnose DIE were included.

Reference standards. Studies were included only if the diagnostic accuracy of physical examination, TVUS, TRUS or MRI was compared with standard laparoscopic or histopathological examination. Diagnosis using the reference standard must have been made by specialist doctors or trained researchers.

Search strategy. An extensive electronic search of the Medline (PubMed) (https://pubmed.ncbi.nlm.nih.gov/), Scopus (https://www.scopus.com/), Embase (https://www.embase.com/) and Cochrane library (https://www.cochranelibrary.com/) was conducted. A combination of medical subject heading (MeSH) and free-text terms was used to conduct the literature search. The following MeSH terms and free-text terms were used for the literature search. ‘Validation studies’, ‘deep infiltrating endometriosis’, ‘physical examination’, ‘transvaginal ultrasonography’, ‘transrectal ultrasonography’, ‘magnetic resonance imaging’, ‘gynaecological disorders’, ‘sensitivity’, ‘specificity’, ‘diagnosis’, ‘endometriosis’ and ‘diagnostic accuracy studies’. These search terms were used in various combinations using Boolean operators such as ‘AND’, ‘OR’ or ‘NOT’. Additional filters related to timeline of search in the database was between their inception and September 2019. The language of the studies was restricted to English.

Selection of studies. The literature search was performed by two investigators independently. Records were screened by their title, abstract and keywords for possible inclusion in the review. Full texts of relevant studies were extracted and screened further, based on the eligibility criteria for final inclusion in the review. The reference list of all full-text articles was searched manually to identify any missed studies. Any disagreements between the two authors during the entire selection process were resolved either through consensus or consultation with an independent third investigator.

Data extraction and management. The primary investigator extracted the relevant study characteristics for the review from all included studies. The following data were extracted: Author, year of publication, study design, index test, reference standards, comorbidities, number of participants, mean age, inclusion and exclusion criteria, true positives, true negatives, false positives and false negatives. Data was transferred to STATA software version 14.2 (StataCorp LLC) by the primary investigator. Data entry was double-checked for accuracy by the third investigator.

Risk of bias assessment. The risk of bias in all of the included studies was assessed by two investigators independently using the quality assessment of diagnostic accuracy studies-2 (QUADAS-2) tool (12). The following domains were used for the assessment of the risk of bias: Patient selection, index test, reference standard and flow and timing of assessments. Each domain was graded as high, low or unclear risk of bias. Any discrepancies were resolved by consensus or consultation with the third investigator.

Statistical analysis. Meta-analysis was performed using STATA software version 14.2 (StataCorp LLC) to obtain the pooled estimate and 95% confidence intervals (CI) of sensitivity, specificity, positive likelihood ratio (LRP), negative likelihood ratio (LRN) and summary diagnostic odds ratio (DOR) for each of the diagnostic tests. Summary receiver operator characteristic (sROC) curves were constructed and the area under the curve (AUC) was obtained. Each data point in the sROC curve represents a separate study. Test performance accuracy was classified as follows: AUC 0.5±0.7, low; AUC 0.7±0.9, medium; and AUC >0.9, high (13).

Forest plots were used to graphically represent the study-specific and pooled estimates. A Fagan plot was constructed to estimate how much the result of a diagnostic test changes the probability that a patient has DIE. Between-study heterogeneity was assessed using bivariate box plots, χ² tests and F statistics. F values of <25% indicated mild heterogeneity, 25-75% indicated moderate heterogeneity and >75% indicated substantial heterogeneity (14). Publication bias was assessed graphically via funnel plots and the asymmetry of the plot was evaluated using Deek's test.

Results

Selection of studies. In total, 1,210 citations were found, of which 445 were retrieved from Medline, 411 from Scopus, 302 from Embase and 52 from the Cochrane library. After the first stage of screening (title, abstract and keywords), 98 relevant studies were retrieved. The full text of these studies was reviewed based on the eligibility criteria. Bibliographies of the retrieved articles were reviewed, and 7 additional studies were identified. In the final analysis 30 studies with 4,565 participants satisfying the inclusion criteria were included (Fig. 1) (15-44).

Characteristics of included studies. Characteristics of the studies are described in Table I. With the exception of
Chapron et al (19), all included studies were prospective studies. All the studies were conducted in high-income countries as per World Bank income group country classification (45). The majority of the studies (50%) were conducted in Italy followed by the United Kingdom, France, the United States of America and Austria. The mean age of the participants ranged from 28-36 years. In total, 4,565 participants were assessed in the included studies, with 469 participants undergoing physical examination, 3,847 undergoing TVUS, 728 undergoing TRUS and 1,298 undergoing MRI. The total sample size of studies varied from 31-1,440 participants. Among the 30 studies included, 5 studies reported on the diagnostic accuracy of physical examination, 21 on TVUS, 6 on TRUS and 13 on MRI.

Methodological quality of the included studies. The risk of bias assessment of the included studies is presented in Table II and Fig. S1. A total of 13 studies had a high risk of bias for patient selection, while 5 studies had a high risk of bias for the conduct and interpretation of the index test domain. A total of 12 studies had a high risk of bias due to the conduct and interpretation of reference standards and 12 studies had a high risk of bias for the flow of patients and the time interval between index test and reference standard.

Diagnostic performance of physical examination. In total, five studies assessed the diagnostic accuracy of physical examination for DIE. The pooled sensitivity and specificity of physical examination for the diagnosis of DIE were 71% (95% CI, 60-80%) and 69% (95% CI, 54-82%), respectively. The DOR was 5 (95% CI, 3-12) as indicated by the Forest plot (Fig. 2A). LRP was 2.3 (95% CI, 1.5-3.6) and LRN was 0.42 (95% CI, 0.29-0.61). Fig. 3A depicts the sROC curve for
Table 1. Characteristics of the included studies (n=30).

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Country</th>
<th>Study design</th>
<th>Population size</th>
<th>Type of diagnostic modality</th>
<th>Gold standard comparator</th>
<th>Mean patient age, years</th>
<th>Time interval between index test and standards reference</th>
<th>(Refs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abrao et al, 2007</td>
<td>Brazil</td>
<td>Prospective cross-sectional study</td>
<td>104</td>
<td>Clinical examination, TVUS, MRI</td>
<td>Laparoscopy + histopathology</td>
<td>33.8</td>
<td>3 months</td>
<td>(15)</td>
</tr>
<tr>
<td>Alborzi et al, 2018</td>
<td>Iran</td>
<td>Prospective longitudinal study</td>
<td>317</td>
<td>TVUS, TRUS, MRI</td>
<td>Laparoscopy + histopathology</td>
<td>31.0</td>
<td>Not specified</td>
<td>(16)</td>
</tr>
<tr>
<td>Bazot et al, 2009</td>
<td>France</td>
<td>Longitudinal</td>
<td>92</td>
<td>Clinical examination, TVUS, TRUS, MRI, MRI</td>
<td>Laparoscopy, laparotomy + histopathology</td>
<td>31.8</td>
<td>&lt;12 months</td>
<td>(17)</td>
</tr>
<tr>
<td>Bergamini et al, 2010</td>
<td>Italy</td>
<td>Prospective</td>
<td>61</td>
<td>TRUS</td>
<td>Laparoscopy, laparotomy + histopathology</td>
<td>33.1</td>
<td>Not specified</td>
<td>(18)</td>
</tr>
<tr>
<td>Chapron et al, 2004</td>
<td>France</td>
<td>Retrospective</td>
<td>81</td>
<td>TRUS, MRI</td>
<td>Laparoscopy or laparotomy</td>
<td>31.9</td>
<td>Not specified</td>
<td>(19)</td>
</tr>
<tr>
<td>Charmie et al, 2009</td>
<td>Brazil</td>
<td>Prospective</td>
<td>92</td>
<td>MRI</td>
<td>Laparoscopy + histopathology</td>
<td>33.0</td>
<td>Not specified</td>
<td>(20)</td>
</tr>
<tr>
<td>Dessole et al, 2003</td>
<td>Italy</td>
<td>Prospective</td>
<td>46</td>
<td>TVUS</td>
<td>Laparoscopy, laparotomy + histopathology</td>
<td>30.3</td>
<td>Not specified</td>
<td>(21)</td>
</tr>
<tr>
<td>Eskenazi et al, 2001</td>
<td>Italy</td>
<td>Prospective (study sample)</td>
<td>90</td>
<td>Clinical examination, TVUS</td>
<td>Laparoscopy, laparotomy + histopathology</td>
<td>35.7</td>
<td>34 days</td>
<td>(22)</td>
</tr>
<tr>
<td>Fedele et al, 1998</td>
<td>Italy</td>
<td>Prospective Retrospective (test sample)</td>
<td>140</td>
<td>TRUS</td>
<td>Laparoscopy, laparotomy + histopathology</td>
<td>30.2</td>
<td>1 week</td>
<td>(23)</td>
</tr>
<tr>
<td>Grasso et al, 2009</td>
<td>Italy</td>
<td>Prospective</td>
<td>33</td>
<td>TVUS, MRI</td>
<td>Laparoscopy + histopathology</td>
<td>35.0</td>
<td>1-4 weeks</td>
<td>(24)</td>
</tr>
<tr>
<td>Guerriero et al, 2007</td>
<td>Italy</td>
<td>Prospective</td>
<td>50</td>
<td>TVUS</td>
<td>Laparoscopy + histopathology</td>
<td>33.0</td>
<td>Within 7 days</td>
<td>(25)</td>
</tr>
<tr>
<td>Guerriero et al, 2008</td>
<td>Italy</td>
<td>Prospective</td>
<td>88</td>
<td>TVUS</td>
<td>Laparoscopy + histopathology</td>
<td>33.0</td>
<td>Within 7 days</td>
<td>(26)</td>
</tr>
<tr>
<td>Guerriero et al, 2014</td>
<td>Italy</td>
<td>Prospective</td>
<td>202</td>
<td>TVUS</td>
<td>Laparoscopy + histopathology</td>
<td>34.0</td>
<td>Within 1 month</td>
<td>(27)</td>
</tr>
<tr>
<td>Guerriero et al, 2017</td>
<td>Italy</td>
<td>Prospective</td>
<td>159</td>
<td>TVUS, MRI</td>
<td>Laparoscopy + histopathology</td>
<td>33.0</td>
<td>Surgery was performed over 5 years of analysis</td>
<td>(28)</td>
</tr>
<tr>
<td>Ha et al, 1994</td>
<td>Korea</td>
<td>Prospective</td>
<td>186</td>
<td>MRI</td>
<td>Laparoscopy + histopathology</td>
<td>35.0</td>
<td>Within 2 weeks</td>
<td>(29)</td>
</tr>
<tr>
<td>Holland et al, 2010</td>
<td>United Kingdom</td>
<td>Prospective</td>
<td>201</td>
<td>TVUS</td>
<td>Laparoscopy + histopathology</td>
<td>34.9</td>
<td>Mean duration =37.5 days</td>
<td>(30)</td>
</tr>
<tr>
<td>Holland et al, 2013</td>
<td>United Kingdom</td>
<td>Prospective</td>
<td>198</td>
<td>TVUS</td>
<td>Laparoscopy + histopathology</td>
<td>35.0</td>
<td>Mean duration =36.8 days</td>
<td>(31)</td>
</tr>
<tr>
<td>Hudelist et al, 2011</td>
<td>United Kingdom</td>
<td>Prospective</td>
<td>129</td>
<td>Clinical examination, TVUS</td>
<td>Laparoscopy + histopathology</td>
<td>32.2</td>
<td>Within 3 months</td>
<td>(32)</td>
</tr>
</tbody>
</table>
Table I. Continued.

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Country</th>
<th>Study design</th>
<th>Population size</th>
<th>Type of diagnostic modality</th>
<th>Gold standard comparator</th>
<th>Mean patient age, years</th>
<th>Time interval between index test and standards reference</th>
<th>(Refs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hudelist et al, 2013</td>
<td>Austria</td>
<td>Prospective</td>
<td>117</td>
<td>TVUS</td>
<td>Laparoscopy + histopathology</td>
<td>31.6</td>
<td>Within 2 months (33)</td>
<td></td>
</tr>
<tr>
<td>Manganaro et al, 2013</td>
<td>Italy</td>
<td>Prospective</td>
<td>42</td>
<td>MRI</td>
<td>Laparoscopy + histopathology</td>
<td>28.0</td>
<td>Within 3 months (34)</td>
<td></td>
</tr>
<tr>
<td>Piessens et al, 2014</td>
<td>Australia</td>
<td>Prospective</td>
<td>85</td>
<td>TVUS</td>
<td>Laparoscopy + histopathology</td>
<td>Age range given (18-48)</td>
<td>Within 12 months (35)</td>
<td></td>
</tr>
<tr>
<td>Ribeiro et al, 2008</td>
<td>United Kingdom</td>
<td>Prospective</td>
<td>37</td>
<td>TRUS</td>
<td>Laparoscopy + histopathology</td>
<td>35.8</td>
<td>1-3 months (36)</td>
<td></td>
</tr>
<tr>
<td>Saba et al, 2011</td>
<td>Italy</td>
<td>Prospective</td>
<td>59</td>
<td>TVUS, MRI</td>
<td>Laparoscopy, laparotomy + histopathology</td>
<td>Not given</td>
<td>Not specified (37)</td>
<td></td>
</tr>
<tr>
<td>Saccardi et al, 2012</td>
<td>Italy</td>
<td>Prospective</td>
<td>54</td>
<td>Clinical examination, TVUS, MRI</td>
<td>Laparoscopy + histopathology</td>
<td>32.3</td>
<td>Not specified (38)</td>
<td></td>
</tr>
<tr>
<td>Savelli et al, 2011</td>
<td>Italy</td>
<td>Prospective</td>
<td>69</td>
<td>TVUS</td>
<td>Laparoscopy + histopathology</td>
<td>33.6</td>
<td>Within 1 month (39)</td>
<td></td>
</tr>
<tr>
<td>Stratton et al, 2003</td>
<td>USA</td>
<td>Prospective</td>
<td>48</td>
<td>MRI</td>
<td>Laparoscopy + histopathology</td>
<td>Age range (20-44)</td>
<td>Within 1 month (40)</td>
<td></td>
</tr>
<tr>
<td>Takeuchi et al, 2005</td>
<td>Japan</td>
<td>Prospective</td>
<td>31</td>
<td>MRI</td>
<td>Laparoscopy + histopathology</td>
<td>32.1</td>
<td>Not specified (41)</td>
<td></td>
</tr>
<tr>
<td>Ubaldi et al, 1998</td>
<td>Belgium</td>
<td>Prospective</td>
<td>133</td>
<td>TVUS</td>
<td>Laparoscopy + histopathology</td>
<td>Age range (21-41)</td>
<td>1 day</td>
<td></td>
</tr>
<tr>
<td>Valenzeno Menada et al, 2008</td>
<td>Italy</td>
<td>Prospective</td>
<td>181</td>
<td>TVUS</td>
<td>Laparoscopy + histopathology</td>
<td>32.0</td>
<td>Not specified (43)</td>
<td></td>
</tr>
<tr>
<td>Vimercati et al, 2012</td>
<td>Italy</td>
<td>Prospective</td>
<td>1440</td>
<td>TVUS</td>
<td>Laparoscopy + histopathology</td>
<td>34.0</td>
<td>4-8 weeks (44)</td>
<td></td>
</tr>
</tbody>
</table>

TVUS, transvaginal ultrasound; TRUS, transrectal ultrasound.
physical examination. The AUC was 0.76 (95% CI, 0.66-0.83), which indicates that physical examination has an intermediate level of diagnostic value.

The clinical utility of physical examination for DIE was average and Fagan's nomogram (Fig. 4A) indicated that post-test probability (positive, 37%; negative, 10%) differed significantly from pre-test probability (20%). There was substantial heterogeneity, with an $I^2$ value of 75%, and $\chi^2$ test for heterogeneity was statistically significant ($P=0.009$). A bivariate box plot (Fig. 5A) indicated that there was a study out of the circle, illustrating the heterogeneity between the included studies. As there were <10 studies, publication bias was not tested.

### Diagnostic performance of TVUS

In total, 21 studies assessed the diagnostic accuracy of TVUS. The pooled sensitivity and specificity of TVUS for diagnosis of DIE were 76% (95% CI, 67-83%) and 94% (95% CI, 88-97%), respectively. The DOR was 47 (95% CI, 21-104) as displayed in the Forest plot (Fig. 2B). LRP was 12.0 (95% CI, 6.3-22.7) and LRN was 0.25 (95% CI, 0.18-0.36). Fig. 3B depicts the sROC curve for TVUS. The AUC was 0.92 (95% CI, 0.86-0.96), which indicates the high diagnostic value of TVUS.

The clinical utility of TVUS for DIE was good and Fagan's nomogram (Fig. 4B) revealed that post-test probability (positive, 75%; negative, 6%) differed significantly from pre-test probability (20%). There was substantial heterogeneity, with an $I^2$ value of 99%, and $\chi^2$ test for heterogeneity was statistically significant ($P<0.001$). A bivariate box plot (Fig. 5B) indicated 4 studies out of the circle illustrating heterogeneity between the included studies. The
Funnel plot (Fig. S2) for publication bias was symmetrical with non-significant Deek's test (P=0.60).

**Diagnostic performance of TRUS.** In total, six studies assessed the diagnostic accuracy of TRUS. The pooled sensitivity and
specificity of TRUS for diagnosis of DIE were 91% (95% CI, 74-97%) and 80% (95% CI, 43-95%), respectively. The DOR was 39 (95% CI, 35-43) as shown in the Forest plot (Fig. 2C). LRP was 4.5 (95% CI, 1.1-18.4) and LRN was 0.12 (95%
CI, 0.03-0.45). Fig. 3C depicts the sROC curve for TRUS. The AUC was found to be 0.93 (95% CI, 0.86-0.97), which indicates the high diagnostic value of TRUS.

The clinical utility of TRUS for DIE was good and Fagan's nomogram (Fig. 4C) showed that post-test probability (positive, 53%; negative, 3%) differed significantly from pre-test probability (20%). There was substantial heterogeneity, with an $I^2$ value of 83%, and the $\chi^2$ test for heterogeneity was statistically significant ($P=0.001$). A bivariate box plot (Fig. 5C) indicated that there was a study out of the circle, illustrating the heterogeneity between the included studies. As there were <10 studies, publication bias was not tested.

Diagnostic performance of MRI. In total, 13 studies assessed the diagnostic accuracy of MRI. The pooled sensitivity and specificity of MRI for diagnosis of DIE were 82% (95% CI, 70-90%) and 87% (95% CI, 78-92%), respectively. The DOR was 30 (95% CI, 14-62) as shown in the Forest plot (Fig. 2D). LRP was 6.1 (95% CI, 3.7-9.9) and LRN was 0.21 (95% CI, 0.12-0.35). Fig. 3D depicts the sROC curve for MRI. The AUC was found to be 0.91 (95% CI, 0.80-0.97), which indicates a high diagnostic value of MRI.

The clinical utility of MRI for DIE was good and Fagan's nomogram (Fig. 4D) showed that post-test probability (positive, 60%; negative, 5%) differed significantly from pre-test probability (20%). There was substantial heterogeneity, with an $I^2$ value of 98%, and the $\chi^2$ test for heterogeneity was statistically significant ($P<0.001$). A bivariate box plot (Fig. 5D) indicated 2 studies out of the circle, showing the heterogeneity between the included studies. The Funnel plot (Fig. S3) for publication bias was symmetrical with non-significant Deek's test ($P=0.41$).

Discussion
Various imaging modalities are available for diagnosis of DIE. Clinical history and physical examination of the pelvis may
suggest endometriosis; however, the heterogeneity of clinical presentation, the prevalence of asymptomatic endometriosis (2-50%), and a lack of association between disease severity and presenting symptoms results in significant difficulty in obtaining an accurate diagnosis based solely on patient history and physical examination (4,5,8,9). Several advanced ultrasonographic techniques have been described to identify endometriosis, such as pelvic organ mobility, sliding...
sign, rectal water contrast, tenderness-guided ultrasound and bowel preparation (25,33,43). These techniques can significantly improve the diagnostic performance of TVUS for endometriosis. Furthermore, the MRI ‘jelly method’ appears to have improved diagnostic accuracy compared to older MRI modalities (41). Although these modalities cannot replace histopathology or biopsy as the gold standard for diagnosis, a high diagnostic accuracy of such non-invasive methods may reduce the diagnostic delay and complications associated with invasive diagnostic methods. Hence, it is important to precisely determine the diagnostic accuracy of each of these modalities. In the present review of 30 studies with 4,565 participants, the results suggested that while physical examination has intermediate diagnostic accuracy, TRUS, TVUS and MRI have high diagnostic accuracy for DIE.

To the best of our knowledge, diagnostic accuracy of physical examination for DIE has not been assessed in any previous reviews. The present study results indicated the pooled sensitivity and specificity of physical examination to be 71 and 69%, respectively, with average diagnostic accuracy (AUC = 0.76). TVUS had a pooled sensitivity and specificity of 76 and 94% with high diagnostic accuracy (AUC = 0.92). These findings are in agreement with earlier reviews on the diagnostic accuracy of TVUS, which also indicated the high specificity and high diagnostic accuracy of this technique (AUC > 0.90) (11,46-48). However, compared with the previous reviews, a greater number of studies were included in the present analysis.

In the present analysis, TRUS had a pooled sensitivity of 91% and a specificity of 80% with an AUC of 0.93, which is similar to the diagnostic accuracy of TVUS. To the best of our knowledge, no previous reviews have compared the accuracy of TRUS and TVUS for the diagnosis of DIE. The present study revealed MRI to have lower sensitivity (82%) and specificity (87%) for diagnosis of DIE when compared with the previous reviews of Moura et al (46) and Nisenblat et al (11). Their studies demonstrated ~90% sensitivity and specificity of MRI for diagnosis of DIE. However, compared to these previous reviews, a greater number of studies were included in the present analysis. The results of the present study also suggested that TVUS may be used to exclude the diagnosis of endometriosis with certainty, which was not seen in the previous reviews.

The results of the present study should be interpreted cautiously, considering the inter-study heterogeneity and quality of the included studies. The high heterogeneity limits any strong conclusions on the diagnostic accuracy of the non-invasive modalities for DIE. There was no significant publication bias among studies reporting diagnostic accuracy of TVUS and MRI. However, publication bias for studies reporting TRUS and physical examination could not be assessed due to the limited number of studies included. The overall quality of studies was good, as most of the studies had a low risk of bias with respect to all the four domains measured using the QUADAS tool.

The present study had certain strengths. A comprehensive review including 30 studies with 4,565 participants to evaluate
the accuracy of various imaging modalities and clinical examination in diagnosing DIE was conducted. To the best of our knowledge, this is the first review to provide pooled estimates for four different diagnostic methods of DIE. The lack of any significant publication bias adds credibility to the current results. Nonetheless, the present study has certain limitations. First, some studies included in the review had a high risk of bias in certain domains, which may have influenced individual study results. Secondly, significant heterogeneity was revealed between the studies included in the review, which limited the interpretation of the pooled outcomes.

Despite these limitations, the present study provided valuable insights regarding the diagnostic accuracy of various non-invasive techniques for DIE. Though all the imaging modalities had good sensitivity and specificity, only TVUS was close to fulfilling the SpPin i.e., specific positive in triage test criteria (95% specificity). None of the modalities met the SnNout criteria i.e. sensitive negative for 95% sensitivity of a diagnostic test (11,49). This suggests that TVUS may be used to diagnose DIE, but cannot confirm a woman to be disease-free based on radiological evidence. These findings are in line with the international guidelines for the diagnosis of DIE, which suggests TVUS as a first-line imaging modality following history and physical examination (50,51). However, TVUS does not replace laparoscopic surgery and biopsy, which is still the gold standard for diagnosing DIE.

To summarize, the present study demonstrated that TRUS, TVUS and MRI have good diagnostic accuracy for DIE and have a high sensitivity and specificity. The diagnostic accuracy was most similar for TVUS, TRUS and MRI, with all the modalities having AUC >0.9. Physical examination was also indicated to have reasonable diagnostic accuracy. TVUS was the most specific imaging modality while TRUS was the most sensitive modality for diagnosis of DIE. However, these findings should be interpreted with caution owing to the significant heterogeneity between the included studies. These results suggest that these imaging modalities can be utilized as efficient and time-saving screening tools for DIE.

Acknowledgements

Not applicable.

Funding

No funding was received.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. The relevant data is also available in the referenced studies.

Authors’ contributions

XZ conceived and designed the study. XZ, TH and WS collected the data, performed the literature search and analyzed the data. XZ was involved in the writing of the manuscript. All authors have read and approved the final manuscript.

Ethical approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

ESHRE guideline: Management of women with endometriosis


