

# Diagnosis of pelvic endometriosis: a systematic review and accuracy meta-analysis of non-invasive tests.

Thomas Bainton<sup>1</sup>, Sobanakumari Jeyapala<sup>1</sup>, Javier Zamora<sup>2</sup>, Khalid Khan<sup>3</sup>, and Amer Raza<sup>1</sup>

<sup>1</sup>Chelsea and Westminster Hospital NHS Foundation Trust

<sup>2</sup>Hospital Ramón y Cajal (IRYCIS) CIBERESP

<sup>3</sup>University of Granada

October 26, 2022

## Abstract

Background: Endometriosis is a chronic, often debilitating condition with a current significant delay from symptom onset to diagnosis. Objectives: To investigate the accuracy of symptoms, clinical history and non-invasive tests to predict pelvic endometriosis. Data sources: Medline, Embase, Web of Science and Scopus from conception to September 2022. Selection criteria: Primary test accuracy studies assessing selected non-invasive tests against a reference standard diagnosis for endometriosis. Data extraction and synthesis: Two authors independently conducted data extraction and study quality assessment. Grading of evidence was performed using a novel visual pentagon model. Meta-analyses of test accuracy was estimated using bivariate random effects models. Results: The 125 included studies (250,574 participants) showed mixed quality. Studies applying non-surgical (database/self-reporting) reference standard had a greater risk of bias. In 98 studies applying surgical reference standard, summary diagnostic odds ratios were: dysmenorrhoea 2.56 (95% confidence interval 1.99-3.29); pelvic pain 2.56 (1.73-3.74); dyschezia 2.05 (1.36-3.10); dyspareunia 2.45 (1.71-3.52); family history of endometriosis 6.79 (4.08-11.3); nulligravidity of 2.01 (1.62-2.50); BMI [?]30kg/m<sup>2</sup> 0.37 (0.19-0.68); TVUSS endometrioma 91.2 (44.0-189); TVUSS invasive endometriosis 26.1 (9.28-73.5); and CA-125 >35U/mL 16.0 (8.09-31.7). Sensitivity analysis excluding all high-risk studies found concordant results. Conclusions: This meta-analysis collated the performance of non-invasive tests for endometriosis across a comprehensive and geographically varied population. Study quality was mixed, however results were consistent with high-risk studies excluded. These findings will inform future prediction models for triage in primary care. Funding: This research received no specific funding. Keywords: Endometriosis; diagnosis; laparoscopy; pelvic pain; sub-fertility

Title:

**Diagnosis of pelvic endometriosis: a systematic review and accuracy meta-analysis of non-invasive tests.**

Thomas Bainton<sup>1,2\*</sup>; Sobanakumari Jeyapala<sup>1</sup>; Javier Zamora<sup>3,5</sup>; Khalid S Khan<sup>4,5</sup>; Amer Raza<sup>1,2</sup>

<sup>1</sup>Department of Obstetrics and Gynaecology, Chelsea and Westminster Hospital NHS Foundation Trust, London SW10 9NH, UK

<sup>2</sup>Imperial College London, Exhibition Road, London SW5

<sup>3</sup>Clinical Biostatistics Unit. Hospital Ramon y Cajal (IRYCIS), Madrid 28034, Spain. Institute of Metabolism and Systems Research, University of Birmingham B15 2TT, UK

<sup>4</sup>Department of Preventative Medicine and Public Health, University of Granada, Granada 18016, Spain

<sup>5</sup>Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Madrid 28029, Spain

**\*Corresponding author:**

Dr. T. Bainton

Department of Obstetrics and Gynaecology, Chelsea and Westminster Hospital, 369 Fulham Road, London SW10 9NH. United Kingdom

E-Mail: [t.bainton22@imperial.ac.uk](mailto:t.bainton22@imperial.ac.uk) *orcid.org/0000-0003-1445-9050*

**Abstract:**

**Background:**

Endometriosis is a chronic, often debilitating condition with a current significant delay from symptom onset to diagnosis.

**Objectives :**

To investigate the accuracy of symptoms, clinical history and non-invasive tests to predict pelvic endometriosis.

**Data sources:**

Medline, Embase, Web of Science and Scopus from conception to September 2022.

**Selection criteria:**

Primary test accuracy studies assessing selected non-invasive tests against a reference standard diagnosis for endometriosis.

**Data extraction and synthesis:**

Two authors independently conducted data extraction and study quality assessment. Grading of evidence was performed using a novel visual pentagon model. Meta-analyses of test accuracy was estimated using bivariate random effects models.

**Results:**

The 125 included studies (250,574 participants) showed mixed quality. Studies applying non-surgical (database/self-reporting) reference standard had a greater risk of bias. In 98 studies applying surgical reference standard, summary diagnostic odds ratios were: dysmenorrhoea 2.56 (95% confidence interval 1.99-3.29); pelvic pain 2.56 (1.73-3.74); dyschezia 2.05 (1.36-3.10); dyspareunia 2.45 (1.71-3.52); family history of endometriosis 6.79 (4.08-11.3); nulligravidity of 2.01 (1.62-2.50); BMI  $\geq 30\text{kg/m}^2$  0.37 (0.19-0.68); TVUSS endometrioma 91.2 (44.0-189); TVUSS invasive endometriosis 26.1 (9.28-73.5); and CA-125  $>35\text{U/mL}$  16.0 (8.09-31.7). Sensitivity analysis excluding all high-risk studies found concordant results.

**Conclusions:**

This meta-analysis collated the performance of non-invasive tests for endometriosis across a comprehensive and geographically varied population. Study quality was mixed, however results were consistent with high-risk studies excluded. These findings will inform future prediction models for triage in primary care.

**Funding:**

This research received no specific funding.

**Keywords:**

Endometriosis; diagnosis; laparoscopy; pelvic pain; sub-fertility

## Word Count:

3318

## Funding:

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

## Introduction:

Endometriosis is a chronic condition affecting women of reproductive age characterised by extra-uterine deposits of endometrial tissue with a prevalence of up to 10%.<sup>1-3</sup> It can be found in peritoneal deposits, ovarian endometriomata and invasive disease, with sufferers experiencing pain and sub-fertility.<sup>2,4-6</sup> There is currently a 7-12 years delay between symptom onset and definitive diagnosis, by laparoscopy and biopsy.<sup>7-9</sup> Reducing this has long been a research priority.<sup>10,11</sup> Up to 58% of sufferers present to primary care 10 or more times before diagnosis, demonstrated across a variety of healthcare systems and geographies.<sup>4,7,12-14</sup> A better understanding of the accuracy with which symptoms, clinical history and non-invasive tests diagnose pelvic endometriosis would aid triage and referral.

Many studies have attempted to find replacement tests for diagnostic laparoscopy.<sup>15-17</sup> These are often invasive or otherwise not applicable to primary care.<sup>16,18</sup> Previous meta-analyses have been restricted by a narrow inclusion criteria yielding a small number of eligible studies, limiting findings.<sup>19</sup> Meta-analyses assessing imaging and biomarkers showed that no individual test met their criteria as a replacement or triage test alone, but findings on trans-vaginal ultrasound and serum CA-125 showed a high specificity for disease.<sup>20,21</sup> These studies did not include assessment of other clinical factors. A 2019 narrative review demonstrated the importance of clinical factors in prediction of disease, but primary studies were not assessed for quality and absence of meta-analysis limited quantitative assessment of test performance.<sup>22</sup>

A 2021 case-control study reporting on the accuracy of a simple patient-completed questionnaire identified those at high or low risk of disease with good accuracy reflecting the utility of assessment of patient reported symptoms at a primary care level.<sup>24</sup>

Research related to endometriosis is increasing in volume, with 75% of primary studies published in the last decade.<sup>25,26</sup> Given this, and due to the limitations of previous reviews a new comprehensive systematic review and meta-analysis is required. We performed such an evidence synthesis and determined the accuracy of symptoms, clinical history, a simple low-cost biomarker and first line ultrasound for the diagnosis of pelvic endometriosis by means of a comprehensive systematic review and accuracy meta-analysis.

## Methods:

The protocol was designed and registered with PROSPERO (Registration number: CRD42020187543).<sup>27</sup> Reporting followed the PRISMA guidelines.<sup>28</sup>

## Patient and public involvement

A patient and public involvement meeting was held following an open invitation on social media. The aim to reduce the diagnostic delay was well supported and resonated with their personal experience.

## Literature search and study selection

Medline, Embase, Web of Science and Scopus were searched from conception to September 2022. Search strategies are shown in Appendix S1. Review article references were also screened.

Titles and abstracts were screened independently by two authors (TB; SJ) using EndNote-X9 and duplicate or irrelevant studies removed.<sup>29</sup> Full texts were screened and justification for inclusion or exclusion recorded, differences were resolved by discussion with the senior author (AR).

We included published peer reviewed studies reporting accuracy estimates to predict pelvic endometriosis (peritoneal; ovarian; or invasive disease >5mm retroperitoneal invasion) for one or more index tests in participants with reported presence/absence of endometriosis. Included tests were dysmenorrhoea; pelvic pain; dyschezia; dyspareunia; nulligravidity; BMI  $\geq 30 \text{ kg/m}^2$ ; family history of endometriosis; transvaginal ultrasound finding (TVUSS) of endometrioma; TVUSS finding of invasive disease; and serum CA-125  $> 35 \text{ U/mL}$ .

Target population was reproductive age women excluding pregnancy or systemic co-morbidities. Studies reporting non-reproductive age participants were included only where their data could be excluded from meta-analysis. Studies were included where a 2x2 contingency table for index test(s) could be constructed. We imposed no limits to language, setting, or number of participants. All non-English studies were translated by a medically trained native speaker.

Studies reporting laboratory tests and imaging were included only when performed prior to reference standard and with ultrasound using only standard 2D protocols. Definitions for each test for the purposes of study selection are shown in Appendix Table 1.

We excluded reviews; case reports; studies where information on recruitment or study population was unavailable; letters; and abstracts. Studies reporting non-pelvic endometriosis were included only where pelvic endometriosis was reported separately. Authors were contacted only to obtain full texts or, failing this, they were obtained through the British Library. Studies with incomplete data preventing determination of inclusion, exclusion or test accuracy were excluded.

### Data extraction and quality assessment

For each included study, two authors (TB; SJ) independently recorded information on study characteristics and data was extracted to form 2x2 tables. Where there was unreliability in data extraction from some non-English language studies, these were excluded.

Risk of bias and applicability was assessed independently by two authors (TB; SJ) using the Quality Assessment of Diagnostic Test Accuracy Studies (QUADAS 2) tool.<sup>30</sup> For studies regarding serum CA-125 or TVUSS we included the additional signalling questions: ‘was the index test performed by a single operator?’ to assess inter-observer bias; and ‘was timing in the participants’ menstrual cycle controlled for?’. We adjusted the original question ‘if a threshold was used, was it pre-specified?’ to ‘was there a clear definition of what was considered a positive test?’.

### Data synthesis

Due to differences in design, studies were divided into groups according to application of the reference standard: ‘Complete verification’, all participants received visual inspection of the pelvis at surgery; ‘Partial verification’, all cases received surgical confirmation but controls did not; and ‘Database/self-reporting’, cases confirmed by healthcare coding or self-reporting and controls from healthy populations not known to have endometriosis.

Statistical analysis was conducted using Stata software (version 15)<sup>31</sup> to allow exploration of heterogeneity and statistical pooling using a bivariate random effects model and produced summary accuracy measures and summary receiver operative characteristic curves for each index test. A bivariate random effects model was applied for index tests with  $\geq 5$  contributing studies, and a univariate fixed effects model for index test with  $\geq 4$ .

Index tests were assessed for performance as a ‘rule-in’ or ‘rule-out’ tool with pre-specified threshold summary accuracy of 95% sensitivity/50% specificity or 95% specificity/50% sensitivity respectively.

### Results:

#### Study selection and characteristics

Of 22,016 studies identified 125 met the inclusion criteria involving 250,574 participants (Figure 1). Characteristics of included and excluded studies are shown in Table S1. Details of included studies by number of participants and index test(s) are shown in Appendix Table 1.

Mean number of index tests per study was 2 (range 1-6). A total of 241 were assessed across all studies. Included studies were geographically varied: 45 from Europe, 34 North America, 19 Asia/Oceania, 13 the Middle East/Africa, 12 South America, and 2 transcontinental. Publication date ranged from 1986 to 2022 with 57% since 2010: 4 before 1989; 22, 1990-1999; 18, 2000-2009; and 71, 2010-2022.

Most studies (75) were 'single-gate' design with 50 of 'two-gate' design, including all studies in the partial verification and database/self-reporting groups. The mean prevalence of endometriosis in studies of a 'single-gate' design was 52% (range 9-93%), due to the selection of matched controls, prevalence in 'two-gate' studies was not relevant. There was heterogeneity in population selection, with participants having surgery for a broad range of indications such as uterine fibroids or adnexal cysts as well as pelvic pain or sub-fertility.

In the 61 studies assessing symptom-based tests 20 did so by self-administered questionnaire; 14 by structured interview; 12 by clinical history taking, and 15 were undefined.

## Quality of included studies

### Risk of bias

The assessment of study quality by QUADAS-2 is presented in Figure 2 and Supplementary Figures S2-5. Overall methodological quality was mixed, with 5 studies presenting a low risk of bias across all domains<sup>32-36</sup> and 64 presenting a high risk of bias or applicability in at least one domain.

In patient selection, 22 studies presented a low risk of bias, with 73 and 30 presenting an unclear or high risk respectively. Non-consecutive or non-random selection, two-gate selection for cases and controls, and having a highly selected group of participants (infertility cohort, surgery for a narrow indication etc.) were the main reasons for a high risk of bias.

Symptom based index tests presented an unclear or high risk of bias due to a lack of definition of a positive test and of blinding. Just 9 studies presented a low risk in symptom-based tests across all groups. Index tests applicable to clinical history or investigations performed better, with 66 studies presenting a low risk. Reasons for an unclear or high risk of bias were a lack of pre-specified criteria for a positive test; no blinding to results of the reference standard; and inter-observer variability regarding imaging. Studies in the partial verification group assessed a proportionally higher number of index tests nulligravidity and BMI [ $?/30\text{kg/m}^2$ ], which, less subjective to interpretation presented a lower risk of bias.

The risk of bias regarding the reference standard performed best in the complete and partial verification groups where 74 studies were at low risk of bias. Those with an unclear or high risk lacked information on how likely the surgery was to correctly classify the target condition or operators not blinded to the result of index test(s). In the database/self-reporting group, 5 studies assessed for probable surgical confirmation by means of additional codes at the time of recording and therefore presented a lower risk, all other studies were high risk.

In flow and timing, the complete verification group presented the lowest risk. An unclear or high risk of bias was attributable to a long ( $>12$  months) or unclear time interval, and a high or unclear withdrawal of participants from analysis. All other studies presented a high risk as not all participants received the same reference standard.

### Applicability

In patient selection, 40 studies gave low concern, with 63 and 22 giving unclear or high risk respectively. An unclear or high risk was attributable to the two-gate selection of controls, or the study likely to only classify a limited spectrum of disease (tertiary centres or infertility clinics).

In regard the reference standard, 97 studies showed a low concern. Studies in the database/self-reporting group were deemed high/unclear depending on whether additional coding input for surgery was recorded.

## Test accuracy

Due to heterogeneity in methodology and study quality, meta-analysis was performed on studies from each group separately.

The accuracy of index tests to predicting endometriosis was variable, although results across groups were consistent. Each index test gave a positive likelihood for the presence of pelvic endometriosis, apart from a BMI  $\geq 30 \text{ kg/m}^2$ , which decreased the likelihood of disease. The positive likelihood ratio (LR+) for disease was highest in investigation tests and there was a trend towards a greater specificity than sensitivity. The summary results of bi/univariate meta-analysis are shown in Figure 3. An assessment of confidence in individual sensitivity and specificity of each test is displayed by a visual pentagon model, the methodology for this assessment is described in the discussion and legend shown in Figure 4.

Investigation category tests were the best performing overall and TVUSS finding of endometrioma gave the highest summary LR+ at 21.6, at sensitivity and specificity of 77.2% and 96.4% respectively. Serum CA-125  $> 35 \text{ U/mL}$  showed sensitivity and specificity of 55.8% and 92.7% respectively, with LR+ of 7.63. TVUSS finding of DIE had showed sensitivity and specificity of 86.5% and 80.2% with LR+ of 4.39.

Symptom based tests showed LR+ within a similar range: 1.47 (dysmenorrhoea) to 1.93 (dyspareunia). Symptoms showed a generally higher specificity than sensitivity. Dyspareunia showed the highest LR+ at 1.93 with a sensitivity and specificity of 36.3% and 81.1% respectively.

Family history of endometriosis showed a LR+ of 6.25 with a high specificity (98.5%) but low sensitivity (9.25%). The finding of BMI  $\geq 30 \text{ kg/m}^2$  showed a decreased likelihood of diagnosis of endometriosis (LR+ 0.44).

Hierarchical Summary Receiver Operating Characteristics (HSROC) curves for index tests in each group are shown in Figures S6-8. The HSROC curves show the greatest area under the curve (AUC) for investigation category tests.

In the partial verification group, symptom index tests showed a greater LR+ than the complete verification group, range 2.47 (dysmenorrhoea) to 7.13 (dyschezia). Specificity was also higher, range 69% (dysmenorrhoea) to 92% (dyschezia).

In the database/self-reporting group symptom-based index tests performed similarly to other groups. In partial verification and database/self-reporting groups BMI  $\geq 30 \text{ kg/m}^2$  showed no correlation with disease and had 95% CI crossing 1.0. In all other index tests across all groups the 95% CI was  $> 1.0$ .

The greatest inter-study variability in confidence intervals was shown in Forest plots for the symptom-based tests, notably pelvic pain. The inter study variance for specificity was generally lower than that for sensitivity, as was the overall width of confidence intervals. Forest plots for each index test in each group are shown in Supplementary Figures S9-15.

Sensitivity analysis performed for studies without any high-risk features is shown in Table 1. All studies included are from the complete verification group. Summary accuracy measures are consistent with those in this group for the majority of index tests, although sensitivity for TVUSS finding of endometrioma and DIE reduced to 69.8% and 73.4% respectively.

## Discussion:

### Main findings

This meta-analysis presents an up-to-date, large, and geographically varied data set identifying predictive factors for diagnosis of pelvic endometriosis with a high degree of confidence. Index tests showed a positive association with endometriosis and trended towards a greater specificity than sensitivity, excluding

elevated BMI, which demonstrated an inverse correlation. TVUSS finding of endometrioma reached a desired threshold for use as a ‘rule-in’ test and none achieved a summary sensitivity of >95% for ‘rule-out’. A family history of endometriosis, dyschezia and serum CA-125 >35U/mL showed summary specificity of >90% although low sensitivity. Sensitivity was poor for symptom and clinical history tests, where the best performing was dysmenorrhoea.

## Strengths and limitations

We undertook a thorough search of the current literature, undergoing analysis by two independent reviewers with strict quality assessment. Attempts were made to mitigate inter-study heterogeneity by division of studies into groups. All index tests are relevant to primary care and immediately available without novel techniques or additional training.

There were, however, limitations. Due to difficulties in data extraction from some non-English journals, 15 studies were excluded from the analysis. Some studies, such as Chapron et al 2005, which was seminal in providing a clinical prediction model for moderate/severe endometriosis, were not able to be included due to the inability for construct 2x2 tables.<sup>37</sup> We did not contact authors to obtain individual data that was not available in the published text.

Overall, there was significant methodological variance and population heterogeneity in age; presentation; and stage of disease. Variation in selection of cases and controls may not reflect a clinically representative population. Prevalence of disease was higher than seen in the general female population, which may reflect a high degree of surgical accuracy, but also indicates the selective nature of study populations.

There is the possibility of inappropriate assignment of cases and controls, occurring in both directions due to uneven application of the reference standard, although we attempted to account for this by assigning groups. We included studies that diagnosed endometriosis by visual inspection and there is debate regarding this in the absence of histological confirmation.<sup>38-40</sup>

There was variation in the definition of positive symptom index tests. This is common across many reviews and although there is guidance on symptom reporting, it was not clearly followed in all studies.<sup>41-43</sup> Assessment of symptoms varied, with most studies using a self-administered questionnaire. Although the use of standardised validated tools would better allow for comparison across studies, the nuance and detail acquired through clinical history taking is likely to better grasp the nature and significance of a symptom and its implications.

It is likely that imaging and surgical techniques have developed over time. A trend towards recent studies may mitigate this.

Considering the balances of strengths and weaknesses, however, we believe that our data synthesis presents an objective summary of the current evidence.

## Interpretation

An understanding of the degree of likelihood associated with various symptoms and features in the clinical history can help assessment of patients with possible endometriosis in primary care.

The negative association between elevated BMI and endometriosis shown in the complete verification group is consistent with that demonstrated previously.<sup>44</sup> This was not replicated across other groups. This may reflect a greater negative correlation between elevated BMI in higher risk populations in the all surgical cohorts who may have more severe disease. This possibility is consistent with previous studies, demonstrating a significantly lower BMI in those with severe compared to mild disease and a 12-14% decrease in the likelihood of endometriosis being diagnosed for each unit increase in BMI (kg/m<sup>2</sup>).<sup>32,45</sup> The interplay between BMI and endometriosis pathogenesis, however, remains poorly understood.

The trend of data from the partial verification and database/self-reporting groups to demonstrate better performing accuracy measures was likely a reflection of the selection of controls. This effect seems to outweigh

the possibility of an undiagnosed disease burden in those not exposed to a surgical reference standard. The accuracy of self-reported diagnosis of endometriosis has been assessed and performs well,<sup>46</sup> false attribution of disease in the self-reporting group may therefore only present a small source of bias.

A greater specificity than sensitivity of tests may be associated with their correlation to disease severity. Dyschezia and dyspareunia have been linked to severe disease due to the involvement of a precise anatomical location in invasive disease, for example, but are less often present in mild cases.<sup>47,48</sup> Tests showing a greater sensitivity such as dysmenorrhea were also less specific, which may only become specific for endometriosis in more severe forms.

Previous systematic reviews have similarly highlighted the heterogeneity and poor methodological quality of primary studies, limiting interpretation of findings.<sup>17,49</sup> As our methodology allowed wide inclusion criteria, we applied a novel grading protocol to more quantitatively assess limitations. Grading of evidence for index tests was performed for sensitivity and specificity by application of a visual pentagon model for grading of test accuracy studies described by Rogozinska and Khan.<sup>50</sup> This methodology is described in detail elsewhere but briefly, studies were given a score of 0 to -2 in each of 5 domains: design (study design type); risk of bias (QUADAS 2 risk of bias); indirectness (QUADAS 2 applicability); inconsistency (visual assessment of inter-study variance in confidence intervals); and imprecision (width of confidence intervals). The complete verification group showed the fewest limitations, whilst the database/self-reporting studies showed very serious limitations. There was greater limitation in the investigation category tests due to more highly selective populations and a generally higher inter-study inconsistency and imprecision.

## Conclusion:

### Research recommendations

The need for high-quality studies of predictive factors for endometriosis remains, particularly assessing populations attending primary care. Further multivariate analysis in powerful primary observational studies assessing factors that can be immediately and readily assessed in primary care would be of great value, as we anticipate the index tests assessed in this study to provide a greater degree of accuracy when applied in combination.<sup>4,37,51</sup>

We examined serum CA-125 at a cut off  $>35$  U/mL, considered the upper limit of normal range, meta-analysis from 2016 found a cut off of 30 U/mL gave a sensitivity and specificity of 52% and 93% respectively, but sensitivity dropped to just 24% for detection of minimal disease.<sup>53</sup> Further research assessing the accuracy of CA-125 at different thresholds and in combination with other tests could help improve accuracy.

Two recent studies (Fauconnier et al 2021 and Chapron et al 2022) assessed the accuracy of a patient-completed questionnaire and epidemiological data for the early identification of endometriosis and found it could do so with high diagnostic accuracy.<sup>24,55</sup> Although these studies were conducted in a high-risk population undergoing surgery, the model maintained accuracy in population with a lower endometriosis prevalence of 10%. We do not anticipate a clinical score replacing laparoscopy due to its added therapeutic advantages and requirement to exclude other pathologies. If, however, disease can be predicted with a high degree of accuracy early on, medical therapy may be instigated, and referral made for definitive diagnosis and counselling regarding treatment, prognosis and fertility in a timely manner with the aim of reducing the current extraordinary delay.

### Acknowledgements:

KSK is Distinguished Investigator at University of Granada funded by the Beatriz Galindo (senior modality) program of the Spanish Ministry of Education.

### Disclosure of interests:

The authors declare no conflict of interest in respect to the topic or contents of this review.

### Contribution to authorship:



Study design: TB, KSK and AR. Search strategy: TB and AR. Study selection: TB, SJ and AR. Data extraction: TB and SJ. Statistical analysis: JZ. Manuscript drafting: TB, JZ, KSK and AR. Critical discussion: TB, SJ, JZ, KSK and AR.

### Details of ethics approval:

Ethical approval was not required for this study.

### References:

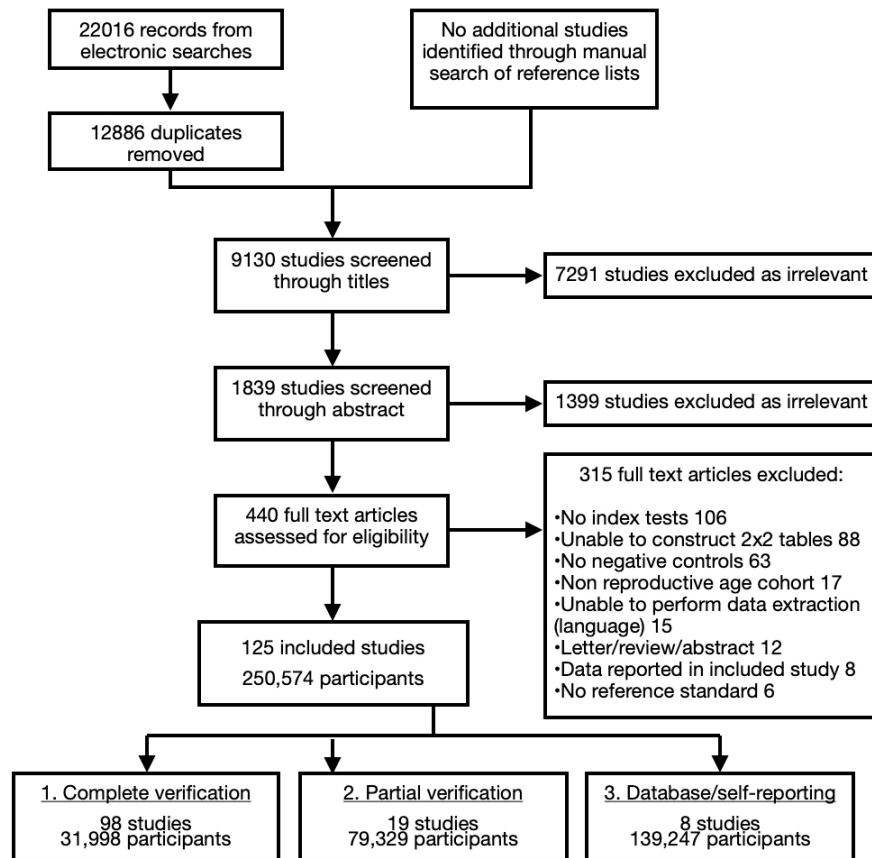
1. Eskenazi B, Warner ML. Epidemiology of endometriosis. *Obstet Gynecol Clin North Am.* 1997;24(2):235-258.
2. Farquhar C. Endometriosis. *Bmj.* 2007;334(7587):249-253.
3. Giudice LC. Clinical practice. Endometriosis. *N Engl J Med.* 2010;362(25):2389-2398.
4. Ballard KD, Seaman HE, de Vries CS, Wright JT. Can symptomatology help in the diagnosis of endometriosis? Findings from a national case-control study—Part 1. *Bjog.* 2008;115(11):1382-1391.
5. Prescott J, Farland LV, Tobias DK, et al. A prospective cohort study of endometriosis and subsequent risk of infertility. *Hum Reprod.* 2016;31(7):1475-1482.
6. Fuldeore MJ, Soliman AM. Prevalence and Symptomatic Burden of Diagnosed Endometriosis in the United States: National Estimates from a Cross-Sectional Survey of 59,411 Women. *Gynecol Obstet Invest.* 2017;82(5):453-461.
7. Nnoaham KE, Hummelshoj L, Webster P, et al. Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries. *Fertil Steril.* 2011;96(2):366-373.e368.
8. Kuznetsov L, Dworzynski K, Davies M, Overton C. Diagnosis and management of endometriosis: summary of NICE guidance. *Bmj.* 2017;358:j3935.
9. Zondervan KT, Becker CM, Koga K, Missmer SA, Taylor RN, Viganò P. Endometriosis. *Nat Rev Dis Primers.* 2018;4(1):9.
10. Rogers PAW, D’Hooghe TM, Fazleabas A, et al. Priorities for endometriosis research: Recommendations from an international consensus workshop. *Reproductive Sciences.* 2009;16(4):335-346.
11. Rogers PA, Adamson GD, Al-Jefout M, et al. Research Priorities for Endometriosis. *Reprod Sci.* 2017;24(2):202-226.
12. Arruda MS, Petta CA, Abrão MS, Benetti-Pinto CL. Time elapsed from onset of symptoms to diagnosis of endometriosis in a cohort study of Brazilian women. *Hum Reprod.* 2003;18(4):756-759.
13. Surrey E, Soliman AM, Trenz H, Blauer-Peterson C, Sluis A. Impact of Endometriosis Diagnostic Delays on Healthcare Resource Utilization and Costs. *Adv Ther.* 2020;37(3):1087-1099.
14. UK All Party Parliamentary Group on Endometriosis Inquiry Report. *Endometriosis in the UK: time for change (October 2020)*. UK All Party Parliamentary Group on Endometriosis;2020.
15. Liu E, Nisenblat V, Farquhar C, et al. Urinary biomarkers for the non-invasive diagnosis of endometriosis. *Cochrane Database of Systematic Reviews.* 2015;2015(12).
16. Gupta D, Hull ML, Fraser I, et al. Endometrial biomarkers for the non-invasive diagnosis of endometriosis. *Cochrane Database Syst Rev.* 2016;4(4):Cd012165.
17. Nisenblat V, Prentice L, Bossuyt PM, Farquhar C, Hull ML, Johnson N. Combination of the non-invasive tests for the diagnosis of endometriosis. *Cochrane Database Syst Rev.* 2016;7(7):Cd012281.
18. May KE, Conduit-Hulbert SA, Villar J, Kirtley S, Kennedy SH, Becker CM. Peripheral biomarkers of endometriosis: A systematic review. *Human Reproduction Update.* 2010;16(6):651-674.

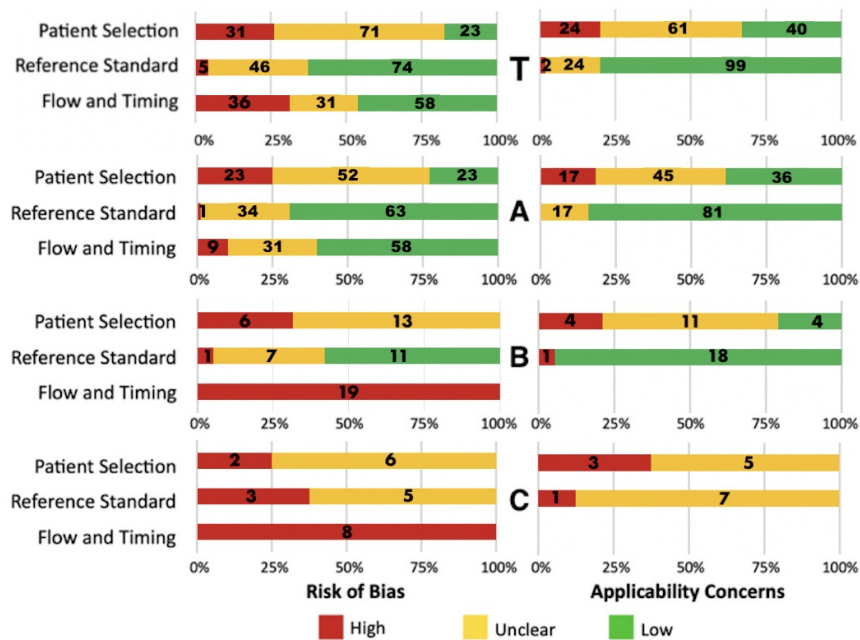
19. Nisenblat V, Prentice L, Bossuyt PMM, Farquhar C, Hull ML, Johnson N. Combination of the non-invasive tests for the diagnosis of endometriosis. *Cochrane Database of Systematic Reviews*.2016;2016(7).
20. Nisenblat V, Bossuyt PMM, Shaikh R, et al. Blood biomarkers for the non-invasive diagnosis of endometriosis. *Cochrane Database of Systematic Reviews*. 2016;2016(5).
21. 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(2):Cd009591.
22. Agarwal SK, Chapron C, Giudice LC, et al. Clinical diagnosis of endometriosis: a call to action. *Am J Obstet Gynecol*.2019;220(4):354.e351-354.e312.
23. Surrey E, Carter CM, Soliman AM, Khan S, DiBenedetti DB, Snabes MC. Patient-completed or symptom-based screening tools for endometriosis: a scoping review. *Arch Gynecol Obstet*. 2017;296(2):153-165.
24. Fauconnier A, Drioueche H, Huchon C, et al. Early identification of women with endometriosis by means of a simple patient-completed questionnaire screening tool: a diagnostic study. *Fertil Steril*.2021;116(6):1580-1589.
25. Bruggmann D, Elizabeth-Martinez A, Klingelhofer D, Quarcoo D, Jaque J, Groneberg D. Endometriosis and its global research architecture: an in-depth density-equalizing mapping analysis. *Bmc Womens Health*.2016;16.
26. Malvezzi H, Marengo EB, Podgaec S, Piccinato CA. Endometriosis: current challenges in modeling a multifactorial disease of unknown etiology. *J Transl Med*. 2020;18(1):311.
27. Bainton T JS, Raza A. A systematic review of the literature to determine the accuracy measures of various predictors of endometriosis.[https://www.crd.york.ac.uk/prospetro/display\\_record.php?RecordID=187543](https://www.crd.york.ac.uk/prospetro/display_record.php?RecordID=187543) . Published 2020. Accessed.
28. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj*.2021;372:n71.
29. *EndNote* [computer program]. Version EndNote X9. Philadelphia, PA: Clarivate; 2013.
30. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536.
31. StataCorp. 2017. Stata Statistical Software: Release 15. College Station TSL.
32. Hediger ML, Hartnett HJ, Louis GM. Association of endometriosis with body size and figure. *Fertil Steril*. 2005;84(5):1366-1374.
33. Paiva P, Lappas M, Barker G, Healey M. Using symptom scores, lifestyle measures and biochemical markers to create a test for endometriosis. *Journal of Endometriosis and Pelvic Pain Disorders*. 2014;6(3):135-143.
34. Backonja U, Hediger ML, Chen Z, et al. Beyond Body Mass Index: Using Anthropometric Measures and Body Composition Indicators to Assess Odds of an Endometriosis Diagnosis. *J Womens Health (Larchmt)*.2017;26(9):941-950.
35. Tanmahasamut P, Preukthanathorn R, Dangrat C. Serum interleukin 6 and cancer antigen 125 in the non-invasive diagnosis of endometriosis.*Journal of Endometriosis and Pelvic Pain Disorders*.2018;10(2):116-122.
36. Kim H, Choi YS, Kim JS, et al. Identification of Serum Biomarkers for Diagnosis of Endometriosis Using Multiplex Immunoassays.*Reproductive Sciences*. 2020;27(5):1139-1147.

37. Chapron C, Barakat H, Fritel X, Dubuisson JB, Bréart G, Fauconnier A. Presurgical diagnosis of posterior deep infiltrating endometriosis based on a standardized questionnaire. *Hum Reprod.*2005;20(2):507-513.
38. Kazanegra R, Zaritsky E, Lathi RB, Clopton P, Nezhat C. Diagnosis of stage I endometriosis: comparing visual inspection to histologic biopsy specimen. *J Minim Invasive Gynecol.* 2008;15(2):176-180.
39. Dunselman GA, Vermeulen N, Becker C, et al. ESHRE guideline: management of women with endometriosis. *Hum Reprod.*2014;29(3):400-412.
40. Lier MCI, Vlek SL, Ankersmit M, et al. Comparison of enhanced laparoscopic imaging techniques in endometriosis surgery: a diagnostic accuracy study. *Surg Endosc.* 2020;34(1):96-104.
41. Chen CX, Kwekkeboom KL, Ward SE. Self-report pain and symptom measures for primary dysmenorrhoea: a critical review. *Eur J Pain.* 2015;19(3):377-391.
42. Doggweiler R, Whitmore KE, Meijlink JM, et al. A standard for terminology in chronic pelvic pain syndromes: A report from the chronic pelvic pain working group of the international continence society. *Neurourol Urodyn.* 2017;36(4):984-1008.
43. Steingrimsdóttir Ó A, Landmark T, Macfarlane GJ, Nielsen CS. Defining chronic pain in epidemiological studies: a systematic review and meta-analysis. *Pain.* 2017;158(11):2092-2107.
44. Liu Y, Zhang W. Association between body mass index and endometriosis risk: a meta-analysis. *Oncotarget.*2017;8(29):46928-46936.
45. Yi KW, Shin JH, Park MS, Kim T, Kim SH, Hur JY. Association of body mass index with severity of endometriosis in Korean women. *Int J Gynaecol Obstet.* 2009;105(1):39-42.
46. Shafir AL, Wise LA, Palmer JR, et al. Validity of self-reported endometriosis: a comparison across four cohorts. *Hum Reprod.*2021;36(5):1268-1278.
47. Seracchioli R, Mabrouk M, Guerrini M, et al. Dyschezia and posterior deep infiltrating endometriosis: analysis of 360 cases. *J Minim Invasive Gynecol.* 2008;15(6):695-699.
48. Fauconnier A, Chapron C. Endometriosis and pelvic pain: epidemiological evidence of the relationship and implications. *Hum Reprod Update.* 2005;11(6):595-606.
49. May KE, Villar J, Kirtley S, Kennedy SH, Becker CM. Endometrial alterations in endometriosis: a systematic review of putative biomarkers. *Hum Reprod Update.* 2011;17(5):637-653.
50. Rogozińska E, Khan K. Grading evidence from test accuracy studies: what makes it challenging compared with the grading of effectiveness studies? *Evid Based Med.* 2017;22(3):81-84.
51. Eskenazi B, Warner M, Bonsignore L, Olive D, Samuels S, Vercellini P. Validation study of nonsurgical diagnosis of endometriosis. *Fertil Steril.* 2001;76(5):929-935.
52. Madhvani K, Fernandez-Felix BM, Zamora J, Carpenter T, Khan KS. Personalising the risk of conversion from laparoscopic to open hysterectomy in benign conditions: Development and external validation of risk prediction models. *Bjog.* 2021.
53. Hirsch M, Duffy JMN, Davis CJ, et al. Diagnostic accuracy of cancer antigen 125 for endometriosis: a systematic review and meta-analysis. *BJOG Int J Obstet Gynaecol.* 2016;123(11):1761-1768.
54. Nnoaham KE, Hummelshoj L, Kennedy SH, Jenkinson C, Zondervan KT. Developing symptom-based predictive models of endometriosis as a clinical screening tool: results from a multicenter study. *Fertil Steril.* 2012;98(3):692-701.e695.
55. Chapron C, Lafay-Pillet MC, Santulli P, et al. A new validated screening method for endometriosis diagnosis based on patient questionnaires. *EClinicalMedicine.* 2022;44:101263.

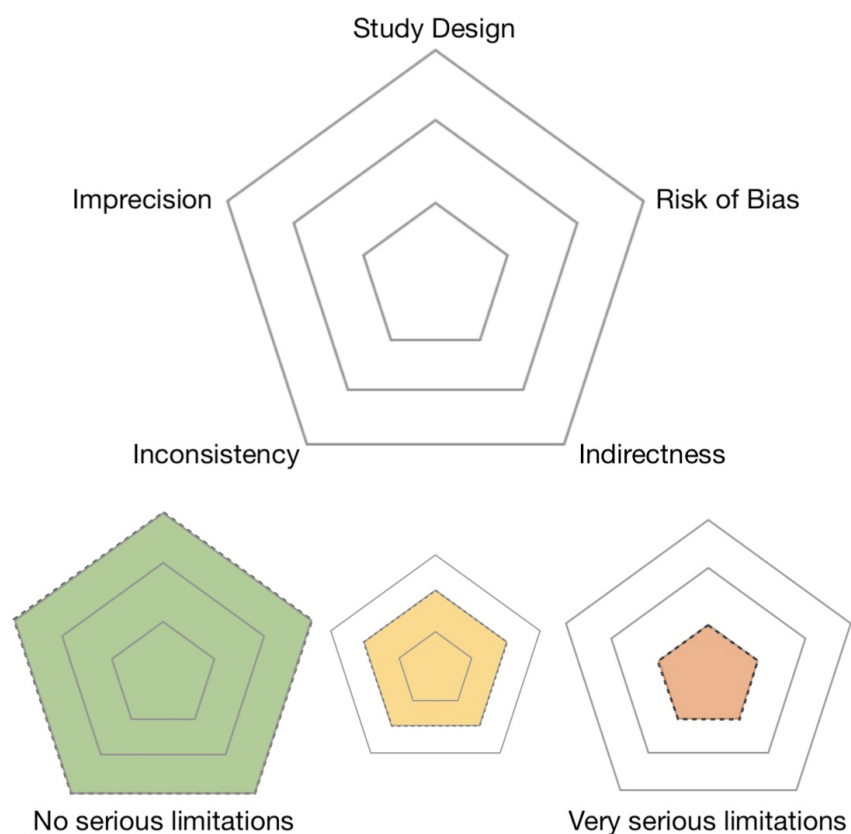
## Hosted file

Table 1.docx available at <https://authorea.com/users/517752/articles/592177-diagnosis-of-pelvic-endometriosis-a-systematic-review-and-accuracy-meta-analysis-of-non-invasive-tests>



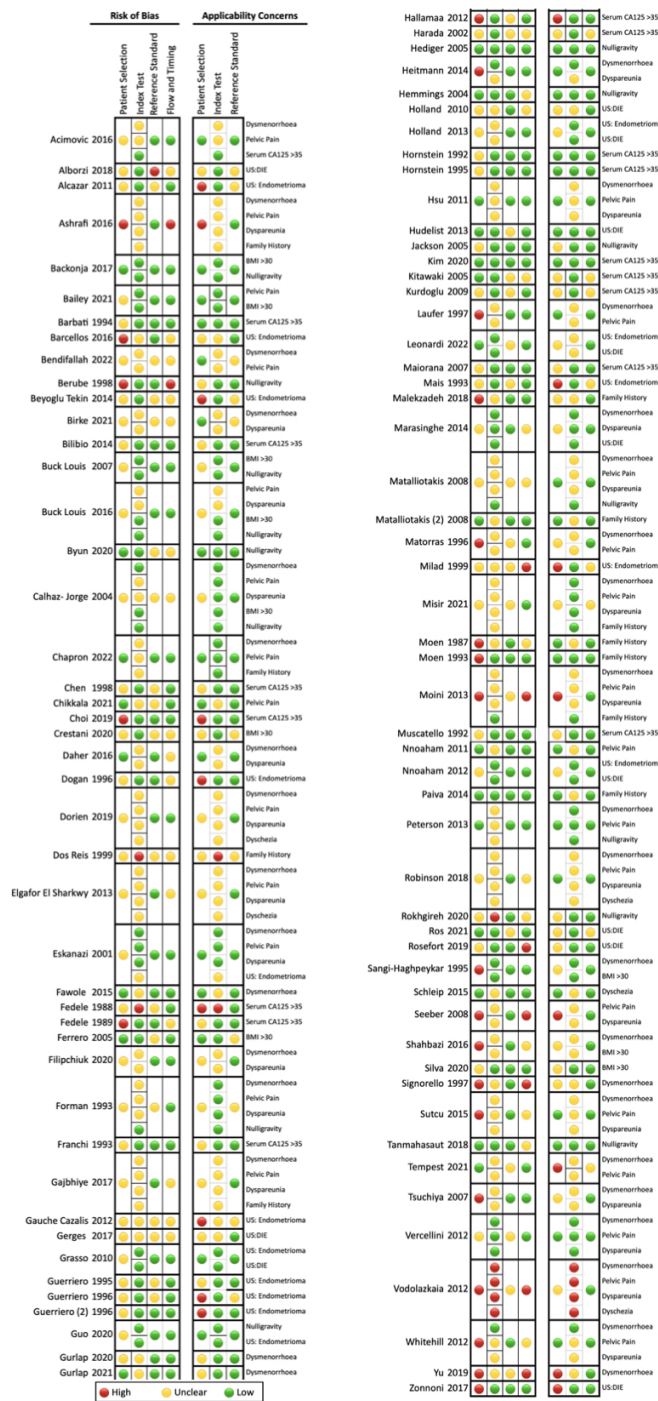


Outcome	No. of Studies (Participants)	Summary Accuracy Measures					Assessment of Confidence		
		Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	Positive LR (95% CI)	Negative LR (95% CI)	Sensitivity Pentagon	Specificity Pentagon	
A	Dysmenorrhoea	35 (10,405)	69.8% (60.7-77.6)	52.5% (41.7-63.1)	2.56 (1.99-3.29)	1.47 (1.27-1.70)	0.57 (0.49-0.68)		
	Pelvic Pain	29 (10,477)	46.9% (33.0-61.2)	74.3% (63.8-82.7)	2.56 (1.73-3.74)	1.83 (1.44-2.33)	0.71 (0.59-0.87)		
	Dyschezia	6 (1453)	19.2% (10.5-32.5)	89.1% (78.1-95.4)	2.05 (1.36-3.10)	1.85 (1.26-2.72)	0.90 (0.84-0.97)		
	Dyspareunia	25 (6832)	36.3% (29.4-43.8)	81.1% (72.2-87.7)	2.45 (1.71-3.52)	1.93 (1.42-2.61)	0.78 (0.72-0.85)		
	Nulligravidity	15 (8187)	55.3% (46.3-64.0)	61.8% (53.0-69.9)	2.01 (1.61-2.50)	1.45 (1.27-1.65)	0.72 (0.64-0.81)		
	BMI ≥30kg/m²	10 (3750)	10.9% (4.58-23.7)	75.1% (63.2-84.1)	0.37 (0.19-0.68)	0.44 (0.25-0.77)	1.23 (1.09-1.30)		
	Family History	11 (5346)	9.25% (7.12-11.9)	98.5% (97.1-99.2)	6.79 (4.08-11.3)	6.25 (3.75-10.4)	0.92 (0.90-0.94)		
	TVUSS: Endometrioma	16 (6819)	77.2% (61.0-87.9)	96.4% (93.0-98.2)	91.2 (44.0-189)	21.6 (11.7-39.8)	0.24 (0.13-0.42)		
	TVUSS: DIE	12 (3264)	86.5% (61.2-96.3)	80.2% (52.3-93.8)	26.1 (9.28-73.5)	4.39 (1.74-11.1)	0.17 (0.06-0.48)		
	Serum CA-125 > 35U/mL	17 (2091)	55.8% (37.5-72.7)	92.7% (82.5-97.1)	16.0 (8.09-31.7)	7.63 (3.73-15.6)	0.48 (0.33-0.69)		
B	Dysmenorrhoea	9 (7718)	76.1% (60.9-86.6)	69.2% (60.0-77.0)	7.13 (3.34-15.2)	2.47 (1.83-3.34)	0.35 (0.20-0.59)		
	Pelvic Pain	4 (1601)	62.2% (45.7-76.3)	88.6% (72.7-95.8)	12.8 (5.89-28.1)	5.46 (2.46-12.1)	0.43 (0.31-0.60)		
	Dyschezia	3* (766)	57% (48-65)	92% (85-99)	2.05 (xx-xx)	7.13 (3.2-65)	0.47 (0.35-0.84)		
	Dyspareunia	4 (2039)	51.3% (46.6-59.0)	90.1% (76.7-96.2)	9.65 (3.91-23.8)	5.21 (2.20-12.3)	0.54 (0.50-0.58)		
	Nulligravidity	8 (48,177)	43.1% (35.1-57.5)	74.9% (62.4-84.2)	2.55 (1.93-3.38)	1.84 (1.43-2.35)	0.72 (0.65-0.80)		
	BMI ≥30kg/m²	7 (69,266)	13.9% (9.70-19.7)	87.3% (82.3-90.9)	1.11 (0.95-1.29)	1.09 (0.96-1.25)	0.99 (0.97-1.01)		
	Family History	5 (44,442)	12.4% (7.98-18.9)	97.8% (93.1-99.3)	6.23 (2.08-18.7)	5.58 (1.91-16.3)	0.89 (0.85-0.95)		
C	Dysmenorrhoea	5 (133,883)	70.1% (41.0-88.8)	66.7% (38.1-86.7)	4.70 (1.85-11.9)	2.11 (1.16-3.82)	0.45 (0.24-0.85)		
	Pelvic Pain	6 (109,376)	44.4% (23.5-67.6)	90.3% (78.0-96.1)	7.47 (4.75-11.8)	4.59 (2.88-7.33)	0.61 (0.44-0.87)		
	Dyschezia	3* (56,042)	38% (0-80)	82% (53-100)	2.78 (xx-xx)	2.11 (0-xx)	0.76 (0.2-1.89)		
	Dyspareunia	5 (105,309)	44.4% (18.6-73.6)	84.2% (55.8-95.8)	4.26 (2.76-6.57)	2.81 (1.57-5.05)	0.66 (0.47-0.93)		
	Nulligravidity	2* (27,819)	61% (60-62)	43% (43-44)	1.18 (xx-xx)	1.07 (1.05-1.11)	0.91 (0.86-0.93)		
	BMI ≥30kg/m²	2* (47,791)	10% (10-11)	92% (91-92)	1.28 (xx-xx)	1.25 (0.99-1.11)	0.98 (0.97-0.99)		



Search strategy for MEDLINE (via PubMed); Embase; Scopus; and Web of Science databases. Searches conducted from database conception to present in July 2020 and updates in September 2022 with an additional filter to exclude results prior to July 2022.

1. (Endometriosis) AND (dysmenorrhoea) AND (sensitivity OR specificity OR accuracy OR likelihood OR prediction)
2. (Endometriosis) AND (dyspareunia) AND (sensitivity OR specificity OR accuracy OR likelihood OR prediction)
3. (Endometriosis) AND (dyschezia) AND (sensitivity OR specificity OR accuracy OR likelihood OR prediction)
4. (Endometriosis) AND (pelvic pain) AND (sensitivity OR specificity OR accuracy OR likelihood OR prediction)
5. (Endometriosis) AND (subfertility OR sub fertility OR sub-fertility OR infertility) AND (sensitivity OR specificity OR accuracy OR likelihood OR prediction)
6. (Endometriosis) AND (family history) AND (sensitivity OR specificity OR accuracy OR likelihood OR prediction)
7. (Endometriosis) AND (ultrasound) AND (sensitivity OR specificity OR accuracy OR likelihood OR prediction)
8. (Endometriosis) AND (CA 125 or CA-125 or cancer antigen 125) AND (sensitivity OR specificity OR accuracy OR likelihood OR prediction)
9. (Endometriosis) AND (BMI OR body mass index) AND (sensitivity OR specificity OR accuracy OR likelihood OR prediction)
10. (Endometriosis) AND (epidemiology)





		Risk of Bias				Applicability Concerns			
		Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard	
Ashrafi 2021									Dyspareunia
									Dysmenorrhoea
Bernays 2020									Pelvic Pain
									Serum CA125 >35
Burghaus 2011									BMI >30
									Nulligravidity
Cramer 1986									Dysmenorrhoea
Darrow 1994									Dysmenorrhoea
									Nulligravidity
Fauconnier 2021									Dysmenorrhoea
									Pelvic Pain
									Dyschezia
									Dyspareunia
Han 1994									Dysmenorrhoea
									Nulligravidity
Kashima 2004									Family History
Lin 2022									Dysmenorrhoea
									Pelvic Pain
Mamdouh 2011									BMI >30
									Nulligravidity
									Family History
Marino 2008									Nulligravidity
Mollazadeh 2019									Dysmenorrhoea
									Dyspareunia
									Family History
Prescott 2016									BMI >30
Ricci 2020									Dysmenorrhoea
									Pelvic Pain
									Dyspareunia
									Dyschezia
Schneider 2020									Dysmenorrhoea
									Pelvic Pain
									BMI >30

		Risk of Bias				Applicability Concerns			
		Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard	
Ballard 2008	Dysmenorrhoea								
	Pelvic Pain								
	Dyspareunia								
	Dyschezia								
	BMI >30								
Flores 2008	Nulligravidity								
	Dysmenorrhoea								
	Pelvic Pain								
Fuldeore 2017	Dyspareunia								
	Dysmenorrhoea								
	Pelvic Pain								
Marino 2009	Dyspareunia								
	Nulligravidity								
Markham 2019	Dysmenorrhoea								
	Pelvic Pain								
	Dyspareunia								
Moen 2019	Dyschezia								
	Pelvic Pain								
Saha 1997	Pelvic Pain								
	BMI >30								
Singh 2020	Dysmenorrhoea								
	Pelvic Pain								
	Dyspareunia								
	Dyschezia								

● High
 ● Unclear
 ● Low

